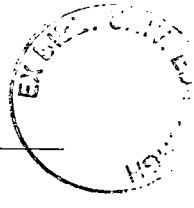
Migration, habitat choice and assortative mating in a *Bombina* hybrid zone

Timothy H. Vines

Thesis presented for the degree of
Doctor of Philosophy
at the University of Edinburgh

2002



A *B. variegata* male (top) in amplexus with a *B. bombina*-like female, from site 200.8 in the Cojocna valley, Romania.

ABSTRACT

How do the differences between diverging taxa contribute to their reproductive isolation? Natural hybrid zones are an ideal opportunity to test hypotheses on the role of ecological, behavioural and postzygotic mechanisms in preventing gene flow. In this thesis I present data from a previously unstudied hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* on the Transylvanian plain of Romania.

The spatial arrangement of toads in Romania resembles a mosaic hybrid zone, with populations of *B. bombina* in isolated ponds surrounded by the more numerous *B. variegata* living in puddles. This structure is in marked contrast with the narrow clinal hybrid zones found in Poland, Croatia and the Ukraine. To understand what might cause these differences, I make a detailed comparison between the zone in Romania with that in Pešćenica, Croatia, and conclude that the distribution of habitat types in the two areas may differ considerably. Furthermore, I found that toads in Romania had a much stronger preference for their respective habitats than in Croatia.

However, the habitat preference is not strong enough to prevent the immigration of pure *B. bombina* individuals into intermediate habitats containing *B. variegata*-like hybrids. This appears to be responsible for the high level of introgression at neutral markers throughout the *B. variegata* population. I used the observed linkage disequilibrium to quantify this migration, and then assessed how much selection against immigrant *B. bombina* alleles is required to conteract this movement. I concluded that adaptive differences could be maintained with plausible levels of selection, but that the differences in marker allele frequencies are probably collapsing.

I then measured the strength of assortative mating and breeding site preference in the Romanian study area, as this could be an important force acting against immigrant *B. bombina* in the intermediate habitats. I found that there was no assortment, but the parents were significantly more *B. variegata*-like than the local adults, implying strong breeding site preference. The latter might be a strong force in maintaining adaptive differences in this mosaic hybrid zone.

Declaration

The work presented in this thesis is my own, apart from where otherwise acknowledged in the text, and the thesis has been written by myself.

Acknowledgments

I would like to thank my supervisor Nick Barton for all his encouragement over the last four years and the immense amount I have learnt from him in that time. His comments on various chapters of this thesis have led to many improvements, and I am also grateful that he always had a moment at the right times.

I also owe an huge debt of gratitude to Beate Nürnberger of LMU Munich: her dedication to making *Bombina* fieldwork actually work, her patience and wizardry through many long months in the lab in Munich, and her enthusiasm for an occasionally recalcitrant study organism. She has also contributed many ideas to all aspects of this thesis, and my conversations with her have greatly enhanced my understanding of *Bombina* in particular and evolution in general.

My fieldwork in Romania would have been impossible without the assistance of Ioan Ghira with everything from charming old ladies to paddling around ponds in the dark, and I am indebted to him for his time, effort and generosity. Gyongyver Mara was an excellent field and city companion. I also thank Lucretia Boonca and Mariuca for their help whilst I was in Fînate.

Sonja Köhler has also contributed enormously to this thesis, both in terms of data and with fieldwork, and the entire study would be greatly diminished without her cooperation. She is also partly responsible for the synthesis of stalled Trabants and cricket commentary. I would also like to thank Martina Thiel for her contribution to the genotyping, Bruni Forg-Brey and Gabi Praetzel for answering my endless questions in either English or German, and Elke Maier for leaving dead fish on the circle line of the U-Bahn. Alexey Yanchukov very kindly gave me access to his data on the *Bombina* hybrid zone in the Ukraine.

I am grateful to Toby Johnson and Stuart Baird for many useful tips and for showing me the magic of maximum likelihood. Loeske Kruuk has also made many helpful comments on both empirical and theoretical matters over the last four years. I would also like to thank Tim Sands for sharing his hard-won mark-recapture data with me, and Andy Gardner, Lucy Odling-Smee, Sarah Reece and Katrina Lythgoe for many stimulating discussions.

Finally, I am indebted to my parents for always knowing that it would work in the end, and for their support and belief throughout my increasingly long stint in academia.

Table of contents

Chapter 1:
Reproductive isolation and hybrid zones

Section 1.1 Overview

An understanding of the mechanisms of reproductive isolation is essential to understanding the diversity of life on earth. In this thesis, I quantify the effects of two such mechanisms in the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* in Romania. I focus on how the environment might influence the structure and dynamics of hybrid zones in Chapters 2 and 3, and quantify the strength of assortative mating between the two taxa in Chapters 4 and 5. Without data on the relative importance of these processes in nature, we cannot hope to arrive at a general theory for the process of speciation.

Progress on this theory has been slow for a number of reasons. Firstly, populations may differ in a huge number of ways, and the effect of each difference depends on the environment. Making quantitative predictions about how a particular difference could reduce the rate of interbreeding is therefore far from easy. The second problem is that differences arise over long timescales in nature, and many of the processes involved are impossible to recreate in an experimental setting (Rice and Hostert 1993). Thirdly, differences arise together, and accumulate more rapidly as isolation increases, and it may be impossible to distinguish their individual contributions. Lastly, one cannot tell from examining the differences between well isolated populations how individual traits contributed to their original separation, as some traits may have diverged after they were no longer capable of interbreeding. I will use this chapter to briefly discuss the ways in which populations can evolve reproductive isolation (Section 1.2), and the role of hybrid zones in quantifying the contribution of different traits (Sections 1.3 & 1.4). Finally, I will introduce the hybrid zone between the fire bellied toads *Bombina bombina* and *B. variegata*, and discuss the previous work on these taxa (Sections 1.5 & 1.6).

Here, I will use the term 'reproductive isolation' to refer to any heritable mechanism that keeps alleles from different populations separate. The traditional definition includes only traits that reduce hybrid fitness directly, but divergence at ecological or sexual traits can also generate reproductive isolation (e.g. Rundle et al. 2000). In fact, recent advances in molecular biology have lead to the identification of an increasing number of the genes responsible for hybrid inviability (e.g. Ting et al. 1998), and it appears that many of the genes responsible for hybrid inviability or sterility diverged as a result of sexual selection (Orr and Presgraves 2000). There is a much greater role for epistasis in hybrid inviability and sterility than in other forms of reproductive isolation, as the former necessarily involves interactions between genes from different populations (Orr 2001). Overall, it is becoming apparent that we cannot understand the evolution of reproductive isolation without first understanding how phenotypic divergence proceeds at the genetic level.

This emphasis on reproductive isolation as the definitive feature of species stems from the biological species concept (Mayr 1942), and although alternatives exist (Mallet 1995), none is as widely accepted among biologists. However, like many important words in biology or ecology, 'species' suffers from excessive attempts at definition, and these so muddy its meaning that it becomes hard to use it in a general sense without inviting trivial disagreement. I will therefore refrain from using it wherever possible for the remainder of this thesis. The root of the problem lies with our inbuilt tendency to put a name to recurring patterns in the natural world, and biological organisms do appear to fall into discrete categories containing similar creatures (Hey 2001). However, as mentioned above, these 'species' do not all arise in the same way, and the process does not occur instantaneously. Why then insist that all known organisms belong to a narrowly defined state known as a 'species'?

Far more progress would be made by concentrating on the different ways in which populations diverge, and the consequences of that divergence for their reproductive isolation. One could argue that this emphasis merely shifts the onus of definition onto 'population', which has to date remained a less constrictive term. However, it is possible to study the isolation between any two groups of organisms one cares to define, and hence our understanding of reproductive isolation is independent of the study unit. The study of some types of population will naturally be more informative than others: for example, those that show significant divergence but can still interbreed can demonstrate how individual traits contribute to the total isolation. This is why hybrid zones are regarded as ideal 'evolutionary laboratories' for the study of reproductive isolation (Harrison 1990). For the purposes of this thesis, I will refer to the two principal study organisms, the fire-bellied toads *B. bombina* and *B. variegata* as members of different taxa, rather than assign them to a particular (and arbitrary) taxonomic level.

Section 1.2 The principal modes of reproductive isolation

As mentioned above, differences between taxa may manifest in a near-infinite number of ways, but ultimately they must stem from differences at the DNA level if they are to be of any evolutionary consequence. For example, if different growth forms for two alternative habitats are entirely induced by the environment, there is then no fitness cost to migration between habitats, and hence no reproductive isolation. So, given that the differences between populations are due to divergence at the genomic level, can we identify the traits that contribute most often to reproductive isolation? Broadly speaking, divergence between populations at any trait can only contribute to their isolation in a limited number of ways, and these can be divided into those that act before the formation of a hybrid zygote, known as prezygotic isolating mechanisms, and after zygote formation, known as postzygotic mechanisms (Dobzhansky 1937). A second aspect that must be considered is the geographic separation of the populations, as close contact will limit the types of difference that can arise and contribute to reproductive isolation.

Prezygotic isolation

Prezygotic mechanisms reduce the frequency of matings between populations, or lower the probability of fertilisation once mating has begun. There are three broad types in the first category: ecological, temporal and sexual isolating mechanisms. Ecological isolation arises when the two populations occupy different habitats, and rarely (if ever) come into contact, which implies that the organisms dispersal range is smaller than the typical habitat patch size, or that they show an active preference for one or other habitat (e.g. *Acyrthosiphon* aphids on alfalfa and clover: Via 1999). The occupation of alternative habitats entails selective differences, and hence the populations will adapt to their local environment, presumably at the expense of

their fitness in the other habitat. However, the populations need not spend their entire lifecycle in their respective habitats: isolation can arise if they merely use different breeding habitats. Populations can thus be in living together almost all year round and still be reproductively isolated, because organisms in mating condition never meet. A logical extension of this is temporal isolation: if the two populations are not ecologically isolated but instead breed at different times of the year, the frequency of matings between them will be greatly reduced. The classic example of this is in *Bufo americanus* and *Bufo fowleri*: they are sympatric and use similar waterbodies, but the former breeds early in the year, and the latter breeds late. Interestingly, in some places, a mid-season breeding population consisting of both taxa and their hybrids can be found, indicating that temporal isolation plays a major role in keeping the two taxa distinct (Blair 1941).

Sexual isolating mechanisms can act when potential mates meet in a reproductive context, and the form they take depends on the mating system involved. The most commonly studied taxa are those in which one sex (usually females) chooses from a set of candidate mates using a secondary sexual trait as a guide. If there is genetic variation for both the preference and the trait, a shift in the preferred phenotype can elicit a rapid response from the male population, and system will move to a new equilibrium (see Panhuis et al. 2001). This shift might occur when the trait differs in detectability between habitats (Schluter and Price 1993); for example, female sticklebacks reared in clear or cloudy water differ in their ability to perceive the colour red, and male sticklebacks in either population are coloured to match their preferences (Boughman 2001). The matching of display traits may cause females to reject the courtship of males from other populations because they do not show her preferred traits. (Note that this may not be the case if male display traits exploit

sensory biases in the females: e.g. Ryan and Rand 1990.) I discuss divergence and sexual selection in greater depth in Section 4.1.

In other mating systems, the opportunities for the evolution of sexual isolating mechanisms are probably reduced. In monogamy, for example, there is much lower variance in reproductive success, and hence less opportunity for rapid evolution of the mate choice system. In fact, the apparently much lower rates of speciation in monogamous compared to lekking bird species is often taken as evidence that sexual selection drives divergence (Barraclough et al. 1995). The potential for such divergence is presumably also much lower in taxa with male-male competition, as selection will always favour increases in size and aggression traits in this case.

The second category of prezygotic isolation involves mechanisms that operate once a mating has begun, but before fertilisation takes place. There are two approximate types, mechanical and gametic. The former includes morphological features, whereas the latter are more biochemical, although examples straddling both types are possible, such as male genital features that remove the sperm of other males prior to copulation. Mechanical isolating mechanisms arise because there is an ongoing conflict between male and female sexual organs, and hence mutations giving either sex an advantage will quickly rise to fixation. In general, males will be favoured by mutations that force an increase in mating frequency, increase short term egg production and decrease the likelihood of females remating subsequently with other males, even if these decrease the fitness of females. An adverse effect on female fitness would mean that any mutations that reverse or dampen these effects are favoured in females, and a perpetual antagonistic cycle results. Populations can thus quickly diverge in reproductive traits, and reproductive isolation ensues (Parker and Partridge 1998). There has been a long running debate over the significance of the diversity of insect genital morphology and its role in

reproductive isolation (reviewed in Arnqvist 1997). However, there is increasing support from studies of waterstriders that morphological coevolution does occur, and this might generate mechanical isolation (Arnqvist and Rowe 2002).

The same sexual conflict that drives mechanical isolation also affects the biochemical interactions between sperm from different males or between the sperm and the female, and these can generate gametic isolation. Mutations bringing an advantage to either sex in these conflicts will be strongly selected for, and generate the same rapid coevolutionary cycles as above. For example, male ejaculate often contains substances that reduce the success of foreign sperm, and males from the same population will be selected to develop counter-measures. These substances may reduce female fitness (e.g. *Drosophila*: Clark et al. 1999). Males from other populations will lack defences against both male and female counter-measures, and will thus be out-competed for fertilisations. This mechanism has been documented in many taxa, including the ground crickets *Allonemobius fasciatus* and *A. socius* (Howard et al. 1998), and between the Louisiana irises *Iris fulva* and *I. brevicaulis* (Emms et al. 1996). Conflicts between the sperm and the female can also result in isolation, as sperm from other populations may lack the necessary traits to survive in the female, or to bind with the egg (see Swanson and Vacquier 2002 for a review). There is evidence that polyandrous insect clades contain more taxa than monandrous ones (Arnqvist et al. 2000). This trend presumably arises because there are greater opportunities for the coevolution of genital morphology, sperm competition or reproductive tract conflicts in polyandrous populations, and these have resulted in the rapid generation of reproductive isolation between these populations.

Postzygotic isolation

Postzygotic isolation mechanisms act after the zygote has been formed, and these determine the fate of interpopulation hybrids. Firstly, there may be alleles present in only one population that are incompatible with alleles fixed in the other population. Since fixing such alleles will involve reduced population fitness when they are at the same locus, they are more likely to appear at different loci. When these incompatible alleles are brought together in the same individual their interaction reduces the viability or fertility of the hybrid. Their potential for generating reproductive isolation was recognised by Dobzhansky (1937), and they have received much recent theoretical attention (Orr 1995, Orr and Turelli 2001). It is currently unclear what kinds of genes are involved in this sort of incompatibility, and whether or not they have arisen as a by-product of adaptive evolution, as there are only two well studied examples at present. One involves the uncontrolled overproduction of melanomas in *Xiphophorus* fish due to negative epistasis between melanin production and its regulatory mechanism (Schartl 1995). The second involves a homeobox gene expressed in *Drosophila* testes (Ting et al. 1998). However, there is other evidence from *Drosophila* that these incompatibilities affect male function more often than female, and thus may evolve faster in males. This higher evolutionary rate in males implies that hybrid inviability or sterility alleles are fixed during adaptation in response to sexual selection (Orr and Presgraves 2000).

There is traditionally a distinction between incompatibilities that reduce the fitness of the hybrid irrespective of its environment and the incompatibilities that arise because the organism carries conflicting alleles adapting it to two distinct habitats (Orr 2001). The latter causes hybrid unfitness because the hybrid is inferior in both parental habitats. This phenomenon has been demonstrated in hybrids between

surface and lake-bed feeding sticklebacks in British Columbia (Hatfield and Schluter 1999), and is expected to occur when hybrids are phenotypically intermediate between their parents, and hence fall between ecological niches. This mechanism requires that there be existing ecological divergence, and that the parents are also unfit in each others habitats (see prezygotic isolation above).

Similarly, if there has been divergence in mating system between the populations, hybrid males are unlikely to be chosen by females from either population. However, hybrid females may have an intermediate mate preference, and, depending on the genetic basis of mate choice, either actively prefer hybrid males or be willing to accept any sort of mate. This mechanism often accompanies habitat divergence, as the phenotypic changes entail a concomitant shift in the preferred male phenotype; for example, mate choice has been shown to reduce the reproductive success of hybrids between the surface and lake-bed feeding sticklebacks mentioned above (Vamosi and Schluter 1999). In other mating systems, sexual selection against hybrids might be less common, although there has been very little work on this issue to date. As mentioned above, there is less opportunity for rapid divergence, and thus there may not be as much scope for discrimination against hybrids.

Where can divergence take place?

If the two populations are geographically isolated (i.e. allopatric), and hence cannot exchange individuals, then differences between them can accumulate through either selection or drift. As the proportion of each population in contact with the other increases, the forces driving divergence must become stronger to ensure that the differences are maintained in the face of migration and recombination. Roughly speaking, there are two ways in which populations can be in contact: either they exchange migrants across a shared boundary (parapatry), or their ranges overlap

considerably (sympatry). However, there is clearly room for intermediate states when their respective habitat patches are interspersed (see Chapter 2).

Ecological isolation is unlikely if the two populations are in identical environments, as the probability of a spontaneous jump from one niche into another is low (Coyne et al. 1997). However, any differences in the ecological communities on either side of the barrier can lead to rapid evolution towards different phenotypic optima (Schluter 2001). For example, variability in the seed-producing plant community on different Galapagos islands has driven rapid divergence in bill morphology in the local finches (Grant 1986). This process can also act when there is opportunity for individuals to move between populations. In this case, selection must be strong in relation to the mixing effects of migration and recombination, and divergence is unlikely to be preserved in sympatry unless there is also some sort of assortative mating by habitat (Via 2001). In this scenario, both immigrants and their hybrid offspring will be unfit, as they both carry alleles adapting them to the other habitat. There is growing evidence that ecological isolation can evolve in sympatry and parapatry given the right conditions, but nonetheless these will always be more restrictive than the conditions for divergence in allopatry.

When more than one gene is responsible for maintaining adaptation to one or other habitat in the face of migration, recombination between the genes must be reduced for the set to be established (Felsenstein 1981). Much recent work has focused on the chromosomal location of these genes, and there is evidence that the differences between taxa are caused by genes either in close physical linkage, or together in a chromosomal inversion. A good example of the former is in *Acyrthosiphon* aphids: the genes for performance on either alfalfa or clover are closely linked to those determining host preference (Hawthorne and Via 2001). There is also evidence from *Drosophila* (Noor et al. 2001) and sunflowers (Rieseberg et al. 1999) that

genes causing hybrid inviability are mainly located on inversions. A recent theoretical study by Navarro and Barton (2002) showed that inversions can delay the spread of new mutations from the other population long enough for local incompatible alleles to rise to fixation, indicating that inversions can play a vital role in protecting sets of diverging genes from recombination.

In the case of sexual isolation, it is clear that the rapid coevolution of male and female traits along different trajectories can lead to isolation, and that this can proceed in allopatry. The evidence that this process has contributed to taxonomic diversity is still under debate, but the required phylogenetic data are becoming steadily more available (Panhuis et al. 2001). However, it not clear that this type of isolation can arise when the two populations are diverging in sympatry (Turelli et al. 2001). If there is a pre-existing ecological difference, different signal types may be favoured in either environment (Schluter and Price 1993). The rate of accumulation of sexual isolation will then be determined by the relationship between selection for different female preferences, migration, and the rate of recombination between the genes responsible for female and male traits. When there is no ecological divergence, sympatric speciation in response to sexual selection seems much less likely. All of the theoretical models developed so far (Turner and Burrows 1995; Payne and Krakauer 1997; Higashi et al. 1999) rely on restrictive circumstances unlikely to arise in natural populations (Turelli et al. 2001). Furthermore, even in cichlid fish, in which taxa differ only in male secondary sexual characters, there is no evidence that this divergence took place whilst the populations were in sympatry (Turner et al. 2001).

Lastly, it has been shown theoretically that mechanical and gametic isolation can evolve rapidly in allopatry (Gavrilets 2000). However, it is hard to imagine how two populations can spontaneously evolve along different trajectories when there is

migration between them, as mutations giving an advantage to one sex will be selected for everywhere. Even if migration is low, advantageous alleles will still cross the barrier and rise to fixation in both populations (Barton 1986), making it unlikely that incompatible phenotypes will become fixed in either population.

In summary, gene flow between the populations restricts divergence by rapid coevolution because advantageous alleles can spread rapidly even when dispersal is limited (Barton 1986). By contrast, if there is a fitness cost to moving into the habitat of the other population, ecological adaptation can continue to accumulate even at high migration rates.

Section 1.3 An introduction to hybrid zones

In principle, it is possible to examine the differences between any two populations of organisms, but if they are fully reproductively isolated from each other the data will be less useful. This is because one cannot distinguish those differences that contributed to their original isolation from those that arose after they could no longer interbreed. Indeed, if the two have never come back into contact since their initial split, there is no way of knowing what effect any of the differences might have, or whether or not they can hybridise at all. Therefore, the most useful populations to study are those that can not only interbreed (i.e. can exchange migrants) but also actually produce hybrids. In these situations, it is possible to infer the level of prezygotic isolation between the populations from the relative rates of inter- versus intrapopulation matings. The strength of postzygotic isolation can also be measured from the fitness of any hybrids produced. Since hybridisation can rapidly erode divergence, it is more likely that hybridising populations will persist when they only share a border and are not fully overlapping. This situation can give rise to many kinds of hybrid zones, depending on the biology of the organisms and the geography of the area.

Origins of hybrid zones

How might hybrid zones arise? The simplest circumstance is the meeting of two populations that had previously been diverging in allopatry: when their ranges come into contact, they begin hybridising. It is also possible that there was no initial geographic split between the two populations, and they somehow evolved partial reproductive isolation when they was still migration between them (see previous section). Given enough time, these two scenarios cannot be distinguished (Endler 1977), because the level of differentiation between the populations will increasingly

depend on the rate of gene flow and not the conditions under which the zone was formed. In any case, unless a very recent origin can be confirmed, it is likely that a pair of populations have spent some time both geographically separated and hybridising across a hybrid zone, such that the signature of their original divergence is erased. Such repeated contact following expansion and subsequent retraction could easily be driven by climatic cycles, such as ice ages: populations diverge for a time in refugia while the area between the populations is covered by ice, and re-expand back into contact when it retreats (Hewitt 2001).

The fate of alleles introgressed from the other population as the ice returns depends on whether the descendants of the individuals from the hybrid zone make it back to the refugia, or are lost *en route*. In general, it is impossible to reliably reconstruct the history of a hybrid zone from present day observations. Instead, workers have concentrated on quantifying the current levels of migration and selection in hybrid zones, and using these to understand how various traits contribute to reproductive isolation.

Fates of hybrid zones

The next aspect to consider is what happens after the initial formation of the zone: will it persist or will it disappear? There are roughly three identifiable fates for hybrid zones. Firstly, if there is insufficient reproductive isolation to counter the mixing effect of migration between populations and hybridisation, all the differences between them will be lost. The initially sharp steps in allele frequencies will flatten out, and most individuals will eventually contain alleles from both populations. On a locus by locus basis, the rate at which this happens depends on the balance between selection on that locus and other loci nearby on the chromosome, and the effective rate of recombination breaking them apart (Barton and Bengtsson 1986). The total barrier to gene flow becomes stronger as more loci

are involved, and as they are spread more evenly through the genome, as then selection tends to act on the genome as a single unit. As selection becomes weaker relative to recombination, loci become more independent, and the selection on each locus individually then determines its rate of introgression (Barton and Bengtsson 1986, Barton and Shpak 2000). Divergence at a locus can be maintained indefinitely by strong selection even when migration is high (see Chapter 3), but without ongoing immigration from pure populations all the differences not involved in reproductive isolation will be eroded, even those close on the chromosome to the selected loci (Barton and Bengtsson 1986). It thus might be argued that hybrid zones are collapsing at all but a few loci, as recombination must eventually break up genomes once hybridisation has started. Under this view, the observations of 'stable' hybrid zones (Barton and Hewitt 1985) are simply those that are collapsing too slowly to be detected over the timescale of a few decades.

Alternatively, if the barrier to gene flow imposed by the hybrid zone is sufficiently strong, the two populations could continue to diverge at most, if not all loci, and the zone may eventually disappear because interbreeding no longer occurs. This is a more plausible outcome when there is ongoing selection towards different optima on either side of the zone (e.g. towards greater ecological or sexual divergence), simply because there is increasingly strong selection against alleles from the other population. The threshold conditions at which the divergence in the two populations switches from slowly collapsing (above) to increasing presumably depend on the proportion of each population in the proximity of the hybrid zone, the strength of divergent selection on either side and the barrier to gene flow imposed by the zone. Naively, this threshold would occur when the effective rate of migration through the zone (i.e. Nm; see Hartl and Clark 1997) is around 1, above this ($Nm > 1$), differences would disappear, and below ($Nm < 1$) they would increase. One could

test this hypothesis by comparing the relative levels of allozyme and neutral marker differentiation between hybridising populations: divergence at marker loci should either be collapsing or increasing faster than allozymes, as the latter may be constrained by selection. This issue is also related to existing theory on the conditions required for parapatric speciation (Slatkin 1973; Turelli et al. 2001).

The final potential fate for a hybrid zone is the attractive idea that selection against hybrids can lead to the evolution of greater mate discrimination, and hence stronger reproductive isolation. The key difference is that the necessary divergence to complete reproductive isolation then comes from inside the zone itself, and not from the populations on either side. This mechanism was originally proposed by Dobzhansky (1940) and the process came to be known as 'reinforcement'. Simply put, pure individuals from either population suffer a fitness cost from producing unfit hybrid offspring, and there will thus be selection for greater discrimination in mate choice within the zone of contact. This can evolve if there is standing genetic variation for mating preferences or if a new mutation arises. Since this additional isolating mechanism further reduces hybridisation, the two populations may become completely reproductively isolated, and the hybrid zone will disappear. Its acceptance as a plausible mechanism has waxed and waned, principally because there are doubts that the field data actually demonstrate reinforcement (Butlin 1989). In contrast, theoretical work has found that the process can proceed under a wider variety of conditions than previously thought (Kirkpatrick and Servedio 1999). Furthermore, Servedio (2001) showed that direct selection on female preferences and on post-mating prezygotic incompatibilities could also lead to reinforcement without the need for reduced hybrid fitness. There is clearly a great deal more to understand about how secondary contact can contribute the

reproductive isolation between populations, and it may be that once a definitive test for the signature of reinforcement is found more examples will emerge (Noor 1999).

Section 1.4 Inferences from hybrid zones

Migration plays a fundamental role in the structuring of hybrid zones (see Chapter 3), but under certain circumstances this role is diminished, and the structure then depends only on the strength of selection. For example, if there is a gradient in the optimum value of some selected trait, then the trait should mirror the optimum provided the gradient is very broad compared to the dispersal distance. This may result in a band in which hybrids have higher fitness than the parental populations on either side, as their intermediate phenotypes can match the trait optimum more closely. This is known as bounded hybrid superiority (Moore 1977). However, unless there is a direct environment-phenotype correspondence, it is possible that a steady gradient in some trait is due to the collapse of a originally narrow cline at the boundary between two populations (e.g. *Orchelimum* crickets: Shapiro 1998). In this case selection against introgression of foreign alleles at those loci was insufficient to prevent their steady spread through both populations (see 'Fates of hybrid zones' above). However, when environmental gradients are sharper, or selection stronger, the relationship between selection and the rate of movement becomes more important. For clarity, I use migration to refer to movement between discrete populations, and dispersal to mean movement across a continuous landscape (c.f. Barton and Bengtsson 1986). Only dispersal includes units of physical distance. There are three distinguishable types of dispersal-dependent hybrid zones, and I discuss these below.

Tension zones and ecotones

When the isolation between populations is principally in the form of postzygotic allelic incompatibilities, such that selection is against hybrids irrespective of the environment, hybridisation will lead to a steep cline. These will tend to minimise

their length, and are hence referred to as 'tension zones' (Key 1968). Many hybrid zones appear to be of this type, as they are narrow in relation to dispersal, and the clines at a wide range of traits are often of similar width (Barton and Hewitt 1985). This implies that selection is acting against the genome as a whole, as otherwise the width of the cline for each trait will reflect their individual selection coefficients (see above). However, since allelic incompatibilities accumulate rather slowly, and then only in allopatry (Orr 1995, but see Navarro and Barton 2002), it is hard to believe that there has been no ecological divergence between the populations, and that this plays no role in the hybrid zone. In fact, steep clines identical to those produced by selection against hybrids can also be produced by two populations adapted to alternate environments meeting at an ecotone (Kruuk et al. 1999b). Furthermore, tension zones will tend to migrate towards areas of low population density, or towards a local barrier to dispersal (Barton and Gale 1993). Since even ecologically identical taxa are unlikely to move freely between alternative habitats, it is entirely possible that a steep cline at an ecotone is maintained entirely by selection against hybrids.

In reality, determining whether a steep cline is driven by selection against hybrids or against alleles in the wrong habitat is only relevant for populations that differ at a small number of loci. A good example of such a cline comes from the butterfly *Heliconius erato*, where hybrid zones exist between colour pattern races divergent at only three loci, although in this case selection is against rare phenotypes (Mallet and Barton 1989). When there is more substantial divergence, both types of selection are likely to contribute. What matters more in this case is the relationship between recombination and total selection, because only strong total selection can hold clines in different traits together and significantly slow introgression at loci not involved in maintaining the cline. It is worth noting that these steep clinal hybrid

zones are only possible when the area covered by both populations is very large relative to their dispersal distance, so that the cline forms between two extensive habitat patches. The movement of animals can then be approximated by diffusion, greatly simplifying the analysis of the zone (Haldane 1948, Nagylaki 1975).

The analytical techniques based around this insight have grown through four stages. 1) Slatkin (1973) showed that clines at a single locus have a width proportional to $\sigma/\sqrt{s}$, where σ is the standard deviation of the distance between the parent and offspring (a measure of dispersal), and s is the strength of selection. This result is fairly independent of the type of selection (Barton and Gale 1993). However, this relationship seems to predict unreasonably weak selection for natural clines, perhaps because dispersal is underestimated in the wild (Barton and Hewitt 1985).

2) The next stage takes into account the interaction between different loci, as loci within hybrid zones rarely segregate independently. Instead, dispersal between areas with different allele frequencies at several loci creates associations between loci, and these are measured as linkage disequilibrium (D). These are halved with every generation of random mating, and therefore high levels of D must be maintained by ongoing dispersal. Linkage disequilibrium can be therefore be used to estimate dispersal, given some assumptions about the allele frequencies in the source populations and the rate of recombination between the loci (Barton and Gale 1993). The same process that generates D between loci also creates covariance between quantitative traits, and hence clines for unconnected traits are often closely concordant through a hybrid zone.

3) If many different loci are contributing to the isolation of the two populations, the total selection may be greater than the sum of the individual selection coefficients. This generates a barrier (B) to the flow of alleles between the populations, and is measured by the ratio of the step in allele frequency at the centre of the cline (on a

logit scale) and the gradient at the edges. This value has units equivalent to the length of unimpeded habitat the allele must pass through to reach the other population (Barton and Gale 1993). Given an estimate of dispersal and the selection for or against the locus in question, an expression for the expected time taken (T) to pass the barrier can be derived ($T \sim (B / \sigma)^2$ for a neutral locus).

4) Dispersal into the hybrid zone by pure individuals brings in unbroken sets of alleles characteristic of that population. After a few generations of hybridisation and recombination, the hybrid genome will contain blocks of genes from one population or the other, and as recombination continues, these blocks get smaller and smaller (Baird 1995). This approach can be used to age a new species formed by hybridisation: Ungerer et al. (1998) found that a *Helianthus* sunflower species was formed by hybridisation between 20 to 60 generations ago.

Overall, the analysis of steep clines provides excellent insights into how genomes of diverged populations interact, although the contribution of each type of isolating mechanism cannot be directly inferred (Kruuk et al. 1999b). However, the inferences of total selection are invaluable in interpreting direct measurements of isolating mechanisms in the zone (e.g. MacCallum 1994, Kruuk and Gilchrist 1997).

Mosaic hybrid zones

When populations differ ecologically, it is rare that their habitat patches are distributed on either side of a sharp boundary, especially when their divergence has involved specialisation onto spatially discrete resources, such as host plants. In these situations, the contact zone can take on a range of structures depending on the balance of migration between habitats, hybrid fitness and the history of the area (see Chapter 2). These are collectively known as mosaic hybrid zones after their patchwork structure (Harrison and Rand 1989). Since the area of contact between

the two populations is much higher in a mosaic, there is greater potential for hybridisation and the erosion of population differences. If the mosaic takes up a large area between the parental populations, the dynamics in the middle may become independent from the edge, and a hybrid swarm may result. Alternatively, the greater contact and the patchy structure may make the likelihood of reinforcement higher than in a cline (Cain et al. 1999). Understanding how the spatial interspersion of habitats can affect the fate of population divergence is important in a number of settings, not least the evolution of reproductive isolation in sympatry, and mosaic hybrid zones provide an ideal setting to test this type of hypothesis. I will discuss mosaic hybrid zones at greater length in Chapters 2 and 3.

Section 1.5 The study taxa: *Bombina bombina* and *B. variegata*

This thesis concentrates on the hybridisation between the fire-bellied toads *Bombina bombina* and *B. variegata* in Romania. These two taxa are capable of producing fertile hybrids even though they are divergent for many traits, and it is this divergence that makes them such a valuable study system. The *Bombina* genus is in the Discoglossidae family, which has recently been supported by molecular data (Hay et al. 1995). I refer to *Bombina* as being 'toads' throughout this thesis; this is actually misleading as 'toad' in the strict sense applies only to amphibians from the genus *Bufo*. There is usually only a common name for *Bombina* in countries where it occurs naturally. For example, in Germany they are known as 'Unken', distinct from Kröten (true toads) and Frösche (frogs in the genus *Rana*).

In appearance, both *Bombina* taxa are small, light brown toad-like amphibians, with distinctive warning coloration on their ventral side. In *B. bombina* this consists of small red spots often less than 2 mm across on a black background (Fig. 1.1a). In *B. variegata*, the pattern consists of large yellow patches on a grey background, sometimes forming a continuous area from throat to the upper leg (Fig. 1.1b). Hybrids have a pattern intermediate between these two, and the extent to which the spots are joined together has been used to classify hybrids for many years (Michałowski and Madej 1969; Gollmann 1984). This 'spot score' is highly concordant with genetic marker loci in both Poland (Szymura and Barton 1986, 1991) and Croatia (Nürnberger et al. 1995). Furthermore, each belly pattern is unique, and it can thus be used to identify recaptures.

Figure 1.1a: *B. bombina* (fire bellied toad)

Figure 1.1b: *B. variegata* (yellow bellied toad)

Figure 1.2a: Typical *B. bombina* habitat

Figure 1.2b: Typical *B. variegata* habitat

Phenotypic differences

The majority of phenotypic differences between the two taxa appear to adapt them to different habitats. *B. bombina* is predominantly found in large ponds and other permanent waterbodies (Fig. 1.2a). They avoid sites with large fish communities, presumably because of the high rate of predation on both adults and juveniles. *B. variegata* is found in temporary sites, such as tractor wheel ruts, flushes and puddles (Fig. 1.2b), but it avoids anything more than 1 m deep (pers. obs.). The phenotypic differences associated with this ecological divergence include:

1) *B. variegata* has a thicker skin than *B. bombina*, presumably so it can better resist dehydration when moving between sites (Nürnberger et al. 1995). The lungs of the former are also more vascularised, which is also associated with preventing water loss (Czopkowa and Czopek 1955).

2) *B. variegata* is generally larger, with relatively longer legs, and has a more robust skeleton than *B. bombina* (Nürnberger et al. 1995), again perhaps as an adaptation to travel between frequently drying sites.

3) The two taxa differ in mating call. *B. bombina* has vocal sacs (Boulenger 1886), and can therefore make a much louder call. Their calls also differ in cycle length, pulse duration and fundamental frequency (Lörcher 1969; Schneider and Eichelberg 1974; Sanderson et al. 1992). *B. bombina* forms large choruses that can be heard many hundreds of metres away, whereas *B. variegata* mating aggregations are only audible over shorter distances, even though they may contain many individuals (pers. obs.).

4) There is some evidence that they have different mating systems (Lörcher 1969). *B. bombina* males appear to guard small territories on the pond surface from which they call to attract females. *B. variegata* probably practices scramble polygyny,

where a group of males will call together from a small waterbody, and then fight over any females that move into the site.

5) The eggs of *B. variegata* are larger than those of *B. bombina* (Rafinska 1991; Nürnberger et al. 1995). This may be an adaptation to permit faster development, as the hatching tadpole is larger. *B. variegata* eggs are laid in clutches of around 20, whereas *B. bombina* lays larger batches of 60-100 smaller eggs. Although nothing is known about the total number of eggs females of each taxa are able to produce, it is possible that *B. variegata* females lay several batches in different location to spread the risk of being caught in a drying puddle. This is supported by the genotypes of eggs from Croatia, where several batches sampled on the same day had the same parental genotypes (Nürnberger et al. 2002a), although these were all within the same waterbody.

6) The tadpoles of *B. variegata* develop faster and are more active (Kruuk and Gilchrist 1997; Vorndran et al. 2002). This is probably in response to the need to escape the site before it dries up. Conversely, the lower activity of *B. bombina* tadpoles may help them avoid visual predators in the pond communities.

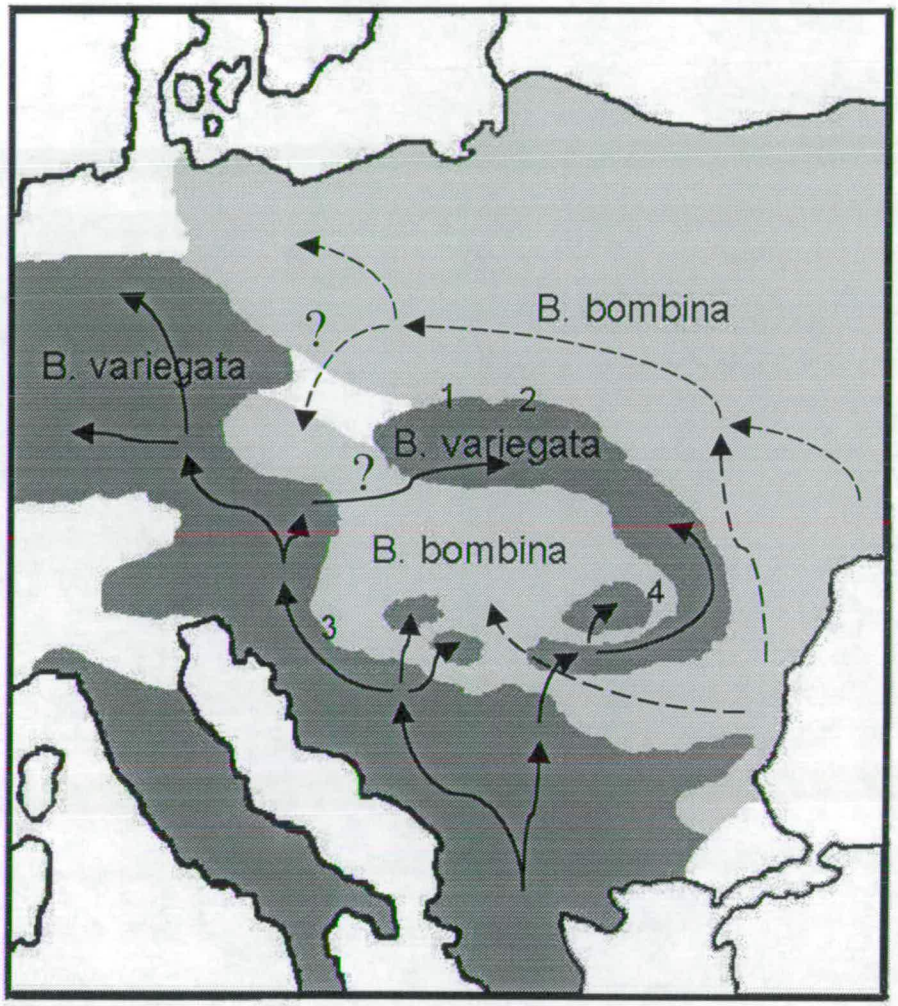
Biogeography of Bombina

The two taxa are generally thought to have diverged in allopatry enforced by the Pliocene ice age 5 to 1.8 million years ago. *B. variegata* appears to have occupied the glacial refugia in Italy and the northern Balkans, whereas *B. bombina* could have retreated into the area around the Black and Caspian Seas (Arntzen 1978; my Fig. 1.3). In the interglacial periods they presumably expanded their ranges into what is now central Europe, and may have occasionally hybridised. In the current interglacial, their distribution suggests that *B. bombina* migrated north west up the Danube onto the Hungarian plain, and a second branch took a more northerly route

around the eastern edge of the Carpathian mountains along the Oresti and Vistula rivers. One subgroup of *B. variegata* seems to have moved north out of the Balkans and then eastwards into the long hook of the Carpathian chain, while another went north-west towards the eastern Alps and the rest of Western Europe. The presence of isolated populations of *B.variegata* on higher ground in the Hungarian Plain suggests that they once occupied lowland areas, and both Arntzen (1978) and Szymura (1993) suggest that they were forced onto higher ground by competition with the encroaching *B. bombina*. However, since they are fairly distinct ecologically, it seems more reasonable that the rising temperatures as the interglacial progressed made puddle habitats in the plains too short-lived for successful reproduction by *B. variegata*, and they migrated to the cooler uplands.

This subdivision into refugia is also supported by molecular evidence. Based on an albumin molecular clock, *B. bombina* and *B. variegata* probably diverged 2 million years ago (Maxson and Szymura 1984), but other proteins suggest this figure is closer to 6.8 million (Szymura 1983). A survey of allozyme loci found that the groups within each taxa are also supported (Szymura 1993): *B. variegata* from the western part of its range in Croatia and from the Carpathians are fairly closely related (Nei's D_N = 0.16; Nei 1972), but these two are distinct from *B. variegata* from Italy and the Balkans (D_N = 0.31). These latter two groups are also different (D_N= 0.24), presumably because the Po river divides their ranges (Fig. 1.3). *B. bombina* are also separated into a Danube Basin and a northern group, although this split is poorly supported (Fig. 10-2 in Szymura 1993).

Figure 1.3: The distribution of *Bombina* in Europe. Studied hybrid zones are at 1. Cracow; 2. Przemysl; 3. Pescenica; 4. Apahida. The arrows show the presumed migration route of *B. bombina* (dashed lines) and *B. variegata* (solid line) after the last ice age.



Section 1.6 Previous work in *Bombina* hybrid zones

Hybridisation in the wild between *B. bombina* and *B. variegata* was first hypothesised by Méhelÿ in 1892, but confirmation only came with molecular techniques decades later. It appears that, despite the wide range of differences between the two taxa, they can still interbreed and produce fertile hybrids wherever their ranges meet. In the field, it is possible to find a very wide range of hybrid phenotypes, indicating that extensive backcrossing has taken place. Over a very wide scale, the ranges of the two taxa almost never overlap, and the two contact zones in Poland (Szymura and Barton 1986, 1991) and Croatia (MacCallum et al. 1998) that were studied intensively showed a narrow clinal transition (<10 km) between them. I briefly summarise the main results from these two areas below.

Poland

Two transects were studied in Poland, one in Cracow and one 200 km away in Przemysl (1 and 2 respectively in Figure 1.3) by Szymura and Barton (1986, 1991). Six independent allozyme loci with alleles diagnostic of one taxon or the other were developed for 6000 individuals collected from 110 waterbodies. In both places they found a stepped cline around 6 km wide, with long tails of introgression on either side. All the loci changed in concordance through both zones, and this suggests that total selection was strong enough hold clines together in the face of recombination within the zone. Most sites were in Hardy Weinberg equilibrium, suggesting that there was effectively random mating despite the differences in mating call between the two taxa (Sanderson et al. 1992). There were, however, strong associations between all pairs of loci, and these are likely to be generated by dispersal. The maximum value of linkage disequilibrium was D = 0.055 (from maximum

likelihood, 2-unit support limits 0.0375 to 0.0725), which could be generated by dispersal of 0.89 km/generation in Cracow and 0.98 km/gen in Przemysl.

These narrow stepped clines could be maintained by a reduction in hybrid fitness of 0.58 (support limits 0.54-0.68), created by a barrier to gene flow at a neutral locus between the two populations equivalent to 51 km (22-81) of unimpeded habitat. There was also direct evidence for selection against hybrids, in the form of developmental abnormalities in tadpoles and adults (Madej 1965; Czaja 1980), and higher embryonic mortality (Koteja 1984; Szymura and Barton 1986). The contact zone also appeared to be stable, as the current position of the cline matched data from samples taken 33 and 55 years earlier (Szymura and Barton 1986). In total, the data suggest that these two contact zones match the definition of a tension zone: selection is mainly against hybrids, there were no strong habitat associations, and it appears to be maintained by a balance between dispersal and selection.

Another study worthy of mention is by Sanderson et al. (1992), which mapped changes in three components of the male call across the Cracow transect. They found that all three traits (pulse duration, fundamental frequency and cycle length) differed significantly between pure populations of either taxa, but only cycle length is sufficiently different to be diagnostic. The clines for each trait are centred with the clines at the allozyme loci, and have a similar width. This suggests that there is not sufficient selection on call traits to shift them out of position relative to each other or the allozyme loci. Interestingly, they found a significant effect of water temperature on both pulse duration and cycle length, but not fundamental frequency, although Schneider and Eichelberg (1974) did find an effect for the latter. If mating call does play a role in mate choice in *Bombina*, the temperature of the waterbody the male calls from might profoundly affect all three call

components, and hence affect the judgement of any listening female. I will consider this issue in greater detail in Chapter 4.

Croatia

The *Bombina* hybrid zone in Pešćenica, Croatia (MacCallum 1994; MacCallum et al. 1998) superficially resembles those in Poland, in that there is a steep stepped cline with long tails of introgression separating the two taxa. However, there was significant heterozygote deficit towards the centre of the zone, and there were marked differences in allozyme allele frequencies between adjacent sites. There also was even more linkage disequilibrium here than in Poland: the maximum D was 0.139, compared to 0.055 in Cracow and Przemysl. All these are driven by an active habitat preference in *B. bombina* for ponds and *B. variegata* for puddles, and the presence of a mixture of different habitat types in the centre of the zone. The heterozygote deficit probably arises because there is migration between adjacent ponds and puddles differing in allele frequency (i.e. the Wahlund effect), and the linkage disequilibrium may be higher partly because the habitat-genotype associations reduce the effective rate of mixing in the centre of the zone, and this slows the break up of parental allele combinations. I make a detailed comparison between the structure of Pešćenica transect and my study area in Apahida, Romania in Chapters 2 and 3.

There have been several other studies on animals from the Pešćenica hybrid zone that are also of interest. Nürnberger et al. (1995) carried out a large scale breeding experiment with Pešćenica toads. They scored a wide range of adults for belly pattern, skin thickness, skeletal characteristics and mating call. Crosses were then made either within populations or between putatively pure individuals from either taxa, and these offspring were scored for egg size, development time, larval survival and metamorph survival. They found that the clines in adult traits were

concordant with each other and with allozyme allele frequencies, but the clines in egg size and development time were shifted in different directions. The latter may be due to strong directional selection on these traits in the zone. This allows them to partially escape the congealing effect of linkage disequilibria and form a cline that follows their own peculiar selection pressures more precisely. Kruuk et al. (1999a) showed that there is also significant selection against some hybrid families in a lab-based rearing experiment, although the strength of ecological selection against alleles in the wrong habitat remains unknown.

Section 1.7 Thesis aims

This thesis will explore two main issues with data from a newly studied *Bombina* hybrid zone in Romania. Firstly, the spatial arrangement of genotypes is much patchier in Romania than in Croatia or Poland, and the situation is closer to a mosaic hybrid zone. In Chapter 2, I describe the distribution of *Bombina* around Apahida, Romania, and make a detailed comparison with the equivalent data from Pešćenica in Croatia. I then discuss why hybrid zones between the same taxa might differ in structure. In Chapter 3 I use the observed levels of heterozygote deficit and linkage disequilibrium for Apahida to infer the rate of migration between different habitat types, and then consider the strength of selection necessary to counteract this migration. I then make predictions about the fate of neutral and selected divergence in the zone.

The second section will attempt to measure assortative mating between the two taxa, which is an important component of reproductive isolation. This has traditionally been extremely difficult to measure in the field, and has only been attempted as a pilot study before (Nürnberger et al. 2002a). Chapter 4 describes the samples of eggs and adults, and considers some of the problems of analysing this type of data. Chapter 5 uses a maximum likelihood model to estimate several assortative mating parameters for these samples, and compares this approach to a second independent method using only the egg batches. Lastly, in Chapter 6 I discuss the wider implications of both threads of this thesis in the context of hybrid zones and speciation research.

Chapter 2:
The Apahida mosaic hybrid zone

Section 2.1 Introduction

This chapter describes the distribution of *Bombina* around Apahida, Romania. The structure of the contact zone here resembles a mosaic hybrid zone, with pure populations of *B. bombina* living in widely scattered ponds, and an extensive *B. variegata*-like population in the surrounding temporary sites. Why should the hybrid zone in Apahida differ so much from the smooth clines found in Croatia and Poland? In this chapter I attempt to quantify how Apahida differs from the zone in Pešćenica (in Croatia), firstly in terms of the spatial distribution of animals, and secondly in terms of the strength of habitat preference. I then explore how these differences arose, and how they can affect hybrid zones.

The data and analyses presented in this and the next chapter are taken from a recently submitted paper entitled 'The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*' by T. H. Vines, S. C. Köhler, M. Thiel, I. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton and B. Nürnberger.

The structure of hybrid zones

Why might hybrid zones differ in structure? When there are no ecological differences between the populations, or they are adapted onto two extensive habitat types that meet at an ecotone, it is probable that a stepped cline will form upon secondary contact (e.g. Fig. 2.1a,b). The width and stability of this cline then depends on the balance between selection against hybrids or against alleles in the wrong habitat, and the rate of dispersal (Barton and Gale 1993). Furthermore, different traits may separate out into clines whose width reflects their individual selection pressures. This depends on how fast parental genomes move into the zone relative to the rate of recombination. If selection is weak, there will be more hybrid

generations, and hence greater opportunity for recombination between sets of loci. A good example of this phenomenon is found in the hybrid zone between *Chorthippus parallelus* and *C. erythropus* in the Pyrenees, where the clines for stridulatory peg number, wing length ratio, song characters and a nuclear organising region all vary in width and position (see Hewitt 1993).

Figure 2.1a: A tension zone formed by selection against hybrids. The boxes represent habitat patches, in this case the environment is the same everywhere.

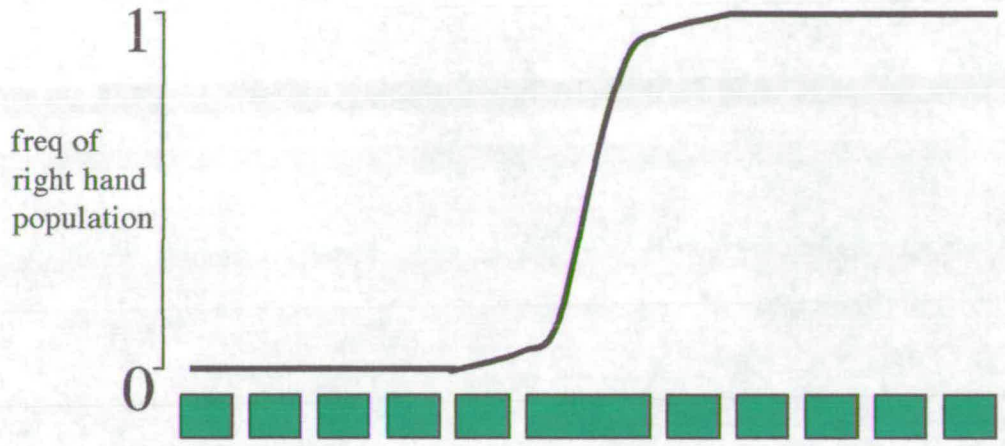


Figure 2.1b: Selection against alleles in the wrong environment at an ecotone. The boxes represent habitat patches, blue is one habitat type and green another.

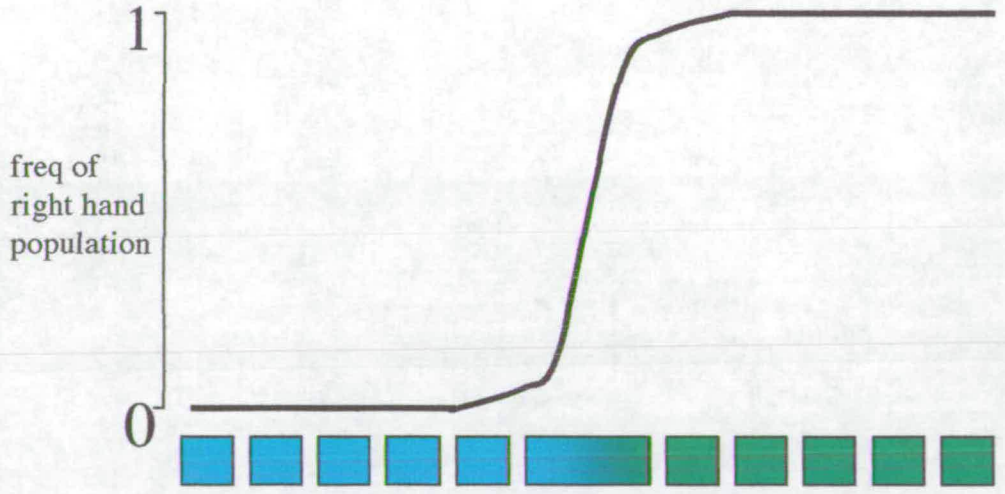


Figure 2.1c: Effect of increasing interspersion of habitats. The boxes represent overlapping habitat types, the blue and green lines show the allele frequency in blue and green habitat patches respectively.

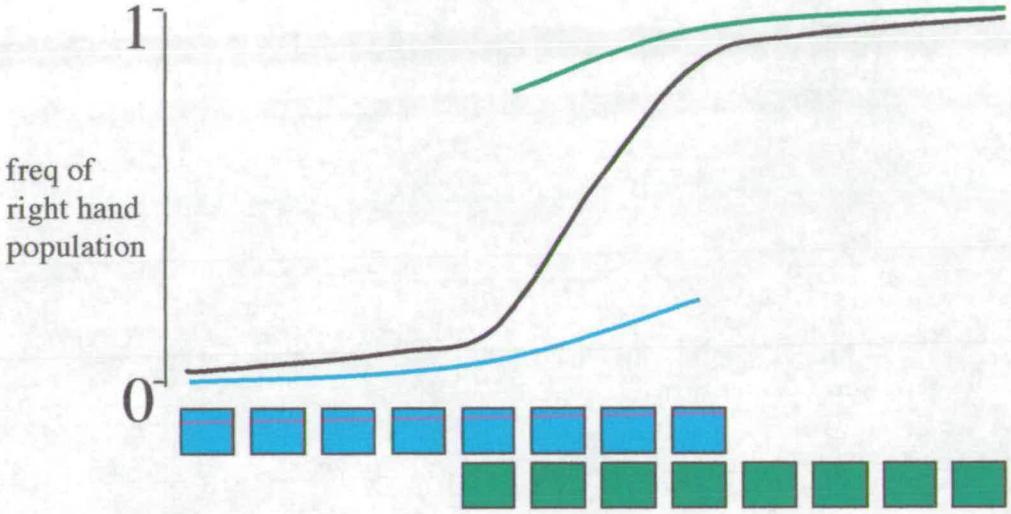
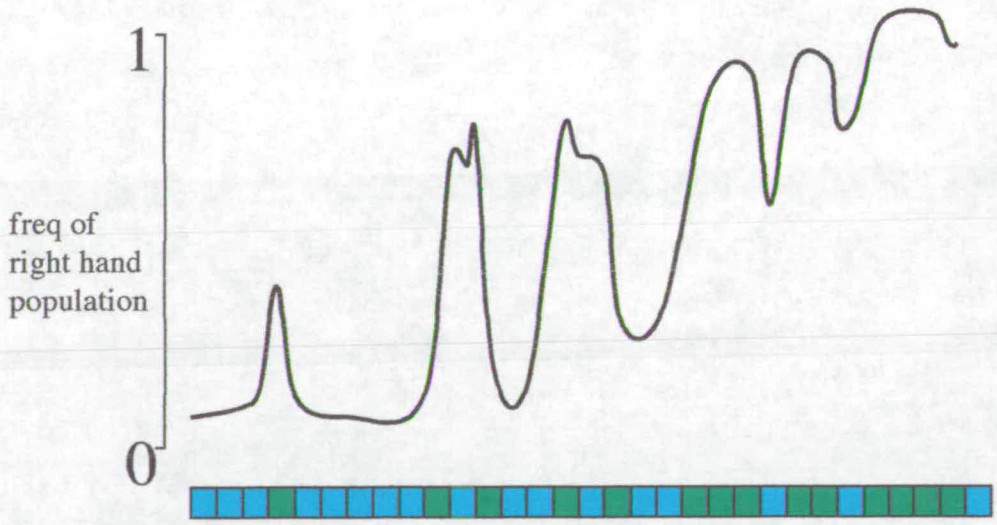


Figure 2.1d: A transect through a mosaic hybrid zone. The boxes represent highly interspersed habitat types



In general, it seems likely that taxa that have been diverging in allopatry will have slightly different ecological requirements, as their environments will inevitably have differed in many ways. Moreover, it is also likely that either habitat type will not cover a large continuous area, but instead be distributed in patches. If the two populations meet in an area where their habitat patches are slightly interspersed then a smooth cline may still form if there is sufficient migration between habitat types. The patches located furthest into the range of the other taxa will be the most introgressed, as they receive the greatest proportion of their immigrants from the other population (Fig. 2.1c). This appears to be the case for *Bombina* in Pešćenica, where the expected allele frequency in the ponds or puddles changes steadily through the zone (MacCallum 1994; my Fig. 2.2).

When the populations ranges overlap more completely, such that the distance between the pure populations on either side is considerably greater than the dispersal range, a much more complex pattern can develop (Fig. 2.1d). Since the population size on a single patch will be much smaller than the extensive pure populations on either side of a narrow hybrid zone, migration will have a much larger effect on allele frequencies, as the ratio of immigrants to locals will be much higher. If individuals move at random between patches, the isolating mechanisms must be strong to prevent the swamping of local adaptation. If isolation occurs mainly prior to mating (e.g. via mate choice), then the migrational load on the local population will be much lower, as postzygotic mechanisms necessarily involve wasted matings or the production of unfit hybrids. However, if there is an active habitat preference, migration between habitat types is reduced, and the within-patch isolating mechanisms need not be so strong. Mate choice and habitat preference are thus two sides of the same coin, as isolation occurs within patches in the former, and between them in the latter.

