Habitat Preference and Selection in a *Bombina* Hybrid Zone

Timothy Sands

PhD

University of Edinburgh

2005

Declaration

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

Abstract

How can diverged populations sharing the same area remain genetically distinct over time when gene flow between them should break down their unique differences? Understanding how reproductive isolation between populations may act to counter gene flow is vital to understanding how speciation may occur in sympatry. Mosaic hybrid zones provide a good opportunity to study the nature of reproductive isolation between sympatric diverged populations in a natural setting. This thesis describes novel data about reproductive isolation between the toad species *Bombina bombina* and *Bombina variegata* in a mosaic hybrid zone at Apahida in north-west Romania.

The hybrid zone at Apahida forms a fine-scaled mosaic, with the genetic composition of subpopulations varying with the aquatic habitat, even over small distances. I conducted a mark-recapture study which showed that adults move between sites at a high rate and move over distances greater than those separating different habitat types. Variation between sites could be maintained with this movement pattern if there is habitat preference. I use mark-recapture data to test for evidence that adult *Bombina* are choosing their sites with a preference that correlates with allele frequency at neutral loci; the conclusions vary depending on the assumptions of the analysis. An adult habitat preference for mating site can result in the continuation of the habitat association in the next generation. I test for evidence of this in the genotypes of the resulting eggs, and find no evidence that habitat preferences create habitat associations.

Morphology also varies between habitats. I examine the distribution and association of quantitative traits across the hybrid zone. This demonstrates that considerable dispersal occurs from sites whose populations are most similar to the pure species to sites of intermediate phenotype. It also provides some evidence that pure species combinations of these traits are favoured over mixed combinations.

These results show that the habitat association of adult *Bombina* may not be as important in preventing introgression as it would first appear. However it also reveals that major changes to the composition of populations occur between egg laying and adulthood, changing the frequency of neutral alleles and generating linkage disequilibria and a deficit of heterozygotes. The details of how this occurs remain unknown but this study does not rule out the possibility that habitat preference and selection are the cause.

These data have implications for sympatric speciation, as these populations show that habitat preference may be ineffective in preventing gene flow. Other mechanisms of reproductive isolation may therefore need to be invoked.

Acknowledgements

I would like to thank my supervisor Nick Barton for his enormous support and encouragement over the course of the PhD. I have learnt a vast amount from him. He has always managed to put a new perspective on any problem.

I am also deeply grateful to Beate Nurnberger of LMU Munich for the endless ways she has helped and encourage me and for her generosity in letting me work in her laboratory and her patience at guiding me through the intricacies of SSCP genotyping. She has also helped immensely in getting field work to work and she has always had great insights into what it is possible to do and what it is wise to attempt in the field.

I would also like to thank Loeske Kruuk, my second supervisor, for her patience in trying to keep me on the straight and narrow and the great help she has been during the writing of this thesis.

Sonja Köehler was a great friend and companion in the field as were Thomas, Lino Ometto, Andy Gardner, Gyongyver Mara, Felix and Suzi. Sonja's willingness to carry on in spite of long days, poor plumbing and bumpy roads should be an inspiration to us all. Many thanks also to her for allowing me to use her egg genotype data. In the first year Ioan Ghira's assistance was invaluable as was his boundless enthusiasm for herpetofauna when the days seemed tough. Gyongy was equally generous with her assisitance and hospitality in the second year and for her dedication and cheerfulness. Thanks also to Andy Gardner, who lugged a ton of equipment across Europe and fraternised with Romanian customs.

In 2002 permission was given to work with *Bombina* by the Romanian Ministry of the Environment. I am thankful to Professor Munteanu for his assistance with this.

In Edinburgh, the Barton lab made a stimulating and entertaining place to work, and I thank everyone for many small and sometimes large pieces of advice. In particular I thank Tim Vines for allowing me access to his data and for telling me what was really going on in *Bombina*. He usually turned out to be right. Thanks also to Kate Byrne, who gave me brilliant advice about lab work and was always willing to take time to look at what I was doing wrong.

Outside of work, Stu, the Dans, Penny, Dave, Andy, Guy and many others who made the time here such a pleasure. Special thanks to Lucy for taking care of me.

Finally, endless thanks to my parents, to whom I am forever indebted for their unceasing support.

This thesis was supported by the Natural Environment Research Council.

Table of Contents

Chapter 4 Quantitative trait variation in the Apahida hybrid zone

Chapter 5 The role of habitat preference in mating site choice and the creation of the offspring genotype distribution

Chapter 6 Conclusions

"I am amazed you are so interested in frogs (sic). To us they are a banality"

-Brother Nicolae, Romanian Orthodox monk, Frata Monastery, Romania

Addendum

In the calculation of the habitat score used in my thesis, ecological variables scored as percentages were incorrectly transformed. An arcsin transformation (arcsin $\sqrt{p}$ where p is a proportion) is used on variables scored as percentages and proportions to improve their normality (Sokal and Rohlf 1995), but in my thesis these variables were transformed just by taking the arcsin of the proportion. This error has been confirmed by Tim Vines, who originally calculated the habitat score. The following analyses attempt to determine the effect this has on the results demonstrated in the thesis.

The effect of the transformation

One assumption of the discriminant function analysis used to quantify the habitat is that all contributory variables are normally distributed. The arcsin transformation is applied to percentage variables (in this case the amounts of submerged and emergent vegetation) to improve their normality (Sokal and Rohlf 1995).

	Untransformed	Incorrect Transformation	Correct transformation
Emergent Vegetation	A^2=6.44, p=0	A^2=6.794, p=0	A^2=5.400, p=0
Submerged Vegetation	A^2=9.274, p=0	A^2=9.800, p=0	A^2=3.626, p=0

Table 1. Tests of normality of the percentage variables. Normality is tested with an Anderson-Darling normality test on untransformed, incorrectly and correctly transformed variables

The arcsin transformation does indeed increase the normality of these variables whereas simply taking the arcsin does not (see Table 1). However, it is notable that the

distribution of none of the variables differs significantly from normality under any of the transformations.

Correctly transforming these variables and entering them into the same habitat discriminant function leads to a reasonably high correlation between the old and new habitat scores at the same sites (r^2=0.70). However if the discriminant function is recalculated using the correctly transformed variables and the resulting discriminant function is standardised to run over the same range of values as the old habitat score (0-1), it can be seen that there is better correspondence between the old and the new habitat scores at each site (r^2=0.79 see Figure1). This difference is marginally significant (homogeneity of correlations, p=0.082).

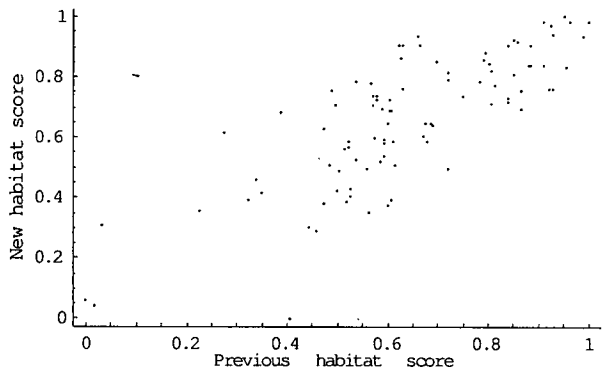


Figure 1. Habitat scores at aquatic sites using the old and new habitat scores

Recalculation of statistics

Shown below are repetitions of some analyses from the thesis that were calculated with respect to the habitat. These are recalculated using the correctly transformed variables and the new habitat discriminant function and are compared with the same results using the previous (incorrect) habitat score.

Firstly the regression of site mean allele frequency on habitat shows a very similar result using either the original or corrected habitat scores (see figure 2). These regressions do not differ significantly in slope ($F_{1,184}$=0.0161, p=0.10) nor in intercept ($F_{1,183}$=0.165, p=0.32).

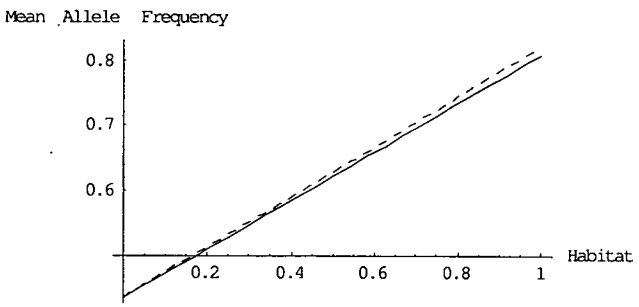
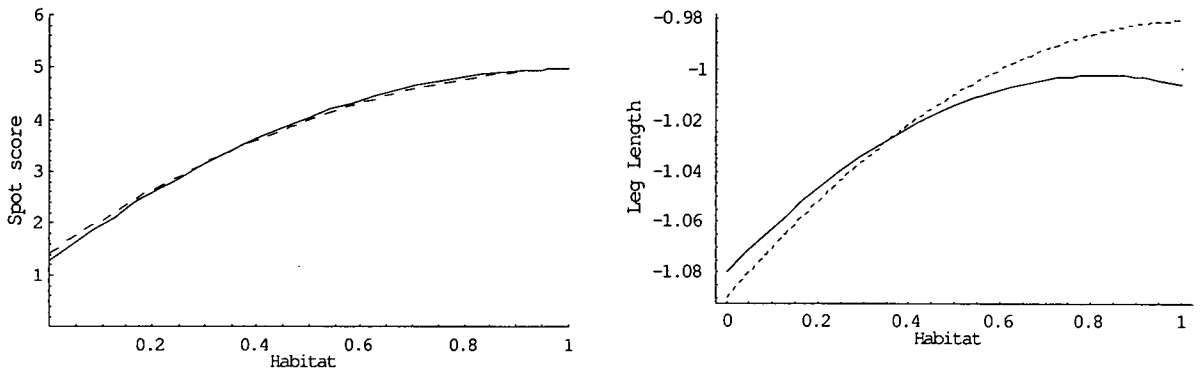


Figure 2. Regressions of the mean allele frequency of sites against their habitat type. The solid line shows the regression using the old habitat scores and the dashed line using the new habitat scores.

Repeating the regressions of the average site spot score phenotype against habitat shows a very similar pattern to that previously observed (figure 2a). The same regression repeated for leg length also shows a similar shape of regression with both habitat scores (figure 2b) but the regressions are slightly more displaced, perhaps reflecting the greater variation inherent in this trait.



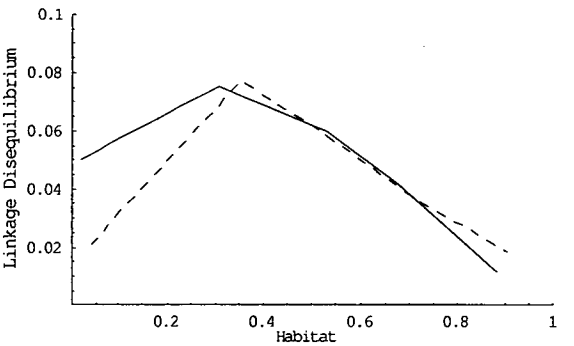
(a) (b)

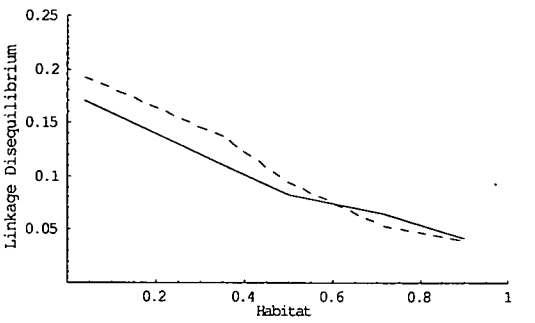
Figure 2. Best fitting quadratic regressions of mean spot score (a) and leg length (b) on site habitat. Solid lines show the relationship with the old habitat scores and the dashed lines with the new habitat scores.

The average linkage disequilibrium between quantitative trait loci with respect to habitat was recalculated (using the methods in chapter 4) for the trait pairs of hybrid index (average proportion of "*variegata*" alleles at marker loci) and spot score, hybrid index and leg length and between leg length and spot score. The results show very similar patterns using both habitat measures (figure 3). The only exceptions to this are found at the habitat extremes, where previously the standard errors were estimated to be very wide and as such are probably not indicative of a significant differences in linkage disequilibrium.

Summary

In summary, the incorrect transformation of the habitat score does significantly change the habitat score at many sites. However, when a habitat function is calculated from properly transformed variables the resulting habitat scores are more similar to those originally used in the thesis. Also when analyses that were calculated with respect to habitat are recalculated with the new habitat scores, the results are generally very similar to the previous results and, where they differ, the general patterns of the results





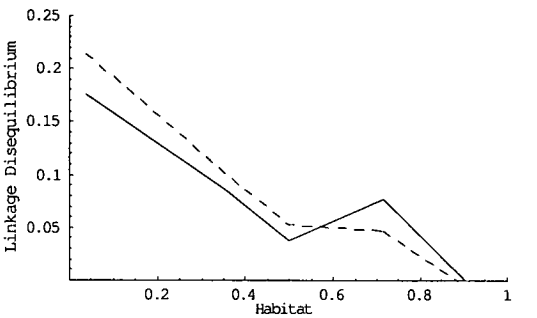


Figure 3. The estimated average linkage disequilibrium between quantitative traits. The patterns are shown between spot score and hybrid index (above left), leg length and hybrid index (above right) and between spot score and leg length (right). The solid lines show the pattern with the original habitat score and the dashed lines with the new habitat score

correspond. As such, I am confident that using the previous habitat function has, in all likelihood, not resulted in any erroneous conclusions being drawn. However, it is clearly desirable to use the corrected habitat scores and it would be necessary to use them in any further publications arising from these results.

Reference

Sokal, R.R. and Rohlf, F.J. 1995. Biometry. 3rd Ed. W. H. Freeman and Co. Ltd., New York, NY.

Chapter 1. Speciation, hybrid zones and *Bombina*

This thesis describes observations and experimental work on the mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *Bombina variegata*. This work aims to quantify mechanisms of reproductive isolation in this hybrid zone, by habitat preference and by selection on several quantitative traits. Such experiments in hybrid zones can give an insight into the processes that occur during the speciation process. In this chapter I introduce the mechanisms underlying the divergence of species. I then describe the characteristics of different types of hybrid zone and how inferences can be made from these about the processes that give an insight into the isolation between the taxa involved. Finally I consider the species *Bombina bombina* and *Bombina variegata* that hybridise in several locations in central Europe and whose hybrid zones have been studied in great detail.

These hybrids are found only in restricted areas where the otherwise parapatric (adjacent and non-overlapping) distributions of these species meet. In this respect such a hybrid zone represents a breach of the most common definition of a species. These "species" can mate and produce viable and fertile offspring. At the same time it is clear that there are restrictions on the success of hybrid offspring as they are only found at the points of meeting of the species ranges, where the opportunity for such mating is readily available. In the particular hybrid zone described here, the picture is even further complicated as the degree of hybridisation appears to vary from one aquatic mating site to the next, even where these are closely spaced.

1.1 Species and Diversity

Life on Earth can be divided and subdivided in many ways but a fundamental unit of its diversity is the species - the diversity of life is measured in the number of extant species. Clearly the species is of great significance and therefore the study of the emergence of

new species forms one of the fundamental problems addressed by evolutionary biology. It is therefore something of a paradox that what constitutes a species is ill-defined.

This is not a new problem. Darwin himself recognised it (while unrepentantly dodging the question): "Nor shall I here discuss the various definitions which have been given of the term species. No one definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species." (Darwin 1859). This problem is ongoing after 145 years despite many attempts at resolution, as this commentary published in June 2004 attests: "There is a crisis in evolutionary biology that is often recognized in theory but almost universally ignored in practice. This crisis can be summarized in two statements. First, our view of how speciation occurs depends on our concept of what species are. Second, biologists disagree about what species are." (Wiens 2004, references removed). Darwin points to the heart of the problem: we all think we know what species are; we just have difficulty providing an adequate definition for all situations.

1.1.2 Defining species

The elusiveness of a definition of species has resulted in a profusion of "species concepts". One estimate puts the number at twenty four (Mayden 1997) and there have been at least two more since this estimate was made (Wu 2001; Wiens 2004). I will only consider the oldest and still the most widely used species concept, the biological species concept (BSC). This was originally described by Dobzhansky (1937) and Mayr (1942, 1963), who defines species as "groups of actually or potentially interbreeding populations, which are reproductively isolated from other such groups." (Mayr 1942).

Under this species definition, speciation is defined by the accumulation of reproductive isolation between the nascent species. This is an important consideration as interbreeding reduces the differentiation between populations. The problem with this

definition is twofold. Firstly it is frequently breached even by what might generally considered to be "good species", either because they are incapable of "potentially interbreeding" as they are asexual or they are not reproductively isolated as they occasionally exchange genes (which is very common in plants and not infrequent in animals). Secondly it relies on a rather vague notion of potentially interbreeding populations. In practical terms this will often be unknowable as allopatric populations cannot be placed together to test whether they can potentially interbreed.

Faced with such a confusing picture it is tempting to abandon species concepts all together (Carson 1985, Cracraft 1989). However there are two reasons why I consider this unnecessary. Firstly to study the process of speciation it is necessary to have at least a working definition of a species and one that can be translated into measurable quantities, as the biological species concept can. Secondly the biological species concept does in a way describe the fundamental change that occurs during speciation: separate members of a population embark on separate evolutionary trajectories that decrease their ability to mate and hence allow the accumulation of further divergence. Therefore any reference to a "species" in this thesis refers to a very loose interpretation of the biological species concept.

1.1.3 Modes of reproductive isolation

The defining feature of species under the biological species concept is the limitation on their ability to interbreed. Other differences between the species may be of evolutionary importance but do not contribute to the continuation of the features that distinguish them. The study of speciation consists largely of determining the mechanisms of reproductive isolation between species of diverging taxa. Such mechanisms can be considered by their time of action, which can profoundly affect the way they evolve. The main division is between those that act before zygote formation or after (pre- and postzygotic isolation), which respectively reduce the rate at which matings occur

between the taxa or the rate at which these mating successfully produce offspring and in preventing the offspring of hybrid matings from themselves successfully breeding. As these modes of isolation are considered in some depth in chapters 4 and 5, I will only briefly outline them here.

Prezygotic Isolation

To prevent the successful production of hybrid offspring, isolation may act either before mating, during mating or after mating. This distinction changes the plausible mechanisms by which prezygotic isolation may occur.

Isolation mechanisms acting before mating can either work by keeping mating populations apart or preventing them from mating where they come together. It might seem most obvious that mating in different locations is the best way of preventing mating between populations, but is not generally considered as prezygotic isolation by definition. The equivalent mechanism in sympatry is the restriction of mating to different habitat types. Restriction to mating in one habitat type can occur through three mechanisms; either each species must exclusively occupy one habitat with the alternative far enough away that dispersal to it is rare, the individuals must make an active choice by some means, to choose one habitat type over the other (generally a behavioural choice in animals) or selection when in the wrong habitat must be very strong.

Just as the diverging populations may be mixed when not breeding but remain distinct by using different locations to breed, if mating occurs at different times for the two populations, either by time of the year or of the day, and thereby live together when not breeding and breed separately (e.g. Miyatake *et al.* 2002).

In circumstances where the populations mate in the same location and at the same time, mating can still be prevented by sexual isolation. There is evidence that sexual isolation is a driving force of speciation, as more species are found in groups of species showing evidence of strong sexual selection (Barraclough *et al.* 1995). Sexual isolation can take the form of assortative mating, whereby one or both species make a choice of mate and this choice is biased towards their conspecfics and indeed any incompatibilities between the mating systems of the populations such as differences in genital morphology can be considered a means of achieving assortative mating. With males frequently contributing only their gametes during mating these incompatibilities provoke a conflict of interest, with females trying to prevent matings with members of the opposite taxa and males trying to defeat the females' defence mechanisms. This can result in a cyclical "arms race" between the sexes. By these sexual conflicts or by the sexual selection underlying mate choice decisions and acting on potentially arbitrary traits, strong isolation can quickly arise between taxa (reviewed by Turelli *et al.* 2001, Parker and Partridge 1998) leading the populations in drastically different directions.

When prezygotic isolation occurs after mating the conflict of interest between parents may be continued, as at this point may make no more material contribution to the mating whereas the female may still have to carry costs of laying and rearing the offspring, it may be better to prevent the male gametes from fertilizing her eggs. This could result in an arms race between the egg and sperm with measures to prevent and ensure successful fertilisation between the species competing. Although the cutoff point between pre- and postzygotic isolation is placed at the point of fertilisation, if the female (and/or male) bears costs after this point then there may be an advantage in preventing the successful survival of hybrid offspring if the resources saved could then be directed towards other offspring from matings with the same species and hence these conflicts of interest can continue after birth.

Post zygotic isolation

Postzygotic isolation acts by reducing the fitness of hybrids relative to individuals of one or other taxon. This form of isolation results from incompatibilities between alleles in the two populations. Incompatibilities that reduce survival and fertility will also reduce the rate of gene flow between the taxa. A special mechanism is required to explain how such an incompatibility can arise. If incompatibilities arise singly then they will be incompatible with the genetic background of their own population as well as with the other. Dobzhansky (1937) and Muller (1942) suggested a mechanism by which these incompatibilities could arise without ever being selected against in their own genetic background. In their model a new allele may arise that is at worst neutral in its own genetic background but is incompatible with an allele at another locus not present in their population. This new allele may then rise to high frequency or fixation by drift or under positive selection. If the allele with which it is incompatible is also neutral or advantageous and arises in an allopatric population, this too may increase in frequency to fixation. When the populations meet and mate the incompatible alleles will segregate together for the first time and the fitness cost will be realised.

Models of this situation make a prediction that the number of such incompatibilities "snowballs" with divergence time (Orr 1995, Orr and Turelli 2001). Consider two allopatric populations in which one such incompatibility has already arisen and in one of the populations another allele that could be involved in a second incompatibility has fixed. There are now two possible loci that any new allele arising in the second population can prove incompatible with. It has been shown that if the fixation of new alleles is linear with time then the number of incompatibilities increases approximately as the square of time (Orr 1995, Orr and Turelli 2001). This makes it more likely that such mechanisms will lead to strong isolation between populations when they meet and mate again.

1.1.4 The geography of speciation

Although distinguishing species by their modes of reproductive isolation explains how species are maintained it does not explain why they originally speciated. Traditionally evolutionary biology has made a major distinction between the modes of speciation that differ in the geographic arrangement of the populations during divergence, which is critical in determining the means by which speciation can occur. It is easy to see why, as this changes more than any other factor the rate at which migration between populations occurs and therefore the homogenising effects of gene flow between the diverging populations. Such gene flow can rapidly break down species barriers as recombination breaks apart allele combinations that promote reproductive isolation (Felsenstein 1981). Speciation modes are therefore categorised as allopatric, parapatric or sympatric. This is essentially a classification by the amount of migration between populations: none, some or equal rates of migration between the populations (Gavrilets 2004).

Allopatric speciation

This is the condition where two populations are isolated during divergence, perhaps separated by a physical barrier. This has long been considered the most likely mode of speciation (Mayr 1942, 1963) and although other modes are increasingly accepted this remains the case (e.g. Lynch 1989, Turelli *et al.* 2001). The reasoning is that in allopatry the populations are evolving independently of one another and hence are more likely to evolve in ways that cause isolation on secondary contact. This evolution could be due to environmental selection, sexual selection or drift. Although the allopatric speciation model does not explicitly require that neither population is extremely small, speciation via founder effects is now considered improbable on theoretical grounds (Barton and Charlesworth 1984) and by experiments (Rice and Hostert 1993, Mooers *et al.* 1999).

The evidence for divergence in allopatry being the dominant form of speciation, though attractive for the reasons outlined above, is rather speculative. There are many clear examples of speciation on islands and in populations split by glaciation but these mechanisms cannot explain all species divergence. Under other, more common conditions, it is far harder to tell that speciation occurred in sympatry rather than allopatry. The evidence for periods of allopatry is often (literally) written in stone: the evidence can come from fossils or geological evidence. Sympatric speciation events would leave no such evidence. However techniques are becoming available that allow inferences about the geography of speciation to be inferred from molecular evidence (Barraclough *et al.* 1998, Barraclough and Nee 2001) and in the few examples available, these demonstrate that the most recently diverged taxa tend to show evidence of allopatric origins (Barraclough and Vogler 2000).

Sympatric speciation

This refers to the situation where speciation occurs in a situation where the diverging populations are not spatially separated for a significant proportion of their range. There are increasingly clear examples of sympatric speciation. The first well tested example of incipient sympatric speciation was in the apple-maggot fly, *Rhagoletis pomonella* (Bush 1994), in which host races are found to both feed and breed on different plant hosts along with evidence that there is genetic divergence between the races (Feder *et al.* 1988). Similar host races have been found in other phytophagous insects that also breed on their host plants, including the pea aphid *Acyrthosiphon pisum* (Via and Hawthorne 2001), the Goldenrod ball-gall fly, *Eurosta solidaginis* (Abrahamson and Weis 1997, cited by Via 2001) and *Zeiraphera diminiana* (Emelianov *et al.* 2001). There is also evidence that the dramatic species radiation in Cameroon crater-lake cichlids occurred by sympatric speciation (Schliewen *et al.* 1994).

Inspired by these findings there have been an ever increasing number of models demonstrating that sympatric speciation was possible in certain specific conditions (e.g. Kaweki 1997, Higashi *et al.* 1999, Dieckmann and Doebeli 2001, Kondrashov and Kondrashov 2001). The common feature of these models is that they are driven by disruptive selection, either for adaptation to different habitats, resource adaptation or preference for a male secondary sexual character. This fact is likely to make sympatric speciation occur quicker than allopatric speciation if the latter occurs by the accumulation and fixation of neutral incompatibility alleles. Where the selection has been in response to environmental differences there are a number of predictions of the genetic make-up and selective regime in supposedly speciating taxa (Via 2001). These predictions have been only been tested in a few cases but the results look encouraging (e.g. Via and Hawthorne 2001).

There are several hurdles that face nascent species diverging in sympatry. Firstly even low levels of gene flow between the adapting populations will tend to spread new alleles to both populations (Barton 1986). In the apple maggot fly and pea aphid this tendency is compensated for by genetic trade-offs between the fitness traits when in the wrong habitat. A second problem occurs where there are multiple loci controlling assortative mating (mating habitat preference in the examples above) and adaptation. If any gene flow occurs between the diverging populations then recombination will tend to break apart the adapted alleles and those promoting the assortment of the populations to which adaptation is conferred (Felsenstein 1981). In the pea aphid this problem is ameliorated by close linkage between the assortative mating and host adaptation loci (Hawthorne and Via 2001).

However, the assortative mating and adaptation loci need not be closely spaced. Navarro and Barton (2002) showed that an inversion promotes the accumulation of postzygotic isolation allele by reducing recombination over a large stretch of the genome. Presumably an inversion could similarly allow the accumulation of alleles causing adaptation to specific habitats and assortative mating of individuals carrying these by

preventing the alleles that control them from being broken apart by recombination. Another mechanism that instantly generates reproductive isolation is the formation of polyploid lineages. Polyploidisation can have the effect of preventing successful mating between newly polyploid individuals and their progenitors. This creates instant reproductive isolation, and such populations are then able to accumulate post-zygotic isolation alleles or sets of adapted loci in sympatry (Soltis and Soltis 1999). This may be a major cause of sympatric speciation, particularly in plants where, for example, around 70% of angiosperm species have undergone polyploidisation at some point (Masterson 1994).

Parapatric speciation

Parapatric speciation is defined the process that creates new species where divergence occurs in only part of a continuous species range and hence the diverged taxa remain in contact with one another. This is postulated to occur as adaptations arise in response to local environmental differences across a broad species range or if incompatibility alleles arise in one part of the species range. Although there will be continual gene flow within continuous population ranges, if selection is strong, there can be local divergence (Turelli *et al.* 2001). If there is a broad environmental gradient across the species range, a cline could be generated across the species range (Slatkin 1973, Endler 1977). Reproductive isolation may be able to evolve via a Dobzhansky-Muller mechanism if the incompatibilities that arise in one part of the range are deleterious only when in the genetic background in a different part of the range. As the differences in the genomes across the species range are likely to comprise only a small part of the genome, this may be less likely to occur than the equivalent mechanism in allopatry.

1.2 Hybrid zones

When populations of diverged taxa are in contact and mate together in nature what is the outcome? This can occur in secondary contact, when the ranges of separated populations change and they come into contact with one another, or at the junction between diverged subpopulations found in a continuous range. The end result of such mating depends on the strength of reproductive isolation between the taxa. If this is weak, then when diverged populations mate they may simply merge. If it is extremely strong then no mating will occur and the populations may either co-exist together in sympatry or, if each species exists to the exclusion of the other, abutting each other in parapatry. A third possibility occurs when the strength of reproductive isolation between taxa lies somewhere between these two extremes. When such taxa meet there will be limited hybridisation between them. The resulting population forms a hybrid zone. Harrison (1993) defines hybrid zones thus: "Hybrid zones occur when genetically distinct groups of individuals meet and mate, resulting in at least some offspring of mixed ancestry". This seems a very broad definition but it in fact encompasses all of the population structures referred to as hybrid zones without reference to how they formed or what form the interactions between populations take.

1.2.1 Hybrid zones and the study of speciation

One of the great difficulties in studying the process of speciation is that, although the genetic differences and the nature of incompatibilities between species may be elucidated in detail (Orr 2001) it is generally impossible to know which of these were important in the early isolation of the taxa. Indeed where species have diverged in allopatry the effectiveness of the incompatibilities between them may never have been tested in the wild.

Hybrid zones offer a model of a key stage of the speciation process: when reproductive isolation is far from complete but gene flow does not homogenise the diverged populations. As such they may offer a unique opportunity to examine the make-up of the differences between the two taxa and how reproductive isolation is maintained in the face of gene flow. In effect the hybrid zone acts as large-scale cross-breeding experiment in the species' natural environment. The effectiveness of prezygotic isolation can be measured from the rates of hybrid mating and the strength of postzygotic isolation can be measured from the fitness of hybrid individuals. Where the hybrid zone has reached an equilibrium between dispersal and selection against hybrids it may be possible to infer the strength of selection against F1 hybrids and even the number of genes under selection from the form of the hybrid zone (Barton and Gale 1993).

1.2.2 The origin of hybrid zones

Hybrid zones by definition form when divergent populations are in contact. The most obvious way this could come about is on the secondary contact between populations that have previously diverged in allopatry. However this divergence may equally occur by adaptation to different trait optima in spatially separated parts of a continuous population to spatially varying environmental differences. Both cases may result in clinal changes in average trait values or allele frequencies in one or more traits between the populations. Therefore it is likely to be impossible to determine by which means a hybrid zone observed in nature originally formed (Endler 1977). To determine the origins of the hybrid zone it is often necessary to rely on non-genetic information such as geological or paleontological information; for example hybrid zones are frequently suggested to have formed at the meeting points of separated populations expanding out of their refugia during interglacial periods (e.g. Arntzen 1978, Hewitt 1993). Another form of evidence pointing towards secondary contact is that clines in a number of differing traits or genes occur in the same location and vary over the same distance (Barton and Hewitt 1985). It is highly unlikely that a collection of different characters

will be under sufficiently similar selection pressures to vary in this way, given the possible variety in function and linkage relations between different genes.

The form taken varies greatly from one hybrid zone to the next . This may be due to differences in characteristics inherent to the taxa such as dispersal and mating behaviours or the strength and genetic architecture of traits under selection. There may also be interactions between the environment and different genotypes, in which case the arrangement of environmental variation may also be important. I describe below three major forms taken by hybrid zones and the inferences that may be drawn from these.

1.2.3 Tension zones and ecotones

Where dispersal of parental types into a hybrid zone is balanced by selection against hybrids (i.e. postzygotic incompatibilities) the hybrid zone that results is referred to as a "tension zone" (Key 1968). As the selection may be intrinsic to hybrid genotypes themselves and not dependent on any particular environmental differences, such hybrid zones may be free to move. To reduce the load imposed by selection these hybrid zones have a tendency to reduce their length (as if under tension). The resulting hybrid zones will tend to form narrow clines between larger areas of the parental populations. In a survey of over 100 hybrid zones Barton and Hewitt (1985) concluded that the majority were tension zones. This conclusion was reached by three lines of reasoning. Firstly, it can be shown that selection acts against hybrids *per se*, either by direct measurement, by observation of reduced fertility or by inference from the abnormalities often observed in hybrid individuals. Secondly it may be shown that dispersal is significant in the cline. Again this may be directly observed or inferred from strong linkage disequilibrium in the centre of the cline. The latter indicates that there is immigration of parental genotypes from outside of this central area, although it could also indicate that there is a selective advantage for parental genotype combinations over a range of hybrid genotypes. Thirdly there is the observation that clines in many traits or the allele

frequency of different marker loci vary consistently across the hybrid zone. Although these could all be following an environmental gradient, it is extremely unlikely that these clines would take the same width and position unless this was the result of association of parental genotypes.

An alternative scenario for narrow clines is that the populations are differentially adapted to different environmental conditions. Where there is an environmental discontinuity in the landscape a cline will found over this ecotone. It may prove impossible to distinguish this form of cline from a tension zone as the cline shape may be very similar (Barton and Gale 1993, Kruuk *et al.* 1999b). Furthermore the presence of an environmental change at the cline location is not conclusive proof that environmental selection underlies the hybrid zone for two reasons. Firstly, there is a tendency for such tension zones to become trapped in one location, particularly where there is a trough in population density or a physical barrier to migration. A good example of this is in the hybrid zone in the grasshopper *Podisma pedestris*, which is trapped by a region of low population density (Barton and Hewitt 1981, Nichols and Hewitt 1986) not related to the ecological habitat (Nichols and Hewitt 1988). Secondly there are quite likely to be slight differences in the environmental adaptation of the parental types, even if intrinsic selection against hybrids is predominant. In this case the tension zone may move until the fitness of the parental types becomes equal, which is likely to occur over an ecotone.

There are also certain rare cases in which different forms of selection are found to act. For example in several hybrid zones in *Heliconius* butterflies the selective mechanism is the frequency-dependent selection on warning colouration (e.g. Mallet 1986, Mallet *et al.* 1990, Jiggins *et al.* 1996). In these cases tension zones are found between different mimicry rings; hybrids whose intermediate warning patterns fall into neither mimicry ring are under strong selection by predation.

Under other hybrid zone models differences in the environmental adaptation of the two parental and the hybrid populations may be the principal factor structuring the hybrid

zone. Two models in which this is the case lead to two differing hybrid zone structures, referred to as bounded hybrid superiority and mosaic hybrid zones respectively.

1.2.4 Bounded Hybrid Superiority

This model, introduced by Moore (1977) supposes that some hybrid genotypes carry an environmental adaptive advantage for the conditions restricted to a specific area between the two parental populations. As hybrids are favoured, unlike in tension zones, such hybrid zones are not dependent of dispersal of parental genotypes into the hybrid zones. However such hybrid zones are necessarily restricted to the uncommon situation in which there are environmental conditions that do favour hybrid types.

There have been several examples described in the literature (Louisiana Iris, Emms and Arnold 1997, Sagebrush, Graham *et al.* 1997, Glaucous-winged and Western Gull, Good *et al.* 2000, Carrion and Hooded Crows, Saino and Villa 1992). This is perhaps less surprising in plants than in animals, where there is a generally higher rate of hybridisation and lower tendency for there to be strong selection against hybrids (Arnold 1996). In fact there are many examples of what appear to hybrid species in plants (e.g. Stebbins 1957, Grant 1966) and some appear to be associated with a novel environmental adaptation (Buerkle *et al.* 2000, Lexer *et al.* 2003).

1.2.5 Mosaic hybrid zones

Often the environmental differences to which hybridising populations are differentially adapted do not vary in a smooth or consistent manner from one to another but rather are distributed in a discontinuous and potentially complex pattern of discrete patches of different environment types. What happens when two hybridising populations are differentially adapted to such environments? The resulting pattern is likely to be determined by a number of factors. Firstly the size of patches relative to dispersal distance, or what might be referred to as the "grain" of the environment and the propensity of each population to move to the alternative habitat determine the potential rate of hybridisation before selection. Secondly the fitness of each parental population and hybrids in the alternative environments will determine the genetic make-up of the adult population within each patch. These factors are likely to vary depending both on the taxa involved and the local environment.

One potential result is that those factors restricting the populations to the habitats in which they are best suited are not strong enough to prevent complete mingling of the populations, collapsing to a "hybrid swarm" consisting of a wide range of hybrid genotype. A second possibility is that there are areas of pure species and the intrinsic barriers to gene flow between them are strong enough to swamp environmental adaptations with a resulting population reminiscent of a tension zone. A third possibility is that the factors fall somewhere in between and there is some degree of association of the parental taxa genotypes with the environment types. Under these circumstances the spatial arrangement of genotypes will be to some degree determined by the arrangement of the environment types. As this will frequently take a patchwork appearance, shadowing the distribution of habitat, this is referred to as a "mosaic hybrid zone" (Harrison and Rand 1989). Although increasing numbers of such hybrid zones are now being identified (e.g. Harrison and Rand 1989, MacCallum 1994, Sites *et al.* 1995, Shoemaker *et al.* 1996, Bridle *et al.* 2001b) it is as yet unclear what proportion of hybrid

zones they make up. Although there are far more examples of clinal hybrid zones in the literature, this is likely due in part to the fact that in tension zones there will be a striking area of hybrids between two distinct populations, which is more likely to be identified by naturalists. As more attention is paid to the fine scale structure of genotypes at the boundaries of parapatric species and in known hybrid zones, it seem likely that more mosaic hybrid zones will be identified.

The maintenance of ecological adaptation is highly dependent on the size of the habitat patch. This is because at the patch edge there may be immigration and extensive hybridisation. Therefore the patches must be larger than about $\sigma/\sqrt{s}$ (Slatkin 1973) where σ is the standard deviation of the per generation dispersal distance distribution and s is the selective advantage within the patch. An exception to this occurs when there is habitat preference that reduces the rate at which mixing occurs in the habitats. A good example of the importance of patch size is given by the ground crickets *Gryllus pennsylvanicus* and *G. firmus*. These taxa are found in a broad hybrid zone (Harrison 1986) but rather than varying in a smooth cline between parental populations the individual species are found to occupy different patches of soil type (Harrison and Rand 1989). It seems likely that there is a selective advantage to this habitat association maintaining this association. At the edges of these patches clines are found between the species (Ross and Harrison 2002) indicative of a balance between migration and selection of some form. Clearly the dispersal distance of the species is much smaller than the patch sizes. As yet there have been no published tests of habitat preference of the parental taxa in these species. The association of genotype and habitat in a number of hybrid zones is discussed further in chapters 3 and 5.