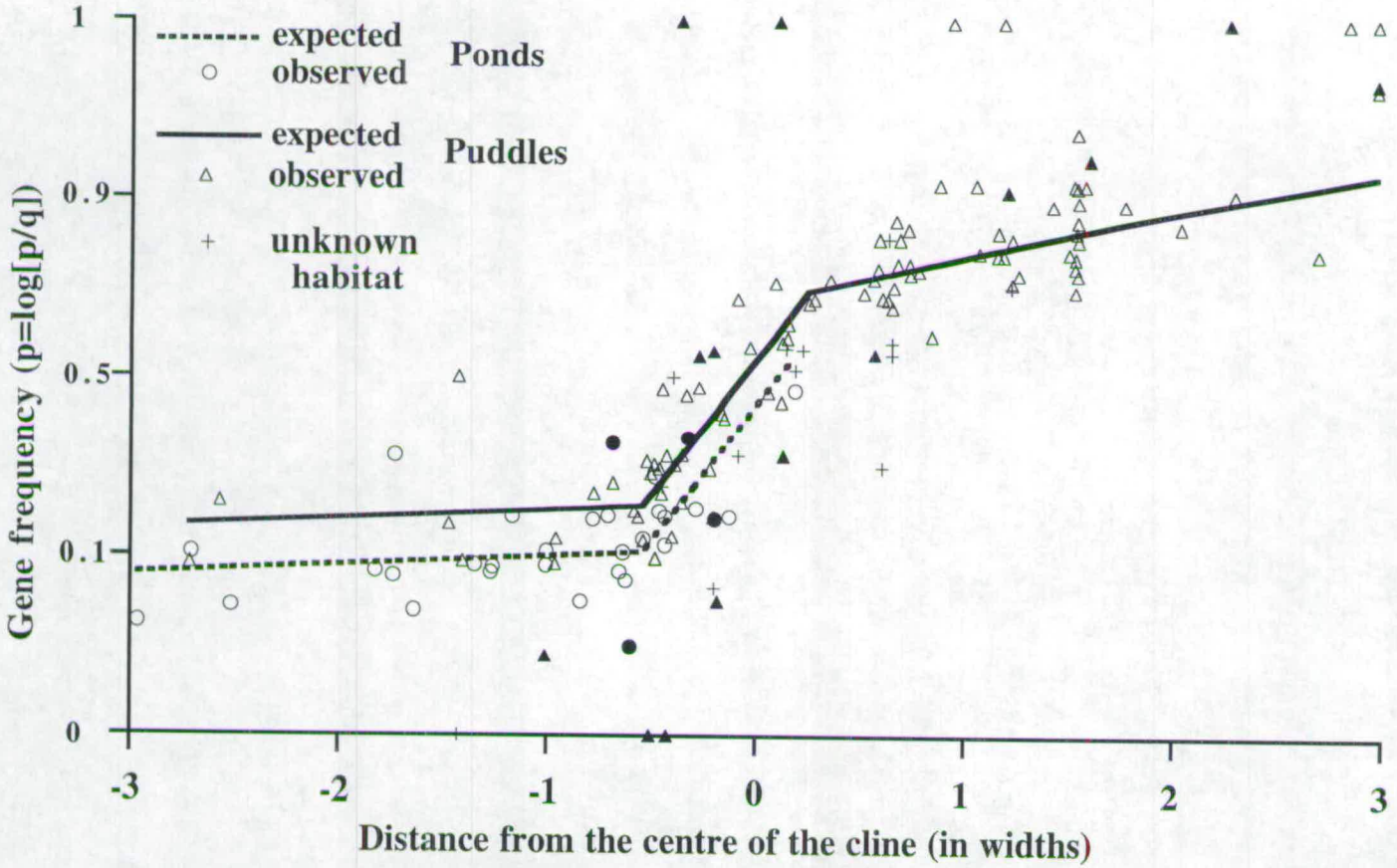


Figure 2.2: The frequency of *variegata* allozyme alleles plotted on a logit scale. This figure is reproduced (with permission) from MacCallum (1994).

Without habitat preference, the locally adaptive allele can only be maintained when the patch is larger than $l \sim \sigma / \sqrt{s}$, where σ is the standard deviation of the distance between the parent and offspring (a measure of dispersal) and s is the selection coefficient (Slatkin 1973). If the patches are sufficiently large, hybridisation occurs at the boundaries between patches, but the centre will remain fixed for the locally adaptive allele. An example of a mosaic of this sort may occur between the ground crickets *Gryllus pennsylvanicus* and *G. firmus* in Connecticut (Harrison 1986; Harrison and Bogdanowicz 1997). Here, although there is a broad north-south gradient, the transition is patchy over the scale of a few kilometres, because the two taxa occupy different soil types (Rand and Harrison 1989). A recent study found that hybridisation only occurred at the transition between two soil type patches (Ross and Harrison 2002). There are unfortunately no data on the strength of either habitat preference or selection against immigrants in this zone, but it is plausible that the latter alone could maintain their differences given the large patch sizes.

If the two taxa have an active preference for their habitat, then divergence can be preserved even when patches are much smaller than the dispersal range, because individuals choose not to move into the other habitat. The structure of this type of mosaic can be much more varied, because there is less constraint on patch size and location. For example, there is considerable fine scale structure in the hybrid zone between two chromosomal races of the fence lizard *Sceloporus grammicus* (Sites et al. 1995). This is probably created by limited female dispersal between their respective habitat types of oak trees and fallen logs, and both these are scattered throughout the contact zone. In the case of *Iris*, individual plants are associated

with either waterlogged or dry habitats over very small scales (<10 m). These patches may be maintained by pollinator preferences (Emms and Arnold 2000), ecological selection (Johnston et al. 2001), and the high prevalence of vegetative reproduction (Burke et al. 2000).

Mosaic patterns can also arise from random drift in an originally homogeneous metapopulation if the number of migrants between demes (*Nm*) or the neighborhood size is small (Wright 1943). A particularly important form of random drift is due to long range dispersal into the space between two expanding populations, although a mosaic generated in this way will disappear after a few generations of migration and selection once secondary contact has been fully established (Ibrahim et al. 1996). Perhaps more importantly, ongoing dispersal by pure individuals into unoccupied patches in the mosaic from the parental populations can sustain a mosaic pattern over longer time periods (Nichols and Hewitt 1994).

Examples of patchwork distributions apparently generated by drift (as opposed to habitat associations) include *Solenopsis* ants (Shoemaker et al. 1996), *Mus* (Hauffe and Searle 1993), *Limnoporus* water striders (Klingenberg et al. 2000) and *Palaemonetes* shrimps (Garcia and Davis 1994). It is difficult to be sure whether the failure to find a genotype-habitat association in these cases represents a failure to measure the right habitat variable or a genuine feature of the zone. Furthermore, mosaics must be mapped in two dimensions to understand their structure, and a number of these studies sample only in one dimension (normally along a road).

There are also few attempts to estimate relevant parameters, such as the dispersal rate, which makes drawing broader conclusions about mosaic hybrid zones from these studies difficult.

One more thoroughly studied example where both habitat associations and drift may be involved is the hybrid zone between the crickets *Chorthippus brunneus* and *C. jacobsi*. In this case habitat associations alone cannot satisfactorily explain the deviations in allele frequency away from a smooth cline. The excess variation could be explained by long distance dispersal from pure populations into the area of contact, either when the zone was initially forming or after the recurrent extinction of hybrid populations (Bridle et al. 2001).

In summary, the major determinant of the structure of a hybrid zone between ecologically differentiated taxa is the distribution of available habitat. If there is a sharp boundary in the availability of each type, a smooth cline may form, but an increasing interspersion of habitat patches will lead to an increasingly mosaic structure. Dispersal into the zone can help maintain local differentiation, but in very wide zones the influence of the pure populations on either side is very small, and the persistence of patches in the middle depends on the local balance of selection, migration and habitat preference.

Hybrid zones in Bombina

This chapter presents data from a mosaic hybrid zone between the fire-bellied toads *B. bombina* and *B. variegata* in Romania. These two taxa have been diverging for around 3-5 million years, and have adapted to different breeding habitats (Chapter 1). While *B. bombina* reproduces in semi-permanent ponds, *B. variegata* is a

typical puddle breeder. Nevertheless, they produce abundant fertile hybrids wherever their ranges meet (Szymura 1993). Several of its features should adapt *B. variegata* to reproduction in and dispersal between ephemeral habitats: a robust skeleton, thicker skin (to reduce water loss), larger and faster developing eggs and tadpoles. *B. bombina* is smaller, and produces a larger number of slow growing and consistently less active tadpoles. The latter trait is thought to reduce their visibility to pond based predators (Nürnberger et al. 1995; Kruuk and Gilchrist 1997; Vorndran et al. 2002). *B. bombina* is found in lowlands and flood plains throughout Central and Eastern Europe, whereas *B. variegata* is found at higher altitudes in Western and Central Europe (Fig. 1.3).

Bombina hybrid zones in Poland, Croatia and the Ukraine consist of narrow clines 5-10 km wide, which separate extensive pure populations (Szymura and Barton 1986, 1991; MacCallum et al. 1998; A. Yanchukov, pers. comm. 2001). These clines are concordant for a range of morphological traits and allozyme markers, suggesting that dispersal is forcing all loci to change in unison through the zone rather than settling at their local optima. This pattern is also consistent with selection being principally against hybrids, and there is some evidence for hybrid inviability in Poland (Koteja 1984) and Croatia (Kruuk et al. 1999a). In Croatia (MacCallum et al. 1998) and the Ukraine (A. Yanchukov, pers. comm.), there is an association between local habitat and genotype. In Croatia, this was shown to be due to an active habitat preference rather than strong ecological selection, as the difference were apparent over much smaller scales than the per season dispersal range (MacCallum et al. 1998). Although no systematic survey of habitats was made in Poland, habitat associations cannot be strong, simply because the hybrid zones fit closely to a set of smooth clines (Szymura and Barton 1991). These clinal patterns contrast sharply with Romania, where there is no steep cline in genotype

frequencies (Figures 2.3 and 2.4). Instead, the distribution of toads reflects the distribution of habitats in the area, with *B. bombina* alleles being associated with pond-like habitats and *B. variegata* with puddles. Other *Bombina* hybrid zones studied in Slovakia (Gollmann 1987) and Kostajnica, Bosnia (Szymura 1988) were also consistent with a mosaic distribution, as in both places *B. bombina* populations were found in *B. variegata*-dominated areas.

The most suitable way to quantify the difference between a clinal and a habitat-based mosaic hybrid zone is in the proportion of the variation in the distribution of genotypes explained by either their geographic location or the habitat type (in this case pond or puddle). At one end of the scale, the large areas containing only one taxon or the other on either side of a cline will create a strong spatial trend in allele frequencies, whereas at the other, in a wide mosaic, unless the habitat also changes along a gradient, the dominant variable will be the habitat type. In a mosaic zone generated by dispersal or drift, none of these variables will satisfactorily explain the distribution of genotypes, mainly because the mosaic pattern itself is the result of stochastic processes. I use this approach to quantify the differences between Romania and Croatia, and suggest reasons why the differences might arise. I then examine the evidence for habitat preference in Apahida.

Section 2.2 Materials and methods

In 1999, Ioan Ghira (Babes-Bolyai University, Cluj-Napoca, Romania) sampled waterbodies extensively around Cluj county, in the NW of the Transylvanian Plain of Romania, and collected 336 toads from 37 sites. Adult toads were caught by hand or with a net and immediately anesthetised in 0.2% MS222 (3-amino benzoic acid ethyl ester, Sigma). A toe was taken as a tissue sample and stored in 99.9% ethanol. I subsequently scored these animals for 3 marker loci: two SSCPs (*Bb7.4* and *Bv24.11*) and one microsatellite (*Bv12.19*), all of which segregate independently (Nürnberger et al. 2002b). Alternate alleles are assumed to be fixed in pure populations of either species, this is supported in *B. bombina* populations from Arad and Satu Mare near the Hungarian Plain (S. Köhler, unpublished data). In *B. variegata* the situation is less clear, as many toads appear to be introgressed; it is thus hard to know whether or not the *B. variegata* alleles were fixed before hybridisation began. Nonetheless, these alleles rise to high frequencies at Lupsa in the Apuseni mountains (S. Köhler, unpublished data). These samples provide a wider regional context for the Apahida area (see Fig. 2.3), although they were not included in the detailed genetic analyses (see below).

Toad Collections

In 2000 and 2001 attention was focused in an area 20 by 20 km, in our main study area near the village of Apahida (Figures 2.3 and 2.4). The south west corner of this area was located at 46°50'N, 23°47'E. Here, the landscape consists of rolling hills dissected by small streams flowing to the nearby Somes river to the north west. The

soil is mostly sandy loam, becoming more clay rich in the valleys. The altitude ranges from 300 to 550 m above sea level, and the vegetation is either small arable strip fields or pasture. There are some small woodlands and few patches of forest in the area, principally beech, oak and hornbeam, although none of the sites were under trees. A total of 745 individuals from 70 sites in 2000 were collected by myself and S. Köhler (Ludwigs Maximillian University, Munich) in 2000. A further 189 toads from 23 new sites were collected in 2001 by S. Köhler and T. Sands (University of Edinburgh). Only three of the sites were large ponds, the majority being either small ditches or tractor wheel ruts. One group of sites consisted of 10 circular excavated holes along the Cojocna road, each was around 3-4 m deep and 3-5 m in diameter (see Figures 2.4d and 4.1a). Several smaller, isolated watering holes were also found in other parts of the study area. As above, adult toads were caught by hand or with a net and immediately anæsthetised in 0.2% MS222. After a photograph was taken of the toads' individual belly pattern, a toe was taken as a tissue sample from either the right (2000) or the left (2001) foot and stored in 99.9% ethanol.

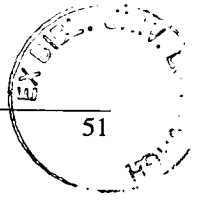
We also measured snout-vent and tibiofibula lengths, and recorded the presence of nuptial pads (found only on breeding males), dorsal warts and dorsal spots. Recaptures were simply re-photographed and released. Up to 15 toads were normally collected on a single visit, and it is unlikely that the number caught was proportional to the number present. This is firstly because capture difficulty is dependent on the habitat type being sampled: large ponds are clearly more difficult to search thoroughly than a puddle. Secondly, *Bombina* appear to spend a

proportion of their time on land, and it may thus be better to view the toads caught in the site as a sample of the toads in the surrounding area. Preliminary mark-recapture data suggest that some areas may contain many hundreds of individuals (T. Sands, unpublished data), only a proportion of which are in aquatic habitat at any one time.

Quantifying the difference in habitat use

B. bombina and *B. variegata* are found in ponds and puddles respectively, and these habitats differ mainly in their persistence through the season. This correlates well with many physical and ecological variables. For example, deeper water will persist longer in hot weather, and in turn will contain more specialised aquatic vegetation. The waterbodies around Apahida are mostly semi-natural (e.g. wheel ruts), or man-made holes. The latter can be roughly split into those created very recently, for example watering holes dug that year, and much older excavations, for example the 10 deep holes mentioned above, the first of which (200.10) was dug in the 1950's and the last (200.3) in the mid 1990's (Ilie Tudorescu, pers. comm.). To quantify the permanence of a site, the following variables were measured: width, depth, % emerged vegetation, % submerged vegetation and % bank vegetation (subdivided into three height classes: % <15 cm, % 15-50 cm, %> 50 cm), along with the location on our 20 by 20 km grid (i.e. the *x*, y coordinates). These variables were also measured in Croatia (MacCallum 1994; MacCallum et al. 1998). The ecological data for each site are given in Appendix II, the *x,y* coordinates for each are in Appendix III.

To enable direct comparison between the Romanian habitats and those sampled in Croatia, we calculated a discriminant function axis on both data sets jointly. I used the approach outlined in MacCallum (1994), in which a subset of sites are chosen as example 'ponds' or 'puddles', and the linear combination of habitat variables that best separates them calculated. All variables were transformed to improve their normality (log for continuous variables, arcsine for percentages: Sokal and Rohlf 1981) and the discriminant function was found for 152 sites (25 ponds and 36 puddles in Romania, 23 and 68 in Croatia) using the stepwise routine in SPSS. The results are given in Table 2.1. The four retained variables were width, % emerged



vegetation, depth and % submerged vegetation. With the exception of substituting % bank cover <15 cm for % submerged vegetation, the variables are the same as those in the axis calculated by MacCallum (1994) for Croatia. The distinction between the two habitat types was highly significant ($F_{4,146} = 91.3$, $P < 10^{-6}$). The function was then calculated for the remaining intermediate sites, and the axis rescaled to run from 0 (ponds) to 1 (puddles), using the most pond-like and the most puddle-like sites in the dataset as a whole as endpoints. I denote this axis by *H*. It correlates well with the original Croatian habitat axis (r = 0.95; MacCallum 1994) and an axis calculated for the Romanian sites alone (r = 0.97).

Table 2.1: Discriminant function coefficients (standardised) and their Wilks λ, which measures the effect of that term on the function. Variables were entered together, and are ordered by their contribution to the function. All Wilks λ gave P < 0.0001.

	Overall	Ponds	Puddles	Wilks λ	Correlation
Width	0.50	2.48	0.79	0.41	0.75
Emergent Veg	0.79	−7.21	−0.22	0.34	0.58
Depth	2.19	−2.24	4.93	0.29	0.57
Submerged Veg	0.66	0.67	−1.58	0.28	0.34
Constant	0.49	−5.60	−5.11		
Group Mean		2.33	−1.05		

Genotyping methods

The 2000 and 2001 animals were scored for the same loci as the 1999 samples, namely *Bb7.4*, *Bv12.19* and *Bv24.11*, along with an additional unlinked microsatellite locus *Bv24.12*. The GenBank accession numbers for these loci are AF472441, AF472423, AF472425 and AF472426 respectively. These markers were chosen because, unlike allozymes, they can be reliably scored in both adults and tadpoles. Upon return to the laboratory, the samples were stored in a -60° C freezer prior to processing. For DNA extraction, the tissue samples were digested overnight with Proteinase K (final concentration: 100 µg/ml) at 55° C in 0.5 ml TNES buffer (0.05 M Tris, 0.4 M NaCl, 0.1 M EDTA, 0.5% SDS). Following the addition of 0.5 ml 2.6 M NaCl, the samples were shaken vigorously for 15 sec and then centrifuged for 10 min. The supernatant was transferred to a fresh tube, extracted once with chloroform, and DNA was precipitated with two volumes of ethanol. The DNA pellet was washed once in 70% ethanol, air-dried and then resuspended in 100 µl of ultrapure water (Merck). Stock solutions were stored at -20° C.

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 units Taq polymerase (rTaq, Amersham Biosciences). $MgCl_2$ concentrations varied among loci between 1.5 mM 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 sec at 94° C, 30 sec at x° C and 30 sec at 72° C for 32-35 cycles, where x is the locus-specific annealing temperature.

The length differences between the microsatellite PCR products were visualised with an ALFexpress automatic sequencer (Amersham Biosciences). 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 hrs (MultiPhor gel rigs, Amersham Biosciences) with an electrode buffer of 2 x TBE and a gel buffer of 1 x Tris-acetate. All samples were denatured in formamide loading buffer at 95° C for 4 min and immediately placed on wet ice before loading. Staining involved the following three steps with intermittent rinses in distilled water: 10% acetic acid (10 min), 0.1% AgNO3 (10 min) and developer (0.375 M NaOH, 2.5 mM sodium borohydride and 0.3 % Formaldehyde, 5-10 min). Thereafter, the SSCP gels (with acetate backing sheet, GelFix™, Serva) were soaked for 30 min in 10% glycerol, air-dried and finally covered with a plain acetate sheet.

I genotyped 479 of the toads from 2000, and S. Köhler and M. Thiel (now at the Technical University, Munich) scored the remaining 269. All 206 toads from 2001 toads were all genotyped by S. Köhler. Two competing ways of scoring locus *Bv12.19* were found for the animals I genotyped in 2000, depending on whether or not weak peaks in the curve output of the automatic sequencer (AlfExpress, Amersham Biosciences) were scored as alleles. Including these faint alleles gave only a weak heterozygote deficit F_{IS}= 0.09, whereas leaving them out gave F_{IS} = 0.54. The former is similar to the deficit at *Bv12.19* for the other 269 samples from

2000 (F_{IS} = 0.08) and the 206 toads genotyped in 2001 (F_{IS} = 0.02). Since there was no reason to believe that *Bv12.19* was subject to greatly different evolutionary forces between the samples, the faint alleles were included.

Based on the analysis of pure populations from Romania (B. Nürnberger, pers. comm.), there were two *B. variegata* alleles and one *B. bombina* allele at *Bb7.4*, *Bv12.19* and *Bv24.11*, whereas *Bv24.12* had five alleles characteristic of *B. variegata*, one characteristic of *B. bombina,* and one of exactly intermediate length. Since the latter was found on only four occasions, and then only in hybrid sites, it was left unscored. (This should have no appreciable effect on the results). The *B. variegata* alleles were combined to make all the loci biallelic; alleles are henceforth labelled either 0 (*B. bombina*) or 1 (*B. variegata*). Mis-scoring rates for these loci have been found to be around 1% in a parallel study (B. Nürnberger, unpublished data; see also Chapters 4 and 5). Since the purpose of this chapter is simply to describe the distribution of genotypes across the study area, we summarise the state of a toad's genome by the total number of '1' alleles carried, creating a hybrid index (HI) which runs from 0 for pure *B. bombina* to 8 for pure *B. variegata.* It is important to bear in mind that this measure is an estimate based on only four markers, and individuals with a 'pure' HI of 0 or 8 may well be introgressed at other loci (Boecklen and Howard 1997).

Section 2.3 Results

The spatial pattern of hybridisation between *Bombina bombina* and *B. variegata* in the area around Apahida differs strikingly from the narrow transition zones that have been described in Poland, Croatia and in the Ukraine (Szymura and Barton 1986, 1991; MacCallum et al. 1998; A. Yanchukov, pers. comm.). There, allele frequency clines were found in bands 6 to 9 km wide, located at ecotones between forested hills and open plains and flanked on either side by extensive areas containing only pure animals. In contrast, the hybrid populations to the east of Apahida (Romania) are about 20 km from pure *B. variegata* populations to the NW in the foothills of the Apuseni mountains and 100 km from the main expanse of pure *B. bombina* in the Hungarian plains (Figures 1.3 and 2.3). Nevertheless, populations of *B. bombina* are found sporadically in large ponds across the entire Transylvanian Plains, often in close proximity to the puddle-dwelling *B. variegata* (Stugren 1959; Stugren and Vancea 1968; pers. obs.).

Our collections across an area of about 40 x 40 km around the city of Cluj confirm this extended mosaic (Fig. 2.3). Isolated *B. bombina* sites in areas dominated by *B. variegata* have been documented before on the Slovakian karst plateau (Gollmann et al. 1988) and in the Mátra mountains of Hungary (Gollmann 1987). Furthermore, in Kostajnica, Bosnia, large allele frequency differences were observed between pond and puddle sites even when they were in close proximity (Szymura 1988, 1993). In the following analyses, I compare the spatial and genetic structure of the Apahida study area to the previously described hybrid zone in Pešćenica, Croatia

(MacCallum et al. 1998; see Fig. 2.5). The latter has a strong clinal structure across an altitudinal gradient, yet also has mosaic pattern at its centre. Around Apahida, in contrast, there is no evidence for a cline over the same spatial scale (Fig. 2.4a), and neither is there is a comparable ecological gradient (see below). It should be noted that the genetic data for Croatia were four allozyme loci, whereas here four neutral DNA markers are used. Since the allozymes *Ldh-B* and *Mdh-1* and SSCP loci *Bb7.4* and *Bv24.11* are closely concordant in a Ukrainian transect (A. Yanchukov, pers. comm.), we consider results based on either allozymes or DNA markers to be directly comparable.

Figure 2.3: Pie map of the 1999 samples around Cluj county (number 4 on Fig. 1.3). Circles show the mean proportion of *variegata* alleles per individual at three marker loci for each site. The box shows the area in which the 2000 and 2001 samples were collected

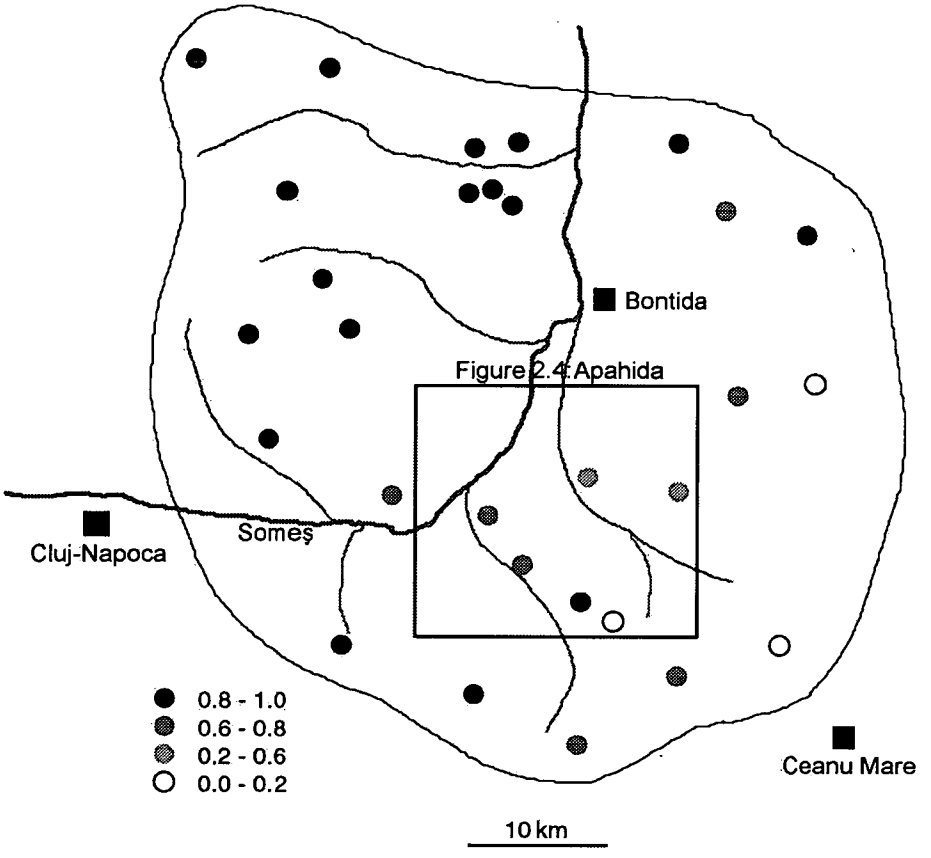


Figure 2.4a: Pie map of the entire Apahida study area, showing only the samples collected in 2000 and 2001. Pies show the proportion of *variegata* alleles per individual at four marker loci for each site. The colour of the shaded part of the pie runs from blue ($H = 0$; ponds) to green ($H = 1$; puddles). The three inserts are magnified in Figure 2.4 b-d.

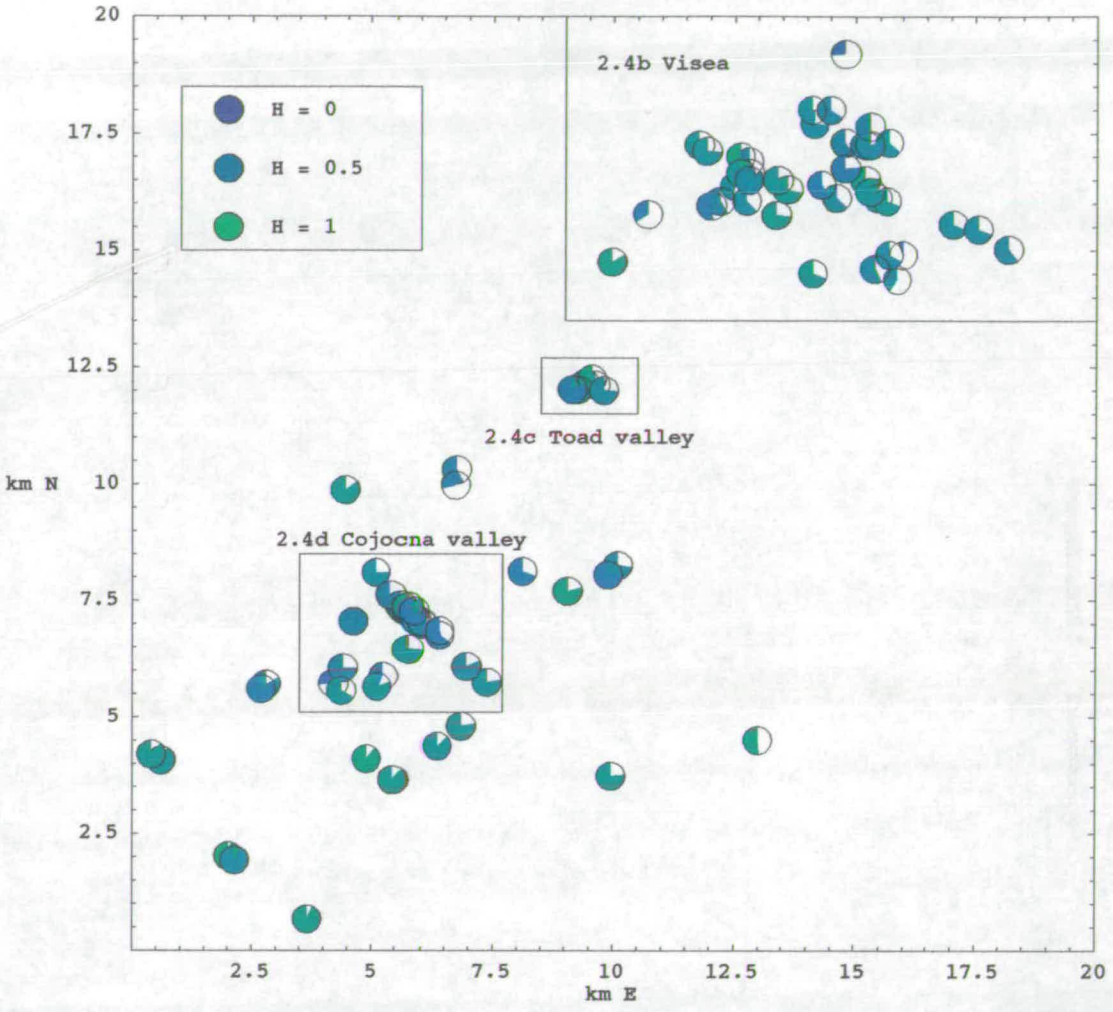


Figure 2.4b: Detail from Fig. 2.4a: the Visea valley. Pies show the proportion of *variegata* alleles per individual at four marker loci for each site. The colour of the shaded part of the pie runs from blue ($H = 0$; ponds) to green ($H = 1$; puddles). The circle is centred on site 290 and encloses a radius of 300 m (see Chapter 4).

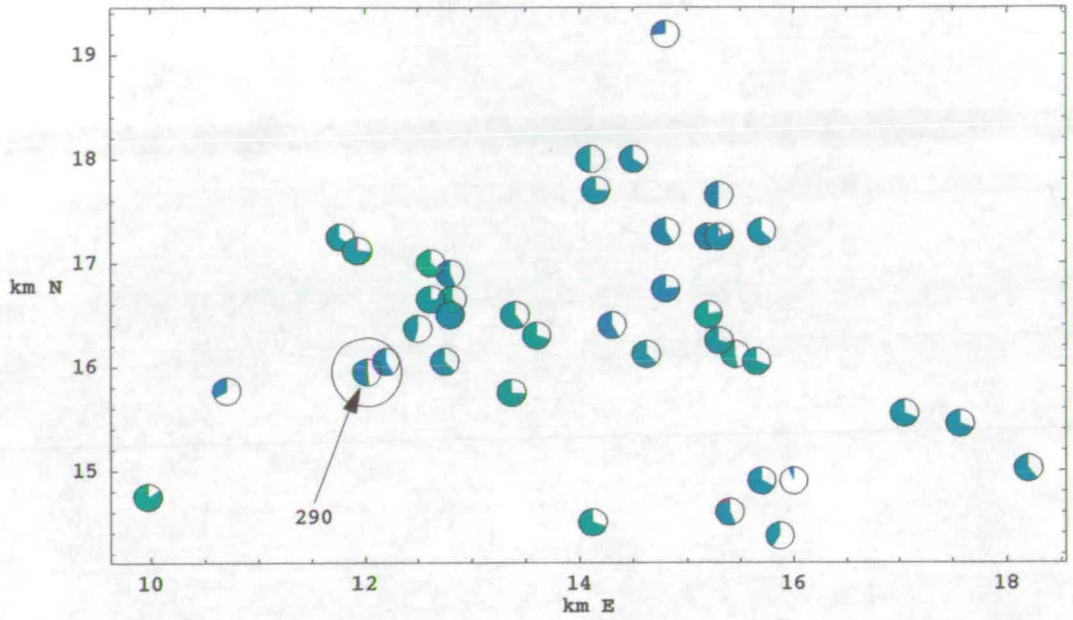


Figure 2.4c: Detail from Fig. 2.4a: Toad valley (in Romanian the 'Valea Broastelei'). Pies show the proportion of *variegata* alleles per individual at four marker loci for each site. The colour of the shaded part of the pie runs from blue ($H = 0$; ponds) to green ($H = 1$; puddles). The circle is centred on site 258 (see Chapter 4) and encloses a radius of 300 m (the scale has been adjusted for clarity).

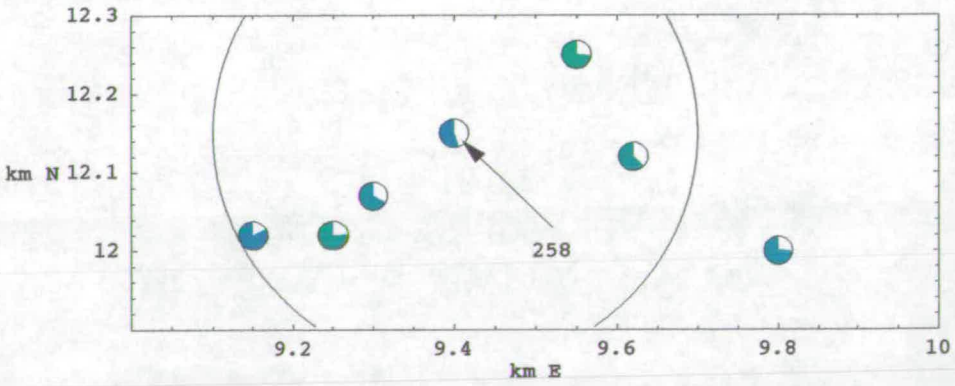


Figure 2.4d: Detail from Fig. 2.4a: the Cojocna valley. Pies show the proportion of *variegata* alleles per individual at four marker loci for each site. The colour of the shaded part of the pie runs from blue ($H = 0$; ponds) to green ($H = 1$; puddles). The circle is centred on site 258 (see Chapter 4) and encloses a radius of 300 m (the scale has been adjusted for clarity).

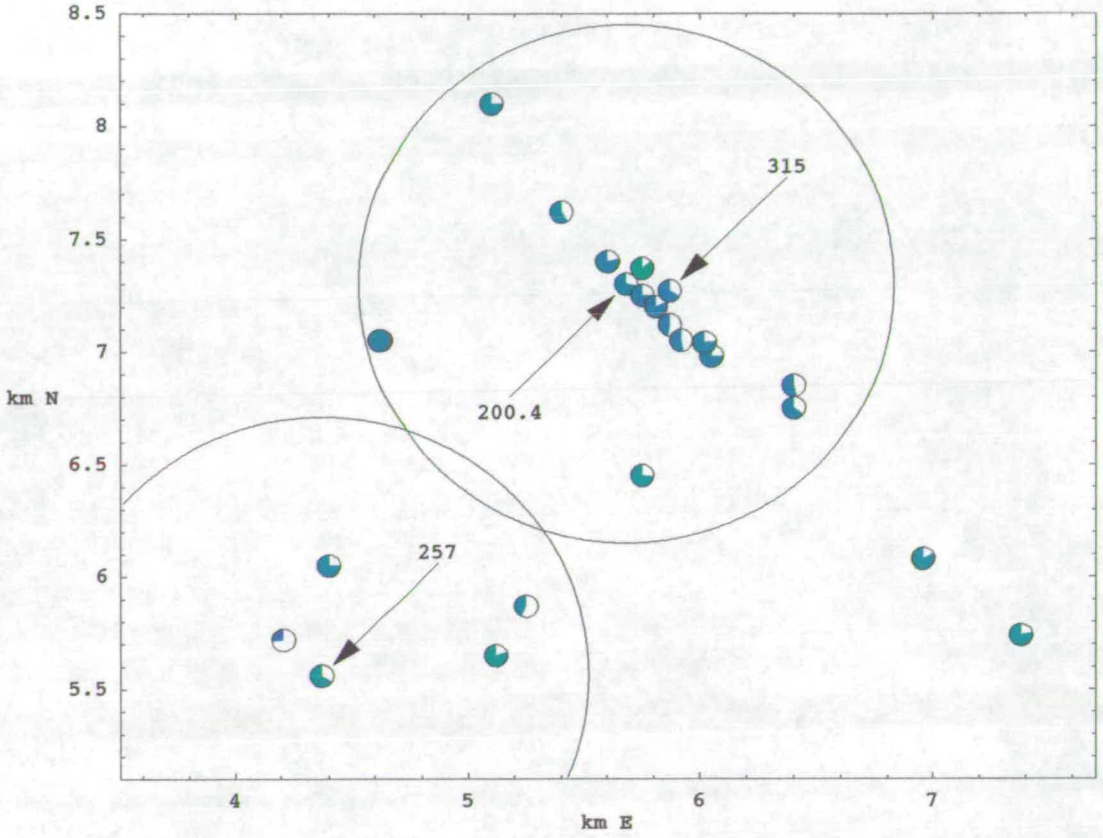


Figure 2.5a: Pie maps for the transect in Pešćenica, Croatia (number 3 on Fig. 1.3). Pies show the proportion of *variegata* alleles per individual at four allozyme loci for each site (data from MacCallum 1994).

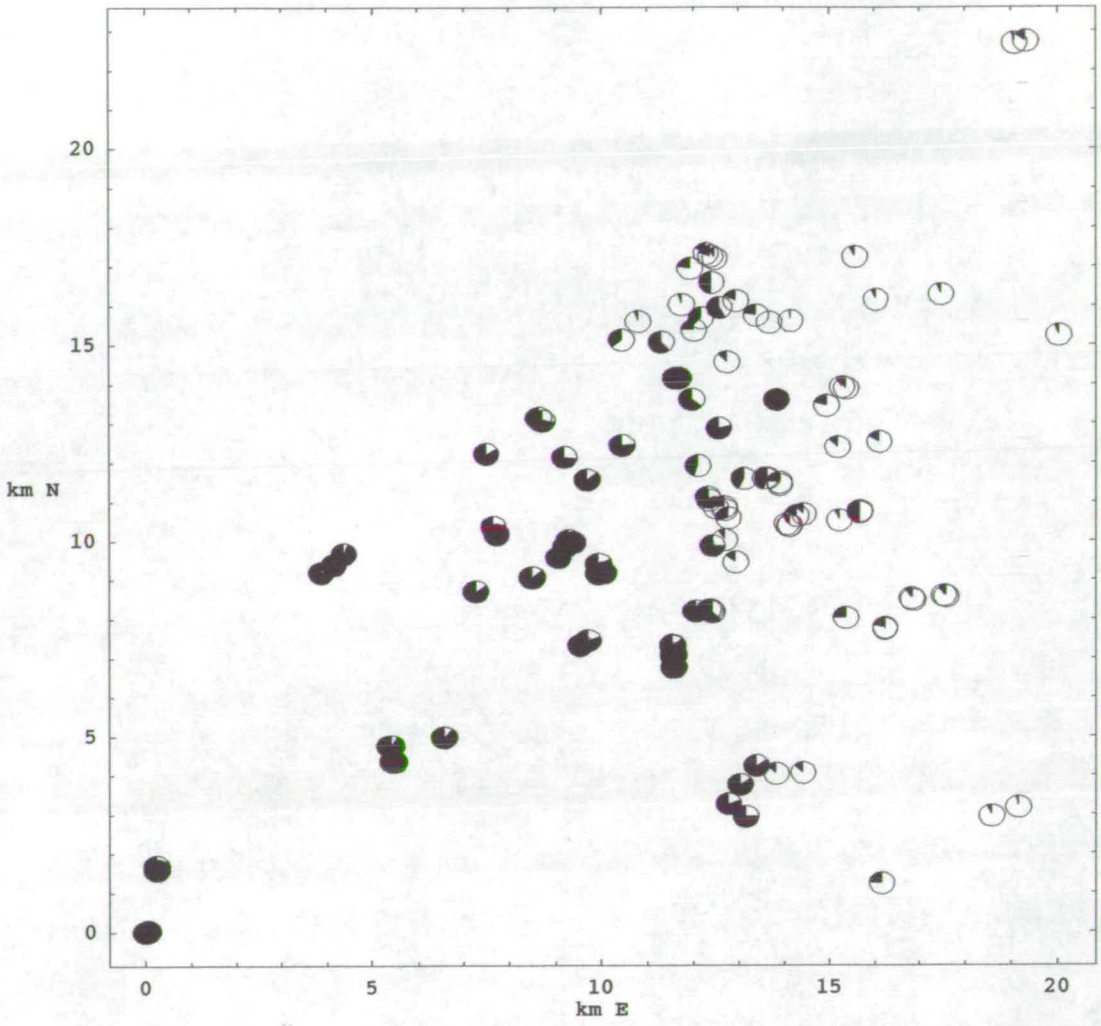
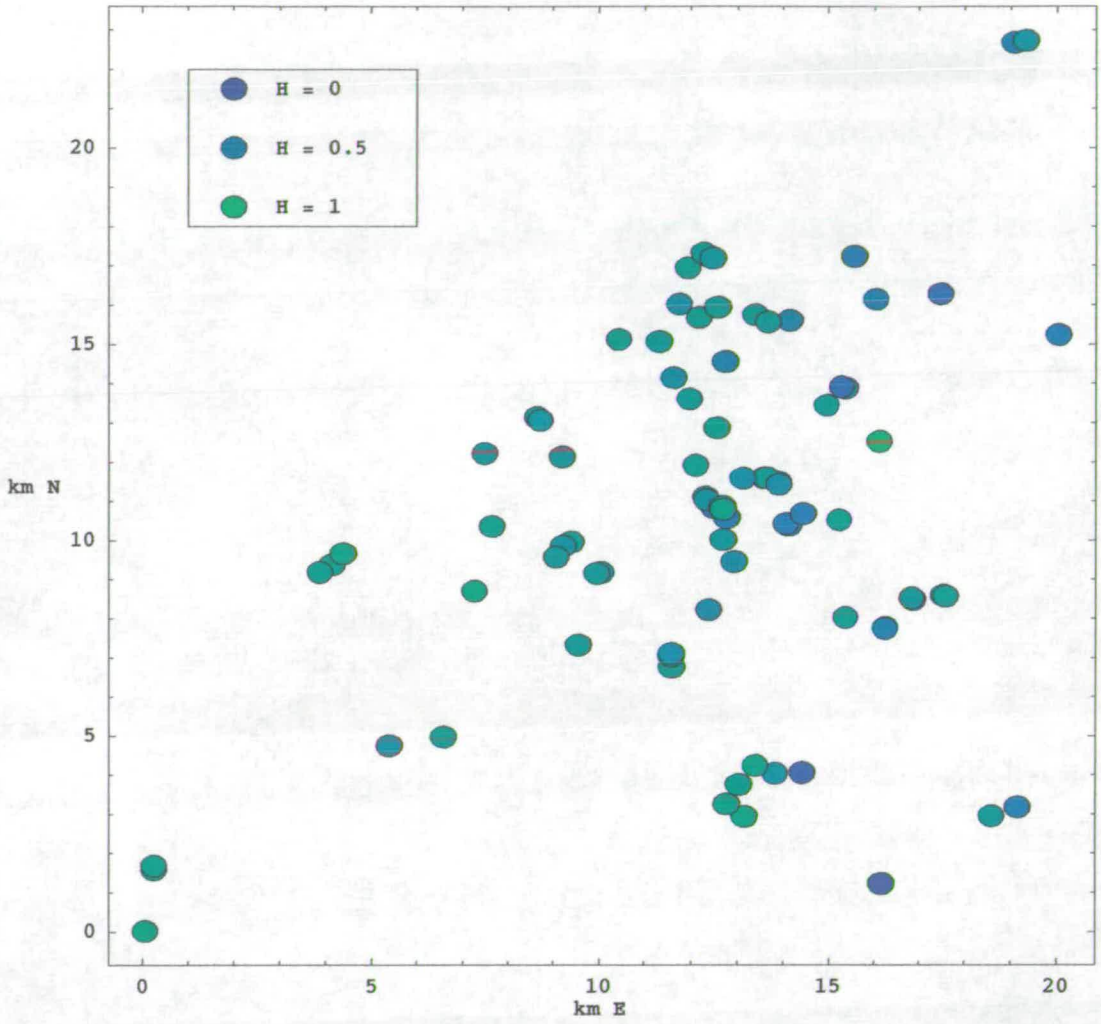


Figure 2.5b: The distribution of habitat for the transect in Pešćenica, Croatia (number 3 on Fig. 1.3). Each circle represent one site, the colour is determined by the habitat score H, this runs from blue ($H = 0$; ponds) to green ($H = 1$; puddles). The original ecological data are from MacCallum 1994, the calculation of the habitat axis is described in the Methods section.



In the majority of sites, the population mean frequency of *B. variegata* alleles ($\overline{p}$) among sampled adults ranged from 0.5 to 0.8. There were no pure *B. variegata* populations. Only four sites had $\overline{p} < 0.4$, and in these, only 13 out of 33 animals contained only *B. bombina* alleles (hybrid index: HI = 0). Overall, the adult sample was dominated by hybrid individuals (Fig. 2.6a, see Appendix III for a complete listing of the genetic parameters for each site in Apahida). The

Figure 2.6a: The distribution of genotypes from the Apahida study area. The histogram shows all 954 individuals classified according to the number of *variegata* alleles that they carry at our four diagnostic marker loci (axis 'ind HI'). Sites were grouped by their mean allele frequency into bins of width 0.05 (axis 'p mean'). The 140 individuals not scored at all four loci were classified in the range 0-8 by interpolation.

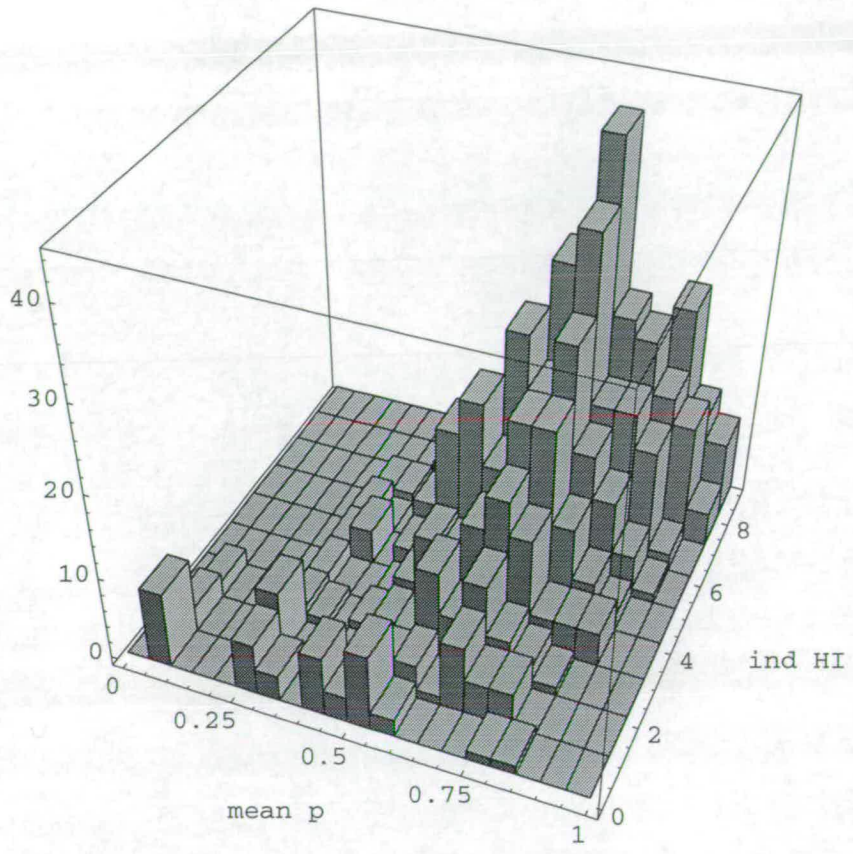


Figure 2.6b: The distribution of genotypes from the Pešćenica study area, the data are from MacCallum (1994). The bars are as in Figure 2.6a, and the data from 1768 toads genotyped at four allozyme loci. A similar chart is shown in Figure 2 of MacCallum et al. (1998).

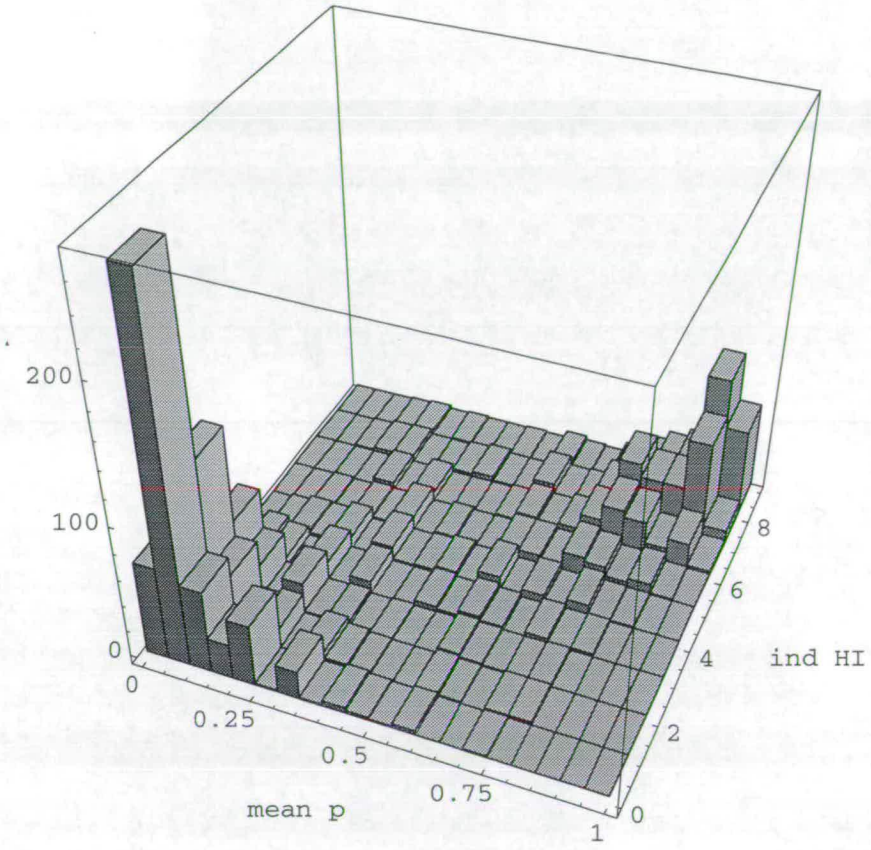
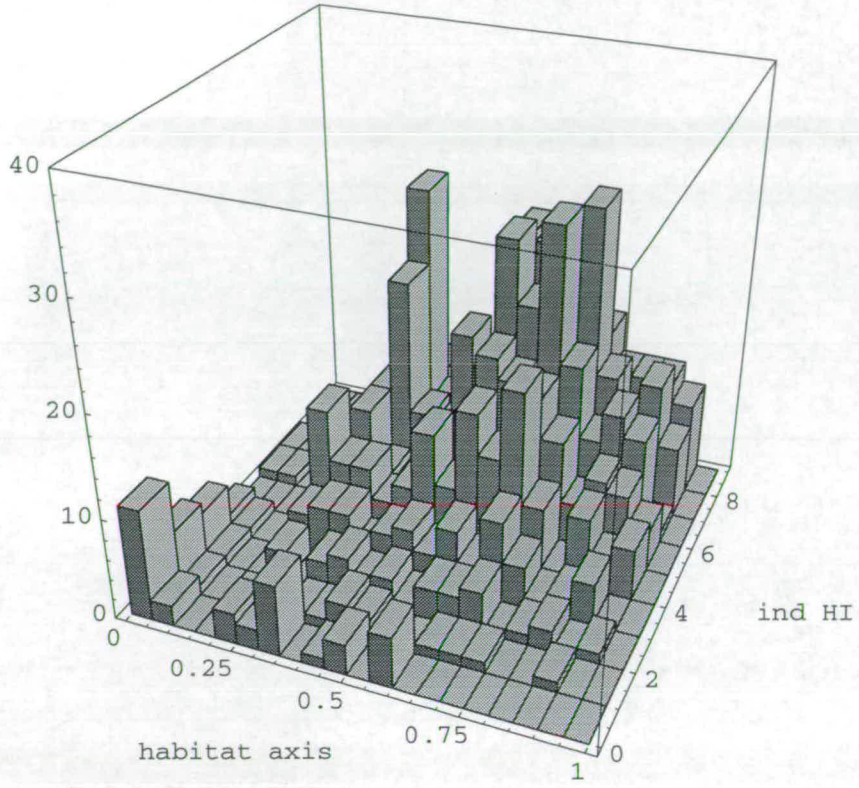


Figure 2.6c: The distribution of genotypes from the Apahida study area by habitat type. The data are as in Fig. 2.6a above, but sites are instead grouped by their habitat score *H* into bins of width 0.05.



Spatial distribution of genotypes and habitats

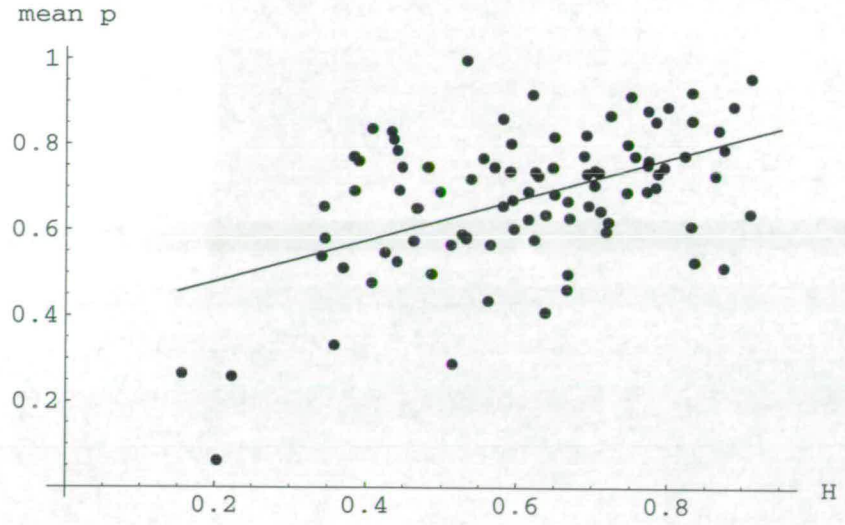
What are the relative roles of spatial location and the site habitat in structuring the Apahida and Pešćencia zones? We can quantify this by exploring the proportion of variation in mean allele frequency explained a) by a clinal transition and b) by the features of the aquatic habitat. It is possible to use a partial Mantel test in this context (Smouse et al. 1986), but since our sites are highly spatially clustered (Fig. 2.4), the null distribution cannot be created from permutations of the pairwise geographic distance matrix (Raufaste and Rousset 2001). Instead, we fitted a multiple regression model to explain the number of *variegata* alleles (z) observed from a possible total of 2 x 4 x n, where n is the number of toads sampled. This binomial error structure is preferable to a Gaussian because it neatly accounts for variation in n, but it implicitly assumes that the residual variation between sites is solely due to sampling error. This may not be entirely valid in this case, as potentially many other factors could create additional variance in allele frequency, such as recolonisation from outside the zone or migration between adjacent sites.

In fact, the multiple regression analysis also fails to account for the effects of spatial clustering of sites. Unfortunately, there is no universally accepted solution to the problem of spatial autocorrelation, and I therefore assess its impact on the reliability of the results below.

The x, y-coordinates and the discriminant habitat score H (see Materials and Methods) for each site were entered simultaneously as independent variables. The fitted regression was $z = 0.037 - 0.05x - 0.01y + 2.28H$. However, as expected,

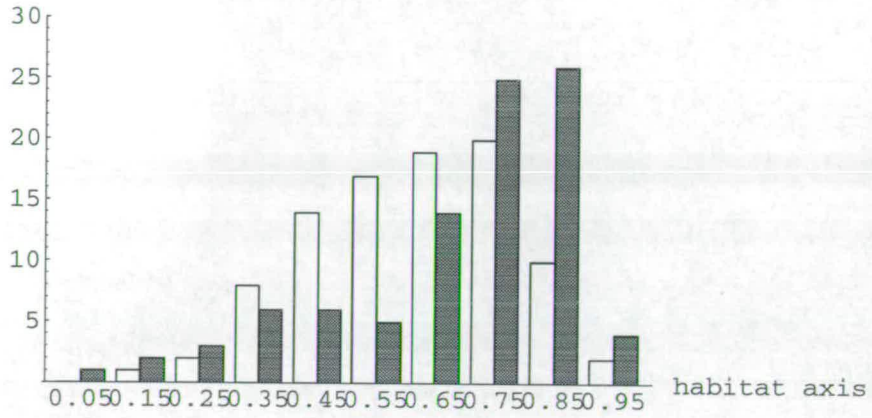
there is an excess of error variance beyond that predicted by the binomial model (model $\chi^2_3 = 367.7$; error $\chi^2_{89} = 540.1$), and this will inflate estimates of significance for the individual model parameters. Since we expect unit error variance per degree of freedom when binomial sampling error is the only source of residual variance, we can factor out the excess by rescaling the χ^2 values for each variable by 540.1/89 = 6.06. (Note that this only adjusts the significance levels and does not properly account for the causes of the excess variance.) One can now test the relationship between the explanatory χ^2 and the rescaled error χ^2 as an F-ratio (Crawley 1993). After this correction, neither the x or the y spatial location axes were significant (P = 0.06 and P = 0.31 respectively), and the habitat variable is of overwhelming importance ($F_{1,89} = 36.9$, $P < 10^{-7}$), as illustrated in Figure 2.7. When the same analysis was applied to MacCallum's (1994) data on the Pešćenica transect (the error variance was $\chi^2_{86} = 709.5$; rescaling by 8.25), the best fit was $z = 3.73 - 0.35x - 0.09y + 1.81H$. Spatial location was highly significant in this case (x: $F_{1,86} = 194.1$, $P < 10^{-10}$; y: $F_{1,86} = 23.1$, $P < 10^{-5}$). This is not unexpected given the gradient in allele frequency running SW-NE across the Pešćenica study area (Fig. 2.5a). The habitat score H was highly significant as well ($F_{1,86} = 18.7$, $P < 10^{-4}$).

Figure 2.7: Mean *B. variegata* allele frequency $\overline{p}$ against the habitat axis H for Apahida. The regression line is $\overline{p} = 0.38 + 0.47H$ ($F_{1,92} = 30.2$, $P < 10^{-6}$).