There remains another mechanism that can generate a mosaic hybrid zone without generating any habitat associations. In a metapopulation located between two pure populations and where patches are frequently vacated or the patch populations go extinct; the few long-range immigrants that enter each site will go on to form a significant proportion or indeed the entire site population (Nichols and Hewitt 1994). If

this immigration is guided by habitat preference then the allele frequency in these sites may remain fairly constant with time, as further immigrants will carry similar genotypes, but if habitat preference is absent then the population allele frequency may be essentially random. However if the pure populations range expand into this metapopulation, the mosaic may disappear as migration-selection dynamics take over (Ibrahim *et al.* 1996). Examples of mosaic metapopulations free of habitat association are intrinsically difficult to prove, as it requires an in depth knowledge of the species ecology to determine that there are not any habitat associations.

Mosaic hybrid zones may prove particularly useful as a model of sympatric speciation. Though they are often the result of secondary contact between populations that evolved their habitat adaptations in allopatry, the questions that arise around mosaic hybrid zones are rather similar to those that are asked about diverging sympatric races: How is the divergence maintained in the face of gene flow? How and why do changes occur that increase isolation between the populations (the accumulation of incompatibilities causing postzygotic isolation and the process of reinforcement increasing prezygotic isolation)?

The problem with studying nascent sympatric species is that it is impossible to know whether the diverged races will speciate completely, remain as differentiated races or become homogenised back into a single population. In many cases mosaic hybrid zones show many of the same features of diverged races but reproductive isolation is further advanced. In cases where strong selection against hybrids is absent, other mechanisms of reproductive isolation must prevent the breakdown of a structured mosaic hybrid zone into an unstructured hybrid swarm. By studying mosaic hybrid zones it may be possible to determine how various mechanisms of reproductive isolation work and hence what factors are significant and therefore give indications of how reproductive isolation could result in full speciation between divergent sympatric races.

Studying mosaic hybrid zones is more useful for this purpose than studying clinal hybrid zones as in a cline, the dynamics of the hybrid zone will ensure that immigrants from areas with pure populations will frequently make up a large proportion of the hybrid populations.

1.3 The Bombina hybrid zone

1.3.1 The Genus Bombina

The fire-bellied toads *Bombina bombina* and *Bombina variegata* are amphibia beloning to the family Discoglossidae which consists of 14 extant species in five genera. The genus *Bombina* contains four species which are distributed throughout Eurasia, although only *Bombina bombina* and *Bombina variegata* are found in Europe (Duellmann and Trueb 1994). The Discoglossids are characterised by a disc-shaped tongue (hence the family name) and a medioventral spiracle in the tadpoles, and are generally small in size and have a largely aquatic lifestyle with breeding in aquatic habitats (Duellmann and Trueb 1994). Adult *Bombina* are small and drab looking with a warty, glandular skin that secretes several toxic substances. The ventral skin is generally darkest but possesses a brightly coloured warning pigmentation. Both species are unpalatable, but *Bombina variegata* is particularly so and demonstrates a reflex response when molested, arching the back to expose the brightly coloured palms and throat (Bajger 1980).

1.3.2 Past and present distributions

The most likely explanation for the origin of the European *Bombina* species is that divergence occurred in allopatry as a result of the splitting of an ancestral population during Pliocene glaciation (Szymura 1993). Molecular clock data suggest a divergence

time between 2 and 6.8 million years ago (Szymura 1983, Szymura and Maxson 1984, Szymura 1988). Furthermore paleontological evidence supports this view (molecular and fossil evidence is reviewed by Szymura 1993).

Within the current populations of *Bombina variegata* there is considerable genetic variation which has been explained as a result of the retreat of populations to separate refugia during periods of Pleistocene glaciation (Arntzen 1978). During these glaciations it is suggested that populations of *Bombina variegata* sought refuge in either the southern or north-western Balkans or in Italy. These subpopulations formed what are now sometimes recognised as subspecies of *Bombina variegata* (*B.v.scabus*, *B.v.variegata* and *B.v.pachypus* respectively). *Bombina bombina* probably retreated to a single refugium on the Black Sea coast (Szymura 1993) which may explain the lower regional variation in *Bombina bombina* populations.

The current distribution of *Bombina* (Figure 1.1) reflects both the post-glacial expansion patterns and a tendency for *Bombina variegata* to be found at higher elevations than *Bombina bombina*. The distributions of the taxa are parapatric, with *Bombina variegata* living to the south and west and *Bombina bombina* to the north and east with their distributions meeting in central Europe. *Bombina variegata* occupies most of Italy and the Balkans and much of Western Europe but also the loop of the Carpathian Mountains and various isolated areas of higher ground in Hungary, Poland, Serbia and Romania. *Bombina bombina* occupies a range from the Urals to the Eastern edge of Germany and South to the Black Sea and also the lowland areas of the Danube basin in Hungary, northern Serbia and southern Romania. The isolated areas of *Bombina variegata* found in the lower Danube would seem to suggest that *Bombina variegata* once occupied much more of this area but were displaced by advancing populations of *Bombina bombina*, with small populations remaining, restricted to enclaves of higher ground (Szymura 1993).

At the junctions of these parapatric distributions hybridisation occurs. This is generally restricted to narrow hybrid zones, often found across altitudinal gradients, but hybridisation is also found across more extensive areas in some locations.

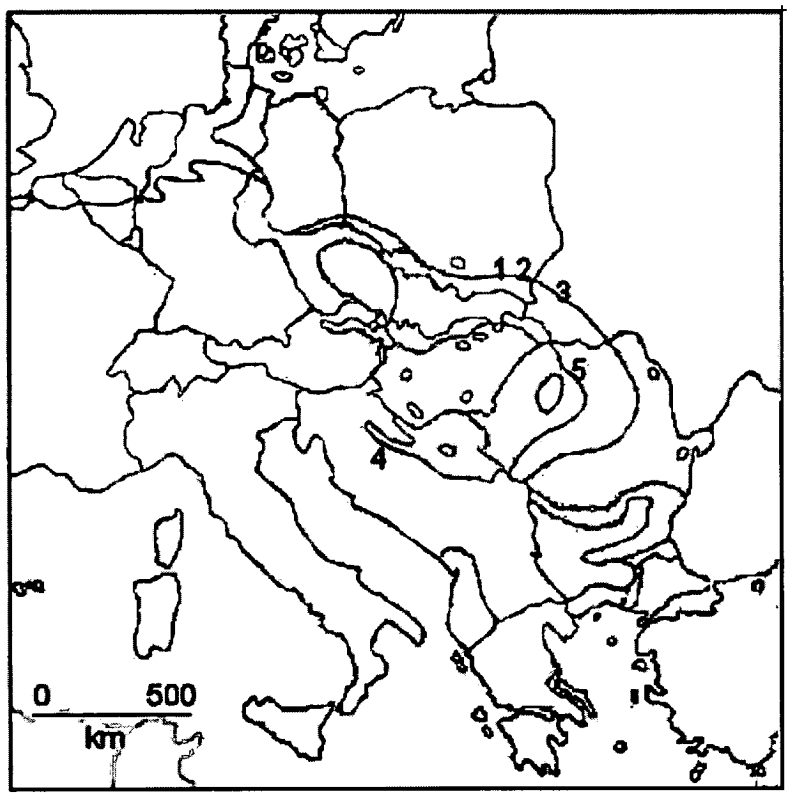


Figure 1.1. The distribution of *Bombina* in Europe. The dark grey region shows *Bombina bombina*, and the lighter grey *Bombina variegata*. The locations of well studied hybrid zones are marked: (1) Cracow (2) Przemysl (3) Stryj (4) Pešćenica (5) Apahida. Map adapted from Arntzen (1978).

1.3.3 Differences between the taxa in morphology and behaviour

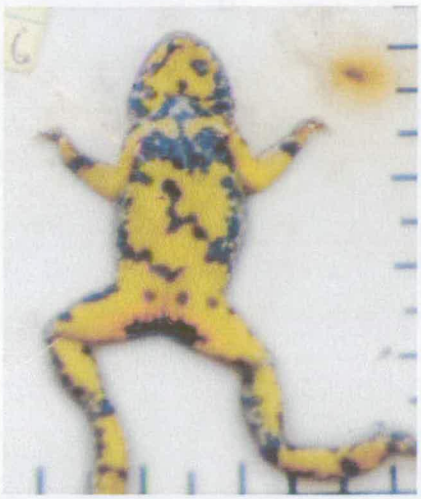
There are a number of consistent differences in morphology and behaviour between *Bombina bombina* and *Bombina variegata*. Many of these differences may be considered to be adaptations to the different lifestyles of the taxa. *Bombina variegata* tends to live on higher ground and occupies small water bodies and that inevitably tend to be temporary. *Bombina bombina* tends to occupy larger and consequently more permanent ponds and is commonly found at lower altitudes.

The best known difference between the taxa is in the ventral colouration (that gives rise to the name fire-bellied toad). In *Bombina bombina* this tends to consist of a number of small red spots on a black background, whereas in *Bombina variegata* there tends to be a greater area of more yellow skin on a paler background (Figure 1.2). Hybrids generally have patterns intermediate between these and this may be used to classify hybrids (Michalowski and Madej 1969, Gollman 1984, see chapter 2), and is highly concordant with allele frequencies of marker loci (Szymura and Barton 1986 & 1991, Nürnberger *et al.* 1995, chapter 4).

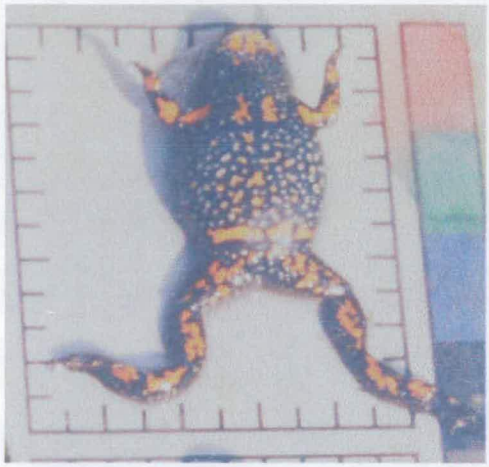
The dorsal skin also differs between the taxa. *Bombina variegata* tend to have rougher and paler skin than *Bombina bombina*. Their skin is also thicker and lungs more vascularised (Czopkowa and Czopek 1955, Nürnberger *et al.* 1995). These may be adaptations to a lifestyle spent more frequently out of the water due to their ephemeral habitat. The skeletal proportions also differ in a manner consistent with *Bombina variegata* possessing longer legs (Michalowksi 1961, Nürnberger *et al.* 1995) which could be also an adaptation to a more terrestrial lifestyle.

The ephemeral nature of puddle habitats must impose a strong selective advantage on individuals whose offspring develop to metamorphosis quickly and therefore more

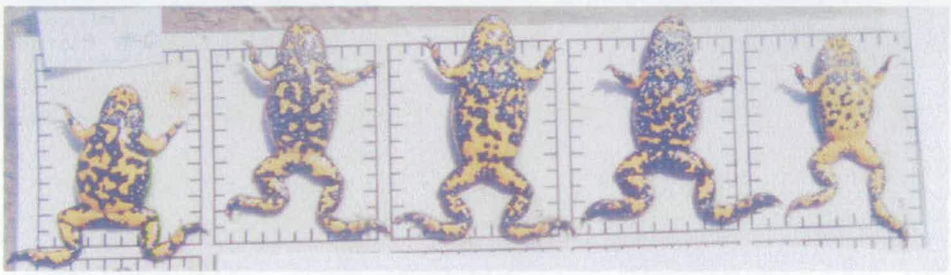
frequently disperse away before the site dries out. This is clearly seen in the morphology and development of *Bombina variegata* eggs, which are both larger when laid and grow and develop to hatching faster than *Bombina bombina* (Rafinska 1991, Nürnberger *et al.* 1995).



Bombina variegata



Bombina bombina



Hybrids of *Bombina bombina* and *Bombina variegata*

Figure 1.2. Typical ventral colouration patterns of *Bombina*

Males of both taxa use a mating call. *Bombina bombina* calls have longer cycle length and pulse duration and a lower fundamental frequency (Lörcher 1969, Sanderson 1994, although only cycle length varies consistently in Pešćenica (Nürnberger *et al.* 1995). The calls of *Bombina bombina* are also louder than those of *Bombina variegata* (Lörcher 1969) as only *Bombina bombina* males posses internal vocal sacs (Günther 1996). This

is perhaps also an adaptation to the habitat of choice. The ponds favoured by *Bombina bombina* are likely to be further apart than the puddles preferred by *Bombina variegata* and the calls of *Bombina bombina* can certainly be heard from a greater distance than those of *Bombina variegata*. The mating systems also differ in other aspects. It appears that *Bombina bombina* males may defend small territories on the water surface (Lörcher 1969), whereas this does not seem to be the case in *Bombina variegata* seems whose mating sites are typically smaller and more densely populated. It is unknown what role the call characteristics play in mate attraction or territory defence. The louder calls of *Bombina bombina* could be an adaptation to territorial mating or to attract females from distant ponds

1.3.4 Genetic differences between the taxa

The genetic divergence between the taxa has been quantified by several studies. Maxson and Szymura (1984) dated the divergence using an albumin clock to 2mya. Another analysis on allozyme divergence gave a divergence of 6mya (Szmura 1983). An allozyme survey (Szymura 1988) revealed significant divergence between the taxa. The difference, expressed as Nei's genetic distance (Nei, 1972) is 0.49. Within the species, there is very little difference between populations of *Bombina bombina*, but subspecies of *Bombina variegata* show considerable divergence; between *Bombina variegata variegata* and the Italian *Bombina variegata pachypus* and Balkan *Bombina variegata* scabus populations there is a genetic distance of 0.31. These populations themselves differ (0.24). Between the *Bombina variegata variegata* populations in the north-west Balkans and the Carpathian populations there is less divergence (0.16). This evidence suggests a divergence between 2 and 7 mya (Szymura 1988).

1.3.5 Transects through hybrid zones

The *Bombina* hybrid zone has been studied within a number of transects. I give details below from four study areas in which in depth studies have been carried out and then

outline some results from other areas. The locations of these transects is indicated in figure 1.1.

Poland

There have been detailed studies of two transects in Southern Poland, at Cracow and Przemysl (Szymura and Barton 1986 & 1991, Sanderson *et al.* 1992), through a hybrid zone that stretches between the Carpathian *Bombina variegata* populations in the South to the northern *Bombina bombina* populations in the north. These studies measured six unlinked allozyme marker loci with alleles diagnostic of the parental *Bombina bombina* and *Bombina variegata* populations (Szymura and Farana 1978), and a number of quantitative traits.

These studies show clear evidence that this hybrid zone follows the tension zone model. Clines in the allele frequency at the marker loci (Szymura and Barton 1986, 1991) coincided with clines in mitochondrial DNA (Szymura 1985) and also in morphological traits (Horbulewicz 1933, Michalowski 1958, Sanderson *et al.* 1992). The coincidence of clines in DNA markers and morphological traits has persisted for over 50 years. The widths of the clines at the two locations are also remarkably similar, estimated at 6.15km and 6.05km at Cracow and Prezmysl respectively. This all suggests that selection in these hybrid zones acts against hybrid individuals as the shape of clines remains the same even in different traits or in different locations. There is direct evidence of this in the form of higher tadpole mortality in hybrids (Koteja 1984, Szymura and Barton 1986) and also a greater rate of morphological abnormalities in tadpoles (Madej 1965, Czaja 1980, Szymura 1993).

Within the breeding sites there was no evidence of deviations from Hardy-Weinberg proportions, suggesting that mating occurred at random, despite there being consistent differences between the pure populations in aspects of their mating calls (Sanderson *et*

al. 1992). There were however strong associations between allozyme loci in sites in the centre of the hybrid zone (D=0.05). These loci are unlinked and the alleles are equally functional (Szymura and Farana 1978). It seems most likely therefore that these linkage disequilibria are generated by dispersal of the parental genotypes into the central sites (with dispersal distances of 0.89 km/gen in Cracow and 0.99km/gen in Przemysl) leading the observed steep step in allele frequencies in the centre of this cline. There are long tails of introgression outside of this step suggesting that slightly introgressed individuals are under weaker selection. The step could be generated by selection against hybrids of 0.58 (0.54-68) relative to the parental taxa, equivalent to the barrier posed to a neutral allele of 51(21-81)km of unimpeded habitat (Szymura and Barton 1991). This also suggests that there are 55(26-88) loci responsible for the loss of fitness in hybrids (Szymura and Barton 1991).

These values are calculated from the linkage disequilibrium observed within tension zones. By approximating dispersal within the hybrid zone by diffusion, the dispersal rate can be estimated from the maximum linkage disequilibrium between genes, the recombination rate between them and the maximal gradients in the allele frequencies of the two loci. Furthermore, assuming selection is against heterozygotes and equivalent at each locus under selection, the selection strength is a function of dispersal distance and width. From the allele frequency gradients at the centre and edges of the cline it is possible to calculate the mean fitness at the centre. Again assuming selection against heterozygotes, it is then possible to estimate the number of genes under selection.

Ukraine

A further hybrid zone has been studied at Stryj in western Ukraine, which has persisted in the same location for 75 years (Yanchukov *et al.* 2003, Yanchukov *et al.* in preparation). Although between the same Carpathian *Bombina variegata* and northern *Bombina bombina* populations as in the Polish transects, a transect through this hybrid zone shows a far narrower width of only 2.3km, which is considered to be a result of

habitat assortment. The cline tail on the *variegata* side is far wider at one point, which seems to be the result of increased migration distances of *B. bombina* due to a river (Yanchukov *et al.* in prep.).

Croatia

A transect in Pešćenica in Croatia, although sharing the same basic cline shape with those in Poland also shows some interesting differences. Firstly the cline is much wider at 9.5km (Szymura 1993). Secondly there is a heterozygote deficit in sites in the centre of the cline (Fis=0.26, MacCallum 1994) and shows much stronger linkage disequilibrium (maximum D=0.139, MacCallum 1994) This appears to result from an association of the habitat type of a mating site and the mean allele frequency of the occupants (MacCallum *et al.* 1998). Mark-release-recapture data seem to show that this is the result of a habitat preference expressed by adults when they move sites (MacCallum 1998 *et al.*). The result of this is that the composition of populations in sites at the centre of the cline are not a random sample from the adult population. This habitat preference also generates stronger linkage disequilibrium by generating a form of assortative mating by the association of more pure genotypes together in the same sites.

Nürnberger *et al.* (1995) performed breeding experiments showing that the clines in several different quantitative traits (the belly colouration pattern, mating call cycle length, the skeletal proportions and skin thickness) are concordant and coincident, as in the Polish transects. However clines in two embryonic traits, egg size and development time were displaced suggesting different patterns of selection on these traits. In order that clines in these traits may be displaced despite strong linkage disequilibria these traits alone must be under significantly different selection pressures, perhaps due to environmental differences in this cline. These experiments also estimated the strength of linkage disequilibrium between the genes underlying these traits, with a result that the maximum linkage disequilibrium is approximately half of that between allozymes

(Nürnberger *et al.* 1995). Kruuk (1997) repeated these analyses on a large number of wild collected adults and eggs, again finding the same concordance and coincidence relationships but linkage disequilibria between quantitative trait loci equivalent to that between allozymes. It seems likely that with a larger sample size that this latter result is more reliable.

Although the habitat clearly plays an important role in structuring the mean genotypes within sites in the centre of the cline, at a larger scale the transect shows the features of a tension zone. Breeding experiments confirm that there is higher mortality in hybrid tadpoles than pure (Kruuk *et al.* 1999a). There is also some evidence that there is differential environmental adaptation in the taxa. *Bombina variegata* tadpoles grow significantly larger and quicker than *Bombina bombina*, giving a significant advantage in their typically short-lived habitats (MacCallum *et al.* 1995). There is no evidence of a similar advantage for *Bombina bombina* tadpoles in their preferred habitat (MacCallum *et al.* 1995), but *Bombina bombina* tadpoles do display do tend to remain motionless in the presence of predators. This is a behaviour which might reduce their rate of predation in their predator-rich preferred habitat (Kruuk and Gilchrist 1997).

Romania

A hybrid zone studied in Apahida, Romania, showed some interesting differences from those in other locations. In at least a 20x20km square extensive hybridisation is seen and there is a mosaic rather than clinal structure (Vines *et al.* 2003). A strong association between the type of aquatic habitat and the mean allele frequency of adults occupying it is seen at four unlinked DNA marker loci (Vines *et al.* 2003). Similarly to the Pešćenica transect there were significant deviations from Hardy-Weinberg proportions (maximum Fis=0.2) and considerable linkage disequilibrium between marker loci (standardised linkage disequilibrium, $D/\sqrt{(pqrs)}$, where p=1-q and r=1- s are the frequencies of the alleles at both loci, is R=0.39 - the equivalent in Pešćenica is R=0.40). This seems to result from migration from the sites with high frequencies of the pure taxa alleles into

more intermediate sites (Vines *et al.* 2003); these sites typically have strongly pond and puddle-like characteristics. The association of a site's habitat and the mean allele frequency of marker loci was much stronger in Apahida (measured as the change in allele frequency per unit of a habitat score, Apahida 0.30, Pešćenica 0.16) apparently showing that habitat preferences are much stronger in Apahida than Pešćenica (Vines *et al.* 2003).

Experiments following the progress of cohorts of tadpoles of the same age have been conducted to test for environmental selection in the "wrong" habitats (Köhler 2003). Surprisingly these seem to suggest that there is little or no selection against *Bombina bombina* or *Bombina variegata* alleles when in the wrong habitat, although this negative result may be mainly due to drawbacks of the experimental method and small sample size (Köhler 2003). Specifically, these experiments have difficulty examining selection in small sites that have to be kept artificially wet for the experiment to run long enough to gather sufficient data. Also these data do not test for selection in the habitats favoured by *Bombina bombina* nor *Bombina variegata*, it is feasible that the most significant selection occurs in such sites. Breeding experiments have shown no evidence of higher rates of mortality in hybrid tadpoles than "pure" (Köhler 2003) in contrast to results from similar experiments on populations from both the Polish and Croatian hybrid zones (Koteja 1983, Kruuk 1999a).

Other hybrid populations

Several other populations have been examined in lesser detail than those described above. Most of these show similar characteristics to these. A cline rather similar to that in Pešćenica has been mapped at Kostajnica, also in Western Croatia (Szymura 1993). There are also areas more reminiscent of the mosaic structure described in Apahida, located in the Slovak Karst Plateau (Gollmann 1986, cited by Szymura 1993). Other populations show interesting differences consisting exclusively of hybrids, mostly in

Hardy-Weinberg proportions (Gollman 1984 cited by Szymura 1993). This may be interpreted as the remnants of a hybrid zone after habitat destruction (Szymura 1993). When the original environment of a hybrid zone breaks down, there is a loss of intermediate populations and the remaining sites are populated largely by immigrants from pure populations hence genotypes are found in Hardy-Weinberg proportions.

1.4 Thesis aims

The aim of this thesis is to examine the nature of reproductive isolation in the *Bombina* hybrid zone near Apahida, Romania. Previous work has shown that this hybrid zone has a mosaic structure that extends over a wide area. It is not clear how such a structure can occur, when in other regions where these species' ranges meet narrow clines are found. The answer must lie in the nature of reproductive isolation between these taxa. Either the populations in Apahida are intrinsically different to those elsewhere or differences in the local conditions result in different expression of reproductive isolation. Understanding how and why reproductive isolation differs between these hybrid zones can give an indication of what underlies different modes of speciation in other taxa.

I attempt to make direct and indirect tests of some hypotheses of how reproductive isolation could be acting to maintain this hybrid zone structure. I also compare these results, where possible, with those obtained in transect at Pešćenica which also shows some mosaic features but also clinal variation, to try to determine how and why they differ.

In chapter 2 I describe the field area in detail and then describe various techniques used in the field and laboratory to quantify the genotypic and phenotypic distributions

In chapter 3 I describe the results of a mark-release-recapture experiment that aims to determine if the genotype-habitat association is the result of adult habitat preference and to quantify the population size and various dispersal parameters.

Chapter 4 describes the distribution of some quantitative traits through the hybrid zone with respect to DNA markers and habitat. This is used as an indirect test of selection on these traits.

Chapter 5 concerns the choice of mating site of adults in the hybrid zone and considers habitat-genotype associations in the offspring and whether this can generate the associations seen in adults generally.

In chapter 6 I describe how the results in this thesis help to explain why the structure of the Apahida hybrid zone differs from those between the same species in other areas. I further explain the relevance these results may have for the study of sympatric speciation.

Chapter 2. The Apahida hybrid zone and methods for describing the distribution of hybrids

This chapters describes the field area near to Apahida in Romania which has been identified as containing a mosaic hybrid zone between the toads *Bombina bombina* and *Bombina variegata* and methods used in the description and analysis of the patterns of their distribution in this area. Descriptions are given of the location, topology and habitats within the field area. Methods are described to quantify aquatic habitat type, measurements taken of adults in the field, methods used to quantify habitats and genetic methods used to characterise tissue samples in the laboratory.

2.1 The Apahida hybrid zone

2.1.1 Location and topology

The Apahida hybrid zone is located near the village of Apahida in Județul Cluj in North-West Romania (see figure 2.1). This is contained within the region of the Transylvanian plain, which lies at roughly 250m above sea level. This region is bounded to the north, east and south by the Carpathian mountain chain. To the west of the studied area is the isolated Bihor (or Apuşeni) mountain region.

The field area itself consists of 20km by 20km area of rolling hills and valleys at elevations between about 200m and 500m above sea level. Running through the northern part of the field area is the wide Someş river valley, which joins with the Danube in Northern Hungary. The soils in the area are generally light loam, becoming

lighter and sandy towards the hill tops and heavier clay in the river valleys. Much of the area is under agriculture with the valley sides being used for mixed-crop arable strip agriculture and the valley floors and hilltops for sheep grazing. Most of the aquatic habitats in which *Bombina* are sampled is within agricultural land, although some populations are also found in the vicinity of small villages. Some areas are covered with woodland, which consist mainly of beech, oak and hornbeam. Such woodland does not appear to contain any sites capable of supporting *Bombina* populations.

Figure 2.1. The location of the Apahida field area within Romania. Map from www.lib.utexas.edu/maps/Romania

This region has been surveyed by B. Nürnberger, T. Vines, A. Hofmann and R. Sieglstetter in 1998 and 1999 and I. Ghira in 1999 (both unpublished data). These surveys revealed that hybrids between *Bombina bombina* and *Bombina variegata* are

found across a wide area and their distribution does not follow the expectation of a cline. A more in depth survey within the field area described above revealed that the average number of *Bombina variegata* alleles (of molecular markers that distinguish the species) within aquatic sites follows the habitat type, despite variation of habitat occurring over very small scales relative to the dispersal distance of the species (figure 2.2), and genotype is unrelated to the geographic location of the sites (Vines *et al.* 2003). This structure is characteristic of a mosaic hybrid zone.

It is not clear why this hybrid zone has taken this form rather than a cline but there are several factors that seem likely. Most of the *Bombina* hybrid zones studied thus far have been located over a clear altitudinal gradient with a pure population of one species on either side (e.g. Szymura and Barton 1986, MacCallum *et al.* 1995, Yanchukov *et al.* 2003), which is not the case in Apahida. Also in contrast to these hybrid zones it is unknown where the parent populations that formed the Apahida hybrid zone came from or are now located, there are several areas nearby that are known to possess populations containing only one species. The Bihor mountains, 20km to the west, and the Carpathian mountains 100-150km to the east and northeast contain populations of only *Bombina variegata* (I. Ghira, G. Mara pers. comm.). The nearest populations of *Bombina bombina* are less clear as much of the Transylvanian plain has been only sparsely surveyed. However 100km to the west is the Danube basin which contains pure populations of *Bombina bombina* (Szymura 1993). The broad Someş valley seems to provide a likely route by which these animals could access this field area. It is plausible that the absence of large pure local populations has resulted in the mosaic structure or that the hybrid zone is the result of a hybrid zone that is collapsing after large populations have withdrawn. It is also plausible that a recent change in land use (deforestation or intensive agriculture) has allowed the colonisation of the area by one or both species.

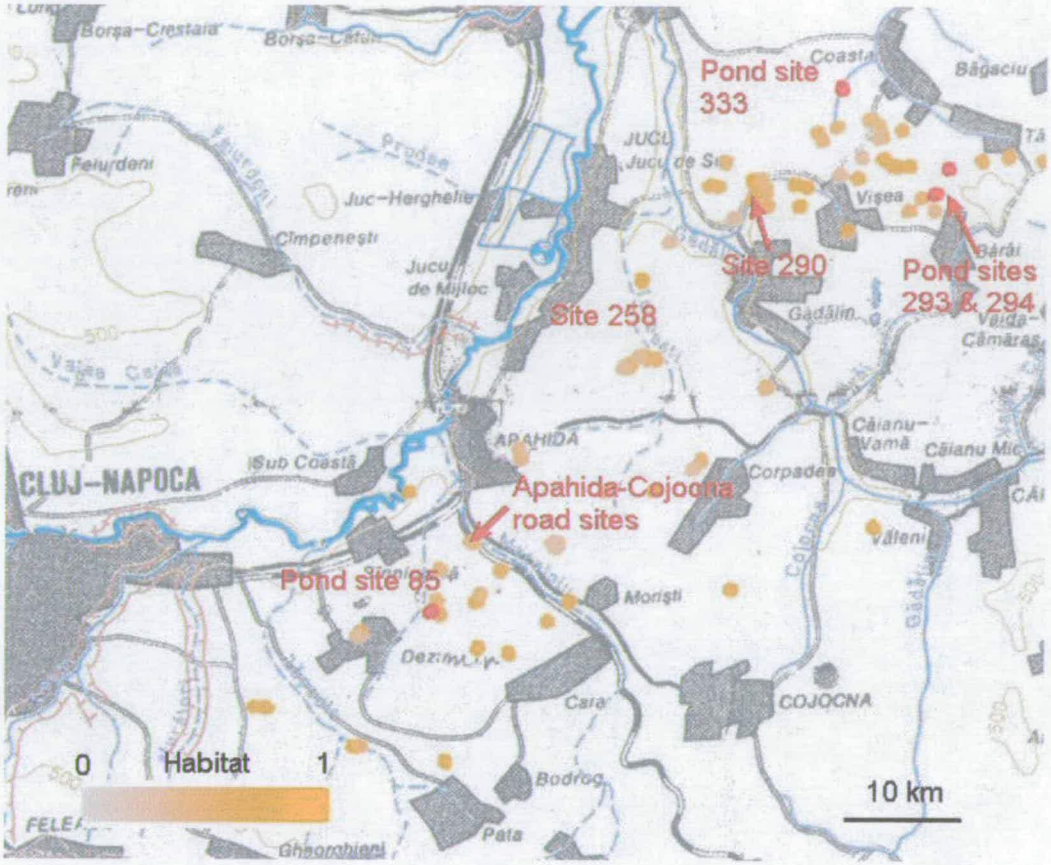


Figure 2.2. The distribution of aquatic habitat in the Apahida hybrid zone. All known occupied sites are indicated. The habitat score is described in section 2.4. Ponds (H<0.25) are marked in red and focal areas are indicated.

2.1.2 The aquatic habitat of Bombina

Toads are sampled only from aquatic habitat. Bombina lay mate and engage in egg-laying exclusively in the water (Duellman and Trueb 1994) although it is likely that they use a wider range of habitat types for other activities. A variety of types of water body were sampled varying in their size, permanence and ecology (see Figure 2.3). Large ponds were infrequent and only four were sampled, all of which were more than 20m

across and probably more than 3m deep (although one was completely dry in the 2002 season) and were heavily vegetated with reeds and various waterweeds. Many sites were found in natural depressions, often near natural springs, and hence varied greatly in size. Vegetation in such sites depended largely on depth and permanence of the water and the soil type. Flooded ditches make up a small number of sites, have habitat use similar to natural depressions. A large number of sites were in shallow flooded depressions, flooded wheelruts and flooded hoofprints. Such sites are typically very shallow and mostly or entirely lacking in vegetation and are typically short-lived, lasting for a few weeks after rain. A final variety of site is found in arable areas; these are the temporary drinking holes dug by farmers for the watering of their working animals. These sites, depending on their age and level of use, may be cloudy and lacking vegetation. A method for quantifying the habitat types is described in section 2.4.

Figure 2.3 Examples of *Bombina* aquatic habitat. (Clockwise from top left): A large pond, ditch, smaller pond, watering hole, flooded natural depression and flooded cattle hoof-prints. These are all sites from which *Bombina* were successfully sampled.

2.2 Field methods

Toads were caught by net from the water surface or by hand. These animals were sedated by brief immersion in a 0.2% solution of MS-222 (3-amino benzoic acid ethyl ester, Sigma), providing approximately 20 minutes of sedation. The end section of a single toe was removed (the left toe in 2000 and 2002 and the right in 2001) to provide a tissue sample and to identify individuals that had been previously sampled. This toe was stored in an Eppendorf tube, preserved in 100% ethanol. Also while sedated toads had their ventral colouration photographed, the "spot score" of this pattern was calculated (See section 2.4), the presence or absence of dorsal spots, warts and nuptial pads noted and measurements were taken of the femur length and snout-anal vent length. Nuptial pads are only present on breeding males. As the majority of males sampled were of mating age this was taken as the indicator of the sex. The toads were then revived and returned to their sites of capture.

Sedation with MS-222 and toe-clipping are common practices in herpetology (Southwood and Henderson 1998) and there is no evidence available to suggest that either has a permanent adverse effect on the toads. The toe clipping provides a convenient way to identify that individuals have been previously caught. Individuals with a clipped toe were only sedated and photographed. Individual identification is possible as the ventral colouration patterns are highly distinctive. Identification was carried out by eye at a later date.

2.3 Quantitative Trait Measurements

The leg length, snout-anal vent length and spot score were measured on all adults. It is known that the leg length and the ventral colouration patters vary between the taxa (Michalowski and Madej 1969, Szymura 1993, Nurnberger *et al.* 1995). Described below are quantitative measures of these traits.

2.3.1 Leg length

The measure of leg length used is of the tibia, not femur as measured in Pešćenica (Nürnberger *et al.* 1995, Kruuk 1997). This is strongly correlated with body length (r^2=0.82), so this is corrected by dividing by the length from the tip of the snout to the anal vent of an animal laying flat on its back to give a measure of leg length relative to body size. The measure used in the analysis is the logarithm of this ratio. I will later refer to this measure as the "leg length". The different leg length measures were used to be compatible with measurements made by other research in the Apahida hybrid zone.

2.3.2. Spot score

Bombina are mostly quite drably coloured except for the on their underside, on which there is a patch of patterned skin featuring very brightly coloured spots and swirls. This is probably an aposematic warning colouration and is displayed by the toads in a typical reflex reaction when disturbed. The patterns vary between the species, with *B. bombina* having a dark belly with a few small red spots, which are generally unconnected, and *B. variegata* a paler belly having a bright yellow colouration, more and larger spots and hence more connectedness between spots. The "spot score" of an individual is a sum of how many of ten sets of spots, characteristically appearing in a similar positions on all toads, are connected by areas of the bright colours (described by Szymura and Barton 1991).

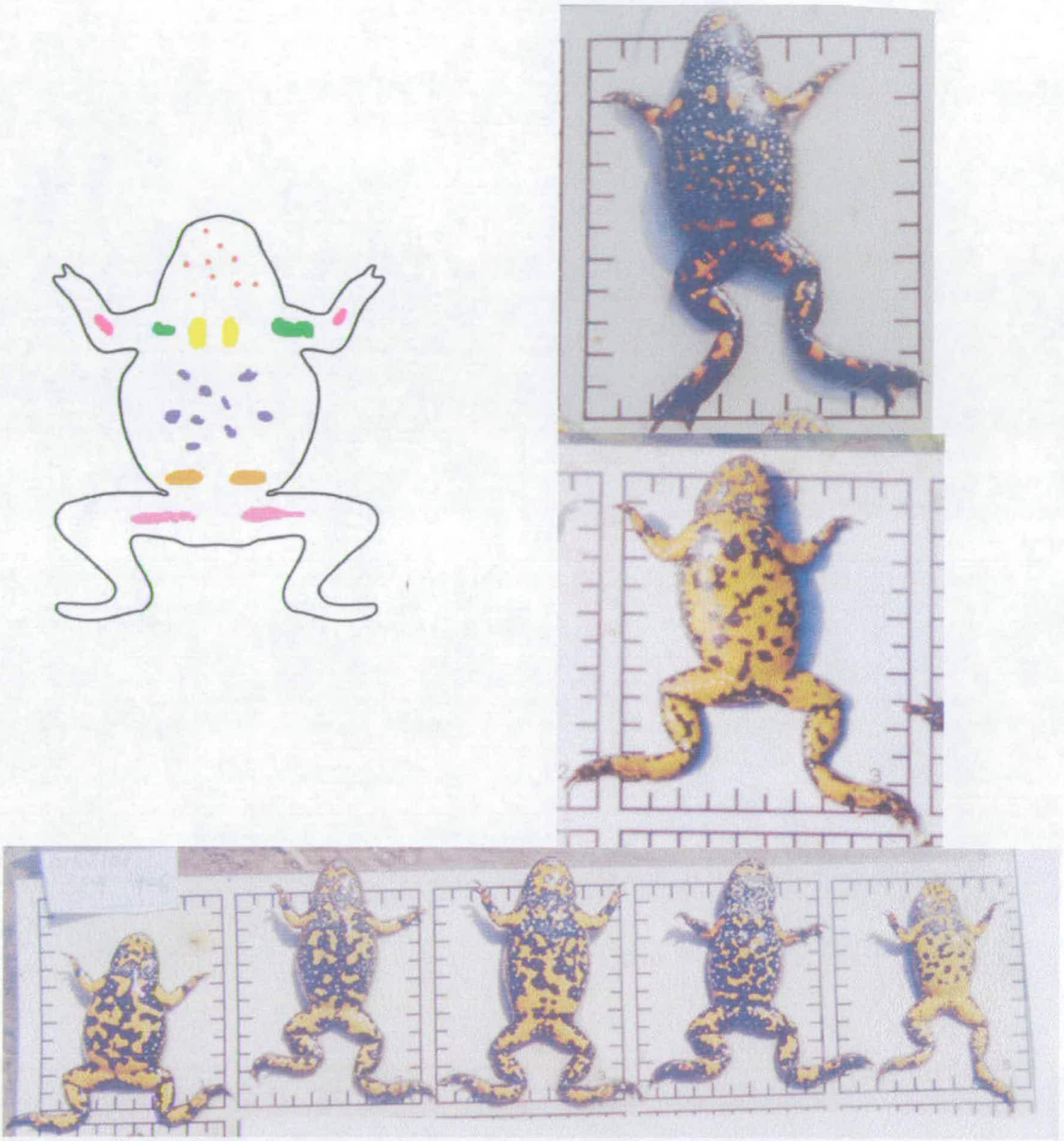


Figure 2.4. Ventral colouration spot score: Connections within and between colour areas (colour coded) sum to form a spot score. Photographs give examples of individuals with low (top right) and high (mid right) spot scores and a range of "hybrid" pattern types (bottom).