Habitat is an major determinant of allele frequency in both Apahida and Pešćenica, and therefore a gradient in the types of habitat available will have an important effect on the structure of the zone. We can detect a gradient in *occupied* habitat with a least squares regression of the habitat axis H against the x and y coordinates. There is a significant relationship between east-west position and habitat type in Pešćenica ($F_{1,92} = 19.7$, $P < 10^{-5}$), but not north-south ($P > 0.1$). There is no relationship in Apahida ($P > 0.1$ for both x and y). This pattern can also be seen in Figure 2.4 (Apahida) and Figure 2.5b (Pešćenica). It is important to note that these data only address occupied sites and not the available habitat. Collecting data on the latter is hampered by the difficulty of defining a 'usable' waterbody, especially from the perspective of a toad. I consider the relationship between available and occupied habitat in both areas below. In addition, the distributions of the range of occupied habitat in both areas differ slightly (Fig. 2.8): intermediate habitats are used most often in Apahida, whereas the majority of occupied sites in Pešćenica are more puddle-like.

Figure 2.8: The distribution of habitat in Apahida (white bars) and Pešćenica (shaded bars).



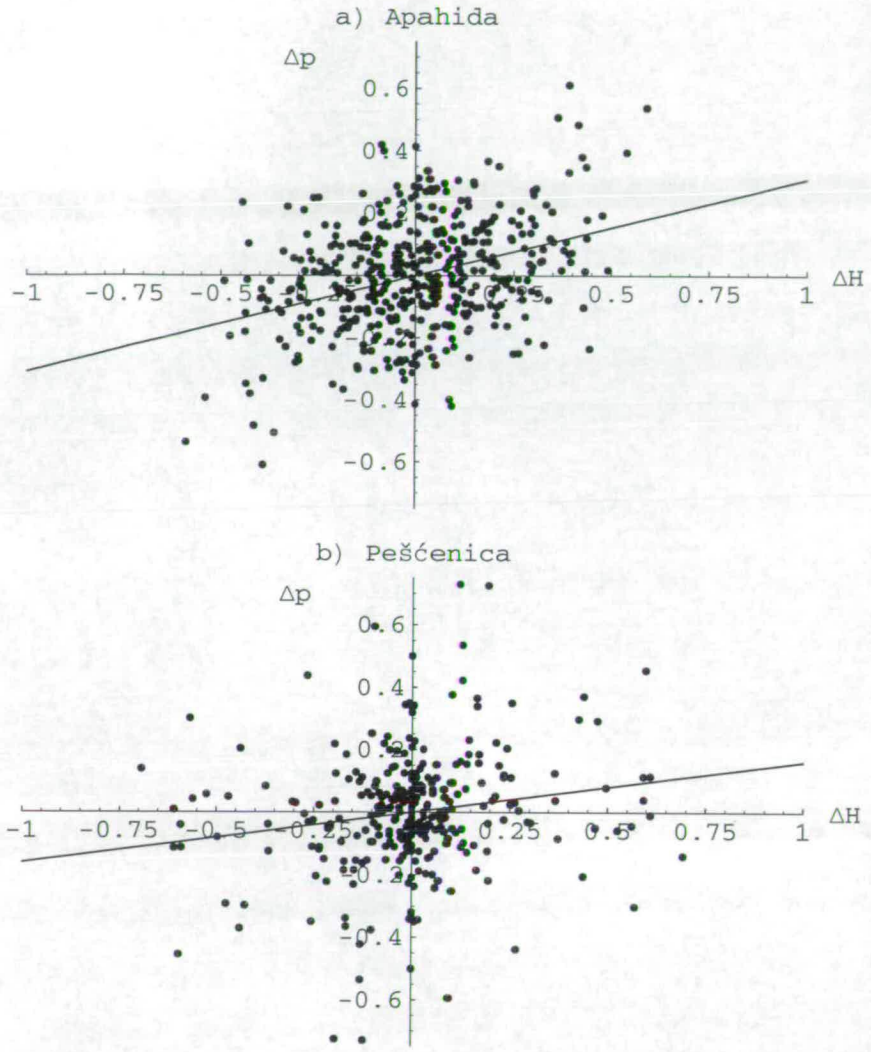
The association between habitat and allele frequency

In Apahida, because there is no clear cline, we can gauge the overall allele frequency difference between ponds ($H = 0$) and puddles ($H = 1$) from the slope of a regression of $\overline{p}$ onto the habitat axis (Fig. 2.7). This gives $\Delta\overline{p} = 0.47\Delta H$. Getting a comparable figure for Croatia is difficult: the steep cline confounds any quantification of habitat associations over large spatial scales, either because habitat types are not equally available, or because the presence of the other taxon precludes their use. We can, however, examine how the difference in $\overline{p}$ corresponds to the difference in habitat score between nearby sites. Since these sites are within easy dispersal distance, differences in allele frequency between them reflect the strength of habitat preference.

For this analysis, I use only those pairs of sites that are less than 1 km apart (an estimate of the lifetime dispersal distance derived from the Polish and Croatian transects, Szymura and Barton 1991; MacCallum et al. 1998), and calculate the pairwise difference in $\overline{p}$ and H (denoted $\Delta\overline{p}$ and ΔH respectively). The data include both possible orderings for each pair, as including only one ordering when $\Delta\overline{p}$ and

ΔH have different signs will bias the slope depending on which ordering is included. (Including both orderings does not bias the regression estimate.) Furthermore, since we expect $\Delta\overline{p} = 0$ when $\Delta H = 0$, the intercept is set to zero. The relationships in Romania and Croatia are $\Delta\,\overline{p} = 0.30\Delta H$ and $\Delta\,\overline{p} = 0.16\Delta H$ respectively (Fig. 2.9). Since the regressions are based on pairs of sites (rather than on independent data) and both orderings of each pair are included, the sample size is greatly inflated, and therefore significance tests on the difference between the slopes would be misleading. However, the plots indicate that the relationship between habitat and allele frequency is stronger in Apahida.

Figure 2.9: Plots of the difference in *B. variegata* allele frequency against the difference in habitat score for pairs of sites under 1 km apart for a) Apahida and b) Pešćenica. The regression line is also shown for each, see text for details.



It is interesting to note that although there is no clear gradient in allele frequency in any direction in Apahida, there is evidence for spatial structure. For example, the difference in allele frequency between extreme habitats (H = 0 and 1) is larger when estimated over larger spatial scales: we estimated $\Delta \overline{p} = 0.30\Delta H$ by comparing pairs of sites under 1 km apart, whereas a regression of allele frequency against habitat for all sites gave $\Delta \overline{p} = 0.39\Delta H$. I consider possible reasons for this below. Estimates of habitat association were also made in two different ways in a

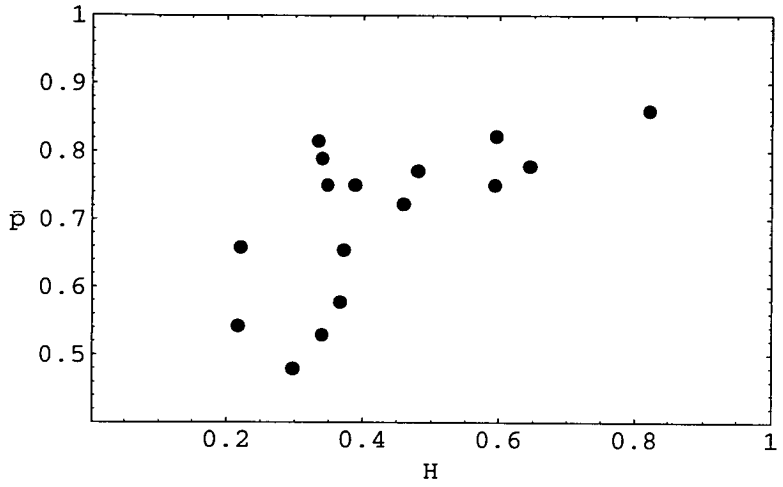
previous analysis of the Croatian data. MacCallum et al. (1998) calculated the difference in allele frequency between adjacent ponds and puddles (≤ 300m apart) as 0.25; this figure is probably less accurate, however, as it is based on only 7 pairs of sites. MacCallum (1994) fitted a model of a sigmoid cline to the whole Croatian dataset, allowing the habitat as an additional explanatory variable. This gave $\Delta \overline{p} = 0.15\,\Delta H$, which is close to our simpler estimate of 0.16.

There was also direct evidence of a habitat preference in Pešćenica: a) significant associations between genotype and habitat were seen on a smaller spatial scale than the within-season dispersal distance; and b) more directly, in two sites, individual movements within a season were correlated with genotype: toads leaving a puddle-like site and travelling to a more pond-like one were significantly more *B. bombina*-like than the ones that remained (MacCallum et al. 1998).

In Apahida, it is more difficult to demonstrate non-random dispersal because suitable sites are generally further apart, making the study of dispersal by mark-recapture less efficient. However, several observations indicate the existence of a preference. The presence of pure *B. bombina* individuals in ponds is striking in a landscape in which *B. variegata*-like animals are found in practically all temporary sites. If they moved between aquatic sites at random, selection against *B. variegata* adults arriving into a pond would have to be very strong indeed in order to eliminate them completely. In addition, in the Cojocna valley there were 16 adjacent man-made sites in various stages of succession (average distance between neighboring sites ~10 m, see Figure 2.4d). Even on this scale, there was a significant correlation between genotype and habitat score ($r = 0.64$, $P = 0.003$; Figure 2.10), which again would otherwise require very strong selection adults in

the wrong habitat. This habitat preference has probably arisen in response to differential adaptation (Rice and Hostert 1993), and thus selection may also contribute to the genotype-habitat association in *Bombina*.

Figure 2.10: Plot of the mean allele frequency against the habitat axis H for 16 sites within a 50 m radius in the Cojocna valley (see Fig. 2.4d).



The effects of selection can be seen over wider spatial scales, as habitat associations depend to some extent on the spatial structure. Specifically, the difference in allele frequency between extreme habitats is larger when estimated over larger spatial scales: the analysis above found $\Delta\,\overline{p} = 0.30$ between pairs of sites under 1 km apart (Fig. 2.9a), but the regression of allele frequency against habitat for all sites gave $\Delta\,\overline{p} = 0.39$ (Fig. 2.7). This may be because sites are not evenly spaced, and the habitat preference is not complete: more isolated sites are less likely to contain casual visitors and ecological selection will cause $\overline{p}$ will reflect H more closely. However, local migration will homogenise allele frequencies between nearby habitat types, and this will further increase the discrepancy between $\Delta\,\overline{p}$ estimates for adjacent pairs of sites and across the study area as a whole. It should be noted that removing the three pond sites from the regression shown in Figure 2.7 gives $\Delta\,\overline{p} = 0.27$, which is much more similar to $\Delta\,\overline{p} = 0.30$ between pairs of sites under 1 km apart. Since it is unclear which of

these values is a more accurate reflection of the genotype-habitat association as a whole, it is difficult to assess the role of ecological selection in maintaining allele frequency differences from these data. Nonetheless, both estimates of $\Delta\overline{p}$ indicate a stronger habitat association than Pešćenica, and this is probably due to a stronger active preference.

Section 2.4 Discussion

This survey of genotype frequencies across the *Bombina* hybrid zone around Apahida is in striking contrast with the narrow clines seen in Poland and Croatia. No steep gradient in allele frequencies was apparent; instead, there was a fine-scaled mosaic with strong divergence in marker frequency between adjacent habitats (Fig. 2.4). The frequency distribution of pure and hybrid populations was asymmetric. Pure *B. bombina* populations were only found in sporadically occurring large ponds, whereas *B. variegata*-like hybrid populations inhabited the much more abundant temporary sites in the surrounding landscape (Figures 2.4 and 2.6a,c). This asymmetry is even more remarkable considering that the nearest extensive pool of entirely pure *B. bombina* lies 100 km away in the Hungarian plains: local strongholds of *B. bombina* appear to cause the observed massive introgression of *B. bombina* alleles into the surrounding *B. variegata*-like gene pool, in which 83% of adults were recombined at our four marker loci (Fig. 2.6a). I have quantified two major differences between Apahida and Pešćenica, in both the spatial distribution of occupied habitat, and in the strength of habitat preference. I now consider reasons why these differences might have arisen.

Why should hybrid zones differ in structure?

In general, there are three processes by which contact zones may arrive at different spatial patterns. Firstly, the toads themselves may differ between areas. In neutral traits, Carpathian (Romania and Poland; Fig. 1.3) and western (Croatia) populations of *B. variegata* differ considerably at 9 allozyme loci (Nei's D_N = 0.16; Nei 1972), compared to D_N = 0.65 between *B. bombina* and *B. variegata*. The maximum

differentiation between *B. bombina* populations is D_N = 0.09 (data from Figure 10-2 in Szymura 1993). These differences must represent a long term reduction in gene flow between *Bombina* groups, but has this resulted in phenotypic divergence that could alter the structure of a hybrid zone? In the case of habitat preference, the regressions of the change in allele frequency against the change in habitat for nearby sites in Apahida and Pešćenica suggest a stronger preference in the former: $\Delta \overline{p} = 0.30\ \Delta H$ and $\Delta \overline{p} = 0.16\ \Delta H$ respectively (Fig. 2.9), although I was unable to test the significance of this difference. Assuming that all habitat types are equally available everywhere, one would expect that a stronger habitat preference might allow the two taxa to move past each other into sympatry upon secondary contact.

When habitat preference is weaker, pure types cannot be preserved in the face of hybridisation for more than a few generations, and establishing pure populations in the range of the other taxa is impossible. In this case, unless there are additional prezygotic isolating mechanisms, the two taxa will probably be parapatric and separated by a hybrid zone. Is this the cause of the differences between Apahida and Pešćenica? It is hard to unravel cause and effect from just two observations, and the weaker habitat association in Pešćenica may instead be a result of the clinal structure rather than its cause. Further work is needed on measuring the strength of habitat preference outside the zone and on a formal model of the conditions necessary for sympatry after secondary contact.

It is also plausible that stronger habitat preferences evolved in Apahida in response to selection against toads migrating into the wrong habitat and producing unfit hybrids. Under this scenario, the toads in both areas would initially be the same, and only have evolved different levels of habitat preference once hybridisation began. Furthermore, reinforcement may occur more easily in a mosaic rather than a cline (Sanderson 1989; Cain et al. 1999), which might explain why the habitat

preference has not evolved in the same way in Pešćenica. One could explore this possibility by comparing habitat associations inside and outside both contact zones, although demonstrating that reinforcement has in fact taken place is much more difficult (Noor 1999).

The second potential cause of the differences between Apahida and Pešćenica is the spatial distribution of habitat. There is good evidence from the regressions of *H* onto spatial location (*x* and *y*) that there is a more clinal structure in occupied habitat in Pešćenica. However, this observation does not take into account the the extent to which the gradient in *occupied* habitat represents a gradient in *available* habitat. This is an awkward issue: data on available habitat are hard to collect for two reasons: firstly, habitats such as puddles or hoofprints are extremely ephemeral, and it is very difficult to quantify their abundance. Second, what might appear to a human observer to be a perfectly good site might in fact be unusable by a toad for any number of reasons. Furthermore, collecting a sufficient number of ecological variables to identify what exactly constitutes a usable yet unoccupied site would be very time consuming, and may well produce spurious relationships.

However, it is possible to make a few generalisations about the Pešćenica transect. Firstly, five years of field work in the area did not reveal any pond-like sites in the hills in the southwest of the transect. This may well explain the lack of *B. bombina* in this area. However, there are abundant puddles on the floodplain to the northeast (L. Kruuk, pers. comm.), and some of these are occupied by *B. bombina* (Fig. 2.5a,b). This raises two questions: why does *B. bombina* live in floodplain puddles but not in those in the hills, and what stops *B. variegata* from utilising the puddles in the floodplain? One possible hypothesis is that *B. bombina* cannot successfully reproduce in puddles, and the population is maintained by the productivity of the pond sites. This could be connected with their much longer larval period (Vorndran

et al. 2002), as puddles rapidly evaporate in the summer. The same factors might apply to the second question: since *B. variegata* does not use ponds, it cannot maintain a viable population in the exposed floodplain puddles, and is excluded from that side of the transect. These hypotheses can only be tested with an in-depth ecological survey of the sites available in the area, and by quantifying the reproductive success of each taxa in different habitat types.

Although there does not appear to be a gradient in the occupied habitat in Apahida, this is not a demonstration that all habitat types are equally available everywhere. In fact, there are three occupied pond sites to the north east of the study area and only one in the southwest, and this probably generates the slightly higher proportion of *B. bombina* in the northeast (Fig. 2.4). However, as with Pešćenica, the majority of occupied ponds in the area were sampled, although there were several other unoccupied fish ponds. The abundance of puddles also varies widely, mainly because the soil becomes much sandier in some places, and correspondingly there are also very few toads in these areas. For example, the south east corner of Figure 2.4a is almost entirely devoid of standing water. Nonetheless, unlike Pešćenica, there are no discernible patterns in puddle occupancy in the main belt of sites running southwest-northeast (pers. obs.). This may be because the area is slightly higher above sea level than Pešćenica, and puddles persist longer all over the study area. For Apahida at least, it seems that the occupied habitat probably reflects the distribution of available habitat, and there is no strong spatial trend in either. For Pešćenica, the matter is unresolved, and more ecological data need to be collected before the situation can be more fully understood.

If the toads are the same and the distribution of available habitat equivalent, Pešćenica and Apahida might also differ because they are either at different stages *en route* to the same equilibrium, or are headed towards alternative stable states. It

is certainly possible that the zones are of different ages, although both must have survived many generations of hybridisation. Alternative stable states may be possible if the formation of mosaic hybrid zones or a cline upon secondary contact is a chance event: some areas might spontaneously become mosaics and others steep clines, although this would require that all other circumstances were equal. More generally, not all the above possibilities are mutually exclusive, and the collection of more ecological data will be an important first step in determining their relative roles in *Bombina* hybrid zone structure.

The rate of hybridisation

As can be seen from Figure 2.6, there is a profound difference in the proportion of hybrid animals between Apahida and Pešćenica. In Apahida, although there appears to be a stronger habitat preference, pure *B. bombina* individuals (HI = 0) do appear in more intermediate sites (see Figure 2.6c). These toads are probably responsible for the high levels of introgression at our four markers into the surrounding *B. variegata* population. It is currently unclear why hybrid animals avoid ponds when they should carry some pond-favouring *B. bombina* alleles at their habitat preference loci. This is especially puzzling since, in the opposite direction, sites containing increasingly *B. variegata*-like toads occupy increasingly puddle-like habitats (Figure 2.7).

More generally, it is apparent that although habitat preference is stronger in Apahida than Pešćenica, it has not protected the *B. variegata* population from massive introgression. The persistence of pure individuals of both taxa in Pešćenica, despite the weaker habitat preference, is probably due to the influence of the extensive pure populations on either side of the hybrid zone, as these ensure a continual supply of pure animals. In Apahida, the effect of these parental populations is much weaker. This effect may also cause the significant excess

variance in the regressions of spatial location and habitat above, as migration into habitat patches from outside the zone may generate random changes in allele frequency (Nichols and Hewitt 1994). This is presumably happening much more often in Pešćenica than in Apahida, as the parental populations are much closer, and correspondingly there is more unexplained variance in Pešćenica (χ^2_{89} = 540.1 in Apahida, χ^2_{86} = 709.5 in Pešćenica). Another potential source of variance is the failure of the habitat axis to accurately summarise the toads habitat choice criteria, a problem which is hard to tackle without collecting large quantities of ecological data.

Overall, since our regressions of habitat score onto spatial location in the two countries show that the gradient in habitat type seen in Pešćenica is not found in Apahida, and that these could plausibly reflect the availability of habitat, I will tentatively conclude that the spatial arrangement of habitat contributes most strongly to the differences between the two hybrid zones. If this is the case, a greater interspersion of habitat types will inevitably lead to higher rates of introgression, and preserving pure populations without an influx of pure individuals from outside the zone is much harder. Given that the Apahida hybrid zone is a mosaic, what will happen to the neutral and selected divergence? Will habitat preference and selection sufficiently reduce gene flow between habitats to preserve divergence, or will the local patches collapse? In the next chapter I use the associations within and between loci to quantify the rate of dispersal between habitats, and explore how much selection is required to preserve divergence at both selected and neutral loci.

Section 2.5 Summary

1) The distribution of genotypes in Apahida (Romania) and Pešćenica (Croatia) is very different: in the former there is a mosaic of *B. bombina* in ponds and *B.variegata*-like hybrids in surrounding puddles and intermediate sites, whereas in the latter there is a smooth cline with some habitat differentiation at the centre.

2) I postulated that this is driven by the availability of the various habitat types, as all habitats are more or less available everywhere in Apahida, but there are no ponds in the hills in Pešćenica, and the floodplain puddles may be too ephemeral for *B. variegata*. This requires experimental confirmation.

3) Habitat preference is stronger in Apahida than Pešćenica: $\Delta \overline{p} = 0.30\ \Delta H$ and $\Delta \overline{p} = 0.16\ \Delta H$ respectively. This may either result from reinforcement of habitat preference in a mosaic hybrid zone, or conversely it could explain why the ranges of *B. bombina* and *B. variegata* can overlap more in Apahida.

4) There are many more hybrids in Apahida, and the rate of introgression may be higher despite the stronger habitat preference.

5) This observation implies that migration must be high between habitat types, and I explore the effects of this in Chapter 3.

Chapter 3:
Dispersal and selection in the Apahida mosaic hybrid zone

Section 3.1 Introduction

In Chapter 2 I described the *Bombina* mosaic hybrid zone around Apahida in Romania. I found that although there was probably a stronger habitat preference there than in Pešćenica, the zone contains a far greater proportion of hybrids. This is not unexpected in a mosaic, because the small populations on each patch can easily be swamped by migration. This effect is exacerbated by the intermingling of habitat patches in a mosaic, as dispersal in any direction can mean moving into the other habitat type. Chapter 2 used four marker loci to summarise the state of an individual's genome, and hence characterise the distribution of genotypes over the Apahida area. This chapter will focus instead on the associations between alleles within and between these loci, as these reveal information about migration and selection. Specifically, I focus on the concordance of loci, the heterozygote deficit and the linkage disequilibrium, and use the latter to quantify the migration between ponds and the surrounding intermediate habitats. This shows that migration is fairly strong (up to 20% of the adults may be immigrants from the other habitat), and this movement also explains the observed heterozygote deficit. I then consider the types of selection needed to counteract this migration and maintain the zone in its current state.

Associations between and within loci

As mentioned in Section 1.4, a high rate of dispersal will create associations between physically unlinked traits. In the case of neutral markers, these can be detected as linkage disequilibrium (see below), or alternatively by a consistent change in allele frequencies at all loci across the range of hybrid populations. The latter, termed concordance, is important in hybrid zones for two reasons: firstly, it indicates that dispersal is sufficiently strong to hold unlinked loci together, and

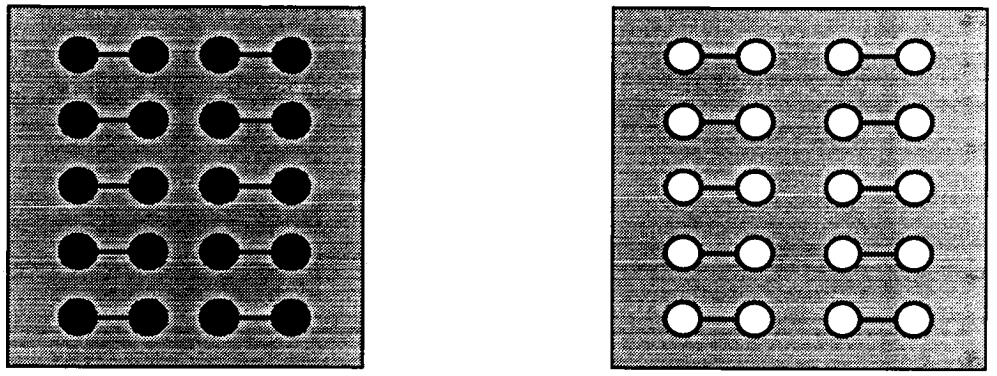
secondly it means that the forces acting on marker loci are similar, so that they can be treated as equivalent in subsequent analyses. Such patterns can also be generated by an environmental selection gradient acting equally throughout the entire genome, or by symmetrical epistasis between many sets of genes. However, neither of these is a plausible explanation for concordance between marker loci in natural populations.

When gametes unite at random, the genotypes at all loci are expected to conform to Hardy-Weinberg proportions. However, a wide range of processes can cause a deficit of heterozygotes, such as migration between populations with different allele frequencies (this is analagous to the Wahlund effect), assortative mating, and selection against hybrids (Hartl and Clark 1997). The linkage disequilibrium in hybrid zones (see below) ensures these processes are detectable with neutral markers, as individuals heterozygous at e.g. an incompatibility locus are more likely to be heterozygous at marker loci. Additionally, since Hardy-Weinberg equilibrium is restored by random mating, the stage of the life cycle at which a deficit is first detected can provide useful information on the dynamics of the zone. For example, a lack of heterozygotes in newly fertilised eggs implies assortative mating, and a lack of heterozygotes in only the adult stages implies either selection later on in the lifecycle (e.g. against metamorphs) or migration between habitats. I will test for heterozygote deficit in egg batches in Chapter 4, whereas here I focus on the adults.

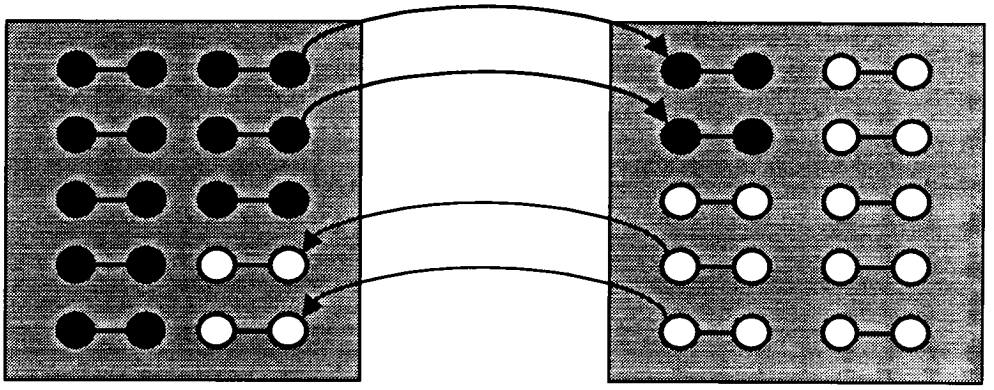
If two populations differ in allele frequency at more than one locus, a newly arrived migrant is more likely to carry alleles characteristic of its source population than its destination. Since these will be contained together in the same individual, alleles at different loci will be non-randomly associated (Figure 3.1). These associations are called linkage disequilibria (D). Associations will also exist between sets of alleles

at different loci if there is epistatic selection against their break-up, and between alleles involved in assortative mating. However, only dispersal is likely to generate genome-wide associations between unlinked loci, and this also easily overwhelms associations driven by other processes (Kruuk et al. 1999b). With every generation of random mating, the linkage disequilibrium between unlinked loci is expected to halve, and will eventually disappear unless new migrants arrive.

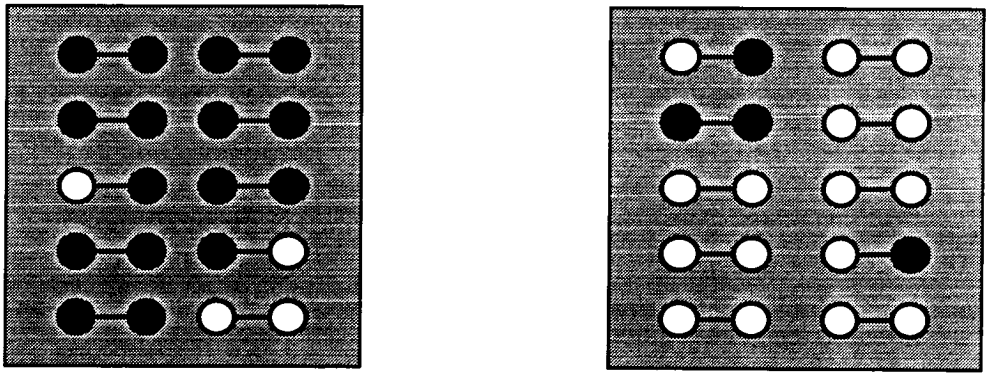
Figure 3.1: The generation of linkage disequilibrium by migration



1) Two haploid populations segregating for black or white alleles at two unlinked loci. Loci in the same individual are joined by a bar.



2) Two individuals in each population migrate, and now alleles at each locus are in complete association (individuals are either black at both loci or white at both loci).



3) After one round of random mating, half the associations are broken up.

The rate of immigration into a population can thus be calculated from the equilibrium level of D and the difference in allele frequency between the source and the destination; in general $D \sim m\,\Delta p^2 / r$, where r is the recombination rate (0.5 for unlinked loci) between the loci, Δp the difference in allele frequency and m the migration rate (Section 3.8). As the strength of habitat preference increases, the migrants between habitats become rarer. However, each individual contributes proportionally more to the linkage disequilibrium, because the difference in allele frequency between the source and the sink is maintained at a higher level (Kruuk 1997). High rates of migration between similar sites and low rates between more divergent ones can only be distinguished by examining associations between 3 or more loci, as these will be stronger in the latter situation (Barton 2000).

In the Polish transects, linkage disequilibrium was assumed to be generated by dispersal across the cline (Szymura and Barton 1986, 1991). Based on a diffusion model, it is expected to equal $D = \sigma^2 / r\,w^2$ when measured after dispersal and before selection. Here, w is the cline width, r the recombination rate and σ the mean dispersal distance between parent and offspring. In both Polish transects, this relationship gave estimates of $\sigma^2 \approx 1$ km gen^{-1} (Szymura and Barton 1991). In Croatia, the argument was complicated by the existence of a mosaic structure in the cline centre. The observed linkage disequilibria are then generated both by the immigration of pure toads from the periphery (i.e. the clinal component) and by dispersal between the two types of habitats in the centre (see Figure 2.5b). Migration at a rate m between nearby habitats which differ by Δp in marker frequency will generate an additional linkage disequilibrium $D = (c\,m\,\Delta p^2)/r$, with

a coefficient c that depends on detailed assumptions about the life cycle (see Section 3.8). The two sources cannot be disentangled from genetic data alone. Assuming (generously) that $m = 0.5$, the two estimates of $\Delta\overline{p}$ in Croatia, 0.15 and 0.25 (see Section 2.3) give estimates of D that can account for 27% and 66% of the observed maximum ($D = 0.094$ at $\overline{p} = 0.39$; MacCallum et al. 1998) respectively. The remaining D presumably arises from the cline itself.

Section 3.2 The maintenance of a mosaic structure

There is a considerable body of theory on how selection can maintain locally adaptive alleles within habitat patches, which I summarise below. I focus on those models where selection is against alleles in the wrong habitat rather than in the wrong genetic background, mainly because this has greater potential to counteract migration (Barton and Shpak 2000; see Section 3.5). For simplicity, I have divided the models into those that deal with discrete patches in a metapopulation, and those that deal with patches in a continuous environment. However, the basic structure and conclusions of the two model types are similar, and I discuss the circumstances under which either is appropriate below.

Maintaining a mosaic in a metapopulation

At a single locus, a locally adaptive allele a may be maintained indefinitely in a habitat patch despite immigration at rate m from populations fixed for allele A, at an equilibrium mA: $(s - m)\ a$, provided that $m < s$, where s is the selective advantage of allele a in that patch (Haldane 1932, p. 210). Both here and throughout this chapter m is defined as the proportion of adults in a population that were not born there. Haldane's result can be extended to a network of populations all equally connected by migration (e.g. Wright's island model: Wright 1931), where allele a is advantageous by s in a fraction of demes α and deleterious by $-\beta s$ everywhere else. In this case a will be preserved below a critical level of migration $m_{crit} < \beta s/(\beta(1-\alpha)-\alpha)$, or alternatively, if it is favoured in a critical proportion of populations $\alpha_{crit} = (1 - s/m)/(1 + 1/\beta)$ (Barton and Whitlock 1997). In addition,

immigration from pure populations outside the area of secondary contact into their respective habitats can also sustain local patches (Nichols and Hewitt 1994).

The conditions for the maintenance of divergence in a metapopulation become more restrictive when more than one locus is involved. This is because each maladaptive allele that arrives into the population requires a selective death to remove it, and so the number of deaths required increases in proportion to mL, where L is the number of loci at which alleles adaptive in either habitat are carried. Even a small m can thus impose a high migration load on the population, and local adaptation relying on a number of loci can easily be overwhelmed by migration (see Barton 1992; Lythgoe 1997). However, this assumes that the incoming loci are strictly independent, which is unrealistic. For example, consider a single immigrant homozygous for locally maladaptive alleles at four loci: if selection eliminates this individual before it can breed, then eight selective deaths have been accounted for. Furthermore, since associations between unlinked loci approximately halve every generation, the loci will not be acting independently for ~3 generations, and selective deaths in this time will on average remove more than one maladaptive allele. Whether or not locally adaptive alleles will be maintained in a patch is thus based on the relationship between the total selection across all loci contributing to immigrant fitness, and the recombination between them, as these together determine the fate of immigrant alleles in the first few generations (Barton and Bengtsson 1986).

Maintaining a mosaic in a continuous environment

If instead the populations are spread across a continuous environment, a locally adaptive allele can be maintained in a habitat patch if the patch is larger than $l = \sigma / \sqrt{s}$ (Slatkin 1973), where σ is the dispersal distance and s is the selective

advantage of the allele (Figure 3.2a). Roughly, the patch has to be wide enough so that the clines in allele frequency on either border do not meet in the middle, such that the locally adaptive allele is fixed in the centre of the patch. In a similar way, if the allele is disadvantageous by $-\beta s$ outside this patch, it can be preserved if the patch is wider than $2\tan^{-1}\left(\sqrt{\beta}\right)l$ (Nagylaki 1975). Local adaptations based instead on a quantitative trait can be maintained when l is greater than $\sigma\sqrt{V_s/2V_g}$, where V_s is a measure of stabilising selection on the trait and V_g is the additive genetic variance (Slatkin 1978). The effect of an increasing number of loci is as above, in that the local adaptation is more likely to be preserved if the total selection can eliminate the immigrant alleles before they have to be individually removed by separate selective deaths. The barrier to gene flow generated by the total selection on a large number of loci is modelled by Barton and Shpak (2000).

Figure 3.2a: The maintenance of a habitat patch. Following Slatkin (1973), l is the width of the patch, s is the selection for the locally favoured allele, and σ is dispersal.

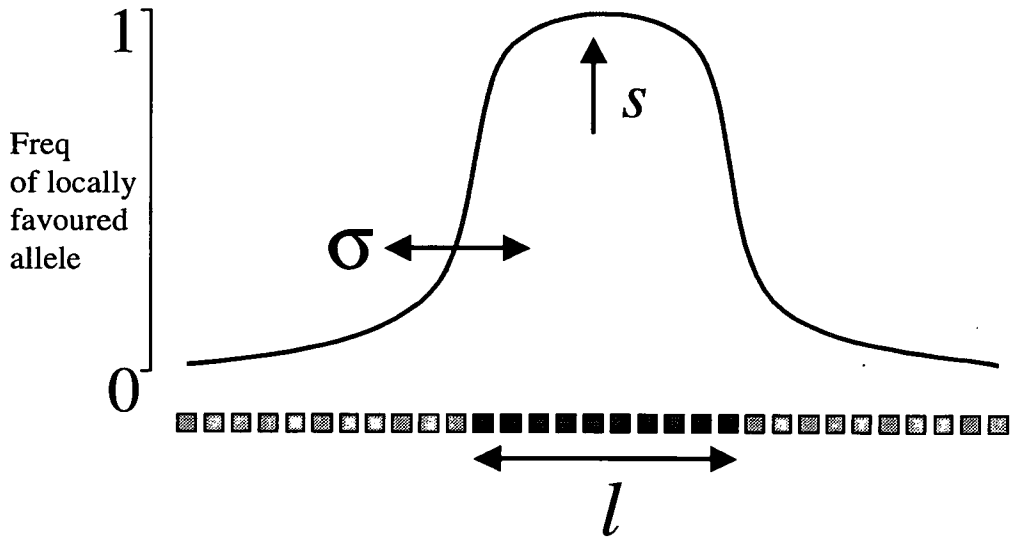
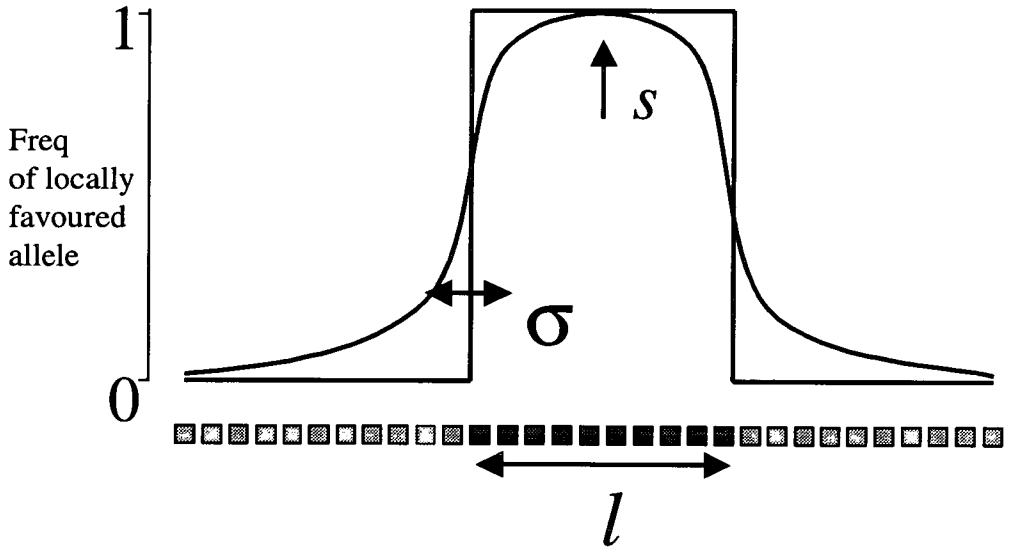


Figure 3.2b: The effect of a strong habitat preference on the maintenance of a favoured allele: migration is reduced across the patch boundary and the boundary is sharper..



The results for continuous populations are based on the diffusion approximation, which assumes the organisms disperse at random (Nagylaki 1975). When instead individuals show a preference for the habitat patch to which they are adapted, locally adaptive alleles can be then maintained on smaller spatial scales, provided that migration between alternative habitats is sufficiently reduced (Figure 3.2b). Habitat preference thus plays a similar role in preserving fine-scaled mosaic hybrid zones as it does in sympatric divergence (Rice 1987). The continuous and metapopulation models then become roughly analogous, as there are now two types of dispersal: within a habitat patch, and between different habitats. In metapopulation models, the former is ignored, and the latter is measured as migration.

When is either model appropriate?

Mosaic hybrid zones generated by habitat associations are probably best described by either a discrete patch or a continuous model, depending on the type of habitat divergence involved. For example, a mosaic based on contiguous habitat patches may well have clines maintained by dispersal and selection at the boundaries (e.g. *Gryllus* crickets: Ross and Arnold 2002, see Section 2.1). When habitat types are more discrete (e.g. waterbodies) such clines will not appear, as migrants can only meet and hybridise within the patches and not at the boundary (e.g. *Limnoporus* waterstriders: Klingenberg et al. 2000). In *Bombina*, it is not clear which model is appropriate. Toads are easiest to sample in their aquatic habitat, but they are also partly terrestrial, especially in winter. Additionally, with the exception of large

ponds, individual waterbodies are usually too small to be considered discrete populations, since they are short-lived and often part of a larger complex of sites. The diffusion approximation fitted well in Poland, because the hybrid zone followed a smooth cline, and there were only small differences in allele frequency between adjacent habitats (Szymura and Barton 1986, 1991). In this case, diffusion can apply equally to discrete sites or a continuous population (Nagylaki 1975). In both Croatia (MacCallum 1998) and Apahida (Chapter 2), there is an active habitat preference, which can generate large allele frequency differences between adjacent habitats. In both these cases, the migration rate between the different habitat types is also important, and may not be a function of the distance between them. The theory necessary to describe mosaic hybrid zones may thus require features from both continuous and metapopulation models (e.g. Barton and Gale 1993; Barton and Whitlock 1997).

How can the Apahida mosaic be preserved?

The mosaic structure in Apahida requires strong isolating mechanisms to maintain it for two reasons. Firstly, migration between the small populations in different habitat patches has a proportionally stronger effect on allele frequencies than when populations are large (i.e. there is a higher migrational load). Second, allele frequency differences are not replenished by the immigration of pure individuals into their respective habitats from outside the zone, as the nearest pure populations are some distance away (Chapter 2). Since the mosaic appears to be generated by the association of *B. bombina* with ponds and *B. variegata* with temporary sites (Chapter 2), its preservation requires that the effective rate of migration between the two habitat types must be very small. The principal way to achieve this is a low

rate of movement between habitats, which in the case of *Bombina* must be achieved by habitat preference, as the toads can ordinarily travel over large distances (Szymura and Barton 1991; MacCallum 1998; Kruuk 1997).

In addition, effective migration can be reduced below the level of migration by assortative mating and the selective removal of immigrant alleles after their arrival. (I loosely term these within-habitat isolating mechanisms). For example, even if toads do move frequently between habitats, they may choose only to breed in their respective habitat type; I attempt to quantify this for *Bombina* in Chapters 4 and 5. Beyond this, immigrant toads may have a reduced mating success due to mate discrimination, although the strength of this may depend on the sex of the immigrant (see Section 1.2). As fertilisation is external in toads, it is harder for either sex to manipulate fertilisation success, and hence there is less opportunity for divergence driven by reproductive conflict. However, it is possible that some sort of sperm-egg recognition traits may have diverged between the two taxa. Even if matings do occur, there may be ecological selection against hybrids or against incompatible allele combinations. Lastly, hybrids may suffer a reduced mating success. I attempt to quantify both assortative mating and breeding site choice in *Bombina* in Chapters 4 and 5. The aim of this chapter is to measure how much migration there is between habitats using the associations between loci, and hence how much more within-habitat isolation is required for the mosaic structure to be maintained.

Section 3.3 Statistical techniques

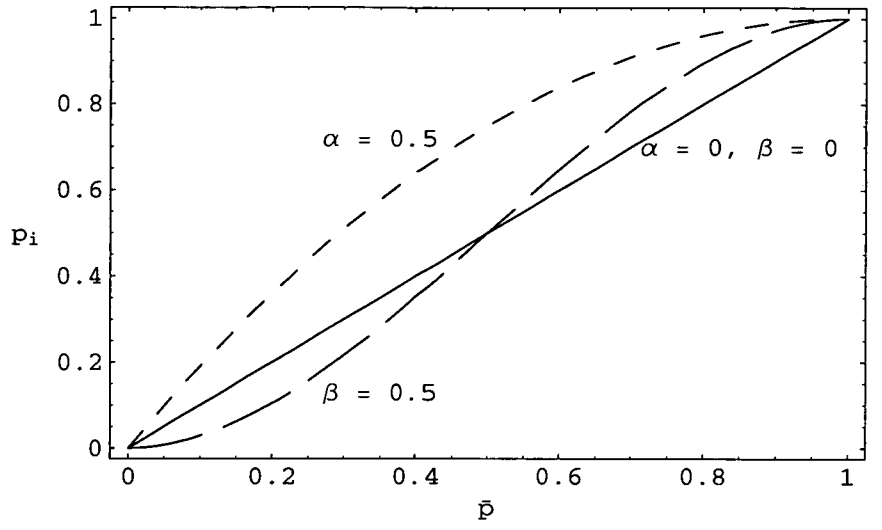
Concordance of loci

If allele frequencies at unlinked marker loci change in unison through a hybrid zone, this implies that dispersal is sufficiently strong in relation to recombination to slow the break up of the parental genomes (see Section 3.1). If recombination is stronger, the allele frequencies at marker loci will be affected more by random drift and selection on genes close by on the chromosome. In a clinal hybrid zone, allele frequencies change along a transect through the zone, but in a mosaic they change from one habitat type to the other. Concordance is quantified with a cubic polynomial model:

$$p_i = \overline{p} + 2\,\overline{p}\,\overline{q}[\alpha_i + \beta_i(\overline{p} - \overline{q})] \qquad [3.1]$$

where p_i is the frequency of *variegata* alleles at the i'th locus in a site and $\overline{p}$ and $\overline{q}$ (= 1-$\overline{p}$) are the average frequencies across all loci for that site (Szymura and Barton 1991). This formula was chosen because necessarily, $p_i = 0$ when $\overline{p} = 0$, and $p_i = 1$ when $\overline{p} = 1$. In a cline, the parameter α_i measures the coincidence of the cline centres at each locus, in a mosaic a positive α represents consistent deviations across the allele frequency range in favour of *B. variegata* (Fig. 3.3). This might arise because that marker locus was linked to a *B. variegata* gene advantageous in the majority of sites, which increases its frequency relative to the other loci.

Figure 3.3: The effects of the model parameters α (coincidence) and β (concordance) on how the allele frequency at a locus p_i changes in relation to $\overline{p}$, the mean frequency across loci (see equation 3.1)



A positive β_i indicates that there is a greater difference in p_i between sites at either end of the allele frequency spectrum compared to $\overline{p}$. In the context of a cline this corresponds to a steepening of the cline away from concordance with other loci; here, it indicates sharper divergence at a locus relative to the average (Fig. 3.3). In either case, a significant positive β_i might be due to strong selection against hybrids at a linked gene. For both α and β, significant values imply that recombination is sufficiently frequent between the marker loci to allow them to follow the selection pressures of whatever genes are nearby on the chromosome. Random deviations away from complete concordance are quantified using the standardised variance in allele frequency F_{ST} (cf. MacCallum et al. 1998); a high level of F_{ST} would indicate that drift within sites was a major determinant of allele frequencies at each locus. The model described by equation [3.1] above was fitted using the Macintosh software package Analyse (by N.H. Barton and S.J.E. Baird).

Heterozygote deficit and linkage disequilibrium

Maximum likelihood techniques were used to estimate heterozygote deficits and linkage disequilibria (Edwards 1972; Hill 1974; see Szymura and Barton (1991) and MacCallum et al. (1998) for details). All calculations were performed in Analyse. Briefly, the goodness-of-fit of an estimate for either heterozygote deficit (F_{IS}) or standardised linkage disequilibrium (R, see below) is given by the natural logarithm of its likelihood (LogL). The difference between two LogL estimates is distributed as $\chi^2/2$ with 1 degree of freedom in large samples, provided that the estimate is not at the boundary. Confidence limits with one degree of freedom around the most likely parameter value contain all those values lying within 2 log likelihood units of the maximum, as the probability of obtaining $\frac{1}{2}\chi_1^2 \geq 4$ is around 5%.

Heterogeneity between sites in F_{IS} can be assessed by comparing log likelihoods when F_{IS} is held constant or allowed to vary between them. The same approach can also detect differences between loci, although linkage disequilibrium makes loci behave non-independently. Following MacCallum (1994), I account for heterozygote deficit when estimating linkage disequilibrium by assuming for a pair of loci that a proportion F_{IS} of the offspring is produced by union of identical gametes, and is thus homozygous. The remainder of gametes, 1 - F_{IS}, unite at random. More complex analyses can estimate all the possible associations within and between loci, but these require larger sample sizes and are time consuming for a large number of sites (Barton 2000; Chapter 5). I standardise pairwise linkage

disequilibrium (D) by the allele frequencies to give $R_{ij} = D_{ij} / \sqrt{p_i\, q_i\, p_j\, q_j}$. This facilitates comparisons between sites with different allele frequencies, although this standardisation does not fully remove the effect of allele frequencies: the full range of R values (-1 to 1) is only possible when $p = q = 0.5$ (Lewontin 1988).

Section 3.4 Results

Concordance of allele frequencies

The four marker loci change approximately in unison across the allele frequency spectrum (Fig. 3.4). This justifies treating them as equivalent in subsequent analyses. The estimates of α and β for each locus and their support limits are given in Table 3.1. Only three of the eight estimates are significant, the largest deviation is at locus *Bv24.11*, where α = -0.08, indicating a consistent excess of *B. bombina* alleles compared to other loci. Locus *Bv24.11* also shows the largest difference in allele frequency between either end of the spectrum, with β = -0.16. It is possible that these are generated by scoring errors at one or more loci, although there was no evidence of this in numerous gel re-runs. None of the deviations from complete concordance are larger than the largest of those found for allozyme loci in Croatia (α = 0.16; β = -0.29, MacCallum et al. 1998) indicating that there is roughly the same level of concordance in the the two areas. In terms of random deviations between sites around the fitted concordance, the average F_{ST} across all sites and loci is 0.033, somewhat larger than the equivalent measure in Croatia for allozyme markers (0.0068), indicating that allele frequencies in Apahida are also slightly affected by drift. In general, however, loci more or less change together across the entire range of sites, and hence are treated as equivalent in the following analyses. This also confirms that they are a good summary of a toad's genome (c.f. Chapter 2).

Figure 3.4: Concordance of allele frequency across loci. Each graph shows the allele frequency at each of the four loci (p_i) plotted against the mean across those loci ($\overline{p}$). The smooth curves show a cubic polynomial regression fitted by the maximum likelihood (see text). The α and β for each regression are given in Table 3.1.

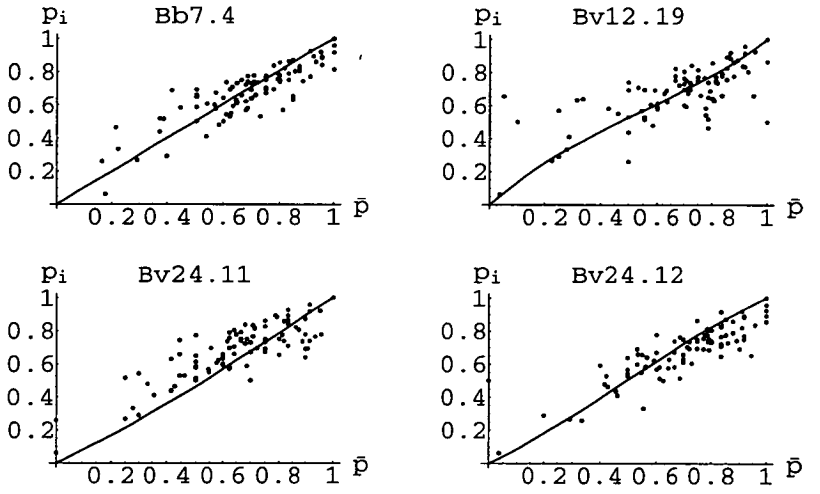


Table 3.1: Patterns at diagnostic loci. Discordance between loci are quantified by α and β (see text and Fig. 3.3). F_{ST} gives the standardised variance in allele frequency due to genetic drift, and F_{IS} is an estimate of the heterozygote deficit. Support limits are given in parentheses and values significantly different from 0 (where ΔlogL >2) are in bold. No correction is made for multiple comparisons.

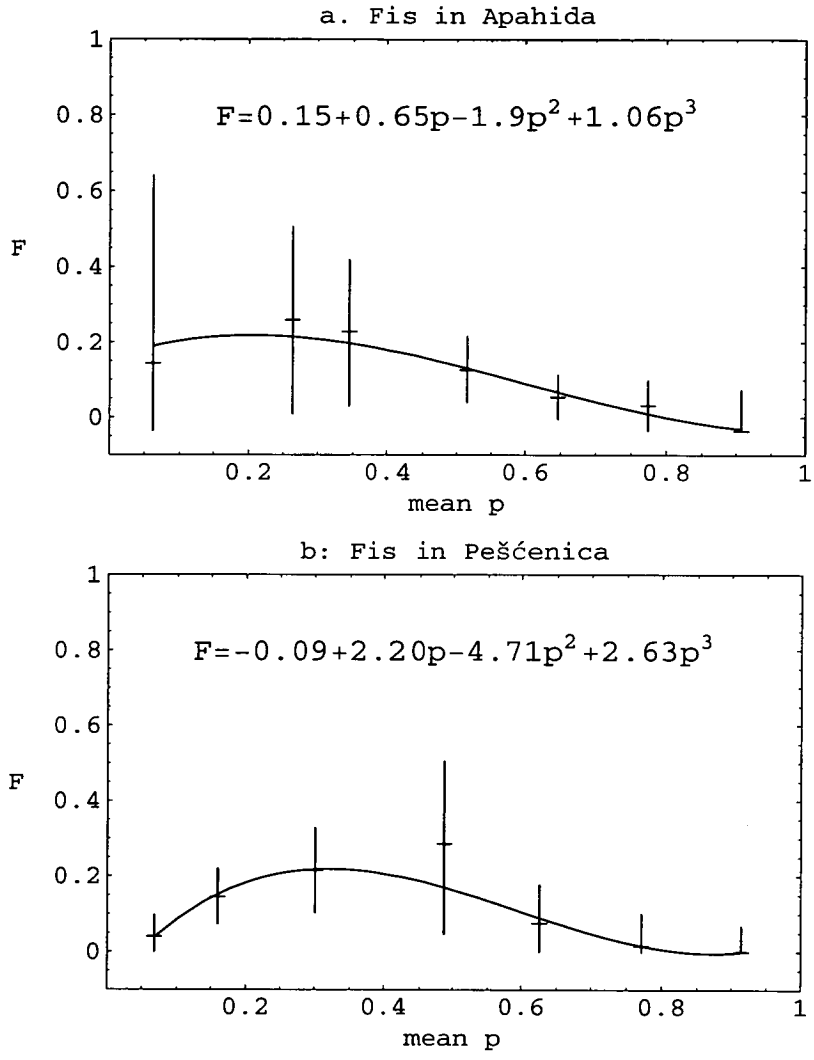
Locus	α	β	F_{ST}	F_{IS}
Bb7.4	0.028 (−0.01, 0.08)	0.041 (−0.08, 0.16)	0.00	**0.016** (0.09, 0.24)
Bv12.19	0.004 (−0.03, 0.03)	−0.025 (−0.15, 0.09)	**0.048**	**0.085** (0.01, 0.15)
Bv24.11	**−0.08** (−0.12, −0.04)	**−0.16** (−0.28, −0.03)	**0.022**	0.011 (−0.03, 0.08)
Bv24.12	0.052 (0.00, 0.09)	**0.15** (0.02, 0.26)	−0.001	−0.011 (−0.03, 0.06)
Overall			**0.033**	**0.06** (0.02, 0.10)

Heterozygote deficit and linkage disequilibria

Pooling all sites, there is a significant deficit of heterozygotes at loci *Bb7.4* and *Bv12.19*; furthermore, there is a significant deficit when F_{IS} is constrained to be the same across all loci (Table 3.1). There is also significant heterogeneity between loci ($\Delta L_3 = 7.63$, $p < 10^{-5}$). The higher deficits at *Bb7.4* and *Bv12.19* are not associated with high levels of β, which might have indicated stronger selection on those loci. It is possible that the higher deficit at *Bv12.19* is due to remaining scoring errors (see Chapter 2). We divided the sites into seven groups by their mean allele frequency and obtained F_{IS} estimates both singly and across all loci within each group. Considered individually, all four loci show a greater heterozygote deficit in more *B. bombina*-like sites (data not shown).

When F_{IS} is estimated as a common value across all loci, heterozygote deficit peaks in sites on the *B. bombina* end of the allele frequency spectrum (max. F_{IS} = 0.22 at $\overline{p} = 0.21$; Fig. 3.5a). Since there are few sites in the three left hand groups (n = 1, 2 and 3 respectively), this asymmetry does not result in a better fit for a cubic model ($F_{3,6} = 1.34$, $P = 0.33$) over a simple linear one ($F_{1,6} = 52.05$, $P = 0.005$) when the means are weighted by the support limits. However, the means for each bin show a similar asymmetric pattern to that seen in Pešćenica (MacCallum et al. 1998; my Fig. 3.5b), where the maximum F_{IS} is 0.23 at $\overline{p} = 0.32$ when a weighted regression is used (the equivalent regressions in MacCallum et al. 1998 are unweighted, the cubic fit is the best for their data with either method). I show the cubic curve in Figure 3.5a for comparison with the Croatian data.

Figure 3.5: F_{IS} for binned sites estimated across all loci. For each group, the point is the maximum likelihood estimate and the support limits are values within two log likelihood units. The curves are cubic regressions on the overall data, fitted by least squares, the regression equation is given in each plot. The figures are a) Apahida and b) Pešćenica (data from MacCallum 1994).



The standardised linkage disequilibria between loci, $R_{ij} = D_{ij} / \sqrt{p_i\, q_i\, p_j\, q_j}$ were estimated across all sites. In these computations, the non-zero estimates of F_{IS} were taken into account in order to remove any undue inflation of disequilibrium through correlations of genes within loci (see Methods). Table 3.2 gives values of R estimated across all sites for each pair of loci. Across all sites and loci, $R = 0.090$

(support limits: 0.083, 0.097). This indicates that combinations of genes found in the parental taxa are in excess, despite their constant break-up by recombination. There is heterogeneity between loci: the reduction in log likelihood of $\Delta L_5 = 13.2$ when the R_{ij} are forced to be the same is significant when compared with the asymptotic χ^2_5 ($P < 10^{-5}$). This is striking, as the observed strong linkage disequilibrium is expected to make loci more similar. From Table 3.2, it appears that the associations involving locus *Bv12.19* are consistently lower than the others, and this may be causing the heterogeneity. However, beyond noting that the conclusions drawn from these analyses may be affected by this heterogeneity, there is little that can be done to account for this problem.

To examine how linkage disequilibrium changes with allele frequency, R was estimated for all loci within each of the seven groups of sites used above (Fig. 3.6). Again, the maximum estimate of $R = 0.38$ at $\overline{p} = 0.28$ was shifted towards *B. bombina*-like hybrid populations (Fig. 3.6a), as in Croatia (there $R = 0.39$ at $\overline{p} = 0.39$; Fig. 3.6b). However, like F_{IS} above, the cubic regression ($F_{3,6} = 4.98$, $P = 0.11$) did not give a significantly better fit than a linear regression ($F_{1,6} = 18.59$, $P = 0.02$). Again, I show the cubic curve fit for comparison with the Croatian plot.

Figure 3.6: *R* for binned sites estimated across all loci. For each group, the point is the maximum likelihood estimate and the support limits are values within two log likelihood units. The curves are cubic regressions on the overall data, fitted by least squares, the regression equation is shown in each plot. The figures are for a) Apahida and b) Pešćenica (data from MacCallum 1994).

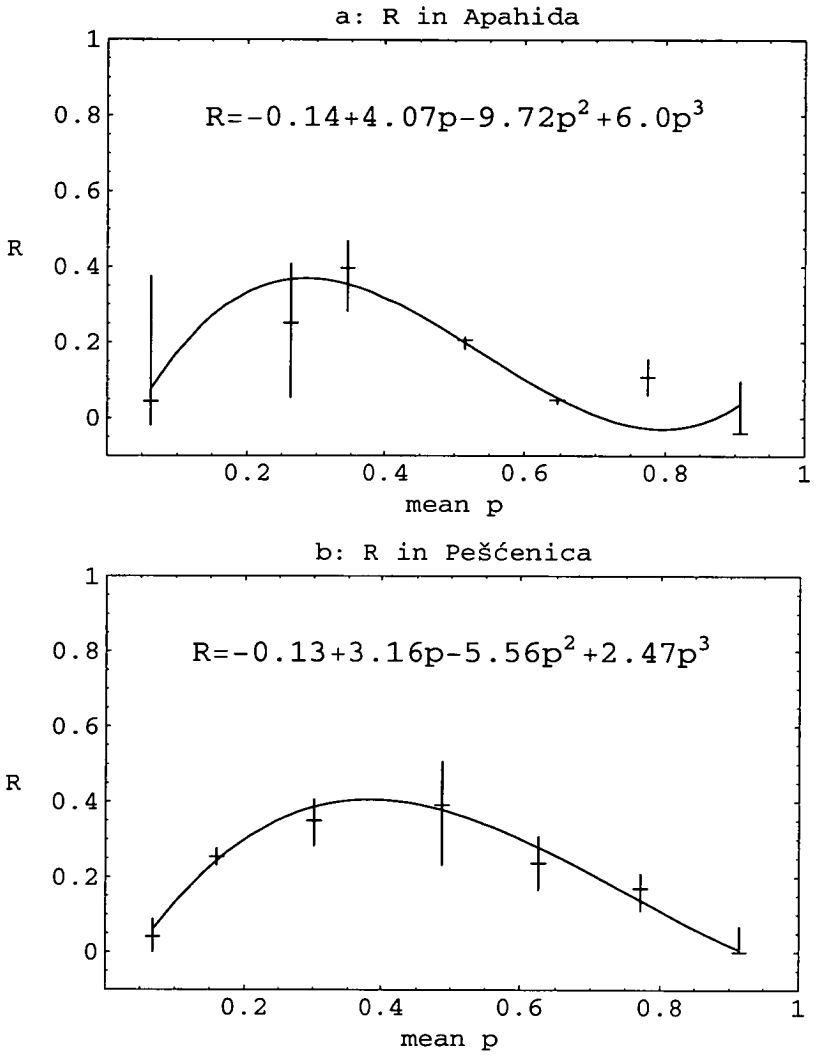
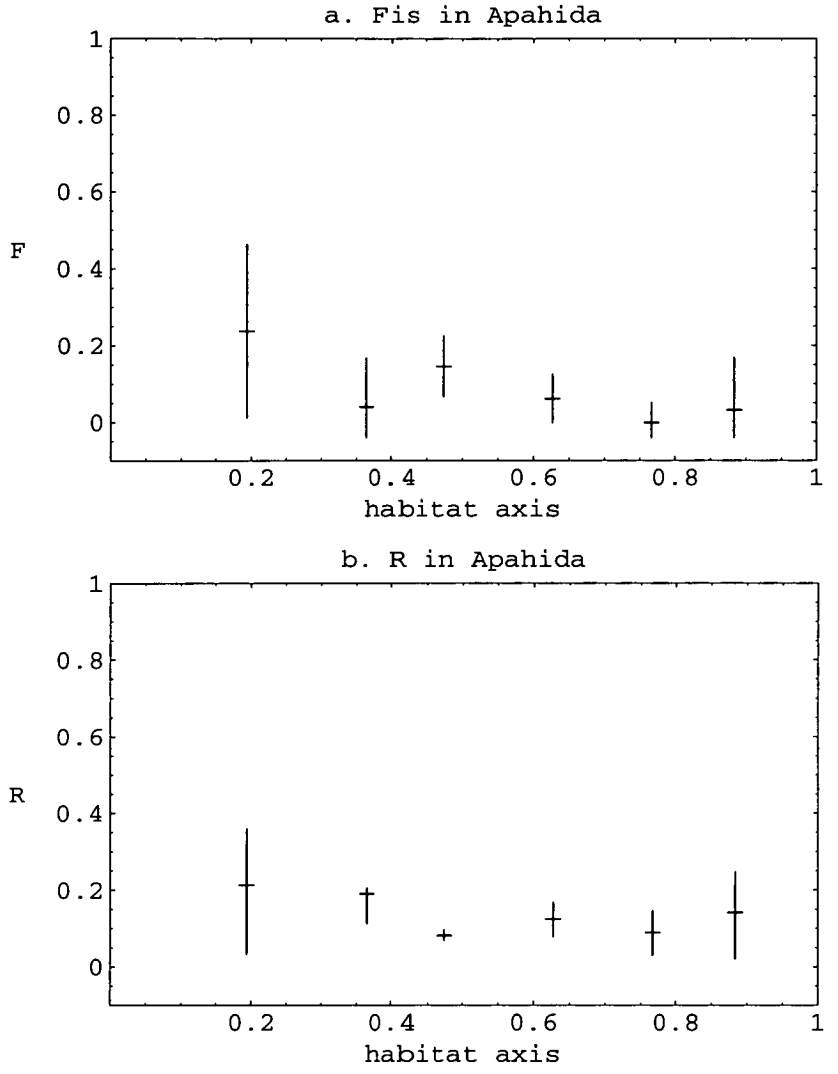


Table 3.2: Pairwise standardised linkage disequilibria across all sites. The maximum likelihood value of *R* for each pair of loci is given in the main body of the table, with support limits in parentheses. The value when R is constrained to be the same across all sites and loci is given at the bottom of the right hand column, and the remainder of this column contains the average for each locus. The *R* value for all pairs of loci were significantly greater than zero (i.e. ΔlogL >2 for all).

	Bb7.4	*Bv12.19*	*Bv24.11*	Average
Bb7.4				0.141
Bv12.19	0.072 (0.04, 0.08)			0.066
Bv24.11	0.204 (0.13, 0.26)	0.091 (0.05, 0.10)		0.146
Bv24.12	0.147 (0.07, 0.21)	0.036 (0.01, 0.04)	0.143 (0.07, 0.21)	0.108
				0.090

Finally, I grouped the sites according to their habitat axis score and repeated the estimation of F_{IS} and *R* for each bin (Figure 3.7). This gave a less definite pattern, although in general both parameters seem to increase towards the *B. bombina* end of the allele frequency spectrum. Once again, the wide support limits at this end reflect the low number on this side: the left-hand bar is based on only three sites. The cubic regression analysis (see above) was not repeated for these data, as there is no clear pattern to quantify.

Figure 3.7: Estimates across all loci for sites binned by habitat axis for a. F_{IS} and b. *R*. For each group, the value and the 2-unit support limits were obtained using maximum likelihood.



To summarise, there is substantial concordance in allele frequencies at all four marker loci, of a similar order to the four allozyme loci in Pešćenica (MacCallum et al. 1998). Since this implied that the forces acting on each locus were similar, they were treated as equivalent in the subsequent analyses. Apahida and Pešćenica were also broadly similar in their genetic structure: when sites were pooled by their allele frequency, the maximum heterozygote deficit was F_{IS}= 0.23 in Croatia and 0.21 in Romania (Fig. 3.5). The analogous maxima for the standardized linkage

disequilibrium, R, were 0.40 and 0.38, respectively (Fig. 3.6). Moreover, both parameters showed an asymmetric pattern in both regions, being stronger in sites with more *B. bombina*-like toads. Lastly, binning the Apahida sites by the habitat axis gave no clear pattern in either parameter, although there may be a slight upwards trend on the *B. bombina* side (Fig. 3.7). I consider the implications of these estimates below.

Section 3.5 Estimating migration

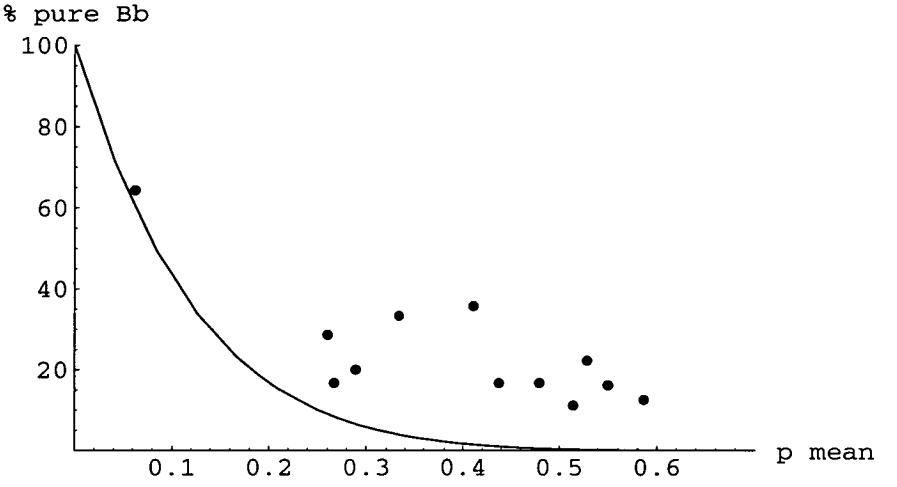
The high levels of linkage disequilibria observed in Apahida show that there is strong mixing between two genetically distinct populations. Furthermore, I found in Chapter 2 that there is a high level of introgression of *B. bombina* alleles into the *B. variegata* population (at least at marker loci). These two observations imply that divergence between the two taxa in the Apahida area may be collapsing rapidly, as not only is there a high rate of movement between habitat types, but the immigrants also appear to contribute to subsequent generations. I now estimate the migration rate of *B. bombina* individuals out of ponds and into the surrounding intermediate sites from the observed linkage disequilibrium, then in Section 3.6 explore how selection must act to preserve divergence at either selected or linked neutral loci in the area.

Estimating the rate of movement by pure B. bombina

Strong linkage disequilibria are seen in *B. bombina*-like hybrid populations and this appears to be caused by the immigration of pure *B. bombina* from ponds. This is suggested by the gap in the distribution of population means between *B. bombina* and hybrid populations (Fig. 2.7) and by the observation of 'pure' (HI = 0) *B. bombina* adults in temporary sites (with $\overline{p}$ as high as 0.59: see Appendix I). Even in hybrid populations with $\overline{p} = 0.2$, these toads are unlikely to have been generated locally (Fig. 3.8) and so can be assumed to be newly arrived immigrants. Their occurrence is determined more by habitat than by the proximity to ponds: the sites in question have a significantly more pond-like habitat score than the remainder of the temporary sites ($t = 4.87$, $p < 0.001$), yet the mean distance to the large pond (site 293) or to the edge of the study area (whichever is smaller) is 2.78 km. The

latter figure is clearly a minimum bound for the actual dispersal distance of these individuals, because several of the sites in question were at the periphery of the study area. In the following, I concentrate on intermediate sites with $0.2 < \overline{p} < 0.6$ and determine for each one of them the immigration rate m of pure *B. bombina* individuals that would generate the locally observed linkage disequilibrium. I then use the mean $\overline{m}$ of these estimates to ask whether plausible strengths of selection could hold this hybrid zone in migration-selection balance.

Figure 3.8: The percentage of toads with pure *B. bombina* marker genotypes as a function of $\overline{p}$ in each site. The line is $(1 - p)^8$, the expected frequency of animals with HI = 0 under Hardy Weinberg proportions.



Since these parameter estimates are based on equilibrium relationships, the demography of the toad populations must be considered. Aside from the large pond (site 293), all sites probably dry up before the end of the summer in most years. Nevertheless, many of them will contain water at the beginning of the breeding season and can thus attract toads for many consecutive years. Judging by their typically higher percentage of aquatic vegetation, the intermediate sites are in a more advanced stage of succession and thus older or less temporary than the remaining sites with $\overline{p} > 0.6$. As pairwise linkage disequilibrium reaches an equilibrium rapidly ($D_{t+1} = 0.5\ D_t$ for unlinked loci), assuming that D has

equilibrated in the sites with $0.2 < \overline{p} < 0.6$ is justifiable, provided they are older than ~10 generations.

From mark-recapture studies in Croatia and in Romania (MacCallum et al. 1998; Nürnberger et al. 2002a; T. Sands, unpublished data), it appears that the total pool of adults that visit a site per year is several times larger than the number present on a given sampling date, and recapture rates across years suggest population sizes of up to several hundred individuals (T. Sands, unpublished data). Furthermore, the adults sampled at a given site are generally representative of those animals that reproduce there, as is seen by a high correlation between allele frequency in adults and eggs (S. Köhler, unpublished data). I assume in the following that the composition of the local adult assemblage is approximately stable on the scale of ~10 generations and that its size is large enough so that deterministic theory can be applied.

In general, migration at a rate m between two populations that differ in allele frequency by $\Delta\overline{p}$ at each of two loci generates pairwise linkage disequilibrium proportional to $(m\,\Delta\,\overline{p}^2)/r$, where r is the recombination rate between the two loci (see Section 3.8). Here, I consider the immigration from pond populations with $\overline{p} \approx 0$, so that $\Delta\overline{p}_i$ for a given intermediate site i is roughly equal to $\overline{p}_i$. In this model (and in the field), linkage disequilibrium is measured after migration and before reproduction. Moreover, I assume that after random mating, strong selection on the offspring returns the $\Delta\overline{p}_i$ to its original value. Then, m_i can be estimated from $D_i = (m_i\,\Delta\,\overline{p}_i{}^2)/r$. For the 23 populations in the range $0.2 < \overline{p} < 0.6$ this yields an average immigration rate $\overline{m} = 0.19$ (s.d. 0.19). Details of the computations are given in Section 3.8, where it is shown that the expression for D_i is a reasonable approximation for this range of allele frequencies and migration rates. The mean F_{IS} in these sites is 0.17; in this model $F_{IS} = m\,p\,/\,q(1\text{-}m)$, which gives 0.16 for $m =$

0.19 and $p = 0.4$. The migration rate inferred from linkage disequilibrium can thus also account for the observed deficit of heterozygotes.

The computation of a single migration rate between two subgroups of populations in Apahida is clearly an oversimplification. For example, actual observations of adults with HI = 0 make up 11% of the samples in intermediate sites, which is about half of the proportion needed to account for the observed linkage disequilibria. The remainder may come from the dispersal of recombinants: disequilibrium in the nine sites with $0.55 < p < 0.6$ can be accounted for by immigration of $\overline{m} = 0.26$ from sites with $\overline{p} = 0.2$. In fact, animals with HI = 0 are seen only in the more pond-like of the intermediate sites ($t = 2.22$, $p = 0.038$). These observations suggest that *B. bombina* alleles are entering the *B. variegata* gene pool along a habitat gradient rather than a spatial cline. Moreover if the offspring of pure *B. bombina* immigrants are contributing to linkage disequilibria in more *B. variegata*-like populations, there can only be a few rounds of recombination before these recombinants reach sites with $\overline{p} > 0.6$. Nonetheless, I stick to the earlier estimate of $\overline{m} = 0.19$ into intermediate sites in the analyses in Section 3.6.

Section 3.6 Preserving divergence in the Apahida mosaic

If, for simplicity, we divide the genome into neutral and selected loci, the question about the stability of this hybrid zone has two parts: a) will differentiation be maintained with respect to those traits that mediate differential adaptation (including habitat preference), and b) for how long will frequency differences at marker loci persist?

Will adaptive differences be preserved?

For a single locus, a locally adapted allele will be maintained if the selection favouring it is stronger than the rate of migration (s > m; see Section 3.2). As the number of selected loci increases, linkage disequilibria build up, and both the per locus selection (s) and the total selection across n loci (with additive selection $S = ns$) determine the dynamics of the hybrid zone (Barton 1983). Moreover, the question of stability depends on precisely how selection acts. Since it is likely that the many differences between *B. bombina* and *B. variegata* are based on many genes, I consider two models in which selection is either purely against hybrids or acts against alleles in the wrong habitat.

As a first approximation, the influx of *B. bombina* alleles into populations with intermediate allele frequency (cf. m = 0.2 into sites with $\overline{p} > 0$, see above) is assumed to be balanced by selection in these sites. Following Barton and Shpak (2000), selection against hybrids is modelled such that the fitness of an individual with a fraction p (=1-q) of *B. variegata* alleles is Exp[$-4\,Spq$]. Then, the critical migration rate above which migration swamps selection is $m \sim S/4n$ when selection is weak, and the population is close to linkage equilibrium (Fig. 3.9a; N. Barton, pers. comm.). However, as selection gets stronger, the critical migration

rate becomes almost independent of the number of loci, and the frequency of *B. variegata* alleles falls from p to $(1 - m)\,p / (1 - 2\,pq)$ over a generation (N. Barton, pers. comm.), simply because some matings must involve *B. bombina* individuals and these leave a small proportion of the offspring. Thus, even if selection is so strong that no hybrids survive, migration rates high enough to account for the observed linkage disequilibria will swamp local differences, simply because some *B. variegata* are involved in matings with immigrant *B. bombina*.

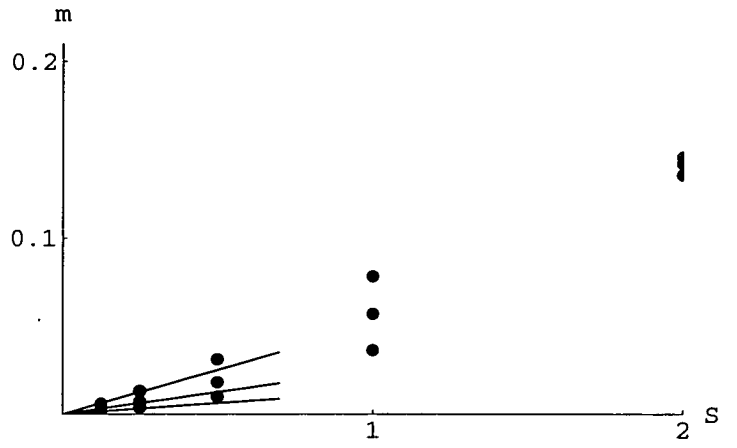
In contrast, if selection favours genotypes that are adapted to the local environment, selection is able to counter higher rates of immigration, simply because the immigrant genotypes have low fitness (Fig. 3.9b). In this case, selection acts before the production of unfit hybrids, and hence the migrational load on the local population is much lower. With fitness Exp[- $2Sq$], arbitrarily strong migration can be balanced by strong selection; for example, $m = 0.2$ can be countered by selection on 20 loci with $S \sim 1.7$, which implies a fitness of immigrant pure genotypes ($q = 1$) of Exp[-$2S$] = 3.3%, and of F1's of 18% (N. Barton, pers. comm.). With more loci, the reduced selection per locus is outweighed by the strengthening of linkage disequilibria, which makes selection more effective overall (Barton 1983).

This model assumes immigration from only one parental taxon (*B. bombina*). Since there must also be migration into the intermediate sites from more *B. variegata*-like populations (Fig. 2.6a; see above), the swamping effect will be reduced, both because allele frequencies are under weaker pressure, and because linkage disequilibria will be higher. Moreover, the high inferred migration rate suggests that the *B. variegata*-like animals may constitute a single large population that uses a range of temporary habitats for reproduction. In this much larger gene pool, the *B. bombina* immigrants make up a smaller fraction ($m \approx 0.05$), so that less selection will suffice to maintain allele frequency differences. Studies are currently

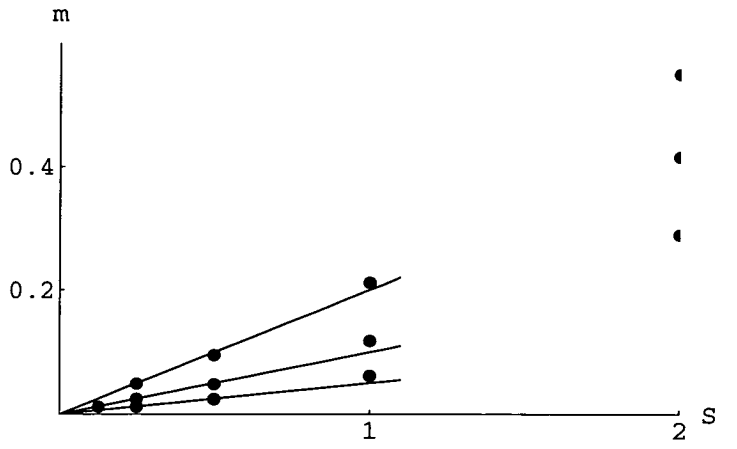
underway to investigate the dispersal patterns among local toad aggregations and selection on tadpole cohorts along the habitat gradient (T. Sands and S. Köhler, unpublished data). For now, I stick with this estimate of habitat-based selection as an upper limit for the required selection strength that could maintain allele frequency differences at selected loci indefinitely. Since the estimate is not implausible, there is no reason to doubt the stability of this hybrid zone as far as selected loci are concerned. I attempt to quantify two potential mechanisms for this selection, namely breeding site choice and assortative mating, in Chapter 5.

Figure 3.9: The critical rate of immigration, m_c, which just swamps selection, S. A: selection against hybrids, $W = \mathrm{Exp}[-4\,Spq]$; B: selection against incoming alleles, $W = \mathrm{Exp}[-2\,Sp]$. Calculations are for n = 5, 10, 20 loci (top to bottom). In each figure, the lines to the left show the approximation assuming linkage equilibrium ($m_c = S/4n$ for selection against hybrids, S/n for selection against incoming alleles). Note the different scales on the y-axes.

A.



B.



How can selection maintain neutral differences?

If indeed the traits that confer adaptation to temporary sites are largely intact in the *B. variegata*-like hybrid populations, then our data on marker alleles demonstrate the ongoing breakdown of neutral divergence: there are hardly any pure *B. variegata* populations in the area. Yet there is still a clear association between alleles and habitat (Fig. 2.7). On their own, neutral allele frequency differences should dissipate very rapidly, over an approximate timescale of $\sim 1/m = 5$ generations. The fact that marker loci are informative in hybrid zones over much longer times scales is due to selection on linked loci, which creates a genetic barrier to gene flow (Barton 1983). In the present context, its effect is to reduce the actual migration rate to a lower effective rate m_e and hence to slow the inevitable decay of neutral divergence. One can therefore ask about the barrier strength of different selection schemes and about the maximum age of the hybrid zone that they could explain.

For the scenario above, selection in intermediate populations against incoming *B. bombina* alleles of $S \sim 1.7$ on 20 unlinked loci would reduce the migration rate at interspersed neutral loci to $m_e = 0.09\,m$ (Barton and Shpak 2000; N. Barton, pers. comm.). Thus, a selection scheme that would maintain the adaptation of the toads to ponds and puddles, respectively, could sustain the divergence at marker loci for only 55 generations or, assuming a generation time of 5 years, for about 280 years. Clearly, selection schemes can be devised that generate larger barrier strengths. For example, Barton and Bengtsson (1986, see their Figure 4a) calculated the reduction in gene flow for a model of weak additive selection on n evenly spaced loci, with the neutral marker embedded at the centre. With 40 selected loci and $S = 1.7$ on one such chromosome, gene flow would be reduced 100 fold. This model implies the existence of one locus every 4 cM with $s = 0.04$ and sustains neutral divergence for

2500 years. If one assumes that a single large *B. variegata*-like population exists in the temporary habitats such that *B. bombina* adults immigrate at a rate $m = 0.05$ (see above), then we can stretch the age of this hybrid zone to 10,000 years. The same effect could be generated by weak assortative mating within sites, or an active choice of breeding habitat (see Chapters 4 and 5).

These computations demonstrate the possibility that this hybrid zone formed soon after the last ice age. However, the models require strong overall selection involving a large number of loci. As an alternative, *B. bombina* may have moved into the area from the Hungarian Plains following the deforestation from the 14th century onwards (Pounds 1994) and so come into contact with the previously established *B. variegata* populations. The latter idea is supported by the fact that *B. bombina* is known to avoid forested areas (MacCallum 1994). It suggests that only 150 generations have passed since the onset of hybridisation, so that current levels of neutral divergence could be explained by selection weaker than S~1.7. Without empirical data on the total strength of selection and on its distribution across the genome, it is not possible to decide between these alternative hypotheses. Instead, the key message is that neutral divergence in this and other mosaic hybrid zones is necessarily transient and poses constraints on feasible combinations of hybrid zone age and selection.

Section 3.7 Summary

Chapters 2 and 3 show that:

1) the fire-bellied toads *B. bombina* and *B. variegata* form an extended mosaic hybrid zone near Apahida, Romania that differs remarkably in its spatial structure from previously analysed clines in Poland and Croatia.

2) Pure *B. bombina* populations occur sporadically in large ponds, and these are the source of the introgressing *B. bombina* alleles into the *B. variegata*-like populations in the abundant temporary aquatic sites throughout the area.

3) Using the linkage disequilibria in intermediate sites, an estimate of this asymmetric migration rate was derived; this also neatly accounts for the observed heterozygote deficit.

4) The minimal selection strengths that are required to stabilise the mosaic structure were explored. While plausible selection coefficients could maintain adaptation to temporary versus permanent breeding sites, the divergence in neutral traits is probably collapsing.

5) If hybridisation began just after the end of the last ice age, this would require that selection creates a very strong barrier to gene flow. Alternatively, this hybrid zone may have formed in the last 500 years following extensive deforestation. In either case, the high rate of introgression implies that non-critical taxon differences are eroding.

Section 3.8 Linkage disequilibrium generated by admixture

This section was written by N. H. Barton (University of Edinburgh) and is Appendix II in Vines et al. (submitted).

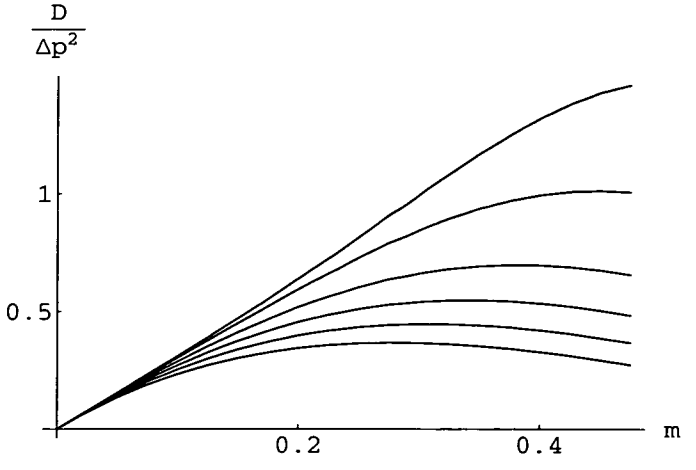
Immigration at a rate m between populations which differ in allele frequency by Δp at each of two loci generates pairwise linkage disequilibrium proportional to $c\,m\,\Delta p^2 / r$. However, the precise relationship (i.e. the value of c) depends on detailed assumptions about the life cycle, the stage at which D is measured, and the forces acting on allele frequency. The formulae given by Asmussen and Orive (2000) and Barton (2000) assume that there is immigration from two source demes, such that allele frequencies remain at equilibrium. Barton and Gale (1993) consider immigration from a single deme, such that allele frequency differences and linkage disequilibrium decay at a steady rate. Here, we suppose that strong selection acts to impede gene flow, so that marker frequencies change slowly despite high rates of immigration. We must therefore find the linkage disequilibrium when immigration is countered by selection at linked loci.

Assume that individuals mate at random within demes, producing zygotes in Hardy-Weinberg proportions. Juveniles are subject to multiplicative selection, which again maintains Hardy-Weinberg proportions. Epistatic selection could generate heterozygote deficit and linkage disequilibrium, and so our assumptions may underestimate these. Note that we need not assume that selection acts directly on the two loci in question: the change in genotype frequencies depends on the marginal fitnesses, which may be determined by selection at linked loci. Adults migrate between demes, and then mate at random to begin the next generation. Migration may represent movement from place to place, or mixing between habitats due to imperfect habitat preference. Genotype frequencies are measured in

breeding adults after migration. We assume that immigrants come from a source deme with allele frequencies $p = 0$ at both loci, and that there is an equilibrium between selection and immigration; this determines the selection coefficients as a function of immigration rate.

The heterozygote deficit in this model can readily be shown to be $mp / q(1-m)$. There is no simple explicit solution for the linkage disequilibrium as a function of immigration rate and allele frequency. However, the relation can be calculated numerically (Fig. 3.10). For low migration rates, $D \sim 3\,m\Delta p^2$, as expected from formulae which assume a quasi-equilibrium between immigration, and with recombination $r = 1/2$ for unlinked loci (Barton, 2000). For higher migration rates, disequilibrium remains close to this expectation when allele frequency is low, but falls below it for higher allele frequencies. In the extreme case, $D \sim 0.5\,m\Delta p^2$ for $m = 0.5$, $p = 0.5$. Since we will be concentrating on migration rates less than m ~ 0.3, and allele frequencies in the range 0.2 to 0.4, we take the relation as $D \sim 2\,m\Delta p^2$.

Figure 3.10: The strength of linkage disequilibrium D scaled relative to difference in allele frequency is plotted against immigration rate ($D / \Delta p^2$ vs. m). The curves are for allele frequency p = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 (top to bottom). Allele frequency and linkage disequilibria are measured in adults, after multiplicative selection and migration.



Chapter 4:
Quantifying mate choice in *Bombina*. I. The data and its preliminary analysis

Section 4.1 Introduction

How often is sexual isolation the principal cause of reproductive isolation? Data from a wide range of taxa suggest that the evolution of assortative mating is rather capricious, as its appearance does not correspond well with the level of differentiation at allozyme markers (Butlin and Tregenza 1998). It appears that the accumulation of sexual isolation is favoured when there is strong sexual selection (e.g. Barraclough et al. 1995), or when divergence of the mating system is driven by natural selection (Schluter and Price 1993). However, even when both these are present strong prezygotic isolation does not always arise, as many taxa exhibiting sexual selection can still hybridise (Harrison 1993). The hybrid zones in *Bombina* are a classic example of this, as *B. bombina* and *B. variegata* have been diverging for over 2 million years without achieving full prezygotic isolation (Szymura 1993). How much assortative mating is there between these two taxa? Chapter 3 found that a strong selective barrier against immigrant *B. bombina* alleles was required to maintain divergence at selected loci in the Apahida mosaic hybrid zone. Since strong ecological selection against immigrant adults is implausible, assortative mating could have a major role in reducing immigrant fitness. In this chapter I present preliminary data on role of breeding site choice and assortative mating in preventing gene flow between *B. bombina* and *B. variegata.*

The methods presented here and in Chapter 5 stem from an analysis of *Bombina* mating behaviour in Pešćenica, Croatia. This work was recently submitted to Heredity in a paper entitled "Mating patterns in a *Bombina* hybrid zone: inferences from full sib genotypes when neither parental genotype is known" by B. Nürnberger, L. E. B. Kruuk, N. H. Barton, and T. H. Vines.

The origins of sexual isolation

The key feature of sexual isolation is the "weakness or lack of mutual attraction between males and females of different species" (Dobzhansky 1937). Why might individuals from separate taxa find each other unattractive? The recognition of suitable mates necessarily involves the coadaptation of male-female signalling systems, and there are two schools of thought on what traits these might involve. Carson (1978) and later Ryan and Rand (1993) proposed that mate recognition systems involved the same phenotypic traits as mate choice, implying that sexual isolation is a by-product of sexual selection. Alternatively, Paterson (1980, 1993) maintained that mate recognition is a choice based on a much wider range of phenotypic cues, which could be under strong stabilising selection.

Unfortunately, there is little empirical evidence for either argument. Ryan and Rand (1993) showed that calling song in *Physaelaemus* toads is used both to distinguish between taxa and to choose a mate, although it is not clear that the same elements of the song phenotype fulfil both roles. Endler and Houde (1995) found extensive variation in female preferences for male guppy colouring and morphology between populations, and females almost exclusively preferred males from their own population. However, as with *Physaelaemus* toads, it was not clear that the same aspects of color pattern were involved in both attractiveness and the rejection of foreign males. In an experiment on a hybrid *Drosophila* population, Blows and Allen (1997) showed that the cuticular hydrocarbon composition involved in population recognition by the parental taxa also determined mating success in the hybrids, illustrating that population specific characters may also have a role in mate choice. By contrast, Boake et al. (1997) found that head width in male stalk-eyed *Drosophila*, although involved in mate choice, apparently plays no part in distinguishing between species.

Since mate recognition is presumably based on a 'search image' for a potential mate, progress might be made on this issue by studying how morphological changes in males affect their acceptance by females of their own population. If mate recognition involves the same traits as mate choice, altering other aspects of male morphology should have little impact on their reproductive success. However, if mate recognition is based on the wider appearance of the male, arbitrary morphological alterations will lead to female discrimination and fewer matings for the altered males.

If, for simplicity, we assume that mate recognition systems are based on sexually selected traits, what might cause divergence in these traits? When females select from a range of displaying males on the basis of a particular trait, sexual selection can rapidly generate sexual isolation because the female preference can evolve to match an arbitrary level of the male trait (Lande 1981). However, this model assumes that female preference is selectively neutral; if selection acts upon female choice, e.g. as a direct cost of searching or in the form of help raising offspring, the optimal level of trait and preference instead converge to a single point (Pomiankowski 1987). Divergence thus requires additional forces to overcome any cost to female preference. Zahavi (1975) showed that female choice is favoured when the male secondary sexual character has a high metabolic cost, and therefore can only be fully expressed in the most fit males. The male trait is thus a guide to overall male viability, and females will gain a fitness benefit from choosing the best males because her sons will achieve more matings (although this benefit may be small: Kirkpatrick and Barton 1997). For example, if there is a parasite load on the population, males resistant to that particular strain will be fitter, and therefore more able to express the costly male trait (Hamilton and Zuk 1982). The coevolutionary cycle between hosts and parasites ensures that only a proportion of

the male population is resistant at any one time, and hence there is always a fitness benefit to female choice. Interestingly, Andersson (1986) showed that this 'handicap' model of sexual selection might also lead to male ornamentation in monogamous species.

Given that male display traits indicate fitness, how then can the trait and preference diverge between populations? Firstly, selection on the female preference can change her mate choice cues: it is conceivable that changes in the parasite population may alter which traits are a more reliable indicator of fitness, and, providing there is heritable variation for choice of male trait, the females may change to preferring a new trait. In addition, if the populations inhabit two different environments, the perceptibility of the male trait may differ in each habitat. In this case, females may switch their preferences onto a less reliable but more detectable male trait, as this reduces her costs of searching (Schluter and Price 1993). An alternative model based on polygenic inheritance of the male trait and female preference was developed by Iwasa and Pomiankowski (1995), in which the fitness cost of greater preference by females leads to cyclical fluctuations in both trait and preference. In this case, populations at different stages of the cycle may well be sexually isolated from each other. However, as Kirkpatrick and Barton (1995) point out, stabilising selection on the male trait stops the cycles. Furthermore, there is no evidence that the levels of male traits fluctuate in the wild; instead completely different traits appear to be stable in different populations (Coyne and Orr 1998). More generally, female mate recognition systems may diverge because male morphology has changed under ecological selection, and the females search image changes in response; or simply as a pleiotropic side effect of selection on traits involved in searching for food or detecting predators (Kirkpatrick and Ryan 1991).

Asymmetry in sexual isolation

Another issue worthy of consideration is asymmetry in sexual isolation, in which interpopulation matings involving one combination of males and females occurs more often than the reciprocal cross. A model for this phenomenon in Hawaiian *Drosophila* was proposed by Kaneshiro (1980), in which the founder effect 'loses' male courtship traits in derived populations. Females in the derived populations accept males of the original population because these males display all the necessary traits, whereas females of the original population reject derived males because they lack crucial courtship elements. This hypothesis has been criticised on several fronts. Firstly, the disappearance of phenotypic traits is an unlikely outcome of the founder effect (Barton and Charlesworth 1984), and secondly, the large number of experimental studies carried out have produced conflicting results (see Arnold et al. 1996). Lastly, differences in mating propensity may account for much of the pattern (Barton and Charlesworth 1984). A more recent study based on Lande's (1981) polygenic model of sexual selection showed that asymmetry in sexual isolation was possible when male traits are close to, but not at, two different female preference optima in the two populations (Arnold et al. 1996). This model predicts greater levels of asymmetry at intermediate levels of divergence, which was supported by their data for *Demognathus* salamanders (Arnold et al. 1996). However, the restrictive assumptions about the neutrality and Gaussian distribution of female preferences in this model make it an unlikely explanation for the widely observed phenomenon of asymmetry in sexual isolation, and a more general model is clearly required.

An intriguing possibility is that the mating system is instead based on the ability of females to resist seductive male traits (Holland and Rice 1998). There is considerable evidence of 'sensory bias' in female mating preferences: females often

prefer male traits not found in their own population, which may arise as an incidental by-product of viability selection on the female sensory system (Basolo 1990; Ryan and Rand 1990). If the males develop such novel traits, the females will initially have no resistance to them, and they will be induced into mating more frequently than is optimal. This in turn generates selection for resistance to the new display. Eventually, the exaggeration of the male trait is checked by natural selection, and the females can evolve full resistance to it. Furthermore, the now ineffective male trait may still be needed to achieve threshold levels of stimulation in females. When two populations meet, one of which has evolved a new male ornament, sexual isolation between them may be asymmetric: females from the new ornament population will reject matings with males with the original ornaments as they are not sufficiently stimulating, whereas females from the original population will be unable to resist the newly ornamented males. Since this model of sexual selection makes several unique predictions (Holland and Rice 1998), it may be possible to test this hypothesis in the more well-known cases of asymmetry (e.g. Hawaiian *Drosophila*).

In summary, there is still much to learn about the evolution of sexual isolation. Sexual selection appears to be important in generating reproductive isolation, not least because there are numerous phylogenetic studies showing that the strength of sexual selection is related to the speciation rate (Panhuis et al. 2001). However, many uncertainties remain. Theoretical work on sexual isolation could perhaps benefit from concentrating on situations where there is a fitness cost to female choice, and by adopting some of the excellent simplifications suggested by Kirkpatrick and Ravigné (2002). In addition, experimental work could gain substantially from understanding the relationship between mate choice and mate recognition (see Ryan and Rand 1993). This will shed light on the circumstances

under which sexual selection can generate sexual isolation as a by-product, and how mate recognition mechanisms can diverge independently of mate choice. Furthermore, it seems that understanding the origins of asymmetry in sexual isolation is important, as it may be possible to identify the traits responsible for greater levels of mate discrimination in these situations. Lastly, important data may come from hybrid zones, as here the individual contributions of sexually selected traits to reproductive isolation can be determined. I discuss sexual selection in hybrid zones below.

Mate choice in hybrid zones

Hybrid zones are an ideal opportunity to examine the relative contributions of different isolating mechanisms to the prevention of gene flow between hybridising taxa (Harrison 1990). However, since sexual selection can rapidly generate complete reproductive isolation (Kirkpatrick and Ravigné 2002; see above), inferring its importance from hybrid zones may be misleading, as these necessarily involve taxa with incomplete prezygotic isolation. Nonetheless, important questions about the evolution of sexual isolation can be addressed with data from hybrid zones. For example, how much assortative mating is generated by morphological changes during adaptation into different habitats? If phenotypic divergence leads to changes in the mate recognition system of each population, their differences are more likely to be preserved on secondary contact. Data on these processes are of fundamental importance to models of sympatric speciation by disruptive ecological selection (Schluter 2001). Another aspect of this issue is whether ecologically isolated populations show more or less sexual isolation than those that still inhabit the same environment, as again this would indicate a role for natural selection in the divergence of mate recognition systems (e.g. forest birds: Endler and Thery 1995). Finally, detailed knowledge of sexual isolation between

hybridising populations is essential for understanding the process of reinforcement, in which selection against hybrids generates selection for greater mate discrimination (Noor 1999).

Empirical work on sexual isolation in hybrid zones has attempted to measure the strength of discrimination against the other taxa or against hybrids using three methods. Firstly, in hybrid zones between taxa with little ecological isolation, the role of habitat structure may be insignificant, and in these cases mate choice trials in the lab are convenient surrogate for field observation. This approach has been used extensively in the *Chorthippus parallelus* hybrid zone. Ritchie et al. (1989) found that there was partial assortative mating between the two populations, perhaps because *C. parallelus parallelus* primarily uses song characters in mate recognition, whereas *C. p. erythropus* uses olfactory cues (Ritchie 1990). The two taxa do not appear to differ ecologically, and hence these results may reflect their behaviour in the field. Other examples of lab-based mate choice trials include *Chauliognathus* beetles (McLain 1985), *Melanoplus* crickets (Orr 1996), *Mus musculus* (Smadja and Ganem 2002), and *Orchelimum* crickets (Shapiro 2001). Generally, there is a trade-off between the opportunities for high levels of replication in the lab and the chance that important environmental factors have been missed. For example, if the two taxa are ecologically differentiated, the habitat in which they meet may have a strong effect on their propensity to mate, and reconstructing this effect in the lab is difficult. A large number of studies of assortative mating in hybrid zones have thus relied on observing of mated pairs in the wild.

A good example of such a study is Johannesson et al. (1995) on *Littorina saxatilis* marine snails, where matings between the smooth/unbanded lower shore morph and the ridged/banded upper shore morph were infrequent (~1 in 60 observed pairings).

This was partially caused by a non-random distribution of the morphs where they meet in the mid shore, and some form of active mate choice. Other studies include Helbig et al. (2001) on *Phylloscopus* chiff-chaffs, Moore (1987) in *Colaptes* woodpeckers, Ribi and Oertli (2000) with *Viviparus* marine snails, and Grant and Grant (1992) with *Geospiza* Galapagos finches. Sætre et al. (1997) used both aviary-based mate choice experiments and field observation of mated pairs in *Ficedula* flycatchers to demonstrate that mate discrimination between *F. albicollis* and *F. hypoleuca* is stronger between sympatric populations. However, there are some circumstances under which there are fitness benefits to a female for choosing a heterospecific mate in these taxa (Veen et al. 2001), which suggests that assortative mating can be based on subtle fitness trade-offs and is not simply a binary mate recognition process. When observation of mating pairs in the wild is difficult, usually when pairings are very brief, or take place out of sight, more indirect measures based on samples of offspring are required.

Inferring assortative mating from offspring combines two (and sometimes three) isolating mechanisms. Firstly, sexual isolation determines what types of matings occur; second, both mechanical and gametic isolation affect what proportion of those matings lead to the fusion of gametes; third, hybrid inviability acting in the early stages after fertilisation will determine which offspring are sampled. Since the latter can seriously affect the estimates of the former two, it is advisable to either conduct separate viability experiments or to sample the offspring as soon after fertilisation as possible. The majority of published studies collect fertilised females from the wild and rear their offspring in the lab, so that inviability can be assessed in a controlled environment.

For example, Hewitt et al. (1987) sampled both mated pairs and gravid females in the hybrid zone between two chromosome races of the grasshopper *Podisma*

pedestris, and karyotyped both the females and the embryos. They found that pairings sampled in the field were not different from random, but fertilisation must be assortative. Mallet et al. (1998) reared field collected larvae from the hybrid zone between mimetic forms of *Heliconius* butterflies, and inferred the rate of interpopulation matings from the wing patterns of the metamorphs. They found that around 5% of the total pairings were between populations, and they postulated that frequency-dependent selection against the rare hybrid warning patterns would be required to stabilise the hybrid zone.

In plants, direct mate choice is expressed via pollinator preferences, and gametic isolation after pollination is also important in determining fertilisation patterns. The rate of interpopulation matings in plants is thus often best estimated from the genotypic composition of the seeds, as direct observation before this stage is difficult. The study by Rieseberg et al. (1998) in *Helianthus* sunflowers employed data from mothers and offspring characterised at seven isozyme loci to infer the relationship between the available pollen and the pollen that finally fertilised the ova for both the pure populations and three categories of hybrid. They found that the pure plants were able to discriminate against pollen from the other population and hybrids, but this system broke down in the more intermediate individuals. Similar studies have be carried out on European oaks (Bacilieri et al. 1996) and in Louisiana irises (Arnold et al. 1993; Hodges et al. 1996). In summary, estimating assortative mating indirectly from the offspring can show how the adult generation contributes to the offspring, which is invaluable for understanding the dynamics of the hybrid zone. However, this includes the effects of several isolating mechanisms, and these cannot be disentangled without additional data.

Section 4.2 Quantifying assortative mating in *Bombina*

The mating system of Bombina

Here and in Chapter 5 I attempt to quantify mating behaviour in *Bombina* toads. Traditionally, mate choice experiments in amphibians have involved monitoring female orientation in response to the calls of males from different taxa (e.g. Ryan and Rand 1990; see Gerhardt 1994 for a review). However, there are a number of reasons why this approach may not be suitable in *Bombina*. Firstly, there is evidence that there are differences between the mating systems of the two taxa (Lörcher 1969; pers. obs.). In the pond breeding *B. bombina*, males fight for small water surface territories in the shallows (< 1m), and attempt amplexus with anything under ~15 cm in length moving within that area. In contrast, the puddle breeder *B. variegata* forms calling aggregations in small waterbodies, and female toads moving into these aggregations can be amplexed by a single male (Seidel 1987; Barandun 1990) or several simultaneously (pers. obs.). This implies that there could be less opportunity for female choice in *B. variegata*, as females cannot easily approach their desired mate without attracting the attention of other males. Additionally, the breeding season for both taxa in Romania extends from early April through to mid July (pers. obs.), and hence there is unlikely to be any temporal isolation between them (see Section 1.2).

The mating pattern is therefore highly dependent on the habitat and the range of animals present, which makes the interpretation of laboratory mate choice experiments difficult. Furthermore, the majority of mating takes place at night (Lörcher 1969; pers. obs.), and thus direct observation of mated pairs would be time consuming and difficult, especially since *Bombina* are challenging to capture in the dark. An additional concern is that females may operate some form of cryptic

female choice by reducing the number of eggs they lay when amplexed by an undesired male (e.g. Reyer et al. 1999). It thus seems most sensible to quantify mating behaviour indirectly from the offspring, by sampling both adults and egg batches across a range of habitat types. This means we cannot distinguish assortative mating from assortative fertilisation, but it is nonetheless useful to have single estimate of within-site prezygotic isolation, as this allows us to make predictions about the offspring composition of a site given a sample of its adults. One additional complication to this approach is that neither parent remains with the eggs after laying, and therefore the paternal genotype cannot be reconstructed from that of the offspring and the mother. Indirect inferences about the parents from the eggs are therefore necessary, and I discuss this in more detail below and in Chapter 5.

One interesting aspect of reproduction in *Bombina* is how eggs are deposited by amplexed pairs. Once a pair has formed, they move around the waterbody laying eggs onto plant stems and other fixed objects. The female manipulates the eggs as they emerge and ensures they are attached, and the male releases sperm to fertilise them. In sites where laying points are scarce, the same fixing point may be used by more than one pair (pers. obs.; L. Kruuk, pers. comm.). There is some evidence that *B. bombina* lays more eggs per batch than *B. variegata* (Lörcher 1969), but there are no data on the total number of eggs females of either taxa can lay. Barandun (1990) hypothesised that *B. variegata* laid fewer eggs in each batch as an adaptation to the risk of a particular puddle drying up, but this has not been shown experimentally. There is certainly no record of amplexed pairs moving over ground between temporary sites, which would lend support to this idea; alternatively, females may escape amplexus and remate once they have moved to another site.

There are good reasons to expect *Bombina* to be capable of active mate choice. Chiefly, their mating calls differ in a number of ways (Sanderson et al. 1992), including in their cycle length (i.e. rate of repetition^{-1}). This trait is involved in mate recognition in numerous other anurans, including *Hyla versicolor* and *H. chrysoscelis* (Blair 1962), *Pseudacris nigrita* and *P. clarki* (Michaud 1962) and *Hyla verreauxi* and *H. ewingi* (Littlejohn and Loftus-Hills 1968). However, it is well known that water temperature affects the calling characters of amphibians (Duellman and Trueb 1985), and both Sanderson et al. (1992) and Schneider and Eichelberg (1974) found that cycle length decreases with rising temperature in *Bombina*. Since ponds and puddles have very different thermal regimes, it is possible that the differences in song between the taxa diminish when they are an unfamiliar thermal environment (i.e. the wrong habitat), although this is not supported by the regressions of song characters against temperature in Sanderson et al. (1992). Additional evidence that there might be assortative mating comes from the observed heterozygote deficit among the adults in both Romania and Croatia, although at least a part of this may be due to either selection against hybrids or migration between habitat types. More definitive evidence for assortative mating would come from a heterozygote deficit in embryos, and I present data on this below.

Defining the adult pools

There is a difficulty over how we define the adult pool from which parents for a collection of eggs from a site might be drawn. I showed in Chapter 3 that there is a high rate of movement between sites ($m \sim 0.2$). This implies that the most representative sample of the potential parents involves those from all the sites in the vicinity, as any of these animals could have visited the site in question and been a parent. This approach has additional advantages: firstly, the larger number of toads

means that the genotypic distribution (i.e. the frequency of each multilocus genotype) can be reliably estimated, this greatly widens the scope of our statistical analyses (see Chapter 5). Secondly, there is no need to worry about the day to day changes in adult composition within this focal site, as these should result from the subsampling of the larger and more stable local adult pool. I define the adult pool as the toads sampled within 300m of the focal site in either 2000 or 2001; individual toads have been observed to cover similar distances in Croatia in only a few days (MacCallum 1994).

Estimating the strength of assortative mating with an adult pool defined in this way is complicated by the active habitat preference (Chapter 2). If there are other habitat types nearby, the adults within a site will be a non-random subset of the local pool. This makes any inference of direct mate choice among all the local adults unreliable, as toads were not able to choose among the entire range of potential mates available. Habitat preference and assortative mating are thus mutually dependent: if active habitat choice is strong, sites will tend to contain similar animals, and there is little potential for discrimination against foreign genotypes. As habitat preference weakens, sites containing a wide range of adults become more frequent and there is more opportunity to express mate preferences.

Since toads may be actively choosing resting or feeding habitats, and simply mating within those, there are three components to assortative mating in *Bombina* that must be estimated: 1) The habitat-genotype association in the adults generated by an active choice of feeding and resting habitat, 2) any additional habitat-genotype association in the eggs generated by breeding site choice by the adults, and 3) direct mate choice within breeding aggregations. I make preliminary estimates of all three in this chapter, and parameterise each more fully in Chapter 5.

Section 4.3 Data collection techniques

Collection of eggs

Sites were visited around the Apahida study area once every 3-4 days as part of the wider adult survey in 2000 and 2001 described in Section 2.2, and each was intensively searched for egg batches by eye. As the season progressed, we concentrated on sites containing a wide range of adult toads, as we have a higher chance of detecting assortative mating in these. Sites that had previously yielded eggs were also visited more frequently than those that had not.

Whenever eggs were found, a sample of each batch was taken for lab rearing. we also took a sample of the adults, using the protocol described in the Section 2.2. These adults are part of the adult survey described in Chapters 2 and 3. Although *Bombina* pairs in amplexus are known to deposit eggs in several locations around a site during a single mating (Lörcher 1969; pers. obs.), we sampled every discrete batch. This meant that whilst we were likely to sample the same mating more than once, we had a better chance of sampling all the recent matings. I consider how batches from the same mating might be detected below.

Bombina bombina clutches contain around 30-50 half-black half-white eggs (Rafinska 1991), although this coloration disappears after about 24 hours (pers. obs.). They can be confused with the eggs of the European tree frog *Hyla arborea*, but the violin-shaped tadpoles of this species make their detection after hatching simple. *B. variegata* lay eggs in smaller clutches of 10-30 eggs, and the two egg hemispheres are coloured light and dark brown respectively. These are easily confused with spawn from the frog genus *Rana*, adults of which were abundant in our sites. *Rana* clutches tend to be larger, often with >100 eggs, but small clusters some distance away from the main group are common (pers. obs.). The jelly is also

slightly stickier than that of *Bombina*, but this feature appears to be inconsistent. Since we only grew the tadpoles up for one week, the characteristic size and fin differences normally used to distinguish *Bombina* and *Rana* had not yet appeared. To ensure no *Bombina* families were missed, I sampled unidentified batches and relied on the non-amplification of *Rana* DNA at the genotyping stage to distinguish the eggs.

Up to sixteen eggs were taken from large (> 30 eggs) batches, and around ten from smaller ones (15-30 eggs). We took all the eggs from batches of 10 or fewer eggs, as although correctly identifying the parental genotypes at each locus becomes difficult with less than ~5 eggs (see Chapter 5), small batches still contain information about changes in allele frequency between adults and offspring. Eggs were taken back to the lab and individually examined under a binocular microscope. I recorded their Gosner stage (Gosner 1960) and measured the diameter for eggs under Gosner stage 12. (After this stage the embryos are elongated and curled up inside the jelly.) Batches were divided into groups of 3-5 eggs and kept in plastic cups containing fresh dechlorinated water in a cupboard (temp range 21-24°C). The water was topped up every three days. Once the tadpoles had reached Gosner stage 23 (around 1 cm in length) around 1 week later they were measured, dabbed dry and stored in a 0.75ml Eppendorf tube with 99.9% Ethanol.

Selection of sites and genotyping of eggs

Over the season we sampled 2986 eggs in 226 batches. Since genotyping this many animals would be impractical, I concentrated on sites that had i) produced a reasonable number of batches over the season and ii) contained a wide range of adults. The remaining eggs are part of a wider study on the correspondence between adult and egg allele frequencies in Romania by T. Sands. The sites with a

wide range of animals correspond to those in Chapters 2 and 3 that contained immigrant *B. bombina* or early generation hybrid animals, and it is here that the assortative mating is most likely to be detected. This is firstly because animals differing at mate choice loci are more likely to also differ at marker loci, and secondly because there will be a wider range of potential mates among which to exercise that choice. Since the egg collections were also part of a parallel study on juvenile survival, an additional criterion was that sites must have produced metamorphs at the end of the season. This left five sites, two from 2000 (0200.4 and 0257), and three from 2001 (1258, 1290 and 1315). The eggs were genotyped at same four loci as the adult samples (*Bb7.4*, *Bv12.19*, *Bv24.11*, *Bv24.12*; see Section 2.2) by Sonja Köhler of LMU Munich (see Nürnberger et al. 2002b for details). Tadpoles that had obviously been feeding (judging by the colour of their stomach) were gutted to prevent DNA contamination. The problem of weak peaks at locus *Bv12.19* in the adult samples (see Section 2.2) was not found in the genotyping of the tadpoles.

Site descriptions

Detailed ecological data on the five focal sites and the remaining 88 sites from 2000 and 2001 can be found in Appendix II. I give summarise the main features of the focal sites here as well.

Site 0200.4 is an excavated pool on the Apahida-Cojocna road (Figure 2.4d and 4.1), with a habitat discriminant axis score of $H = 0.46$. It measured 5 m long and 4 m wide, and it was approximately 1.5 m deep, although the opaque muddy water obscured everything 10 cm below the surface. There was no aquatic vegetation except for long grass shoots trailing into the water from the bank. It is near the north west end of the series of fifteen excavated pools described in Section 2.4 and shown in Figure 2.4d. It was probably dug in the mid 1990's (Ilie Tudorescu, pers.

comm.). The older and more vegetated sites to the south east are more *B. bombina*-like, and these individuals are quite capable of migrating the 30 m to mate in 0200.4. In total, there were ten sites sampled around 0200.4 in 2000 (0200.3, 0200.5, 0200.6, 0200.7, 0200.8, 0200.9, 0200.10, 0244, 0245 and 0271, and six (including 271) were sampled in 2001 (1271, 1315, 1317, 1318, 1334 and 1335).

Site 0257 is an area of tractor wheel ruts and puddles in a valley to the north of Dezmir (Figure 2.4d). The maximum depth was 20 cm, and the total surface area of water is about 10 m^2, and the presence of a flush on the adjacent hillside maintained water in the site until sampling ended in late June. It contained a moderate amount of vegetation, mostly reeds and grasses typically associated with wet areas. None of the plants normally associated with permanent water were present. The discriminant habitat axis score for 0257 was H = 0.65, which is towards the puddle end of the scale (see Section 2.2). The large pond 985 is 100 m away to the north west, which may be the source of the *B. bombina* like adults found in 0257 (Appendix I), and 200 m from the puddle complex at 0256. There were also two small pools (0252 and 0251) 300 m to the east, on the adjacent ridge.

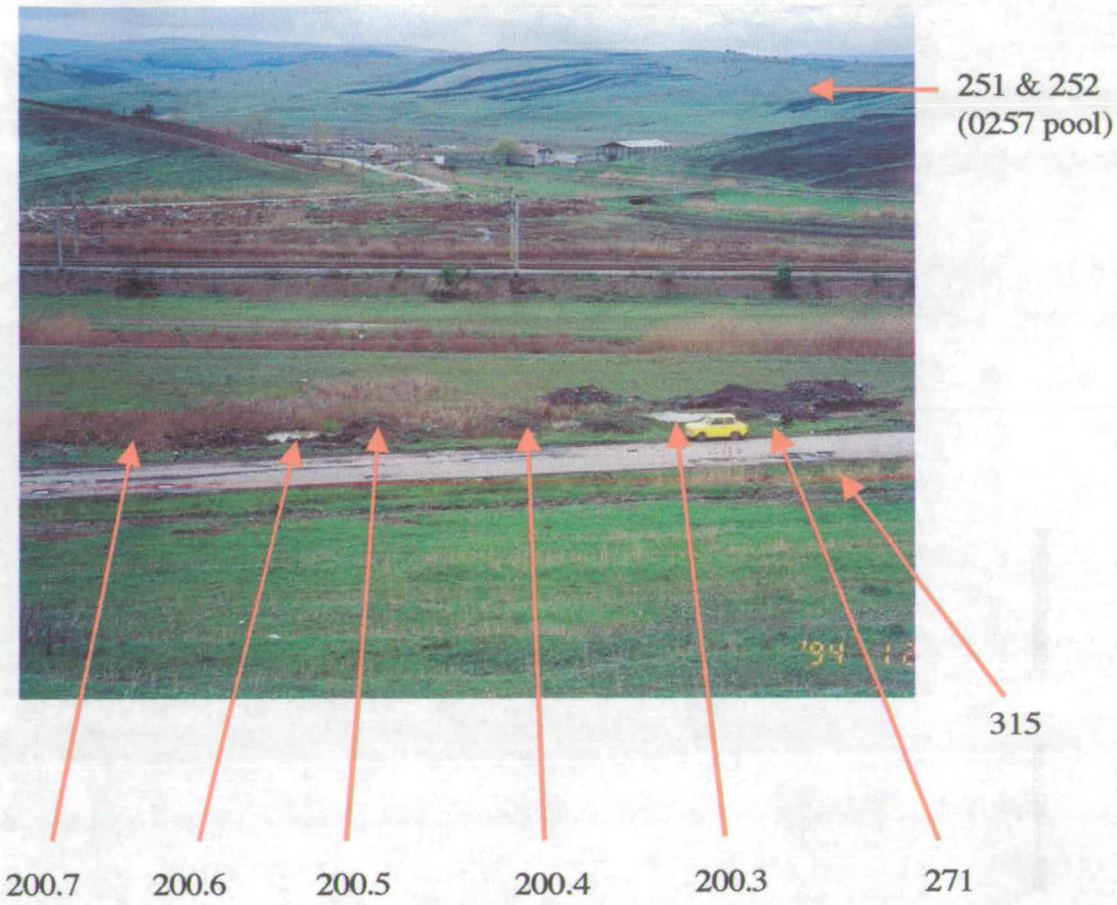
Site 1258 consisted of a small pool adjoining a shallow reeded area in the Valea Broastelei (the Valley of the Toads), which is a small valley to the west of the Gadalin river (Figure 2.4c). The pool side was 4.3 m across and almost round, and at most 50 cm deep. There was only a little aquatic vegetation, and the substrate was dark mud. The reedy area, by contrast, was 5.4 m long, 3.4 m wide and around 10 cm deep, and was mostly covered with dense clumps of *Juncus* reeds. The discriminant habitat axis score for 0258 was H = 0.32. The Valea Broastelei contained five sites sampled in 2000 (0247, 0248, 0258, 0260 and 0270), all these were within 300 m of 258. The eggs analysed here come from when the site was sampled in 2001 (hence the number given to the site is 1258).