2.4 Quantifying aquatic habitat

As mentioned in section 2.2, aquatic sites show a great number of differences in the ecological and physical characteristics. It has long been recognised that there is an association between each species and different habitat types with *Bombina bombina* being associated with larger and more permanent sites than *Bombina variegata* (Madej 1973). To measure the association between the species and their habitats it is first necessary to reduce the complicated characteristics of habitat to a single quantifiable measure. There are various techniques of multivariate statistics to do this and the method chosen for the *Bombina* habitat data is a discriminant function analysis. This is a technique that allows the reduction of a large number of variables used to describe a site's habitat to a smaller number (in this case one) and to classify sites into categories on this basis. This method was used to classify sites as either "pond-like" or "puddle-like" (MacCallum 1994, MacCallum *et al.* 1998). The calculated discriminant function can also be used to describe a site with a single value on a continuous habitat scale, giving a habitat score for each site which may then be scaled from 0 to 1 (from the most pond-like to most puddle-like site in the sample).

This method is applicable because habitat may already be classified into one of two categories, pond or puddle, on a subjective basis. This method produces a linear combination of separately weighted ecological variables (described in section 2.4.1) that best classifies sites into these categories. The function minimises the ratio of within to between category variance (the Wilk's λ). This proceeds in a step-wise fashion, including the variable that increases the Wilk's λ the least, if this produces a significant increase in the explained variation. This analysis was performed on habitat measurements from the Apahida hybrid zone (the measured variable are described below) using the program SPSS and is described by Vines (2002).

2.4.1 The ecological variables

A number of variables were measured in each site, quantifying the size, aspect and vegetation within the site. These essentially measure features important in determining and determined by the permanence of a site. The larger a site is, the less likely it is to dry early within the breeding season and the more permanent the site, the more likely it is to have typical aquatic vegetation (as opposed to submerged terrestrial plants) and possess dense populations of animal predators of eggs and tadpoles. These variables are described in Table 2.1. Other variables were rejected in an earlier analysis on the basis that they were too unreliably measured (MacCallum 1994). These variables and those that contributed least to the habitat discriminant function in Pescenica were also disregarded in Apahida (Vines 2002).

Variable	Measurement	Transformation
Length	Length on longest axis (m)	Log
Width	Length on perpendicular axis (m)	Log
Depth	Depth at deepest part (m)	Log
Emergent vegetation	% surface covered with emergent plants	Arcsin
Submerged/floating vegetation	% surface covered with submerged or floating plants	Arcsin
Bank Vegetation <15cm	% site bank with plants <15cm tall	Arcsin
Bank Vegetation 15-50cm	% site bank with plants 15-50cm tall	Arcsin
Bank Vegetation >50cm	% site bank with plants >50cm tall	Arcsin

Table 2.1 Ecological variables used in the calculation of the habitat score.

One of the key assumptions of the discriminant function analysis that the variables are multivariate normally distributed. The variables were transformated to improve their normality: natural logarithms were taken of continuous measures and the arcsins of percentages (Note that the correct transformation is an Arcsin transformation ($\sqrt{x}$),

Sohkal and Rohlf 1995. See Addendum for the effect of this incorrect use of the arcsin transformation).. Only the length, width and depth were directly measured, other variables were assessed subjectively.

2.5 Genetic methods

Genetic analysis of *Bombina* tissues is based on genotyping of four codominant genetic marker loci. These loci are presumed to be neutral in all analyses. Alleles at each locus are classed as being characteristic of originating from a *Bombina bombina* or *Bombina variegata* genetic background. This is based on the assumption that all the alleles found in this hybrid zone originated from the various pure populations that hybridised during the formation of the hybrid zone. The alleles present in pure populations were identified by genotyping of such populations from within Romania (B. Nurnberger, unpublished results), although the identity of the actual populations that contributed to this hybrid zone is debatable.

2.5.1 Marker loci used

The genetic analysis uses two microsatellite loci (designated 24.12 and 12.19) and two SSCPs (single-strand conformation polymorphism loci, 7.4. and 24.11). These loci were described by Nürnberger *et al.* (2002, primer sequences are available from Genbank). Microsatellite loci consist of tandem repeats of short sequences (2-4 nucleotide) and alleles exist as different repeat numbers. When amplified these sequences may be separated electrophoretically, separating the alleles by size and allowing the identification of maternal and paternal alleles. Different alleles in SSCPs consist of single or a few nucleotide sequence differences. These are detected by the effect they have on the folding properties of their amplified DNA fragments. When denatured fragments are exposed to specific chemical or temperature conditions they fold in different ways (show different shape conformations). The different properties of these

conformations allows the genotypes to be resolved by electrophoresis (revealing four bands in a heterozygote). For the loci described here low temperature conformations are separated by electrophoresis on cooled acrylamide gels.

2.5.2 DNA extraction

DNA is extracted from tissue samples using a standard chloroform extraction protocol. Ethanol preserved samples are first allowed to air-dry (in some cases the ethanol had evaporated off at some point, which did not prevent successful DNA extraction and gentoyping). The tissues were then digested overnight at 37°C in approximately 0.2ml of proteinase K in 1.5ml TNES (0.05 M Tris, 0.4 M NaCl, 0.1 M EDTA, 0.015 M SDS in H_2O, at pH 7.5). These were then mixed with 1.5ml of 2.6M NaCl, shaken vigorously for 30 seconds and centrifuged at 13,000 rpm for 10 minutes to pellet any cell debris. The supernatant was removed to another tube, mixed with an equal volume of chloroform and vigorously shaken. This was placed on ice for 15 minutes, after which the solution had separated into phases. The upper, aqueous phase was removed to a new tube. This was mixed with 1ml 99.9% ice-cold ethanol and left at -20°C for 1 hour, to precipitate the DNA. This was centrifuged (13,000 rpm for 15 minutes) to pellet the DNA. The supernatant was discarded and the pellet washed twice in 70% ethanol. The pellet was then allowed to air-dry and was then resuspended in 200µl of molecular biology grade purified water. The DNA content of a proportion of samples (diluted 1:20) was measured in a spectrophotometer at 260nm and 280nm. Samples were then diluted to a concentration of 10ng/µl assuming an equal DNA concentration across samples.

2.7.3 Polymerase Chain Reaction (PCR)

The PCR was set up in a volume of 30 µl per reaction. This contains 5µ (approximately 5ng) of template DNA, 1.5 µl of NH_4 PCR buffer, a variable quantity of 50mg/ml $MgCl_2$ (see table 2.2) and 0.5 µl TAQ polymerase (all PCR solutions are BioTAQ, Bioline), 1µl

dNTPs (0.2 mM per nucleotide) and 10 pm of each primer. The volumes were made up to 30 µl with molecular biology grade water (Sigma). For microsatellite loci one fluorescently labeled and one unlabeled primer was used (ABI-Prism primers with LIZ and FAM dye labels, Applied Biosystems, for 12.19 and 24.12 respectively). For SSCP loci both primers are unlabeled. Amplification was carried out on an MJ Research Dyad thermocycler with an adhesive film overlay. Initial denaturation was at 94°C for 3 minutes. The temperature regime thereafter was: 30 seconds denaturing at 94°C, 45 seconds primer annealing (initially at 65°C), followed by 1 minute elongation at 72°C. The annealing temperature was dropped by 1°C per cycle until it reached 50°C, whereupon amplification was continued using this annealing temperature for 21-23 cycles.

To test the success of PCR reactions (loci 7.4 and 24.11 only), 7µl of each PCR product was electrophoresed on a 2% agarose gel for ~30 minutes. DNA was visualized by Ethidium bromide staining. The microsatellite loci were not tested as fluorescent primers disrupt agarose electrophoresis.

Locus	Genbank Acc. No.	Marker type	$MgCl_2$ Conc. (mM)	Gel Temp.	Gel run Time	Voltage
7.4	AF472441	SSCP	2.5	2°C	~2.5 hr	500V
24.11	AF472425	SSCP	1.5	4°C	~4 hr	500V
12.19	AF472423	Microsat.	1.5	n/a	n/a	n/a
24.12	AF472426	Microsat.	2.5	n/a	n/a	n/a

Table 2.2 PCR and native polyacrylamide gel running conditions. Primer sequences are available through Genbank.

Genotypes were determined by electrophoresis; alleles were identified by running samples with known genotypes (genotyped by B. Nürnberger). For all loci there is only one allele characteristic of *Bombina bombina* but two or more characteristic of *Bombina variegata.*

2.5.4 Identification of genotypes – SSCP loci

Alleles at SSCP loci were determined by electrophoresis on native polyacrylamide gels. The gels contained 8% acrylamide (Roth). The gel buffer was 1xTA at pH 7.5. PCR products were diluted 1:1 with (100% formaldehyde, 0.05μM xylene cyanol). This was denatured at 95°C and then shock-cooled on wet ice prior to gel loading. Electrophoresis was carried out on cooled horizontal gel rigs (MultiPhor, Pharmacia) with a 2xTBE electrode buffer on SSCP buffer strips (ETC GmbH). The rigs were cooled to the temperatures specified in Table 2 by a continuous liquid coolant system. The samples were loaded onto the gel at 125V for 40 minutes and then run at the voltage and for the time specified in Table 2.2. The gels were then fixed for 30 minutes in 10% acetic acid then stained for 30 minutes in 60mM Silver Nitrate. The stain was visualized with a solution of 0.375M Sodium Hydroxide, 2.2mM Sodium Borohydride and 0.12% formaldehyde. The gels were then softened in 10% glycerol for 15 minutes, air-dried overnight and sealed with clear plastic film. The alleles present were identified from the furthest migrating strand pair.

2.5.5 Identification of genotypes – Microsatellite electrophoresis

Samples were diluted 100:1 with Hi-Di formamide (Applied Biosystems). Samples of both microsatellite loci were mixed and electrophoresed on an ABI 3730 sequencer (Applied Biosystems) by the sequencing staff of ICAPB, University of Edinburgh. The fragment sizes were analysed using GeneMapper software version 3.0 (Applied Biosystems, 2002). Alleles were assigned to taxa by comparison with individuals of known genotype (genotyped by T. Vines). For both loci there was only one allele characteristic of *Bombina bombina* but four (12.19) and five (24.12) characteristic of *Bombina variegata.*

Chapter 3.
Dispersal and habitat preferences

3.1 Introduction

Dispersal is one of the fundamental processes involved in determining the composition of populations along with birth and death (i.e. selection). By controlling the extent to which diverged populations come into contact with each other, different dispersal rates change the rate of gene flow between populations. In a heterogeneous environment the preference of individuals for different habitat types may also act to alter the genotypic composition of subpopulations. Dispersal rates and habitat preference can be inferred from genetic differences between populations or by direct observation.

This chapter uses data from a mark-release-recapture experiment to make direct estimates of the average dispersal rate of individual *Bombina* in the Apahida hybrid zone and to test whether there are differences in habitat preference related to individual genetic composition. It further uses these data to estimate the population size of aquatic sites, to determine the magnitude of movements in and out of sites during a sampling period during one breeding season. This study is complementary to that of Vines *et al.* (2003), estimating by direct methods those parameters inferred by indirect methods in this previous study. Whereas indirect methods illustrate the ultimate result of unknown processes, a direct study gives a clearer picture of some of the actual processes involved while not necessarily giving a full reflection of the overall result of these process. The results of this experiment are also compared with similar results obtained from the *Bombina* hybrid zone in Pešćenica, Croatia (MacCallum 1994, MacCallum *et al.* 1998), which, although following an overall cline shape, shows features of habitat-associated mosaicism when considered at fine scales.

3.1.1 Dispersal and the maintenance of population differences

If populations show differences in allele frequencies then migration between them will tend to break down these differences. The preservation or loss of a locally adaptive allele in a population subject to immigration is a result of a balance of migration and the selective advantage, such that the allele will be preserved if s>m (Haldane 1932), where *m* is the migration rate as a proportion of the population and *s* is the selection coefficient of the advantage. In a continuous habitat, the preservation of local adaptations depends on whether the size of the patch in which they are advantageous exceeds a critical minimum (Slatkin 1973, Nagylaki 1975). In such models dispersal is often approximated as a diffusion process and hence the appropriate measure of dispersal is the mean of the distribution of dispersal distances σ.

In a hybrid zone two populations meet that typically differ in the frequency of alleles at a number of loci. The form of such a hybrid zone will be determined by selection on different combinations of these alleles and recombination between them. The dispersal rate determines the rate of mixing between populations and hence the propensity for the formation of new gene combinations through recombination. In narrow hybrid zones, the cline width is determined by a balance of dispersal, selection and recombination (Barton 1983). In addition there may be an effect resulting from long-range migration into the hybrid zone which introduces "pure" genotypes from outside the contact zone (Rousset 2001).

3.1.2 Habitat heterogeneity in hybrid zones

Many hybrid zones are associated with an environmental heterogeneity. Some form broad clines, tracking similarly broad environmental gradients, such as in *Mus musculus* and *Mus domesticus* in Denmark or the northern flicker *Colaptes auratus* that has a broad hybrid zone across the Great Plains of North America (Moore and Price 1993). Others, making up a large fraction of all hybrid zones, are found across narrower

ecotones (Barton and Hewitt 1985). This does not even imply that the differential response to the environment is the primary cause of the structure. In fact, in such a situation there need not even be different environmental selection or habitat preference in the two populations, as a third habitat type associated the ecotone may lead to low population density; this could trap a tension zone which has only intrinsic selection against hybrids (Barton and Hewitt 1985). It is also plausible that in this third habitat type intervening between the two larger habitat ranges hybrids could actually be at a selective advantage. Such a condition is referred to as "bounded hybrid superiority" (Moore 1977, Moore and Price 1993, Good *et al.* 2000, Arnold 1997).

When there are selective differences between the hybridising taxa in different environments, genotype differences may track the changes in the environment. As environmental heterogeneities frequently have a patchy distribution this will be reflected in the distribution of genotypes and such hybrid zones are referred to as being "mosaic" in structure (Harrison and Rand 1989).

The precise form of a mosaic hybrid zone itself depends on the strength and types of selection and the dispersal patterns as well as the distribution of habitat types. Perhaps the best described mosaic hybrid zone is that formed between the ground crickets *Gryllus pennsylvannicus* and *G. firmus*. In a hybrid zone between these two species, each clearly maps over the distribution of one of two soil types (Harrison and Rand 1989, Rand and Harrison 1989). In this case it is clear that large areas of each pure species are maintained in this mosaic because the species are differentially selected on different soil types and that the patches of soil types are large relative to the dispersal distance of these species. This is confirmed by looking over the boundaries between soil type patches, where clines between the species are observed (Ross and Harrison 2002).

In the hybrid zone between chromosomal races of the lizard *Sceloporus grammicus,* the karyotypes are associated with different habitats. An important factor in maintaining this hybrid zone is the scattered distribution of habitat, which limits between site migration (Sites *et al.* 1995) and hence reduces rates of interbreeding between the races.

A hybrid zone between the grasshoppers *Chorthippus jacobsi* and *C. brunneus* takes a different form again. In this case there is a cline between the two species but there is also a significant effect of habitat variation causing deviations of the population composition from from the predictions of a clinal model without habitat (Bridle *et al.* 2001a).

3.1.3 Habitat-genotype association in *Bombina*

The influence of habitat on hybrid zones between *Bombina bombina* and *Bombina variegata* seems to vary from location to location. In hybrid zones in Prezmysl and Krakow, both in Poland, there is no evidence that habitat variation changes site composition in the hybrid zones (Szymura and Barton 1991). In a nearby cline at Stryj, in Ukraine, there is again a smooth cline but also there appears to be an association of genotype and habitat in the centre of the cline (A. Yanchukov unpubl.). In the hybrid zone at Pešćenica, in Croatia, there is a strong influence of habitat. Although the hybrid zone forms a narrow cline, there is great variation in the centre of the cline which is caused to a significant degree by habitat variability resulting in an association of the mean genotype in a site and the habitat (MacCallum 1994, MacCallum *et al.* 1998).

This habitat-genotype association results partly from an active preference of each species for a different habitat (MacCallum *et al.* 1998), *Bombina bombina* for ponds and *Bombina variegata* for puddles, but it may also be due partly to ecological selection for the taxa in the same habitat types. Nürnberger *et al.* (1995), MacCallum *et al.* (1995), Kruuk and Gilchrist (1997) and B. Nürnberger, S. Nell and F. Zajitschek (2003, unpublished data) demonstrate plausible mechanisms for such ecological selection. These include larger egg volumes and hence quicker development time in *variegata* aiding development before site drying in puddles, predation avoidance behaviours in *bombina* tadpoles and differential survival of tadpoles in temperature regimes typical of ponds and puddles. The range over which the habitat-genotype association is observed is far smaller than the range of lifetime dispersal (MacCallum *et al.* 1998) clearly showing

that whatever mechanisms create these association they are must be very effective to maintain divergence in the face of the homogenising effects of migration.

The *Bombina* hybrid zone at Apahida in Romania shows a different structure again, with a mosaic of habitat types explaining a large proportion of the variation in allele frequency between sites whereas there does not appear to be any significant clinal component (Vines *et al.* 2003). Currently there is no clear evidence of ecological selection, at least not at the tadpole stages when one might expect this to be at its most effective (Köhler 2003). As large pure populations are absent from this hybrid zone, and hence hybrids may have been through many more generations of hybridisation than in Pešćenica, it is difficult to see how any loci underlying habitat preferences and marker loci could remain associated for many generations (resulting in the observed habitat genotype association) in the absence of ecological selection. However more definitive experiments are required to determine whether ecological selection is indeed lacking in this hybrid zone.

A strong association between a site's habitat (measured by a function of ecological variables described in chapter 2) and the mean frequency of the alleles characteristic of *Bombina variegata* is seen in both the Pešćenica and Apahida hybrid zones. However this must largely be constrained to a narrow central band in Pešćenica as it is only there that there are both a range of available habitats and a range of genotypes - elsewhere the allele frequencies in the different habitat types converge (MacCallum 1994). In Apahida the area of the mosaic extends widely and the habitat-genotype association is maintained throughout (Vines *et al.* 2003, see figure 3.1). This association is approximately of equal strength if only the more "pure" populations (very pond or puddle-like sites) or the "hybrid" populations are considered. The slopes of linear regressions of site mean allele frequency on habitat are 0.40 and 0.46 respectively considering those sites with habitat scores either less than 0.4 or greater than 0.7 or the intervening sites.

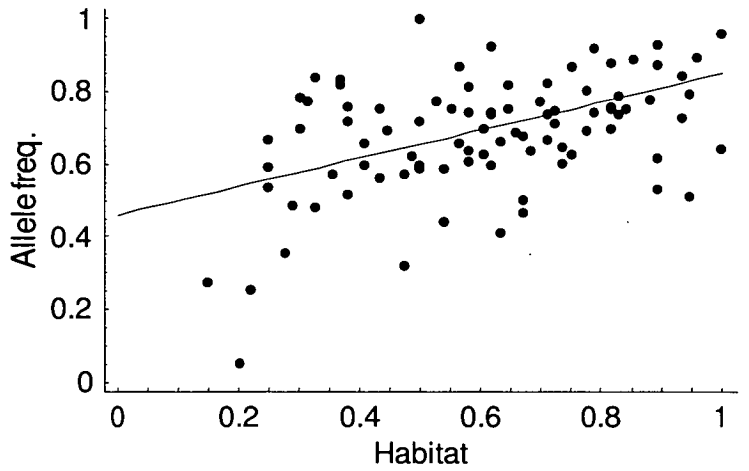


Figure 3.1. The association of *variegata* allele frequency (p) and habitat type (H) in aquatic sites in the Apahida hybrid zone. The regression is p=0.46+0.39H ($F_{1,92}$=30.2, $p<10^{-6}$). This figure is redrawn from Vines *et al.* (2003).

A habitat-genotype association in adults can be generated by one or more of three means:

1) There is differential adaptation of the two species for different habitat types. Between occupation of a site and sampling selection acts in such a way as alter the distribution of genotypes within sites consistently with respect to their habitat.

2) There is habitat preference expressed by adults in their movements between sites. The preferred habitat varies with genotype such that the observed association results.

3) Habitat association results from a preference for the natal habitat. Breeding adults choose the habitat of their mating site with respect to their genotype and the habitat of preference of their offspring is imprinted onto them by their natal habitat. As adults, the offspring choose sites according to this preference.

The third explanation is unlikely to be a sole explanation. As habitat association is not perfect and as a wide range of habitats is available to most individuals, with this mechanism alone the association of genotype and habitat will tend to break down. In each generation the association as adults will be weaker than at birth. The habitat preference of adults in their mating site is explored in chapter 5.

It also seems unlikely that selection is the sole agent of the association. If selection is to create such an association from a offspring mated at random from a variety of parental genotypes, it would have to act rather strongly (Vines *et al.* 2003). There was no evidence of this in the survival of tadpoles of different genotypes in sites of different habitats in Apahida (Köhler 2003). However, this was in a rather small sample of sites and tadpoles. On the other hand selection at the pre-metamorphosis stages is only relevant if the offspring consistently return to the same site as adults. Currently there is no evidence for or against this. If individuals choose their site only at an adult stage, then there may only be a very short time in which selection may act. Experiments to test this have yet to be conducted.

The great attraction of active adult habitat preference as an explanation is that it is capable of producing an association immediately on first habitat choice, and could be maintained or strengthened throughout the adult life as the animals move around. This is the hypothesis that is tested in this chapter.

3.2 Chapter aims

The main experimental method used in this chapter is a mark-recapture experiment with three main aims:

1. To estimate sizes of populations and rates of population turnover. This should indicate whether there is significant movement between sites of varying habitat type.

2. To estimate average distance of dispersal between sites and rates of movement between sites. This suggests the range over which mixing effects occur and hence over which differences in mean allele frequency between sites would be disrupted by random site choice.

3. To test for the presence and estimate the strength of habitat preference seen in the choice of habitat of individuals moving between sites.

3.3 Methods

3.3.1 Data collection

Animals were caught by hand in aquatic sites throughout the Apahida hybrid zone during the 2000 and 2001 field seasons (April-June). Upon capture these animals were sedated with 0.2% MS-222 (3-amino benzoic acid ethyl ester, Sigma), then had the middle-toe of a rear foot clipped (the right foot in 2000 and the left in 2001) and had

their belly patterns photographed and when revived were released back into the site of capture. These toes were preserved in individual Eppendorf tubes in 100% ethanol.

Throughout each field season sites were resampled repeatedly, approximately between every three days and a week for each site but with unequal sampling rates and effort at different sites. On these occasions the above procedure was repeated. Checks were made for the presence of missing toes and animals thus identified as recaptures were only sedated and photographed.

The local area around all sites was searched extensively for new sites, with searches repeated after rain. At the smaller sites all the animals in residence are easily captured but catching any adults at all in large ponds is challenging. This is despite there being large populations at least of male toads, judging from the mating chorus. At sites between the two extremes of size the proportion caught depends on the depth and cloudiness of the water and the amount of vegetation.

Resampling of each site continued until they were found to dry out (and again after any rain). Sampling finished when the majority of sites had dried. These collections were carried out in 2000 by Sonja Köhler and Tim Vines and in 2001 by Sonja Köhler , Tomas Alfert, Lino Ometto and myself.

The site habitats are quantified using the habitat score described in chapter 2 giving a habitat score that is scaled to run from 0 to 1 (describing the range of sites from the most "pond-like" to the most "puddle-like"). The distance between sites was calculated from co-ordinates given by a GPS or from estimation of the location on a map. This slightly overestimates the distance between very close sites, in which the shortest distance between the water will be closer than the centre point of the sites (in some cases the water of site pairs are only around 3m apart but the site centres are separated by up to 15m).

The identity of recaptures was determined later from photographs. *Bombina* have a distinctive pattern of bright red or yellow colouration on a background of dark skin on their undersides. This is used as the basis of the spot score (described in chapter 2). Although variations in the patterns of coloured spots tend to reflect certain types, they appear to be individually unique. These patterns are then used to identify the recaptures.

The toe tissue was genotyped for a sample of adults from each site (minimum 12 per site) and all recaptures at the four unlinked genetic loci using the methods described in chapter 2 (loci designated 7.4, 12.19, 24.11 and 24.12). The genotyping of most adults was carried out by Tim Vines (2000 season adults) and Sonja Köhler (2001 season adults) as part of a survey of adult genotypes. Adults identified as recaptures that had not previously been genotyped were genotyped by myself.

3.3.2 Dispersal distances

The average length of dispersal distance is taken as the average distance between one capture and the next. This is calculated both for all recaptures and for only those that move site between captures. It is calculated separately for males, females and for the "pure" taxa and "hybrids", which I define as those individuals with mean marker allele frequencies of <0.25, 0.25-0.75 and ≥ 0.75 respectively. As the distribution of dispersal distances is unknown, the significance of any differences between these groups is tested using non-parametric Mann-Whitney U-tests (for the paired sample comparisons) or Kruskal-Wallis ranks test (for comparing the three mean allele frequency classes).

3.3.3 Distance moved against time

The distance individuals move with time allows us to see if migration follows a random walk or is more directed. The relationship of distance moved and time is examined by looking at the distance moved by individuals between captures against time between

captures. Any relationship is tested by fitting a linear regression. Any tendency to consistently move away from the original site is examined in those individuals caught on three of more occasions.

It is not appropriate to look at the proportion of individuals staying in the same site over time, as this would fail to account for the changes in habitat during this period. In most sites the water level decreases and the water temperature increases during the season and it is not clear at what point a site becomes uninhabitable. These changes are characteristic of puddle environments so this might be genotype biased.

3.3.4 Population sizes

It is impossible to take a direct estimate of the population size of a given aquatic site for two reasons. Firstly, as mentioned above, in only a few sites is it possible to catch every animal. Even then, on occasions when the whole area of a site has been meticulously hand-searched, animals have later been seen in the site. The second problem is one of the definition of the population of a site. Although the aquatic habitat clearly plays an important role in the life of a toad, animals that frequent a site may also inhabit the surrounding area. If these animals are to be included in the site's population then they must also be sampled and finding animals outside of the aquatic habitat is difficult.

Instead there are a number of statistical methods that can be used to estimate the population size of a given site from the number of new captures and recaptures at repeated visits to a site. The stochastic method of Jolly and Seber (Jolly 1965, Seber 1965, Seber 1982) allows the estimation of a number of population parameters. This is based on the assumption of an open population (i.e. one in which there may be both immigration and emigration from the population). This is appropriate for this analysis as there may be both immigration into and emigration from these sites.

The main estimator of the Jolly-Seber analysis is that of the current population size, calculated as:

$$N_t = \frac{M_t}{\alpha_t}$$

where N_t is the estimate of population size at sample period t, M_t is the estimated number of marked individuals in the population and α_t is n_t/m_t where n_t and m_t are the number of animals and marked animals caught at time t. The estimate of the number of marked individuals is given by:

$$M_t = \frac{a_t Z_t}{R_t} + r_t$$

where a_t is the number of animals previously released, R_t is the number of animals released at period t and subsequently recaptured, α_t is the marked proportion at sample period t and Z_t is the number not caught at sample period t but caught subsequently. It is also possible to estimate the survival probability ϕ (or the proportional loss of individuals 1-ϕ) and the gain of individuals β:

$$\phi_t = \frac{M_{i+1}}{M_t - r_t - a_t}$$

$$\beta_t = N_{t+1} - \phi_t (N_t - n_t + a_t)$$

To obtain all these measures it is necessary to have recapture data from three subsequent sampling periods. Where this has not been possible, parameter estimates or their standard errors have been omitted.

The gain of individuals into a population may be through recruitment of new individuals by birth or immigration. Similarly loss of individuals may be through death or emigration. When applied to a population of *Bombina* within the period of a single field season, I consider immigration as the only source of population increase.

Whether mortality is causing a significant decline in populations size compared to emigration is less clear. Individuals first caught in 2000 make up only 10% of the 2001 season recaptures despite making up nearly half of all marked animals. This could indicate a high rate of mortality in the intervening period. However individuals may have migrated away from the original sites but not to other sampled sites, i.e. they are still alive but will evade capture. Furthermore high over-winter mortality does not necessarily imply that there is any significant mortality during a single field season. For these reasons I will make the assumption that mortality is largely negligible during a single field season.

A further assumption of the Jolly-Seber methods is that individual mortality rate is age-independent during the sampling period. This seems likely to be true of *Bombina*, which are known to be long-lived in captivity (Duellman and Trueb 1994). This is not to imply that wild *Bombina* are equally long-lived, just that given their potential longevity it is unlikely that their mortality rates changes drastically during the experiment. Although it is almost certain that *Bombina* suffer higher mortality in the wild than in captivity, this higher mortality is likely to be a consequence of predation or disease, which are not likely to change greatly during a single field season, rather than ageing.

3.3.5 Sites used for estimates

The Jolly-Seber methods were applied to recapture data from four areas. Two of these contain multiple sites in close proximity. The two pooled populations are the Apahida-Cojocna road sites and sites 372 and 373. Sites 258 and 290 were relatively isolated from other sites and are considered as single isolated sites.

The Apahida-Cojocna road sites are a collection of twenty two sites stretching for approximately 100m along the sides of a road. The most striking sites are ten large and flooded man-made holes. The original purpose of these holes is unknown but it is known that they vary in age with the first having been dug in the 1960s up to the latest which have been dug in the last decade. These holes vary from around three to fifteen metres

across and up to three metres deep, and there are also several other associated sites formed in shallow scrapes and the flooded wheel-tracks of heavy machinery. The vegetation in these sites varies according to the size, shape and age of sites. The oldest site (200.3) is extensively vegetated with reeds and pondweed, whereas the newer sites (e.g. 200.5) have only patchy vegetation. Sites vary in their water quality from highly turbid to quite clear, with clarity increasing with age but also affected by disturbance by animals and passing traffic. Several of the sites also have only a small area of shallow water and rather steep sides. The water quality and steepness of the sides do not contribute to the habitat score but may be of importance to the toads and varying water quality and bank steepness are found in sites of a variety of habitat types. Two nearby sites are not included in this analysis (sites 334 and 335). Although separated from the nearest other sites by only around 100m, this distance includes a deep drainage ditch with running water and a railway line and embankment which may act as significant barriers to migration between these areas. Parameters for the Apahida-Cojocna road sites are estimated for all individuals marked or recaptured pooled for each week of the field season to reduce the number of parameters calculated.

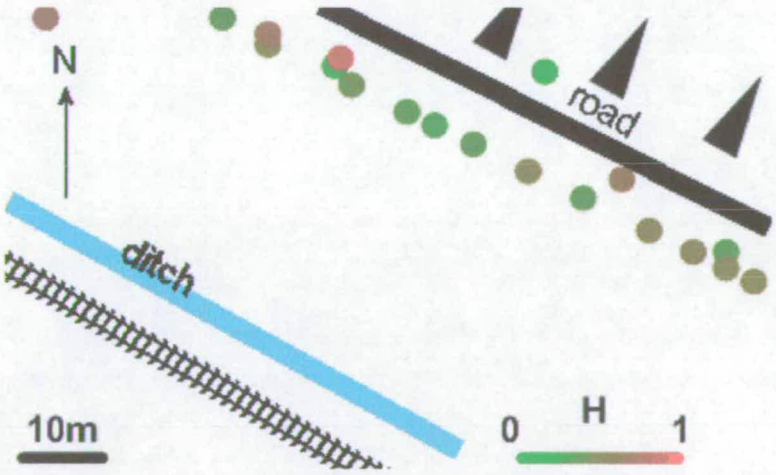


Figure 3.2. The location and topography of the Apahida-Cojocna road sites. The habitat score of sites are indicated by their colour.

Site 258 is a small, natural pond contained in a hollow on the side of a small valley, in an area of lightly grazed, grassy meadow. The valley has numerous such sites apparently fed from natural springs (appropriately it is called Valea Broaştelei, the valley of toads). However during the 2001 field season all other sites were dry.

Sites 372 and 373 are located in a small valley containing mixed-crop fields, meadows and woodland. Site 372 is a steep-sided, man-made watering hole (~1m in diameter and 60cm deep, lacking in vegetation and disturbed daily by horses that drink from it. Site 373 is a shallow drainage ditch (around 50cm across containing water for about 15m of its length). It is connected to site 372 by a shallow and completely dry drainage channel.

Site 290 is a medium sized pond located next to and fed by a well (which itself contains fish, some frogs of the genus *Rana* but no *Bombina*). It is almost completely covered by emergent aquatic vegetation. It is located close to two sites in ditches (289 and 285) which were dry throughout the 2001 season and hence are excluded from this analysis.

3.3.6 Observed and expected rates of recapture

In a "closed" population (i.e. one with no immigration or emigration) with no mortality, the probability of recapturing marked individuals increases with time as the marked proportion increases. One way of determining if populations are in fact "open" with individuals moving in and out of them is to compare the observed and expected numbers of animals that are new at every sampling period. Under the "open" population assumptions of the Jolly-Seber analysis, an excess of marked individuals can only indicate emigration or mortality biased towards unmarked individuals. If there is no recruitment of juveniles into the population, a deficit of marked individuals can only indicate immigration.

Under the null hypothesis (a closed population), the proportion of the catch that is unmarked declines at the rate Pr (recap)^s where pr(recap) is the probability of recapture

at any time and *s* is the sampling period (each week of the field season). The probability of an individual being recaptured during each time interval is estimated by the Jolly-Seber method and I take the recapture probability at any given point to be the average over all intervals. The proportion of captured animals that are found to be unmarked is noted at each sampling period. The number of marked individuals in the population (i.e. those marked adults that haven't died or emigrated) is estimated as part of the Jolly-Seber procedure. Differences from expectation are calculated using a χ^2 test.

3.3.7 Measuring habitat preference

Are the choices of habitat the result of preferences for different habitats depending on individual genotype? There is a strong association of habitat and marker genotype for all adults in the site of their first capture (see figure 3.1, r^2=0.39) but at the first capture it still remains a possibility that this results from ecological selection acting differentially on varying individual genotypes, depending on the site habitat. Assuming that any genotype-biased ecological selection during the recapture experiment is negligible, recapture data allows us to test whether adults are making choices about their preferred habitat and thus whether this association may be generated by habitat choice alone.

MacCallum *et al.* (1998) examined this hypothesis in the Pešćenica hybrid zone transect by looking for genotype bias in migration between close pond-puddle pairs. This is not possible in the Apahida hybrid zone as the large pond sites (site nos. 293, 294, 85 & 333) happen to have no surrounding puddles and the few surrounding sites yielded few recaptures or were dry in 2001.

3.3.8 The effect of migration on the habitat genotype association

The equivalent to the above in Apahida is to test for the association of genotype and habitat after migration. If migration is random with respect to genotype and site habitat it will tend to break down the association observed in adults before migration. The habitat-genotype association is assessed by regression of individual mean allele frequency on habitat. Habitat preference is tested by testing the significance of the association of genotype and habitat at second or later captures. Significance of this regression cannot be tested by standard methods (i.e. testing that the regression coefficient is significantly different from zero) as this could be biased by the range of local genotypes and by correlation between habitat types in each area. Significance is instead tested by a randomisation method. Replicate randomly chosen populations are generated by assigning to each individual a site at random from those sites that are available and repeating the regression. The significance of the observed regression is assumed to be the proportion of cases in which the magnitude of the observed regression exceeds that in the random replicate. This is repeated for 1000 replicates. I define an available site as one that lies less than 120m from the original. This distance encompasses all the observed within-season dispersal distances.

Though it might be considered that only individuals that move between sites can express a habitat preference, it is also plausible that remaining within the same site between capture represents a habitat choice in itself. There is a risk that the inclusion of these recaptures fails to exclude the possible effects of ecological selection that may have created the earlier habitat-genotype association. However, if philopatry is itself a habitat choice, then excluding these individuals throws away information about habitat preference and potentially even excludes those individuals that are most satisfied with their current habitat choice. For this reason the regression is carried out twice, once on all recaptures and once on only those that choose a site different from their original.

A somewhat different measure is to compare the habitat-genotype association at initial and later recaptures. This essentially tests whether habitat preference is maintaining the association originally observed. The significance of differences between the regression at initial and later capture is tested by an ANOVA testing the variation amongst regression slopes (Sokal and Rohlf 1995). If this test finds that the slopes are not significantly different then deviations of y-intercepts of the regressions are tested by ANCOVA (Sokal and Rohlf 1995).

3.3.9 A likelihood model of dispersal and habitat preference

The measurements of dispersal distance and habitat preference used above have two drawbacks. Firstly the discrete nature of the sampled sites will tend to reduce the apparent mean dispersal. For instance if two alternative sites were available, one 10m from the source site and one 1km, then the nearer site might receive the bulk of migrants, even if the mean dispersal distance in a continuous environment were, say, 300m. Secondly they ignore interactions between dispersal distance and habitat preference. Consider a situation where there are two sites around a source population, with one nearer to the source than the other, with the further site having a preferable habitat for the majority of occupants of the source than the nearer. Whether the further site can be exploited depends on the dispersal range and habitat preference; the further site might receive proportionately more immigrants than expected given the mean dispersal distance as the nearer site is rejected.

The method presented below attempts to overcome these drawbacks by estimating mean dispersal distance and habitat preference simultaneously. The model gives the likelihood of the observed pattern of recaptures in terms of two parameters; the mean distance of dispersal and the global habitat preference strength.

The shape of the distribution of dispersal distances in *Bombina* is unknown. Although there have been several studies of migration in other anurans (reviewed by Beebee 1996)

but there is only scant data on the distribution of dispersal distances, measurement of which is known to be methodologically problematic (Koenig *et al.* 1996). The spatial distribution of individuals dispersing by a random walk from a fixed point, in a given length of time, can be described by a Gaussian curve (Skellam 1951); the standard deviation of this distribution increases with time. However, the distance moved by *Bombina* from the original site is not correlated with the time since previous capture (see figure 3.5), which would seem to suggest that their movement does not follow a random walk. However, a broadly similar distribution is expected when recaptures are obtained from a geographically bounded area (Koenig *et al.* 1996). This may be the case for the sampled adults as there sites are strongly clumped and the distance between the groups of sites far greater than the average dispersal distance.

Also dispersal distributions are frequently platykurtic, as there are occasional long-distance migrants. However for simplicity and in the absence of other information on the distribution of dispersal distances in *Bombina* I will assume that the distribution of *Bombina* dispersal distances is Gaussian in form.

Given a distribution of dispersal distances in continuous space, what is the expected distribution in an environment with discrete habitats? The answer could depend on variables such as the landscape between one location and the next (Wien 2001). It could also depend on any number of behavioural factors (e:g. is the distance to surrounding sites known or detectable by dispersers? Does dispersal distance vary from individual to individual and is it correlated with other variables?). Again I use perhaps the simplest approach which is to suggest that the number of individuals ending up in one site, having moved from another, is proportional to the height of a normal curve (with a standard deviation equal to the mean dispersal distance and mean of zero) at the distance between these sites. This measure is independent of the time between first and later capture.