Site 1290 was a small pool next to a dipping well towards the mouth of the Visea valley (see Figure 2.4b). It was 8.8 m long and 5.6 m wide, and around 0.6 m deep, and was well vegetated with algae and tall reeds growing in a substrate of dark mud. Its habitat score was $H = 0.25$. It is 20 m away from the culvert ditch 0289, and the adult pool consists of this site and the adult sample from 290 in 2000.

Site 1315 was a long ditch on the opposite side of the road to the sites described in 0200.4 above (Fig. 4.1), approximately 100 m long, 1.8 m wide and 0.25 cm deep (see Figure 2.4d). It was extensively vegetated with reeds and other semi-aquatic vegetation, and was almost completely dry in 2000. Its habitat score was $H = 0.22$. The adult pool for this site is the same as that for 0200.4 above.

There are 15 sites in the vicinity of 0200.4 and 1315 in which adults were sampled, ten in 2000 and five in 2001. Recapture data shows that many toads returned to the area in consecutive years (T. Sands, pers. comm.), and hence I will combine the adults from both years. This means that the same adult pool can be used for both site 0200.4 and site 1315. I refer to this combined adult pool as "0200.4/1315". I use the same approach with data from other sites sampled in consecutive years: the 0258 animals are included in the 1258 adult pool, and the animals from 0289 and 0290 will be included in the 1290 adult pool.

Figure 4.1: Site 1315 and others from the area shown in Figure 2.2d (the car has moved from Fig 4.1a).



Section 4.4 The egg data: initial analysis

The egg data

The egg data consist of 584 eggs in 64 batches from 5 sites. Seven of these batches had fewer than four eggs, and since batches of this size contain too little information to make reliable inferences about their parents, they were discarded. This left 567 eggs in 57 batches (Table 4.1). Here and in Chapter 5 I describe a single locus genotype in terms of the frequency of *B. variegata* alleles, i.e. 0,1 or 2. The actual genotype of an adult or an egg batch is always contained with curly brackets; for example a completely heterozygous individual at four loci would be {1,1,1,1}. The adult pair of genotypes at a single locus is denoted e.g. (1,1) and the resulting egg genotypes a $\{x,y,z\}$ batch, where x, y and z are the numbers of 0,1 and 2 genotypes respectively.

Before any analysis of mating patterns can be attempted, it is important to check that each batch is the product of only one mating, and contains no genotyping errors. Generally, both of these problems will widen the range of genotypes present in the batch, increasing the number of matings that appear to have involved heterozygous parents. For example, a batch created from one parent homozygous for the *bombina* allele and a heterozygote (denoted a (0,1) mating) mixed with a batch from a (1,2) mating will contain all three genotypes in the offspring, and hence would be wrongly assigned to a (1,1) mating. Incorrectly genotyping an egg as a '0' in a batch containing only '1' or '2' eggs will have the same effect. Both of these can be evident as a large departure from Mendelian segregation (which I test for below), although mixed batches containing equal numbers of eggs from either pair of parents can appear to have segregated normally. A third issue is that our

sample may contain two or more batches from the same mating, as *Bombina* pairs in amplexus may lay eggs in several places around a site.

Table 4.1: Summary of data for each batch. Singletons in segregations containing 4 or more eggs are highlighted in bold. The column 'N' refers to the number of eggs per batch.

Site	Batch	Date	N	mean p	Locus			
					Bb7.4	*Bv12.19*	*Bv24.11*	*Bv24.12*
0200.4	1	27.4.00	10	0.96	{0, 0, 10}	{0, 3, 7}	{0, 0, 10}	{0, 0, 10}
0200.4	2	27.4.00	18	0.85	{5, 12, 0}	{0, 0, 18}	{0, 0, 18}	{0, 0, 18}
0200.4	4	1.5.00	14	0.83	{3, 4, 7}	{0, 0, 13}	{0, 4, 10}	{0, 5, 9}
0200.4	5	1.5.00	15	0.91	{0, 5, 9}	{0, 0, 15}	{0, 6, 8}	{0, 0, 15}
0200.4	6	14.5.00	4	1	{0, 0, 4}	{0, 0, 3}	{0, 0, 2}	{0, 0, 4}
0200.4	7	18.5.00	9	0.79	{0, 3, 6}	{0, 0, 9}	{3, 6, 0}	{0, 0, 9}
0200.4	8	2.6.00	5	0.65	{0, 3, 2}	{0, 0, 5}	{2, 3, 0}	{0, 4, **1**}
1315	3	28.4.01	6	0.81	{0, 0, 6}	{0, 3, 3}	{**1**, 3, 2}	{0, **1**, 5}
1315	4	4.5.01	5	0.88	{0, 0, 4}	{0, 0, 2}	{0, 0, 4}	{1, 1, 0}
1315	6	21.5.01	8	0.77	{0, 0, 8}	{0, 0, 8}	{3, 3, 2}	{0, 6, 2}
1315	7	21.5.01	5	0.85	{0, **1**, 4}	{0, 0, 5}	{0, 5, 0}	{0, 0, 5}
1315	8	21.5.01	9	0.76	{0, **1**, 7}	{0, 5, 4}	{0, 7, 2}	{0, 3, 5}
0257	1	24.4.00	15	0.82	{0, 6, 9}	{0, 8, 7}	{0, 7, 8}	{0, 0, 14}
0257	2	24.4.00	13	0.71	{0, 0, 13}	{2, 8, 3}	{0, 0, 13}	{7, 4, 2}
0257	4	24.4.00	10	0.81	{0, 0, 10}	{0, 4, 5}	{0, 6, 4}	{0, 5, 5}
0257	5	24.4.00	7	0.7	{0, 4, 3}	{4, **1**, 2}	{0, 4, 3}	{0, 0, 7}
0257	6	2.5.00	7	0.77	{0, 0, 7}	{0, 3, 4}	{0, 7, 0}	{0, 3, 4}
0257	7	11.5.00	4	0.94	{0, 0, 4}	{0, 0, 4}	{0, 2, 2}	{0, 0, 4}
0257	8	11.5.00	11	0.7	{4, 6, **1**}	{0, 3, 7}	{0, 5, 6}	{0, 0, 11}
0257	9.1	11.5.00	11	0.86	{0, 0, 9}	{0, 0, 9}	{0, 3, 6}	{0, 0, 9}
0257	10	11.5.00	11	0.77	{0, 0, 11}	{0, 10, 0}	{**1**, 8, 2}	{0, 0, 11}
0257	11	15.5.00	11	0.74	{0, 0, 11}	{0, 11, 0}	{2, 8, **1**}	{0, 0, 11}
0257	12	15.5.00	8	0.89	{0, 2, 6}	{0, 0, 8}	{0, 5, 2}	{0, 0, 8}
0257	13	15.5.00	10	0.57	{0, 7, 2}	{0, 0, 9}	{3, 5, 2}	{5, 5, 0}
0257	14	23.5.00	13	0.82	{0, 0, 13}	{0, 0, 13}	{0, 13, 0}	{0, 6, 7}
0257	15	23.5.00	12	0.92	{0, 0, 12}	{0, 8, 4}	{0, 0, 12}	{0, 0, 12}
0257	16	23.5.00	13	0.58	{0, 13, 0}	{**1**, 7, 5}	{0, 13, 0}	{0, 9, 4}
0257	18	23.5.00	14	0.63	{0, 3, 11}	{2, 9, 3}	{2, 7, 5}	{**1**, 12, **1**}
0257	19	23.5.00	9	0.78	{0, 0, 9}	{0, 0, 8}	{**1**, 5, 3}	{2, 4, 2}
0257	20	23.5.00	11	0.86	{0, 0, 11}	{0, 6, 5}	{0, 6, 5}	{0, 0, 11}
0257	21	23.5.00	13	0.89	{0, 0, 13}	{0, 0, 13}	{0, 5, 8}	{0, 6, 7}
0257	22	23.5.00	8	0.48	{0, 8, 0}	{2, 4, 2}	{0, 8, 0}	{2, 5, **1**}
0257	23	23.5.00	16	0.67	{0, 0, 16}	{0, 0, 15}	{4, 10, **1**}	{7, 8, 0}
0257	24.1	23.5.00	9	0.88	{0, 0, 9}	{0, 2, 6}	{0, 3, 6}	{0, 0, 7}
0257	25	23.5.00	4	0.59	{0, 0, 4}	{0, 4, 0}	{2, **1**, **1**}	{0, 4, 0}
0257	26	23.5.00	11	0.72	{0, 8, 3}	{0, 0, 11}	{0, 8, 3}	{3, 3, 5}

Table 4.1 continued

Site	Batch	Date	N	mean p	Locus			
					Bb7.4	*Bv12.19*	*Bv24.11*	*Bv24.12*
1258	3	29.4.01	9	0.48	{6, 0, 0}	{0, 7, **1**}	{**1**, 4, 3}	{0, 3, 2}
1258	4	5.5.01	8	0.48	{2, 4, 2}	{3, 5, 0}	{0, 3, 3}	{**1**, 3, 0}
1258	5	3.6.01	6	0.76	{0, 3, 3}	{0, 0, 6}	{**1**, 3, 2}	{0, 3, 2}
1258	7	3.6.01	6	0.48	{1, 2, 3}	{**1**, 4, 0}	{3, 0, 2}	{2, 2, **1**}
1290	1	28.4.01	13	0.73	{6, 4, 3}	{0, 6, 7}	{0, 4, 7}	{0, 0, 12}
1290	4	22.5.01	7	0.68	{**1**, 3, **1**}	{0, 6, **1**}	{0, 0, 5}	{0, 1, 1}
1290	7	22.5.01	8	0.66	{**1**, 2, 3}	{3, 3, 2}	{0, 0, 4}	{0, 2, 2}
1290	8	22.5.01	5	0.91	{0, **1**, 4}	{0, **1**, 4}	{0, 0, 4}	{0, 1, 2}
1290	9	22.5.01	20	0.76	{0, 4, 14}	{**1**, 12, 5}	{0, 5, 11}	{0, 6, 2}
1290	10	22.5.01	8	0.66	{0, 2, 6}	{0, 5, **1**}	{0, 3, 5}	{4, **1**, **1**}
1290	11	22.5.01	14	0.81	{0, 2, 12}	{0, 6, 8}	{0, 4, 9}	{0, 8, 3}
1290	12	22.5.01	9	0.41	{4, 5, 0}	{9, 0, 0}	{0, 0, 9}	{2, 5, 0}
1290	13	22.5.01	11	0.81	{**1**, 6, 4}	{0, 0, 11}	{0, 3, 7}	{**1**, **1**, 3}
1290	14	22.5.01	9	0.8	{0, 0, 9}	{0, 5, 4}	{0, 3, 5}	{0, 4, 0}
1290	15.1	22.5.01	11	0.41	{3, 5, 3}	{0, 5, 6}	{10, 0, 0}	{4, 6, **1**}
1290	16	22.5.01	11	0.71	{0, 11, 0}	{0, 6, 4}	{2, **1**, 8}	{0, 0, 6}
1290	17	22.5.01	13	0.71	{6, 6, **1**}	{0, **1**, 12}	{0, 0, 13}	{0, 8, 0}
1290	18	22.5.01	9	0.5	{3, 5, **1**}	{4, 5, 0}	{0, 5, 4}	{**1**, 0, 3}
1290	19	22.5.01	10	0.47	{5, 5, 0}	{2, 4, 4}	{0, 0, 7}	{6, 4, 0}
1290	20	22.5.01	10	0.66	{3, 4, 3}	{0, 0, 10}	{0, 2, 8}	{6, 3, **1**}
1290	21	22.5.01	8	0.86	{0, 4, 3}	{0, 0, 8}	{0, 2, 6}	{0, 2, 3}

Detecting genotyping errors as singletons

An initial examination of the number of {0,1,2} genotypes at a locus within a batch shows that most appear to have segregated in roughly Mendelian ratios. However, three batches stand out. In site 0257, batch 9 at locus *Bb7.4* the segregation is {9,1,1} for 0,1, and 2 genotypes respectively, which is highly skewed in favour of *B. bombina* alleles. It appears that this is a mixed batch, as there were two eggs (individuals 1 and 8) that were at a slightly different stage when they were first brought into the lab. Since there are nine others, I will simply discard these two eggs. This modified batch will be referred to as ' 0257-9.1'. The segregation at *Bv24.12* in family 24 of site 0257 is {1,0,7}, which is also strange. This appears to

be an isolated genotyping error for individual 9, or less plausibly an single egg from another batch. Since this egg is unexceptional at the other three loci, I will only discard the score at this locus. This is referred to as batch '0257-24.1'. The third bizarre segregation occurs in site 1290 in family 15 at locus *Bv24.11*. Here, the 13 genotypes are arranged {3,0,10}, which is highly implausible for a (1,1) mating. This is probably another mixed batch (although this is not reflected in the developmental stages of the eggs on sampling), and thus these three eggs (individuals 1, 2 and 3) will be discarded, making batch '1290-15.1'.

As noted above, not all mixed batches or errors will be this easy to spot. However, the less noticeable these problems are, the smaller the distortion they cause on inferences made from the affected batch. The segregations within the remaining batches are all much more feasible under Mendelian segregation, and it is much harder to identify any particular one as wrong. It is possible, however, to test for overall departures from Mendelian expectations. If there have been isolated genotyping errors, these will cause most problems when they change the expected parental genotypes for a batch, for example by mis-scoring a '1' in a batch containing only '2' genotypes changes the expected mating from (2,2) to (1,2). If scoring errors are rare, then more than one error per locus per batch is unlikely, and hence they will appear as a single anomalous genotype; I refer to these as 'singletons'. We can test for an excess of these across all the loci and batches by comparing the numbers observed with the numbers expected under normal Mendelian segregation with no errors. For a (0,1) or a (1,2) mating with a segregation ratio $s:r$, and a family of size n, the probability of getting $j = 1$ of either genotype is:

$$\frac{nr\,s^{n-1}}{1-r^n-s^n} \qquad [4.1]$$

(excluding cases $j = 0$ and $j = n$). The numerator refers to the probability of seeing one event r (nr) and the remainder as s (s^{n-1}). The denominator truncates the overall probability by excluding the cases where $j = 0$ or $j = n$ (i.e. s^n and r^n). The generating function for the underlying probability distribution is $1 + \frac{n p q^{n-1}}{1-p^n-q^n} (z - 1)$, and so the generating function for the total number of singletons in a sample of families of size n_i is:

$$f[z] = \prod_i \left(1 + \frac{n_i p q^{n_i - 1}}{1 - p^{n_i} - q^{n_i}} (z - 1)\right) \qquad [4.2]$$

(N. Barton, pers. comm.). With N families, this is an N'th order polynomial in z; the probability of observing k singletons in the whole sample is $\partial_{z,k} f[0]/k!$ (i.e. the probability of seeing k is recovered by differentiating the generating function k times and setting $z = 0$; N. Barton, pers. comm.). There are 14 (0,1) segregations in total; three of these have a singleton '0' genotype, and one has a singleton '1' genotype (see Table 4.1). The singleton '0's might have been produced by the misclassification of an heterozygote as a *bombina* homozygote, and vice versa for the singleton '1's. I assume that $s = r = 0.5$, as genotyping errors alone cannot distort the underlying process of Mendelian inheritance (although selection might: see Section 4.5). The probability of observing 4 singletons in either direction is 0.31, indicating that there is no excess of singletons in (0,1) segregations.

There are 69 (1,2) segregations, seven of these have a singleton heterozygote and five a single *variegata* homozygote. These may have been created by errors in a similar way to the (0,1) segregations above. There is no significant excess in either direction: the probability of observing five or more singletons in 69 matings is 0.44, whereas the probability of seven or more is 0.10. The probability of seeing 12 or more in either direction is 0.06.

In the case of (1,1) segregations, there are 13 from 43 with a singleton *bombina* homozygote, and 13 with a singleton *variegata* homozygote. Since the chance of producing a homozygote in a (1,1) mating is 0.25, r is set to 0.25 (s = 0.75) when considering *bombina* singletons, and vice versa for *variegata*. The probabilities of seeing 13 or more singletons is 0.20 for 13 or more on either side. There is a significant excess when the total number in (1,1) segregations is considered: the probability of seeing 26 or more in 43 segregations is 0.0006. In general, it is impossible to say which singletons are valid and which are not in all but a few extreme cases, and hence it is more appropriate to view each genotype as a potential error. I expand on this approach in Chapter 5.

Segregation patterns in the eggs

Under Mendelian expectations, the eggs will have the same allele frequency as the inferred parents, as both *bombina* and *variegata* alleles have an equal chance of being represented in the eggs, and each segregation should conform to Hardy-Weinberg frequencies. There are three plausible ways in which this can be distorted. Firstly, is there is selection increasing the probability of finding one allele over the other in the egg batch, either before zygote formation through an increased probability of inclusion in a gamete, or afterwards via selection against animals carrying a greater proportion of one genome or the other. This kind of selection could also be connected with habitat, for example if a batch has segregated for a gene adapting tadpoles to either habitat, those carrying the *bombina* version (i.e. pond adapted) in a puddle will be at a disadvantage. This will be detected at neutral markers because recombination will roughly halve the association between traits and marker alleles rather than removing it entirely.

The second type of distortion can arise through selection against the break-up of coadapted gene complexes in hybrids. Since hybrids are expected to be more

heterozygous at marker loci, their elimination will lead to a heterozygote deficit in the offspring. Thirdly, as mentioned above, a heterozygote deficit can also arise when there is assortative mating. This approach has wider applications: for example, Harushima et al. (2001) used the segregation ratios across a high density neutral marker map to locate incompatibility loci in rice strains.

The expected number of each genotype for (0,1) and (1,2) segregations in a given batch size n is given by the truncated binomial distribution. This is obtained from a standard binomial distribution and excluding the cases when $j = 0$ or n, and where j:(n - j) is the observed segregation. The truncation step is necessary because the segregations (0, n) or (n, 0) would be counted as having come from a different type of mating. The variable s ($r = 1 - s$) is the segregation ratio, and for all segregations, including (1,1), s can be viewed as the probability that the *variegata* allele will be sampled in the offspring. The general form of the binomial distribution is:

$$Pr[j] = \binom{n}{j} s^{(n-j)} r^{j} \qquad [4.3]$$

This truncated by renormalising without the cases where $j = 0$ or $j = n$.

$$Pr'[j] = \frac{Pr[j]}{1 - (Pr[0] + Pr[n])} \qquad [4.4]$$

For (1,1) matings, in a family of size n with segregation probabilities s^2,$2sr$, r^2, the probability of seeing (U,V,W) of the three possible (0,1,2) genotypes is:

$$Pr[U, V, W] = (n; U, V, W)\ (s^2)^U\ (2sr)^V\ (r^2)^W \qquad [4.5]$$

where (n;U,V,W) is the number of different ways of drawing U, V and W of each genotype from a total of n (i.e. the multinomial coefficient). This is truncated by excluding the cases where $U = 0$, $W = 0$ and $U+W = 0$:

$$Pr'[U, V, W] = \frac{Pr[U, V, W]}{\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} Pr[i, (n-i-j), j]} \qquad [4.6]$$

The best fit value for s was estimated with maximum likelihood. The most likely values are those that maximise $\sum_{i=1}^{N} \text{Log}[\Pr'(s \mid j_i, n_i)]$, where N is the total number of segregations under consideration, j is the observed number of '0' genotypes in (0,1) segregations and '1' in (1,2) segregations and n is the size of that batch. For (1,1) segregations, this function is $\sum_{i=1}^{N} \text{Log}[\Pr'(s \mid \{U_i, V_i, W_i\})]$, where $\{U_i, V_i, W_i\}$ are the numbers of each genotype in batch i.

We expect that selection for one allele or the other will vary in direction between sites, as *variegata* marker alleles are likely to be associated with *variegata* alleles at adaptive loci in puddles, and vice versa. By contrast, endogenous selection against hybrids will reduce the proportion of heterozygotes in all types of crosses, decreasing s for (0,1) segregations, leaving it unchanged for (1,1) and increasing it in (1,2). Since dispersal has created significant linkage disequilibrium between unlinked loci in the adult generation (see Section 3.3, Table 4.4), the forces acting on neutral loci should be roughly equivalent across the genome. Differences in s between marker loci should then only arise through drift or because a proportion of markers are linked to loci involved in hybrid fitness. There is slight heterogeneity in s across loci ($\Delta L_3 = 4.75$, P = 0.023) which is mainly because locus *Bv24.12* gives $s = 0.46$, whereas the remaining three give $s > 0.5$. However, only *Bb7.4* and *Bv24.11* are significantly different from 0.5 ($s = 0.55$, P = 0.017 and $s = 0.55$, P = 0.015 respectively). For *Bv12.19*, $s = 0.54$, P = 0.11.

Table 4.2 Estimates of s for each site and segregation. * indicates $P < 0.05$ and ** $P < 0.005$. No correction is made for multiple comparisons.

site	segs			All
	(0, 1)	(1, 1)	(1, 2)	
0200.4	0.68*	0.61*	0.64	0.63**
1315	0.29	0.54	0.50	0.53
0257	0.52	0.50	0.49	0.50
0258	0.77*	0.33	0.53	0.52
1290	0.53	0.61**	0.49	0.54**
All	0.59*	0.55**	0.50	0.53*

The most likely values for s for each site and segregation type separately are given in Table 4.2. What is immediately striking is that the majority of values are greater than 0.5, although most are not significantly so. Only in site 1315 do the s estimates fit the pattern expected for selection against heterozygotes, although only one of the three s values is significantly different from 0.5. There is slight heterogeneity in s between sites combining all segregations: $\Delta L_4 = 4.84$, $P = 0.046$. This may represent differing effects of habitat on the relative viabilities of *bombina* and *variegata* alleles, although genetic drift or scoring error could equally cause the slight differences. There is also slight heterogeneity between the segregation types ($\Delta L_2 = 3.12$, $P = 0.044$) when all sites are combined.

Over all sites, loci and segregations the best estimate of s is 0.53, which is significantly different from 0.5 ($P = 0.009$). This indicates an excess of *variegata* alleles are found in the eggs. This may be caused by systematic mis-scoring of *bombina* alleles as *variegata* or by selection against eggs and early stage tadpoles containing *bombina* alleles. Since mis-scoring is unlikely to be consistent in direction at different loci (see above), the latter explanation is more feasible. As most sites have $\overline{p} > 0.5$ in the adults, selection is presumably against *bombina*

alleles at trait loci introgressing into a *variegata* background, and the genome-wide linkage disequilibrium ensures that this is detectable with our markers as an increase in *s*. I include the potential for variance in segregation ratio and occasional genotyping errors into the more formal model of assortative mating in Chapter 5.

Testing for mixed batches

We presented a test in Nürnberger et al. (2002a) for detecting batches consisting of two or more distinct segregations. These are plausible in *Bombina* as more than one adult pair can deposit their eggs on a given plant stem. This need not be apparent in the field from the developmental stage, but mixed batches can be uncovered using the genetic data. For unlinked loci, we expect no correlations between genotypes at two different loci within the family. However, a mixing of families does generate correlations, which increase in strength with increasing difference between the two pairs of parents. I use two measures to investigate mixing: the variance in the hybrid index 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 hybrid index, averaged over all loci. To obtain null distributions, the observed genotypes at each locus are randomised among the offspring of the family. Both measures of correlation are computed for 1000 such shuffled datasets for each family (Nürnberger et al 2002a). The test statistic is the proportion of randomised batches (M) that give a higher variance or covariance than the original batch.

This approach is much more sensitive than the methods traditionally used to detect multiple matings, which normally rely on improbable genotype combinations (e.g. Akin et al. 1984; Trexler et al. 1997; Zane et al. 1999). However, more sophisticated techniques are becoming available (e.g. Emery et al. 2001) and I discuss these more fully in Section 5.1.

In Nürnberger et al. (2002a), the egg batches were genotyped at six loci, and simulations with artificial sets of mixed families showed that these gave sufficient power to detect about half of the mixed batches. This low level of power was not a large problem in that case, as the undetected mixtures are presumably those in which the pairs of parents are similar, and hence these instances of mixing should not unduly distort any inferences made assuming they were a single batch. In the extreme case, the mixing of families with identical parental genotypes would be completely undetectable. There are only four loci in this dataset, and these may not give sufficient resolution to make the test worthwhile.

The proportion of mixed batches detected in simulations under a variety of mixing schemes for four loci are shown in Table 4.3. The data are the proportion of 1000 mixed batches detected in simulations by the variance in Hybrid Index (var HI) or the sum of squared covariance (ssq CV) methods (see Nürnberger et al. 2002a). The former quantity is only sensitive to the mixing of families with distinct genotypic composition. The latter statistic is more sensitive, and should detect mixtures when the average hybrid index of the families is similar. This is because the mixing is detected from the correlation in allelic state within individuals in a mixed batch, and not from the overall HI of each individual. In the simulations, the parents were drawn from an adult population in linkage equilibrium or from a population with strong pairwise linkage disequilibria (D = 0.05). There were three types of family, one with an even mix of five eggs from one pair and five from the other, a second with nine from one pair and one egg from the other, and third an unmixed batch. The latter was included to show the proportion of unmixed batches that are detected as mixed by mistake (i.e. the Type I error rate).

Table 4.3: the proportion of 1000 mixed batches detected in simulations by the variance in Hybrid Index (var HI) or the sum of squared covariance (ssq CV) methods. The adults mated to form the batches were drawn from a population in linkage equilibrium (LE) or from a population with moderate pairwise linkage disequilibria (D = 0.05), denoted LD. The batch type column gives the mixture proportion for each batch. The column 'M' refers to the proportion of shuffled batches giving a greater statistic than the original (i.e. the significance level).

		var HI		ssq CV	
		LE	LD	LE	LD
batch type	M				
5 : 5	0.05	0.067	0.245	0.3	0.354
	0.01	0.02	0.155	0.153	0.208
9 : 1	0.05	0.039	0.118	0.083	0.121
	0.01	0.004	0.025	0.01	0.029
10 : 0	0.05	0.019	0.012	0.022	0.027
	0.01	0.003	0.001	0.004	0.007

There is at best a 35% chance of detecting a mixture (with a 5:5 mix and linkage disequilibrium), and therefore I will not attempt to detect mixed batches in these data. This is because the probability of mistakenly detecting an unmixed batch as mixed (~3%) approaches that for correctly detecting a mixed one (10-30%), and the risk of a Type I error becomes unacceptably high. Using more polymorphic loci may increase the the resolution, and I therefore return to the egg genotype data before the alleles are assigned as either *bombina* or *variegata* (see Section 2.3). There are three common alleles at *Bv7.4*, *Bv1219* and *Bv24.11*, and five at *Bv24.12*.

These contain more power for identifying mixed batches as there is a lower chance that the two segregations in a mixed batch involved exactly the same alleles. The egg batches in this format are presented in Appendix IV, where each is sorted into alphabetical order by the alleles in each column successively. When the eggs are presented in this way, it is relatively easy to search for mixed batches by eye, as distinct blocks of eggs across loci will stand out. It is also relatively simple to extend the above method to loci with more than two alleles by generalising the notation.

I represent the genotype of an individual at an n-allelic locus as a column vector with n elements: e.g. the genotype vector $\mathbf{G}$ for triallelic locus i would be $\mathbf{G}_i = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, where x ,y and z = 0 ,1 or 2 and x+y+z = 2. For example, $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ represents a heterozygote for the first and second alleles. The mean genotype across individuals $\overline{\mathbf{G}}_i = \begin{pmatrix} (\sum x)/2n \\ (\sum y)/2n \\ (\sum z)/2n \end{pmatrix}$, and the sum of the squared covariances between two loci $\mathbf{G}_1$ and $\mathbf{G}_2$ (which need not have the same number of alleles) is $\sum \left((\mathbf{G}_1 - \overline{\mathbf{G}}_1)(\mathbf{G}_2 - \overline{\mathbf{G}}_2)^{\mathrm{T}}\right)^2$, where T represents the transpose of that vector (T. Johnson, pers. comm.). The final test statistic is given by the mean over all pairs of loci. The critical feature of this scheme is that the test statistic (i.e the sum of the squared covariances) must be an monotonically increasing function of our belief that the family is not mixed (Hacking 1965; Sprott 2000). Extensive experimentation with artificially created batches have so far failed to find an exception to this assertion.

As above, null distributions for each family are generated by randomising the alleles at a locus between individuals, and the probability that the batch is mixed expressed by the proportion M of the randomisations that give a larger covariance.

This method is expected to be more powerful than when all loci are constrained to be biallelic, as here there is less chance that two subfamilies share the same alleles, increasing the similarity in state across loci in mixed batches. Simulations to quantify this improvement are underway. For the egg data, this approach detects none of the families as mixed; the batch with the most significant test statistic (0257-1) gave $M = 0.002$, which is unsurprising given 57 tests. For comparison, the mixed batches detected in Nürnberger et al. (2002a) all gave $M = 0$.

Joining batches from the same segregation

Bombina mating pairs are known to deposit their eggs in a variety of locations around the breeding site, and it is thus possible that some of our batches come from the same mating. There are two approaches to this problem. Firstly, it is possible to infer the segregations at each locus, giving a 'joint parental genotype' or jpg; these are denoted by e.g. {(0,1),(1,2),(0,0),(1,1)} for four loci. Batches that share the same jpg and that were at a compatible developmental stage when collected can be joined together into a single batch. Alternatively, it is plausible that batches coming from the same mating do not share a jpg. For example, both a (0,1) and a (1,2) segregation could have come from two heterozygous parents. In the same way, a batch entirely homozygous for *variegata* alleles could have arisen from a (2,2) mating, a (1,2) or a (1,1), although the latter two involve increasingly unlikely Mendelian segregations. This implies that the same batch can have more than one jpg: in the extreme case, any batch could be produced by improbable (1,1) segregations at all loci.

Using the binomial or multinomial distribution, it is possible to work out the probability that a particular mating type gave rise to any set of offspring genotypes; I discuss this in greater depth in Section 5.2. Whether or not two batches should be joined could then be determined by comparing the probability that they share a

joint parental genotypes with the probability they came from separate jpg's. However, the threshold difference in probability above which batches should be joined is unclear. One cannot use the χ^2 approximation for differences between two likelihoods in this case, as the two hypotheses are not based on estimates of a parameter. On a more practical note, the probability of two distinct pairs of parents giving the same jpg is quite high, especially when the adults are genotypically similar. For example, there are only 324 unique joint parental genotypes at four loci from a total of 3240 distinct pairs of parents.

The second approach is simpler and more pragmatic: do nothing. Since not all batches need be joined, mainly because they come from different laying dates (Table 4.1), the statistical consequences of a slight inflation of the number of matings under consideration will be small. Furthermore, any statistic calculated considering the smaller sub-batches separately will give them a lower statistical weight, reducing the impact of leaving them unjoined. By comparison, the consequences of mistakenly joining unrelated batches are more serious, as the proportion of heterozygotes in the inferred parents will be artificially inflated. I will therefore consider each batch as a separate entity, and ensure that any statistic calculated for the eggs takes the batch size into account.

Section 4.5 Changes through the lifecycle

I now move on to consider the relationships between the adults and the offspring. This can be divided into the three stages laid out in Section 4.2 above. First, how does the observed active habitat preference affect the composition of the mating adults in the focal sites? There are two extreme scenarios: either the toads choose their waterbodies on the basis of their preferred habitat for resting and feeding, and simply mate with whoever else is present, or they make a deliberate journey to the mating site of their choice, which might be totally unlike the habitat they normally occupy. Under the former scenario, the allele frequency of the adults sampled in the focal site will match that of the joint parental genotypes inferred from the egg batches (see above), whereas under the latter they may be totally unrelated. The real situation, of course, will be somewhere in between the two. Nürnberger et al. (2002a) found that the breeding adults in a puddle in Croatia were much more *B. variegata*-like than the adults present as a whole, indicating that the *B. bombina* animals may have chosen not to breed there, and there may well be a similar pattern here. To explore this, I first gauge the strength of habitat preference from the differences in allele frequency between the adult pool and the focal site. I then examine the relationship between the focal adults and the inferred parents. Lastly, to quantify the strength of direct mate choice within habitats, I look at the heterozygote deficit of the offspring, which could indicate non-random mating (or selection: see Chapter 3).

Are we likely to detect assortative mating in these data? If the adult pools are genotypically homogeneous, female toads may be unable to express a mate preference because their preferred phenotype is not present. In some taxa, this might result in an overall reduction in the frequency of matings (as females reject all the available mates). Unfortunately, the necessary data on genotypic

composition and reproductive output has not been collected in *Bombina*. Alternatively, the females may just take any mate in the absence of her preferred type. These two scenarios correspond to the absolute and relative mate preference models described by Lande (1981). In either case, there is no opportunity to detect assortative mating. Only when there is a wide range of genotypes in the breeding aggregation can mate choice be fully expressed, and can only be detected at neutral markers if there is genome-wide linkage disequilibrium generated by dispersal (see Section 3.1). Fortunately, the mixing of distinct gene pools simultaneously generates a wide range of local genotypes and linkage disequilibrium, and this appears to be the case for three of the four adult pools: 0200.4, 1315, 0257 and 1258 all have significant linkage disequilibrium when R is constrained to be the same across all sites (see Section 3.3; Table 4.4). I estimate the full set of associations within and between loci in Section 5.3.

Table 4.4: Heterozygote deficit F_{IS} and linkage disequilibrium R for the adult pools, the focal adults and the eggs. A significant deviation from neutral expectations is denoted by * for $p < 0.005$.

Adult Pool	pool Fis	focal Fis	egg Fis	pool R	focal R
0200.4	0.06	0.01	0.19*	0.05*	0.28*
1315	0.06	0.29	0.004	0.05*	0.04
0257	0.16*	0.11	0.001	0.17*	0.18*
0258	0.02	−0.05	0.04	0.18*	0.06
1290	0.07	0.19	0.18*	0.05	0.05

In Chapter 3, I showed that the inferred migration rate of $m = 0.2$ between ponds and intermediate sites could also explain the observed heterozygote deficit in the latter. When F_{IS} is estimated for the adult pools the pattern is inconsistent: only the 1258 adult pool has a significant deficit of heterozygotes. Nonetheless, the F_{IS} generated by assortative mating in previous generations is only expected to

persist if there is no adult migration, which is clearly not the case. I will therefore not place much emphasis on heterozygote deficit in the adults, as the uneven patterns presumably reflect the limited sample sizes and the effects of migration rather than the effects of assortative mating.

Table 4.5: Table of parameters for each adult pool. 'n' gives the number of sites in each adult pool (including samples from different years in the same site), 'Np' the total number of animals in each adult pool, and 'Nf' gives the number of individuals in each focal site. In the remaining columns, '$\overline{p}$' refers to allele frequency, F_{IS} to the habitat score. The 'ΔH; $\Delta \overline{p}$' shows the difference in H and the difference in $\overline{p}$ between the pool and the focal site, and 'ΔH;ΔIP' shows the same for the difference between the pool adults and the inferred parents.

Pool	n	Np	Nf	pool H	focal H	pool $\overline{p}$	focal $\overline{p}$	Inf Par
0200.4	17	143	12	0.44	0.45	0.71	0.72	0.83
1315	17	143	10	0.44	0.22	0.71	0.65	0.78
0257	5	72	28	0.47	0.65	0.58	0.66	0.76
1258	7	99	31	0.55	0.31	0.67	0.59	0.53
1290	3	30	12	0.31	0.25	0.51	0.46	0.66

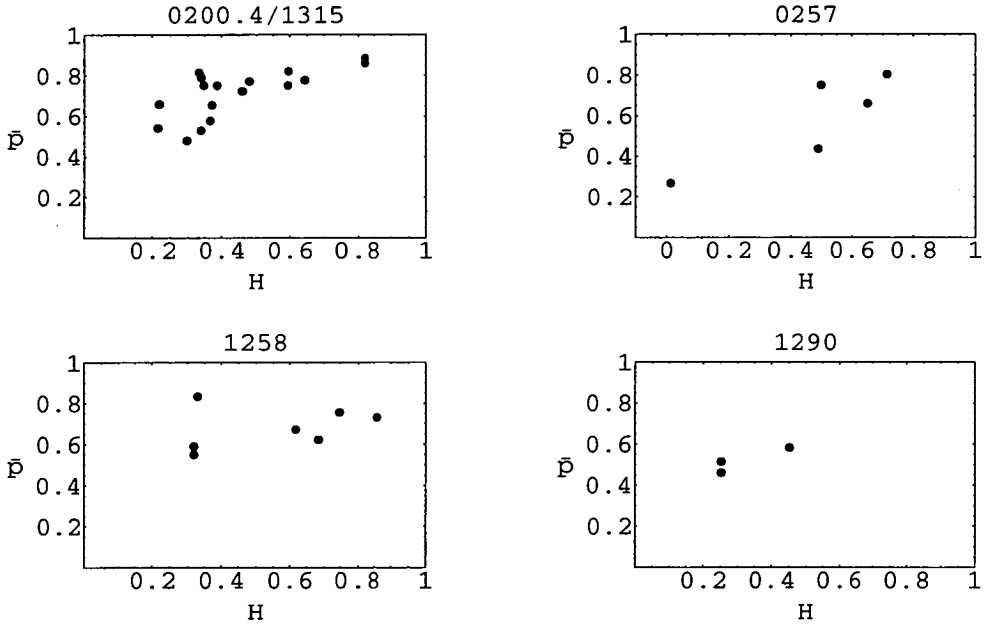
Pool	ΔH; $\Delta \overline{p}$	ΔH; ΔIP
0200.4	0.01; 0.01	0.01; 0.12
1315	−0.22; −0.06	−0.22; 0.07
0257	0.18; 0.08	0.18; 0.18
1258	−0.24; −0.08	−0.24; −0.14
1290	−0.06; −0.05	−0.06; 0.15

Adult pools to focal adults: habitat preference

The relationship between habitat and allele frequency in each is shown in Figure 4.2, and the genotypic composition of both the adult pools and the focal sites in summarised in Table 4.5. As in Chapter 2, there is evidence for an active habitat preference, particularly in the 0200.4/1315 and the 0257 pools, and the correlation between $\overline{p}$ and H is significant in both ($p < 0.005$ and $p < 0.05$ respectively). The other two adult pools are less clear, although a roughly positive trend can be seen in both cases. This implies that the adults in the focal sites will not be a random

subsample of the pool as a whole. When the habitat type of the focal site differs from the mean of the sites contributing to the adult pool, there should a corresponding difference in allele frequency between the pool and the focal site as well. This is only true for four of the focal sites, as 0200.4 has an almost identical habitat score to the pool average. The deviations are shown in the '$\Delta H;\Delta\overline{p}$' column of Table 4.5.

Figure 4.2: Plots of mean *variegata* allele frequency 'p' against habitat score 'H' for each adult pool. Note that several of the points are samples from the same habitat in different years, and the data for adult pool 0200.4/1315 is also shown in Figure 2.10.

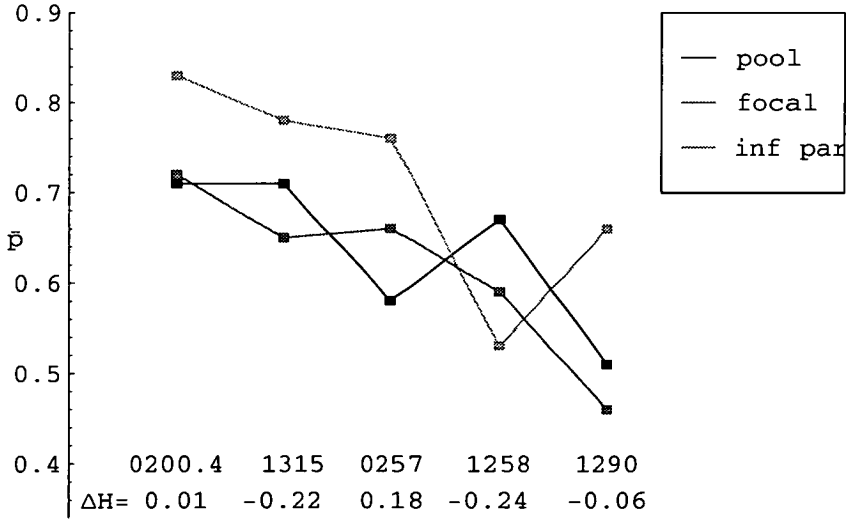


This habitat preference may considerably reduce the range of males from which a female may choose, as only a non-random subset will be found in a particular habitat. However, the adults sampled in the focal habitat may be totally unrelated to the animals that chose to breed there the night before (see above). This can be roughly quantified by the difference in allele frequency between the focal adults and the inferred parents, and I address this issue next.

Adults to inferred parents: breeding site preference

It is possible to infer the alleles in the parents from the segregation in the offspring, and from this construct a joint parental genotype across the marker loci. Since neither parent has been sampled with the eggs, it is impossible to assign the alleles to being in the male or the female, but the jpg's nonetheless contain useful information. In the analyses in this chapter I consider only the most plausible jpg: for example, a batch containing a mixture of '0' and '1' is most plausibly due to a (0,1) mating, although a (1,1) mating producing no '2' genotypes is also possible. I will take this ascertainment bias into account in the fuller analysis in Chapter 5. The joint parental genotypes for all the batches of eggs from a focal site are known as the inferred parents, and although they cannot be used to estimate linkage disequilibrium in the parents, they are a reliable indicator of both their allele frequency and their heterozygote deficit.

Figure 4.3: The relationship between the allele frequencies in the adult pools, the focal sites and the inferred parents ('inf par') in each of the adult pools. 'ΔH' gives the difference in habitat score between the pool and the focal site.



The relationship between the the allele frequencies in the adult pools, the focal sites and the inferred parents is shown in Figure 4.3, the values themselves are in Tables

4.5 and 4.6. Since there is little correspondence between $\overline{p}$ in the focal adults and the $\overline{p}$ of the inferred parents, it seems unlikely that the adults within a focal site mate at random. Instead, the $\overline{p}$ of the inferred parents does not correspond well with either the pool or the focal adults. This suggests that mating aggregations are a non-random subset of the wider adult pool, and that most toads move away from their breeding site before they can be sampled. It is interesting to note that the difference in habitat score between the pool and focal site (ΔH in Figure 4.3) predicts the difference between the pool and focal adults quite well, but does not correspond well with the inferred parents. This may be because the habitat axis is a poor reflection of breeding habitat choice, or because a female chooses to lay her eggs in a site for other reasons. I come back to this result in the discussion section of Chapter 5.

Table 4.6: Changes through the lifecycle in the focal sites. Numbers shown are the frequencies of 0, 1 and 2 genotypes at each stage. 'Inf Par' refers to the parental genotypes inferred from the batch data (see text), and 'Exp Eggs' refers to the expected frequencies if the inferred parents had produced offspring in Mendelian ratios. An asterisk indicates $p < 0.05$.

			genotypes			Fis
site	stage	p	0	1	2	
	Pool Adults	0.71	0.12	0.32	0.54	0.06
	Focal Adults	0.72	0.07	0.40	0.52	0.01
0200.4	Inf Par	0.83	0.05	0.23	0.71	0.011
	Exp Eggs	0.82	0.06	0.22	0.71	–
	Eggs	0.85	0.04	0.19	0.75	0.192*
	Pool Adults	0.71	0.12	0.32	0.54	0.06
	Focal Adults	0.65	0.11	0.45	0.43	0.29
1315	Inf Par	0.78	0.05	0.32	0.62	−0.008
	Exp Eggs	0.78	0.03	0.34	0.61	–
	Eggs	0.79	0.04	0.31	0.63	0.004
	Pool Adults	0.58	0.24	0.34	0.41	0.16*
	Focal Adults	0.66	0.11	0.44	0.43	0.11
0257	Inf Par	0.76	0.06	0.34	0.58	0.000
	Exp Eggs	0.75	0.06	0.35	0.58	–
	Eggs	0.75	0.05	0.36	0.57	0.001
	Pool Adults	0.65	0.13	0.41	0.44	0.02
	Focal Adults	0.59	0.16	0.48	0.34	−0.05
1258	Inf Par	0.53	0.15	0.62	0.21	0.000
	Exp Eggs	0.53	0.24	0.43	0.31	–
	Eggs	0.54	0.21	0.47	0.30	0.043
	Pool Adults	0.51	0.27	0.42	0.30	0.07
	Focal Adults	0.46	0.28	0.49	0.21	0.19
1290	Inf Par	0.66	0.08	0.50	0.41	−0.02
	Exp Eggs	0.65	0.13	0.42	0.44	–
	Eggs	0.67	0.14	0.35	0.50	0.184*

Inferred parents to eggs: assortative mating

Once the breeding aggregation is formed (by whatever process), the toads will form pairs and deposit eggs around the site. Features of the mated pairs can then be inferred from the offspring data. For example, under random mating, the parents will be a random subset of the adults in the breeding site. If there is some sort of mate choice, the pairs of parents will be more similar to each other than would otherwise be expected. This will generate a heterozygote deficit at neutral markers in the eggs, but not in the inferred parents (as all types of animals have equal mating success). There may also be discrimination against animals with a hybrid phenotype, which will increase the relative contribution of the more extreme phenotypes relative to the intermediate hybrids. This will generate a deficit of heterozygotes in both the inferred parents and the offspring, as intermediate animals are expected to be more heterozygous when there is linkage disequilibrium (see above). Table 4.6 shows the genotype frequencies for the adult pool, the focal adults, the inferred parents and eggs for each site. The 'Exp Eggs' rows show the genotype frequencies in the eggs under Mendelian expectations.

There is no evidence in Table 4.6 for consistent deviations away from Mendelian segregation in the inferred parents, showing that, for these data at least, there is no evidence of mate discrimination against hybrids. I make separate estimates for both the heterozygote deficit and linkage disequilibrium in the parents in Section 5.5, where I will discuss these estimates in greater depth.

The slight differences between the Exp Eggs and the Eggs rows could be due to sampling error between parents and offspring (i.e. genetic drift). Additionally, genotyping errors and biased segregation ratios will also cause deviations. I found little evidence in Section 4.3 for an excess of singletons, which implies that random scoring errors are not very common. There is good evidence that *B. variegata*

alleles occur in the offspring more often than expected. This might be due to systematic mis-scoring at three of the four loci (see Table 4.2), or due to selection against *B. bombina* alleles at linked loci. Whilst this might indicate that gametic or postzygotic isolation is present in *Bombina*, their effects are an irritation when trying to infer assortative mating, as this might also generate a heterozygote deficit in the eggs. I develop a model for assortative mating that accounts for both segregation bias and genotyping errors in Chapter 5.

There is a heterozygote deficit in the eggs in all sites, but it is only strong in 0200.4 and 1290, where F_{IS} is 0.19 and 0.18 respectively (see Table 4.4). Whilst this may be due to a lack of power in 1315 and 1258, there are 24 batches in 0257, and here F_{IS}= 0.001. Is this evidence for assortative mating in 0200.4 and 1290? Firstly, both these sites have significantly skewed segregation ratios (Table 4.2), and this may be contributing to the deficit. Secondly, because the eggs are sampled in batches, there is greater potential for sampling error to create strong heterozygote deficit or excess when the number of parents is small. It is therefore unclear at this stage what might have driven these difference in F_{IS} between sites. If assortative mating is the main cause, why might its strength be so variable between sites? The same argument applies to selection against hybrid offspring and systematic genotyping errors. In Chapter 5 I present more sophisticated estimates of assortative mating within each site, and obtain an independent estimate from a modified version of the moment-based appraoch. I also conduct simulations to test the possibility that this level of F_{IS} arose by sampling error alone.

Section 4.6 Summary

Since Chapter 5 will also be devoted to quantifying assortative mating, albeit with slightly different techniques, I will defer a full discussion of the results presented here until the end of that chapter.

1) The role of habitat in the mating system of *Bombina* precludes the use of lab-based mate trials in assessing assortative mating, and catching amplexed pairs is also too difficult. Indirect methods based on the genotypes of the offspring must therefore be used. One further complication is that neither parent is sampled with the eggs, and thus it is impossible to recreate the parental genotypes.

2) We collected eggs and adults throughout the breeding season in 2000 and 2001 from five sites in the Apahida hybrid zone, and these were genotyped at four marker loci by S. Köhler.

3) As toads are highly mobile, I included adults from sites within 300 m of the egg collection site with the local adults, forming a larger pool of potential parents.

4) One can only make accurate inferences about mating behaviour from error-free egg batches that come from a single segregation, and therefore I check for genotyping errors appearing as an excess of singletons, and present several methods for finding mixed batches. I also checked for skew in the segregation ratios, and found a slight bias in favour of *B. variegata* at three of the four loci.

5) There was significant linkage disequilibrium in three of the four adult pools, but the pattern for heterozygote deficit was less clear.

6) I analysed the changes in allele frequency between the adult pool, the focal adults, the inferred parents and the eggs, and found that the largest differences lay between the pool and the inferred parents, implying that breeding aggregations were a non-random subset of the local adults, although the effect of habitat on mating behaviour was unclear.

7) There was some support for assortative mating among the parents, as there was a heterozygote deficit in two of the five egg sites. However, alternative explanations for this pattern cannot be ruled out at this point.

Chapter 5:
Quantifying mating behaviour in *Bombina*. II. Models of mating and moment estimation

Section 5.1 Introduction

In Chapter 4 I used changes in allele frequency between adult and the inferred parents to quantify how the parents related to the adult pool or the focal adults, and tested for assortative mating with the heterozygote deficit. This approach discards much of the information contained within the egg batches, and in this chapter I present two techniques exploiting this to uncover more about the mating behaviour of the parents. The first method is based on parentage inference, in which the segregations in the offspring are used to identify how the parental pairs relate to the sampled adults. I first estimate the best fitting segregation ratio and the genotyping error rate, and then use the batch data and an estimate of the adult genotype frequencies to fit parameters for breeding habitat choice and assortative mating. The second uses the inferred parents for each batch to directly estimate their mean heterozygosity, their linkage disequilibrium and the correlation between their hybrid indices without reference to the adult sample. In this section I will present the background and underlying logic of each method, and review how my approach relates to other analyses of offspring genetic data.

Parentage inference

The techniques for inferring parentage were originally developed in the late 1930's for excluding candidate males as the father of a particular child (Essen-Möller 1938, in Smouse and Meagher 1994). The basic principle is that the father must possess the correct combination of alleles to have produced the child, given knowledge of the mother and the child's genotype. The probability of excluding a falsely accused male $P[E_i]$ on the basis of a single codominant locus with allele frequencies p and q (=1-p) is $P[E_i] = pq(1-pq)$, and hence loci with intermediate alleles

frequencies are more useful as parents are more likely to carry different alleles (Neel and Schull 1954, cited in Chakraborty et al. 1974).

Multiplying these probabilities over many independent loci can generate a higher probability of excluding uninvolved males. This is given by $P[E] = 1 - \prod (1 - P[E_i])$. Initially, the data were restricted to blood groups and a relatively small number of allozyme variants, and the probability of excluding a non-parent was only around 95% (Chakraborty et al. 1974). With the advent of highly polymorphic DNA markers such as microsatellites, the potential for excluding non-fathers became much higher, mainly because the chance of an uninvolved male carrying the same alleles as the true father becomes very small. Nonetheless, it is much harder to exclude all but one of the candidate males when any of them could have been father, as is the case for samples from natural populations (Chakraborty et al. 1988).

Even if a male is not excluded, the segregations required for him to have been the father may be very improbable, and it is obviously useful to take this into account when assigning paternity. Likelihood methods (Edwards 1972) are the best way to achieve this, as they provide a natural 'measure of belief' that a particular male transmitted the alleles observed in the offspring, based on Mendelian segregation. However, if the genotype of the father is unknown, it must be inferred from his offspring and their mother. For example, when the offspring are homozygous at a biallelic locus, and the mother is heterozygous, the likelihood of the father being a homozygote or a heterozygote is 0.5 and 0.25 respectively (see Section 5.2). These likelihoods can be multiplied across unlinked loci to give the total likelihood that a particular male was the parent given his multilocus genotype. The candidate males

can then be ordered by their likelihood of having been the father. Thompson (1975) showed that this approach was useful in inferring human relationships in a pedigree, and Meagher (1986) extended it to paternity inference in wild populations. Since the ultimate goal is the assignment of paternity to a single individual and the exclusion of all others, these are collectively known as 'categorical' methods.

Significant advances have been made in paternity methods in the last five years. For example, Marshall et al. (1998) introduced a Monte Carlo simulation procedure for determining confidence intervals for the assignment of parentage to an individual. Perhaps more importantly, they considered the possibility that their sample of males did not contain all the fathers, and instead worked from the assumption that the distribution of reproductive successes was similar in both the sampled and unsampled males. The role of genotyping errors in the exclusion of potential males has also been increasingly discussed. Marshall et al. (1998) modelled errors as the probability that a particular allele would be incorrectly observed as another allele in proportion to their frequency in the population, whilst Sanchristobal and Chevalet (1997), following Chakraborty et al. (1974), used a fixed probability 1-ϵ that the observed allele is the true one, and $\frac{\epsilon}{a-1}$ that it was any other of the a alleles.

Another thread of parentage analysis is the fractional assignment of paternity to a known population of males, principally when the parameter of interest is the distribution of male reproductive success. It was developed to eliminate the bias towards selecting homozygotes as parents found in categorical methods, and the original model assigned males to offspring with a known mother (Devlin et al. 1988). This was extended to cope with situations where neither parent was sampled with the offspring by Roeder et al. (1989). Both the latter study and Smouse and Meagher

(1994) essentially combine the probabilities that a particular individual sired each of the offspring, and assign it a total probability of having contributed. This is equivalent to a Bayesian procedure in which all males are given an equal prior probability of paternity, which involves the assumption that all individuals are equally likely to be fathers. This is implausible for natural populations, and therefore Neff et al. (2001) developed a more explicitly Bayesian version of the fractional method which can include non-genetic data such as morphology into an individuals prior probability of parentage.

The fractional method can also be used to test hypotheses on the determinants of male reproductive success. This involves multiple regression of male fertility onto male traits to assess their contribution to fecundity (cf. Lande and Arnold 1983). For example, Smouse et al. (1999) found a high variance in male reproductive success in a population of *Chamaelirium luteum*, but inter-mate distance was a better predictor of paternity than floral morphology. Morgan and Conner (2001) further refined this approach by introducing Newton-Raphson optimisation for the selection coefficients, and they show with simulations that this method is statistically powerful in closed populations of 400 or more individuals.

One major restriction of the above fractional paternity models is that all the parents must be included in the sample, which is almost impossible for a natural population. The earliest models that take account of unsampled parents are those that infer the rate of extra-pair parentage in bird nests from the presence of alleles not carried by the attendant mother and father (Westneat et al. 1987; Wrege and Emlen 1987). These were extended to provide estimates for the proportion of a batch of offspring fathered or mothered by a putative parent by Neff et al. (2000a,b), given the allele

frequencies in the adults. Marshall et al. (1998) also used the adult allele frequencies to generate the genotype frequencies of unsampled males and compare the likelihoods that the father is a specific sampled animal versus an unknown unsampled one, which assumes that reproductive success is the same for sampled and unsampled males. More recently, Nielsen et al. (2001) modified the full Bayesian version of the fractional paternity model to include the possibility that some males were unsampled. They used a decision-theoretic framework, in which the candidate male was only assigned the father if he was the most likely out of the sampled males and his probability of being the father was larger than the posterior probability that the father was not sampled.

Several features of *Bombina* preclude using the models outlined above to measure assortative mating, and hence I developed my own approach. This had to simultaneously overcome several of the problems addressed by individual models above. Firstly, neither the male or the female is sampled with the offspring; second, given the high rate of movement between sites (T. Sands, unpublished data; Chapter 3), it is unlikely that all the parents will be in the sample. Furthermore, none our four marker loci has more than five alleles, and thus there is very little power for the direct assignment of parents from the sampled adults (see Nielsen et al. 2001). In addition, these data are from a hybrid zone, and there is strong linkage disequilibrium generated by admixture. This means that multilocus genotype frequencies cannot be calculated from the product of their individual allele frequencies, and nor can they be estimated from even large samples of adults (Barton 2000). Lastly, there may well be genotyping errors at all four of our loci, and there is some bias in the segregation ratios (see Chapter 4).

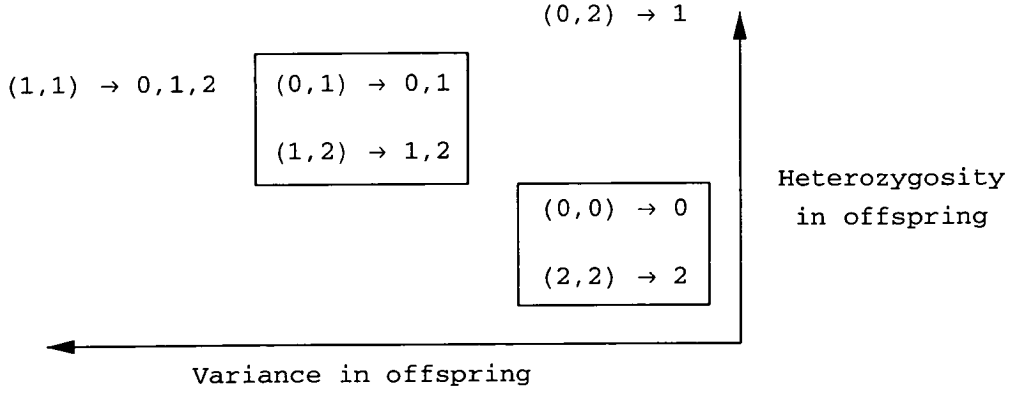
The low resolution of our markers, the undersampling of the adult pool and the lack of a definite maternal or paternal reference genotype means that mating behaviour cannot be inferred from the relationship within and between pairs of assigned parents. The one advantage of these data over that used in the methods above is that the offspring are sampled as batches coming from a single mating, and thus the alleles present in the parents can be reliably inferred, given a segregation ratio and the mis-scoring rate. However, unlike Neff et al. (2001) and Nielsen et al. (2001), it is not worth finding pairs of adults carrying compatible sets of alleles and assigning them as the parents, because our resolution is so low. Instead, I calculate the multilocus genotype frequencies in the adults from estimates of all the within and between locus associations in the sampled adults. This is the original application of the moment estimation approach laid out in Barton (2000) and briefly summarised in Section 5.3; a modified version is used to infer associations in the parents later on in Section 5.5 (see below). I use these adult genotypes frequencies as the prior probability that a particular pair of parental genotypes were present to produce a given batch. The most likely values of a range of parameters describing the mating behaviour of the adults can then be found conditional on these prior probabilities. The inference of parental genotypes is described in Section 5.2, the estimation of the adult genotype frequencies with the original moment estimation approach is presented in Section 5.3, and I fit a likelihood model of mating behaviour in *Bombina* in Section 5.4.

The estimating moments for the parents

This second approach to the data is motivated by the uncertainties of the adult sample, and was developed by N. Barton as an extension of the original moment

estimation procedure (Section 5.3). It relies on the associations within and between loci in a batch of eggs to infer features of the parents without reference to the adult sample. The flow of information from the parents to the offspring can be illustrated by considering the six possible mating types involving a single biallelic locus. As in Chapter 4, I describe a single locus genotype in terms of the frequency of *B. variegata* alleles, i.e. 0,1 or 2. For example, a mating between a *B. bombina* homozygote and a heterozygote, denoted a (0,1) mating in the notation of Chapter 4, gives '0' and '1' genotypes in the offspring in equal proportions (denoted a 0,1 batch). The variance in allelic state within a batch is thus a function of the heterozygosity of the parents, as only heterozygous parents can produce segregations in the offspring. In the same way, the correlation between the number of *B. variegata* alleles carried across loci in the parents is given by the mean heterozygosity across all loci in the offspring. The interplay between these two is shown schematically in Figure 5.1.