The probability of an individual ending up in the observed site, if habitat is irrelevant, is therefore the height of the normal curve at the observed distance from the source, as a

proportion of the sum of probabilities for all sites (most of which are distant and the probability of reaching them will be approximately 0 if the dispersal distance is not very large):

$$pr(site \mid \sigma) = \frac{pr(x \mid \sigma)}{\sum_{i}^{nsites} pr(x_i \mid \sigma)}$$

(equation 3.1)

where σ is the standard deviation of the dispersal distribution and x is the distance moved between sites. If habitat preference is entirely under genetic control then the preference for a given habitat depends on these factors: the individual genotype at habitat preference loci, the destination site habitat and a preference function. Of these, only the site habitat can be measured. I make the assumption that the genotype at habitat preference loci may be described by the allele frequency at marker loci. I assume the preference function decays as an exponential function of the deviation of the actual habitat type from that predicted:

$$preference\ (site \mid\ genotype, habitat, H) \propto e^{-H(habitat - predicted_{genotype})^2}$$

(equation 3.2)

where H is the global habitat preference strength and the predicted optimal habitat for a genotype is given by the observed adult regression of allele frequency on habitat at first capture. The preference of site habitats relative to the most favoured under a range of habitat preference strengths is shown in figure 3.3. Note that a negative habitat preference is also possible, if the expected optimal habitat is in fact the least preferred.

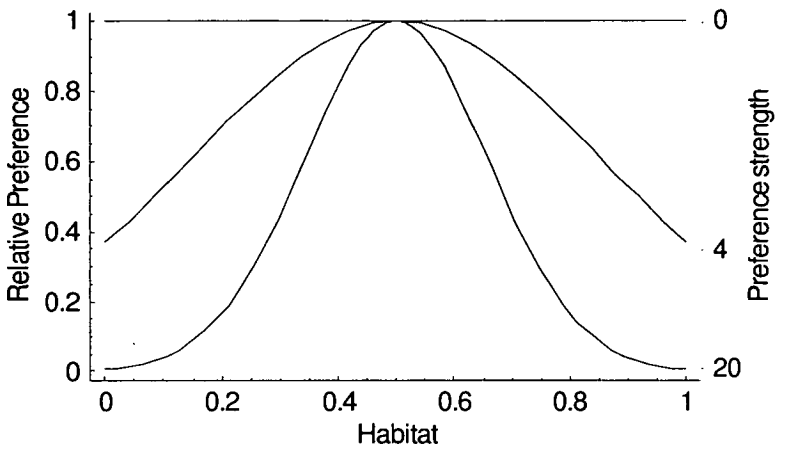


Figure 3.3. The habitat preference function for an individual with the greatest preference for sites with habitat 0.5 with no habitat preference (strength 0, dashed line), intermediate preference(strength 4, dotted line) or strong preference (strength 20, solid line).

The probability of an individual choosing the actual destination site, irrespective of distance, is given as the preference for this site relative to the preference for all other available sites. Ignoring the restrictions of limited dispersal distance for the moment, available sites can be defined as those whose habitat the toads are capable of detecting, rather than having to be visited. Sinsch (1990) reviews the orientation cues that amphibians use to find suitable habitat during migration which include olfactory, geomagnetic and landscape cues, possibly in conjunction with a "map sense". The toad *Bufo bufo* has been known to migrate up to 3km to annual breeding pools even after they have been destroyed (McMillan, 1963), suggesting it uses landscape cues for orientation. As *Bombina* have also been known to migrate long distances (MacCallum 1994) it seems likely that they are similarly capable of locating known distant sites. Also it is generally possible to predict the habitat of an unseen site from the local topography and the sound of the mating chorus so it seems likely that *Bombina* have this ability too.

The range over which these senses are reliable is crucial to these assessments as it determines the number of available sites (if the habitat preference strength is zero, the probability of choosing a site is given as 1/(number of available sites) by equation 3.3).

However the clumping of sites observed in Apahida means that if the detection range is between approximately 100m and 1km the same number of sites will be available. The analyses are conducted under the assumptions that toads are capable of detecting the habitat of all occupied sites within 120m. To test the validity of this figure, analyses were carried out where the detection range was allowed to vary as a multiple of σ, the mean dispersal distance. When the multiplication factor was small and therefore the detection distance possibly much lower than 120m the likelihood surface (described below) was extremely rough with multiple likely values of σ. As this factor increased, such that detection ranges were around 100m, the likelihood surface smoothed out to give a single most likely value of σ. Increasing the multiplication factor still further made no difference to estimates.

The probability of the observed choice of site is the product of the probability of choosing the site based on its habitat multiplied by the probability of migrating the distance to that site. Therefore the likelihood of the site choice of an individual, given a value of mean dispersal distance and habitat preference, is this probability proportional to the probability of all other available sites (equation 3.3).

$$lik(site_x, genotype, habitat_x \mid \sigma, H) = \frac{\Pr(site_x \mid \sigma) * pr(site_x \mid habitat_x, H, genotype)}{\sum_{i}^{nsites} \Pr(site_i \mid \sigma) * pr(site_i \mid habitat_i, H, genotype)}$$

where nsites is the number of available sites

(equation 3.3)

The overall likelihood is the product of the likelihoods of all migrant site choices.

$$lik(\sigma, H \mid all\ choices) = \prod_{i}^{nmigrants} lik(site_i \mid \sigma, H)$$

(equation 3.4)

Again the issue arises of how to treat individuals that fail to move site. It is possible to make a separate estimate of the probability of philopatry. Bailey *et al.* (2003) model such an effect as a fixed rate of philopatry. Bernstein *et al.* (1991) model the decision to leave a site as a judgement that the quality of the current site falls below the mean of all detectable sites. In this second model, describing the site choice of a predator, the decision to move is based upon the depletion of limited resources. Although this could also be the case in *Bombina* it is not clear what resources are depleted, how and when. Therefore I have chosen a simpler approach that regards the choice of the current site in exactly the same way as any other site choice.

The relative probability of choosing the same site is given by the relative height of the normal curve and the preference by the habitat preference function. This view seems to tally with the strong habitat association of such recaptures. However the dispersal distance estimated under these assumptions will likely be reduced, so I make separate analyses with and without these individuals. The most likely value of mean dispersal distance and habitat preference are found by entering values of mean dispersal distance and habitat preference into equation 3.4. The approximate 95% confidence area is described by the range of values that give a 2-unit drop in the logarithm of the likelihood (Mangel and Hilborn 1997).

This method was tested on simulated data sets. The probability of each individual choosing any particular site was calculated by equations 3.1 and 3.2 and the site choice of individuals was drawn proportionally to these probabilities. Recaptures remaining in the same sites were created either with the assumption that the current site was an

available choice, the same as any other, or with a fixed rate of philopatry increasing the probability of staying. The results are shown below in table 3.1. It can be seen that the when habitat preference is absent or weak its estimates vary, and may even give false indications of the presence of habitat preference. At stronger habitat preferences (strength 10) the estimates are also quite variable but are generally quite close to the true value and in all case correctly identify the presence of habitat preference. The estimation of dispersal distance is somewhat less convincing, overestimating short dispersal distances and underestimating the larger. The latter is almost certainly the result of a large dispersal distance compared with the distribution of possible between site migration distances. When analysing data generated under the fixed rate of philopatry assumption habitat preference estimates are reasonable but all dispersal distance estimates are greatly overestimated.

Dispersal \ Hab. Pref.	0	4	10
50m (a)	**4 (1-7)**	**4 (0-6)**	**4 (5-11)**
	50m (40-90)	130m (100-160)	50m (42-62)
100m (a)	**3 (0-5)**	**4 (2-8)**	**5 (3-8)**
	150m (130-185)	140m (120-160)	110m (100-160)
500m (a)	**4 (2-6)**	**7 (3-10)**	**11 (7-11)**
	450m (400-550)	250m (200-325)	280m (200-320)
Fixed rate of philopatry and dispersal 50m (b)	**5 (1-9)**	**6 (3-9)**	**6 (3-10)**
	130 (110-155)	130m (110-150)	130m (110-150)

Table 3.1. Estimates and 95% confidence limits of habitat preference and dispersal parameter estimates on single simulated data sets. Simulations (a) were generated randomly under the same assumptions as the statistical method. Simulations (b) assuming that an individual's preference for their current site were doubled. Habitat preference estimates are in bold and dispersal estimates in feint type.

3.4 Results

In the 2000 season, of 1064 adult captures, 15 were found during the 2001 field season (sampling of which included the same sites at least once, where they were extant). Of 1457 adult captures within the 2001 season, 196 were recaptured later within the same field season. 34 adults were recaptured more than once (31 twice, 2 three times and 1 four times). Of these recaptures, 69 were in sites other than that in which they were first caught. Details of the sites and dates on which individuals were caught and the numbers recaptured from each site (along with the groupings of sites) are given in the appendix. 176 recaptured adults have genotypes at the marker loci.

3.4.1 Dispersal measures

The average dispersal distances are given below in table 3.2 and the distribution of distances between captures within 2001 shown in figure 3.4. There are separate measures for all adult recaptures within the 2001 season, only those recaptures in sites different from previously occupied and the same between the 2000 and 2001 seasons. These measures are repeated for the sexes separately and for individuals with *variegata* allele frequencies in the ranges <0.25, 0.25-0.75 and ≥ 0.75. In all cases multiple consecutive movements by the same individual are included as separate data points (the non-independence of these data points is not a problem as significance is determined by a randomisation test). The tests of differences between sexes or allele frequency categories showed no significant difference between any groups, either including or excluding non-moving individuals (Mann-Whitney U-tests to compare the sexes and individual hybrid index classes, and a Kruskal-Wallis ranks test for *bombina,* "hybrid" and *variegata* classes).

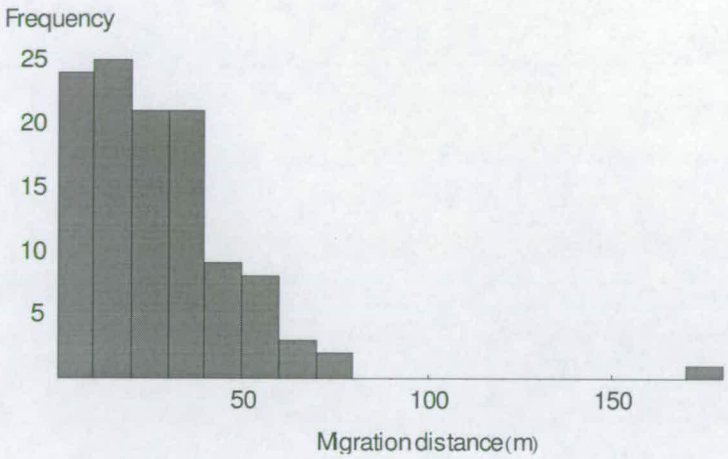


Figure 3.4. The distribution of individual migration distances during the 2001 field season.

	Within 2001 (Apr-June)	Within 2001 (moved site only)	Between year (all)	Between year (moved site only)
All	**7.2m (5-9m)**	**30.7m(25-36m)**	**20.1m (4-36m)**	**30.1m (7-54m)**
	n=229	n=64	n=23	n=15
Males	**5.2m (2-8m)**	**38.4m(25-52m)**	**13.6m (6-21m)**	**22.8m(13-32m)**
	n=96	n=13	n=20	n=12
Females	6.3m (4-9m)	29.2m (22-36m)	63.0m (0-304m)	63.0m (0-304m)
	n=111	n=24	n=3	n=3
$p \leq 0.25$	6.2m (0-13m)	**18.6m (6-32m)**	2.5m (0-34m)	5m
	n=12	n=4	n=2	n=1
$0.25<p<0.75$	**8.8m (5-13m)**	**33.6m(23-44m)**	10.9m (0-24m)	**18.2m (4-32m)**
	n=80	n=21	n=5	n=3
$p \geq 0.75$	**9.5m (5-14m)**	**33.8m(25-42m)**	29.4m (0-61m)	50.4m (0-103m)
	n=64	n=18	n=12	n=7

Table 3.2. Mean dispersal distances between recaptures with standard errors. Significant values are shown in bold type For all individuals, by sex and marker genotype class. Comparison of the dispersal distance of groups, including and excluding non-movers are shown overleaf. Statistics are Mann-Whitney U and Kruskal-Wallis H values with respective significance probabilities.

Male/female comparison	All	U= 4928.5	P=0.164
	Move only	U=125	P=0.337
All allele frequency classes	All H= 0.556, p=0.767		
	Move only H=1.068, p=0.586		

Allele freq. class comparisons (All and move only)	**Hybrid**	**Variegata**	**pure**
bombina	U=399 , p=0.509 U=17 , p=0.443	U=395.8 , p=0.508 U=4978 , p=0.635	-
hybrid	-	U=19.5 , p=0.297 U=337 , p=0.666	U= 5369, p=0.707 U=522 , p=0.953

Table 3.2. Continued. See previous page for full legend

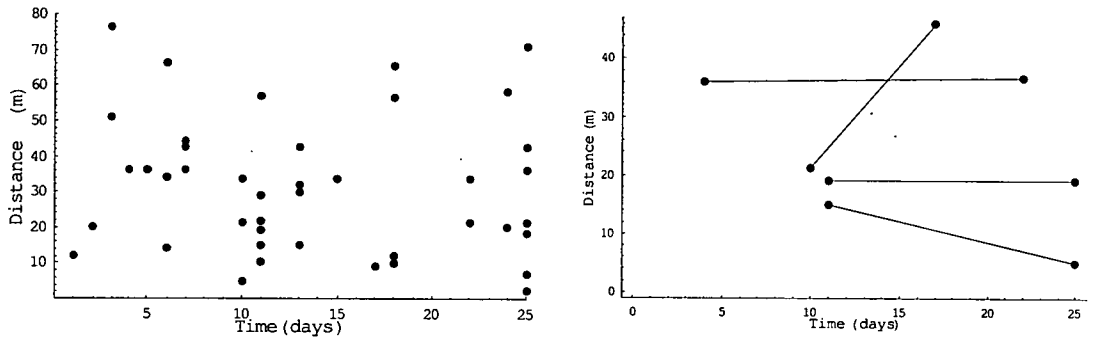
Data from a similar study in Pešćenica, Croatia (MacCallum 1994) similarly show no significant differences between any groups. However the mean distances recorded were much larger: the within-season average dispersal distance, despite capture over a shorter period, was 358m in Pešćenica (30.7m in Apahida) and between seasons was 929m (30.1m in Apahida). The Apahida recaptures very rarely made long-distance movements, with no movements over 100m and only 9 (~5%) greater than 50m. In contrast all movements in Pešćenica were greater than 50m and 35% were over 150m. Between seasons in Apahida there was only one movement greater than 100m (175m), whereas in Pešćenica there were numerous movements between seasons greater than 1km.

It seems likely that this is at least in part due to the distribution of sites in Apahida. Although there are numerous sites within a few tens of metres of one another in both hybrid zones, in general the next nearest site in Apahida is likely to be more than a kilometre distant. Therefore it is not clear to what extent the lower migration rate is a result of a tendency to migrate shorter distances and what results from a lack of sites to migrate to within a reasonable distance, with more distant sites being linked only by infrequent long-distance migration.

3.4.2 Distance and Time

The distance moved by individuals with respect to the length of time between their captures within 2001 is shown in figure 3.5 as is the distance moved from the starting position by the first and second recaptures (note that these exclude individuals that do not move site between one or more of their captures). In the first case there is no significant regression of distance moved against time (distance (m) =0.037-0.0004days, p=0.201). It is also noteworthy that the longest migration of 76.5m takes place in only three days. Clearly the toads are capable of very fast migration if this is desirable. There is a slight increase in the average distance gained from the start point at the second

recapture over the first (23m at the first, 30m at the second) but migration back toward the first site also occurs. This is as would be expected if the migrations are random with respect to the previous movement.



(a) (b)

Figure 3.5. The relationship of time between first and later capture and the distance moved from the original site, for all within-season recaptures (a) and those caught three times (b), whose movements of individuals are joined

3.4.3 Population size estimates

The parameter estimates from the Jolly-Seber analysis are given below in table 3.3. For each site or site grouping and time interval estimates and 95% confidence intervals are given of population size, the probability of an individual being recaptured in the intervening time, the estimated "survival rate" and the estimated population size change. The indicated "survival rate" is really a measure of the rate of population change so includes loss through emigration and gain through immigration (and hence can be greater than one).

Site 258	*Week 1*	*Week 2*	*Week 3*	*Week 4*	**Average**
Population Size	-	**55 (3-106)**	309.8 (0-787)	87.8 (0-192)	***150.7***
Recapture Probability	-	0.182	0.058	0.246	***0.152***
Φ ("survival rate")	**1.57 (1.2-2)**	3.07 (0-6.3)	0.32 (0-0.67)	-	*1.66*
β (population increase)	-	140.95	-13.24	-	*63.86*

Site 290	*Week 1*	*Week 2*	*Week 3*	*Week4*	*Week 8*	**Average**
Population Size	-	-	**28.6 (6.8-50.3)**	**22.5 (4.9-40.1)**	9	20.03
Recapture Probability	-	-	0.280	0.311	0.111	0.23
Φ ("survival rate")	**0.75 (0.68-0.82)**	0.65	**0.81 (0.67-0.95)**	0.088	-	0.575
β (population increase)	-	13.99	-	-	-	13.99

Table 3.3. Table continued and full legend overleaf

Sites 372/3	*Week 1*	*Week 2*	*Week 3*	**Average**
Population Size	-	**47.44 (23.8-71.0)**	**14.0 (7.9-20.1)**	30.72
Recapture Probability	-	0.443	0.214	0.33
Φ ("survival rate")	**1.078 (1-1.15)**	**0.6 (0.58-0.62)**	-	0.84
β (population increase)		41.84		41.84

Apahida-Cojocna road sites	*Week 1*	*Week 2*	*Week 3*	*Week 4*	*Week 5*	*Week 6*	**Average**
Population Size		521.5 (0-1783)	4203.25 (0-19303)	552 (0-1536)	145.0 (0-313.5)	316.4	1147
Recapture Probability		0.288	0.0102	0.080	0.262	0.212	0.12
Φ ("survival rate")	**0.709 (0.23-1.2)**	2.37 (0-12)	**0.19 (0.15-0.23)**	**0.38 (0.34-0.42)**	1.04 (0-5.2)	-	0.94
β (population increase)	-	2967.5	-265.5	-65.6	257.2	-	723.4

Table 3.3. Estimates from Jolly-Seber analysis of 2001 recapture data from sites 258, 290, sites 372 & 373 together and the Apahida-Cojocna road collection of sites. Estimates are made on recaptures in blocks of one week. Variables calculated are population size, marked proportion of the population, Φ (the survival rate), and β, the number of new individuals entering the site) with 95% confidence interval where possible (see methods). Significant values are in bold type.

The estimates of the population size in each of these sites fluctuate greatly from one sample period to the next. It is difficult to determine how closely these changes reflect the reality as the estimates generally come with large standard errors, reflecting the low power of these methods when there is sporadic success in recapturing adults (Southwood and Henderson 1998). This could plausibly be indicative of fluctuations in population size, consistent with highly fluid population composition in the aquatic sites.

3.4.4 Observed and expected rates of recapture

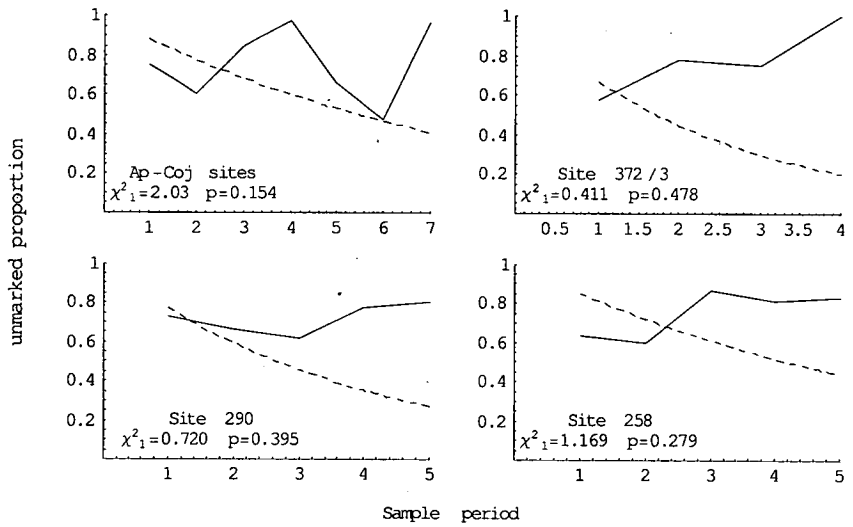


Figure 3.6. The expected (dashed line) and observed proportions of marked adults caught at each time period and test of significance of fit

Comparisons of the observed and expected rates of recapture under a closed-population assumption are shown in Figure 3.6. It can be seen that the observed unmarked fraction in each case differs significantly from expectation and at most points is greater than would be expected in a population without immigration or emigration. Where the observed values are lower than expectation they are only slightly so in contrast to observations large than expectation that are frequently much larger. There are two interpretations either indicating an influx of individuals during the season, increasing the pool of unmarked individuals (as they are likely to be unmarked), or that there is a preferential emigration of marked individuals. It is not implausible that the act of

marking individuals disturbs them enough that they are more likely to move although it seems unlikely that this would cause such a marked effect. In sampling in Pešćenica the same effect was seen (MacCallum 1994).

3.4.5 Is habitat choice random with respect to genotype?

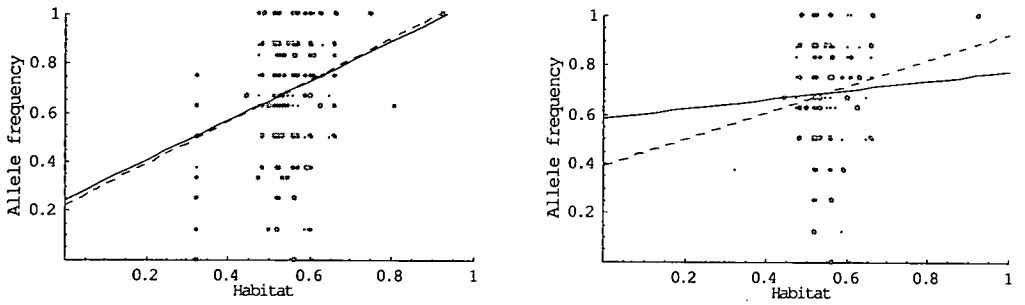
There is a strong association of genotype with habitat in Apahida adults (Figure 3.1) and it can be seen from previous results in this chapter there is copious migration. Therefore if this association is maintained through the field season, after dispersal, then this clearly indicates that adults are expressing preferences for habitats according to their genotype. The association of genotype and habitat after migration is shown in figure 3.56, along with the same relationship before migration. Although the association appears to be approximately half as strong after migration as before, neither regression slope is significantly different from that under random habitat choice. If only those sites in which there is a choice of more than one alternative site are removed then the result is even more striking (the regression after movement is 0.72-0.07x, p=0.89). This clearly indicates that there is no habitat-genotype association created by these migrations. However, if the individuals that remain in their original sites are included too, then the association is both extremely strong and remains almost unchanged (Figure 3.7a)

There is evidence that those members of the population that move site between captures and those that do not move site are qualitatively different. Before any migration has been observed, if habitat preference is entirely absent, it would be expected that there would be no difference in the habitat-genotype association between those that will later be found to have moved site and those that will not. However the regression of genotype on habitat in those that will not move site is significant, which is not true for those that will later move (non-moving 0.18 + 0.92x, $p<10^{-5}$, moving 0.39 + 0.53x, p=0.21). This suggests that a poor fit of genotype and habitat could be a condition causing individuals to move site, which would itself represent a sort of habitat preference. If this is the case then the inability to improve the fit with the habitat through migration would seem

curious. One possible reason why this is the case could be that individuals are actually quite poor at detecting the habitat type they are approaching. Perhaps it is more likely that the habitat they would prefer is not available perhaps as a result of changes in some habitat parameters through the field season. Alternatively, where there are a plethora of available sites, perhaps several are sampled before one is finally chosen.

One plausible explanation for the apparent lack of habitat preference becomes apparent in figure 3.8; those moving site are almost exclusively confined to the sites with the most intermediate habitats (fewer than 1.5% are in sites outside a habitat range 0.4-0.65 after movement, whereas around 5% are in those that do not move).

This use of intermediate sites is significant because they contribute almost none of the explained variation. If, for example, we remove from the analysis all individuals who inhabit sites at their last recapture that have habitat scores less than 0.4 or greater than 0.65 then the regression slope drops from 0.8 to 0.26. This is only true of the limited recapture sample – in adults generally the association is equally strong in the sites of intermediate habitat as the extremes, as described in section 3.1.3). The failure to recapture animals in extreme habitat has several likely causes. Sites of lower habitat score are more difficult to sample successfully and those with higher scores degrade in quality over the course of the experiment (ultimately drying completely), with a result that recaptures in these habitat types are less likely.



(a) (b)

Figure 3.7. The relationship of individual marker allele frequency with site habitat in migrant at first capture (grey point, dashed lines) and at later capture (black points, solid lines). These are shown for all recaptures (a) and only those move site between first and later capture (b). Detail and comparison of regression lines given in Table 3.4.

	First capture	*Later capture*	*First capture (moved only)*	*Later capture (moved only)*
Regression	0.22+0.85x	0.25+0.81x	0.39+0.53x	0.58+0.19x
Significance	P=0	P=0	P=0.177	P=0.475
Test for equal slopes	$F_{1,420}$=0.0279, p=0.132		$F_{1,128}$=0.29, p=0.21	

Table 3.4. Comparison of regressions of individual allele frequencies against the habitat of the sites of captures. Significance of regressions is calculated by a randomisation method.

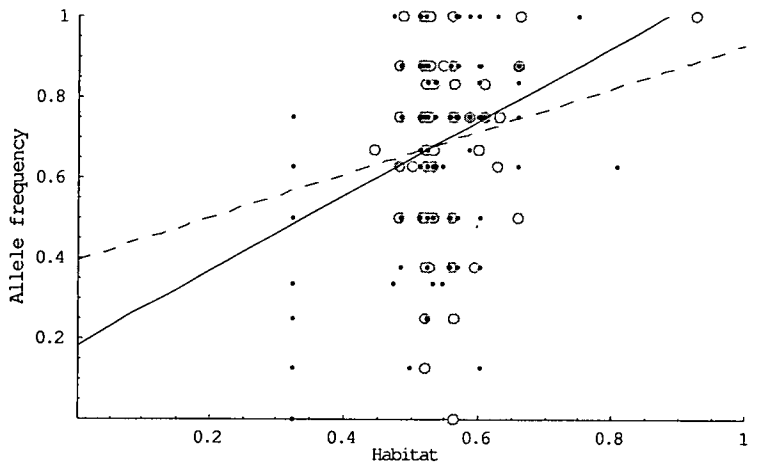


Figure 3.8. The relationship of individual allele frequency and site habitat (of site of initial capture) of individuals later recaptured in different sites (dashed line, grey points) and in the same site (solid line, black points). The regression lines are: move site – 0.39+0.53x $F_{1,64}$=1.577 p=0.214, r^2=0.024; don't move site – 0.18+0.92x $F_{1,144}$=23.3 $p<10^{-5}$, r^2=0.139.

In contrast to these results from Apahida, in the Pešćenica hybrid zone there was clear evidence of an active habitat preference in the movements of adults during a field season (MacCallum *et al.* 1995). A number of migrants were observed moving between two puddle sites and a pond and there was a clear tendency for the more *bombina*-like to move to the pond and *variegata*-like between the puddles (MacCallum *et al.* 1998). A similar analysis to that above would not be possible in this hybrid zone, because due to the clinal structure and the location over a transition from high to low ground both habitat availability and genotype are also determined by geographic location. Repeating the Pešćenica analysis was not possible in Apahida because there were no examples of paired ponds and puddles with migrants between them.

3.4.6 Maximum likelihood methods

As in previous analyses there remains a question of whether philopatry represents a habitat choice in itself. Maximum likelihood methods therefore give parameter estimates

both to those cases in which the individual moved site (model 1) and to all recaptures (model 2). The two estimated parameters are the mean dispersal distance and the habitat preference strength. The latter is presumed to be the same across all individuals. The maximum likelihood values together with 2-Log likelihood intervals (approximating 95% confidence intervals) are given in Table 3.5. These were obtained by a grid search of the parameter space.

	Model 1 – excluding philopatry	**Model 2 – including philopatry**
Mean dispersal distance	41m	26m
	(32-57m)	(22-30m)
Habitat preference strength	3.1	7.8
	(-5.7 – 13.0)	(4.1-11.8)

Table 3.5. Maximum likelihood estimates of dispersal distance and habitat preference under two models of adult site choice with 2-log-likelihood support limits with a habitat detection range of 120m

If habitat preference is constrained to be zero the same dispersal distances are estimated. The increase in likelihood by including the habitat preference parameter is highly significant only for the model including philopatry (Testing against ½ χ^2_1: Model 1 ΔLog-lik=0.25, p=0.76 , model 2 Δ Log-lik=10.92 p=0.0019).

In comparison to the equivalent dispersal distance estimated previously, this method gives a significant increase in estimated mean dispersal distance under both models (the equivalent measures are 30.7m and 7.2m). This difference is particularly noticeable when the non-moving individuals are included as in the likelihood model, non-moving individuals are expected whatever the dispersal distance.

The habitat preference estimated in model 2 is significantly greater than zero. These maximum likelihood estimates would suggest that the preference of a pure *bombina* or *variegata* for sites with the least appropriate habitat is 24% of that for the most preferred habitat type, and the preference for sites with habitat 0.5 is 33% of that for the most preferred habitat. Similarly an individual with half *bombina* and half *variegata* alleles would have a relative preference of 14% for the most pond or puddle-like sites. In the five most puddle-like sites (with habitat scores over 0.95) there is only one individual with allele frequency less than 0.5 and in the five most pond-like sites (habitat types 0-0.107) there are 10% pure *variegata*. Only the former tallies well with the predicted low preference of "pure" individuals for the wrong habitat type.

Vines *et al.* (2003) showed that there were around 20% pure *bombina* in sites of intermediate allele frequency (and about half of this for pure *variegata* in the same sites, unpublished results). The relative preference of pure *bombina* for sites with mean allele frequency 0.5 with the estimated habitat preference is 37%, and 27% for pure *variegata*. Of course the proportion of intermediate populations made up of "pure" animals depends not only on their preference for such sites but also the availability of these individuals and of all habitat types, but it seems encouraging that these predictions seem broadly similar to the observed frequencies.

3.5 Discussion

Estimates have been made from multiply-caught individuals of dispersal distance of adults, the rates of population turnover and the extent of adult habitat preference over a two month period. These can be compared with similar results from the *Bombina* hybrid zone in Pešćenica. From these comparisons we can ask some questions: To what extent do these quantities vary depending on the environmental conditions specific to each hybrid zone? Are the toads themselves different? Can the differences in form between different *Bombina* hybrid zones be due to such differences?

3.5.1 The dispersal distances

Perhaps the most striking difference seen in these analyses between the Apahida populations and those in Pešćenica is that in dispersal distance. From the averages of the distance moved between instances of capture or the maximum likelihood estimates, the Apahida populations are estimated to move ten times shorter distances than in the equivalent measures from recaptures in Pešćenica and the maximum distance moved was twenty times smaller (over a similar length of sampling period). Similarly in another *Bombina variegata* population a third of movements are greater than 150m (Barandun 1995 cited by MacCallum *et al.* 1998), which is greater than the furthest migration observed in Apahida.

To what extent is this due to an intrinsic tendency for shorter dispersal and how much is it controlled by the availability of sites? As noted in the results section there appears to be a difference in the spacing of sites between the two areas, with isolated clumps of sites being further apart in Apahida than in Pešćenica. Sites spaced 100-1000m apart, between which the bulk of the migrations occur in Pešćenica are almost entirely absent in Apahida. The likelihood method explicitly takes the availability of sites into account and although the maximum likelihood estimates of dispersal distance are greater than

those from simple averages, they are barely any closer to the values measured in Pešćenica.

If the aim is to estimate the mean dispersal distance that would be seen in an environment with continuous habitat then the assumptions implicit in the model may reduce its ability to make an accurate estimate. It assumes that the relative proportions of the population found in each site are proportional to the dispersal distribution. This need not be the case with these recaptures. If the toads are aware that there are no sites beyond their local area and remain locally, then the expected distribution of dispersal distances will be much as observed (Koenig *et al.* 1996), merely because there are a greater number of possible short dispersals than long in a spatially enclosed environment.

Philopatry is a strong feature of movement patterns with over half of individuals remaining in the same site between captures. However it is clear that this is not a long term trend as it is uncommon for individuals caught multiple times to be found in the same site repeatedly (where there is a choice). Where there was a choice of sites, no adults were caught in the same sites between years. This does not seem to indicate that the choice of the same site is related to avoiding any costs of movement. The increase in estimated habitat preference strength when including the individuals remaining within the same sites seems rather to support the idea that remaining in the same site is a habitat choice in itself.

It has frequently been noted that the dispersal distance estimated by direct methods is lower than that from indirect estimates, due mostly to the failure of recapture experiments to detect long range dispersal (Murray 1967, Slatkin 1985) or a more general failure to account for the kurtosis of the dispersal distribution (Koenig *et al.* 1996). What is less clear is the effect of habitat patchiness on these measures. Where resources are continuous but also highly clumped, predator-prey dynamics are essentially identical to those in a discontinuous habitat (Arditi and D'Acorogna 1988).

Even if there are useable resources between aquatic sites, it seems likely that the site structure would force different patterns of dispersal on the *Bombina* populations.

In the Polish hybrid zones the dispersal distance was estimated from the levels of linkage disequilibrium to be around 1km per generation (Szymura and Barton 1991). This compares with the annual average dispersal distance in Pešćenica of 929m obtained from recapture data. These clines are geographically distant from one another, are between somewhat divergent races of *Bombina* and the cline in Pešćenica shows differentiation between habitats not seen in Poland. These differences give no reason to expect the dispersal distances to be identical, but it is notable that they are rather close to each other, if there are only one or two years of migration in each generation. The comparable measure in Apahida is considerably lower.

Dispersal distances estimates inferred from cline shape are generally larger than those estimated directly. For example the dispersal distance estimated from the cline shape of the hybrid zone between *Chorthippus jacobsi* and *C. brunneus* in Northern Spain is estimated to be 1.34km per generation (Bridle and Butlin 2002), but direct methods estimate a dispersal distance of only 7-33m per generation (Bailey *et al.* 2003). In this case it is considered that prezygotic isolation between the taxa has increased the indirect estimate of width (Bailey *et al.* 2003). In contrast, in the hybrid zone between chromosomal races of *Sceloporus Grammicus* lizards, dispersal distance estimates that were made indirectly from cline shape and directly from mark-recapture studies are more closely matched (indirect - 160m per generation, direct - ~80m per generation) (Sites *et al.* 1995).

This effect cannot be the cause of the low dispersal distance estimates in Apahida, as they are smaller than both the indirect estimates from Polish hybrid zones and direct estimates from Pešćenica. Although it is possible to infer the migration rate from the hybrid zone structure in Apahida (of the order of 20% of the population per generation, Vines *et al.* 2003), indirect estimates of dispersal distance are impossible due to the absence of a clinal structure.

The dispersal distance measure most commonly used in models of hybrid zone dynamics is not that during a single breeding season but that over a generation (which may be around 5 years). This may comprise the sum of a number of movements and any additional movements that occur between metamorphosis and adulthood. It is likely that the movement rate during the juvenile period differs from that in adults but these have not been directly estimated for juvenile *Bombina* probably up to the age of 2 years, as these are very rarely caught and are more difficult to identify as their belly patterns have not yet stabilised.

Furthermore the dispersal distance during this experiment may actually have been lower than at other times. During other years, which have been wetter, sites have appeared that have had abundant toads but are clearly temporary, such as flooded field, and which have not been seen since (B. Nürnberger and T. Vines pers. comm.). It is conceivable that longer distance migrations take place in wetter years or indeed during rainy periods when there are abundant wet areas and temporary sites.

What is the significance of the isolation of many sites by low migration? Generally a low rate of migration between sites would aid the continued divergence in genotype of different sites in a mosaic by reducing the rate of gene flow between divergent populations. However, a low rate of migration between distant sites does not seem to explain why there are differences in allele frequency between sites on a far smaller scale. For instance there are two sites within the Apahida-Cojocna road sites that are separated by around 1m and that differ in mean allele frequency by 0.26 and within this site group of sites (separated by 110m at most) there is a strong habitat-genotype association and distance will provide little barrier to movement between such sites. Vines *et al.* (2003) estimated a average immigration rate of pure *bombina* into intermediate sites in Apahida as 0.19. Unfortunately with no examples of large ponds connected by migration to intermediate sites, no comparable estimate of migration rate from direct measures is possible.

3.5.2 Population size

It is difficult to draw strong conclusions from the population sizes estimated from the Jolly Seber analysis as the individual estimates have such wide confidence limits. However it seems that the fraction of the population of each site that are caught is rather small. Whether the uncaught animals are hidden in the sites or are in the surrounding land is unknown. The fluctuations in population size and the apparently continuous immigration do give the impression of highly mobile populations.

3.5.3 The association of genotype and habitat - habitat preference and environmental selection

There is a strong association of the mean of individual marker allele frequencies within sites and the site's habitat in both Apahida and Pešćenica (Vines *et al.* 2003, MacCallum 1994, MacCallum *et al.* 1998) and indeed in Apahida this association is much stronger than in Pešćenica (Vines *et al.* 2003). In Pešćenica there is clear evidence that there is active expression of habitat preference that could result in habitat-genotype associations. The evidence in Apahida is mixed depending on whether it can be assumed that individuals remaining in the same site are aware of the habitat of other sites available to them and make a choice to remain in the same habitat.

It is possible that various forms of selection could create all these associations but it is questionable whether selection alone does. The selection strength required to maintain such neutral divergence is quite strong (Vines *et al.* 2003) but there is no evidence of environmental selection acting on the eggs or tadpoles to create this in Apahida (Köhler 2003). In contrast there is clear evidence that there is selection against hybrid eggs and tadpoles in Pešćenica (Kruuk *et al.* 1999a). Although this does not rule out selection being the principal cause of the genetic differences between the populations, the selection would have to act at later stages of development. It seems likely that most

selection would occur at the early stages of development when major changes are occurring that may be particularly vulnerable to environmental conditions. However it is still possible that the patterns of selection change at a later stage perhaps as result of differential adaptation to aspects of semi-aquatic life in different habitat types, or specifically for an amphibian, at the metamorphosis stages at which a further major body rearrangement occurs.

The results from between site migrations give no clear evidence either way. During the two month duration of the 2001 field season around 35% of the adults caught were later found in other sites. Between 2000 and 2001 there was no reduction in the habitat-genotype association (2000 correlation =0.32, 2001 correlation =0.39) despite the evidence of a high rate of migration between sites between the years. Despite such high levels of between site migration during such a short period it is not entirely implausible that environmental selection could create the adult habitat association as sampling is restricted to a small part of each year.