Figure 5.1: The six possible mating types at a biallelic locus ranked in terms of a) the proportion of heterozygous offspring and b) the allelic variance within sibships. Boxed mating types behave identically. This figure is taken from Nürnberger et al. (2002a).



In Nürnberger et al. (2002a) the associations (referred to here as cumulants) between alleles in the same or different loci in the offspring were used to estimate cumulants of interest in the parents, such as the correlation between their hybrid indices (a measure of assortative mating) or their linkage disequilibrium. Here I take a slightly simpler approach and estimate the cumulants directly from the inferred parents within each site. This has the advantage that it is not affected by non-Mendelian segregation in the offspring, although it is more sensitive to sampling error in small batches than the inferences made indirectly from the offspring. Nonetheless, simulations suggest that the two give similar answers for the sample sizes used here (N. Barton, pers. comm.). There are three cumulants of particular interest: the heterozygosity of the parents, their pairwise linkage disequilibrium and the correlation between their hybrid indices, (i.e. their total number *B. variegata* alleles across all four marker loci). I present the calculation of these three in Section 5.5, a more general account of moment estimation for the parents can be found in Appendix II of Nürnberger et al. (2002a).

This approach is related to a number of other methods involving estimating the relationships between sets of offspring in the absence of parental information, and there are three identifiable threads to this work. Firstly, there are techniques that estimate the relatedness of individuals based on their similarity at marker loci, usually as an average over all the possible pairwise combinations of the sample (Queller and Goodnight 1989; Ritland 1996; Lynch and Ritland 1999). Data of this type can be used to estimate the contributions of additive and dominance genetic variance to a quantitative trait (e.g. Lynch and Walsh 1998), or for testing hypotheses over the role of relatedness in the evolution of social behaviour (see Frank

1998). The work by Ritland (1996) is of particular relevance here because he infers inbreeding from a moment-based estimate of the intra-locus associations between alleles.

The second thread is similar in that it involves mostly pairwise statistics, but these are instead used to classify individuals by their relatedness, for example into full- or half-sibs and unrelated individuals. The use of maximum likelihood for this was pioneered by Thompson (1975), and continued by Painter (1997). Recently, these analyses have been extended to allow the reconstruction of whole families of full- and half-sibships using likelihood-based Markov Chain Monte Carlo techniques (Thomas and Hill 2000; Smith et al. 2001). Another noteworthy contribution is from Emery et al. (2001), who used a Bayesian method to partition the eggs in squid egg strings into full and half-sib families using only five microsatellite loci. Their approach has features in common with both the moment-based estimates and the detection of multiply mated batches (e.g. Neff et al. 2000a,b).

The third thread is related to the assignment of individuals into families, but in this case the assignment is into population of origin. The majority of methods are Bayesian in structure, and can use even slight allele frequency differences between populations to identify immigrants or their offspring provided enough loci are available. For example, Rannala and Mountain (1997) used 50 RFLP loci to demonstrate that human immigrants were detectable even when the population differentiation was slight (F_{ST} = 0.056). Pritchard et al. (2000) extended this approach to the detection of population structure from a large sample of individuals; this method simultaneously optimises the number of populations and assigns individuals to them. Dawson and Belkhir (2001) considerably improve on this method, and dis-

cuss the statistical issues involved in partitioning groups of individuals into populations. Unfortunately, their method is computationally too intensive to cope with large samples of individuals or many possible populations. One feature shared by these models is the assumption that migration does not generate significant linkage disequilibria, which is less tenable when each population is small.

The extension of moment estimation approach to inferring features of the parents presented here is roughly analogous to many of the above techniques, in that it uses the relationships within and between loci in different individuals to make inferences about the nature of their parents (their identities, where they came from etc.). However, since the offspring are clearly all full sibs, much more information about the parents in general can be derived from the batches. I present the original method in Section 5.3 and the extension to dealing with family data in Section 5.5.

Section 5.2 The likelihood of parents

Joint parental genotypes

The first step in inferring assortative mating parameters from the eggs batches is to work out what parental genotypes were involved at each locus. In this section I show how this can be calculated taking into account potential bias in segregation ratio and genotyping error. I then make maximum likelihood estimates for both the segregation ratio and the error rate from the egg batches. As mentioned in Section 4.4, it is possible to express the probability that a particular mating type gave rise to a set of egg genotypes. The probability of observing a batch containing $\{U,V,W\}$ of each genotype for each mating type is proportional to:

$$\begin{aligned}
&\Pr[(0,0) \mid \{U,V,W\}] \propto 1^U\, 0^V\, 0^W \\
&\Pr[(0,1) \mid \{U,V,W\}] \propto \left(\frac{1}{2}\right)^U \left(\frac{1}{2}\right)^V 0^W \\
&\Pr[(0,2) \mid \{U,V,W\}] \propto 0^U\, 1^V\, 0^W \\
&\Pr[(1,1) \mid \{U,V,W\}] \propto \left(\frac{1}{4}\right)^U \left(\frac{1}{2}\right)^V \left(\frac{1}{4}\right)^W \\
&\Pr[(1,2) \mid \{U,V,W\}] \propto 0^U \left(\frac{1}{2}\right)^V \left(\frac{1}{2}\right)^W \\
&\Pr[(2,2) \mid \{U,V,W\}] \propto 0^U\, 0^V\, 1^W
\end{aligned} \qquad [5.1]$$

This approach also accounts for the ascertainment bias mentioned in the previous chapter: a set of n eggs consisting only of '2' genotypes could come from a (2,2) mating, a (1,2) mating or a (1,1) mating with probability 1, 0.5^n, and 0.25^n respectively. Since our ultimate aim is to compare the probabilities of each mating given $\{U,V,W\}$, the multinomial coefficient will be constant for all mating types, and for brevity is left out.

Marshall et al. (1998) and Sanchristobal and Chevalet (1997) both model genotyping errors on the basis that an allele is correctly observed with probability 1-ϵ, and mistakenly scored as one of the others with probability ϵ. Since my loci are biallelic, I use a slightly simpler model whereby the genotypes (and not alleles) that may or may not be correctly observed. In this case, 1-ϵ is the probability that the correct genotype has been observed, and ϵ that we have observed either of the other two. This type of error may arise because individual samples are mistakenly assigned to a different batch, or from data entry errors. This model does not cover assigning alleles to the wrong taxa, which is included in the segregation ratio, and nor does it account for confusing the identity of larger groups of samples, which was hopefully avoided through stringent lab procedure. Including an error rate is important because it deals with the problem of singletons: without it, a batch with {0,0,0,0,0,1} would always be counted a (0,1) mating, but with the inclusion of the potential for errors, a (0,0) mating is more probable than a (0,1). For example, when ϵ = 0.05, the likelihood of a (0,0) mating for this batch is 0.019 and a (0,1) is 0.013.

Additionally, we have also observed slightly biased segregation ratios (Section 4.4), and these should also be included into the inference of the most likely parental pair. A biased segregation ratio can indicate either systematic mis-scoring of a locus, or selection on alleles carried at linked loci. Both segregation bias and random genotyping errors (as measured by ϵ) can affect the inference of parents from a batch of eggs, and their effects are not independent. For example, a biased segregation ratio may increase the frequency of singleton genotypes involving the unfavoured allele. As in Section 4.4, I set the variable s (r =1-s) as the segregation

ratio, and for all segregations, including (1,1), s can be viewed as the probability that the *variegata* allele will be sampled in the offspring. For example, in a (1,1) segregation, the expected frequency of the three genotypes given the error rate ϵ is:

$$(Pr[0],\ Pr[1],\ Pr[2]) \propto \begin{pmatrix} 1-\epsilon & \frac{\epsilon}{2} & \frac{\epsilon}{2} \\ \frac{\epsilon}{2} & 1-\epsilon & \frac{\epsilon}{2} \\ \frac{\epsilon}{2} & \frac{\epsilon}{2} & 1-\epsilon \end{pmatrix} . (r^2,\ 2\,sr,\ s^2)$$

i.e.

$$\begin{aligned} Pr[0] &\propto r^2\,(1-\epsilon) + \frac{2\,sr\,\epsilon}{2} + \frac{s^2\,\epsilon}{2} \\ Pr[1] &\propto 2\,sr\,(1-\epsilon) + \frac{r^2\,\epsilon}{2} + \frac{s^2\,\epsilon}{2} \\ Pr[2] &\propto s^2\,(1-\epsilon) + \frac{r^2\,\epsilon}{2} + \frac{2\,sr\,\epsilon}{2} \end{aligned} \qquad [5.2]$$

This means that the frequency of e.g. '0' genotypes is made up of the eggs that are really '0' minus those that have been mis-scored, $r^2(1-\epsilon)$, and those that are really either '1' or '2' but have been mis-scored as '0' (proportions $\frac{2\,sr\,\epsilon}{2}$ and $\frac{s^2\,\epsilon}{2}$ respectively). The probability that each mating type gave rise to $\{U,V,W\}$ genotypes in the eggs is then:

$$Pr[(0,\ 0) \mid \{U,\ V,\ W\},\ \epsilon,\ s] \propto (1-\epsilon)^U \left(\frac{\epsilon}{2}\right)^V \left(\frac{\epsilon}{2}\right)^W$$

$$Pr[(0,\ 1) \mid \{U,\ V,\ W\},\ \epsilon,\ s] \propto \left(r\,(1-\epsilon) + \frac{s\epsilon}{2}\right)^U \left(s\,(1-\epsilon) + \frac{r\epsilon}{2}\right)^V \left(\frac{r\epsilon}{2} + \frac{s\epsilon}{2}\right)^W$$

$$Pr[(0,\ 2) \mid \{U,\ V,\ W\},\ \epsilon,\ s] \propto \left(\frac{\epsilon}{2}\right)^U (1-\epsilon)^V \left(\frac{\epsilon}{2}\right)^W$$

$$Pr[(1,\ 1) \mid \{U,\ V,\ W\},\ \epsilon,\ s] \propto \left(r^2\,(1-\epsilon) + \frac{2\,sr\epsilon}{2} + \frac{s^2\,\epsilon}{2}\right)^U \left(2\,sr\,(1-\epsilon) + \frac{r^2\,\epsilon}{2} + \frac{s^2\,\epsilon}{2}\right)^V \left(s^2\,(1-\epsilon) + \frac{r^2\,\epsilon}{2} + \frac{2\,sr\epsilon}{2}\right)^W$$

$$\Pr[(1, 2) \mid \{U, V, W\}, \epsilon, s] \propto \left(\frac{\epsilon}{2}\right)^{U} \left(\frac{1-\epsilon}{2} + \frac{\epsilon}{4}\right)^{V} \left(\frac{1-\epsilon}{2} + \frac{\epsilon}{4}\right)^{W}$$

$$\Pr[(2, 2) \mid \{U, V, W\}, \epsilon, s] \propto \left(\frac{r\epsilon}{2} + \frac{s\epsilon}{2}\right)^{U} \left(r(1-\epsilon) + \frac{s\epsilon}{2}\right)^{V} \left(s(1-\epsilon) + \frac{r\epsilon}{2}\right)^{W} \qquad [5.3]$$

A likelihood model for errors and segregation bias

We can rearrange the expressions in [5.3] to give the likelihood of particular values of ϵ and s and a parental segregation S for a batch containing $\{U, V, W\}$ of each genotype because $L[S, \epsilon, s \mid \{U, V, W\}] \propto \Pr[S \mid \{U, V, W\}, \epsilon, s]$. These likelihoods can be multiplied across loci and across egg batches, and different hypotheses on the values of ϵ and s can be compared using a χ^2 distribution with one degree of freedom for every additional parameter that is allowed to find its own optimum value (Edwards 1972). There are six possible mating types at a biallelic locus when the order of the genotypes is ignored, and there are $6^4 = 1296$ possible joint parental genotypes for four such loci. Since this is not a forbiddingly large number, and we have no information about which of the jpg's is most probable, the different jpg's are treated as alternatives, and the likelihood summed across them for each locus within a batch. The log likelihood (LogL) of ϵ and s across the entire batch dataset X is given by:

$$\mathrm{LogL}[\epsilon, s \mid X] \propto \mathrm{Log}\left[\prod_{i=1}^{57} \prod_{j=1}^{4} \sum_{k=1}^{6} \Pr[S_k, \epsilon, s \mid \{U_{ij}, V_{ij}, W_{ij}\}]\right] \qquad [5.4]$$

where $\{U_{ij}, V_{ij}, W_{ij}\}$ is the number of 0,1,2 genotypes in batch i at locus j, and the term S_k refers to number k of the six segregation types [i.e. (0,0), (0,1),... (2,2)]. The maximum likelihood values of ϵ and s were found using *Mathematica* (Wol-

fram 1999), entering values by hand and working towards the highest log likelihood score. In Section 4.4 I found significant heterogeneity in s between sites, segregation types and loci, and therefore it is sensible to include this possibility into this joint model for ϵ and s. In the same way, ϵ might also vary between loci depending on how easy each is to score, or between sites because individual samples may have been mislabelled. The potential for ϵ to vary over segregations is not included in this model.

When s is fixed at 0.5, the maximum likelihood estimate for ϵ is 0.0061, which is marginally different from 0 ($\Delta L_1 = 2.76$, $P = 0.018$). Allowing s to vary has a much stronger effect ($\Delta L_1 = 7.42$, $P = 0.0001$). Their joint maximum likelihood estimates are $\epsilon = 0.0077$, and $s = 0.563$ ($\Delta L_2 = 10.66$, $p < 10^{-4}$). The improvement in the LogL when ϵ is allowed to take different values over loci is small ($\Delta L_3 = 2.26$, $P = 0.21$). I therefore hold ϵ constant at 0.0077 in the remaining analyses. This level of error effectively discards ~1 in 13 genotypes per batch.

When s is allowed to take independent values across loci, there is a slight improvement in the likelihood ($\Delta L_3 = 5.68$, $P = 0.009$). However, there is no strong improvement when s can vary between segregation types ($\Delta L_2 = 0.89$, $P = 0.41$) or between sites ($\Delta L_4 = 5.72$, $P = 0.022$). Because the threshold significance level after so many comparisons becomes increasingly unclear, the choice of level at which to stop adding parameters must be determined by biological reality. Since both selection and systematic scoring errors can easily vary in strength between loci, it seems sensible to allow for this in the estimates of s. This is supported by the data, as locus *Bv24.12* is differs considerably from the others: $s = 0.46$ compared to 0.60, 0.55 and 0.60 for *Bb7.4*, *Bv12.19* and *Bv24.11* respectively. I will therefore fit

separate values of s across loci in the likelihood model presented in Section 5.4; the values for ϵ and s and their 2-unit support limits are shown in Table 5.1.

Table 5.1: The best-fit parameter values and their 2-unit support limits (SLs) for the genotyping error rate ϵ and the segregation ratio s for the entire dataset.

	ϵ	s			
		Bv7 .4	Bv12 .19	Bv24 .11	Bv24 .12
value	0.0077	0.60	0.55	0.60	0.46
SLs	(0.001, 0.016)	(0.54, 0.66)	(0.48, 0.62)	(0.54, 0.65)	(0.40, 0.53)

Making parental pairs from the joint parental genotype

The next step is to identify the most likely mating type for each locus (i.e. the joint parental genotype), and from this work out the possible parental genotypes. The joint parental genotype only indicates which alleles were present in either parent, and when more than one locus is involved the alleles can usually be distributed in a number of ways between the male and the female. For example:

the best joint parental genotype for this batch $\begin{pmatrix} \{0,\ 1,\ 1,\ 1\} \\ \{0,\ 2,\ 1,\ 2\} \\ \{1,\ 2,\ 1,\ 2\} \\ \{1,\ 2,\ 1,\ 1\} \\ \{0,\ 1,\ 1,\ 0\} \\ \{1,\ 1,\ 1,\ 0\} \\ \{0,\ 2,\ 1,\ 1\} \end{pmatrix}$

is {(0, 1), (1, 2), (0, 2), (1, 1)},

and the possible parental genotype combinations are :

{0, 1, 0, 1} x {1, 2, 2, 1}
{1, 1, 0, 1} x {0, 2, 2, 1}
{0, 2, 0, 1} x {1, 1, 2, 1}
{0, 1, 2, 1} x {1, 2, 0, 1}

The next step involves determining the probability that each possible parental genotype from the joint parental genotype was present to be involved in a mating.

For this we need the frequencies of each multilocus genotype in each adult pool. The techniques for calculating these are presented in the next section.

Section 5.3 The adult genotype frequencies

Estimating genotype frequencies

If the adult population were close to Hardy-Weinberg and linkage equilibrium, finding the frequency of a multilocus genotype would simply involve multiplying the frequencies of the relevant alleles. Once there are significant associations between alleles both within and between loci, this is no longer accurate, and a more complex method is required. One could estimate the frequency of each genotype from the number of times it appears in the adult sample, but since there are 3^n (= 81) distinguishable multilocus genotypes for $n = 4$ loci, the frequency of many genotypes would be zero even with very large samples. Such an estimate would imply that only the observed genotypes were present, ruling out unseen adult genotypes as potential parents of any of the egg batches. This would be unduly restrictive, as then parentage would be forced onto an improbable sampled adult rather than a more probable unsampled genotype.

Instead, I use the moment estimation approach outlined in Barton (2000). This makes several simplifying assumptions: in populations generated by admixture, associations will tend to be between alleles from the same source population, and they will tend to be of similar strength between unlinked marker loci (i.e. the loci can be treated as equivalent). The population can then be described in terms of a set of allele frequencies p, a set describing the divergences in p between the source populations δp, and a matrix of cumulants $\kappa_{J,K}$, which quantifies the correlations between sets of genes. The latter are divided into (n-1) K th order within-genome associations $\kappa_{0,K}$ and $n(n+1)/2$ between-genome associations involving J genes from one parent and K from the other, $\kappa_{J,K}$. These are shown in Table 5.2. Note

that $\kappa_{0,0} = 1$ because it involves only terms to the power of zero, and $\kappa_{0,1} = \kappa_{1,0} = 0$, because these represent the average deviation of the allele frequencies away from their mean. In total, the 3^n genotype frequencies are determined by $(n^2 + 5n - 4)/2$ parameters (Barton 2000).

Table 5.2: Complete list of the cumulants $\kappa_{J,K}$ needed to describe the four locus genotypes in a population created by admixture.

		K				
		0	1	2	3	4
J	0	1	0	$\kappa_{0,2}$	$\kappa_{0,3}$	$\kappa_{0,4}$
	1	0	$\kappa_{1,1}$	$\kappa_{1,2}$	$\kappa_{1,3}$	$\kappa_{1,4}$
	2	$\kappa_{2,0}$	$\kappa_{2,1}$	$\kappa_{2,2}$	$\kappa_{2,3}$	$\kappa_{2,4}$
	3	$\kappa_{3,0}$	$\kappa_{3,1}$	$\kappa_{3,2}$	$\kappa_{3,3}$	$\kappa_{3,4}$
	4	$\kappa_{4,0}$	$\kappa_{4,1}$	$\kappa_{4,2}$	$\kappa_{4,3}$	$\kappa_{4,4}$

For example, the cumulant $\kappa_{1,1}$ describes the correlation between two alleles at a single locus (which is related to F_{IS} by pqF), and $\kappa_{0,2}$ is the correlation between two loci inherited from the same parent (i.e. the gametic linkage disequilibrium), both of these are second order cumulants. Higher order cumulants represent correlations between more than two genes and may involve loci both within and between the haploid (i.e. maternal and paternal) genomes of a diploid animal. In this context cumulants are more useful than the more familiar moments (e.g. Barton and Turelli 1991), as they quantify the association over and above that expected given the lower order cumulants. Cumulants of the same kind are expected to be of similar magnitude, and higher order cumulants should be small in a population generated by admixture.

Estimating these parameters individually may lead to incompatible combinations, which in turn may generate negative genotype frequencies. The solution is to find the optimum values for all the parameters simultaneously; i.e. to find the set of parameters with the highest joint maximum likelihood. There are two broad strategies to optimising this many variables (Kirkpatrick et al. 1983): 1) divide the problem into smaller sections, optimise the parameters within each and combine them to gain the overall solution. This method relies on the absence of interactions between the subgroups. 2) the system can be kept whole, and the parameters gradually optimized through a process of iteration. One key problem facing both methods is finding the global optimum whilst avoiding local optima; the latter become more common as the number of parameters involved increases (Kirkpatrick et al. 1983). A good analogy is finding the highest peak on an adaptive landscape when many inferior peaks are present. One of the best methods for overcoming this difficulty is the version of the Metropolis algorithm laid out by Kirkpatrick et al. (1983), as it incorporates both of the above strategies into finding the optimal solution.

The Metropolis algorithm was originally developed to simulate the behaviour of atoms at equilibrium for a given temperature (Metropolis et al. 1953), but was extended by Kirkpatrick et al. (1983) to deal with a much wider range of problems. In the original algorithm, a small random change was made to a configuration of atoms, and the resulting change in energy calculated. If the change results in a decrease in the energy of the system, it is accepted. If the change increases the energy, it is accepted with probability $\mathrm{Exp}(-\Delta E / k_B T)$ where ΔE is the change in energy, T is the temperature and k_B is Boltzmann's constant. If the atoms are

replaced by the values of a set of parameters, and the energy by the 'cost' of any particular arrangement, this procedure can be used to optimise any multivariate problem by a process of 'simulated annealing' (Kirkpatrick et al. 1983). This involves slowly decreasing the temperature T until the global optimum is found, similar to the process of crystal formation in solids. Decreasing T infinitely slowly will always lead to a global optimum, but in reality finding one from a number of near optimal 'peaks' usually suffices.

In our case, the parameters p_i, δp_i and the $\kappa_{j,k}$ are being optimised, and the 'cost' of a set of values is given by the log likelihood, LogL. Since the frequency of genotype X, $g[X]$ is given by a linear combination of the parameters, summing $\log(g[X])$ over all genotypes gives the overall log likelihood for that set of parameters. Their value is optimised as follows: a random change is made to parameter k; this change is drawn from a symmetrical uniform distribution, which has maximum values at $\pm\Delta_k$. If the new parameter set is valid (i.e. all $g[X] \geq 0$), and increases the LogL, it is accepted. If it is valid and decreases the likelihood by a factor $\theta < 1$, it is accepted with probability $\theta^{1/T}$. The size of the random perturbations is also optimised by increasing Δ_k slightly if a change in k is accepted, and reducing Δ_k if the change is rejected. Repeated iterations of this cycle lead to a random walk with a probability distribution $L^{1/T}$ (Barton 2000). The entire process is controlled by varying the 'temperature' T. When T is high, changes decreasing the likelihood are accepted more often, and more of the multi-dimensional landscape can be visited. As the temperature falls to 0, changes decreasing LogL are rejected more often, and the algorithm enters a 'hill climbing' phase up whatever peak it happens to be on. This is similar to the 'divide and conquer' strategy mentioned above: the algorithm first

moves widely over the landscape to find the best approximate solution and then homes in on the exact values as T falls towards zero.

Using the Metropolis algorithm

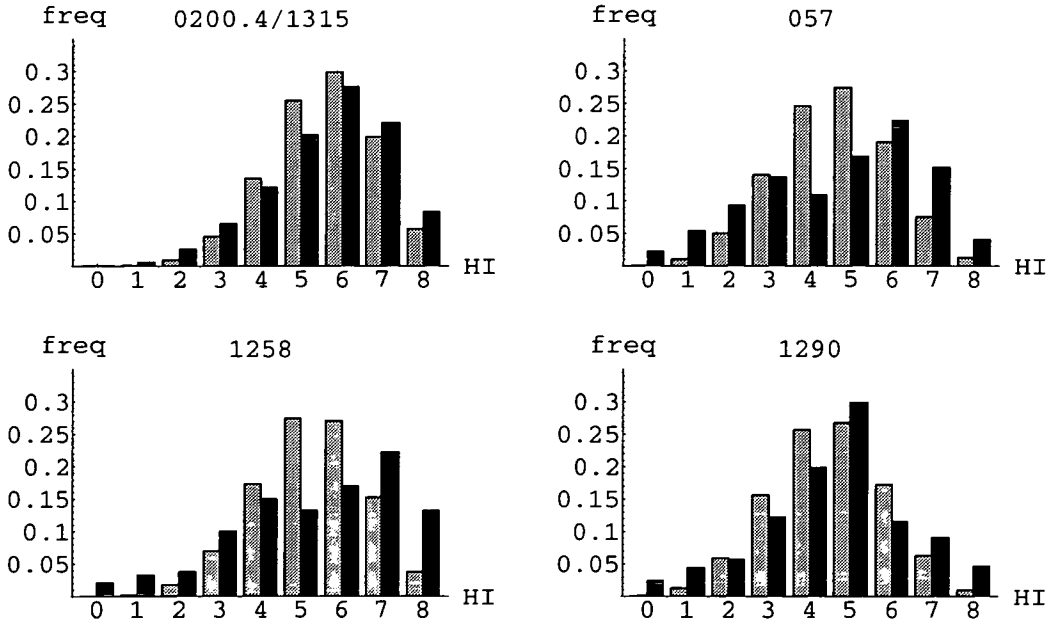
The Metropolis algorithm used here was written by N.H. Barton in *Mathematica* (Wolfram 1999), and is available from http://helios.bto.ed.ac.uk/evolgen/. Full details of the Metropolis algorithm as applied to estimating genotype frequencies are given there and in Barton (2000). The initial parameter values at the start of a run were as follows. The allele frequencies p_i were set at the mean values for each adult pool, and the δp_i set to 1. All the cumulants $\kappa_{j,k}$ set to 0, with the exception of $\kappa_{0,0}$, which is necessarily 1. The boundaries of the initial uniform distribution from which the first random change is drawn (i.e. Δ_k above) were set as follows. For the allele frequencies, the bounds were $p_i \pm 0.05\sqrt{p_i(1-p_i)}$. Since the adult pools are generated by admixture, and marker loci are diagnostic and unlinked (B. Nürnberger, pers. comm.; Nürnberger et al. 2002b), we set the divergences at each locus to be 1 (i.e. the maximum permitted change was zero), which considerably saves on the number of parameters needing to be simultaneously estimated. Finding the most useful initial step for each cumulant is more important, as avoiding changes that lead to negative genotype frequencies saves a lot of computer time. These are tricky to calculate, and since no single method is universally appropriate, I will not present the workings here. The calculation routine can be found in the 'Multilocus' *Mathematica* package by N. H. Barton and is available on the web at http://helios.bto.ed.ac.uk/evolgen/.

There are two potential ways to estimate the genotype frequencies across a group of discrete sites. First, one can lump all individuals into a single group and estimate

the moments from these. However, since the sites differ in allele frequency, this lumping will generate a Wahlund effect, artificially raising the associations between alleles and loci. Second, one could constrain both the allele frequencies and the matrix of moments to be the same across all the sites, and find a single set of values that best decribe the associations in all sites simultaneously. This avoids the Wahlund effect, but is unrealistic in assuming equal allele frequencies in different habitats (see Figure 4.2). I use the latter approach.

The algorithm was iterated 600 times, 200 times each at T = 1, 0.15 and 0 respectively. This procedure was repeated five times, and the most likely set of best fit parameters from the five were fed back in as starting values for another 600 cycles, to see if further improvement could be made. The best fit values for the p_i, δp_i and $\kappa_{j,k}$ are given in Table 5.3. The genotype frequencies for all of the 3^nobservable genotypes were then calculated as in Barton (2000) with the functions in the Multilocus *Mathematica* package mentioned above. The frequencies of individuals with 0,1...8 *B. variegata* alleles (i.e. the Hybrid Index) under the calculated genotype frequencies and under linkage equilibrium for each adult pool is shown in Figure 5.2.

Figure 5.2: Frequency of each Hybrid Index class under linkage equilibrium (light bars) and the observed multilocus linkage disequilibrium (dark bars).



There is only weak evidence for a bimodal distribution of genotypes in the 0257 and 1258 adult pools; these also have the highest levels of linkage disequilibrium (Table 4.4). The estimates made separately for heterozygote deficit and pairwise

linkage disequilibrium in Chapter 4 are similar to more sophisticated estimates for $\kappa_{1,1}$ (= pqF) and $\kappa_{0,2}$ (= D), although there are some discrepancies (Table 5.4).

Table 5.4: Comparison of the direct estimates and the moment-based estimates for heterozygote deficit (pqF and $\kappa_{1,1}$) and linkage disequilibrium (D and $\kappa_{0,2}$).

site	pqF	$\kappa_{1,1}$	D	$\kappa_{0,2}$
0200.4/1315	0.007	0.009	0.001	0.011
0257	0.038	0.048	0.015	0.028
1258	0.005	0.078	0.016	0.013
1290	0.019	0.035	0.01	0.012

As noted in Section 4.3, females can only exercise a mate preference when there is a range of males to choose from, and this choice can only be detected with neutral markers when there is linkage disequilibrium between male trait and female preference loci and the markers. It is therefore possible that there is an insufficient range of animals in each adult pool to allow assortative mating, and I return to this point in Section 5.6. The frequency of each multilocus genotype within the adult pool should, under random mating, be equal to its probability of having been a parent. However, if there is assortative mating or breeding site choice, the probability that a particular genotype contributed to the next generation will be some function of its genotype and its frequency. In the next section I fit parameters for these functions using a maximum likelihood model.

Section 5.4 The likelihood model of mating behaviour

Section 5.2 described how to calculate the likelihood that a particular joint parental genotype produced a batch of eggs, given values for the genotyping error rate ϵ and the segregation ratio *s*. If the batch contains several segregations, or if ϵ is greater than zero, there will be more than one candidate jpg. If these are separated out into the possible pairs of genotypes, we then have a list of likelihoods that each particular pair created a batch of eggs (Section 5.2). The probability that a genotype was present in the adult pool is given by its frequency; these were calculated in Section 5.3.

Parameterising the model

To model direct mate choice I use Lande's (1981) model of sexual selection, in which the probability of two individuals mating is an exponential function of the difference between the females preference and the males trait, I scale this effect with the parameter α. The probability that an individual will mate in a site at all is expressed in a similar way, by the difference between the individuals trait or preference and the mean in the adult pool. This is scaled by the parameter β. An exponential function is preferable to linear one in this situation because the latter can generate negative probabilities when the preference is strong. Since there are no data on the phenotype of the animals, I assume that an animals hybrid index is an accurate summary of its state at the loci determining breeding site or female preference and the male trait. This is most likely to be true in hybrid populations with strong linkage disequilibrium between neutral markers, as this implies that the marker alleles are still in association with their respective mating behaviour alleles. I denote the

hybrid index of individual i as Z_i. The expression for the likelihood of the strength of direct mate choice α and the breeding site preference β is:

$$L[\alpha, \beta \mid B, \epsilon, s] \propto$$
$$C^{-n} \prod_{i=1}^{n} \sum_{j} \sum_{k} \Pr[G_j] \Pr[G_k] \Pr[B_i \mid (G_j, G_k), \epsilon, s] \exp[$$
$$\alpha (Z_i - \bar{Z}) (Z_j - \bar{Z})] \exp[\beta (Z_i - \bar{Z})] \exp[\beta (Z_j - \bar{Z})] \quad [5.6]$$

where $\prod_n$ is the product over the n families, G_j and G_k denote each of the $3^4 = 81$ distinguishable multilocus genotypes, $\Pr[G_i]$ denotes the frequency of genotype i in the adult population, and $\Pr[B_i \mid (G_j, G_k), \epsilon, s]$ gives the probability that G_j and G_k produced batch B_i given ϵ and s. C is a numerically calculated normalising constant. The $\bar{Z}$ in both the α and β functions is the mean HI in the adult pool not the focal adults, as it was apparent from Section 4.5 that the latter bore little relation to the inferred parents. As with the estimation of ϵ and s above, α and β were entered by hand into a *Mathematica* function until the best fitting values and their support limits were found.

Estimating the parameters

The improvement in the likelihood by fitting the assortative mating parameter α instead of setting it zero is small: $\alpha = 0.05$, $\beta = 0$; $\Delta L_1 = 1.22$, $P = 0.11$). Furthermore, setting $\beta = 0$ and allowing α to vary over sites produced no further improvement ($\alpha = \{0.021, -0.005, 0.066, 0.042, -0.001\}$ in each site respectively, $\beta = 0$; $\Delta L_4 = 0.29$, $P = 0.98$). I therefore set $\alpha = 0$ in the following. By contrast, fitting β had a large effect ($\alpha = 0$, $\beta = 0.31$; $\Delta L_1 = 14.52$, $p < 10^{-7}$), and allowing β to take different values in each site makes a further improvement in the fit: $\Delta L_4 = 7.15$, $P = 0.006$. Further subdivisions of the data are possible but would be hard to interpret,

such as assessing assortative mating on a per locus basis. I therefore stick to $\alpha = 0$ and β variable across sites. The best fit values and their support limits are given in Table 5.5.

Table 5.5: The best-fit parameter values for the mate choice parameter α and the breeding habitat choice β and their 2-unit support limits (SLs).

	α			β		
		0200.4	1315	0257	1258	1290
value	0	0.57	0.43	0.44	−0.25	0.25
SLs	–	(0.12, 1.09)	(−0.09, 1.00)	(0.25, 0.65)	(−0.48, 0.07)	(0.04, 0.46)

Assessing the fit with simulations

It is possible to assess how well the fitted mating behaviour model matches the observed egg data with simulations. These draw pairs of parental genotypes from the adult genotype frequencies for each site, and mate them together to produce the offspring. The proportion of 0,1,2 genotypes in the simulated parents and offspring can then be compared to the observed data, along with the proportion of each segregation type in the inferred parents. The first simulation drew individuals at random from the adult pool (i.e. $\beta = 0$), paired them up at random ($\alpha = 0$), and produced offspring under Mendelian segregation (i.e. $\epsilon = 0$ and $s = 0.5$ for all loci). The second simulation used the estimated values for β (Table 5.5) to draw adults from the adult pool, paired them at random (since α is set to 0), and then also produced offspring using normal Mendelian segregation. These two simulations provide a yardstick for evaluating the effects of the fitted parameters on the results. The third simulation uses β to sample adults, and the fitted values for ϵ and s for each locus to produce the offspring (see Table 5.1). All three simulations were written in *Mathematica*.

For each site, 10,000 pairs of adults were simulated, and 10 offspring created for each. The frequencies were calculated by pooling across all individuals and loci. I do not use varying sizes of batch in the simulations, as these introduce an interaction between the identity of the parents and the batch size, such that an order of magnitude more replication is needed to account for this effect. The results of the simulations are in Table 5.6.

In general, the fitted model and the observed data match each other well. By contrast, the null model is a poor fit, especially in terms of the allele frequencies in the inferred parents and eggs. Fitting the breeding site preference parameter β alone makes a large improvement, although it is apparent that the simulated eggs fit the observed more closely when ϵ and s are included. This shows that the fitted model is explaining these observed data quite well. It is hard to discern from the simulated data what other parameters could be fitted to improve the model, and it is doubtful that an extra parameter is sustainable for a dataset of this size in any case.

These simulations also provide a way of assessing the significance of the strong heterozygote deficit seen in the eggs of sites 0200.4 and 1290 found in Section 4.5. Under random mating, the eggs should on average conform to Hardy-Weinberg expectations, but sampling batches of eggs can generate a wide variance in the proportion of heterozygotes depending on the genotypes of the parents. For very large number of batches, the parents should reflect the allele frequencies in the adults, but for smaller sample sizes considerable sampling error can arise. To see if this is the case for the observed batches of eggs, I used the simulations described above to generate the sampled number of batches (7 and 17 respectively), each with 10 eggs using the fitted ϵ, s and β parameters. I then found the most likely value for

the heterozygote deficit F_{IS} when pooling genotypes across all loci and batches. This was repeated 1000 times for each site. The observed value of F_{IS}= 0.19 in 0200.4 was exceeded by 5.8% of simulated results, and in 1290, the heterozygote deficit F_{IS}= 0.184 was exceeded by 2.2% of the simulated values. The simulated F_{IS} values were distributed roughly equally on either side of the observed value in the other three sites.

Both 0200.4 and 1290 were found to have a significantly biased segregation ratio (Table 4.2), and hence the simulations were repeated using the per-locus best fit values for s in these two sites separately in place of the overall estimates (see Section 4.4). This had little effect on the simulations for 0200.4, as the observed F_{IS} was exceeded by only 3.3% of simulated values, but in 1290 the observed F_{IS} was now exceeded by 36% of the values. It thus seems likely that the heterozygote deficit in 1290 is an artefact of sampling error and biased segregation, whereas the strong F_{IS} in 0200.4 is explained by neither of these effects. It is possible that there was in fact assortative mating in 0200.4, but the estimation of α above somehow affected by the genotype frequencies in the adult pool. To test this, I turn to the moment-based approach, which can estimate features of the parental generation (including the strength of assortative mating) without data from the adults.

Section 5.5 The moment-based estimates for the parents

The basic premise of this approach is that the eggs contain information about the parents: for example, the variance in hybrid index across individuals is an indicator of the mean heterozygosity of their parents, as only matings involving heterozygotes can generate segregations in the offspring (see Section 5.1). The levels of variance and heterozygosity in the offspring associated with each segregation type is shown in Figure 5.1. Nürnberger et al. (2002a) directly inferred the cumulants in the parents from the cumulants in the offspring, I take a slightly simpler approach here and calculate the cumulants directly from the inferred parents. The cumulants in this case are very similar to the $\kappa_{J,K}$ estimated by the original moment-based approach in Section 5.3 above. In that section they were calculated for a sample of individual adults, whereas here I am dealing with pairs of parents.

Calculating cumulants

The cumulants for a sample of individuals are calculated as follows: Each diploid individual carries two genes at each locus, and the allelic states of genes from the maternal and paternal genomes at locus i are labelled X_i and X_i^* respectively. We can only observe the combined value $X_i + X_i^* = 0,1,2$, as the underlying genotypes $\{0,1\}$ and $\{1,0\}$ cannot be distinguished in heterozygotes. The mean allelic states are just the allele frequencies in males and females, which are both assumed be p_i. The deviation from the population mean in an individual is given by $\zeta_i = X_i - p_i$ for the maternal and $\zeta_i^* = X_i^* - p_i$ for the paternal contribution. Two useful quantities can be defined for each locus. First, the additive effect is given by $\zeta_i + \zeta_i^* = X_i + X_i^* - 2\,p_i$. Second, the deviation from Hardy-Weinberg proportions is described by $\zeta_i\,\zeta_i^*$, which takes values $\{p_i^2,\ -p_i\,q_i,\ q_i^2\}$ for $X_i + X_i^* = 0,1,2$. These

two quantities are then combined in different ways to estimate the various cumulants for that sample.

In this section the focus is on the inferred parents. Since we have no information on how the parental genotypes at each locus were distributed between the mother and father, the cumulants must be estimated for pairs of individuals together. The cumulants in this case are denoted $\kappa_{a,b,c,d}$. The subscripts a and b refer to the maternal and paternal gamete of the mother, respectively, and c and d the maternal and paternal gamete of the father. The deviations ζ for joint parental genotype i are calculated as above; those in the mother are denoted ζ_i for the maternal and ζ_i^* for the paternal gamete, and those in the father are ζ_i' for the maternal and $\zeta_i^{*\prime}$ for the paternal gamete. Estimating a single cumulant for both the mother and the father simultaneously assumes that the allele frequencies in the males and females were equal, and that they were also equal in the previous generation (i.e. the grandparents of the eggs).

Here, I will focus on three cumulants of particular interest. The first is the heterozygote deficit in the parents, which involves the correlation in state between an allele inherited from the mother and an allele from the father. When heterozygote deficit is strong, the same allele will be carried by both gametes, and hence the correlation between them will be higher. Because we are estimating the $\kappa_{1,1}$ cumulant across both the mother and the father, the estimated value could either be the mean correlation between alleles at the same locus in the mother (denoted $\kappa_{1,1,0,0}$) or in the father ($\kappa_{0,0,1,1}$). I use the latter notation for consistency with Barton (2000). The value of $\kappa_{0,0,1,1}$ is given by the mean of $\frac{1}{4}(\zeta_i \zeta_i^* + \zeta_i' \zeta_i^{*\prime})$ across all of the joint parental genotypes within that site, and it is equivalent to *Fpq* in the parents.

The second cumulant of interest is the linkage disequilibrium D in the parents, which is the correlation in allelic state between two alleles at different loci within the same gamete. For the samples of adults in Section 5.3 this cumulant is denoted $\kappa_{0,2}$; when it is estimated in the parents from joint parental genotypes it is $\kappa_{0,0,0,2}$. As with $\kappa_{0,0,1,1}$, this is assumed to be the same in the gametes in the mother ($\kappa_{2,0,0,0}$ and $\kappa_{0,2,0,0}$) and in the father ($\kappa_{0,0,2,0}$ and $\kappa_{0,0,0,2}$). For joint parental genotypes with n loci it is calculated as:

$$\kappa_{0,0,0,2} = \frac{1}{4}\left(\sum_{i\neq j}^{n} \frac{(\zeta_i + \zeta_i^* + \zeta_i' + \zeta_i^{*\prime})\,(\zeta_j + \zeta_j^* + \zeta_j' + \zeta_j^{*\prime})}{n!/(n-2)!\,2!} - 2\sum_{i}^{n} \frac{(\zeta_i + \zeta_i^*)\,(\zeta_i + \zeta_i^*)}{n} - 2\sum_{i}^{n} \frac{(\zeta_i\,\zeta_i^* + \zeta_i'\,\zeta_i^{*\prime})}{n}\right) \qquad [5.7]$$

Equation 5.7 approximates to subtracting the heterozygote deficit ($\kappa_{0,0,1,1}$) and the covariance between the parental Hybrid Indices ($\kappa_{0,1,0,1}$) from the overall additive effect ($\kappa_{0,0,0,1}$).

The last cumulant is potentially the most informative, as it measures the correlation in allelic state between one gamete in the mother and one in the father, and hence is a measure of assortative mating. It is denoted $\kappa_{0,1,0,1}$, and, as above, is assumed to be equivalent in strength over all four gametic combinations. It relates to the assortative mating parameter α estimated in Section 5.4 through the variance in hybrid index v in the adult pool: $\kappa_{0,1,0,1} \approx v\,\alpha$ (N. Barton, pers. comm.). The value of $\kappa_{0,0,1,1}$ is given by the mean of $\frac{1}{4}(\zeta_i + \zeta_i^*)(\zeta_i' + \zeta_i^{*\prime})$ across all of the joint parental genotypes within that site.

Observed and simulated cumulants

To assess the significance of the estimates for these three cumulants, I use the same simulation approach as above. Pairs of parents are sampled from the adult pool genotype frequencies either at random or according to the fitted values for β described above. In each site, I simulate the same number of pairs as in the original sample, so that sampling error associated with that sample size is included. The parental pairs are then mated to produce 10 offspring each, either under normal Mendelian segregation, or using the fitted values for ϵ and s, and the cumulants calculated from the joint parental genotypes for each batch. As in Section 5.5, I do not use varying sizes of batch in the simulations, as this introduces an interaction between the identity of the parents and the batch size. The simulation was written in a *Mathematica* notebook and repeated 1000 times; the cumulants for both the observed and the simulated data are shown in Table 5.7.

Table 5.7: The three cumulants estimated for parents in each site, and the three simulated datasets (see text). The first column is from random mating and Mendelian segregation ratio (Null model), the second uses the estimated strength of breeding site preference β, and the third uses both β and the estimated error rate ϵ and the segregation bias s. For these columns the data is mean (s.d.).

cumulant	$\kappa_{1,1}$		$\kappa_{0,0,1,1}$		
	Adults	Parents	Null model	β only	β, ϵ, s
Site					
0200.4	0.011	0.012	0.009 (0.028)	0.001 (0.018)	−0.023 (0.013)
1315	0.011	−0.006	0.006 (0.032)	0.001 (0.026)	−0.028 (0.019)
0257	0.028	0.004	0.031 (0.017)	0.011 (0.014)	−0.026 (0.01)
1258	0.013	−0.065	0.005 (0.044)	0.009 (0.051)	−0.036 (0.041)
1290	0.012	0.03	0.016 (0.021)	0.011 (0.019)	−0.032 (0.015)

cumulant	$\kappa_{0,2}$		$\kappa_{0,0,0,2}$		
	Adults	Parents	Null model	β only	β, ϵ, s
Site					
0200.4	0.009	−0.054	0.021 (0.053)	0.012 (0.038)	0.024 (0.037)
1315	0.009	−0.018	0.031 (0.062)	0.022 (0.051)	0.03 (0.048)
0257	0.048	0.03	0.046 (0.031)	0.016 (0.028)	0.026 (0.024)
1258	0.078	0.02	0.093 (0.057)	0.111 (0.057)	0.092 (0.06)
1290	0.035	0.028	0.038 (0.036)	0.031 (0.035)	0.035 (0.034)

cumulant		$\kappa_{0,1,0,1}$		
	Parents	Null model	β only	β, ϵ, s
Site				
0200.4	0.021	−0.006 (0.02)	−0.004 (0.013)	0.001 (0.012)
1315	−0.006	−0.009 (0.023)	−0.007 (0.019)	0.001 (0.015)
0257	−0.004	−0.002 (0.015)	−0.001 (0.01)	0.008 (0.008)
1258	0.028	−0.013 (0.032)	−0.016 (0.04)	0.016 (0.026)
1290	0.009	−0.003 (0.015)	−0.002 (0.014)	0.013 (0.011)

There are several interesting patterns in Table 5.7. First, the estimation of $\kappa_{0,0,1,1}$ is affected quite strongly by non-Mendelian segregation, as only in site 1258 is the observed value within one standard deviation of the simulated mean when ϵ and s are included. Second, none of the cumulants are particularly large, especially when compared to the values in Nürnberger et al. (2002a). There, $\kappa_{0,0,1,1}$ was estimated as 0.065 and $\kappa_{0,0,0,2} = 0.191$. Only 1258 comes close, with $\kappa_{0,0,0,2} = 0.111$, but this estimate is only based on four batches. The heterozygote deficit and linkage disequilibrium is estimated to be smaller in the parents compared to the adults. This suggests that there is little or no discrimination against matings with hybrids, as this would lead to greater levels of D in the parents than the adults.

Lastly, the value of $\kappa_{0,1,0,1} = 0.021$ in 0200.4 does not match any of the simulated data, and, since it is positive, it may indicate positive assortative mating in this site. The maximum likelihood estimate of α in this site is also 0.021, but when this value is scaled by the variance in hybrid index in the adult pool, it suggests that $\kappa_{0,1,0,1}$ should be ~0.021 x 2.26 = 0.047. The discrepancy may arise because this relationship assumes that the distribution of hybrid index in both males and females is normal, and the estimate of α includes segregation bias and genotyping errors. There is also a significant deficit of heterozygotes in this site, even when sampling error is taken into account. It is thus possible that assortative mating has actually taken place in this case. Nonetheless, such a conclusion cannot be supported without more data from this site.

Section 5.6 Discussion

This section brings together the results of Chapters 4 and 5. I firstly cover the initial checking of the eggs batches for Mendelian segregation and the presence of mixed batches. I then examine the evidence that the different types of prezygotic isolating mechanism (e.g. Section 1.2) are operating in *Bombina*. Lastly, I discuss how these results fit with the existing knowledge of *Bombina* mating behaviour in particular, and assortative mating in hybrid zones in general.

Collection and checking of egg data

Eggs were collected from sites all over the Apahida study area in 2000 and 2001 by myself, Sonja Köhler and Tim Sands. From these, 567 eggs from five sites were genotyped at four marker loci by S. Köhler. Since the inference of mating behaviour from the offspring normally assumes Mendelian segregation, I checked for three potential problems. Mislabelling samples or mis-scoring alleles can generate single anomalous genotypes in a segregation, which undermine inferences about the parents. Since some singletons can arise by chance, I tested for an overall excess by comparing the observed number with the proportion expected under the multinomial distribution, and found that there was only an excess in segregations involving 1,1 segregations (Section 4.4).

Mendelian expectations can also be affected by a bias in the segregation ratio, and I quantified this by estimating the probability that a particular allele will be transmitted to the offspring (Section 4.4). Two types of bias are possible: first, the proportion of the favoured allele may be increased in all segregations, perhaps by systematic mis-scoring or selection. Second, selection removing heterozygotes prior to

sampling will generate bias in opposing directions in 0,1 and 1,2 segregations, while leaving 1,1 unchanged. The pattern in 1315 was weakly consistent with the selection against heterozygotes. However, there was good evidence for consistent bias as *B. variegata* alleles were in excess at three of the four loci (Table 4.2). Locus *Bv24.12* showed an excess of *B. bombina* alleles, although this was not significantly different from 0.5. It is impossible to distinguish systematic scoring errors from selection against linked loci from these data, although further work should perhaps first concentrate on the validity of the allele scoring before considering alternative explanations. Repeated rescoring of these loci in Nürnberger et al. 2002a) found that the genotyping error rate was around 1%.

I estimated both the rate of random genotyping errors (ϵ) and the strength of segregation bias (s) from the data itself with maximum likelihood. This gave $\epsilon = 0.0077$ (support limits 0.001-0.016), and I therefore included ϵ when inferring the joint parental genotypes in the full likelihood model in Section 5.4. This slightly significant error rate implies that discarding singletons from batches with ~13 or more eggs greatly improves the fit of the model. If singletons were a significant problem in the data, the error rate would be correspondingly higher so that smaller batches would also be included. As it is, there are only 14 batches with >12 eggs, and only six segregations within these contain a singleton. The best fit model kept the same value for s across segregation types but varied it across loci, not unlike the pattern when s is estimated alone (compare Table 4.2 and Table 5.1). I therefore also included variable segregation ratios when inferring joint parental genotypes in Section 5.4.

The third potential problem is the mixing of batches from separate matings, which can cause significant problems when making inferences from egg data. This was found in three of the 18 batches in Nürnberger et al. (2002a), but since their test has little power with only four loci I extended their approach to search for mixing using the more polymorphic original alleles. Neither this technique or a visual inspection of the data produced evidence for mixed batches, and it is likely that any undetected mixtures were generated from similar pairs of parents, which hopefully will have little impact on our inferences. I also considered joining batches of eggs with compatible genotypes, but since the consequences of joining unrelated families are worse than leaving related ones apart, I decided to keep the eggs in their original families.

Prezygotic isolation in Bombina

I now consider the evidence for different types of prezygotic isolation in *Bombina* from Apahida. Ecological isolation was amply discussed in Chapter 2, and there is good evidence for an active preference from both Pešćenica (MacCallum 1994; MacCallum et al. 1998) and Apahida (Chapter 2). This phenomenon was also supported by the distribution of adult genotypes between the sites within the adult pools, especially in 0257 and 0200.4/1315, where there are a wide range of habitat types (Figure 4.2). The relationship between the mean allele frequency of the adult pool and the adults in the focal site also corresponded well to the difference in habitat score (Figure 4.3; Table 4.5). Since the preference is not strong enough to prevent migration between habitats, I concluded in Chapter 3 that fairly strong selection against immigrant adults was also required to counteract the influx of *B. bombina* alleles from ponds.

An active preference for alternative habitats can only generate ecological isolation if those habitats are also used for breeding, and it is not clear that this is the case for *Bombina*. The relationship between the allele frequency in the adult pool and that in the inferred parents was consistent with the direction of habitat preference in three of the five egg sites (1315, 0257 and 1258), but was in the opposite direction in 0200.4 and 1290. The difference in $\overline{p}$ between the adult pool and the inferred parents was also supported by the estimates of β in Table 5.5. There was also no visible relationship between the adults in the focal site and the inferred parents, implying that the parents often vacate a site before they can be sampled. The lack of correspondence between β and the habitat score is unexpected if breeding site preference is based on the same criteria as habitat preference, although the observed pattern could arise by chance if the preference is weak.

It is also possible that the criteria for choosing a resting or feeding habitat are different from those for breeding habitat, and that the latter correlate poorly with our habitat axis. Additionally, a breeding aggregation may form because both males and females move towards large choruses, and individuals may trade-off their breeding site preference against the opportunity to choose from a wider range of animals. It is hoped that a more extensive survey of eggs and local adults currently being analysed by S. Köhler and T. Sands will shed more light on the relationship between resting and breeding habitat. It may also be interesting to compare the nocturnal and diurnal composition of a range of sites to quantify the effect of calling males on the number and types of toad present.

I now consider sexual isolation in *Bombina*. Since I have relied on the offspring to infer the mating behaviour in the parents, there are several isolating mechanisms

that may have operated in between the formation of the breeding aggregation and the sampling of the eggs. Fertilisation is external in toads, and hence there is less opportunity for mechanical incompatibilities between the male and female genitals, although size differences may have some effect. However, there may well be potential for incompatibility between gametes. There have been no direct studies on fertilisation in *Bombina*, but gametic isolation may not be very strong because hybrid eggs can be produced quite easily in the lab (J. Szymura and B. Nürnberger, pers. comm.). Additionally, selection against hybrid eggs before they are sampled can also affect the inference of mating behaviour. Fortunately, this mechanism will usually distort segregation ratios in the offspring, which can be detected when the offspring are collected in batches.

I found above that there is a significant bias in favour of *B. variegata* alleles in segregations, although this could instead be due to scoring errors. There was no consistent evidence for selection against hybrids from the segregation patterns. There have been direct measurements of inviability selection against hybrid offspring, and it appears that only a proportion of hybrid batches have reduced fitness (Kruuk et al. 1999a). For the tadpoles collected in Apahida, only 185 of the 2986 died before they could be sampled, and some of these deaths may be due to interactions between tadpoles in the same cup. Since the remainder of the batch survived in all but 9 cases, it seems unlikely that selection will distort our inference of mating patterns from the egg data via an under-representation of hybrid families.

Is there evidence for sexual isolation in *Bombina* from these data? Including the assortative mating parameter α in the likelihood model did not result in a significant improvement, nor was there any effect of allowing it to vary across sites.

Furthermore, the lack of heterozygote deficit in the eggs in the majority of sites suggests that mate choice is not an important isolating mechanism in *Bombina*. This was also apparent in the estimates of the correlation between parental hybrid indices $\kappa_{0,1,0,1}$, none of which were more than two standard deviations away from the mean of the values simulated under random mating. Interestingly, there was a strong heterozygote deficit in 0200.4 which could not be explained by either sampling error or segregation bias, and although both α and $\kappa_{0,1,0,1}$ are positive in this site, neither are significantly different from zero. It is possible that there was assortative mating here, but it could not be demonstrated with only seven batches.

A perhaps more important issue is the composition of the adult pools, as females can only exercise a preference if there is a wide range of males present. It is possible that the four adult pools or the breeding aggregations formed from them were not sufficiently variable to permit the detection of assortative mating with neutral markers. There is only weak evidence for a bimodal distribution of genotypes in the 0257 and 1258 adult pools (Figure 5.1), and the other two are unimodal, although the overall spread in each is wide. Nonetheless, the adult pools and the focal sites were selected because they contained the widest ranges of animals in the Apahida area, and if there is no assortative mating within these it seems unlikely that it would operate elsewhere.

An earlier analysis of *Bombina* mating patterns in Pešćenica (Nürnberger et al. 2002a) provides encouraging support for my conclusions: there was no assortative mating ($\alpha = 0.007$), but a strong effect of breeding site preference ($\beta = 0.23$). Even though the site was puddle like ($H = 0.81$) there were many *B. bombina*-like individuals, although the parents are mostly *B. variegata*. There was significantly stronger

linkage disequilibrium in parents in this case, suggesting that hybrids were not achieving as many matings as purer animals, contrary to the estimates for $\kappa_{0,0,0,2}$ in Apahida. However, there is overall much stronger linkage disequilibrium in the adults in the Pešćenica site (D = 0.122), which may be partly responsible for this effect. Interestingly, the wider range of animals available in this site still did not result in assortative mating. The close similarity between the results from Pešćenica and those from Apahida suggests that the lack of assortative mating may be a general feature of *Bombina.*

The wider picture

There is generally very little concrete data about mating behaviour in *Bombina*, and hence interpreting these results in light of what is already known is difficult. The two taxa differ in call in Poland, but there is no evidence that females can distinguish between the taxa on the basis of this call. Furthermore, the anecdotal evidence that *B. bombina* males hold calling territories and *B. variegata* males engage in scramble polygyny does not necessarily entail a lower reproductive success for visiting males from the other taxa, as there is potential for other 'sneaky' mating strategies. I therefore concentrate identifying potential avenues for research into *Bombina* mating behaviour. Firstly, we need to demonstrate that male calling has a role in mate choice or mate recognition, perhaps with experiments under semi-natural conditions. Second, data on the strength of gametic isolation in *Bombina* would confirm the validity of inferring mating patterns from offspring data. Lastly, understanding the role of habitat in determining the composition of breeding aggregations is of great importance, as ecological isolation appears to be the strongest prezygotic mechanism in this hybrid zone.

In their 1985 survey, Barton and Hewitt found assortative mating in 10 out of the 20 hybrid zones in which mating behaviour had been measured, although one of these is *Bombina*. The observation of effectively random mating in the remainder raises an interesting question: if there is potential for mate choice based on male characters, why does this not translate into assortative mating? Unfortunately, direct sexual selection on the males has not been demonstrated in *Bombina*, but it is reasonable to assume that the energetically expensive calling displays must have some role in reproductive success. There is clearly a continuum between pairs of taxa that are isolated on the basis of a single sexually selected trait (e.g. between several taxa of *Physaelaemus* frogs: Ryan and Rand 1993), and those that differ by many traits potentially important in mate choice but that mate at random (e.g. *Colaptes*: Moore 1987; *Bombina*: this study). Understanding the circumstances leading to either case will be an important step in understanding the role of sexual selection in the formation of species.

Section 5.7 Summary

In Chapter 4 and 5 I have tried to quantify the mating behaviour of *Bombina* in the Apahida study area, and I summarise the main conclusions below.

1) The egg data showed a significantly biased segregation ratio, this may be driven by selection or systematic scoring errors. There is also a low level of random scoring errors. However, I could find no evidence for mixed batches.

2) There was good evidence for an active habitat preference among the sites within each adult pool, consistent with the observations of Chapter 2.

3) There were strong differences between the mean allele frequency of the adult pool and the inferred parents of the eggs, but these did not correspond well with the difference in habitat. The location of breeding aggregations may be at least partly independent of the habitat type.

4) Even though the parents generally had a more extreme allele frequency than their adult pool, they generally had lower levels of linkage disequilibrium and heterozygote deficit, suggesting that hybrids and purer types had roughly equal mating success.

5) There was no evidence for strong assortative mating anywhere, especially in relation to the strength of breeding site preference. There was a significant heterozygote deficit in site 0200.4, but the sample size is perhaps too small to demonstrate

significant assortment.

6) The observation of effectively random mating within sites despite the large number of differences between the taxa implies there is much more to understand about the role of sexual selection in generating reproductive isolation.

Chapter 6:
Conclusions and future directions

This thesis has considered two distinct but related topics. In this chapter I summarise the conclusions of each and suggest several avenues for further work, both in *Bombina* and in the study of reproductive isolation in general.

Chapter 1 started by proposing that speciation research must discover how each aspect of divergence between two populations contributes to their reproductive isolation. This entails shifting the focus of research away from explaining the existence of species and subspecies onto understanding how reproductive isolation accumulates between groups of organisms. Hybrid zones are a fundamental part of this programme, as they are a unique opportunity to measure how the divergence between two taxa affects the rate of gene flow between them in a natural context.

How might the 'natural context' affect the structure of hybrid zones? Chapter 2 described a hybrid zone between the fire-bellied toads *B. bombina* and *B. variegata* in the Transylvanian plain of Romania. The toads here were patchily distributed across the study area, and there was no steep cline in any direction. This structure is in sharp contrast to the steep clines found in the *Bombina* hybrid zones in Poland, Croatia and the Ukraine. It appears that the differences are due to the distribution of available habitat in the two areas, but other explanations are possible. Interestingly, the active preference of *B. bombina* for ponds and of *B. variegata* for puddles was almost two times stronger in Romania than in Croatia.

To understand these differences in structure and habitat preference more fully, we need a model describing how the distribution of habitat affects the outcome of

secondary contact between ecologically diverged taxa, and what strength of habitat preference is required for them to exist in sympatry. The model of Kruuk (1997) would be an excellent starting point. To be able to compare the results of such a model with field data for *Bombina*, we must know what constitutes suitable habitat from the perspective of a toad, and we should understand how this is distributed in each country. Lastly, data on the strength of habitat preference outside the zone of contact is necessary to understand how the presence of the other taxon affects habitat use.

I further explored the dynamics of the Apahida mosaic in Chapter 3 through the associations within and between the marker loci. Both the linkage disequilibrium and the heterozygote deficit appear to be driven by the movement of pure *B. bombina* toads into intermediate habitats. Here they meet and hybridise with *B. variegata*-like animals. The selection required to prevent the swamping of the locally adaptive *variegata* alleles by pond-adapted *B. bombina* alleles is not implausibly strong. This selection must be against immigrant alleles and not based on hybrid inviability, as the latter requires too many selective deaths in the local population. However, ecological selection against immigrant adults is unlikely, and we need to identify how else selection is operating in *Bombina*. Only then can we understand how the adaptive differences between the two taxa contribute to the maintenance of their reproductive isolation.

In contrast with the selected loci, the neutral marker differences in the area can only be maintained by implausibly strong selection, and hence they are probably collapsing. This is evident from the high level of introgression into the *B.variegata* pool at the marker loci: there are hardly any pure animals in the study area. With

more marker loci, it may be possible to locate the selected loci with a QTL analysis, and then assess how many generations of recombination are needed for the observed level of introgression at nearby neutral loci. A similar study in sunflowers found that a hybrid speciation event had taken place between 20 and 60 generations before (Ungerer et al. 1998).

One potentially important source of selection against immigrant *B. bombina* is mate discrimination. I quantified this by fitting parameters describing how the local adult population paired up to produce an observed set of egg batches. These showed that there was little or no assortative mating between the parents in these intermediate sites, but there was a pronounced shift in allele frequency towards *B. variegata* in four of the five sites. This suggests that there is a strong breeding site preference in *Bombina*, although how this preference corresponds to the habitat type is unclear. More detailed estimates of how the parents relate both to the surrounding adult pool and the local habitat would allow us to assess the effect of this isolating mechanism on the stability of the Apahida mosaic.

Other work on *Bombina* mating behaviour would also be interesting, such as assessing the ability of females to distinguish between the calls of either taxon. Alternatively, it would be interesting to see how the rate of hybridisation depends on the mating system. For example, it is possible that there is less opportunity for female choice in the scramble polygyny practised by *B. variegata* than in the territorial system of *B. bombina*. There are also wider issues raised by these findings: why is there apparently so little sexual isolation between phenotypic distinct populations, and what does this tell us about the traits used for mate recognition? The study of sexual selection and speciation has often neglected the

distinction between mate choice and mate recognition (see Ryan and Rand 1993), and experiments validating the present assumptions are long overdue.

In summary, this thesis has demonstrated two new facts about *Bombina*. First, the hybrid zones between the two taxa can take on a range of structures, and their dynamics can be quantified with a small number of markers. Second, the mating behaviour of *Bombina* can also be measured effectively even when neither parent has been sampled. Hopefully, both these results will contribute towards a wider understanding of the role of the environment and assortative mating in the evolution of reproductive isolation.

References

Akin, E., H. Levene, L. Levene, and R. Rockwell. 1984. A conservative procedure for the estimation of multiple insemination in *Drosophila*. Am. Nat. 124:723-737.

Andersson, M. 1986. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. Evolution 40:804-816.

Arnold, M. L., J. L. Hamrick, and B. D. Bennett. 1993. Interspecific pollen competition and reproductive isolation in *Iris*. J. Hered. 84:13-16.

Arnold, S. J., P. A. Verrell, and S. G. Tilley. 1996. The evolution of asymmetry in sexual isolation: a model and test case. Evolution 50:1024-1033.

Arnqvist, G. 1997. The evolution of animal genitalia: distinguishing between hypotheses by single species studies. Biol. J. Linn. Soc. 60:365-379.

Arnqvist, G., M. Edvardsson, U. Friberg, and T. Nilsson. 2000. Sexual conflict promotes speciation in insects. Proc. Natl. Acad. Sci. USA 97:10460-10464.

Arnqvist, G., and L. Rowe. 2002. Correlated evolution of male and female morphologies in water striders. Evolution 56:936-947.

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

Asmussen, M. A., and M. M. Orive. 2000. The effects of pollen and seed migration on nuclear-dicytoplasmic systems. I. Nonrandom associations and equilibrium structure with both maternal and paternal cytoplasmic inheritance. Genetics 155:813-831.

Bacilieri, R., A. Ducousso, R. J. Petit, and A. Kremer. 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 50:900-908.

Baird, S. J. E. 1995. The mixing of genotypes in hybrid zones: a simulation study of multilocus clines. Evolution 49:1038-1045.

Baradun, J. 1990. Reproduction of yellow-bellied toads *Bombina variegata* in a man-made habitat. Amphibia-Reptilia 11:277-284.

Barraclough, T. G., P. H. Harvey, and S. Nee. 1995. Sexual selection and taxonomic diversity in passerine birds. Proc. Roy. Soc. Lond. B 259:211-215.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454-471.

Barton, N. H. 1986. The effects of linkage and density-dependent regulation on gene flow. Heredity 57:415-426.

Barton, N. H. 1992. On the spread of new gene combinations in the third phase of Wright's shifting balance. Evolution 46:551-557.

Barton, N. H. 2000. Estimating multilocus linkage disequilibria. Heredity 84:373-389.

Barton, N. H., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357-376.

Barton, N. H., and B. Charlesworth. 1984. Founder effects, genetic revolutions and speciation. Ann. Rev. Ecol. Syst. 15:133-164.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13-45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Ann. Rev. Ecol. Syst. 16:113-148.

Barton, N. H., and M. Shpak. 2000. The effect of epistasis on the structure of hybrid zones. Genet. Res. Camb. 75:179-198.

Barton, N. H., and M. Turelli. 1991. Natural and sexual selection on many loci. Genetics 127: 229-255.

Barton, N. H., and M. C. Whitlock. 1997. The evolution of metapopulations. Pp 183-210 *in* I. Hanski and M. E. Gilpin, eds. Metapopulation dynamics: ecology, genetics, and evolution. Academic Press, San Diego, CA.

Basolo, A. L. 1990. Female preferences predates the evolution of the sword in swordtail fish. Science 250:808-810.

Blair, A. P. 1941. Variation, isolating mechanisms, and hybridization in certain toads. Genetics 26:398-417.

Blair, W. F. 1962. Non-morphological data in anuran classification. Syst. Zool. 11:72-84.

Blows, M. W., and R. A. Allen. 1997. Levels of mate recognition within and between two hybrid *Drosophila* species and their hybrids. Am. Nat. 152:826-837.

Boake, C. R. B., M. P. DeAngelis, and D. K. Andreadis. 1997. Is sexual selection and species recognition a continuum? Proc. Natl. Acad. Sci. USA 94:12442-12445.

Boecklen, W. J., and D. J. Howard. 1997. Genetic analysis of hybrid zones: number of markers and power of resolution. Ecology 78:2611-2616.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411:944-948.

Boulenger G. A. 1886. On two European species of Bombinator. Proc. Zool. Soc. Lond. 1886:499-501.

Bridle, J. R., S. J. E. Baird, and R. K. Butlin. 2001. Spatial structure and habitat variation in a grasshopper hybrid zone. Evolution 55:1832-1843.

Burke, J. M., M. R. Bulger, R. A. Wesselingh, and M. L. Arnold. 2000. Frequency and spatial patterning of clonal reproduction in Louisiana *Iris* hybrid populations. Evolution 54:137-144.

Butlin. R. K. 1989. Reinforcement of premating isolation. Pp. 158-179 *in* Otte D. and Endler J. A., eds. Speciation and its consequences. Sinauer Associates, Sunderland, MA.

Butlin, R. K., and T. Tregenza. 1998. Levels of genetic polymorphism: marker loci and quantitative traits. Phil. Trans. Roy. Soc. B 353:187-198.

Cain, M. L., V. Andreasen, and D. J. Howard. 1999. Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. Evolution 53:1343-1353.

Carson, H. L. 1978. Speciation and sexual selection in Hawaiian *Drosophila*. Pp 93-107 *in* P.F. Brussard, ed. Ecological genetics: the interface. Springer, New York, NY.

Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. Expected proportion of offspring with unambiguous parentage. Genetics 118:527-536.

Chakraborty, R., M. Shaw, and W. J. Schull. 1974. Exclusion of paternity: the current state of the art. Am. J. Hum. Genet. 26:477-488.

Clark, A. G., D. J. Begun, and T. Prout. 1999. Female x male interactions in *Drosophila* sperm competition. Science 283:217-220.

Coyne, J. A., and H. A. Orr. 1998. The evolutionary genetics of speciation. Phil. Trans. Roy. Soc. Lond. B. 353:287-305.

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. A critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643-671.

Crawley, M. J. 1993. GLIM for ecologists. Blackwell Scientific Publications, Oxford.

Czaja, A. 1980. Studies on the variability of the oral apparatus of tadpoles of *Bombina bombina*, *Bombina variegata*, and their natural hybrid populations [in Polish]. M. S. thesis, Jagiellonian University, Cracow.

Czopkowa, G., and J. Czopek. 1955. The vascularisation of the respiratory surfaces in *Bombina variegata.* Bulletin Academia Polonica Scientatis Cl II:313-318.

Dawson, K. J., and Belkhir, K. 2001. A Bayesian approach to the identification of panmictic populations and the assignment of individuals. Genet. Res. Camb. 78:59-77.

Devlin, B., K. Roeder, and N. C. Ellstrand. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. Theor. App. Genet. 76:369-380.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York, NY.

Dobzhansky, T. 1940. Speciation as a stage in the evolutionary process. Am. Nat. 74:312-321.

Duellmann, W. E. and L. Trueb. 1985. Biology of amphibians. McGraw-Hill, New York, NY.

Edwards, A. W. F. 1972. Likelihood. Cambridge Univ. Press, Cambridge.

Emery, A. M., I. J. Wilson, S. Craig, P. R. Boyle, and L. R. Noble. 2001. Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. Mol. Ecol. 10:1265-1278.

Emms, S. K., and M. L. Arnold. 2000. Site-to-site differences in pollinator visitation patterns in a Louisiana iris hybrid zone. Oikos 91:568-578.

Emms, S. K., S. A. Hodges, and M. L. Arnold. 1996. Pollen-tube competition, siring success and consistent asymmetric hybridization in Louisiana irises. Evolution 50:2201-2206.

Endler, J. A. 1977. Geographic variation, speciation and clines. Princeton University Press, Princeton, NJ.

Endler, J. A., and A. E. Houde. 1995. Geographic variation in female preference for male traits in *Poecilia reticulata.* Evolution 49:456-468.

Endler, J. A., and M. Thery. 1996. Interacting effects of lek placement, display behavior, ambient light, and color patterns in three neotropical forest birds. Am. Nat. 148:421-452.

Essen-Möller, E. 1939. Die Beweiskraft der Ähnlichkeit im Vaterschaftweis. Mitt. Anthrop. Ges. Wien. 68:9-53.

Frank, S. A. 1998. Foundations of social evolution. Princeton University Press, Princeton, NJ.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution 35:124-138.

Garcia, D. K., and S. K. Davis. 1994. Evidence for a mosaic hybrid zone in the grass shrimp *Palaemonetes kadiakensis* (Palaemonidae) as revealed by multiple genetic markers. Evolution 48:376-391.

Gavrilets, S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403:886-889.

Gerhardt, H. C. 1994. The evolution of vocalization in frogs and toads. Ann. Rev. Ecol. Syst. 25:293-324.

Gollmann, G. 1984. Allozymic and morphological variation in the hybrid zone between *Bombina bombina* and *Bombina variegata* (Anura, Discoglossidae) in northeastern Austria. Z. Zool. Syst. Evolut.-forsch. 22: 51-64.

Gollmann, G. 1987. *Bombina bombina* and *B. variegata* in the Mátra mountains (Hungary): New data on distribution and hybridization (Amphibia, Anura, Discoglossidae). Amphibia-Reptilia 8:213-224.

Gollmann, G., P. Roth, and W. Hödl. 1988. hybridization between the fire-bellied toads *Bombina bombina* and *Bombina variegata* in the karst regions of Slovakia and Hungary. J. Evol. Biol. 1:3-14.

Gosner, K. L. 1960 A simplified table for staging anuran embryos and larvae with notes on identification. Herpetologica 16:183-190.

Grant, P. R. 1986. Ecology and evolution of Darwin's finches. Princeton University Press, Princeton, NJ.

Grant, P. R., and B. R. Grant. 1992. Hybridization of bird species. Science 256:193-197.

Hacking, I. 1965. Logic of statistical inference. Cambridge University Press, Cambridge.

Haldane, J. B. S. 1932. The causes of evolution. Cornell University Press, Ithaca, NY.

Haldane, J. B. S. 1948. The theory of a cline. J. Genetics 48:277-284.

Hamilton, W. D., and M. Zuk. 1982. Heritable true fitness and bright birds: a role for parasites? Science 218:384-387.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. Heredity 56:337-349.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Oxford Surv. Evol. Biol. 7:69-128.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp 110-133 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, MA.

Harrison, R. G., and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493-505.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics (3rd Edition). Sinauer Associates, Sunderland. MA.

Harushima, Y., M. Nakagahra, M. Yano, T. Sasaki, and N. Kurata. 2001. A genome-wide survey of reproductive barriers in an intraspecific hybrid. Genetics 159:883-892.

Hatfield, T., and Schluter D. 1999. Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. Evolution 53:866-873.

Hauffe, H. C., and J. B. Searle. 1993. Extreme karyotypic variation in a *Mus musculus domesticus* hybrid zone: the tobacco mouse story revisited. Evolution 47:1374-1395.

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904-907.

Hay, J. M., I. Ruvinsky, S. B. Hedges, and L. R. Maxson. 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. Mol. Biol. Evol. 12:928-937.

Helbig, A. J., M. Salomon, S. Bensch, and I. Seibold. 2001. Male-biased gene flow across an avian hybrid zone: evidence from mitochondrial and microsatellite DNA. J. Evol. Biol. 14:277-287.

Hewitt, G. M. 1993. After the ice: *parallelus* meets *erythropus* in the Pyrenees. Pp. 140-164 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Hewitt, G. M. 2001. Speciation, hybrid zones and phylogeography- or seeing genes in space and time. Mol. Ecol. 10:537-549.

Hewitt, G. M., R. A. Nichols, and N. H. Barton. 1987. Homogamy in a hybrid zne in the alpine grasshopper *Podisma pedestris*. Heredity 59:457-466.

Hey, J. 2001. Genes, categories and species. Oxford University Press, Oxford.

Higashi, M., G. Takimoto, and N. Yamamura. 1999. Sympatric speciation by sexual selection. Nature 402:523-526.

Hill, W. G. 1974. Estimation of linkage disequilibrium in randomly mating populations. Heredity 33:229-239.

Hodges, S. A., J. M. Burke, and M. L. Arnold. 1996. Natural formation of *Iris* hybrids: experimental evidence on the establishment of hybrid zones. Evolution 50:2504-2509.

Holland, B., and W. R. Rice. 1998. Chase-away sexual selection: antagonistic seduction versus resistance. Evolution 52:1-7.

Howard, D. J., P. G. Gregory, J. M. Chu, and M. L.. Cain. 1998. Conspecific sperm precedence is an effective barrier to hybridization between closely related species. Evolution 52:511-516.

Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansions. Heredity 77:282-291.

Iwasa, Y., and A. Pomiankowski. 1995. Continual change in mate preferences. Nature 377:420-422.

Johannesson, K., E. Rolán-Alvarez, and A. Ekendahl. 1995. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. Evolution 49:1180-1190.

Johnston, J.A., R. A. Wesselingh, A. C. Bouck, L. A. Donovan, and M. L. Arnold. 2001. Intimately linked or hardly speaking? The relationship between genotype and environmental gradients in a Louisiana *Iris* hybrid zone. Mol. Ecol. 10:673-681.

Kaneshiro, K. Y. 1980. Sexual isolation, speciation and the direction of evolution. Evolution 34:437-444.

Key, K. H. L. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14-22.

Kirkpatrick, M., and N. H. Barton. 1995. *Déjà vu* all over again. Nature 377:388-389.

Kirkpatrick, M., and N. H. Barton. 1997. The strength of indirect selection on female mating preferences. Proc. Natl. Acad. Sci. USA 94:1282-1286.

Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: models and experiments. Am. Nat. 159:S22-S35.

Kirkpatrick, M., and M. J. Ryan. 1991. The evolution of mating preferences and the paradox of the lek. Nature 350:33-38.

Kirkpatrick, M., and M. Servedio. 1999. The reinforcement of mating preferences on an island. Genetics 151:865-884.

Kirkpatrick, S., C. D. Gelatt and M. P. Vecchi. 1983. Optimization by simulated annealing. Science 220: 671-680.

Klingenberg, C. P., J. R. Spence, and C. K. Mirth. 2000. Introgressive hybridization between two species of waterstriders (Hemiptera: Gerridae: *Limnoporus*): geographical structure and temporal change of a hybrid zone. J. Evol. Biol. 13:756-765.

Koteja, E. 1984. Embryonic mortality of the fire-bellied toad (*Bombina bombina*), the yellow-bellied toad (*Bombina variegata*), and their hybrids [in Polish]. M. S. thesis, Jagiellonian University, Cracow.

Kruuk, L. E. B. 1997. Barriers to gene flow: a *Bombina* (fire-bellied toad) hybrid zone and multilocus cline theory. Ph. D. diss., University of Edinburgh.

Kruuk, L. E. B., and J. S. Gilchrist. 1997. Mechanisms maintaining species differentiation: predator mediated selection in a *Bombina* hybrid zone. Proc. Roy. Soc. Lond. B 264: 105-110.

Kruuk, L. E. B., J. S. Gilchrist, and N. H. Barton. 1999a. Hybrid dysfunction in fire-bellied toads (*Bombina*). Evolution 53:1611-1616.

Kruuk, L. E. B., S. J. E. Baird, K. S. Gale, and N. H. Barton. 1999b. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics 153:1959-1971.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. USA 78:3721-3725.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 36:1210-1226.

Lewontin, R. C. 1988. On measures of gametic disequilibrium. Genetics 120:849--852.

Littlejohn, M. J., and J. J. Loftus-Hills. 1968. An experimental evaluation of premating isolation in the *Hyla ewingi* complex (Anura: Hylidae). Evolution 22:659-663.

Lörcher K. 1969. Vergleichende bio-akustische Untersuchungen an der Rot- und Gelbbauchunke, *Bombina bombina* und *B. v. variegata*. Oecologia 3:84-124.

Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766.

Lynch, M., and J. B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

Lythgoe, K. A. 1997. Consequences of gene flow in spatially structured populations. Genet. Res. Camb. 69:49-60.

MacCallum, C. J. 1994. Adaptation and habitat preference in a hybrid zone between *Bombina bombina* and *Bombina variegata* in Croatia. Ph. D. diss., University of Edinburgh.

MacCallum, C. J., B. Nürnberger, N. H. Barton, and J. M. Szymura. 1998. Habitat preference in a *Bombina* hybrid zone in Croatia. Evolution 52:227-239.

Madej, Z. 1965. Variations in the sacral region of the spina in *Bombina bombina* (Linnaeus, 1761) and *Bombina variegata* (Linnaeus, 1758) (Salientia, Discoglossidae). Acta Biol. Cracow 8:185-197.

Mallet, J. 1995. A species definition for the Modern Synthesis. Trends Ecol. Evol. 10:294-299.

Mallet, J., and N. H. Barton. 1989. Strong natural selection in a warning color hybrid zone. Evolution 43:421-431.

Mallet, J., W. O. McMillan, and C. D. Jiggins. 1998. Estimating the mating behavior of a pair of hybridizing *Heliconius* species in the wild. Evolution 52:503-510.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7:639-655.

Maxson, L. R., and J. M. Szymura. 1984. Relationships among discoglossid frogs: an albumin perspective. Amphibia Reptilia 5:245-252.

Mayr, E. 1942. Systematics and the origin of species. Columbia University Press, New York, NY.

McLain, D. K. 1985. Clinal variation in morphology and assortative mating in the soldier beetle, *Chauliognathus pennsylvanicus* (Coleoptera: Cantharidae). Biol. J. Linn. Soc. 25:105-117.

Meagher, T. R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum.* I. Identification of most-likely male parents. Am. Nat. 128:199-215.

Méhelÿ, L. 1892. Beitrage zur Kenntnis der Bombinator-Arten, sowie deren Standorte und Verbreitung in Ungarn. Math. Nat. Ber. Ungarn. 10:55-79.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087-1095.

Michałowski, J., and Z. Madej. 1969. Studies on the relationship of *Bombina bombina* (Linnaeus) and *Bombina variegata* (Linnaeus). III. Taxonomic characters of both species and of tadpoles obtained from crosses under laboratory conditions. Acta Zool. Cracow 14:173-200.

Michaud, T. C. 1962. Call discrimination by females of the chorus frogs, *Pseudacris clarki* and *Pseudacris nigrita*. Copeia 1962:213-215.

Moore, W. S. 1977. An evaluation of narrow hybrid zones in vertebrates. Q. Rev. Biol. 52:263-277.

Moore, W. S. 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, M. T., and J. K. Conner. 2001. Using genetic markers to directly estimate male selection gradients. Evolution 55:272-281.

Nagylaki, T. 1975. Conditions for the existence of clines. Genetics 80:595-615.

Navarro, A., and N. H. Barton. 2002. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. Genetics, in review.

Neel, J. V., and W. J. Schull. 1954. Human Heredity. University of Chicago Press, Chicago, IL.

Neff, B. D., J. Repka, and M. R. Gross. 2000a. Parentage analysis with incomplete sampling of candidate parents and offspring. Mol. Ecol. 9:515-528.

Neff, B. D., J. Repka, and M. R. Gross. 2000b. Statistical confidence in parentage analysis with incomplete sampling: how many loci and offspring are needed? Mol. Ecol. 9:529-539.

Neff, B. D., J. Repka, and M. R. Gross. 2001. A Bayesian framework for parentage analysis: the value of genetic and other biological data. Theor. Pop. Biol. 59:315-331.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283-292.

Nichols, R. A., and G. M. Hewitt. 1994. The genetic consequences of long distance dispersal during colonization. Heredity 72:312-317.

Nielsen, R., D. K. Mattila, P. J. Clapham, and P. J. Palsbøll. 2001. Statistical approaches to paternity analysis in natural populations and applications to the north Atlantic humpback whale. Genetics 157:1673-1682.

Noor, M. A. F. 1999. Reinforcement and other consequences of sympatry. Heredity 83:503-508.

Noor, M. A. F., K. I. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA 98:12084-12088.

Nürnberger, B., N. H. Barton, C. J. MacCallum, J. Gilchrist, and M. Appleby. 1995. Natural selection on quantitative traits in a *Bombina* hybrid zone. Evolution 49: 1224-1238.

Nürnberger, B., L. E. B. Kruuk, N. H. Barton, and T. H. Vines. 2002a. Mating patterns in a *Bombina* hybrid zone: inferences from full sib genotypes when neither parental genotype is known. Heredity, in review.

Nürnberger, B., J. M. Szymura, A. MacLean, B. Förg-Brey, G. Praetzel, C. Abbott, and N. H. Barton. 2002b. A linkage map for the hybridising toads *Bombina bombina* and *B. variegata* (Anura, Discoglossidae). Heredity, in review.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805-1813.

Orr, H. A. 2001. The genetics of species differences. Trends Ecol. Evol. 16:343-350.

Orr, H. A., and D. C. Presgraves. 2000. Speciation by postzygotic isolation: forces, genes and molecules. Bioessays 22:1085-1094.

Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55:1085-1094.

Orr, M. R. 1996. Life-history adaptation and reproductive isolation in a grasshopper hybrid zone. Evolution 50:704-716.

Painter, I. 1997. Sibship reconstruction without parental information. J. Agric. Biol. Env. Stat. 2:212-229.

Panhuis, T. M., R. K. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends Ecol. Evol. 16: 364-371.

Parker, G. A., and L. Partridge. 1998. Sexual conflict and speciation. Phil. Trans. Roy. Soc. Lond. B. 353:261-274.

Paterson, H. E. H. 1980. A comment on "mate recognition systems". Evolution 34:330-331.

Paterson, H. E. H. 1993. Evolution and the recognition concept of species: collected writings. Johns Hopkins University Press, Baltimore, MD.

Payne, R. J. H., and D. C. Krakauer. 1997. Sexual selection, space and speciation. Evolution 51:1-9.

Pomiankowski, A. 1987. The cost of choice in sexual selection. J. Theor. Biol. 128:195-218.

Pounds, N. J. G. 1979. A historical geography of Europe 1500-1840. Cambridge Univ. Press, Cambridge.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Queller, D. C., and K. F. Goodnight. 1989. Estimating relatedness using molecular markers. Evolution 43:258-275.

Rafinska, A. 1991. Reproductive biology of the fire-bellied toads, *Bombina bombina* and *B. variegata* (Anura: Discoglossidae): egg size, clutch size and larval period length differences. Biol. J. Linn. Soc. 43:197-210.

Rand, D., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432-449.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA 94:9197-9201.

Raufaste, N., and F. Rousset. 2001. Are partial Mantel tests adequate? Evolution 55:1703-1705.

Reyer, H.-U., G. Frei, and C. Som. 1999. Cryptic female choice: frogs reduce clutch size when amplexed by an undesired male. Proc. Roy. Soc. Lond. B 266:2101-2107.

Ribi, G., and S. Oertli. 2000. Frequency of interspecific matings and of hybrid offspring in sympatric populations of *Viviparus ater* and *V. conectus* (Mollusca: Prosobranchia). Biol. J. Linn. Soc. 71:133-43.

Rice, W. R. 1987. Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. Evol. Ecol. 1:301-314.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637-1653.

Rieseberg, L. H., S. J. E. Baird, and A. M. Desrochers. 1998. Patterns of mating in wild sunflower hybrid zones. Evolution 52:713-726.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713-727.

Ritchie, M. G. 1990. Does song contribute to assortative mating between subspecies of *Chorthippus parallelus* (Orthoptera: Acrididae)? Anim. Behav. 39:685-691.

Ritchie, M. G., R. K. Butlin, and G. M. Hewitt. 1989. Assortative mating across a hybrid zone in *Chorthippus parallelus*. J. Evol. Biol. 2:339-352.

Ritland, K. 1996. Estimators for pairwise relatedness and inbreeding coefficients. Genet. Res. Camb. 67:175-186.

Roeder, K., B. Devlin, and B. G. Lindsay. 1989. Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities. Biometrics 45:363-379.

Ross, C. L., and R. G. Harrison. 2002. A fine scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvaticus*. Evolution, in press.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287:306-308.

Ryan, M. J., and A. S. Rand. 1990. The sensory basis of sexual selection for complex calls in the túngara frog *Physalaemus pustulosus* (sexual selection for sensory exploitation). Evolution 44:305-314.

Ryan, M. J., and A. S. Rand. 1993. Species recognition and sexual communication as a unitary problem in animal communication. Evolution 47:647-657.

Sætre, G-P, T. Moum, S. Bureš, M. Král, M. Adamjam, and J. Moreno. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. Nature 387:589-592.

Sanchristobal, M., and C. Chevalet. 1997. Error tolerant parent identification from a finite set of individuals. Genet. Res., Camb. 70:53-62.

Sanderson, N. 1989. Can gene flow prevent reinforcement? Evolution 43:1223-1235.

Sanderson, N., J. M. Szymura, and N. H. Barton. 1992. Variation in mating call across the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. Evolution 46: 595-607.

Schartl, M. 1995. Platyfish and swordtails: a genetic system for the analysis of molecular mechanisms of tumor formation. Trends Genet. 11:185-189.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372-380.

Schluter, D., and T. Price. 1993. Honesty, perception and population divergence in sexually selected traits. Proc. Roy. Soc. Lond. B. 253:117-122.

Schneider, H., and H. Eichelberg. 1974. The mating calls of hybrids of the fire-bellied toad and yellow bellied toad, (*Bombina bombina*, *Bombina variegata*, Discoglossidae, Anura). Oecologia 16:61-71.

Seidel, B. 1987. Breeding of a *Bombina variegata* population in a habitat with temporary pools. Pp 353-356 *in* J. J. van Gelder, H. Strijbosch, and P. J. M. Berg-

ers, eds. Proceedings of the fourth ordinary general meeting of the Societas Europaea Herpetologica. Nijmegen Univ. Press, Nijmegen.

Servedio, M. 2001. Beyond reinforcement: the evolution of premating isolation by direct selection on female preferences and postmating, prezygotic incompatibilities. Evolution 55:1909-1920.

Shapiro, L. H. 1998. Hybridization and geographic variation in two meadow katydid contact zones. Evolution 52:784-796.

Shapiro, L. H. 2001. Asymmetric assortative mating between two hybridizing *Orchelimum* katydids (Orthoptera: Tettigoniidae). Am. Midl. Nat. 145:423-427.

Shoemaker, D. D., K. G. Ross, and M. L. Arnold. 1996. Genetic structure and evolution of a fire ant hybrid zone. Evolution 50:1958-1976.

Sites, J. W., N. H. Barton, and K. M. Reed. 1995. The genetic structure of a hybrid zone between two chromosome races of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in Central Mexico. Evolution 49:9-36.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733-756.

Slatkin, M. 1978. Spatial patterns in the distributions of polygenic characters. J. Theor. Biol. 70:213-228.

Smadja, C., and G. Ganem. 2002. Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. Behav. Ecol. 13:312-320.

Smith, B. R., C. M. Herbinger, and H. R. Merry. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. Genetics 158:1329-1338.

Smouse, P. E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. 35:627-632.

Smouse, P. E., and T. R. Meagher. 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). Genetics136:312-3222.

Smouse, P. E., T. R. Meagher, and C. J. Kobak. 1999. Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): why do some males have higher reproductive contributions? J. Evol. Biol. 12:1069-1077.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. W.H. Freeman and Co., New York, NY.

Sprott, D. A. 2000. Statistical inference in science. Springer-Verlag, New York, NY.

Stugren, B. 1959. Eidonomische Untersuchungen an Bombina Oken (Amph., Discoglossidae) aus dem Gurghiu-Tale. Zool. Jb. Syst. 86:382-394.

Stugren, B., and S. Vancea. 1968. Geographic variation of the Yellow Bellied Toad (*Bombina variegata*) (L.) from the Carpathian mountains of Romania and the USSR. J. Herpetol. 2:97-105.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137-144.

Szymura, J. M. 1983. Genetic differentiation between hybridizing species *Bombina bombina* (L.) and *Bombina variegata* (L.) (Salientia, Discoglossidae) in Poland. Amphibia Reptilia 4:137-145.

Szymura, J. M. 1988. Regional differentiation and hybrid zones in the fire-bellied toads *Bombina bombina* and *B. variegata* in central Europe [in Polish]. Rozprawy Habilitacyjne No. 147, Uniwersytet Jagiellonski, Krakow.

Szymura, J. M. 1993. Analysis of hybrid zones with *Bombina*. Pp. 261-289 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. Evolution 40:1141-1159.

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

Thomas, S. C., and W. G. Hill. 2000. Estimating variance components in natural populations using inferred relationships. Heredity 84:427-436.

Thompson, E. A. 1975. The estimation of pairwise relationships. Ann. Hum. Genet. 39:173-188.

Ting, C. T., S. C. Tsaur, M. L. Wu, and C. I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501-1504.

Trexler, J. C., J. Travis, and A. Dinep. 1997. Variation among populations of the sailfin molly in the rate of concurrent multiple paternity and its implications for mating system evolution. Behav. Ecol. Sociobiol. 40:297-305.

Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16:330-343.

Turner, G. F., and M. T. Burrows. 1995. A model of sympatric speciation by sexual selection. Proc. Roy. Soc. Lond. B 260:287-292.

Turner, G. F., O. Seehausen, M. E. Knight, C. J. Allender, and R. L. Robinson. 2001. How many species of cichlid fishes are there in African lakes? Mol. Ecol. 10:793-806.

Ungerer, M. C., S. J. E. Baird, J. Pan, and L. H. Rieseberg. 1998. Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95:11757-11762.

Vamosi, S. M., and D. Schluter. 1999. Sexual selection against hybrids between sympatric stickleback species: evidence from a field experiment. Evolution 53:874-879.

Veen, T., T. Borge, S. C. Griffiths, G.-P. Sætre, S. Bureš, L. Gustafsson, and B. C. Sheldon. 2001. Hybridization and adaptive mate choice in flycatchers. Nature 411:45-50.

Via, S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. Evolution 45:827-852.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16: 381-390.

Vines, T. H., S. C. Köhler, M. Thiel, I. Ghira, T. R. Sands, N. H. Barton, and B. Nürnberger. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. Evolution, in review.

Vorndran, I. C., E. Reichwald, and B. Nürnberger. 2002. Does differential susceptibility to predation on tadpoles stabilize the *Bombina* hybrid zone? Ecology 83:1648-1659.

Westneat, D. F., P. C. Frederick, and R. H. Wiley. 1987. The use of genetic markers to estimate the frequency of successful alternative reproduction tactics. Behav. Ecol. Sociobiol. 21:35-45.

Wolfram. S. 1999. The *Mathematica* book 4th Edition. Wolfram Media / Cambridge University Press, Cambridge.

Wrege, P. H., and S. T. Emlen. 1987. Biochemical determination of parental uncertainty in white-fronted bee-eaters. Behav. Ecol. Sociobiol. 20:153-160.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Zahavi, A. 1975. Mate selection - a selection for a handicap. J. Theor. Biol. 53:205-214.

Zane, L., W. S. Nelson, A. G. Jones and J. C. Avise. 1999. Microsatellite assessment of multiple paternity in natural populations of a live-bearing fish, *Gambusia holbrooki*. J. Evol. Biol. 12:61-69.

Appendix I

The adult genotypes. For *Bb7.4*, *Bv12. 19* and *Bv24.11*, the 'v' and 'x' alleles are both *B. variegata*, and the 'b' *B. bombina*.

In *Bv24.12* the 'v', 'a', 'd' and 'f' alleles are assigned to *B. variegata* and 'b' to *B. bombina*. A '-1' indicates missing data.

Site	ind	7.4	12.19	24.11	24.12
0200.3	1	vb	vb	vv	2
0200.3	2	vv	vv	vv	1
0200.3	4	vv	vb	xx	2
0200.3	5	vv	vv	vv	2
0200.3	6	vb	-1	bx	1
0200.3	7	vv	vv	bb	2
0200.3	8	vv	vx	vv	2
0200.3	9	vx	-1	bx	2
0200.3	10	vv	xx	bx	2
0200.3	11	vv	xx	bx	2
0200.3	12	bb	-1	bb	1
0200.3	13	vv	bx	bx	2
0200.4	1	vv	xx	xx	2
0200.4	2	vv	xx	vx	1
0200.4	3	-1	bx	-1	2
0200.4	4	vv	vb	xx	2
0200.4	5	vv	vb	xx	2
0200.4	6	bb	vv	bx	1
0200.4	7	vv	bx	bb	2
0200.4	8	vv	bx	bv	2
0200.4	9	vb	vv	bv	2
0200.4	10	vb	-1	bb	1
0200.4	11	vb	vb	vx	2
0200.4	12	bb	vb	vv	1
0200.5	1	vv	vv	vv	2
0200.5	2	vv	vv	vv	1
0200.5	3	vv	vb	xx	2
0200.5	4	vb	vx	bv	1
0200.5	5	bb	vb	bv	1
0200.5	6	vb	vx	bv	1
0200.5	7	vv	vb	bx	2
0200.5	8	vv	vv	bv	2
0200.6	1	vv	-1	bv	2
0200.6	2	vx	vv	bx	2
0200.6	3	vv	bx	bv	2
0200.6	4	vv	vx	xx	0
0200.6	5	vv	bx	vx	2
0200.6	6	vv	vb	vx	2
0200.6	7	vv	vv	bx	1
0200.7	1	vv	vx	bv	1
0200.7	2	vb	vb	bx	1
0200.7	3	vv	vv	bv	2
0200.7	4	bb	vv	bv	0
0200.7	5	bb	bb	bb	0
0200.7	6	xx	bb	bb	1
0200.8	1	bb	vv	vv	2
0200.8	2	bb	vb	bx	2
0200.8	3	bb	bb	bb	0
0200.8	4	vv	vv	bv	1
0200.8	5	-1	bb	bb	0
0200.8	7	vv	vv	bv	2
0200.8	8	vv	bb	vx	1
0200.8	9	vx	bx	bb	0
0200.8	10	vx	vv	bx	1
0200.9	1	vv	vb	xx	1
0200.9	2	vv	vv	bb	2
0200.9	3	vv	vb	bb	2
0200.9	4	vv	vx	bv	2
0200.9	5	vv	vb	vx	2
0200.9	6	vb	vv	bx	2
0200.101		xx	vb	vv	2
0200.102		vb	vb	bb	0
0200.103		vv	vv	xx	2
0200.104		vv	-1	vx	1
0200.105		vb	-1	vv	2
0244	1	bb	vb	vx	2
0244	2	vb	bx	bx	1
0244	3	vx	vv	bx	2
0244	4	vx	xx	bx	2
0244	5	vv	vb	vx	2
0244	6	vb	vb	vx	2
0245	1	vb	vb	bb	0
0245	2	vv	vv	bx	2
0245	3	vv	-1	bv	2
0245	4	bb	bb	bx	0
0245	5	vb	-1	xx	2

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0245	6	bb	vb	bb	1	0250	8	vv	vv	vv	1
0245	8	vx	vv	vx	2	0250	9	vb	-1	bx	2
0246	1	vb	bx	bx	1	0250	10	vx	-1	xx	1
0246	2	bb	bx	bb	0	0250	11	vb	vv	xx	1
0246	3	vv	-1	vx	1	0250	12	vb	bb	bx	1
0246	4	vb	bb	bb	0	0251	1	vx	vx	bv	2
0246	5	bb	bb	bb	0	0251	2	bb	bx	bb	1
0247	1	vb	vv	bx	2	0251	3	vb	vb	bv	1
0247	2	vb	bb	bb	1	0251	4	vb	bx	vx	2
0247	3	vb	bx	vx	1	0251	5	bb	xx	bx	0
0247	4	vx	xx	vv	2	0251	6	bb	bb	bb	0
0247	5	vx	-1	vx	1	0251	7	bb	bb	bb	0
0247	6	-1	vb	bv	1	0251	8	vb	bb	bv	0
0247	7	vv	vb	bb	1	0251	9	bb	vb	bb	0
0247	8	vb	vb	bx	0	0251	10	vb	bx	bx	2
0247	9	vx	bx	bx	-1	0251	11	vb	vv	xx	1
0247	10	vb	vv	bx	1	0251	12	vv	vb	bx	2
0247	11	vv	vv	vx	2	0252	1	vb	vb	bv	0
0247	12	bb	vb	bx	1	0252	2	vx	vx	bx	2
0248	1	vv	vb	xx	1	0252	3	vx	vv	bb	2
0248	2	bb	bx	xx	1	0252	4	vx	xx	xx	2
0248	3	vv	vv	vx	2	0252	5	xx	vx	xx	1
0248	4	vv	vx	vv	1	0252	6	vb	vv	vx	2
0248	5	vb	-1	vv	1	0252	7	vx	xx	bx	2
0248	6	xx	-1	bv	2	0253	1	vb	vb	bv	1
0248	7	vv	vv	bv	2	0253	2	vx	bb	bx	0
0248	8	vb	bb	bb	0	0253	3	vv	-1	vx	2
0248	9	vb	bx	bx	1	0253	4	vb	vv	bv	2
0248	10	vv	vb	bv	1	0253	5	bb	vb	bb	1
0249	1	bb	bb	bb	0	0253	6	vb	bb	bv	0
0249	2	vb	vb	vv	2	0253	7	vv	vb	vx	1
0249	3	vx	vv	vx	1	0253	7	vv	-1	bv	2
0249	4	vb	vb	bx	2	0253	8	vv	vb	vx	2
0249	5	bb	-1	vv	1	0254	1	vb	vv	bv	2
0249	6	vv	-1	xx	2	0254	2	vv	vb	vx	1
0249	7	vb	-1	bv	1	0255	1	vv	vx	vv	2
0249	8	vb	vv	bx	0	0256	5	vv	bb	vv	1
0250	1	bb	vb	bx	1	0256	6	vv	xx	vv	2
0250	2	vx	-1	xx	2	0256	7	vv	vv	bb	2
0250	3	vv	vv	vv	2	0256	12	bb	bb	bx	1
0250	4	vb	vx	vx	1	0256	13	vv	vv	vv	2
0250	5	vb	xx	-1	1	0256	18	vv	vx	bx	1
0250	6	vv	-1	bv	1	0256	19	vv	vv	bb	2
0250	7	vv	-1	vv	2	0256	20	vv	vx	vx	2

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0256	21	vb	bb	bv	1	0258	14	xb	xx	vx	1
0256	22	vv	vv	bx	1	0258	15	vx	bx	xx	1
0256	23	vv	vb	xx	2	0258	16	vv	bb	vx	2
0256	24	xb	vx	xx	2	0258	17	vb	vb	vv	0
0257	1	bb	vb	bv	0	0258	18	vx	vx	bv	1
0257	2	vx	vx	bx	1	0258	19	bb	bb	bb	0
0257	3	bb	bb	bx	0	0258	20	xx	-1	bv	2
0257	4	vb	bb	bv	1	0258	21	xb	vb	bb	1
0257	5	vb	bx	bv	0	0258	22	bb	vb	bx	1
0257	6	vx	bx	bb	0	0258	23	vv	-1	bx	2
0257	7	vb	vv	vx	1	0258	24	vb	bx	bx	2
0257	8	vb	vv	xx	1	0258	25	vb	vb	bv	1
0257	9	vx	vb	vx	2	0258	26	vv	xx	bx	2
0257	10	vx	vb	bx	2	0258	27	vb	bb	bv	1
0257	11	vv	vv	xx	2	0258	28	vb	bb	bb	0
0257	12	vv	vx	bx	1	0258	29	xx	vv	vv	1
0257	13	vb	vb	bv	1	0258	30	vv	-1	bx	2
0257	14	vv	-1	vx	1	0258	31	vv	bx	vv	2
0257	15	bb	vb	bb	1	0258	32	bb	bb	bv	1
0257	16	vb	vv	bx	1	0259	1	vv	vv	vx	1
0257	17	bb	vv	bx	1	0259	2	vv	vb	bx	2
0257	18	vv	xx	vv	2	0259	3	vv	bx	vx	2
0257	19	vb	vv	bv	2	0259	4	vb	bx	bx	2
0257	20	vx	-1	bx	0	0259	5	bb	bx	bv	1
0257	21	vv	vv	bv	1	0259	6	vb	bx	bx	1
0257	22	vv	vv	bx	1	0259	7	vv	vx	vx	2
0257	23	vv	vx	vx	2	0259	8	vx	vx	vx	2
0257	24	vx	vv	vx	2	0259	9	vb	-1	bx	1
0257	25	vx	vx	vx	0	0259	10	bb	vv	xx	2
0257	26	vv	vv	vx	2	0260	1	bb	bb	bb	1
0257	27	vb	vb	bb	0	0260	2	vx	vx	vv	1
0257	28	vv	-1	vx	2	0260	3	vx	vv	bx	1
0258	1	vx	vv	vx	1	0260	4	vx	vv	vv	2
0258	3	bb	bb	bb	0	0260	5	vb	vv	vv	2
0258	4	bb	bb	bb	0	0260	6	vx	vb	bx	1
0258	5	vb	vb	bb	2	0260	7	vb	-1	bb	1
0258	6	bb	bb	bb	0	0260	8	vv	vx	vx	2
0258	7	bb	bb	bb	0	0260	9	vv	vv	bx	-1
0258	8	vx	vv	vx	1	0260	10	vb	vv	bx	2
0258	9	vv	bx	vv	1	0260	11	vv	vv	vx	2
0258	10	vb	vb	bx	1	0260	12	vv	-1	bv	2
0258	11	vv	vx	vx	2	0261	1	vv	vx	bv	2
0258	12	vv	-1	bv	1	0261	2	vv	vv	bx	2
0258	13	vv	vb	vx	2	0261	3	vv	xx	xx	2

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0261	4	vv	vx	bv	2	0265	6	vv	vx	bx	2
0261	5	vx	vx	xx	2	0265	7	vv	vv	vx	2
0261	6	vx	bb	vv	2	0265	8	vv	vx	xx	2
0261	7	vb	vx	vx	2	0265	9	vv	vv	xx	2
0262	1	vv	vv	bx	2	0265	10	vv	vx	xx	2
0262	2	vv	vx	vx	1	0265	11	vv	vv	vx	2
0262	3	xx	vx	vx	1	0265	12	vv	vv	vv	2
0262	4	vx	vx	vv	2	0266	1	vx	bx	vx	2
0262	5	vx	vb	vv	1	0266	2	vv	vb	xx	2
0262	6	vb	vv	vx	1	0266	3	vv	vb	vx	2
0262	7	vx	-1	bb	2	0266	4	vv	vx	bv	2
0262	8	vx	vb	vv	2	0266	5	vx	vx	vx	2
0262	9	vv	xx	vx	2	0266	6	vv	xx	vx	1
0262	10	vx	vx	bb	2	0266	7	vv	vv	xx	2
0262	11	vb	vb	bv	2	0266	8	vv	vx	vx	2
0262	12	vb	-1	xx	2	0266	9	vb	xx	vx	2
0263	1	vb	-1	bx	2	0266	10	vv	vx	vx	1
0263	2	bb	-1	bv	2	0266	11	vv	vv	vv	2
0263	3	vv	vv	bx	2	0267	1	vv	vx	vx	2
0263	6	vv	vv	bv	2	0267	2	vb	-1	bv	2
0263	8	vb	vv	bv	1	0267	3	vv	vx	vx	2
0263	9	vv	vv	vv	2	0267	4	vv	vb	bx	2
0263	9	vv	vb	bx	2	0267	5	vv	bb	vx	2
0263	10	vb	vx	bx	1	0267	6	vv	vv	vx	2
0263	11	vv	vb	vx	2	0267	7	vv	vv	bv	1
0263	12	vx	vv	vx	1	0267	8	vv	vv	bx	2
0264	1	bb	vx	bx	2	0267	9	vx	vx	vv	2
0264	2	vv	bx	bv	1	0267	10	vv	xx	vv	2
0264	3	vv	vv	vx	1	0267	11	vv	-1	vx	1
0264	4	vb	vx	bv	2	0267	12	vx	vv	vx	2
0264	5	vx	vb	bv	2	0268	1	vv	vx	vv	2
0264	6	vv	bb	bv	1	0268	2	vv	xx	vv	2
0264	7	vv	vv	bx	2	0268	3	vv	vx	bv	0
0264	8	vb	vx	xx	2	0268	4	vx	bb	bv	2
0264	9	vx	vx	vx	2	0268	5	vv	xx	vv	2
0264	10	vv	vb	vv	2	0268	6	vb	vv	vv	2
0264	11	bb	-1	bx	1	0268	7	vx	vx	vv	1
0264	12	vv	vv	vx	2	0268	8	vv	vv	vv	2
0265	1	vv	vb	xx	2	0268	9	vv	vv	vv	2
0265	2	vv	xx	bx	2	0268	10	vx	bb	vx	2
0265	3	vv	xx	vx	2	0268	11	vv	bb	xx	1
0265	4	vv	bx	vv	2	0268	12	vx	-1	bb	2
0265	5	vv	vx	vv	2	0270	1	vv	vx	vv	1
0265	5	vv	vx	xx	2	0270	2	xx	vv	vx	2

Site	ind	7.4	12.19	24.11	24.12
0270	3	xx	vv	bv	2
0270	4	vb	vx	bv	2
0270	5	vb	vx	vx	2
0270	6	vv	bx	bx	1
0270	7	vb	vx	bv	2
0270	8	vv	vx	xx	2
0270	9	vv	vb	xx	1
0270	10	vx	vx	bb	2
0270	11	vv	vv	bx	2
0270	12	vv	vv	xx	1
0271	1	vv	vv	bx	2
0271	2	vv	vb	xx	1
0271	3	vb	vv	bx	2
0271	4	vv	xx	vx	2
0271	5	vv	vb	bx	2
0271	6	vx	vv	vv	2
0271	7	vv	xx	bx	1
0271	8	vv	vv	vv	2
0272	1	vb	vv	bx	2
0272	2	vv	vv	vv	2
0272	3	vv	vv	vx	1
0272	4	vv	xx	vv	1
0272	5	vb	vb	bv	2
0272	6	vv	vv	vv	2
0272	8	vv	vx	xx	2
0272	9	vv	vb	xx	2
0272	10	vb	-1	xx	1
0272	11	vv	xx	xx	2
0273	1	vv	vv	bx	2
0273	2	vv	vb	vx	2
0273	3	vv	xx	xx	1
0273	4	vv	vv	xx	2
0273	5	vx	vv	xx	2
0273	6	vv	vv	vv	1
0274	1	vv	vb	vx	2
0274	2	vb	-1	bx	2
0274	3	xx	vv	xx	2
0274	4	vv	vx	xx	2
0274	5	vv	vx	vv	2
0274	6	vv	vx	bx	2
0274	7	vv	vv	vx	2
0274	8	vx	vv	bv	1
0274	9	vv	vb	vx	2
0274	10	vv	vv	vv	2
0274	11	vv	vx	bb	2
0274	12	vv	vb	xx	1
0275	1	vv	-1	vv	2
0275	2	vv	vx	xx	2
0275	3	vv	vx	bx	2
0275	4	vb	vx	bx	2
0275	5	vv	vx	vx	2
0275	6	vb	vx	bx	2
0275	7	vv	xx	xx	2
0275	8	vv	vv	vv	2
0275	9	vv	vv	bv	2
0275	10	vv	vx	xx	2
0275	11	vx	vb	vx	2
0275	12	vv	vx	vx	2
0276	1	1	1	0	2
0276	2	1	1	1	1
0276	3	1	0	2	0
0276	4	2	1	2	2
0276	5	1	0	1	1
0276	6	2	1	2	1
0276	7	2	2	2	1
0276	8	2	2	0	1
0276	9	1	0	0	1
0276	10	1	2	0	0
0276	11	1	2	2	1
0277	1	vv	vv	xx	1
0277	3	bb	vb	bx	0
0277	3	vx	vb	bx	1
0277	4	vx	vb	vx	2
0277	5	vb	vx	bb	2
0277	6	vv	vb	vv	2
0277	7	-1	vv	bx	2
0279	1	vb	vv	vx	1
0279	3	vx	xx	xx	2
0279	4	vv	vb	bx	1
0279	5	vx	-1	vv	1
0279	6	vb	-1	xx	1
0280	1	bb	bb	bx	2
0280	2	-1	vv	vv	2
0280	3	vv	qq	vx	2
0280	4	vb	vx	vv	2
0280	5	vv	vv	vx	2
0280	6	vv	vx	xx	2
0280	7	vx	vv	vx	2

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0280	8	vx	vb	vx	2	0285	6	2	1	2	2
0280	9	vv	vx	bv	2	0285	7	0	2	1	2
0280	10	vb	vb	bv	2	0285	8	2	2	1	2
0280	11	vv	xx	bx	2	0285	9	2	2	1	1
0280	1	vb	bx	bx	2	0285	10	1	2	1	2
0281	2	vv	vb	bb	2	0285	11	vv	xx	bx	1
0281	3	vv	vv	-1	2	0285	12	vx	xx	bv	2
0281	4	vb	-1	bx	2	0285	13	vb	vv	bx	2
0281	5	-1	vx	vv	2	0285	14	vv	-1	bv	1
0281	6	vb	vv	xx	2	0285	15	vv	vv	vv	1
0281	7	vv	vx	vx	2	0285	16	vx	vv	bx	2
0281	8	-1	vv	xx	1	0285	17	vv	vv	bx	1
0281	9	vb	vv	bx	2	0285	18	vv	vb	bv	2
0281	10	-1	vv	bx	2	0285	19	vb	vv	xx	2
0281	11	vv	vv	vx	2	0285	20	vb	vb	xx	-1
0281	12	vv	vv	bx	2	0285	21	vx	xx	bx	2
0282	1	0	0	0	0	0285	22	vb	vv	vx	2
0282	2	0	0	0	0	0285	23	bb	vb	xx	1
0282	3	0	1	1	2	0285	24	vv	vx	bx	2
0282	7	1	1	0	2	0285	25	vv	vb	bb	0
0282	8	0	0	0	0	0285	26	vx	vv	vx	2
0282	9	1	0	1	2	0285	27	vv	vv	xx	2
0282	10	2	1	2	1	0285	28	vv	bb	bx	1
0282	11	0	1	1	2	0285	29	bb	xx	bv	1
0282	12	0	1	0	1	0285	30	vv	vx	bx	2
0283	1	0	2	1	2	0285	31	vb	vb	bx	2
0283	2	0	2	1	0	0285	32	vv	vb	bx	2
0283	3	1	2	0	1	0285	33	vv	vb	xx	2
0283	4	0	2	1	0	0285	34	vb	xx	vx	2
0283	5	2	1	1	1	0286	1	0	0	1	2
0283	6	0	2	1	1	0286	2	0	2	0	1
0283	7	0	0	1	1	0286	3	0	2	1	-1
0284	1	0	2	2	2	0286	4	2	2	0	1
0284	2	0	1	0	2	0286	5	2	2	1	2
0284	3	1	2	1	2	0286	6	1	2	1	1
0284	4	2	2	2	2	0286	7	2	2	0	1
0284	5	0	2	1	2	0286	8	2	2	1	1
0284	6	2	2	0	0	0286	9	1	2	1	1
0284	7	2	0	1	-1	0286	10	0	1	0	2
0285	1	2	2	0	1	0286	11	2	2	1	0
0285	2	0	2	1	2	0286	12	vv	vb	xx	1
0285	3	1	2	1	1	0286	13	-1	vb	bv	2
0285	4	1	0	1	0	0286	14	vv	vb	xx	1
0285	5	0	2	1	2	0286	15	bb	vv	bb	2

Site	ind	7.4	12.19	24.11	24.12
0286	16	vv	vx	bv	2
0286	17	vv	vb	bv	1
0286	18	vx	vv	bv	1
0287	1	1	1	0	2
0287	2	2	2	1	1
0287	3	1	1	2	2
0287	4	2	1	1	1
0287	5	2	2	1	2
0287	6	0	1	1	1
0287	7	2	0	1	0
0287	8	1	2	1	1
0287	9	1	1	1	2
0287	10	1	1	1	2
0288	1	1	1	2	1
0288	2	1	2	2	2
0288	3	1	2	2	2
0288	4	2	2	2	0
0288	5	1	0	2	2
0288	6	1	2	2	1
0288	7	2	2	1	1
0288	8	1	2	2	2
0288	9	1	0	1	2
0289	1	1	2	0	1
0289	2	0	2	1	0
0289	3	1	1	1	2
0289	4	1	1	2	1
0289	5	1	2	1	2
0289	6	1	1	0	1
0289	7	2	2	1	0
0289	8	0	1	1	2
0289	9	1	2	2	2
0290	1	1	2	2	1
0290	2	0	1	2	1
0290	3	1	0	1	1
0290	4	0	0	0	0
0290	5	0	1	0	2
0290	6	2	2	2	1
0290	7	0	1	1	2
0290	8	2	2	0	1
0290	9	1	1	1	2
0291	1	2	1	1	1
0291	2	1	2	1	1
0291	3	2	1	1	2
0291	4	2	2	1	2
0291	5	2	1	1	2
0291	6	1	2	2	2
0291	7	1	1	1	2
0291	8	2	1	1	1
0291	9	1	2	1	0
0292	1	1	0	2	2
0292	2	2	1	1	1
0292	3	1	1	2	1
0292	5	2	1	1	2
0292	6	0	0	0	0
0292	7	1	2	1	2
0292	8	2	1	0	2
0292	9	2	1	0	1
0292	10	0	0	0	0
0292	11	2	0	1	1
0292	12	2	1	2	1
0292	13	0	0	0	0
0292	14	0	0	0	0
0292	15	0	0	0	0
0293	1	0	0	0	0
0293	2	0	0	0	0
0293	3	0	0	0	0
0293	4	1	0	0	0
0293	5	0	0	0	0
0293	7	0	0	0	0
0293	8	2	0	0	0
0293	9	1	1	0	0
0293	10	0	0	0	0
0293	11	0	0	0	0
0293	12	0	0	0	0
0293	13	0	0	0	0
0293	14	0	0	0	1
0293	15	1	0	0	0
0294	1	1	1	1	1
0294	2	0	0	1	0
0294	3	1	1	2	1
0294	4	2	1	2	1
0294	5	1	2	2	2
0295	1	0	1	1	1
0295	2	0	1	1	2
0295	3	0	1	0	1
0295	4	2	1	2	2
0295	5	1	2	2	2
0295	6	1	1	2	1