We can get a rough idea of the amount of selection required to restore the original habitat-genotype association. For the correlation between habitat and genotype in the sample after migration between sites, to equal that in the recapture sample before migration requires the removal of the third with the worst fit. If we can assume that this proportion would have to be removed from the population after moving site and assuming that the observed rate of migration is true across the hybrid zone, this implies that around 10% of the population is removed each year by environmental selection.

Can we determine whether the habitat-genotype association seen at the start of the field season is the result of habitat preference? The fact that many sites are completely devoid of animals during the summer by virtue of being absolutely dry, would suggest that the association arises by habitat preference as sites are repopulated. However, it is unknown what animals enter these sites as they reflood and indeed it is unknown whether most sites are depopulated over the winter. Holenweg and Reyer (2000) have shown that a large fraction of individuals of *Rana lessonae* and *R. esculenta* hibernate away from the

pond environment. If the same is true of *Bombina* then the habitat association of different genotypes at the start of the field season reflects the sites chosen by individuals re-entering the sites. Therefore the strong habitat-genotype association seen at the start of the season would represent the result of a behavioural habitat preference and cannot be attributed to ecological selection within the sites.

Although the hibernation behaviours of *Bombina* are largely unknown, it they do frequently choose a different site between years, and if the sites they choose in the following year are not necessarily the same as those from which they left, then the habitat preference at the start of the year is clear evidence of an active habitat preference. Where recaptures had a choice of sites, none were found in identical sites between years. Although there were only a small number of recaptures between seasons

I would suggest that directly following movement patterns of individual toads during the season and during hibernation would seem to address several of the shortcomings of these analyses. For instance Holenweg and Reyer (2000) followed hibernation behaviours by radiotracking individuals as they began to hibernate. Such a finding could directly answer whether the initial site choice in spring shows behavioural habitat preference. A similar experiment during the breeding season could answer important questions about habitat usage and the make-up of mating assemblages.

As they stand these results give an unsatisfying picture of habitat preference and dispersal. This results less from weakness in the results presented here than from a lack of knowledge about a number of behaviours of the toads. If animals are having to make active choices to end up in the observed sites of first capture then habitat preference is stronger than in Pešćenica, but if this is not an active choice then I have shown that there is at most a weak habitat preference.

Implications for mosaic hybrid zone structure

In the light of the difficulties in determining the rates of dispersal between sites and detecting habitat preference, it is difficult to make any general statements about the implications these results have for mosaic hybrid zone structure. Although the high rates of dispersal and strong association of genotype and habitat give the impression that they are an important component of reproductive isolation between the taxa, it cannot be assumed that this is the case. In fact, direct observations of the effect of habitat preference on the genotypic composition of the offspring shows that habitat preference does not cause the association amongst genetically similar animals during mating (chapter 5).

However it must be borne in mind that the these studies were mostly carried out in sites with intermediate to high habitat scores, for purely practical reasons. It may be the case that the habitat preference of the occupants of these sites is of trivial importance and that adults with more pure genotypes do show strong habitat preferences that do translate into the strong association of parental genotypes at mating and reproductive isolation. This implies that either the intermediate genotype animals have a low preference for the "pure" sites or that they are otherwise excluded from successfully mating in them. The linkage disequilibrium seen in these sites (chapter 4) implies that there are some immigrants of intermediate allele frequencies in these sites but that the bulk of the populations are more similar to the pure species.

The high rates of migration from ponds and puddles to intermediate sites strongly suggest that these sites are acting as sources of offspring that feed other sites. It might be that most animals seen in the intermediate populations have undergone relatively few generations of hybridisation and therefore they may still possess associations between loci underlying habitat preference and neutral markers. Further work is needed to clarify this situation. The results of several studies have been unable to determine what causes the habitat-genotype association and this remains one of the most important outstanding questions in research into *Bombina* hybrid zones.

Chapter 4

Quantitative trait variation in the Apahida hybrid zone

4.1 Introduction

Much of the variation observed within and between species is in the form of quantitative traits – those that vary either by degree on a continuous scale such as height or weight, or approximate this by falling into one of many possible discrete categories, such as the number of bristles on a body segment. Relative to simple traits controlled by a single gene, the genetics of polygenic quantitative traits may be opaque.

What patterns do such traits take in hybrid zones? Across smooth clines quantitative trait variation often follows similar patterns to those found in neutral markers. This generally seems to be the case in clines in *Bombina* (Sanderson *et al.* 1992, Kruuk 1997, Nürnberger *et al.* 1995) with exceptions found in only a few traits (Nürnberger *et al.* 1995). This chapter examines the patterns of variation in quantitative traits and associations between them in the Apahida hybrid zone. As the Apahida hybrid zone, with a mosaic structure, differs from all others in which studies of these kinds of traits have been conducted, it is possible that this hybrid zone differs greatly in this respect, which would give some insight into the dynamics maintaining this hybrid zone.

4.1.1 The study of quantitative traits

Statistical methods allow quantitative trait variation to be decomposed into the components due to additive, dominance, epistatic and other genetic components and random environmental deviations (Falconer and Mackay 1995). Advances in molecular

techniques have provided a large enough number of molecular markers to locate and permit the identification of the major genes underlying quantitative variation (quantitative trait loci, or QTL), and determination of the genetic effects of variation in these traits (Lynch and Walsh 1997).

Of course fitness itself, if considered as a trait in its own right, is quantitative in nature (Hartl and Clark, 1995). However, assessing the fitness of quantitative traits is a difficult task. Fitness profiles for various quantitative traits, that is a description of the change in relative fitness with a change in phenotype, have been measured in the lab experimentally (e.g in *Drosophila,* Mackay 1985). There are also methods for the estimation of the strength of selection in natural populations and the strength of indirect selection due to correlations between traits (Arnold and Wade 1984a,b, Lande 1979, Lande and Arnold 1983). However in reviews of natural selection in the wild by Endler (1986) and later by Kingsolver *et al.* (2001) many studies have been shown to be inadequate. One conclusion that may be drawn from this extensive latter review was that, due to the difficulty of accurately measuring selection in the wild, the power of a large proportion of the studies to detect selection was very low and that as a consequence many of the estimates of the strength of selection were not significant.

4.1.2 The genetic basis of species differences in quantitative traits

In comparisons of closely related pairs of taxa it is apparent that individuals may differ in a consistent manner in a host of characters. If a pair of diverged taxa may be successfully crossed, it is possible to make deductions about the genetics of trait differences between the taxa. The techniques of QTL mapping allow the estimation of the number and size of QTL effects and to estimate the chromosomal position of these genes. There are now an increasing number of studies using these methods to describe the genes underlying the differences between species pairs, and between cultivated and wild strains of crop plants. For example there have been studies of the differences and contribution to reproductive isolation of species pairs (*D. simulans* and *D. mauritiana*,

Palopi and Wu 1994, various species of monkeyflowers, *Mimulus*, e.g. Macnair and Cumbes 1989; *Helianthus annuus* and *H. petiolaris* Rieseberg *et al.* 1999) and between strains of crops (maize, *Zea mays mays* and Teosinte, *Zea mays parviglumis* e.g. Westerbergh and Doebley 2002, rice subspecies, *Oryza sativa* Li *et al.* 1997). In some cases these studies also map factors contributing to reproductive isolation that are ecologically mediated (*Mimulus lewisii, M. cardinalis,* Bradshaw *et al.* 1994; *Aquilegia formosa, A. pubescens*, Arnold *et al.* 2002).

Perhaps surprisingly, there aren't any strong patterns of similarity in the genetics of these quantitative trait differences between various pairs of taxa (in terms of QTL numbers, the distribution of QTL effect sizes and interactions between them). However, one common feature of most cases is that a great proportion of the difference between species can be explained by a single or a small number of QTL, particularly where these differences are considered relative to standing phenotypic variation (Orr 2001). What this suggests is that the species differences are due to new mutations of large effect arising in the diverging populations rather than a gradual divergence to alternative optima.

The nature of the selective differences between diverged taxa is also considered by a different kind of study, in hybrid zones. In these cases, whether the isolation arose in sympatry or allopatry may be unknown and potentially the two are indistinguishable (Endler 1977). In many cases these studies have been used to estimate the number of genetic differences with selective significance between the hybridising taxa and the loss of fitness inherent in hybridisation. Furthermore some studies give a different kind of insight into the genetics of these taxa in the form of hybrid strains that outcompete their parents in certain environments, where a lottery of genetic mixing in the hybrid zone has created a novel gene combination that enjoys enhanced fitness (e.g. *Helianthus* Rieseberg *et al.* 1999, *Iris*, Burke *et al.* 1997)

4.1.3 Species differences and the nature of reproductive isolation

When differences between species are observed in conjunction with the reproductive isolation between taxa (present by definition if we are using the biological species concept), it is natural to consider whether these differences play an important role in creating this reproductive isolation. The mechanisms proposed for the action of reproductive isolation vary greatly in their mode of action.

Perhaps the most fundamental division is in the time of action of the reproductive isolation which may be broadly divided into those that act pre or post-zygotically, that is those that reduce the rate of successful matings between groups or reduce the success of their hybrid offspring. These will further differ in whether the presence of isolation depends on the state of the external environment or is relatively unchanging in different environment (respectively extrinsic and intrinsic isolation). Prezygotic isolation is discussed in chapter 5.

The formation of intrinsic post-zygotic isolation in allopatry is most commonly considered to proceed along the lines of a model described by Dobzhansky (1937) and Muller (1942). This model explains how an incompatibility can arise between two loci without ever being selected against in the population in which it arises. It supposes that a novel allele is fixed at one locus in one population and another new allele at a different locus in a separate population; these new alleles are selected against when in the genetic background of the other population. This model is also attractive as, once an initial incompatibility forms, each new incompatibility adds to the number of available targets with which changes in the other population may clash, and hence the number increases by a "snowball" process (Orr 1995), to ever increase the strength of isolation.

4.1.4 Quantitative traits in hybrid zones.

Hybrid zones may frequently be recognised from the observation of phenotypic hybrids between taxa. The phenotypic differences between populations are often seen to vary in smooth clines from one population to another. It has been observed in many cases that clines in different traits are highly concordant and coincident (Barton and Hewitt 1985) i.e. that there are found in the same location and have the same width. Such clustering of clines is evidence that there is selection against hybrids *per se* as linkage disequilibrium between the loci under selection and quantitative trait loci acts to pull multiple clines together (Nürnberger *et al.* 1995). If selection were against hybrid phenotypes then the clines might be expected to vary in width and position depending on the selection on particular traits. Displacement of clines may therefore be taken as evidence of greatly varying selection on different traits. An example of this is in the hybrid zone between *Chorthippus parallelus* and *C. erythropus* in the Pyrenees, where clines in many morphological traits are displaced and show variation in width (Hewitt 1993) suggesting very different selection pressures.

One interesting example of a rather different form of selection acting against intermediate phenotypes is found in various species of *Heliconius* butterflies (e.g. Mallet 1986, Mallet *et al.* 1990, Jiggins *et al.* 1996). In these cases pure species or races belong to different mimicry rings but hybrids between them carry intermediate warning colouration patterns. The effectiveness of warning pattern is density-dependent, as the more frequent a pattern is the more likely it is to be learned by predators. The hybrids within intermediate phenotypes are therefore more likely to suffer strong predation as their individual patterns are rare.

4.1.5 Phenotypic differences between Bombina bombina and Bombina variegata

There are numerous phenotypic differences between *Bombina bombina* and *B. variegata* in morphology, development and behaviour, including differences in the ratios of the sizes of various body parts, skin thickness, texture and aposematic colouration, mating calls and mating behaviour, egg volume, egg development time, tadpole development and various others (Michałowski and Madej 1969, Szymura 1993, Nürnberger *et al.* 1995, Lörcher 1969, MacCallum 1994, Kruuk and Gilchrist 1997, Vorndran *et al.* 2003).

Clines in many traits are concordant with those in molecular markers. In the Cracow transect, mating call components were concordant with allozyme allele frequency clines (Sanderson *et al.* 1992). In the Pešćenica transect, the same was found of belly colouration, mating call cycle length, skeletal proportions, egg volume, skin thickness and femur length (Nürnberger *et al.* 1995, Kruuk 1997). One exception was in larval development time where clines were displaced (Nürnberger *et al.* 1995) suggesting strong selection on this trait independent of any selection on other loci.

Differences in these traits between the taxa may have significance for the adaptation to the habitats in which each is found, when they occur together in hybrid zones. Other trait differences may have a role in pre- or post-zygotic isolation. However without empirical support such differences are only conjecture, and it is also possible that many of the traits have no role in creating isolation or the differences are of no selective importance, at least in the habitat in which both species are found together.

Experimental evidence strongly suggests that there are selective differences between the taxa during the egg and larval stages of the life-cycle. Development rate differences, anti-predator differences and adaptation to different thermal environments observed

between the taxa (Kruuk and Gilchrist 1998, Nürnberger unpublished data, Vordran *et al.* 2003) all appear to vary in ways appropriate to the habitats in which each of the taxa is typically found in hybrid zones. Selective differences between genotypes have been directly tested in the stages from egg to tadpole (in both laboratory and natural conditions) by a cohort analysis, but there was no strong evidence of intrinsic or extrinsic selection varying between genotypes (Köehler 2003).

4.1.6 Associations between traits in hybrid zones

Linkage disequilibrium between genetic markers has commonly been reported in hybrid zones (e.g. Kocher and Sage 1986, Rand and Harrison 1989, Szymura and Barton 1991, Mallet *et al.* 1990, Mousseau and Howard 1998, Bridle *et al.* 2001a, Dasmhapatra *et al.* 2002, as a selection from many such studies). The association of parental combinations of markers may be taken as evidence of mixing of parental populations and selection against hybrids. The linkage disequilibrium created by mixing between sites of differing allele frequencies has been modelled by Li and Nei (1974) and Prout (in Mitton and Kroehn 1973). Barton and Gale (1993) give an estimate of the equilibrium linkage disequilibrium generated by mixing between two sites and Kruuk (1997) for the linkage disequilibrium in a central site receiving migrants from two donor sites.

The linkage disequilibrium between markers may be reflected in the allelic state of quantitative trait loci. As a result the trait value of individual quantitative traits will tend towards that of the parental populations. When linkage disequilibrium exists widely through the genome this will be equally true for all quantitative traits and hence individuals will tend have the phenotype of one parent population in all traits. This will be observed within subpopulations as a covariance between trait values across traits.

However there are alternative explanations for the patterns of linkage disequilibrium. For instance assortative mating between parental types and epistatic selection for

parental combinations of alleles could also create linkage disequilibrium (the role of assortative mating in hybrid zones is considered in the next chapter).

Epistatic selection in hybrid zones could act to prevent the break-up of co-adapted gene complexes in each parental population. Epistatic interactions in hybrid unfitness have been reported from lab crosses (e.g. Palopoli and Wu, 1994, Rieseberg *et al.* 1996, Li *et al.* 1997, Orr and Irving 2001, Burke *et al.* 1997) suggesting that there are co-adapted gene complexes in these cases. Fritz *et al.* (2003) give an example where the increased susceptibility of hybrids of *Salix eriocephala* and *S. sericea* willows to herbivory was a combination of breakdown of two different trait types, controlling resistance to herbivory and herbivore host recognition cues. In natural hybrids the fitness advantage of some hybrid combinations (e.g. Rieseberg *et al.* 1999, Grant and Grant 1997) and examples of "bounded hybrid superiority" (Moore 1977, Moore and Price 1993, Good *et al.* 2000, Arnold 1997) suggest that certain gene combinations that arise by hybridisation can have a fitness advantage.

Distinguishing the various causes of linkage disequilibrium in hybrid zones may be rather difficult. Any combination of effects may have contributed to the associations and these are not immediately apparent from the distribution in the hybrid zone and must be tested for separately. For example the effects of assortative mating can be estimated from the genotypes of hybrids in zygotes before selection (Nürnberger *et al.* in press). Models of generation of linkage disequilibrium by migration could give predictions of the linkage disequilibrium, if the migration rate and the allele frequency difference between the source populations are known. Therefore, with detailed knowledge about migration patterns, it may be possible to determine if the observed linkage disequilibrium is consistent with the predictions of these models but such details may be difficult, if not impossible, to obtain. In only one study is there clear evidence of linkage disequilibria in hybrids, not created by mixing between populations. Gardner *et al.* (2000) used a genetic mapping approach to study hybrids of *Helianthus petiolaris* and *H. annuus*, and after accounting for the effect of generational structure (i.e. the maintenance of blocks of parental genome type) they found significant epistasis between

many pairs of loci, principally those involved in fertility. Such an analysis requires a large number of diagnostic loci, which is a limiting factor in many systems.

4.2 The aims of this study

Studies of hybrid zones in *Bombina* from Cracow in Poland and Pešćenica in Croatia have found that clines in a number of phenotypic traits are highly concordant and there are strong linkage disequilibria between them in the centre of the cline (Sanderson *et al.* 1992, Nürnberger *et al.* 1995, Kruuk 1997). The maximum magnitudes of these linkage disequilibria, in the larger study of wild-caught animals (Kruuk 1997), are not significantly different from that between genetic markers, leading to the conclusion that these were generated by mixing of immigrant adults in the centre of this cline. In the earlier study of Nürnberger *et al.* these estimates were approximately half of those between markers. This chapter aims to compare these patterns with the equivalent measures in the Apahida *Bombina* hybrid zone.

The different structure of this hybrid zone, in comparison to that studied in Pešćenica, may promote a greater effect of selection. This is because there are no large populations of "pure" parents feeding individuals with parent-species type genotypes into the hybrid zone every generation. Therefore it is likely that there have been more generations since the original immigration of the parental species into the hybrid zone and therefore more time for parental genotype combinations to have been broken apart by recombination. This is observed as a greater proportion of hybrid individuals in Apahida than in Pešćenica (Vines 2003). Also different gene combinations may be formed as there is a greater chance for mixing between populations that are very different in genetic composition. Selective differences between traits may be revealed as the new gene combinations are formed and tested by selection.

Furthermore, in a mosaic hybrid zone, reinforcement may be more effective (Cain *et al.* 1999) and therefore pre-zygotic isolating mechanisms, which may underlie the habitat associations (this is discussed in chapter 5), may become more effective. If any quantitative traits are associated with habitat choice or environmental selection then this too may be revealed as a pattern different from than in other traits.

By examining and comparing the patterns of various traits and molecular markers in this hybrid zone it may be possible to determine if there are major differences in selection between these populations and those previously studied and whether these are important factors resulting in the different structures.

4.3 Methods

4.3.1 Quantitative traits used in this analysis

Of the trait differences between the taxa described above, three are used in this analysis. These are the ventral colouration pattern, tibia length and egg size. Except for egg size, these have been described in detail in chapter 2. Molecular marker genotypes at four loci are available for 1007 toads from 96 sites, spot scores for 2313 toads from 143 sites, leg lengths for 2281 toads from 143 sites and 21 measurements of egg batches from 14 sites. Habitat scores have been calculated for 137 of these sites.

The measure of egg size used is the length of an embryo along its longest axis. As the size of the embryo increases with time the eggs were measured at the same stage of development. The progress of their development was assessed according to the scheme of Gosner (1960). All measurements were taken at the 19^{th} development stage. This is later than ideal as the embryos are no longer completely spherical and so there will be variation introduced into the measurements by the amount of curvature there is in individual embryos but this was necessary in order to include as many eggs as possible. This later stage does have the advantage that it is easier to accurately assess the state of development (personal observation). The use of this measure of egg size was used rather than the pre-gastrula measurement used in experiments in Pešcenića (e.g. Nürnberger *et al.* 1995) as eggs were often developed past this stage when they were returned to the laboratory.

4.3.2 Methods of data collection

Adult collections

Adult toads were caught during the spring of 2000 (by Tim Vines and Sonja Köhler) and 2001 (by Sonja Köhler, Thomas Alfert, Lino Ometto and myself) from a large number of distinct water bodies. The leg length and spot score measures were taken *in situ*, egg measurements in the laboratory in Babeş-Bolyai University in Cluj-Napoca, Romania. The toe clippings were later analysed for genotype at four unlinked genetic marker loci (details of loci and genotyping protocols are given in chapter 2, all genotyping was conducted by Tim Vines or Sonja Köhler for 2000 and 2001 samples respectively. The proportion of *Bombina variegata* alleles at all the markers scored for each individual is referred to as the individual's mean allele frequency and the average over all individual means within a site is referred to as the site mean allele frequency.

Egg collections

All aquatic sites were regularly searched for eggs, which are usually found attached to blades of grass or reeds. Where batches of eggs were found, 5-10 eggs from each batch were collected in small glass and plastic vials with a small volume of water. As the recapture study in chapter 3 shows, there are large populations within most sites and given the high turnover of adults within seasons these batches have been assumed to be from different parental pairs, though there is evidence this may not be the case (chapter 4, Köhler 2003). These vials were then kept in a container of water to try to maintain a low temperature and reduce the rate of development before analysis in the laboratory.

During the same day the eggs had their developmental stage assessed according to the scheme of Gosner (1960) by observation under a binocular dissecting microscope and their total length measured using an eyepiece graticule (see chapter 2). The tadpoles were then transferred to plastic cups and kept alive in the laboratory for around 10 days. This allows all tadpoles to grow to the same development stage and be measured and

also to grow in size sufficiently to allow reliable DNA extraction and allow confirmation of the species.

It is quite difficult to determine in the field which species laid any particular egg batch as there are no completely distinctive morphological differences between the eggs of several of the species native to this region (particularly *Hyla arborea*) and identification relies principally on the number of eggs and the shape of egg batches. At later stages of development, post-hatching, species differences in morphology become quite apparent. During the growth of the tadpoles their developmental stage was checked every day or two and length measurements taken.

4.3.3 Methods of analysis

One way to determine if individual traits are behaving differently from others is determine if different traits vary in unison with one another. In a clinal hybrid zone clines in separate loci or traits can differ in two ways: by having a different width or a different location. In a mosaic hybrid zone it is possible to make a similar determination, but with some differences as sites of varying trait values do not change in a simple geographic pattern. Instead, the mean value of each trait is compared with respect to the mean of others in the same sites. Where a trait varies over its full range of observed values and a comparable trait varies over less than its full range in the same sites, this may be considered as equivalent to the first trait having a narrower cline than the second.

The concordance of individual traits is therefore tested by fitting a regression of mean trait value within sites against mean *variegata* allele frequency or the mean value of another trait. This regression is fitted by least-squares, weighted by the sample size within each site. Linear, quadratic and cubic terms are fitted and the best-fitting regression is determined by an F-ratio test of explained to unexplained variation. A similar regression is performed for variance in trait within each site. For egg size, the measure used is mean size of all eggs within a single clutch averaged over all clutches within each site.

The best fit regressions can further be used to make predictions of the trait mean and variance values in sites with mean allele frequencies of 0 and 1 as there are no such sites within the studied area. Although there are examples of sites from nearby regions fixed for alleles of one or other taxon at marker loci, these cannot be used as a measure of the trait value in "pure" sites within the hybrid zone as there is no evidence that migrants from these populations contribute to the hybrid zone or that these are representative of pure populations within the hybrid zone.

4.3.4 Sex differences

It is not guaranteed that the sexes will be under similar selection or other pressures, and therefore the distribution of these traits may differ in the sexes. Male marker allele frequency is regressed on female within the same site to determine if there is a difference between distribution of male and female genotypes. To determine if the same pattern of quantitative traits is observed in males and females, the above regressions are repeated separately for each sex. The equality of slope was tested.

4.3.5 Habitat associations

The habitat type of a site explained a large amount of the variation in marker frequency across sites (Vines 2003). Also as there is a strong correlation between the type of habitat and the mean allele frequency of the occupants (Vines *et al.* 2003), these regressions are repeated with respect to the site habitat.

4.3.6 Association between traits

Linkage disequilibrium describes the association in populations between the alleles at separate loci. By the most commonly used definition, linkage disequilibrium is the covariance in state between pairs of loci (Hartl and Clark 1995). The QTL underlying

different quantitative traits may also be in linkage disequilibrium. However without knowledge of these loci and without a means to detect their allelic states, the strength of any linkage disequilibria cannot be estimated. However if the linkage disequilibria are between alleles that affect both traits in the same direction then this will be observed as a positive covariance between the traits. Under certain assumptions about the genetics of the traits, Sanderson *et al.* (1992) showed that it is possible to estimate the average strength of linkage disequilibrium between pairs of loci controlling different traits. This makes the assumption that the loci that affect both traits are independent of each other, no loci have a pleiotropic effect on both traits and that any environmental deviations are independent (Nürnberger *et al.* 1995).

The phenotypic value of two traits, i and j, are defined as z and z' respectively. It is assumed that that these phenotypes are determined by the additive effects of sets of n and n' genes and by random environmental deviations. The allelic state of each locus is defined by an indicator variable x, and the positive genotypic effect α. The phenotype results from the sum of these effects and also an environmental deviation ε:

$$z = \sum_{i=1}^{n} \alpha_i x_i + \varepsilon$$
$$z' = \sum_{j=1}^{n} \alpha'_j x'_j + \varepsilon'$$

Therefore, as the environmental deviations are assumed to be independent, the covariance between the traits is entirely due to the covariance between the allelic states:

$$\mathrm{cov}(z, z') = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \alpha'_j \, \mathrm{cov}(x_i, x'_j)$$

As the covariance is assumed to be entirely due to genetic factors, i.e. linkage disequilibrium between the loci, $\mathrm{cov}(x_i\, x_j) = D_{ij}$ then:

$$\mathrm{cov}(z, z') = \sum_{i=1}^{n} \sum_{j=1}^{n'} \alpha_i \alpha'_j \, D_{ij}$$

Clearly as we know nothing about the individual underlying loci then the factors α_i and α_j will be unknown. Therefore these have to be replaced by measurable quantities to estimate the linkage disequilibrium. However, we can estimate the effect of all the loci underlying each trait if we assume that in the pure parental populations all these loci fixed for the same species allele at every locus. In this case we can estimate the average linkage disequilibrium between pairs of loci. These values, Δz and $\Delta z'$ can be estimated by extrapolation of the regression of the trait value on site mean allele frequency to mean allele frequencies of 0 and 1. To estimate the disequilibrium in subsets of the populations Δz and $\Delta z'$ represent the part of the regression of trait on mean allele frequency over the range of site mean allele frequencies sampled, Δp.

As $\Delta p=1$ if the trait difference alleles are fixed across the hybrid zone:

$$\Delta z = \sum_{i=1}^{n} \alpha_i 2\Delta p = 2\sum_{i=1}^{n} \alpha_i$$

And therefore:

$\text{cov}(z,z') = \frac{1}{2}\Delta z \Delta z' E[D]$

$$E[D] = \frac{2\text{cov}(z,z')}{\Delta z \Delta z'}$$

If these assumptions are met, it is possible to estimate the mean pairwise linkage disequilibrium between loci controlling the quantitative traits from the estimates of the covariance between traits. I consider it likely that the assumptions will be met in this case on the grounds that the genes underlying the development of pigmentation and leg length are likely to share few developmental similarities.

Although it not clear to what extent these traits respond to environmental differences, it is perhaps more likely that there is a common environmental effect e.g. developing in

warm water could conceivably lead to longer legs and more belly pigmentation causing non-genetic covariance of traits. It is clear that such an effect would also reduce any covariance between marker genotype and traits but leave covariance between markers unchanged as genotype can clearly not respond to environmental differences. I consider this unlikely given the results from a similar analysis in the Pešćenica cline show a close coincidence between values of linkage disequilibrium between trait pairs, marker pairs and between traits and markers (Kruuk 1997). It therefore seems highly improbable that quantitative traits were greatly affected by environmental differences in this cline and, if the genetic control of these traits remains similar in Apahida, environmental differences will have little effect there either.

To test the significance of these estimates, and of differences between the estimates for different pairs of traits, linkage disequilibria are calculated by maximum likelihood. The distribution of a p x p covariance matrix estimated from a sample of a known size is described by a Wishart distribution (Kendal 1980). For a pxp covariance matrix, as estimated between two traits, and given the estimates of the variance in the traits in the parental populations and the difference in traits between the parental populations, the expected covariance in intermediate populations can be calculated (assuming linkage disequilibrium is the cause of the covariance). Integrating over the range of possible variances gives an expression for the likelihood of a value of D (N. Barton pers. comm.). The maximum log likelihood value is the calculated numerically and also the values of D giving a drop of 2 log likelihood units which approximates a 95% confidence interval of the estimate of D (Mangel and Hilborn 1997). Values are considered to be significantly different from zero if the confidence intervals do not encompass zero and estimates from different traits are considered significantly different if the confidence intervals of both traits do not overlap.

The average linkage disequilibrium between pairs of marker loci can be estimated from the variance in "hybrid index" (the sum over all loci of the number of alleles characteristic of one population) as the association of these alleles (i.e. linkage disequilibrium) leads to an increase in the variance within a site over that under Hardy-

Weinberg expectations (Barton and Gale 1993). This average value of all pairwise linkage disequilibria is given below:

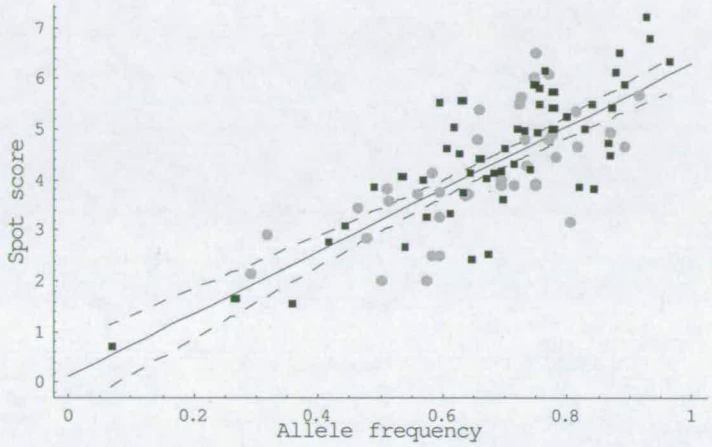
$$\overline{D} = \frac{\mathrm{var}(HI) - 2\sum_{i=1}^{L} p_i q_i}{2L(L-1)}$$

where L is the number of loci, and $p_i = 1 - q_i$ is the allele frequency at the i^{th} locus (Kruuk 1997).

4.4 Results

4.4.1 Site means and variances of traits

(a)Spot score



(b) Log leg length: body length ratio

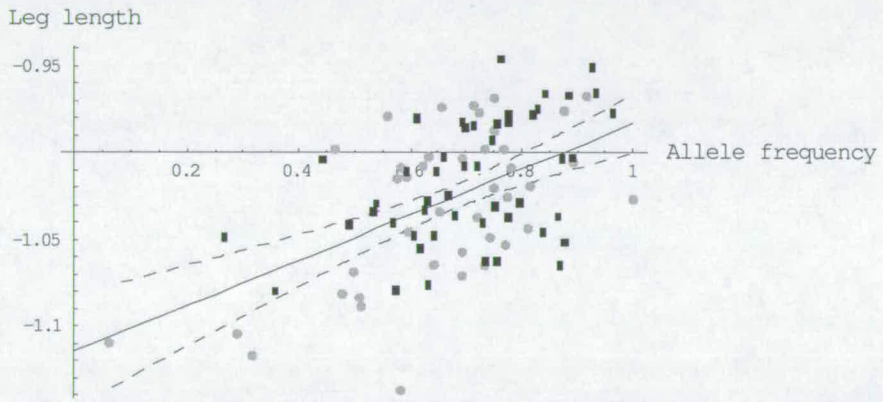


Figure 4.1. Regressions of quantitative traits on site mean allele frequency

Continued overleaf.

(c)Embryo length (mm)

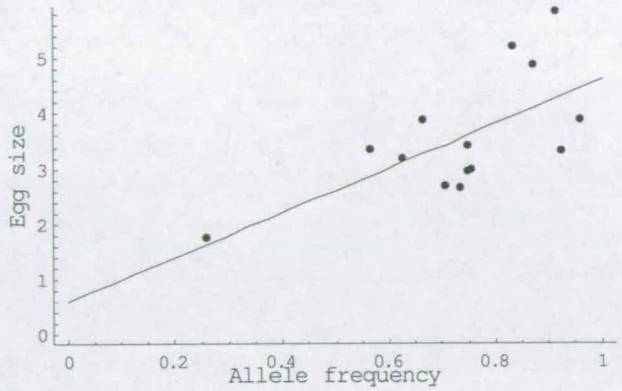


Figure 4.1. continued. Regressions of quantitative traits on site mean allele frequency:

(a) Spot Score. The regression line is 0.12+6.15x, $F_{1,90}$ =151 p<<0.001, r^2=0.627.

(b) Leg length The regression line is -1.115+0.131x, $F_{1,88}$ =35 p<<0.001, r^2=0.257.

(c) Egg size. The regression line is 0.61+3.99x, $F_{1,12}$ =9.2 p=0.0104, r^2=0.434.

Black points are sites with ≥10 individuals, grey <10

For all traits the site mean phenotypic value increases as a straight line function of site mean allele frequency at the genetic markers. Plots of these relationships together with best fitting regression lines and significance values are shown in figure 4.1. All relationships are highly significant and in all cases a straight line fit is more significant than either a quadratic or cubic fit. This includes data for egg length, for which there are a small number of sites and very limited data for each site.

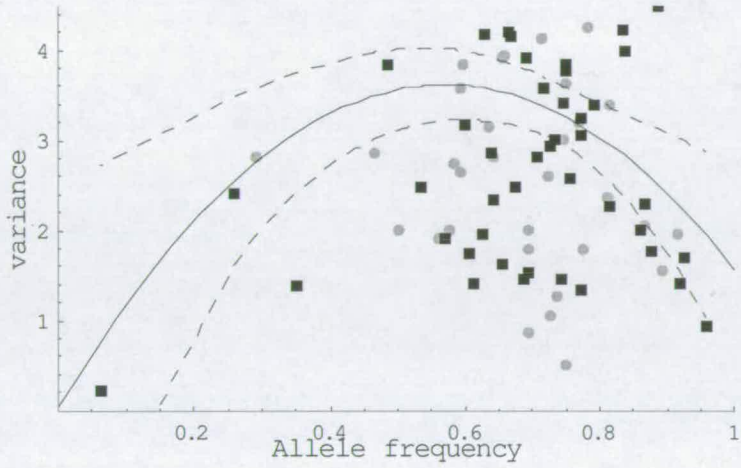
This indicates very strong concordance of quantitative traits with differences in mean allele frequency indicating that all traits follow the pattern of mean allele frequency. This does not hold for the regression of leg length on spot score which is not significant (Spot=-1.05+0.013 Leg, p=0.184). Although this could reflect a different relationship of these two traits it is perhaps more likely a refection of the greater variance within these traits.

Only one direct comparison with estimates from Pešćenica is possible. The mean

difference in spot score value across the hybrid is very similar (5.7 in Romania, 5.2 in Croatia). Leg length and egg size measures are not directly comparable as they are based on somewhat different measures of these traits (relative tibia length and egg volume at the pre-gastrula stage were measured in Pešćenica). However the relatively shorter legged animals have femurs roughly 87% of the relative length of the longer-legged in Pešćenica and the shorter tibias are a very similar 88% of the length of the longer tibia in Romania.

The concordance of individual marker loci with others can be assessed by regression of site mean allele frequency of individual loci against the mean of all other loci. Vines (2003) showed that there were few deviations from complete concordance between loci in the Apahida hybrid zone. However these methods are inappropriate for use with quantitative traits; there is variation in quantitative traits even in individuals fixed for one allele at all marker loci. Therefore the trait is assumed to be non-concordant with the markers if the regression of trait upon allele frequency is significantly non-linear. This is not the case for any of these traits (Figure 4.1).

(a) Spot score



(b) Leg length

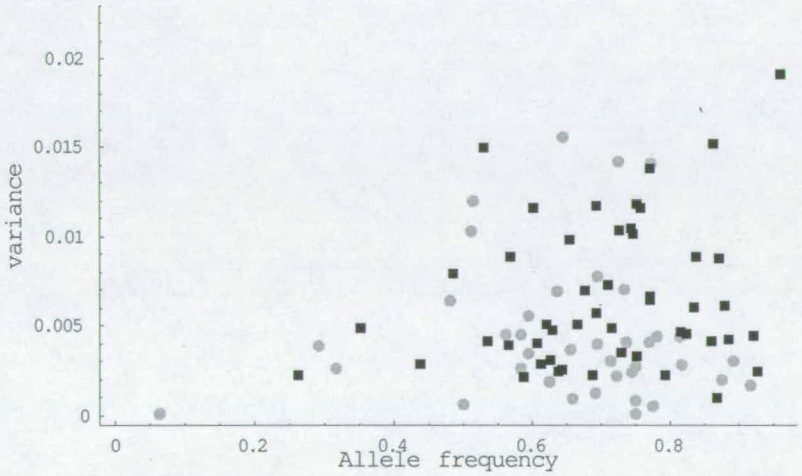


Figure 4.2. Regression of site variance in traits on site mean allele frequency. Continued overleaf.

(c) Embryo length

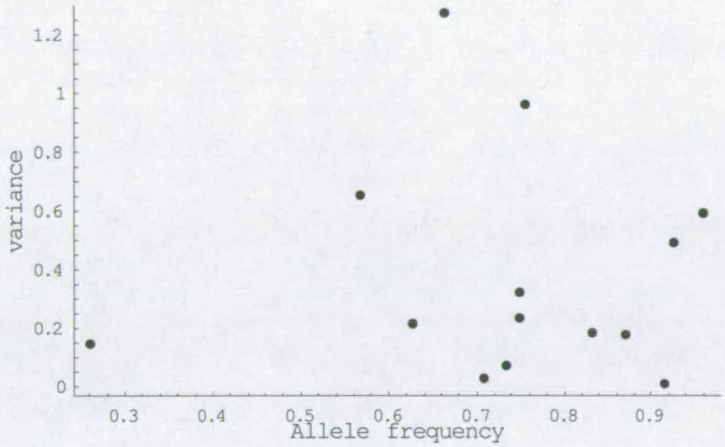


Figure 4.2. Regression of site variance in traits on site mean allele frequency. Black points are sites ≥10 individuals, grey <10:

(a) Spot score. The regression line is $0.030+12.58x-11.04x^2$, $F_{2,89}$ =5.59 p=0.000516, r^2=0.111.