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0295	7	1	2	2	1	0298	15	0	2	1	1
0295	8	2	2	0	2	0299	1	2	0	1	2
0295	9	1	2	2	2	0299	2	2	0	1	1
0295	10	2	2	2	1	0299	3	1	1	1	1
0295	11	0	2	2	2	0299	4	1	1	1	2
0295	12	0	2	2	2	0299	5	1	1	2	0
0296	1	1	0	2	2	0299	6	1	1	0	0
0296	2	0	1	2	2	0299	7	2	2	0	2
0296	3	0	1	2	1	0299	8	1	1	1	1
0296	5	0	2	2	2	0299	9	2	1	1	2
0296	6	1	2	1	1	0299	10	0	2	0	0
0296	7	1	0	1	2	0299	11	2	2	1	0
0296	8	1	2	2	2	0299	12	2	1	1	0
0296	9	1	2	2	1	0299	13	2	0	2	0
0296	10	2	2	2	2	0300	1	2	1	2	2
0296	11	2	0	2	1	0300	2	1	1	1	0
0296	12	2	1	2	1	0300	3	1	1	2	0
0297	1	2	1	0	2	0300	4	1	1	2	1
0297	2	2	1	2	2	0300	5	2	2	2	2
0297	3	0	0	2	2	0300	6	2	0	1	1
0297	4	0	0	2	2	0300	7	1	1	1	1
0297	5	0	1	1	1	0300	8	2	1	2	1
0297	6	2	2	2	1	0300	9	2	2	2	1
0297	7	2	1	2	2	0300	10	2	1	2	0
0297	8	1	1	2	1	0300	11	0	2	2	2
0297	9	1	2	0	1	0300	12	2	2	2	0
0297	10	0	0	0	1	0300	13	2	1	2	2
0297	11	1	2	2	1	0300	14	1	2	2	2
0297	12	1	2	0	2	0300	15	0	1	2	1
0297	13	2	2	0	1	0301	1	1	2	2	2
0298	1	2	2	2	1	0301	2	1	2	2	2
0298	2	1	2	1	1	0301	3	2	0	1	1
0298	3	2	0	1	1	0301	4	1	1	0	2
0298	4	0	2	2	2	0301	5	2	2	2	1
0298	5	1	2	2	2	0301	6	1	2	2	2
0298	6	1	2	0	1	0301	7	1	0	1	2
0298	7	1	1	1	1	0301	8	1	0	2	2
0298	8	1	1	2	1	0301	9	1	1	1	2
0298	9	2	2	2	1	0301	10	1	1	2	1
0298	10	2	2	0	1	0301	11	1	1	2	2
0298	11	2	1	1	1	0301	12	1	1	2	1
0298	12	2	2	2	2	0301	13	2	2	2	1
0298	13	1	2	1	0	0301	14	2	1	2	2
0298	14	0	1	0	1	0301	15	1	0	2	2

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
0302	1	2	1	2	1	0306	2	vx	vv	vv	1
0302	2	1	2	1	2	0306	3	vb	vv	vv	2
0302	3	2	1	2	2	0306	4	vv	bb	bv	2
0302	4	1	2	0	1	0306	5	vv	vv	bv	2
0302	5	2	1	1	1	0306	6	vx	vb	bx	1
0302	6	1	2	2	0	0306	7	vv	vv	vv	2
0302	7	1	2	1	0	0306	8	vb	vb	bb	1
0302	8	1	2	1	1	0306	9	vb	xx	bx	2
0302	9	1	1	1	1	0306	10	vv	vx	bb	2
0302	10	1	2	1	0	0306	11	vv	vx	vv	2
0302	11	2	2	0	2	0306	12	bb	vv	vv	2
0302	12	0	2	1	1	0307	1	bb	-1	bb	2
0302	13	0	1	1	2	0307	2	vx	vv	bb	1
0302	14	2	0	2	0	0307	3	bb	bb	bb	1
0302	15	0	0	2	2	0307	4	bb	xx	bb	1
0303	1	vv	vb	bb	2	0307	5	vb	vb	bv	1
0303	2	vv	vx	bx	1	0307	6	bb	vv	bv	2
0303	3	vx	vv	vx	1	0307	7	vb	xx	xx	1
0303	4	vv	vb	bx	2	0307	8	vx	xx	bb	2
0303	5	bb	vx	vv	2	1085	1	-1	bb	vx	bb
0303	6	vx	vb	bb	2	1085	2	bb	bb	-1	bb
0303	7	vv	bx	bx	2	1085	3	vb	vb	vx	vb
0303	8	vx	vv	bb	2	1085	4	bb	bb	bx	bb
0303	9	vv	-1	bv	2	1085	5	vb	bb	bb	vb
0304	1	vv	vb	bx	2	1085	6	vb	vb	bb	vb
0304	2	vb	-1	bx	2	1085	7	bb	vb	bb	bb
0304	3	vx	bx	vv	1	1085	8	vb	bb	-1	vb
0304	4	vv	bb	vx	1	1085	9	bb	-1	bb	vb
0304	5	vb	vb	bv	2	1085	10	vb	bb	bb	bb
0304	6	vv	bb	bb	0	1085	11	vv	vv	bb	vv
0304	7	vv	bb	bv	2	1085	12	bb	bb	bb	bb
0304	8	vv	-1	xx	1	1258	1	vb	-1	vb	vv
0304	9	vv	bb	bx	2	1258	2	vb	vb	bx	vd
0304	10	vb	vx	bx	1	1258	3	vb	vb	bx	vv
0305	1	xx	xx	xx	1	1258	4	vv	vb	vx	vf
0305	2	vb	bb	bv	1	1258	5	vv	vb	vb	vb
0305	3	vb	vb	bv	1	1258	6	vb	vb	vx	vb
0305	4	vv	vb	xx	1	1258	7	vb	vb	bx	bf
0305	5	vv	bb	vv	2	1258	8	vx	vx	bx	vv
0305	6	vx	bb	vv	2	1258	9	bb	vb	bx	vb
0305	7	vb	xx	bv	1	1258	11	vv	vv	bb	vv
0305	8	vv	xx	bv	2	1258	12	vb	bb	bb	bb
0305	9	vb	vx	bv	2	1258	13	vx	vb	vx	bb
0306	1	vv	vv	vx	2	1271	1	vv	vb	vx	vf

Site	ind	7.4	12.19	24.11	24.12
1271	2	vv	vv	-1	vf
1271	3	vb	vv	vb	vv
1271	4	vv	vv	vx	vv
1282	1	bb	-1	bb	vb
1282	2	bb	vb	bb	vb
1282	3	bb	vb	bx	vb
1282	4	bb	vb	bb	bb
1282	5	vb	vb	vx	vb
1282	6	vb	vb	bx	vb
1282	7	bb	vb	xx	vb
1282	8	bb	vb	-1	vb
1282	9	bx	vb	vb	vb
1282	10	bb	vb	bx	vb
1282	11	vb	bb	vv	vb
1282	12	bb	bb	bb	vb
1290	1	vv	vb	vb	vb
1290	2	bb	bb	vx	dd
1290	3	vb	bb	bb	bb
1290	4	vv	vv	vx	vv
1290	5	vb	bb	bb	bd
1290	6	vv	vb	-1	vd
1290	7	vb	bb	bb	vd
1290	8	vv	vb	-1	vd
1290	9	bb	bb	bb	bd
1290	10	vv	bb	vb	vb
1290	11	bx	-1	vb	vb
1290	12	vb	bb	vx	vv
1315	1	bb	vb	bb	vb
1315	2	bx	vv	bb	bb
1315	3	-1	vv	bx	vb
1315	4	xx	vv	bb	bb
1315	5	vb	vv	vx	vv
1315	6	vx	vv	bx	bb
1315	7	vv	vv	xx	vv
1315	8	vv	vv	-1	vv
1315	9	vv	vv	vb	vv
1315	10	bb	vv	bx	vb
1317	1	vx	vv	bx	vv
1317	2	vx	vv	vx	vb
1317	3	vv	vv	vv	vv
1317	4	bb	vv	vx	bb
1317	5	vb	vb	vx	vb
1317	6	vb	vv	vx	vv
1317	7	vx	vb	vx	vd
1317	8	vb	vb	vx	vd
1317	9	vb	bb	vx	bf
1317	10	vb	vb	vx	vv
1317	11	vv	bb	vx	vv
1317	12	vv	vv	vx	vv
1318	1	vb	vb	xx	vv
1318	2	vx	vb	vb	vv
1318	3	vx	-1	vx	vb
1318	4	vb	-1	bx	vb
1318	5	vv	vv	vx	vv
1318	6	vv	vv	vx	vv
1318	7	vb	vb	vx	vv
1318	8	vv	vb	vx	vv
1318	9	vv	vb	vx	vv
1318	10	vb	vb	vb	vv
1318	11	vv	vv	vx	vv
1321	1	vv	vb	bb	vv
1321	2	bb	vx	bb	vv
1321	3	vb	bb	-1	vb
1321	4	bb	vv	bb	vb
1321	5	vb	vb	-1	vb
1321	6	vv	vb	vx	vb
1321	7	bb	vb	xx	vb
1321	8	vv	bb	xx	vb
1321	9	-1	vv	xx	vv
1321	10	bb	vv	xx	bf
1327	1	vb	vv	vx	vv
1327	2	vb	vv	bx	vv
1327	3	vv	vv	xx	vv
1327	4	vv	vv	vx	bb
1327	5	vx	vv	bx	ba
1327	6	vx	vv	xx	aa
1330	1	vv	vv	vb	vb
1330	2	vb	bb	vx	vv
1330	3	vv	vv	bx	vv
1330	4	bb	bb	bx	bb
1330	5	-1	vv	-1	bb
1330	6	vb	xx	vx	bb
1330	7	vv	vv	bx	bb
1330	8	vv	vb	vx	bf
1330	9	vx	vv	-1	bf
1330	10	vx	vb	bb	vv
1330	11	bb	vb	bx	bb
1330	12	bb	vv	bx	vb

Site	ind	7.4	12.19	24.11	24.12	Site	ind	7.4	12.19	24.11	24.12
1333	1	-1	vv	-1	vb	1344	8	vb	vv	-1	vf
1333	2	vv	vb	bb	vv	1345	1	bb	vv	vb	vd
1333	3	bb	vv	bb	vb	1345	2	vb	vv	vb	vv
1333	4	bb	vb	bb	vf	1345	3	vb	vv	vx	vd
1333	5	bb	bb	bb	bb	1345	4	vv	vv	vb	vv
1333	6	bb	bb	bb	bb	1345	5	-1	vv	-1	vb
1333	7	bb	vb	bb	bb	1345	6	vv	vb	-1	vb
1334	1	vb	bb	xx	vv	1345	8	bb	-1	bx	vb
1334	2	vx	vb	vx	vv	1347	1	vb	vv	bx	bb
1334	3	vv	bb	vx	vv	1347	2	vb	vv	bx	bb
1334	4	vx	bb	bx	vb	1372	1	vb	vv	bx	vv
1334	5	vv	bb	xx	vv	1372	2	bb	vb	bx	vv
1334	6	vb	bb	bx	vv	1372	3	vb	bb	vb	vv
1334	7	vv	bb	vx	vv	1372	4	vb	xx	vx	vd
1334	8	vb	bb	bb	vv	1372	5	vv	vv	bx	vv
1334	9	vv	bb	vv	vv	1372	6	vv	vv	vv	vb
1334	10	vv	bb	vv	vf	1372	7	vv	bx	vv	vv
1335	1	bx	bb	bb	aa	1372	8	vv	vb	vx	vd
1335	2	vb	vv	bx	bb	1372	9	vv	vv	xx	vd
1335	3	bx	-1	vb	va	1372	10	vb	vb	bb	vb
1335	4	bx	vb	bb	vb	1374	1	vb	vb	vb	vb
1335	5	vv	vb	bx	vb	1374	2	vb	vv	bb	vb
1335	6	vv	xx	bb	vv	1374	3	vv	vv	xx	vv
1335	7	vb	vv	vb	vb	1374	4	vx	vv	xx	vv
1335	8	vb	vv	bx	vb	1374	5	bb	vb	vx	vv
1335	9	vb	vv	bb	vb	1374	6	vv	vv	xx	vf
1335	10	-1	vv	bx	vb	1374	7	vx	vv	vx	vv
1342	1	bb	vb	xx	vb	1374	8	vb	vb	vx	vb
1342	2	vv	vb	vx	vv	1377	1	vv	vv	bx	bd
1342	3	vv	vv	-1	vv	1377	2	vb	vv	bx	vb
1342	4	vb	vb	vx	vb	1377	3	-1	-1	vx	vv
1342	5	vv	vb	-1	vf	1377	4	xx	-1	vx	vv
1342	6	vb	vv	vx	vv	1377	5	-1	vv	bx	va
1342	7	vb	vb	xx	vb	1377	6	bx	vb	bb	vf
1342	8	vx	vb	vx	vb	1377	7	vx	bb	vx	df
1342	9	vv	vv	vx	vb	1377	8	vb	-1	vb	vb
1342	10	vv	vv	bx	vd	1379	1	bb	vb	vb	bd
1344	1	vv	vb	vx	bb	1379	2	vv	-1	vb	vd
1344	2	vv	vv	-1	vv	1379	3	vv	bb	vv	vv
1344	3	bb	vv	xx	vv	1379	4	vb	vv	xx	vv
1344	4	vv	vb	xx	vf	1379	5	vv	-1	xx	vv
1344	5	vb	vb	bb	bd	1379	7	vb	vb	vv	vd
1344	6	vb	vv	vb	vv	1379	8	vb	vb	bb	vv
1344	7	vv	vv	bx	vv	1379	9	vb	-1	vx	vb

Site	ind	7.4	12.19	24.11	24.12
1380	1	vv	vb	vv	vb
1380	2	vb	bb	vv	vb
1380	3	-1	bb	vx	vv
1380	4	vx	vv	xx	vv
1380	5	vx	vb	vx	vb
1380	6	vb	vv	bx	vv
1385	1	vx	bb	vv	vb
1385	2	vb	bb	vv	vb
1385	3	vv	-1	vv	vv
1385	4	vv	-1	vv	bb
1385	5	bb	vv	vb	vb
1388	1	vx	vv	vv	dd
1388	2	vv	vb	bx	vv
1388	3	vv	vb	vx	vv
1388	4	vb	vv	vx	vb
1388	5	-1	vb	-1	vb
1388	6	-1	-1	-1	vv
1389	1	-1	vb	-1	vv
1389	2	-1	-1	vx	va
1389	3	vx	bb	vb	va
1389	4	vx	bb	-1	vf
1389	5	bb	bb	bx	vb
1389	6	bb	-1	bx	bb
1389	7	vv	bb	xx	vv
1396	1	-1	vb	bb	-1
1396	2	vv	vb	xx	vv
1396	3	vv	vv	xx	vv
1396	4	bb	vv	vx	bf
1396	5	vb	-1	vb	vb
1396	6	vb	vv	-1	vb
1396	7	vb	vb	-1	vb
1396	8	vb	vb	vx	vb
1397	1	vv	vv	xx	va
1397	2	vb	vb	xx	vf
1397	3	vx	bb	vb	vf
1397	4	vb	bb	vb	vv
1397	5	vb	bb	bb	bb
1397	6	vv	bb	bx	vb
1397	7	vv	vb	vv	vv
1397	8	vv	bb	vb	vb

Appendix II

Ecological data for all the sites sampled in 2000 and 2001. The column 'type' refers to how the habitat was assigned in calculating the discriminant function (1 = pond, 2 = puddle, 3 = unassigned), and '*H*' is the habitat score. The last five columns describe the vegetation: '% em' and '% sub' are the percent of the surface area covered by submerged and emergent vegetation respectively; and the 'bk' columns refer to the percentage of bank cover less than 15 cm, between 15 and 50 cm, and more than 50 cm high respectively.

Sites	type	*H*	width	depth	% em	% sub	bk <15	bk 15-50	bk >50
85	1	0.14	40.0	1.7	60	20	0	10	90
200.3	3	0.43	2.7	2	10	5	40	10	30
200.4	3	0.52	3.2	0.7	5	5	30	10	0
200.5	3	0.47	3.0	1.2	10	5	10	20	10
200.6	1	0.42	2.0	1.7	20	20	20	40	40
200.7	1	0.39	4.0	1.5	20	20	30	20	10
200.8	1	0.43	4.0	0.95	20	30	70	20	10
200.9	1	0.54	3.1	0.6	0	20	25	25	0
200.10	1	0.43	4.4	1.2	10	20	40	20	10
244	3	0.63	1.5	0.35	5	0	100	0	0
245	1	0.45	3.0	0.4	50	20	100	0	0
246	1	0.50	1.5	0.4	50	5	20	0	10
247	2	0.71	3.0	0.1	5	0	70	0	30
248	3	0.65	3.6	0.1	10	40	100	0	0
249	3	0.52	1.5	1.2	5	5	80	0	0
250	3	0.69	1.2	0.15	15	5	80	10	0
251	3	0.55	5.2	0.12	30	60	80	10	10
252	3	0.73	1.5	0.05	35	0	100	0	0
253	3	0.61	1.3	0.2	10	80	50	30	0
254	2	0.78	0.4	0.17	5	0	0	90	10
255	3	0.52	1.2	0.2	60	40	15	45	40
256	2	0.56	1.2	0.25	50	5	80	10	10
257	3	0.68	0.9	0.2	15	10	70	30	0
258	1	0.41	8.0	0.45	40	20	80	0	0
259	2	0.85	0.2	0.1	10	0	80	10	0
260	3	0.76	0.6	0.15	5	5	60	20	20
261	2	0.87	0.2	0.1	0	0	20	40	40
262	3	0.68	0.2	0.2	50	5	80	20	0
263	2	0.74	0.6	0.2	5	0	40	0	0
264	2	0.76	0.4	0.2	5	0	80	0	0
265	2	0.90	0.4	0.05	0	0	10	0	0
266	3	0.61	3.6	0.1	40	0	50	0	30
267	3	0.79	0.6	0.1	10	0	30	20	10
268	3	0.39	7.6	0.4	45	40	10	80	10
269	3	0.83	0.4	0.1	5	0	90	10	0
270	3	0.42	8.0	0.3	50	20	70	10	0
271	2	0.82	0.4	0.1	5	5	30	10	0
272	3	0.71	0.3	0.4	5	5	10	80	0
273	2	0.74	0.3	0.1	30	50	90	10	0
274	2	0.76	0.5	0.2	0	0	60	0	0
275	2	0.82	0.3	0.1	10	10	60	30	5
276	3	0.50	0.6	0.4	20	10	70	10	0
277	3	0.77	1.0	0.1	5	0	50	40	0

Sites	type	*H*	width	depth	% em	% sub	bk <15	bk 15-50	bk >50
279	2	0.81	0.5	0.1	5	0	0	10	0
280	3	0.77	0.7	0.1	10	10	10	0	30
281	2	0.85	0.2	0.12	0	5	50	0	0
282	1	0.35	15	0.5	50	10	20	60	20
283	3	0.65	0.5	0.12	50	30	5	55	0
284	2	0.70	0.5	0.4	0	0	5	75	0
285	3	0.68	1	0.3	0	5	0	5	0
286	3	0.60	2.5	0.4	0	5	0	0	2
287	2	0.82	0.5	0.1	0	5	70	0	0
288	3	0.77	0.5	0.15	5	0	20	60	20
289	1	0.52	2.5	0.6	20	0	10	20	60
290	1	0.36	6	0.5	50	60	0	30	70
291	2	0.76	0.5	0.15	5	20	0	45	0
292	3	0.62	5	0.2	0	5	30	50	20
293	1	0.19	70	2	20	40	85	5	5
294	1	0.09	50	2	60	50	95	0	5
295	3	0.64	0.3	0.2	50	30	0	50	50
296	3	0.60	2	0.12	40	30	85	15	0
297	3	0.58	1.5	0.4	5	50	20	50	20
298	2	0.90	0.2	0.05	10	20	10	40	50
299	2	0.82	0.4	0.1	5	0	30	60	0
300	3	0.73	1.5	0.1	5	20	10	90	0
301	2	0.69	1.5	0.2	0	10	60	30	0
302	3	0.70	0.25	0.15	20	80	50	45	0
303	3	0.69	0.5	0.4	0	10	40	30	30
304	3	0.66	1.3	0.3	5	0	50	30	5
305	3	0.37	6.8	2	10	10	40	0	5
306	2	0.86	0.2	0.1	5	0	40	50	10
307	2	0.86	0.2	0.1	5	0	90	0	0
315	3	0.33	1.8	0.25	90	5	5	5	90
316	3	0.35	5.6	0.23	80	5	100	0	0
317	2	0.67	0.4	0.13	1	0	50	0	50
318	3	0.63	0.6	0.26	2	0	90	10	0
319	3	0.41	1.5	0.15	80	0	100	0	0
320	2	0.69	0.4	0.09	2	0	100	0	0
321	3	0.33	4.1	0.28	85	5	100	0	0
323	3	0.56	12	0.2	3	5	100	0	0
324	3								
325	1	0.31	10	0.13	90	10	100	0	0
327	3	0.57	0.26	0.1	50	0	100	0	0
328	3	0.61	1.3	0.24	5	0	100	0	0
329	3	0.59	0.13	0.08	50	0	98	2	0
330	1	0.51	16	0.4	10	5	100	0	0
331	3	0.67	0.8	0.1	0	0	100	0	0
332	3	0.57	5.1	0.26	5	3	100	0	0
333	1	0.21	60	1.2	85	5	20	70	10
334	3	0.45	5	0.23	50	10	100	0	0
335	1	0.33	45	0.35	70	0	40	60	0
336	3	0.52	1.1	0.43	25	30	100	0	0
337	2	0.69	0.87	0.05	4	0	100	0	0
338	3	0.54	1.2	0.19	30	30	40	0	60
339	3	0.64	1.05	0.1	10	0	50	50	0
340	3								
341	3	0.62	5	0.1	0	0	100	0	0
342	3	0.38	2.1	0.27	80	0	90	10	0
343	3	0.57	5.2	0.26	5	3	100	0	0

Sites	type	*H*	width	depth	% em	% sub	bk <15	bk 15-50	bk >50
344	2	0.58	0.25	0.08	50	0	100	0	0
345	1	0.57	2.2	0.12	25	10	100	0	0
346	3	0.63	1.9	0.15	2	2	95	5	0
347	3	0.65	0.8	0.13	2	2	95	5	0
348	3	0.64	1.2	0.09	10	5	60	30	10
349	3								
371	2	0.68	0.25	0.13	5	0	50	50	0
372	3	0.61	0.95	0.36	0	2	100	0	0
373	3	0.64	0.5	0.1	15	10	0	100	0
374	3	0.37	2	0.27	80	10	100	0	0
375	1	0.45	5	0.28	50	5	10	30	60
376	3	0.61	2.4	0.13	10	0	10	20	70
377	1	0.43	4.9	0.24	45	55	2	98	0
378	3	0.58	3.6	0.31	1	10	30	70	0
379	3	0.48	2	0.11	60	0	0	100	0
380	3	0.61	1.4	0.28	2	1	70	30	0
381	2	0.44	0.33	0.13	80	10	1	50	50
382	2	0.73	0.12	0.08	0	0	100	0	0
383	2	0.61	0.6	0.01	60	0	60	40	0
384	3	0.43	2.5		0	0	0	0	0
385	3	0.62	5	0.1	0	1	80	20	0
386	2	0.70	0.25	0.1	0	0	0	10	0
387	2	0.70	0.25	0.1	0	0	0	10	0
388	3	0.58	2.6	0.5	0	0	30	70	0
389	3	0.47	0.9	0.2	60	10	10	60	30
390	3	0.44	2	0.28	60	10	10	60	30
391	2	0.70	0.25	0.1	0	0	0	10	0
392	2	0.67	0.35	0.14	1	1	100	0	0
393	2	0.69	0.25	0.1	5	1	80	20	0
394	3	0.32	2.9	0.8	80	10	10	30	60
395	1	0.46	5.8	1.7	15	20	100	0	0
396	3	0.58	3.4	0.4	0	0	0	0	100
397	3	0.55	1.9	0.35	15	15	10	60	30
398	3	0.60	3.8	0.2	0	0	100	0	0
399	3	0.65	0.25	0.35	0	0	90	5	5
400	2	0.66	0.23	0.08	20	0	10	80	10
401	3	0.64	0.45	0.25	5	0	0	85	15
402	3	0.27	0.55	0.15	100	1	0	0	100
403	1	0.37	3	1.2	60	10	80	20	0

Appendix III

Distribution of genetic parameters across the Apahida study area. Sites with four or more individuals are included. 'N' is the number of animals genotyped, 'p' is the mean frequency of *variegata* alleles, 'x' and 'y' are the coordinates of the site on the 20 by 20 km grid (see Materials and Methods). 'H' refers to the habitat axis score. 'F_{IS}' and 'R' are the maximum likelihood estimates for the heterozygote deficit and standardised linkage disequilibrium respectively, values significantly greater than 0 (ΔLogL >2) are in bold. 'HI' refers to the hybrid index, the values are the number of animals at each site with 0,1...8 *variegata* alleles. The few individuals not scored at four loci were classified in the range 0-8 by interpolation.

Site	N	p	x	y	H	F_{IS}	R	HI								
								0	1	2	3	4	5	6	7	8
0293	14	0.06	16.0	14.0	0.04	0.15	0.04	9	3	2						
1333	7	0.26	15.0	18.0	0.06	0.38	0.25	2	1	1	1		1	1		
1085	12	0.27	4.3	5.5	0.01	0.21	**0.25**	2	3	3	2		1	1		
0246	5	0.29	6.9	9.6	0.43	0.05	**0.57**	1	2			1			1	
0282	9	0.33	11.0	15.0	0.24	0.18	**0.31**	3		1		4		1		
0292	14	0.41	16.0	14.0	0.58	**0.31**	**0.43**	5				2	4	3		
0251	12	0.44	5.4	5.6	0.49	0.15	**0.42**	2	1	2	1	1	1	3	1	
0283	7	0.46	13.0	16.0	0.62	0.04	−0.09			1	2	2	2			
0200.7	6	0.48	7.1	6.6	0.3	0.21	**0.4**	1			2	1		1	1	
1389	7	0.5	16.0	17.0	0.39	0.29	**0.19**			2		2		2		1
0290	9	0.51	12.0	15.0	0.25	0.14	0.16	1			2	2	2	1	1	
0307	8	0.52	13.0	4.3	0.87	0.21	0.09		1		2	1	2	2		
0200.8	9	0.53	7.1	6.6	0.34	**0.48**	0.23	2			1	1	1	3	1	
0299	13	0.53	16.0	15.0	0.83	0.05	−0.14			2		6	3	2		
1335	10	0.54	7.1	6.6	0.22	−0.02	0.27			1	1	3	3	2		
0258	31	0.55	9.7	12.0	0.32	**0.22**	**0.47**	5	1	1	3	4	1	6	9	1
0276	11	0.57	13.0	16.0	0.43	0.04	0.13			1	3	2	1	2	2	
1397	8	0.57	16.0	14.0	0.49	0.04	**0.25**		1			4		1	1	1
0245	7	0.58	6.9	9.9	0.37	0.36	**0.41**		1	2					3	1
0253	8	0.58	4.7	6.8	0.56	0.03	**0.3**			2	1	1		2	1	1
1321	10	0.58	15.0	16.0	0.22	0.29	−0.09				2	3	2	2		1
0289	9	0.58	12.0	15.0	0.45	−0.02	−0.02				2	2	3	1	1	
0249	8	0.59	5.5	7.3	0.45	0.13	**0.44**	1				3	1	1	1	1
1330	12	0.59	15.0	17.0	0.44	**0.34**	0.02		1	1		2	4	1	2	1
0302	15	0.6	14.0	16.0	0.68	0.08	−0.2					8	3	3	1	
0297	13	0.61	19.0	14.0	0.53	**0.29**	0.02		1		1	3	4	1	3	
0287	10	0.61	13.0	16.0	0.82	−0.04	0.02				2	1	4	2	1	
0247	12	0.62	9.9	12.0	0.68	−0.05	**0.4**			1	2	2	2	2	1	2

Site	N	p	x	y	H	F_{IS}	R	HI								
								0	1	2	3	4	5	6	7	8
0286	18	0.63	13.0	15.0	0.55	0.07	−0.11				3	2	6	5	2	
0304	10	0.63	15.0	15.0	0.62	−0.01	−0.1			1			5	3	1	
1385	5	0.64	16.0	17.0	0.58	0.26	−0.03					2	1	1		1
0298	15	0.64	13.0	16.0	0.92	0.00	0.15			1		5	4	1	3	1
0284	7	0.65	12.0	17.0	0.67	**0.55**	0.01				1	2	1	2		1
1334	10	0.65	7.1	6.5	0.37	0.04	0.03				1	2	1	5	1	
1315	10	0.66	7.1	6.6	0.22	0.3	**0.19**			1	1	2	1	1	2	2
1345	7	0.66	15.0	17.0	0.51	−0.01	−0.12				1	1		4	1	
0257	28	0.66	4.5	5.3	0.65	0.12	**0.37**		1	3	3	3	1	9	2	6
0248	10	0.67	9.5	12.0	0.62	−0.01	**0.41**		1			2	1	2	3	1
1396	8	0.67	16.0	14.0	0.53	−0.03	0.46			1		2	1	2	1	1
0295	12	0.69	17.0	15.0	0.6	0.1	0.18			1	1	1	1	4	4	
0300	15	0.69	16.0	15.0	0.71	0.00	0.2				1	4	2	3	4	1
1379	8	0.69	16.0	17.0	0.41	0.02	0.00				1	1		4	1	1
0296	11	0.69	18.0	15.0	0.55	0.04	−0.03					2	4	3	1	1
0305	9	0.69	8.3	7.8	0.27	−0.02	0.08				1	1	1	4	2	
0291	9	0.69	14.0	14.0	0.75	−0.02	−0.09					1	4	2	2	
1377	8	0.7	16.0	17.0	0.34	−0.03	0.33				1	1	1	2	1	2
0277	7	0.7	14.0	16.0	0.76	0.11	0.12			1			2		4	
0301	15	0.71	16.0	16.0	0.66	−0.01	0.00					3	5	1	6	
0200.4	12	0.72	7.1	6.6	0.46	0.01	0.07				1	2	1	4	3	1
1372	10	0.73	10.0	7.9	0.57	0.00	0.16				1	2		1	5	1
0259	10	0.73	9.8	12.0	0.86	−0.02	**0.45**				1	2	1	2	2	2
0285	34	0.74	12.0	16.0	0.65	0.05	0.05			1	1	3	6	9	12	2
0288	9	0.74	14.0	15.0	0.77	0.03	−0.05					1	2	3	3	
0250	12	0.74	5.9	6.2	0.67	−0.01	**0.37**				2			5	2	3
1380	6	0.74	10.0	11.0	0.57	0.00	0.00					1		4		1
1344	8	0.74	14.0	17.0	0.52	0.21	−0.01				1		1	3	2	1
0303	9	0.74	13.0	16.0	0.66	0.13	−0.12						2	5	2	
0256	12	0.75	4.5	5.8	0.5	**0.31**	**0.22**			1	1		1	4	2	3
0244	6	0.75	5.5	7.4	0.59	0.00	0.17					1	1	1	3	
0200.5	8	0.75	7.1	6.6	0.39	−0.01	0.28				1		2	1	3	1
0200.10	5	0.75	7.1	6.6	0.35	0.25	0.32			1					3	1
0260	12	0.76	9.5	12.0	0.75	0.14	**0.4**		1		1		1	2	4	3
1342	10	0.76	15.0	16.0	0.28	0.00	0.02					1	2	2	4	1
0264	12	0.77	10.0	3.6	0.75	0.03	0.13				1	1	2	2	4	2
0200.9	6	0.77	7.1	6.6	0.48	0.00	−0.08						1	3	2	
1374	8	0.77	10.0	7.7	0.27	0.06	**0.5**					2	2			4
0263	10	0.78	7.0	4.6	0.73	0.00	−0.03					1	2	2	4	1
1317	12	0.78	7.0	6.6	0.64	0.16	−0.1					2	1	3	4	2
0279	5	0.78	16.0	16.0	0.81	0.00	0.32						1	2	1	1
0200.3	12	0.79	7.2	6.5	0.34	0.06	0.18			1		1		3	5	2
0306	12	0.79	9.3	7.4	0.87	0.21	0.17				1		2	3	3	3

Site	N	p	x	y	H	F_{IS}	R	HI								
								0	1	2	3	4	5	6	7	8
0252	7	0.8	5.2	5.4	0.72	0.13	0.24				1			1	4	1
1388	6	0.81	16.0	17.0	0.53	0.00	−0.14					1		2	1	2
0200.6	7	0.81	7.1	6.6	0.33	0.00	−0.3							3	4	
1318	11	0.82	7.1	6.6	0.6	0.00	**0.41**					1	1	3	3	3
0262	12	0.83	2.9	5.5	0.65	0.02	−0.14						1	4	5	2
0270	12	0.83	9.4	12.0	0.33	0.00	−0.1						1	4	5	2
0281	12	0.84	10.0	14.0	0.87	0.00	0.06						2	2	4	4
0268	12	0.84	2.8	5.4	0.3	**0.58**	−0.01						3	2	2	5
0280	11	0.86	4.6	9.5	0.76	0.17	**0.39**				1		1		4	5
0271	8	0.86	7.1	6.6	0.82	0.00	0.06							4	1	3
1327	6	0.86	13.0	16.0	0.51	0.14	−0.2							2	2	2
0272	10	0.87	6.5	4.2	0.69	0.00	0.18						1	2	3	4
0274	12	0.88	0.4	4.0	0.75	0.00	0.03							4	3	5
0267	12	0.89	5.6	3.5	0.78	0.06	0.12							4	2	6
0261	7	0.89	5.0	3.9	0.89	0.22	−0.1							1	4	2
0273	6	0.92	0.6	3.9	0.72	0.00	−0.08								4	2
0266	11	0.92	2.2	1.8	0.56	0.00	−0.04								7	4
0275	12	0.93	3.8	0.6	0.82	0.00	0.13							2	3	7
0265	12	0.96	2.1	1.9	0.92	0.00	−0.08								4	8

Appendix IV

The egg genotypes. For *7.4*, *12. 19* and *24.11*, the 'v' and 'x' alles are both *B. variegata*, and the 'b' *B. bombina*. In *24.12* the 'v', 'a', 'd' and 'f' alleles are assigned to B. variegata and 'b' to B. bombina. A '-1' indicates missing data.

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
200.4	1	1	vx	vv	vx	vv
200.4	1	2	vv	vx	vx	vv
200.4	1	3	vx	vx	xx	vv
200.4	1	4	vv	vv	vx	vv
200.4	1	5	vx	vv	xx	vv
200.4	1	6	vx	vb	xx	vv
200.4	1	7	vx	vx	xx	vv
200.4	1	8	vv	vv	xx	vv
200.4	1	9	vv	vb	vx	vv
200.4	1	10	vv	vb	vx	vv
200.4	2	1	bb	vx	vx	vv
200.4	2	2	vb	xx	vx	df
200.4	2	3	bb	xx	xx	vd
200.4	2	4	vb	vv	xx	vd
200.4	2	5	vb	vv	vx	df
200.4	2	6	vb	vx	xx	vf
200.4	2	7	vb	vx	xx	vd
200.4	2	8	vb	vv	xx	vf
200.4	2	9	-1	vv	xx	vv
200.4	2	10	vb	vx	xx	df
200.4	2	11	vb	vv	xx	df
200.4	2	12	vb	xx	vx	df
200.4	2	13	bb	vx	xx	vv
200.4	2	14	bb	vx	vx	vv
200.4	2	15	vb	vx	vx	df
200.4	2	16	vb	vx	xx	vv
200.4	2	17	bb	vx	xx	vf
200.4	2	18	vb	vv	xx	vf
200.4	4	1	vb	vv	xx	vv
200.4	4	2	vb	-1	vx	vv
200.4	4	3	bb	vv	xx	vd
200.4	4	4	vv	xx	vx	vd
200.4	4	6	vv	xx	vx	vd
200.4	4	7	vb	xx	vx	vd
200.4	4	9	vv	vv	bx	vd
200.4	4	10	vv	vv	bx	vb
200.4	4	11	bb	vx	xx	vb
200.4	4	12	vv	vv	bx	vb
200.4	4	13	bb	vv	vx	vd
200.4	4	14	vb	vv	xx	vb
200.4	4	15	vv	vx	xx	vb
200.4	4	16	vv	vv	bx	vd
200.4	5	1	vx	vx	vx	vv
200.4	5	2	vx	vx	vx	vv
200.4	5	3	vb	vv	vb	vv
200.4	5	4	vv	vx	vb	vv
200.4	5	5	bx	xx	bx	vv
200.4	5	6	vx	vx	xx	vv
200.4	5	7	vx	xx	vx	vv
200.4	5	8	xx	xx	xx	vv
200.4	5	9	vb	vv	vx	vv
200.4	5	10	vx	vv	vx	vv
200.4	5	11	vv	xx	bx	vv
200.4	5	12	bx	vv	vx	vv
200.4	5	13	vx	vx	bx	vv
200.4	5	14	vb	xx	bx	vv
200.4	5	15	-1	xx	-1	vv
200.4	6	1	vv	-1	-1	vv
200.4	6	2	vx	vx	-1	vv
200.4	6	3	vx	vx	xx	vv
200.4	6	4	vv	vx	vv	vv
200.4	7	1	vv	vx	bx	vv
200.4	7	2	vb	vv	bb	vv
200.4	7	3	vb	vx	bx	vv
200.4	7	5	vv	vv	bb	vv
200.4	7	6	vv	vx	vb	vv
200.4	7	7	vv	vx	vb	vv
200.4	7	8	vv	vv	bb	vv
200.4	7	9	vv	vx	bx	vv
200.4	7	10	vb	vx	bx	vv
200.4	8	1	vb	vx	bb	vb
200.4	8	2	vx	xx	bx	vb
200.4	8	3	vx	vx	bb	vb
200.4	8	4	vb	xx	bx	vb
200.4	8	5	vb	xx	bx	vv
200.4	9	1	vv	vx	vx	vf
200.4	9	2	vb	vx	vx	vf
200.4	9	3	vv	vx	xx	vv
315	3	1	vb	bb	vv	vb
315	3	2	vv	vb	vx	vv
315	3	3	vv	vb	vx	vv
315	3	4	vb	vx	vx	vv
315	3	5	vb	vb	vx	vv
315	3	6	vv	vx	vv	vv
315	4	1	vv	xx	vx	bb
315	4	3	vv	xx	vx	vb
315	4	4	-1	-1	vx	-1
315	4	6	-1	vv	vx	-1
315	4	7	-1	vv	-1	-1
315	5	1	vv	vv	bx	vf
315	5	2	vv	vv	bx	vv
315	5	3	vv	vv	bb	vv
315	6	1	vv	bb	vx	vb
315	6	2	vx	bb	vx	vb
315	6	3	vx	bb	xx	vb
315	6	4	vv	vv	vx	vv
315	6	5	vv	vv	vx	vb
315	6	6	vv	vb	vx	vb
315	6	7	vx	vb	vv	vb
315	6	8	vx	vb	vv	vv
315	7	1	vv	vb	bx	vv
315	7	2	vv	vb	vx	vv
315	7	3	vv	vb	vx	vv

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
315	7	4	vv	vb	xx	vv
315	7	5	vv	vb	xx	vv
315	8	1	vb	vb	vx	vv
315	8	2	vv	vb	vx	vb
315	8	3	vv	vb	vx	vb
315	8	4	vv	vb	vx	vv
315	8	5	vb	vv	vx	vv
315	8	6	vv	vb	bx	vv
315	8	7	vb	vb	vx	vv
315	8	8	vb	vb	vx	vb
315	8	9	vb	vv	-1	-1
257	1	2	vb	vb	bx	df
257	1	3	vb	vv	bx	vf
257	1	4	vv	vb	xx	vv
257	1	5	vb	vv	bx	vd
257	1	6	vv	vv	xx	vv
257	1	7	vv	vv	xx	vv
257	1	8	vx	vb	bx	vf
257	1	9	vx	vb	xx	vf
257	1	10	vb	vb	bx	vf
257	1	11	vv	vv	xx	vd
257	1	12	vb	vv	xx	-1
257	1	13	vb	vb	bx	vd
257	1	14	vv	vb	bx	vv
257	1	15	vv	vv	xx	vv
257	2	1	vx	vv	vv	vb
257	2	2	vx	vb	vv	bf
257	2	3	vx	vv	vv	bb
257	2	4	vx	bb	vv	bb
257	2	5	vx	vb	vv	bb
257	2	6	vx	vb	vv	vf
257	2	7	vx	vb	vv	vb
257	2	8	vx	vv	vv	bb
257	2	9	vx	vb	vv	bb
257	2	10	vx	vb	vv	bf
257	2	11	vx	bb	vv	vf
257	2	12	vx	vb	vv	bb
257	2	13	vx	vb	vv	bb
257	3	1	vb	-1	vv	vb
257	3	2	vb	vv	vv	vd
257	3	3	vb	vv	vv	bd
257	4	1	vx	vb	bx	vb
257	4	2	vv	vb	bx	vv
257	4	3	vv	vb	xx	vv
257	4	5	vv	-1	xx	vb
257	4	7	vx	vb	xx	vb
257	4	8	vv	vx	bx	vb
257	4	13	vv	vx	bx	vb
257	4	14	vx	vx	bx	vv
257	4	15	vx	vx	xx	vv
257	4	17	vx	vx	bx	vv
257	5	1	vb	bb	bx	vf
257	5	2	vb	bb	vx	vf
257	5	3	vv	bb	xx	vf
257	5	4	vb	vb	xx	vf
257	5	5	vv	vv	vb	vv
257	5	6	vb	vv	vb	vv
257	5	7	vv	bb	vb	vv
257	6	1	vv	bx	vb	vv
257	6	2	vv	xx	vb	bf
257	6	3	vv	xx	bx	vf
257	6	4	vv	bx	bx	vb
257	6	5	vv	xx	vb	vv
257	6	6	vv	xx	vb	vv
257	6	7	vv	bx	bx	vb
257	7	1	vv	vx	vb	vv
257	7	2	vv	vx	vx	vv
257	7	3	vv	vx	vx	vv
257	7	4	vv	vv	bx	vv
257	8	1	vb	vb	vx	vv
257	8	2	bb	vx	vx	vv
257	8	3	bb	vx	vx	vv
257	8	4	bb	vx	bx	vv
257	8	5	bb	bx	vb	vv
257	8	6	vv	vv	vb	vv
257	8	7	vb	vv	bx	vv
257	8	8	vb	vx	bx	vv
257	8	9	vb	vv	vx	vv
257	8	10	vb	vb	vx	vv
257	8	11	vb	-1	xx	vv
257	9	1	bb	vx	bx	vv
257	9	2	vv	vx	vx	vv
257	9	3	vv	vx	vx	vf
257	9	4	vv	vx	vb	vv
257	9	5	vv	vx	vb	vv
257	9	6	vv	vv	xx	vv
257	9	7	vv	vv	xx	vv
257	9	8	vb	vv	bx	vv
257	9	9	vv	vx	vx	vv
257	9	10	vv	vv	xx	vf
257	9	11	vv	vx	vb	vv
257	10	1	xx	vb	xx	vf
257	10	2	xx	vb	bx	vf
257	10	3	xx	vb	bb	vv
257	10	4	vx	vb	bx	vv
257	10	5	vx	-1	bx	vf
257	10	6	xx	vb	bx	vf
257	10	7	vx	vb	bx	vv
257	10	8	vx	vb	bx	vf
257	10	9	vv	vb	xx	vv
257	10	10	vx	vb	bx	vf
257	10	11	vx	vb	bx	vv
257	11	1	vv	vb	vb	vv
257	11	2	vv	vb	bb	vv
257	11	3	vv	vb	bb	vv
257	11	4	vv	vb	vb	vv
257	11	5	vv	vb	vb	vv
257	11	6	vv	vb	vb	vv
257	11	7	vv	vb	vb	vv
257	11	8	vv	vb	vv	vv
257	11	9	vv	vb	vb	vv
257	11	10	vv	vb	vb	vv
257	11	11	vv	vb	vb	vv
257	12	1	vv	vv	vb	vv

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
257	12	2	vv	vv	-1	vv
257	12	3	vv	vv	vb	vv
257	12	4	vv	vv	vx	vv
257	12	5	vb	vv	bx	vv
257	12	6	vv	vv	vv	vv
257	12	7	vv	vv	vb	vv
257	12	8	vb	vv	vb	vv
257	13	1	vb	vv	bx	bb
257	13	2	vb	vx	bx	bb
257	13	3	vb	vx	bx	vb
257	13	4	vb	xx	bx	bb
257	13	5	vv	vv	bx	vb
257	13	6	-1	vx	xx	vb
257	13	7	vb	vx	bb	bb
257	13	9	vb	vx	xx	vb
257	13	10	vv	vx	bb	bb
257	13	11	vb	-1	bb	vb
257	14	1	vv	vx	bx	vb
257	14	2	vv	vx	vb	vb
257	14	3	vv	xx	vb	vv
257	14	4	vv	vv	bx	vv
257	14	5	vv	vx	bx	vv
257	14	6	vv	vv	vb	vv
257	14	7	vv	vx	vb	vb
257	14	8	vv	vv	vb	vb
257	14	9	vv	xx	vb	vv
257	14	10	vv	vx	bx	vv
257	14	11	vv	vx	bx	vv
257	14	12	vv	vx	vb	vb
257	14	13	vv	vx	bx	vb
257	15	1	vx	vb	vv	vd
257	15	2	vx	vb	vv	vv
257	15	3	vx	vb	vv	vv
257	15	4	vv	vb	vv	vd
257	15	5	vx	vb	vv	vd
257	15	6	vx	vv	vv	vv
257	15	7	vv	vb	vv	vv
257	15	8	vx	vb	vv	vv
257	15	9	vx	vv	vv	vd
257	15	10	vv	vb	vv	vv
257	15	11	vx	vv	vv	vd
257	15	12	vv	vv	vv	vd
257	16	1	vb	vv	vb	vv
257	16	2	vb	vv	vb	vb
257	16	3	vb	vb	vb	vv
257	16	4	vb	vb	vb	vb
257	16	5	vb	vb	vb	vb
257	16	6	vb	vb	vb	vv
257	16	7	vb	vv	vb	vb
257	16	8	vb	vb	vb	vb
257	16	9	vb	vv	vb	vv
257	16	10	vb	vb	vb	vb
257	16	11	vb	bb	vb	vb
257	16	12	vb	vb	vb	vb
257	16	13	vb	vv	vb	vb
257	17	7	vv	vx	vb	vb
257	17	8	vv	vx	vb	vb
257	17	9	vv	vx	vb	vb
257	18	1	vv	vb	vx	vb
257	18	2	vv	vb	vx	vb
257	18	3	vv	bb	bb	vb
257	18	4	vb	vv	vx	bb
257	18	5	vv	vv	bx	vb
257	18	6	vv	bb	bx	vb
257	18	7	vv	vb	bx	vb
257	18	8	vv	vb	vx	vb
257	18	9	vb	vb	vb	vb
257	18	10	vv	vb	vx	vb
257	18	11	vv	vb	bb	vb
257	18	12	vb	vb	vb	vv
257	18	13	vv	vv	bx	vb
257	18	14	vv	vb	bx	vb
257	19	1	vx	vx	vx	vf
257	19	2	vv	vx	bb	bf
257	19	3	vx	vv	vb	bf
257	19	4	vv	vx	vb	vb
257	19	5	vv	vv	vb	bf
257	19	6	vx	vv	bx	bb
257	19	8	vv	vx	vx	vf
257	19	9	vv	vv	bx	bb
257	19	10	vv	-1	vx	-1
257	20	1	vv	vv	vb	vf
257	20	2	vx	vb	vx	vf
257	20	3	vx	vb	vb	vv
257	20	4	vx	vb	xx	vf
257	20	5	vv	vb	vb	vv
257	20	7	vx	vb	vx	vv
257	20	8	vx	vv	bx	vv
257	20	9	vv	vv	bx	vf
257	20	10	vv	vb	xx	vf
257	20	11	vx	vv	xx	vv
257	20	12	vx	vv	bx	vf
257	21	1	vv	vx	bx	vf
257	21	2	vx	vv	xx	vf
257	21	3	vx	xx	vx	vb
257	21	4	vx	vx	xx	vb
257	21	5	vv	xx	bx	vb
257	21	6	vv	vv	xx	vb
257	21	7	vv	vx	bx	vb
257	21	8	vv	vx	vb	vv
257	21	9	vv	vx	vx	vv
257	21	10	vx	xx	xx	vf
257	21	11	vx	xx	vx	vb
257	21	12	vv	xx	xx	vf
257	21	13	vv	vv	vb	vv
257	22	1	vb	bb	vb	vb
257	22	2	vb	vb	vb	bb
257	22	3	vb	vb	vb	vb
257	22	4	vb	vv	vb	vb
257	22	5	vb	vb	vb	vv
257	22	6	vb	vv	vb	vb
257	22	7	vb	bb	vb	vb
257	22	8	vb	vb	vb	bb
257	23	1	vv	vx	-1	bb

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
257	23	2	vv	xx	bx	bb
257	23	3	vv	vx	bb	bb
257	23	4	vv	xx	xx	vb
257	23	5	vv	xx	bx	vb
257	23	6	vv	vx	bx	vb
257	23	7	vv	xx	bb	vb
257	23	8	vv	vx	bb	vb
257	23	9	vv	vx	bx	bb
257	23	10	vv	vx	bx	vb
257	23	11	vv	vx	bb	bb
257	23	12	vv	vx	bx	vb
257	23	13	vv	xx	bx	bb
257	23	14	vv	xx	bx	vb
257	23	15	vv	vx	bx	bb
257	23	16	vv	-1	bx	-1
257	24	1	vv	vv	xx	vv
257	24	2	vx	vb	xx	vv
257	24	3	vv	vv	vb	vf
257	24	4	vv	-1	bx	-1
257	24	5	vv	vb	xx	vv
257	24	6	vx	vv	bx	vv
257	24	8	vx	vv	xx	vf
257	24	9	vv	vv	vx	vv
257	24	10	vv	vv	vx	bb
257	25	1	vv	vb	bx	bf
257	25	2	vx	vb	bb	bf
257	25	3	vx	vb	bb	bf
257	25	4	vv	vb	vx	bf
257	26	1	vb	vv	vx	bb
257	26	2	vv	vx	bx	vv
257	26	3	vb	xx	xx	vv
257	26	4	vb	vx	vx	bb
257	26	5	vb	xx	vb	vv
257	26	6	vv	vx	vb	vv
257	26	7	vb	vv	vb	vv
257	26	9	vb	vx	vb	bb
257	26	10	vv	vv	vb	vb
257	26	11	vb	xx	vb	vb
257	26	12	vb	vx	vb	vb
258	3	1	vb	vb	bb	vb
258	3	2	vv	vb	bb	vb
258	3	3	vb	vv	bb	bd
258	3	4	vb	vb	bb	bd
258	3	5	vb	vv	bb	vb
258	3	6	-1	-1	-1	vv
258	3	7	bx	vv	-1	vd
258	3	8	bx	vb	-1	-1
258	3	9	vb	bb	bb	vv
258	4	1	bb	vv	bx	bd
258	4	2	bb	vb	bx	vd
258	4	3	bb	vb	bb	bd
258	4	6	bx	-1	vb	vb
258	4	7	bx	vv	bx	bb
258	4	8	bx	vb	bb	vb
258	4	9	bx	vv	vx	vb
258	4	10	bx	-1	vx	-1
258	5	1	vv	vb	bx	vv

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
258	5	2	vv	vb	bx	vb
258	5	3	vv	vb	bx	bd
258	5	4	vv	bb	xx	vb
258	5	6	vv	vv	vx	vb
258	5	7	vv	vv	vx	vv
258	7	1	bb	bb	vx	bb
258	7	2	-1	bb	vb	-1
258	7	3	vb	bb	vb	vb
258	7	4	bx	vv	vx	vb
258	7	5	bx	vv	bb	bb
258	7	8	bx	-1	vx	vv
258	8	2	bx	vv	bb	vb
258	8	3	vb	vv	bb	vb
290	1	1	vx	vb	bb	vv
290	1	2	vb	vb	bb	vv
290	1	3	bx	vv	bb	vv
290	1	4	vx	vv	bx	vv
290	1	5	vx	vv	vv	vv
290	1	6	vx	vv	vx	vv
290	1	7	vb	-1	vx	-1
290	1	8	vx	-1	bx	vv
290	1	9	vx	vb	bx	vv
290	1	10	vb	vv	bb	vv
290	1	11	bx	vv	bx	vv
290	1	12	vv	vv	bb	vv
290	1	13	bx	vb	bb	vv
290	4	1	vb	vv	-1	-1
290	4	2	vb	xx	bx	vv
290	4	3	vb	xx	vx	va
290	4	4	vb	vv	vb	va
290	4	5	vb	-1	bb	vb
290	4	6	vb	vv	bx	va
290	4	7	vv	-1	-1	-1
290	5	1	vx	xx	bx	bb
290	6	1	vx	vv	vb	vv
290	6	2	vv	vv	vx	vb
290	6	3	vx	vv	bb	vb
290	7	1	bb	xx	vx	vv
290	7	2	vb	vx	vb	-1
290	7	3	vv	xx	bb	vb
290	7	4	vb	xx	bx	vv
290	7	5	vb	-1	vx	-1
290	7	6	bb	-1	-1	-1
290	7	7	vv	-1	-1	-1
290	7	8	bb	-1	vx	vb
290	8	1	vx	vv	vb	vb
290	8	2	vv	vv	vx	vv
290	8	3	vv	vv	vx	vv
290	8	4	vb	-1	vv	-1
290	8	5	vv	vv	vv	vd
290	9	1	vb	vv	vx	vb
290	9	2	vb	vx	vx	vb
290	9	3	vb	-1	vx	vb
290	9	4	-1	-1	vx	vv
290	9	5	-1	-1	vx	vv
290	9	6	vb	vv	vb	-1
290	9	7	vv	vv	vb	vd

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
290	9	8	bb	vv	vv	-1
290	9	9	vv	vv	vv	vd
290	9	10	vb	-1	-1	-1
290	9	11	vb	vv	vb	vd
290	9	12	vb	vv	vb	-1
290	9	13	vv	vb	vx	vd
290	9	14	vv	vb	-1	vb
290	9	15	vb	vv	vx	vd
290	9	16	vv	vb	vx	vb
290	9	17	vb	vb	vx	vb
290	9	18	vb	vb	vv	-1
290	9	19	vb	vv	vv	vd
290	9	20	vb	vv	vv	-1
290	10	1	-1	vb	bx	bb
290	10	2	-1	vv	vx	bb
290	10	3	vb	vv	vx	vb
290	10	4	vb	vb	vv	bb
290	10	5	vb	vb	vv	-1
290	10	6	vb	vv	vx	vv
290	10	7	vv	vv	vv	bb
290	10	8	vb	vv	vb	vd
290	11	1	vv	vv	vx	vb
290	11	2	vv	vv	vx	vb
290	11	3	vb	vv	vx	vb
290	11	4	vv	vv	vb	vb
290	11	5	vv	vv	vx	-1
290	11	6	vv	vb	vx	vd
290	11	7	vb	vb	vx	vd
290	11	8	vb	vb	vx	vv
290	11	9	vb	vb	vx	vv
290	11	10	vv	vv	vx	vv
290	11	11	vv	-1	vx	vb
290	11	12	vv	vv	vx	vb
290	11	13	vb	vv	vb	vb
290	11	14	vb	vv	vx	vb
290	12	1	bb	xx	bb	vb
290	12	2	bb	vx	vb	vb
290	12	3	bb	vx	vb	vb
290	12	4	bb	vv	vb	vb
290	12	5	bb	vv	vb	bb
290	12	6	bb	vx	bb	vb
290	12	7	bb	vx	vb	bb
290	12	8	bb	vv	bb	vd
290	12	9	bb	vx	bb	bd
290	13	1	vv	vb	vb	vd
290	13	2	vv	vb	bx	vv
290	13	3	vv	vv	bb	bd
290	13	4	vv	vv	bx	bb
290	13	5	vv	vv	vb	vv
290	13	6	vv	vv	bx	vd
290	13	7	vv	vv	vx	bd
290	13	8	vv	vb	vv	vb
290	13	9	vv	vv	vv	vd
290	13	10	vv	vv	vv	vv
290	13	11	vv	-1	vb	-1
290	14	1	vb	vb	vx	vb
290	14	2	vb	vb	vx	vd

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
290	14	3	vv	vb	vx	vb
290	14	4	vv	-1	vx	vb
290	14	5	vv	vv	vx	vb
290	14	6	vb	vv	vx	vd
290	14	7	vb	vv	vx	vd
290	14	8	vb	vv	vx	vd
290	14	9	vv	vv	vx	vd
290	15	1	vb	xx	bx	vb
290	15	2	vv	xx	bx	vb
290	15	3	vb	xx	bx	vb
290	15	4	vv	-1	bx	bb
290	15	5	vv	bb	bb	vb
290	15	6	vv	bb	bb	vb
290	15	7	vb	bb	bx	vb
290	15	8	vb	bb	bx	vb
290	15	9	vv	bb	bb	bb
290	15	10	vb	bb	bx	vb
290	15	11	vb	bb	vv	vb
290	15	12	vv	bb	vv	bb
290	15	13	vv	bb	vv	bb
290	15	14	vb	bb	bx	vv
290	16	1	vv	vv	vb	vv
290	16	2	vb	vv	vb	vv
290	16	3	-1	vv	vb	vf
290	16	4	vv	bb	vb	vv
290	16	5	vb	bb	vb	vv
290	16	6	vv	vb	vb	-1
290	16	7	vb	vv	vb	vv
290	16	8	vb	vv	vb	vf
290	16	9	vb	vv	vb	vf
290	16	10	vb	vv	vb	vv
290	16	11	vv	vv	vb	vf
290	17	1	vx	vv	bx	vb
290	17	2	vv	vv	bb	vb
290	17	3	vv	vv	vx	vb
290	17	4	vv	vv	vb	vb
290	17	5	vv	vv	bb	vd
290	17	6	vx	vv	bx	vd
290	17	7	vx	vv	bx	vb
290	17	8	vx	vv	bx	vb
290	17	9	vv	vv	bb	vb
290	17	10	vv	vv	vb	vb
290	17	11	vb	vv	bb	vd
290	17	12	vv	vv	bb	vd
290	17	13	vv	vv	bb	vd
290	18	1	vb	vv	vx	bf
290	18	2	vb	vv	bb	bb
290	18	3	vb	vb	bb	vv
290	18	5	bb	vv	vb	vd
290	18	6	bb	vb	bx	bd
290	18	7	bb	vb	bb	vv
290	18	8	bb	vb	vb	bd
290	18	9	vb	vb	bx	vv
290	18	10	vb	vv	vb	bd
290	19	1	vx	-1	bb	vb
290	19	2	vx	-1	bb	bb
290	19	3	xx	-1	bx	bb

Site	Batch	Egg	*7.4*	*12.19*	*24.11*	*24.12*
290	19	4	vb	vx	bx	bb
290	19	5	vb	xx	bx	bb
290	19	6	vv	xx	bx	bb
290	19	7	vb	xx	bb	bb
290	19	8	vb	xx	bb	vb
290	19	9	bb	xx	bb	vb
290	19	10	bb	xx	bx	vb
290	20	1	vv	vv	vb	bb
290	20	2	vv	vv	bb	bb
290	20	3	vv	vv	bx	vb
290	20	4	vv	vv	vx	vb
290	20	5	vv	vv	bb	bb
290	20	6	vv	vv	vx	bb
290	20	7	vv	vb	bb	vb
290	20	8	vv	vv	vx	vv
290	20	9	vv	vb	vb	bb
290	20	10	vv	vv	bx	bb
290	21	1	vv	vb	vx	bd
290	21	2	vv	vb	vb	vb
290	21	3	vv	vv	vv	vv
290	21	4	vv	vv	-1	-1
290	21	5	vv	vv	bx	vd
290	21	6	vv	vv	vb	vb
290	21	7	vv	vv	vb	vv
290	21	8	vv	vv	vv	vv