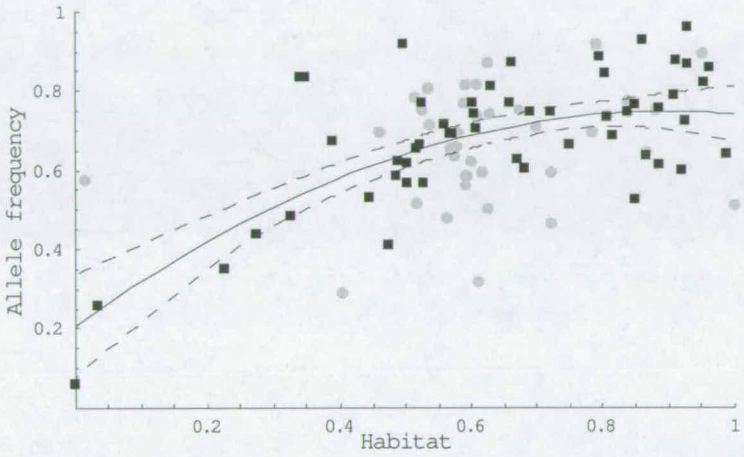
(b) Leg length. There is no significant regression ($-0.015+0.13x+-0.23x^2+0.13x^3$, p=0.35, r^2=0.037)

(c) There is no significant regression ($-0.23+2.19x-1.74x^2$, p=0.77, r^2=0.046)

The best fit of variance in mean spot score with respect to mean allele frequency is a quadratic with the peak variance slightly towards the variegata end of the mean allele frequency spectrum. The variance in the most *bombina*-like sites is nearly zero, but there is almost half the maximum variance in the most *variegata*-like sites. That there is little variation in *bombina*-like sites is not surprising, as the spot score is almost zero in these sites. In sites with high mean allele frequency the mean spot score is only 5.6, so there are a greater variety of possible spot patterns available. The pattern for leg length is less clear; there is no significant regression through individual sites. There is also no significant regression of variance in egg length with respect to site allele frequency.

4.4.2 Trait changes with habitat

(a)
Allele
frequency



(b)
Spot score

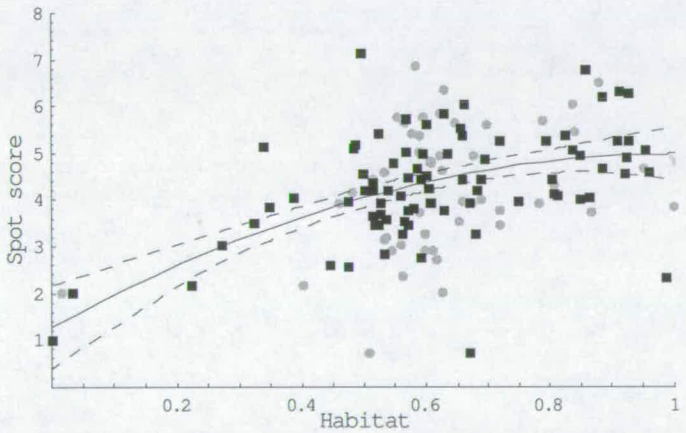
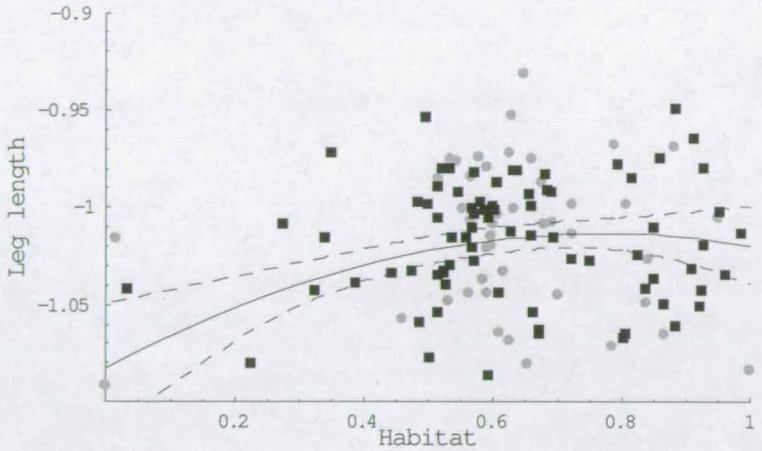


Figure 4.3. Regressions of marker allele frequency and quantitative traits on site habitat.

(c)

Leg length



(d)

Embryo length

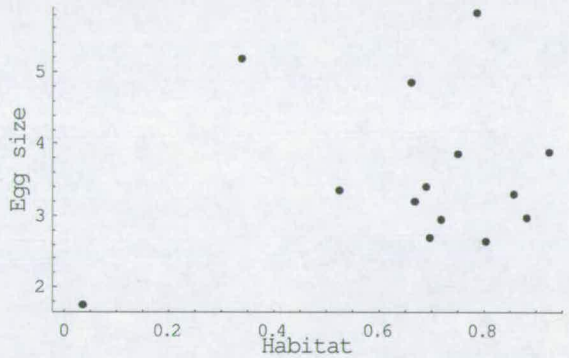


Figure 4.3. Contd. Regressions of marker allele frequency and quantitative traits on site habitat. Black points are sites ≥10 individuals, grey <10:

(a) Mean allele frequency. The regression line is $0.205+1.21x-0.678x^2$, $F_{2,89}=34$ $p<<<0.001$, $r^2=0.432$.

(b) Spot score Regression line is $1.26+7.40x-3.69x^2$, $F_{2,128}=33$ $p<<<0.001$, $r^2=0.344$.

(c). Leg Length Regression line is $-1.08+0.179x-0.116x^2$, $F_{2,124}=7.84$ $p=0.0006$, $r^2=0.112$.

(d) Egg size There is no significant regression $p=0.50$, $r^2=0.037$

The regression of site mean allele frequency or trait means against habitat is significant and positive in all cases except for egg size (figure 4.3). All the significant regressions show a very similar relationship. The slope of the regression in sites of lowest habitat score, if extrapolated as a straight line, has a slope approximately equivalent to that seen in the regression of trait on mean allele frequency. However the slope decreases from around the middle of the habitat range to around 0, indicating no difference in mean trait value across this range of habitats. In all cases the difference in mean trait value between

the two most extreme habitat types predicted by these regression is approximately half that between sites fixed for one or other allele type (predicted from of trait on site mean allele frequency)

4.4.3 Traits in males and females

There is a strong pattern of differences in trait value between males and females. In all cases the regression of mean female on mean male trait value has a slope significantly lower than 1, expected if the relationships are equivalent (Table 4.1). The regression of female on male allele frequency in the same sites shows a slope of 0.57 (Figure 4.4). Though this represents a consistent tendency for female allele frequencies to change less between sites than male, it is not symmetrical around the allele frequency 0.5; in the sites where males are fixed for *bombina* alleles females are have an average allele frequency of 0.30 and in those fixed for *variegata* allele in males, female allele frequency is predicted to be 0.88 and the allele frequency at which the mean allele frequencies are the same is 0.70. The sex difference indicates either that the cline in female allele frequency is much wider than that of males or that there is more variability in female allele frequency within sites. The deviation of males and females is repeated in other traits too: in spot score and leg length (Table 4.1) the regression slope of female on male trait is around 0.5.

	Allele Frequency	Spot score	Leg Length
Regression slope	0.57	0.48	0.59
Deviation from slope 1	$T_{81} = -4.38$, p=0.00017	$T_{121} = -7.53$, p<<0.0001	$T_{119} = -4.95$, p<0.0001

Table 4.1. Results of regressions of female on male mean traits within sites and tests of the significance of deviations from an equivalent relationship

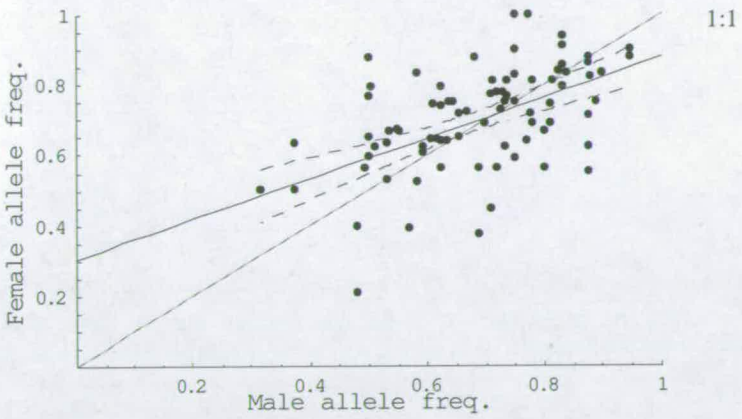
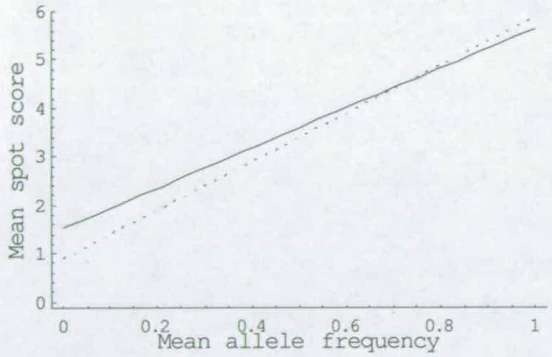
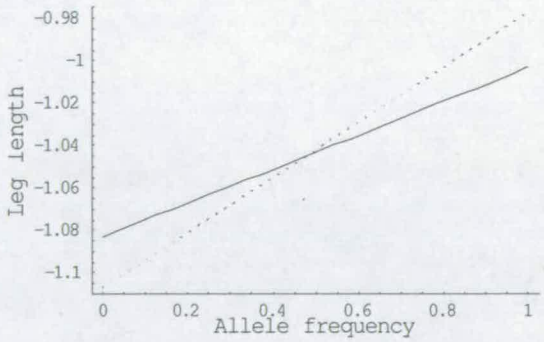


Figure 4.4. Regression of female mean marker allele frequency on male mean marker allele frequency. Regression line is 0.304+0.572x, $F_{1,79}$ =34.5 p<<<0.001, r^2=0.30. Also marked is the dotted line representing a 1:1 relationship of male an female trait values.

The differences between the sexes are also demonstrated as deviations of the relationships of trait means on allele frequency: there is a significantly steeper regression of male traits on allele frequency than female (Figure 4.5, Table 4.2). Allele frequency explains a much higher proportion of the variance in male trait compared to female traits, suggesting that it is stronger variability in females that is generating the apparently wider cline in female traits than male.

In Pešćenica, there were slight differences between male and female cline width (in allozyme frequency) but these were not nearly so marked and were in the opposite direction indicating a narrower female cline, which it was suggested may have been due to greater male dispersal (Kruuk 1997).





Spot score Leg Length

Figure 4.5. Regression of female mean trait (solid line) and male mean trait (dashed line) against site mean marker allele frequency for the traits spot score and leg length. Regression lines are:

Spot score: female - 1.37+4.47x, $F_{1,82}$ =43 p<<<0.001, r^2=0.34; male - 0.10+6.16x, $F_{1,82}$ =171 p<<<0.001, r^2=0.67

Leg length: female: -1.099+0.110x, $F_{1,81}$ =13.7 p=0.0038, r^2=0.145; male: -1.116+0.142x, $F_{1,81}$ =58.1 p<<<0.001, r^2=0.417

	Spot score	Leg Length
Regression slope (female, male)	4.47 6.16	0.110 0.142
Significance of differences in slope and intercept	$F_{1,1685}$ = 20.0, p<<0.0001 $F_{1,1686}$ = 1951, p<<<0.0001	$F_{1,1678}$ = 4.66, p=0.0309 $F_{1,1679}$ = 669, p<<<0.0001

Table 4.2. Regressions of traits on site mean marker allele frequency for females and males and the significance of differences between the regressions.

4.4.4 Disequilibrium between traits

Estimates of linkage disequilibrium were calculated for groups of sites binned by the mean allele frequency of their occupants. For all traits this shows a unimodal distribution with a maximum in sites of intermediate allele frequency (Figure 4.6). If it is assumed that allele frequencies of trait loci are the same as the site mean marker allele then these estimates can be standardised ($R=D/\sqrt{pqrs}$, where p=1-q and r=1-s are the allele frequencies at the two loci) to show these as a proportion of the maximum possible linkage disequilibrium. The maximum values of R remain in the same sites as the maximum values of D. This is a pattern that mirrors the patterns in linkage disequilibria between markers (in Croatian, MacCallum *et al.* 1998, and Romanian studies, Vines *et al.* 2003) and between other quantitative traits in Pešćenica (Nürnberger *et al.* 1995, Kruuk 1997).

The maximum value of disequilibrium between markers in this analysis was 0.172, which is comparable to that observed in Pešćenica of 0.16 (Kruuk 1997). These results are increased by the heterozygote deficit within sites, and are correspondingly greater than those estimated by a maximum likelihood method that attempts to remove this effect (Pešćenica: D=0.093 MacCallum *et al.* 1998, Romania: D=0.077 Vines *et al.* 2003).

The maximum values of disequilibrium between markers and spot score and leg length give very similar values which do not lie outside the other's confidence intervals in any instance (figure 4.6). The pattern observed between spot score and leg length is somewhat different. Although the observed values of linkage disequilibrium in the sites of intermediate allele frequency are approximately equivalent to those in other comparisons, in sites of lower and higher mean allele frequency the estimated values do not decrease (although the confidence intervals in the sites of lower mean allele

frequency are extremely wide). In sites of allele frequency 0.7 and above this difference is significant but not in sites of low allele frequency.

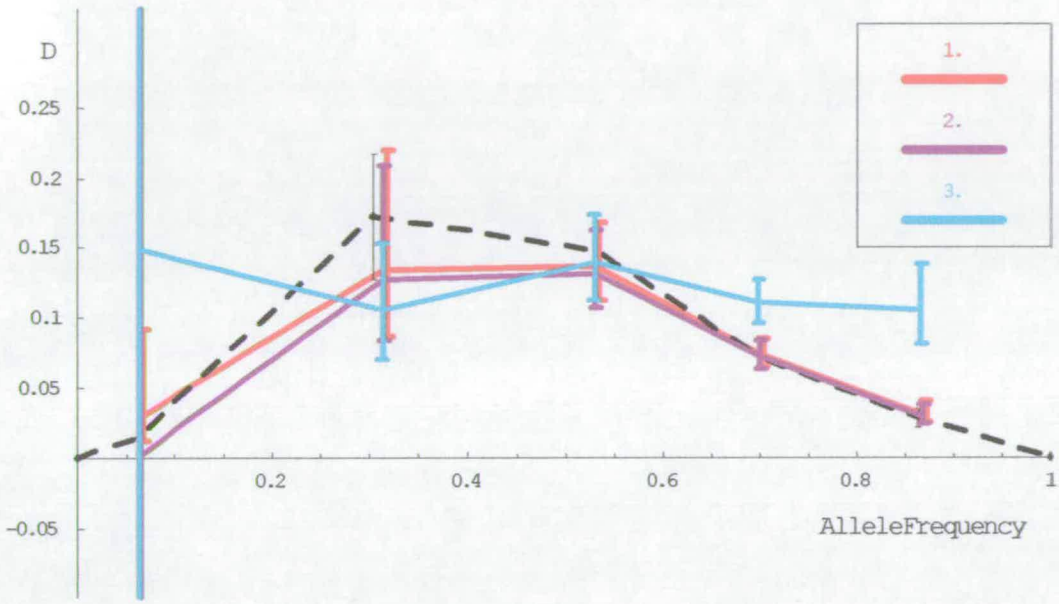


Figure 4.6. Linkage disequilibrium between trait pairs: 1) Spot score and allele frequency 2) Leg length and allele frequency 3) Spot score and leg length, and calculated from variance in allele frequency (dashed line) within sites grouped by their mean allele frequencies 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8 & 0.8-1. Error bars show the approximate 95% confidence intervals.

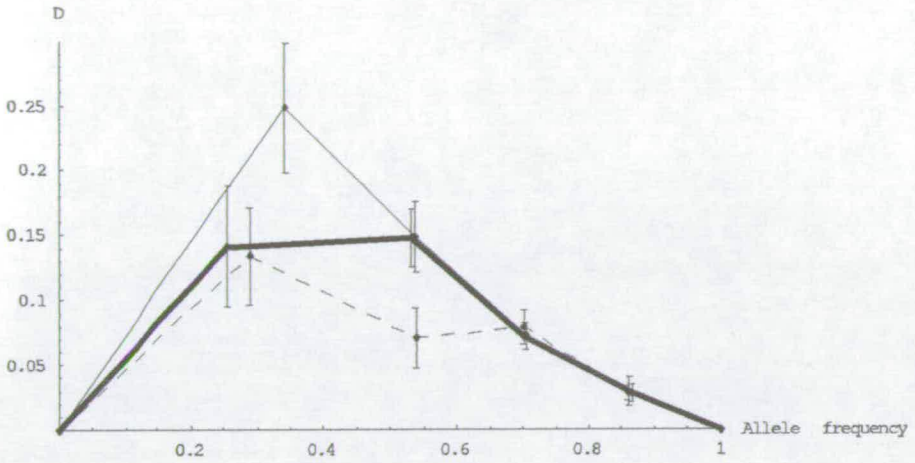


Figure 4.7. Linkage disequilibrium for females (solid line), males (dashed line) and sexes combined (heavy solid line) against the site mean marker allele frequency of site bins. Calculated from the variance in allele frequency. Values are estimated for sites binned by mean allele frequency of the appropriate sex.

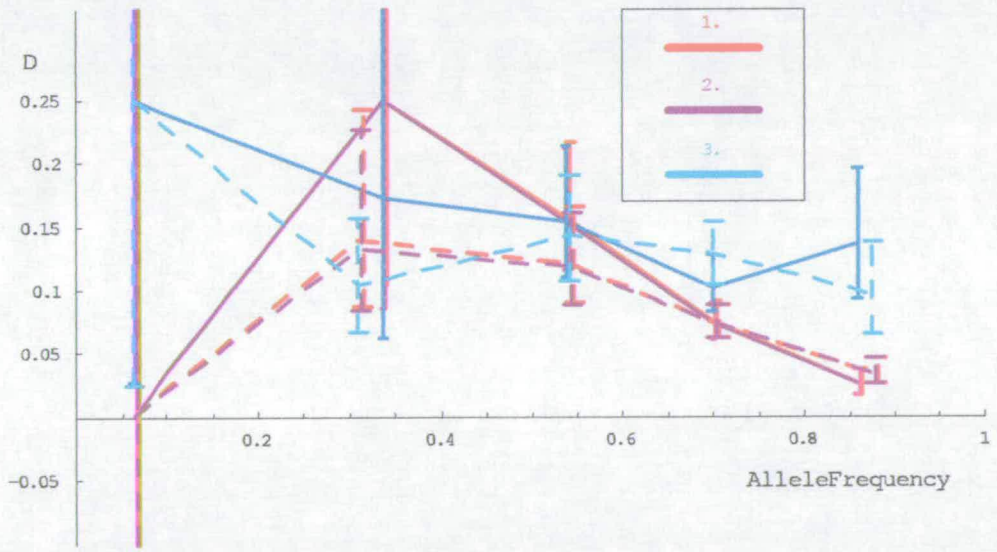


Figure 4.8. Disequilibrium between quantitative trait pairs, by sex: 1) Spot score and allele frequency 2) Leg length and allele frequency 3) Spot score and leg length. In site binned by mean allele frequencies 0, 0.2, 0.4, 0.6, 0.8 & 1. Error bars show the approximate 95% confidence intervals. Solid lines are female data, dashed lines are male.

Estimation of values separately for males and females shows that the linkage disequilibria between markers in males and in the combined-sex sample are broadly similar (Figure 4.7. The maximum value of disequilibrium in males is 0.116 as opposed to 0.137 in the combined sample) but that the maximum disequilibrium in females is extremely strong (and significantly stronger) in sites between mean allele frequencies 0.2-0.4.

When the sexes are considered separately the pattern seen between other traits quite closely follows that observed between markers for that sex (figure 4.8), again with the exception of the linkage disequilibria between spot score and leg length, which doesn't decrease in sites at the extremes of mean allele frequency. In the sites of intermediate allele frequency males show consistently weaker linkage disequilibrium.

Finally, the mean disequilibrium was estimated for sites grouped by their habitat score and shows a distinctly different pattern. There is very little difference in disequilibrium estimates between trait pairs over the range of sites of habitat scores 0- 0.6 and this then decreases slightly with increasing habitat score (Figure 4.9).

There is also some suggestion that the decrease in disequilibrium with habitat is smaller between spot score and allele frequency than for other comparisons. Split by sex, it can be seen that the disequilibrium shows a maximum in sites of intermediate habitat but in females is at its greatest in the most pond-like sites and is progressively lower as the sites become more puddle-like. In males the linkage disequilibrium in sites of low mean allele frequency is consistently (but not significantly) lower.

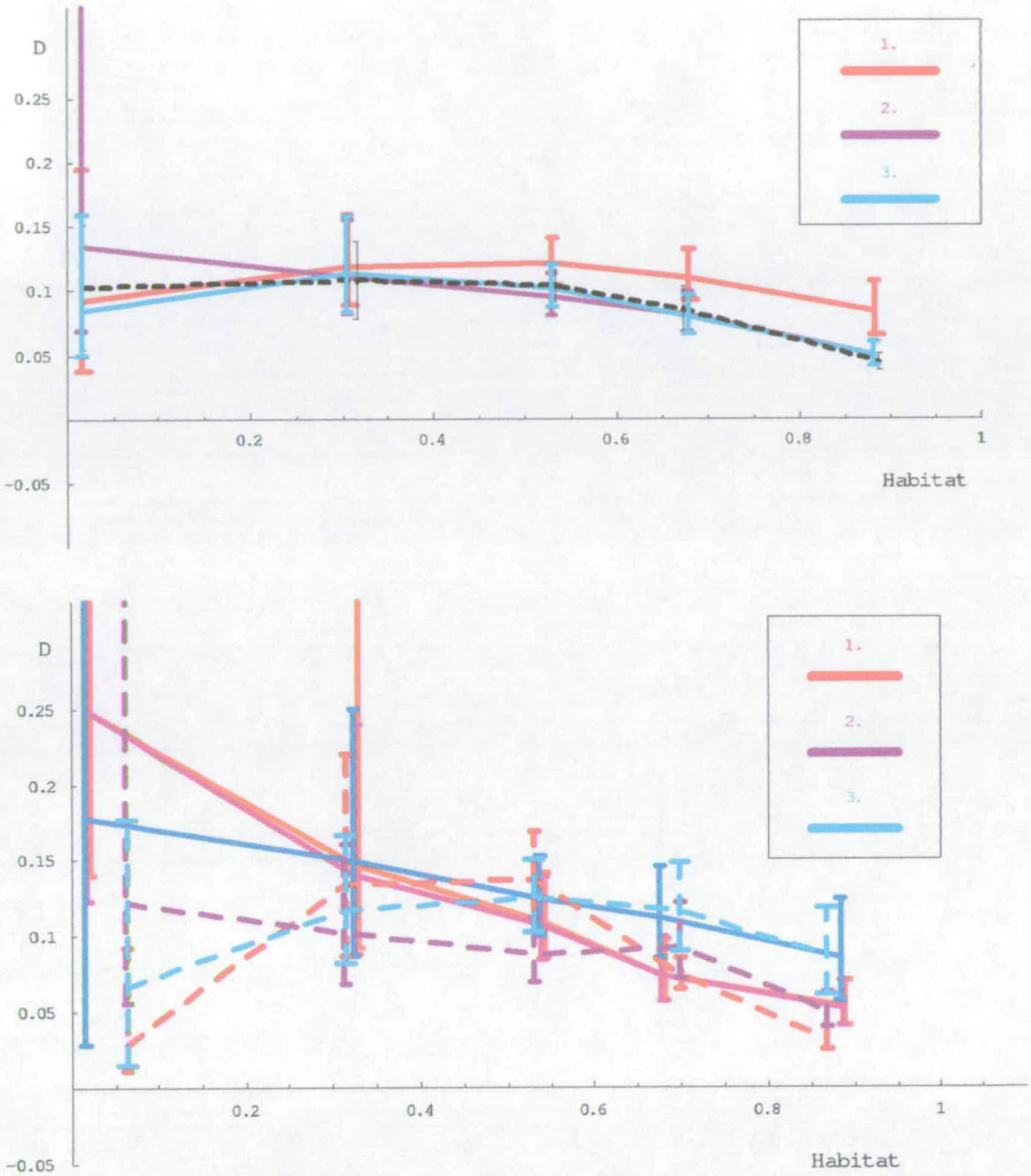


Figure 4.9. Linkage disequilibrium between trait pairs: 1) Spot score and allele frequency 2) Leg length and allele frequency 3) Spot score and leg length and disequilibrium calculated from variance in allele frequency (dashed line). In site binned by site habitat scores in the ranges 0, 0.2, 0.4, 0.6, 0.8 & 1. Error bars show 95% confidence intervals. (top) Combined sexes (bottom) Female (solid lines) & male (dashed lines)

4.5 Discussion

In a large part the distributions of quantitative trait means, variances and disequilibria in this survey appear very similar to those seen across a cline between *Bombina bombina* and *B. variegata* in Pešćenica, Croatia (Nürnberger *et al.* 1995, Kruuk 1997, MacCallum *et al.* 1998). This is despite there being a clear distinction between the structures of the Apahida and Pešćenica hybrid zones, the former possessing a mosaic of sites of varying mean allele frequencies across an extended area, as opposed to the smooth clines seen in the latter.

Although it is not immediately obvious why the distribution of quantitative traits should be different in a mosaic as opposed to a cline, there are some aspects in which the two areas differ greatly. Below I consider the significance of these similarities and differences in the evolution of this hybrid zone.

4.5.1 Concordance of traits

There are straight line fits between the site means of all traits and mean allele frequency in the same sites. Nürnberger *et al.* (1995) showed that where selection follows a spatially varying optimum and is relatively weak, small differences in selection strength between traits affects cline width only slightly. The result is based upon a model of a cline with optimum phenotype that varies about a point x=0 (the centre of the cline). The cline of the mean then takes the form:

$$\frac{\partial z}{\partial t} = 0 = \frac{\sigma^2}{2}\frac{\partial^2 z}{\partial x^2} + \frac{\sigma^2}{2r}\frac{\partial z}{\partial x}\frac{\partial Log(\overline{W})}{\partial x} - s'V_g(z - z_{opt})$$

(Equation 1 of Nürnberger *et al.* 1995, where σ is the standard deviation of the dispersal distance, z the mean phenotypic value and z_{opt} its optimum value, r the mean

recombination rate between loci, x the distance across the cline and W the fitness, V_g the genotypic variance and s the selection coefficient)

Is such a result informative in a mosaic? This cline has three parts, the first two describing the effect of spatial structure in phenotype and the spatial structure of the fitness of populations - in a mosaic such structure is essentially random. Whether this result extends to a mosaic hybrid zone would demand further theoretical examination.

To observe strong concordance in a mosaic, it seems implausible that these traits are following spatially varying optima as these would have to also be closely concordant for these very different traits. It is more conceivable that different traits follow spatially concordant optima across a cline such as that in found in Pešćenica as this exists over an ecotone and thus many different environmental differences are occurring over a similar region. In a mosaic this is less likely as the environmental changes are not spatially correlated but depend on a patchwork of local environmental differences. Alternatively strong selection against hybrids in one focal trait (i.e. selection in intermediate phenotypes for the trait of interest) would tend to reduce the range of values of other traits over which the focal trait changed (as in figure 4.1c). This is not observed suggesting that there is not strong selection against intermediate phenotypes.

The concordance relationship could occur if the selection on different traits varied in the same fashion, which again seems highly implausible. Epistatic selection maintaining combinations of trait values might tend to maintain a concordant relationship between traits, however it would have to act on extensive proportions of the genome to account for the close relationship of each of these traits and four unlinked marker loci, so it is also considered implausible as a complete explanation.

Genes of pleiotropic effect (or the very close linkage of genes affecting different traits) would cause a pair of traits to vary concordantly as the change in genotype has an equivalent effect on both traits. I consider this rather unlikely as the traits are rather

dissimilar and it seems unlikely that the same genes would have a major effect on variation in these traits.

The only reasonable explanation for strong concordance is that it is a result of the mixing of populations maintained at different allele frequencies. If populations where marker allele frequencies are high also have the allele for quantitative trait loci at high frequency then immigration from sites fixed with differing allele frequencies would generate covariance between the traits, increased variance and strongly concordant trait means, all of which are observed.

This leaves two problems: why would immigrants mix in different proportions to produce sites with a range of mean allele frequencies and what part is played by later generation hybrids in whom recombination has broken down the associations between traits? There is evidence of habitat preference in this hybrid zone (Vines *et al.* 2003, chapter 3), which could explain why a site would receive a different proportion of immigrants of different trait values. How the genetic composition of later generation hybrids affects the expression of habitat preference is completely unknown. If recombination in these hybrids results in the disassociation of quantitative traits and the disruption of habitat preference, it might tend to reduce the strength of the relationship between trait and habitat but might not change that concordance relationship. Also further explanation is needed of why the sites of high allele frequencies are maintained in the face of matings with immigrant hybrids. It has been suggested that the integrity of these sites may be maintained by plausibly strong selection (Vines *et al.* 2003) but this has yet to be tested experimentally.

4.5.2 Linkage disequilibrium

The strong average linkage disequilibrium between markers observed in Romania has been attributed to the mixing of populations at different allele frequencies (Vines et al. 2003). The high concordance between traits and individual marker loci suggests a null hypothesis that they are under similar selective pressures to the genetic markers and

therefore a similar strength of linkage disequilibrium is expected in these traits. These results at least partially appear to confirm this.

There is evidence of strong linkage disequilibrium between mean allele frequency and quantitative traits and between the traits themselves. In a single, random mating population these linkage disequilibria are broken down by recombination every generation by a proportion equal to the recombination rate between the loci. The linkage disequilibria between the unlinked marker loci are reduced by the recombination rate of 0.5 every generation. Although the linkage relationships of the quantitative trait loci are unknown, unless they are on average closely linked, the genetic covariances between traits will also quickly break down. Strong linkage disequilibrium between these loci is evidence of either a recent episode generating these disequilibria or a continual process doing so.

Linkage disequilibria may be increased either by the mixing of individuals from sites of differing allele frequencies (Li and Nei 1974, Prout 1973, Barton 2000) or by epistatic selection on trait combinations (or on closely linked selected loci in the case of markers, as these are presumed to be under, at most, very weak selection).

Estimates of mean pairwise linkage disequilibria between markers and between marker allele frequencies and traits have similar maximum values in sites of intermediate allele frequency. This pattern of variation with site mean allele frequency matches closely that observed in Croatian hybrid populations (closely following the results of Kruuk 1997, rather than Nürnberger *et al.* 1995 where the strengths of disequilibria between markers were approximately half that seen between marker loci). This strongly supports the hypothesis that mixing generates these patterns. Mixing of individuals from populations at high frequencies of the marker alleles and trait alleles of each parental population creates disequilibria of equal strength for varied traits. In contrast disequilibria created by epistatic selection depend on the strength of selection on particular combinations, which are therefore likely to vary from one another.

In the case of the trait pair of spot score and leg length, there is linkage disequilibrium significantly greater than zero in sites nearly fixed for *variegata* alleles. Equivalent linkage disequilibrium is absent in the Croatian data. What could cause this deviation? The simplest explanation is that these populations mostly consist of individuals fixed for *variegata* alleles at most loci and the disequilibrium is due to immigrants into these sites from sites with generally lower trait values. Clearly this cannot be repeated for trait-marker pairs or disequilibrium between markers as there is by definition little marker variation in these sites. However, if immigrant individuals, by chance, have low frequencies of *variegata* alleles then this would lower the mean allele frequency of the sites and hence would be grouped with sites of lower mean allele frequencies. The other explanation is that the covariances between these traits are being created within the site either by selection on available variation or by processes that generate correlation between traits – pleiotropy or correlated environmental variation.

If there is selection acting for the parental species trait combinations in these sites, then this must not be true of combinations of the same alleles at marker loci. As there is little genotypic variance in these sites (Vines *et al. 2003*) selection would have to act largely on the environmental variation in these traits. If this explanation is correct, it would also require that this selection is relaxed in sites of lower mean allele frequency, or otherwise an effect would be seen to act on environmental variation in sites of intermediate allele frequency too and create an excess of covariance between quantitative trait pairs there also. Similarly, if there is an effect of pleiotropy causing correlation between trait variation or environmental variation in separate traits, then this would have to be absent in intermediate sites.

These latter explanations seem improbable compared to the simple previous explanation that this is created by immigration without selection, particularly when it is known that between site migration is frequent (Chapter 3). However it is further possible to distinguish these when sites are grouped by habitat rather than mean allele frequency.

When disequilibria are estimated from sites grouped by habitat the explanation above of excess linkage disequilibrium between quantitative traits in the extreme sites is created by mixing cannot hold as any individuals with high trait values and low allele frequencies do not result in the site being moved into a separate habitat classification. There may be variation in both allele frequency and trait values in the sites at the end of the habitat spectrum, therefore as expected we see more consistent linkage disequilibrium across all habitats than across sites grouped by mean allele frequency (Figure 4.9). This would be consistent with linkage disequilibrium created by mixing, selection or pleiotropy.

However in sites with high habitat scores there is a significant excess of linkage disequilibrium between leg length and spot score. This can only have resulted from pleiotropy or selection for parental species phenotype combinations. If it were due to pleiotropy then this would have to act only in the sites with high habitat scores for the reasons discussed above. The more likely explanation is that there is generally a selective advantage for parental-species type trait combinations. However it must be borne in mind this is a result from only one trait pair and to test the generality of any hypotheses, many more pairs of quantitative traits would need to be measured.

Whichever is the cause of this linkage disequilibrium, this effect would have to be absent from the Pešćenica populations as there are at least equal amounts of variation in quantitative traits at opposite sides of the cline there without linkage disequilibrium being observed in these sites (Nürnberger *et al.* 1995, Kruuk 1997). The clinal structure of the Pešćenica hybrid zone means that any local immigration is most likely to be from genetically similar sites.

4.5.3 Habitat and quantitative traits

The habitat has previously been shown to be a strong correlate of a site's mean allele frequency in both Apahida and Pešćenica (Vines *et al.* 2003). In this analysis the relationship of all traits with habitat is consistent – in the most pond-like half of sites the

mean trait value rises with increasing habitat score at approximately the same rate as for the equivalent relationship of trait value and site mean allele frequency, but in the more puddle-like half of sites the increase with increasing habitat score declines and varies very little with habitat. This either reflects a genuine reduced strength of habitat assortment in sites with high mean allele frequencies or simply that the habitat score is less able to distinguish the features of the environment that are correlated with allele frequency in these sites. Comparisons with Pešćenica are not possible as the available habitat and patterns of site occupancy are strongly influenced by the ecotone and clinal structure.

4.5.4 Sex differences

In all cases female trait means have less extreme values than the male equivalent in the same sites. The slope of a regression of female site mean allele frequencies on male or the female site mean values of all traits on male was much less than one and the slope of the fit of mean trait value on allele frequency for females was much lower than for males. This contrasts quite strongly with the same analysis in Pešćenica where a weak effect was observed that female clines were narrower than male (effectively the opposite result).

In the Apahida hybrid zone the results could reflect stronger selection on males between birth and adulthood (which would be supported by Haldane's rule that the heterogametic sex is likely to suffer more from hybrid incompatibilities (Haldane 1922)). Alternatively it could reflect higher female dispersal rates or weaker habitat preference in females. Female dispersal distances were not found to be significantly different from male in the recapture tests (chapter 3) but no specific tests were made of habitat preference in the sexes.

These observations are also consistent with the observed patterns in linkage disequilibrium. Looking at the relationship of linkage disequilibrium with mean allele frequency, the sexes separately show the same general pattern as the sexes combined but

the strength of linkage disequilibrium in females appears to be stronger than that of males in the same sites. Of the previous explanations of sex differences, higher female dispersal rate is most likely to create this pattern. The arrival of females with parental allele combinations migrating into sites of intermediate allele frequency would generate stronger linkage disequilibrium. If more rapid change in male traits were the result of stronger selection on the males then this would result in greater linkage disequilibrium in males.

4.5.5 Drawbacks of the methods of linkage disequilibrium estimation

There are several drawbacks of this method of calculating linkage disequilibrium that could give misleading estimates and underestimate the confidence interval of these estimates. The estimated value and confidence interval estimation relies on estimates of Δz (the difference in trait value between populations fixed for the alleles of the two taxa) and the variance in the traits at the edge of the hybrid zone, and therefore the accuracy of these estimates depends on the accuracy of these measures. The value of Δz is estimated from an extrapolation of the observed regression between mean marker allele frequency and trait value and there is both error in the estimation of the slope of this regression and a lack of data for sites of extreme allele frequency and therefore knowledge about the expected value of these traits in the sites. The estimation of variance in the pure sites also relies on extrapolation of a regression of variance on mean allele frequency, which has wide confidence intervals, but also on the assumption that the variance is the same in sites at the *Bombina bombina* and *Bombina variegata* extremes, which may not be the case.

Errors in the estimates of these parameters will give poor estimates of linkage disequilibrium and no account is taken of the error assumed in these parameters. For these reasons there remains some doubt about the accuracy of the disequilibrium estimate and the confidence intervals are likely to be underestimated.

Chapter 5

The role of habitat preference in mating site choice and the creation of the offspring genotype distribution

5.1 Introduction

Prezygotic isolation is undoubtedly of tremendous importance in the reproductive isolation of many taxa. This is a fact apparently recognised by Darwin ("It is not surprising that the degree of difficulty in uniting two species, and the degree of sterility of their hybrid offspring should generally correspond, though due to distinct causes", Darwin 1859). However the conditions that lead to it becoming important in the isolation between one pair of taxa and not another are unclear. One general condition that seems to promote speciation is where there is strong sexual selection (e.g. Barraclough *et al.* 1995). Another is that the process of reinforcement is likely to increase the strength of prezygotic isolation in taxa that exist in sympatry rather than allopatry (Coyne and Orr, 1989). However, unlike postzygotic isolation, the strength of prezygotic isolation is not correlated with genetic distance (Butlin and Tregenza 1998). These observations seem to point to a conclusion that prezygotic isolation only becomes important when there are selective pressures directly favouring it (sexual or other forms of selection) and that this may be sensitive to conditions specific to the taxa involved.

The hybrid zones of *Bombina* offer an opportunity to examine what effect of difference in local conditions on the evolution of prezygotic isolation. This chapter examines on mechanism of prezygotic isolation in the Apahida mosaic hybrid zone – the association of adults with similar genotypes in sites together during mating. This is compared with studies of similar associations in the clinal hybrid zone in Croatia. The difference in structure between these hybrid zones changes the dynamics of hybridisation: the parental

populations are in closer contact with each other, and there is more potential for mating between the two in a mosaic. Theory has proposed that under these conditions reinforcement might act to strengthen prezygotic isolation. The comparison of these two different hybrid zones tests whether these predictions hold up in this hybrid system.

Considering the importance placed on prezygotic isolation, the genetics has been poorly characterised in natural populations relative to that of postzygotic isolation. The absence of a unified theoretical framework describing the evolution of prezygotic isolation is a hindrance in finding evidence for general mechanisms. Instead there are a profusion of varying theoretical models exploring different aspects of prezygotic isolation and data that test the predictions of these individual models yet it is unclear how these models relate to one another and what common ground they share.

In contrast, the origin of postzygotic isolation has a solid theoretical basis which is reflected in the quality of evidence for its evolution. In some cases this has been very well described and understood, with genetic incompatibilities between divergent populations having been narrowed down in some cases to the gene level (examples were considered in the previous chapter and have been recently reviewed by Orr, 2001).

This difference in emphasis is unfortunate given that pre-zygotic isolation is likely to be more efficient than post-zygotic in the creation of the reproductive isolation of nascent species (Turelli *et al* 2001, Kirkpatrick and Ravigné 2002, Coyne and Orr 1997). For example prezygotic isolation under sexual selection is apt to increase rapidly (e.g Liou and Price 1994). Even when sexual selection is absent, where there is already postzygotic isolation there it is possible for the strength of prezygotic isolation to increase under selection (see section 5.1.5). Acting under natural selection, reinforcement of pre-zygotic isolation works by trading off a loss of fitness incurred by increasing the effectiveness of assortative mating against a lowering of the fitness cost of

producing inviable or infertile offspring. Therefore we should expect prezygotic isolation to play a major role in the isolation of many examples of diverging populations.

The lack of knowledge of the genetics of pre-zygotic isolation may stem from difficulties in detecting and quantifying the strength of pre-zygotic isolation, and in difficulties in detecting the effects of reinforcement (Noor 1999). The detection and quantification of pre-zygotic isolation requires detailed controlled experiments and observations of the ecology of diverging taxa, their mating and mate choice behaviours and their reproductive biology. In contrast, to identify postzygotic isolation may involve experiments as simple as a cross between two populations and noting infertility or inviability in the offspring.

5.1.1 The classification of theory of prezygotic isolation

The evolution of pre-zygotic isolation may be classified in a number of ways. Most commonly the geographic circumstances of diverging populations are the basis of the principal division of different mechanisms. It seems that this has more to do with an analogy with postzygotic isolation - in the case of postzygotic isolation the Dobzhansky-Muller model provides a good explanation of allopatric divergence but is not applicable in sympatry and for this reason geographic separation is an important way to classify cases of diverged taxa. Although this may be a convenient basis on which to devise models of prezygotic isolation it is less convincing than for postzygotic isolations as the crucial features for speciation in geographically separated populations, the cessation of gene flow between populations, is also in effect a means of prezygotic isolation (if mating occurs primarily in the separated habitats). However geographic separation is generally excluded as a means of prezygotic isolation by definition.

Kirkpatrick and Ravigné (2002) offer a more rational division of the problem, proposing five elements of speciation: the source of disruptive selection, the means of isolation, the means by which the first influences the second, the genetics of isolation mechanisms and initial conditions of divergence. These have a great advantage that they are largely independent of one another (in that it is conceivable that two pairs of diverging populations could differ in one element but remain similar in all others). Under this classification geographic isolation may be considered as simply another form of assortative mating (Kirkpatrick and Ravigné 2002).

Mechanisms of pre-zygotic isolation themselves fall into three distinct categories:
(1) Acting to prevent matings by keeping diverging populations apart, so that mating cannot occur.
(2) Reducing the frequency of between population matings when they are otherwise free to occur.
(3) Preventing the success of any hybrid matings that do occur.

These differences are essentially demarcated by their time of action. The first acts prior to mating and is therefore potentially a side-effect of factors unconnected to the mating process, the second occurring as a part of the mating system and the third acting post-copulation.

In the first instance, divergent populations may be separated spatially through geographical isolation or the occupation of different habitats within the same region or they may be temporally isolated by mating at different times of day or year. The second instance covers instances of assortative mating between individuals of one or both populations, occurring through divergent mate preferences or by incompatibilities of the reproductive systems such that matings cannot occur. The third refers to factors that cause mating between populations to fail to produce a zygote and includes sperm selection and problems of fertilisation by sperm from the divergent population.

5.1.2 Spatial and temporal isolation

A simple way to prevent breeding between populations is to ensure that they never meet or that they only come together in a context in which there is no mating. Clearly each case where sister taxa are allopatric provides an example of this. Less complete isolation will occur where there are habitat preferences that vary across the populations i.e. where alternative habitats are available to both subpopulations but each uses only one. Such habitat preferences prevent the divergent populations from coming into contact as often as in the absence of habitat separation and thereby create a kind of assortative mating across the whole range of the populations.

Clear examples come from certain insect species that use specific plant species for feeding and also breed upon their host-plant. The first example to be thoroughly studied was the apple-maggot fly, *Rhagoletis pomonella* (Feder *et al* 1988), but several other cases have also been identified including *Acyrthosiphon pisum* (Hawthorne and Via 2001) and *Zeiraphera diminiana*, (Emelianov *et al.* 2001). The effect of breeding on the host plant is similar to that of assortative mating in a mixed population and is effective to the extent to which populations adapted to different hosts are faithful to their host species.

A similar pattern is created when the separation is temporal and not spatial. The time of breeding often varies between sister taxa, whether that be the time of year (e.g. between genetically divergent populations of Chinook salmon *Onchrhyncus tshawytscha* (Quinn *et al.* 2000)) or time of day (e.g. in *Bactrocera cucucubitae* flies that achieve partial isolation between short and long circadian rhythm types by mating five hours apart

(Miyatake *et al.* 2002)). Again if these differences are characteristic of separate populations then this will result in assortative mating within the groups.

5.1.3 Sexual isolation

Sexual isolation between taxa refers to both assortative mating and sexual selection or indeed any process during mating that reduces the rate of mating between the populations, when the populations have free access to one another. The difference is that while sexual selection is potentially based upon preference for arbitrary traits, the traits of preference under assortative mating are shared by all members of one population specifically for differentiating them from the other population. The difference between these two processes may have a profound effect on the course of the divergence of taxa. Whereas assortative mating may increase in effectiveness only through improvement in quality of mate recognition, preferences under sexual selection open the possibility of being a runaway process, in which the end point is unclear. Therefore distinguishing between assortative mating and sexual selection requires the identification of the preference target of mate recognition systems.

Ryan and Rand (1993) suggest that the cues used in mate recognition and those used in sexual selection are frequently the same. In contrast, Paterson (1982, cited by Sanderson 1989) argues that mate recognition systems are likely to be under stabilising selection. The profusion of species in clades demonstrating stronger sexual selection (Passerine birds, Barraclough *et al.* 1995; flowering plants, Hodges and Arnold 1995), apparently signifies that sexual selection acts to increase the rate of speciation. These observations at least seem to suggest that in many cases that there is not stabilising selection on mate recognition traits or if there is, that it does not act as a brake on speciation.

Divergence of the female preferences may also evolve to reflect changes in the environment (Schluter and Price 1993). In combination with the potentially rapid co-evolution of female preference and male trait (Lande 1981) this can lead to great divergence between populations that become isolated in different environments, greatly increasing the likelihood of speciation in allopatry or by reinforcement on secondary contact (Liou and Price 1994).

5.1.4 Post-mating prezygotic isolation

Although female choice may act by behavioural mechanisms before mating, there are numerous mechanisms by which mating between individuals from different populations can be unsuccessful. This may occur through incompatible genital morphology or other sperm-blocking mechanisms, or by the biochemical incompatibility of gametes that prevents fertilisation.

A recent example is a study by Geyer and Palumbi (2003) on the gamete recognition systems of *Echinometra oblonga* sea urchins. In populations where *E. oblonga* is sympatric with *Echinometra* species C there are signatures of positive selection in the sequence that codes for bindin, the protein that attaches sperm to the egg, but none at other neutral loci. Individuals from allopatric populations of these species hybridise readily in the laboratory and there is no evidence of selective changes to any loci. The authors consider this mechanism to be the reason no hybrids are found in areas of sympatry. This appears to be a case where reinforcement has acted at a late stage post-mating in a species where behavioural assortative mating is not possible.

Post-mating selection can also act as a form of cryptic female choice. In areas containing no pure adults of the waterfrog *Rana ridibunda*, matings between *R. lessonae/ridibunda*

hybrids fail. The hybrid failure results when hybrids eliminate the *lessonae* genome, producing only *ridibunda* gametes, as *R. ridibunda* gametes carry loci causing inviability. Therefore while *R. lessonae* females should seek to avoid mating with hybrids at all costs (their genome will not be passed on by hybrid offspring) the mating assortment appears to be random (Reyer *et al.* 2003). This is perhaps because they have a form of coercive mating –fertilisation is external and a male amplexing a female is likely to fertilise all her eggs. The means that females use to reduce this cost of unwanted hybridisation is by reducing the number of eggs she releases when amplexed by a hybrid male (Reyer *et al.* 2003)

Clearly when there are blocks to fertilisation soon before the formation of the zygote it becomes extremely difficult to distinguish prezygotic mechanisms from the failure of embryos very early in development. The staple studies of observation of mating patterns in the field and simple mating experiments give no clue of whether isolation occurs before zygote formation or soon after by selection, as the results will look identical. However the significance of the distinction may be more than just semantic. Consider the case of a species in which there is no parental care but where there is postzygotic isolation. Up until the moment of zygote formation there may be a conflict between males seeking to fertilise eggs and females, preferring a different mate, to thwart this. Upon fertilisation of eggs amelioration of the effects of postzygotic isolation becomes advantageous to both parents. Therefore prezygotic isolation may be subject to the effects of sexual conflict, sexual and natural selection but after zygote formation the effects of sexual selection and sexual conflict become ineffective. However in many cases the costs of producing offspring continue after fertilisation, at least for one parent. In these cases the conflict may continue after fertilisation, as the parent carrying the cost of caring for the offspring could instead direct these resources into remating with a preferred mate.

5.1.5 Reinforcement

It has often been suggested that pre-zygotic isolation provides a means to complete the formation of new species. This process, termed "reinforcement" by Dobzhansky (1940), posits that whenever diverging populations are hybridising and their hybrid offspring suffer a fitness cost, there is a potential trade-off between any fitness costs of increasing the strength of prezygotic isolation and the increase in fitness of offspring this achieves. Therefore under suitable conditions prezygotic isolation is expected to increase in efficacy.

It seems that Darwin and other contemporaries appreciated the significance of the process, at least in terms of post-mating prezygotic isolation: "At one time it appeared to me probable, as it has to others, that the sterility of first crosses and of hybrids might have been slowly acquired through the natural selection of slightly lessened degrees of fertility, which, like any other variation, spontaneously appeared in certain individuals of one variety when crossed with those of another variety" (Darwin 1859). However, reinforcement has had a somewhat turbulent journey from widespread acceptance, through rejection as theoretically implausible then more recently largely resurrected by new experimental data (e.g. Coyne and Orr 1989, Rice and Hostert 1993) and a number of new theoretical studies demonstrating the plausibility of the process and increasing empirical support from a large number of new studies testing predictions of these models (the progress of arguments about reinforcement has been thoroughly reviewed by Noor 1999).

One of the main planks of empirical support for this comes from studies comparing the strength of mating preference exhibited by one or a few pairs of species from areas where they exist in sympatry or parapatry and those where they are allopatric. The definition of reinforcement would seem to imply that, all else being equal, there should

be lower levels of pre-zygotic isolation in populations not subject to hybridisation, but that levels of post-zygotic isolation should be little different between such populations.

However it is easy to conceive of situations where there might be differences between populations, unrelated to reinforcement, that could generate differences in sexual isolation (Butlin 1995). For instance Day (2000) considers a situation in which the female preference varies across a species range which then generates differences between populations in reproductive characters, similar to that expected under reinforcement. Many studies consider only a small number of species pairs and use the strength of assortative mating behaviour as the only measure of pre-zygotic isolation. The argument above makes it difficult to conclude that reinforcement has generated these patterns in individual cases, but combined across many groups of organisms this provides convincing support for the reinforcement hypothesis.

Therefore the most convincing evidence of reinforcement comes from studies with large numbers of comparisons. Coyne and Orr (1989, 1999) surveyed a large number of *Drosophila* species and found that the strength of assortative mating was greater between sympatric species pairs than between allopatric pairs of similar genetic differentiation. The same pattern was not repeated for post-zygotic isolation. Although this only considers one form of many forms of pre-zygotic isolation it is clear evidence of reinforcement.

Similarly in the *Drosophila willistoni* group, sexual isolation evolves faster than postmating isolation and song evolves fastest of all perhaps suggesting (if song is used for assortative mating) a selection pressure to increase premating isolation (Gleason and Ritchie 1998).

Within species there may be similar patterns between divergent populations. Tilley *et al.* (1990) surveyed populations of *Desmognathus ochrophaeus* salamanders and found that

there was greater ethological isolation between geographically closer pairs than similarly diverged but more distant located pairs. There was no relationship between ethological isolation and the genetic distance of pairs of with the same geographic separation.

5.1.6 Pre-zygotic isolation in hybrid zones

Hybrid zones by definition have incomplete reproductive isolation. Although there are examples of strongly bimodal hybrid zones (Jiggins and Mallet 2000) in which there are mostly parental types and sterile F1 individuals, in many cases a profusion of hybrid genotypes can be found. This indicates that hybrids are both viable and fertile and are mating together. However as these hybrid zones persist there must also be either or both prezygotic and postzygotic isolation to resist gene flow.

Therefore hybrid zones may provide an interesting example of where isolating mechanisms are effective over a long period but are fairly weak. Therefore it is intriguing to know what forms of reproductive isolation are found in these hybrid zones and the contribution that they each make in the maintenance of hybrid zone structure

The occurrence together of hybrid unfitness and frequent hybridisation between parapatric populations might suggest that hybrid zones are prime sites for finding evidence of reinforcement. However, theoretical studies show that the conditions under which reinforcement is successful in increasing assortative mating in a tension zone are restrictive (Sanderson 1989, Cain *et al.* 1999). However it is apparent that reinforcement may increase temporarily in tension zones before being swamped by gene flow and decreasing in frequency (Sanderson 1989, Cain *et al.* 1999) and also the conditions in a "mosaic" hybrid zone under which alleles causing reinforcement may become established are much more forgiving (Cain *et al.* 1999). Therefore reinforcement may be detected in new hybrid zones or those showing a mosaic structure.

Table 1 lists some of the main components of pre-zygotic isolation in a number of studies of hybrid zones: association of parental populations with different habitats, assortative mating through mechanisms of mate choice, the presence of differentiation between mating signals of parental populations and evidence of post-mating assortment. Clearly there is some overlap between these categories. For instance assortative mating behaviours cannot be successful without differentiation between the parental populations in some distinguishing characters but it is quite possible for populations to be divergent in mating signals and not display assortment. For these reasons there is an entry only where each category has been explicitly tested for or they are quite apparent from observations.

There are few common patterns between these studies and there are few systems in which all possibilities have been considered. Population or genotype specific habitat assortment seems to be a very common feature as does behavioural assortative mating and these frequently, but far from always, occur in the same hybrid zone. Post-mating, prezygotic isolation has been examined less frequently but appears to occur in the absence of other mechanisms only rarely.

Habitat assortment in hybrid zones

The association of different combinations of genotypes typical of the parental species with different habitats seems to be a common feature of many hybrid zones. The cause of this variation may lie either in ecological selection acting differently on the opposing genotypes in the alternative habitats, or the expression of an active habitat preference of individuals of the opposing genotypes for these habitats. However the observation of such habitat assortment within a hybrid zone does not in itself demonstrate that it generates reproductive isolation. For such habitat assortment (generated by either

mechanism) to induce reproductive isolation, firstly mating must be contained within these habitats (Maynard Smith 1966) and this has not always been explicitly tested for.

The association of parental genotypes with different habitats does not always demonstrate habitat preference or necessarily involve isolation. The following are respectively examples of where habitat preference does and does not generate reproductive isolation. The grasshoppers *Chorthippus jacobsi* and *C. brunneus* are found to vary across a narrow hybrid zone in Northern Spain (Bridle *et al.* 2001a) but within this cline there is some variation in genotypic composition between habitat types (Bridle *et al.* 2001b, Bridle and Butlin 2002). The isolating effect of this habitat association is enhanced by different seasonal variation in density between the species and has been shown to contribute heavily to prezygotic isolation between them (Bailey *et al.* 2004). The mussels *Mytilus edulis* and *M. galloprovincialis* form a mosaic of patches of the pure species and hybrid populations (Bierne *et al.* 2003a) despite the potentially very high rates of gene flow that could result from the close association of populations that spawn into open water and have planktonic larvae. In this case habitat assortment in adult mussels does not create prezygotic isolation, which is generated instead by assortative fertilisation success and mating asynchrony (Bierne *et al.* 2002) and patterns generated by preferential settlement rates of larvae and selection on settled larvae (Bierne *et al.* 2003b).

Assortative mating in hybrid zones

Assortative mating behaviours are commonly addressed by mate choice experiments in captivity or in the wild by observation of mating pairs or by collecting fertilised eggs and seeds. The first method can only be a substitute for observation of wild mating if the mating preferences and mating propensity remain the same in captivity as in the wild.

These may be in opposition: in no-choice trials, where there is only one choice of mate provided by the experimenter, reduced mating propensity is accounted for but mate preferences are more likely to be skewed than in choice experiments where reduced mating propensity in one parental type may alter the apparent assortment.

Observation of matings in the wild perhaps offers a better insight into mating assortment. However this method is not without its problems – matings may be missed or the importance of matings overstressed. For example, in waterfrogs, females may reduce the number of eggs laid when amplexed by an undesired male (Reyer *et al.* 1999). In the marine snail *Saxatilis littorina* where there is apparent assortative mating between low-, mid- and upper-shore morphs, allowance needs to be made for male-male matings, female rejection of males despite mounting and non-random distribution of morphotypes within parts of the shore (Johannesson *et al.* 1995).

A final method of assessing the importance of assortative mating is to infer parental genotypes from the offspring. This is a particularly useful approach in plants, where mating preferences are likely either be expressed by pollen preferences by the pollen recipient or through the preference of pollinator species. Study of seed genotypes, at RAPD markers, of wild hybrids (Cruzan and Arnold 1994) or crosses (Emms *et al.* 1996) between species of Louisiana Iris (*Iris fulva* and *I. brevicaulis*) revealed an asymmetric pollen preference, with *I. brevicaulis* seeds showing no evidence of assortative mating but *I. fulva* seeds showing a strong assortment of parental genotypes.

Rieseberg *et al.* (1998) studied rates of selfing and outcrossing between the sunflowers *Helianthus petiolaris* and *H. annuus*, and the success of pollen from the different species using seven isozyme markers and found reduced fertilisation by conspecifics in hybrid populations. In fertile hybrids preference was expressed for the closely related parental genotype and but intermediate hybrids showed no preferences.

The genotypes of offspring have also been examined in animal hybrids. Howard *et al* (1998) tested for assortment between the ground crickets *Allonemobius fasciatus* and *A. socius* in cages containing both taxa. Observation of matings showed significant behavioural assortment between the taxa only when one dominated the cage. However, genotyping of offspring at allozyme loci reveals that there is a strong fertilisation advantage for conspecific sperm. Mallet *et al.* (1998) looked for evidence of assortative mating between the butterflies *Heliconius erato* and *H. himera* in eggs collected from females in the wild. The proportion of alleles from each species can be deduced from adult phenotype, revealing highly assortative mating, with 5% mating of eggs resulting from matings between the species. Any contribution to this result from sperm preference can be ruled out in this case as there is only fertilisation by a single male. It is unclear from these results whether females could have also prevented fertilisation of eggs from mating by undesirable males as the frequency of unfertilised eggs artificially expressed from females is not stated.

The presence of diverged mate recognition or sexually selected traits is often taken as evidence that there is assortative mating without testing this assertion. Although for behavioural assortative mating to be successful then there must be divergence between the populations, there is also a possibility that these cues are not used during mating for species differentiation. For example, in the *Chorthippus parallelus parallelus/C.p. erythropus* hybrid zone there is no evidence of assortative mating in the wild (isolation is largely by biased fertilisation) but there is differentiation in the cuticular hydrocarbon composition, which is frequently used in mate recognition in insects.

Reinforcement in hybrid zones

The flycatchers *Ficedula albicollis* and *F. hypoleuca* share an area of sympatry within their ranges. In areas of sympatry there is divergence in plumage colouration and there is a resulting increase in preference for a conspecific male in mate-choice tests on females from sympatric populations despite the divergent mating character being a marker of lower viability in allopatry (Sætre *et al.* 1997). This pattern is not entirely consistent across all populations. In hybrid zones in the Baltic island plumage divergence between the taxa is less pronounced and interbreeding rates are higher possibly due to the young age of these islands and the higher hybrid fitness in these populations (Sætre *et al.* 1999). Also, in certain conditions and in certain populations heterospecific matings are favoured (Veen *et al.* 2001).

The impression that reinforcement in hybrid zones is dependent on some details specific to the taxa in question is confirmed in the hybrid zone of the quail species *Callipepla californica* and *C. gambelii*. In these taxa mating occurs in coveys (individual breeding flocks consisting of a number of families) that may contain only one or both species. Mating experiments showed that in coveys from areas of allopatry there was preference for conspecifics (Gee 2003). However in mixed coveys, contrary to the expectations under reinforcement, mate preference appeared to be for heterospecific mates, probably as a means of reducing the costs from the inbreeding inherent in mating within a limited covey population (Gee 2003).

Hybridising taxa	Habitat preference/ temporal separation	Behavioural assortative mating	Mate choice or mate signal differentiation	Post-mating, pre-zygotic assortment
Butterflies:				
Heliconius himera/erato (a)		Yes		No
Heliconius melopomene/ cydno (b)	Yes	Yes	Yes	
Crickets:				
Chorthippus parallelus parallelus/erythropus (c,d,e)	No	No	Yes	
Chorthippus jacobsi/ brunneus (f,g,h)	Yes	Yes	Yes	
Allenomobius socius/ fasciatus (i,j)		No		Yes
Melanoplus (k)	Yes	Yes		
Orchelimum nigripes/ pulchellum (l)		Yes	Yes	
Molluscs:				
Littorina saxatilis (m)	Yes	Yes	Yes	
Mytilus edulis/ galloprovincialis (n,o)	Yes			Yes
Mammals:				
Mus musculus musculus/ domesticus (p)		Yes		
Birds:				
Corvus corone/cornix (q, r)		Yes		
Callipepla californica/ gambelii (s)	Yes	+ve in captivity and allopatry. -ve in sympatry	Yes	
Dendroica occidentalis/ townsendii (t)		Yes	Yes	

Table 5.1. Characteristics of pre-zygotic isolation in hybrid zones. Continuation and full legend overleaf

Continuation of Table 1.				
Hybrising taxa	**Habitat preference/ temporal separation**	**Behavioural assortative mating**	**Mate choice or mate signal differentiation**	**Post-mating, pre-zygotic assortment**
Birds contd.				
Ficedula hypoleuca/ albicollis (u,v,w)	Yes		Yes	
Hippolais polyglotta/ icterina (x)	Possibly		No – collapsed differences	
Amphibians				
Triturus vulgaris/ montadoni (y)	No	Yes	Yes	
Bombina bombina/ variegata (z, aa, bb, cc)	Yes	No	Yes	Possibly
Plants				
Iris fulva/brevicaulis/ hexigona (dd, ee)	Yes		Yes	Yes
Helianthus paradoxus/ annuus (ff)	Yes			Yes

Table 5.1. Characteristics of pre-zygotic isolation in studies of hybrid zones
References: (a) Mallet *et al.* 1998; (b) Jiggins *et al.* 2001; (c) Butlin *et al.* 1987; (d) Ritchie *et al.* 1989; (e) Butlin *et al.* 1992; (f) Bridle and Butlin 2002; (g) Bridle *et al.* 2002; (h) Bailey *et al.* 2004; (i) Howard and Gregory 1993 (j) Howard *et al.* 1998; (k) Orr 1996; (l) Shapiro 2001 (cited by Vines 2002); (m) Johannesson *et al.* 1995; (n) Bierne *et al.* 2002; (o) Bierne *et al.* 2003; (p) Smadja *et al.* 2004; (q) Palestrino and Rolando 1996; (r) Risch and Andersen 1998; (s) Gee 2003; (t) Pearson and Rowher 2000; (u) Sætre *et al.* 1997; (v) Sætre *et al.* 1999; (w) Veen *et al.* 2001; (x) Secondi *et al.* 2003; (y) Babik *et al.* 2003 (z) Sanderson *et al.* 1992; (aa) MacCallum *et al.* 1998; (bb) Nürnberger *et al.* in press; (cc) Vines 2002;(dd) Cruzan and Arnold 1994; (ee) Emms *et al.* 1996; (ff) Rieseberg *et al.* 1998.

5.1.7 Prezygotic isolation in *Bombina* hybrid zones

Often divergent taxa found in sympatry show evidence of prezygotic isolation (Coyne and Orr 1998). Furthermore many species pairs of anuran amphibians show assortative mating on the basis of call characteristics (Blair 1964, Gerhardt 1994, Ryan and Rand 1993). *Bombina bombina* and *Bombina variegata* are known to differ in their mating system in several respects. They vary in the pulse duration, cycle length and fundamental frequency of their mating calls (Lörcher 1969), *Bombina bombina* possess vocal sacs that *Bombina variegata* lack, and *Bombina bombina* form mating territories (Lörcher 1969) whereas *Bombina variegata* have scramble mating (Barandun 1990). This might suggest that the taxa show prezygotic isolation when found together in hybrid zones. However analyses of several different hybrid zones show a somewhat more complicated picture.

Across a clinal hybrid zone between *Bombina bombina* and *Bombina variegata* near Cracow, Poland, the cline in mating call characteristics was of the same width as the cline in allozyme allele frequency, suggesting these traits are neutral (Sanderson *et al.* 1992), and there was no deficit of heterozygotes at allozyme loci suggesting there was no assortative mating (Sanderson *et al.* 1992).

In a clinal hybrid zone at Pešćenica, Croatia, heterozygote deficit was found in allozymes at the centre of the cline (Nürnberger *et al.* 1995, MacCallum *et al.* 1998). However this can largely be attributed to an association of different genotypes with alternative habitat types and a mosaic structure of habitat at the centre of the cline (MacCallum *et al.* 1998). This may be due to an active habitat preference found in the movements of adult toads in these taxa (MacCallum 1994). In a single site in the centre of this hybrid zone, no associations were found between the inferred genotypes of

parents of eggs families, suggesting that there was no assortative mating (Nürnberger *et al.* in press).

The hybrid zone in Apahida has a starkly different structure, with a mosaic of habitat types across a wide area and a strong association of genotype with habitat across the mosaic (Vines *et al.* 2003). A simulation study of assortative mating in hybrid zones showed that the conditions under which reinforcement might be effective are much more lenient if the hybrid zone is a mosaic rather than a cline (Cain *et al.* 1999) as the strong differences in allele frequency between neighbouring demes created by the mosaic structure mean that there are more opportunities for mating between the parental genotypes compared to a cline which promotes reinforcement. Therefore we might expect to see a greater effect of prezygotic isolation in such a hybrid zone.

When compared over the scale of the per generation dispersal distance, there is an association twice as strong in the difference between site pairs in their habitat and the mean allele frequency of their occupants in Apahida as in Pešćencica (Vines *et al.* 2003). If this habitat association represents prezygotic isolation then it would support the suggestion that reinforcement has increased the efficacy of habitat preference in this hybrid zone. Direct measurements of habitat preference from the movements of individuals show no genotype preference for particular habitat types. However with high rates of migration between nearby sites (Chapter 2 and Vines *et al.* 2003) only habitat preference or strong ecological selection can explain the strong association of habitat and genotype. There is no evidence of strong environmental selection against maladapted immigrant tadpoles in *variegata*-like and intermediate habitat types (Köhler 2003), or on quantitative traits in sites of intermediate habitat type (Chapter 3).

In five sites of intermediate allele frequency in Apahida, there is no evidence of assortative mating between adults (Vines 2003). However there is evidence that the adult population caught in sites does not match well with the population mating within the

same site (Vines 2003) and that the population mating within some sites is a non-random sample of all the adults in the local area that are able to make up the mating population (Vines 2003), suggesting that there is some form of behaviour causing the association of adults of similar genotypes in these sites, perhaps resulting in a form of prezygotic isolation.

5.2 Aims of the analysis

The aims of this chapter are threefold:

(1) To compare the association of habitat type and genotype in the Apahida adult populations and only in the mating adults.
(2) To determine if adult habitat associations generate prezygotic isolation.
(3) To determine if there are differences in the genetic characteristics of the eggs and adults in the hybrid zone.

This is to determine firstly if the mating patterns of *Bombina* can explain the association of habitat and genotype observed in adults in Apahida. Secondly it is to determine if the association of adults of similar genotypes within different sites generates reproductive isolation. Finally, it is determined if there are differences between the offspring and adult populations that must be resolved between these life-stages by processes that have not or cannot be directly measured or tested experimentally. Such results are critical in a mosaic hybrid zone. Whereas in a clinal hybrid zone, the source of pure genotypes is known, in a mosaic it is far less clear why relatively pure populations should exist. Also it is unknown why such a strong habitat assortment should be present in a mosaic when the prevalence of hybrid genotypes might suggest that any genes underlying habitat preference loci may have lost their association with specific alleles at marker loci through a history of interbreeding and recombination.

Finding both habitat assortment and assortative mating in the wild requires knowledge of the make-up of the parent population. Ideally this would be through direct sampling of the successfully mating adult pairs. However there is no guarantee that a pair in amplexus represent a successfully mating pair. For instance Reyer *et al.* (2003) find that *Rana ridibunda* and *Rana lessaonae* females can reduce the number of eggs laid if the amplexing male is undesirable. Furthermore it is possible that females may be able to escape amplexus with unwanted males. This approach is also not favourable because observation of mating pairs is quite rare.

An alternative approach is to infer the genotypes of parents from the genotypes of their offspring (Nürnberger *et al.* in press). However, the data available here do not allow this (see methods). Therefore I do not make any direct assessment of assortative mating as this requires accurate information about the genotypes of the parents or reliable inferences of them. However this information is suitable for measuring the general properties of populations.

As the population of adults that mates in a site appears to show some difference from the adults caught in the sites (Vines 2003), I make the distinction between two different populations in the same sites. I refer to the adult that commonly reside within the site during the daytime and hence that are caught during sampling as the adult "population." I refer to the adults that mate within the sites as the "mating adult population". I also refer to the adults within an area around a given site that are likely to be able to migrate to and from the site quite freely, and thus, all else being equal, effectively forming a panmictic population as the pool of local adults.

Instead of directly sampling of mating adults, inferring the mating population from their offspring has the advantage of giving information about only those adults that were successful in mating. Additionally it gives an indication of the genetic make-up of the next generation of toads. The downside is that the genotypes in the offspring can be

affected by other processes and therefore give a false indication of parental genotype. For example, the creation of gametes may be non-random with respect to genotype (i.e. carrying more copies of loci showing meiotic drive), and similarly the fertilisation of eggs could be non-random with respect to sperm genotype and lastly selection against offspring before egg collection or during egg rearing and this could be genotype biased, perhaps due to selection against hybrid offspring or selection against certain genotypes under specific environmental conditions.

In this chapter the last approach is used: inferring parental mating patterns from the genotypes of their offspring. Tests are made of genotype assortment and habitat preferences in adult mating site choices. However as direct inference of parental genotypes is not possible, for reasons stated in the methods, assortative mating is not examined. Finally consideration is made of the genetic make-up of the offspring populations relative to the adult population.

5.3 Methods

5.3.1 Data collection

Eggs were collected from a number of sites across the Apahida hybrid zone during the spring of 2000 (collected by Sonja Kohler and Tim Vines) and 2001 (by Sonja Kohler, Thomas Alfert, Lino Ometto and myself). The frequency with which sites were checked for eggs varied depended on their use in other studies – they were sampled on one or only a few occasions sufficient to obtain the required number of egg batches or alternatively every 1-3 days with every new clump of eggs sampled. Around 10 eggs were removed from each batch found within a site and the locations marked and remaining eggs counted so that newly laid eggs were identifiable.

It is difficult to successfully extract and amplify DNA from eggs so these must be grown at least until hatching. Therefore after collection each egg batch was placed in a plastic vial and stored in cool water and transported to the laboratory. Once there, the eggs were placed into plastic cups of dechlorinated tap water and allowed to develop. Every few days a small proportion of the water was changed. Upon hatching tadpoles were transferred to individual cups an allowed to develop further. These tadpoles were fed *ad libitum* on dried powdered nettle leaves. After around 10 days (approximately upon reaching stage 25 using the scheme of Gosner (1960)), the tadpoles were anaesthetised, briefly wiped dry on tissue paper, and then up to three tadpoles from a batch were stored in Eppendorf tubes preserved in 100% ethanol. Where multiple tadpoles were stored in a single tube generally each maintained its integrity so that cross-contamination of DNA from one tadpole to another is likely to have been minimal. Tubes containing tadpoles that had disintegrated during storage were discarded

5.3.2 Sample sizes

There were four suitable species diagnostic marker loci available to genotype the tadpoles (same loci as for adults, two microsatellites, loci 12.19 and 24.12, and two SSCPs, loci 7.4 and 24.11). Between three and twenty two individual tadpoles were genotyped per batch and between six and twenty two batches per site. These sample sizes were chosen to allow estimation of the mean and variance in allele frequency and heterozygote deficit of parents at the maximum number of sites (so sites spanning a range of adult mean allele frequencies and habitat types can be used). These were chosen to be as small as possible so as to include as many sites from the collections as possible (most sites had only a small number of batches sampled). The suitability of these sample sizes was determined by simulations in which egg batches of offspring were generated from randomly paired adults from sites with high variance in adult genotype, showing that these batch sizes give an excellent estimate of the allele frequency in the parent population.

After genotyping was completed new results revealed that batches of mixed paternity were in fact relatively common (from genotyping of a highly polymorphic microsatellite locus that may be used as a marker of parent identity, Sonja Köhler and Beate Nürnberger pers. comm.). Further simulations showed that when batches were of mixed parentage batches estimates of parental mean allele frequency remained very good (table 5.2). The predictions from these simulations are likely to represent an overestimate of the sampling error in many sites from which more than the minimum number of batches were sampled.

	Site 3	Site 4	Site 258	Site 372
Actual mean allele frequency	0.795	0.781	0.597	0.746
Estimated allele freq. (Standard error)	0.800 (0.002)	0.781 (0.002)	0.592 (0.003)	0.742 (0.002)
Allele freq. est. mixed batches (Standard error)	0.801 (0.003)	0.782 (0.003)	0.588 (0.004)	0.738 (0.003)

Table 5.2. The average estimate of parent population allele frequency estimates from 1000 replicates of 6 egg batches and 3 eggs per batch. Egg batches are from a single parent pairs and mixed parent pair egg batches.

5.3.3 Genetic methods

In total 1203 eggs from twenty four sites were genotyped. 288 were genotyped using the protocols described in chapter 2 with the amendments described below. The remaining eggs were genotyped by Sonja Köhler using the same protocols (details given in Köhler 2003). Eggs from around twenty more sites were rejected as there were insufficient egg batches. The only difference in genotyping protocol between that described in chapter 2 for the genotyping of adult tissue is in the preparation of tadpoles. Where possible, the gut was dissected from the tadpoles prior to DNA extraction to minimise the amount of foreign DNA contamination. As the tadpoles were grown from
eggs in the laboratory and fed only plant food from a known source, the risk of cross contamination of toad DNA from one tadpole to another is considered minimal. All genotype analysis by Sonja Köhler was carried out at LMU Munich (described fully in Köhler 2003 and in the appendix). DNA extraction and the genotyping of loci 12.19 and 24.12 were carried out by myself at Edinburgh University and the genotyping of loci 24.11 and 7.4 at LMU Munich according to the protocol outlined in chapter 2.

5.3.4 Data preparation

The data in its raw form may contain errors of a number of kinds, some of which it is possible to detect and which can be corrected or removed. Principally these are due to genotyping errors and the possibility of egg batches being of multiple parentage.

Certain aspects of the mating behaviour of *Bombina* may alter the chances of finding a batch of eggs of mixed paternity. *Bombina bombina* males tend to defend a territory (Lörcher 1969) and hence can monopolise any females entering it reducing the chances of fertilisation by other males. Furthermore males of both taxa grasp a female in amplexus ensuring that they are the closest male to hand when she releases her eggs.

However in the smaller sites typically occupied by *Bombina variegata*, looser mating aggregations are found (Barandun 1990) and thus there may be less opportunity for males to monopolise one female and for an amplexing male to fertilise an entire egg batch. Eggs may be laid in several locations around the site, so each batch collected may not represent an independent family, particularly in puddle sites where there may be a shortage of vegetation on which eggs may be laid. Therefore there is no guarantee that a clump of eggs has only single paternity nor that they are an egg batch from a single female.

Genotyping errors may apparently show alleles not actually present in the parents. Single genotyping errors can only be detected in an individual when a single locus shows an anomalous genotype (i.e. showing an allele unlikely to have not occurred in other siblings, or that suggests three or more alleles at one locus in one parent). These errors can be removed by genotyping the sample again. In small batches it may be impossible to determine which individual is carrying a wrong allele.

Nürnberger *et al.* (in press) describe statistical methods for detecting single egg batches that in fact result from multiple matings. However these methods are only of use when used with larger batch sizes and preferably more (and more polymorphic) marker loci than are available for the eggs in this study.

Another potential error not detected by these methods is that during a single mating eggs are laid in several batches and collected under the assumption that they are from multiple matings. However, this error is unlikely to be biased towards any particular parental pair genotype so will only tend to increase the variability in measures of the average parental genotype.

The small batch sizes in many sites make the above methods inappropriate for determination of errors in batch genotypes. Therefore I consider it unwise to attempt to estimate a joint parental genotype at each locus. Instead I calculate the average joint parental genotype as the mean over all loci in the offspring. The simulations introduced in section 5.4.6 show that under a variety of conditions the average of these parental genotypes within sites gives a good estimate of the true parental mean allele frequency and allele frequency variance.

Assumptions underlying joint parental genotypes

The estimate of a joint parental genotype relies on three assumptions: firstly that there is no meiotic drive and therefore random segregation of alleles in the gametes, secondly that there is no genotype bias in fertilisation rates and thirdly that there has been no genotype-biased selection on the eggs.

There is some doubt over whether these assumptions are actually met. The results of laboratory breeding experiments show that *variegata* alleles occur more frequently than

expected (approximately 3.5% more frequent) in tadpoles of heterozygote-homozygote matings, without evidence of a reduction in numbers of heterozygotes (Köhler 2003). The author thinks that this is more likely to be attributable to epistatic selection against individuals carrying a minority of *bombina* alleles rather than loci creating meiotic drive linked to three of four separate marker loci. These matings took place in a laboratory setting so it is not clear if this is repeatable in naturally fertilised and laid batches (eggs that fail to develop cannot usually be genotyped successfully, so this cannot distinguish between mortality in the eggs and those that are unfertilised). The stages at which selection could act differently in wild fertilised batches are either on sperm before fertilisation, the fertilisation process or very early in egg development. Unfortunately there is no way to determine this from the available data. As this effect is relatively minor I will ignore it in these analyses.

5.4 Analysis methods

5.4.1 Habitat Association

The strength of the habitat association of mating adults is tested by fitting a least-squares regression of the mean of all parent genotypes against the site habitat function score (described in chapter 2) with each site weighted by the sample size in that site. This regression is compared with the regression of the sampled adult genotypes on habitat (similarly weighted). Deviation of the slopes of the regressions is tested by an ANOVA testing the variation amongst regression slopes and on the assumption of homogeneity of regression slopes, deviations of y-intercepts of regressions are tested by ANCOVA (Sokal and Rohlf 1995).

5.4.2 Prezygotic reproductive isolation

The analysis in the previous section looked for habitat assortment in the parents. However such an association would not unequivocally demonstrate that this will result in the isolation of populations on a local scale. For instance it is possible for parent populations to show habitat association without this resulting from a habitat preference in the choice of mating site if the genotypes of adults available for mating in a local area are correlated with the habitat in that area. The following section explicitly tests the assumption that the parents are non-randomly drawn from the available adults against a null hypothesis that the parents are a random assortment of the available adults. Such a test has two facets - the characterisation of the available adult mating pool and testing the probability that the observed site of choice are chosen non-randomly from all those available. I firstly define the pool of mating adults and then two methods of estimating the probability of non-random assortment in the choice of mating site.

I define the local adult mating pool as an area within the local neighbourhood in which all animals may be considered to have equal opportunity of access to the site for the purpose of mating. This is the pool of adults that under the null hypothesis of random assortment and random mating would form a panmictic unit. Vines *et al.* (2003) considered all adults within a distance of 0.3 km from a focal site as a reasonable estimate, given the per-generation and observed within-season dispersal distances and the between-site dispersal rate. In fact it is unlikely that all adults would have an equal opportunity to mate within a given site, but as it is unknown what form the availability would take, and for consistency with earlier analyses I use the same definition. As noted in chapter 3, the arrangement of sites in this hybrid zone means that any choice of local neighbourhood size makes very little difference if the diameter lies between 100m and 1km. Also, to be informative this adult pool must be a range of adults and a number of alternative breeding sites available. There are two areas with closely spaced sites. In the first are (the Apahida-Cojocna valley sites. Site nos. 3- 14, 271, 315-319, 334, 335, 392

& 408, described more fully in chapter 3), there are six sites from which eggs have been sampled, from local area of twenty two sites, of which fifteen have had adult genotypes measured. The second area has a large pond and a nearby puddle and ditch (containing sites 85, 256 & 257).

5.4.3 Quantifying the adult mating site choice

The association of habitat and genotype observed in adults generally need not be reflected in an association between genotype and habitat in mating adults. It is known that migration is prolific and habitat choice in this migration may not show habitat-genotype association (chapter 2). Particularly if males are capable of multiple matings then there may be pressure for males to attempt mating in sites regardless of their habitat.

If habitat is a factor in mating site choice, the mean allele frequency of adults mating within the sites may differ from those in all the sites around them. The effect of the assortment into sites of similar parental genotypes is calculated as:

$\overline{p}_{site} - \overline{p}_{neighbourhood}$

where $\overline{p}_{site}$ is the estimated parent mean allele frequency and $\overline{p}_{neighbourhood}$ is the mean allele frequency of all adults collected in sites within 300m of the focal site. The significance of differences between the site and neighbourhood is calculated using a one-sided Welch's approximate t-test (Sokal and Rohlf 1995) assuming unequal variances. This tests for the significance of the magnitude but not the direction of deviation from the neighbourhood mean. To account for this each site is classified as deviating in the expected direction (i.e the parent mean allele frequency deviates from the pool frequency in the same direction in which the site habitat deviates from the pool habitat). The neighbourhood habitat type is defined as the average of the habitat scores

of all neighbourhood individuals. This is tested over all sites of interest and the significance of the number of deviations in the expected direction is tested by a sign test.

5.4.4 Habitat assortment and habitat availability

The previous method makes no distinction between sites that vary a little or greatly from the average in habitat or allele frequency and fails to account for the habitat available for occupation thus giving an incomplete picture. As a better test, it is possible to determine directly whether the observed sites are significantly different from under random assortment.

Nürnberger *et al.* (in press) and Vines (2002) fit a model to the data that simultaneously attempts to quantify the isolation due to variable propensity to mate in individual sites and assortative mating. The likelihood of an egg batch, given the genotypes of the parental pair, is proportional to the product of the association of the "hybrid index" of the parent pair (the sum of variegata alleles), and the common deviations of parent genotypes from the population mean, with each weighted by a separate parameter. This is summed over all possible parent pairs, weighted by how likely each pair is to have given rise to the batch, and over all batches within a site. Maximum likelihood values and support limits of the strengths of these associations were obtained for each site individually. The probability of assortment of adults into sites together and the mating propensity within the site given are given respectively by the following functions:

$$\exp(\delta(z_i - z_{pop})\ (z_j - z_{pop}))$$

$$\exp(\gamma\ (z_i - z_{pop}))^* \exp(\gamma\ (z_j - z_{pop}))$$

where z_i and z_j are the proportion of *variegata* alleles in each of the putative parents and z_{pop} the mean *variegata* proportion of all adults in the whole population. A positive value of γ indicates that the parent population has on average a higher proportion of *variegata* alleles than the average population member. A positive value of δ indicates that the parents deviate from the average in the same direction. Therefore the parameters contain a contradiction in that a positive value of δ only indicates assortative mating when γ is 0. When $\gamma \neq 0$ part of any assortment of parental genotypes is created by the non-random associations of adults. Put another way, if $\gamma \neq 0$, then z_{pop} is not the mean allele frequency of the mating population. To provide an accurate reflection of both processes, then the assortative mating parameter must be calculated relative to the population of adults mating within the site. For these reasons I do not use this method of assessing assortative mating and mating habitat preference.

As an alternative it is possible to estimate the probability of obtaining the observed parental mean allele frequency given a null hypothesis of random assortment into habitat types and random mating. One potential way to do this would be to calculate the probability of a parent bearing any genotype, given the population from which it is sampled. In this circumstance the probability of any set of parental genotypes (and hence parent mean allele frequency) may also be estimated. However as there are 81 (3^4) distinguishable genotypes at four loci and up to 46 parents per site this requires calculations over an unfeasibly large number of potential parent genotype sets. Instead I use a Monte Carlo approach and randomly sample parents to create a simulated parental populations under an assumption of random assortment. Assortative mating cannot be tested under these assumptions, as it does not allow for estimation of the individual parent genotypes.

5.4.5 Defining the properties of the mating pool

Continuing to define the adult pool as consisting of all adults in the local neighbourhood (i.e. sites within 300m of each other) there are only 192 adults with complete genotypes at three or four marker loci and their genotypes cover only a small fraction of the possible genotype combinations. Furthermore it is quite apparent that these do not represent all the adults in the pool as there are numerous adults that have been captured but have not been genotyped. To maximise the number of genotypes that can make up the simulated mating populations, instead of drawing adults randomly from the observed population I estimate the allele frequency and associations between loci and therefore estimate the frequency of all genotypes and draw random adults from the same genotype distribution.

As adult populations in Apahida have been shown to have both significant heterozygote deficit and linkage disequilibrium (Vines *et al.* 2003), the frequencies of genotypes at each locus cannot be calculated from the allele frequencies alone; linkage disequilibria must be considered also. A diploid population with two alleles at n loci may be fully described by 2^{2n} genotype frequencies or alternatively by the allele frequencies and a set of cumulants or associations between loci (Barton 2000). Estimating these cumulants allows us to determine the probability of a given genotype arising in a population even where this genotype is not directly observed and therefore to draw random adults from this population.

Barton (2000) describes methods of estimating the multilocus cumulants in a population created by admixture of two source populations that differ in the frequency of alleles at a number of loci. The assumption that hybrid genotypes are the result of mixing greatly

simplifies the process as loci may be assumed to be equivalent and thus the state of the population can be described by the allele frequencies, the divergence in allele frequency between the source populations, within genome associations of orders up to the number of loci and between genome associations between all combinations of one up to the number of genes from the maternal and similarly in the paternal genome (Barton 2000).

Under these assumptions the frequency of any genotype can be given as a function of the allele frequencies at the loci involved and a number of statistical associations between loci. These could take the form of multilocus moments but Barton instead uses multilocus cumulants, which are polynomial functions of the moments (Barton 2000). These are defined such that the cumulants of a set of genes describe the association between loci above that created by the lower order cumulants. These associations are designated $\kappa_{0,K}$, which represents a Kth order within genome associations and $\kappa_{J,K}$ an association between K genes in one genome and J in another.

The probability of obtaining a specific genotype may then be described in term of allele frequencies and cumulants. The log likelihood is obtained by summing the log of these probabilities for all genotypes sampled. The likelihood of possible sets of magnitudes of allele frequencies and cumulants can therefore be calculated. Methods of obtaining the maximum likelihood values of cumulants are implemented in the program Mathematica (Wolfram 1999) using the package "Multilocus" written by Nick Barton (available from http://www.icapb.ed.ac.uk/evolgen, along with in depth explanation of the methods it uses). A Metropolis algorithm is used to efficiently search the parameter space to obtain the maximum likelihood estimates of the cumulants.

This analysis was carried out on the combined population of adults from the focal neighbourhood using a divergence between the two source populations of 1 (fixed for alternate alleles) and estimated allele frequencies from the adults, testing for the significance of fitting different allele frequencies, heterozygote deficit, linkage

disequilibrium and higher-order cumulants. The significance of increases in log likelihood given by a more complex hypothesis over a simpler one can be tested against a ½ χ^2_ν distribution where ν is the difference in the number of degrees of freedom between the models (Mangel and Hilborn 1997).

5.4.6 Simulating a parent population

When the population has been described, it is possible to calculate the probability of any given genotype (also calculated using functions of the package "Multilocus"). This enables populations to be drawn at random from a population sharing the same characteristics as the adult pool but giving a more varied range of genotypes. Random parent pairs are drawn to simulate each parent pair in a site under random site choice and under random mating. Sufficient parent pairs are generated to recreate batches within a site and the mean allelic state of these parents is averaged to give an estimate of the parental allele frequency of a site under these conditions.

However these methods do ignore the effect of a small segregation bias seen in previously analyses (Vines 2002, Köhler 2003). This will have a small effect on the mean of the offspring and reduces the variance therefore not increasing the probability of a type I error.

5.4.7 The test statistics

The observed parental mean allele frequencies are compared with simulated populations generated under the null hypothesis using two test statistics. The aim of these tests is to assess three issues:

1) Could the mating populations be a random subset of the adult pool population?

2) Are deviations of site parent population allele frequencies from the adult pool consistently in the expected direction?

3) Could the association of habitat and genotype in the mating populations arise from the same preference shown by the general adult population?

These are assessed by calculating three test statistics on the observed parent population and in replicate parent populations simulated under the assumption of random assortment and random mating. These populations are randomly drawn and this is repeated for 1000 replicate populations.

To first test whether the observed populations are differ significantly from random samples from the adult pool, the difference between a site's average joint parental genotype and the mean allele frequency in the adult pool. These values are summed over all sites in both observed and simulated populations.

To address the second hypothesis, test 1 is repeated but where the deviation from the pool mean allele frequency is in the opposite direction to the deviation of a site mean allele frequency from the average, the deviation is negative. This is tested under the same criteria as test 1.

The test of the third hypothesis is the deviation of each site mean joint parental genotype from the predicted allele frequency, given the site habitat. The habitat association is taken from the regression of all site adult mean allele frequencies against habitat (0.043+0.830*habitat score). This is summed over all sites.

One thousand replicate populations are simulated per site and the test statistics calculated for each replicate. Significance is taken as the proportion of cases in which

the observed value is exceeds (tests 1) or is smaller than (tests 2&3) that in the replicate population.

5.4.8 Genetic composition of the offspring population

If we take the egg sample as being representative of the population of the new generation then we can compare the characteristics of this population with the adult population. Changes between the adult population and the offspring may be generated by habitat assortment or assortative mating. If the hybrid zone is stable then the genetic characteristics of the adult population must be restored between the egg stage and adulthood. This must reflect patterns of selection and migration during these stages.

Concordance of loci

This tests whether the change in one locus from sites of low to high mean *variegata* allele frequency occurs in parallel for all marker loci. Departures of one locus from others may be indicative of difference patterns of selection on the locus (see chapter 3). This is determined by fitting a cubic polynomial model, $p_i = \overline{pq} + 2\overline{pq}(\alpha + \beta(\overline{p} - \overline{q}))$ where $p_i = 1 - q_i$ -is the mean allele frequency at a given locus and $\overline{p}$ is the mean at all loci

This will give an indication of any patterns of selection acting specifically on one locus between laying and genotyping or the consistent bias for one allele at each locus during segregation. A positive value of α indicates on average more *variegata*-like state in the focal locus. A positive β indicates that the focal locus changes over a smaller range of allele frequencies than comparable loci. Examples of possible relationships are shown in figure 5.1.

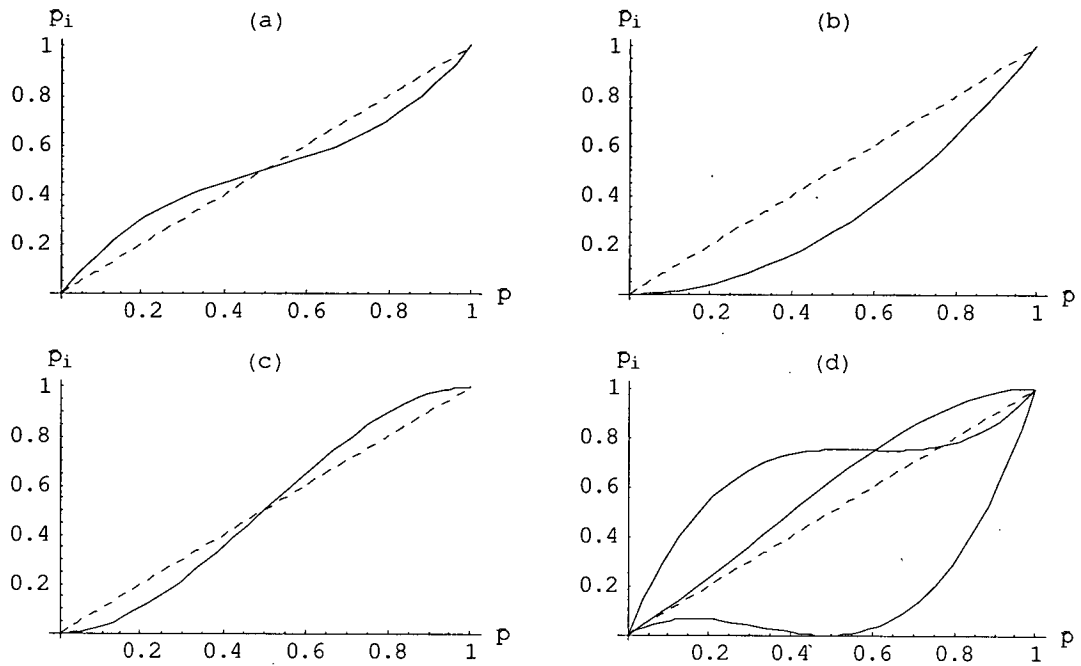


Figure 5.1. Examples of deviations from perfect concordance (dashed lines) of locus p_i from other loci.

a) β=-0.5 b) α=-0.5 c) β=0.5 d) Various possible values of α and β.

Heterozygote deficit and linkage disequilibrium

Departures of the offspring populations from Hardy-Weinberg proportions and linkage equilibrium are determined by maximum likelihood methods (Szymura and Barton 1991, MacCallum *et al.* 1998). These are calculated using the computer package Analyse 1.3 (Barton and Baird 1996, http://www.icapb.ed.ac.uk/evolgen.) which gives parameter estimates and log likelihood scores for heterozygote deficit and linkage

disequilibrium. These can be estimated for all sites individually and across all pooled sites and also independently or commonly for all loci.

The improvement of likelihood is tested against a ½ χ^2_ν distribution where ν is the number of degrees of freedom lost in the more complex model. Support limits are given by the parameter values encompassing a 2 log likelihood unit drop from the maximum likelihood, which approximates to a 95% confidence interval around the parameter.

5.5 Results

5.5.1 The data set

In total 1182 egg batches gave useable genotypes, representing 180 egg batches ranging in size from three to twenty two eggs (Figure 5.2). The average batch allele frequency is 0.691 (s.d. 0.174) and there is a strong skew towards the *variegata* end of the genotype spectrum, at least partially reflecting the bias in collection towards smaller sites. The lowest mean allele frequency is 0.104 and the greatest 0.958.

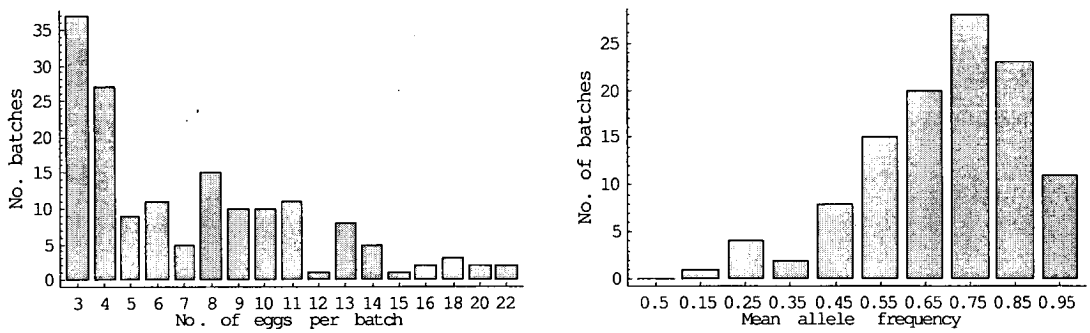


Figure 5.2. Frequency of sizes of egg batch and distribution of mean allele frequencies of egg batches.

5.5.2 Habitat association

As in adults, there is a strong association between the habitat of a site and the mean allele frequency of egg batches within the site. Figure 5.3 shows the regression of mean allele frequency on habitat in both egg batches means and adults within the same sites. In both cases a linear regression gives better fit than quadratic or cubic regressions. It

can be seen that the slope of the regression of parent mean on habitat is shallower than that of adults and that the two regression lines cross at an intermediate habitat (with a habitat score of 0.56). The differences in slope between the two regressions are significant, (Slope: b_{parent}= b_{adult}: b_{parent}=0.33 (s.e. 0.087), b_{adult}=0.62 (s.e. 0.073), $F_{1,60}$=6.19, p=0.0131). As the regression slopes are significantly different significance tests of differences in intercept cannot be calculated. The mean genotypes of adults and parents are not significantly different (mean difference=0.0021, two-sided t-test, t_{62}=0.11, p=0.914).

These results indicate that the parental genotypes and adult genotypes are drawn from a population with a different relationship between site habitat and allele frequency. Habitat explains a greater proportion of the variation in adult site mean allele frequency than the equivalent measure in parents (regression r^2=0.70 in all adults, r^2=0.42 in the parents).

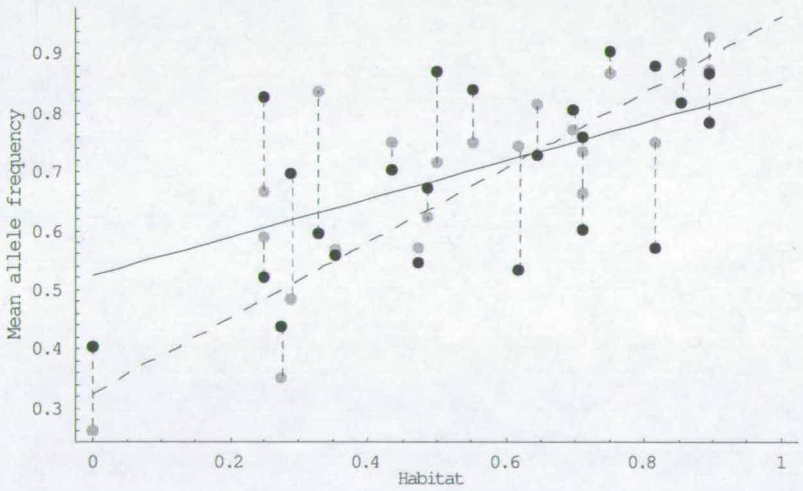


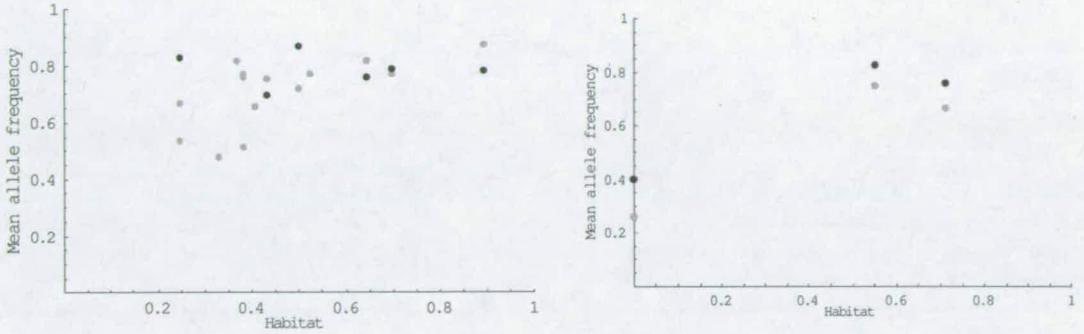
Figure 5.3. Relationship of habitat and genotype in parents. Black points – parent genotypes, grey points – adult genotypes. The vertical dashed lines connect parents and the whole adult populations within the same sites. The solid line shows the best fitting least-squares regression of joint parental genotype on habitat (0.52+0.33x, r^2=0.32, P=0.77x10^{-3}) and the dashed line the regression of mean adult genotype on habitat (0.32+0.64x, r^2=0.72, p=1.02x10^{-9}).

5.5.3 Deviations of sites from the total adult pool

The results of the tests of the direction of deviations of site allele frequencies from pool averages are given in Table 5.2 and deviations shown in two site groups in figure 5.4. It can be seen that most deviations of mean allele frequency do not represent significant deviations from the pool population. Eight of nine of the deviations from the mean are in the direction suggested by the difference in habitat type (two-tailed sign test: p=0.0039). However as the parents tend to be more *variegata*-like in a given habitat type and as the majority of these sites are of a more puddle-like habitat type, this could just be an artefact of these shifted allele frequencies.

Site (group 1)	4	5	271	315	317	318
Δ Habitat	0.026	-0.04	0.42	-0.22	0.22	0.17
Δ p.	0.157	-0.018	0.071	0.11	0.076	0.047
Signif. Δp	p=0.002	n.s. p=0.45	n.s. p=0.33	p=0.001	n.s. p=0.25	n.s. p=0.33
Same direction	+	+	+	-	+	+
Site (group 2)	85		256		257	
Δ Habitat	-0.36		0.20		0.36	
Δ p.	-0.14		0.28		0.215	
Signif. Δp	n.s. p=0.09		p=0		P=0.0001	
Same direction	+		+		+	

Table 5.2. Deviations of focal site habitats and mean allele frequencies from the average of all sites in the local pool



(a) (b)

Figure 5.4. The relationship between site habitat and (a) adult mean allele frequency in black and (b) parent mean allele frequency in grey for the two local neighbourhoods described above, (i) Cojocna-Apahida roadside sites (ii) sites 85, 256,257

5.5.4 Effectiveness of habitat assortment

The results of the simulation method are presented in table 5.3, giving the results of the three tests over all data and tests of significance of the deviations of the parent allele frequencies from the local pool average in individuals sites. It can be seen that by none of the three tests are the observed parent populations significantly different from those that could have been obtained by random sampling of adults from the local pool (test 1), including when the directionality of the deviations is included (test 2). The genotypes of mating adults are also no closer to that predicted given the site habitats (test 3), indicating that there was no habitat preference in the choice of mating site.

Considering the allele frequency of mating adults in individual sites, in only two of six sites are the allele frequencies significantly different a random sample of adults from the local pool.

	Test 1 Deviation of sites from pool mean allele frequency		Test 2 Deviations of site from pool mean in expected direction		Test 3 Deviation of parent allele freq from the "optimum"	
	Observed	*Simulated*	*Observed*	*Simulated*	*Observed*	*Simulated*
Averages	*0.433*	*0.282*	*0.145*	*-0.001*	*0.522*	*0.539*
Test statistics	**Sim≥ obs**	**0.053**	**Sim≥ obs**	**0.146**	**Sim≤ obs**	**0.48**
Site	***271***	***4***	***5***	***315***	***317***	***318***
Observed site deviation	*0.053*	*0.138*	*0.028*	*0.095*	*0.075*	*0.0026*
Sim≥obs.	**0.53**	**0.01**	**0.59**	**0.034**	**0.196**	**0.99**

Table 5.3. Comparisons of simulated and observed parent populations. (Top) Three tests statistics calculated over all sites under simulated random assortment and observed. (Bottom) Test 1 carried out on individual site populations.

5.5.5 Characteristics of the offspring generation

The four marker loci are highly concordant in the egg population (Table 5.4). Although the deviations from perfect concordance are generally slightly larger than those over all Apahida sites (table 3.1 of Vines 2003), none are significantly different from zero and they are not greatly different from those seen if the same is repeated for the adults only in the same sites as these (result not shown). Also any deviations from a straight line are not in consistent directions between adults and eggs in different loci. If differential selection on different loci were responsible then non-zero fitted parameters would be expected to vary in the same direction in eggs and adults (a reversal in the direction of selection on a locus between egg hatching and adulthood would be otherwise be required), which is not observed in this case.

Heterozygote deficit and standardized linkage disequilibrium (F_{IS} and $R=D/\sqrt{(p_1q_1p_2q_1)}$, where p_1, q_1, p_2, q_1 are the allele frequencies at the two loci) are estimated across all sites

and values are given in Table 5.5. However the increase in likelihood from fitting separate values for each locus are not significant (F_{IS}: $\Delta Log(L)=3.47 \rightarrow \frac{1}{2}\chi^2_3$, p=0.16, R: $\Delta Log(L)=3.78 \rightarrow \frac{1}{2}\chi^2_5$, p=0.21). Assuming the same values across loci, the maximum likelihood values of F_{IS} is 0.0239 (support limits 0-0.0553) and of R is 0.0417 (support limits 0.0167-0.0668). In comparison, the adults in the population show a heterozygote deficit of 0.06 (0.02-0.10) and linkage disequilibrium of 0.09 (0.083-0.097).

The support limits for the estimate of heterozygote deficit in the eggs encompasses zero and the most likely estimate is much smaller than in the adult population. As random mating reduces heterozygote deficit to zero, this would appear to suggest that assortative mating within sites is either absent or very weak. The *variegata*-biased segregations seen in heterozygote-homozygote matings (Vines 2002, Köhler 2003) would also cause a deficit of heterozygotes (homozygous *variegata* offspring would be over represented). Any heterozygote deficit in the offspring population is likely to have arisen by selection against heterozygotes in the early stages of development. However considering the change in heterozygote deficit between offspring and adult and if the value of heterozygote deficit remains relatively stable between generations, then this increase clearly indicates that this is being generated at later stages, either by immigration from sites of adults with pure parental genotypes or by selection against heterozygotes.

The estimates of linkage disequilibrium are approximately as expected given random mating in the adults. Linkage disequilibria between unlinked loci are expected to decrease by ½ per generation and the value of R expected in the offspring generation (0.0417) is half the adult estimate (0.090). This is clear evidence of a difference between the adult populations and their offspring. If levels of linkage disequilibrium and heterozygote deficit are stable with time, this implies that considerable amounts of both must be generated by processes occurring between egg laying and adulthood (Figure 5.8).

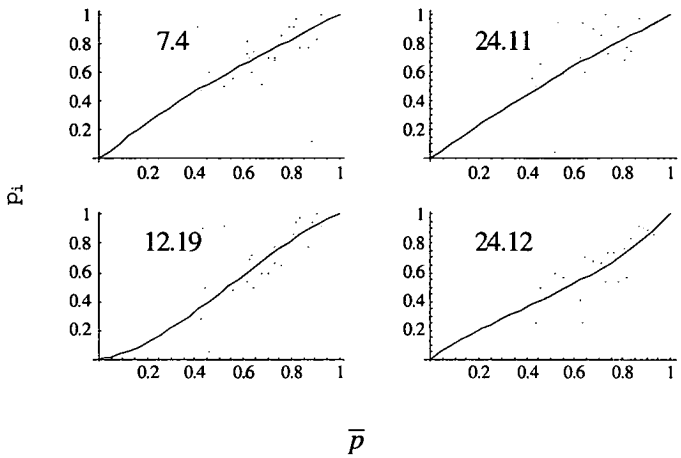


Figure 5.5. Concordance of the frequency of each of four marker loci, 7.4, 24.11, 12.19 and 24.12, to mean over all loci within sites. The lines are least-squares best fits of the model: $p_i = \overline{p} + 2\overline{pq}(\alpha + \beta(\overline{p} - \overline{q}))$

	α	β
24.12	-0.120 (-0.299,0.059)	-0.205 (-0.693,0.284)
12.19	-0.104 (-0.325,0.117)	0.281 (-0.323,0.885)
7.4	0.127 (-0.011, 0.264)	-0.051 (-0.426,0.325)
24.11	.0.097 (-0.156,0.349)	-0.025 (-0.714,0.663)

Table 5.4. Best fit parameters and 95% confidence intervals of concordance parameters for the offspring all loci

Locus	24.12	12.19	7.4	24.11	All Loci
24.12	**0.0962**	-	-	-	F_{IS}=0.0239
12.19	0.0201	0.0257	-	-	R=0.0417
7.4	**0.0776**	0.0683	0.012	-	
24.11	0.0234	**0.0756**	0	0	
Average R	0.0404	0.0547	0.0486	0.0330	

Table 5.5. Maximum likelihood estimates of heterozygote deficits and pairwise standardised linkage disequilibria ($R=D/\sqrt{(p_1q_1p_2q_1)}$) between loci and averaged over all sites. Values of heterozygote deficit are on the diagonal and linkage disequilibria on the off-diagonals. Values significantly different from zero are in bold.

5.5.6 Adult and offspring mean allele frequencies

Figure 5.7 shows the relationship between offspring and adults within the same sites. The slope of the regression of offspring on adult mean allele frequency is significantly less than one (slope=0.69 (s.e. 0.101), p(slope<1), t_{30}=3.04, p=0.0049) and intersects with the expected regression under the null hypothesis ($p_{offspring}=p_{adult}$) at approximately p=0.61. This indicates that the offspring within a site tend to be of more intermediate allele frequency than the adults with a slight tendency towards more *variegata*-like allele frequency. If the offspring allele frequencies are to return to the adult values then the genetic make up of the population must be altered between the egg stage and adults either by emigration of migrants from the pure sites or by selection against hybrids in the sites exhibiting high parental allele frequencies (either intrinsic selection against hybrids or environmental selection in the habitat types associated with these allele frequencies).

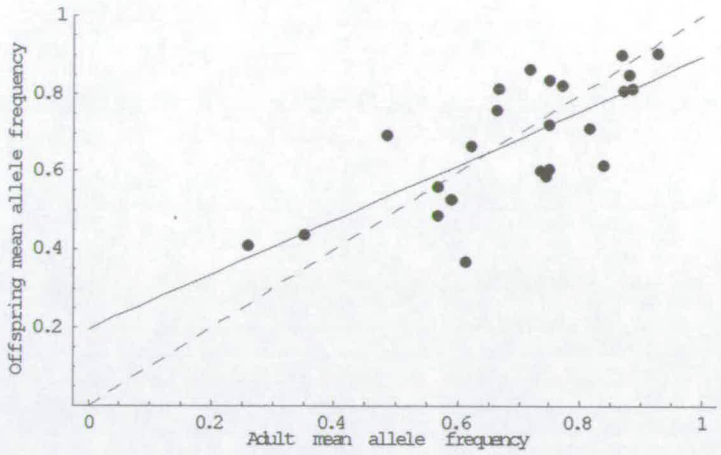


Figure 5.7. The relationship between offspring and adult mean allele frequencies in the same sites. The solid regression line is 0.20+0.69x, , $F_{1,31}$=46.6, $p<10^{-6}$, r^2=0.61. The dashed line shows the relationship where the adult and offspring population are the same.

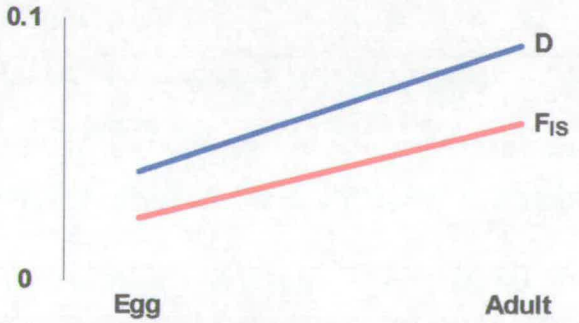


Figure 5.8. The difference in linkage disequilibrium and heterozygote deficit between juveniles and adults within the same sites.

5.6 Discussion

The aims of the analyses in this chapter were to test whether the association of adult genotype with habitat seen in the Apahida hybrid is continued through to mating and, if so, whether this causes prezygotic isolation. The second aim was to test whether the offspring resulting from these matings were shared the same genetic characteristics as the adults that preceded them,

Previous observations of a strong association between the average adult genotype within a site and the habitat of the site, which are consistent even over small spatial scales relative to dispersal (Vines *et al.* 2003) naturally suggest that a similar association might exist in the mating patterns of adults. A previous analysis of the genotypes of wild-laid *Bombina* eggs from a limited number of sites indicated that there was a degree of assortment of parental genotypes into sites but show no evidence of assortative mating of adults within these sites (Vines 2002).

In this chapter wild laid eggs from sites across the Apahida hybrid zone were analysed. In comparison to the early similar study conducted in the same hybrid zone by Vines (2003), a far greater number of sites have been sampled although the samples sizes within each site were much smaller, with a result that joint parental genotypes can only be estimated over all marker loci, not individually, limiting the analyses to those that can be carried out on the mean allele frequency of parents within sites or on the genotypes of individual offspring.

5.6.1 The association of offspring genotype and habitat

A strong regression of average adult genotype on habitat across many sites was demonstrated in parents average allele frequencies within sites against habitat. There is no significant difference between the mean allele frequency of the parents and other adults within the same sites. This suggests either that the mating population is the same as the general adult population in the same sites or that there is equivalent habitat preference in the choice of mating site as that which results in the habitat genotype association in the first instance. Either way this is surprising given the results of the mark-recapture experiment. Individuals move prolifically within a season but their choice of site appears to be random with respect to habitat (chapter 3). Although it seems likely that the mating population remains within the site, it cannot be proven that they do and hence the habitat preferences seen in all adult between site movements may be irrelevant

Tests of genotype-dependent habitat assortment

The tests of genotype-dependent habitat assortment aimed to determine whether, within an area that offers a choice of potential mating sites to a range of adults, overall mating populations form that quite possibly have been assembled at random, but that some sites were unlikely to have been populated at random. These mating assemblages were further tested to test the assumption that habitat preferences were important in site choices. The available data are not ideal for this purpose as they do not give information about any particular parent so it is difficult to test whether individuals showed preference. Instead the site parental mean allele frequencies were tested against the results of simulated populations of identical size.

These simulations showed that the observed populations could have been generated by random assortment from the adult population in a significant fraction of cases but again

there is convincing evidence that some sites were not populated at random. Habitat preference was tested under three hypotheses. Firstly, that populations should deviate allele frequency from the parent pool more than would be expected by taking random parent samples. Secondly, that these deviations should be in the direction expected given the deviation of the focal site habitat from the average habitat. Thirdly that the deviations from the site mean allele frequency may be predicted from the regression of adult site allele frequency on habitat. With all of these tests the allele frequencies of site eggs could have resulted from random assortment from the local adult pool, although individual sites varied significantly in their allele frequency from their local pools.

Therefore under these tests it seems there is no evidence that there is any habitat assortment nor in the choice of mating habitat. The highly significant deviation of some individual sites from the mean would seem to suggest that there is habitat preference expressed by some adults, but this is not enough to create a significant association of parent genotypes over all.

In comparison, an earlier analysis made direct estimates of the genotypes of adults within mating pairs and therefore has a greater ability to detect the association of genotypes of adults within sites and associations of genotypes within the adult pair. However of five sites only one was an association much greater than random noted (Table 5.5 of Vines 2002, site 257 with a value of 0.44 for the habitat assortment parameter). Even this is quite a weak effect; for example, in an area with a mean adult allele frequency of 0.5 where all genotypes exist at equal frequency, an individual homozygous for *variegata* alleles at all four marker loci is around 25% more likely to end up in a site with another individual fixed for *variegata* allele than one with half and half *variegata* and *bombina* alleles.

In both the Monte Carlo method presented here and the likelihood methods of Vines (2003) the assumption is made that the sampled adult population is an accurate reflection

of the population which may make up the mating populations. It is possible that this is not the case and that individuals are included who in fact are not able to mate. Furthermore the definition of "optimum" allele frequency for each habitat type may be inappropriate for egg populations. Errors in either of these affects the chance of detecting assortment in adults.

5.6.2 Genetic composition of the offspring population.

The degree of concordance, heterozygote deficit and linkage disequilibrium observed in the offspring are very much as we would expected from a randomly mating adult population (table 5.4 and 5.5). The observed values are consistent with there having been very little genotype-specific selection between mating and harvesting of eggs and there is little evidence from the heterozygote deficit of either assortative mating or segregation bias as predicted from the previous evidence of these being either weak or absent. Linkage disequilibrium is approximately halved between the adults and offspring, again consistent with an absence of selection. However it is apparent (Figure 5.6) that the mean allele frequencies is sites has been shifted towards being more hybrid. The lack of evidence for genotype-biased selection is consistent with experimental evidence on intrinsic selection in the laboratory and ecological selection on an egg cohort in the field which also showed no genotype-bias in mortality (Köhler 2003).

However, the observed rate of decline of linkage disequilibrium, if it is not regenerated between egg laying and adulthood, would imply that only around 5 generations of hybridisation have occurred. This provides very strong evidence that some processes are creating large amounts of linkage disequilibrium between markers every generation. The results from this chapter further suggest that the mechanism is not preference of adults for different habitats, leaving only epistatic selection favouring parental allele combination or immigration from sites maintained at high allele frequency.

The evidence for which of these are responsible for creating this linkage disequilibrium is not yet conclusive. Although the development and selection during the egg stage are relatively well known (e.g. Barandun 1990, Kruuk *et al.* 1999a, Köhler 2003), the behaviour and ecology of the juvenile stage is something of an unknown (juveniles are rather rarely caught and the youngest sampled adults are estimated to be at least two or three years old). Also during the embryonic and larval stages from Apahida there is no evidence of genotype-dependent intrinsic or extrinsic selection (Köhler 2003). Therefore to recreate the allele frequencies, linkage disequilibria and heterozygote deficit seen in the adult stage from the values estimated in eggs, either the selection on eggs and tadpoles has remained undetected (natural selection in the wild being notoriously difficult to detect) or there are changes in the compositions of populations at the later juvenile or adult stages.

There is evidence that the linkage disequilibrium and heterozygote deficit observed in adult populations is due in part to an excess of "pure" adults in intermediate sites (an excess of adults carrying only *B. bombina* alleles (Vines *et al.* 2003) and *B. variegata* alleles (results not shown)) and occasional hybrids in sites otherwise at very high frequencies of one or other allele type. This strongly suggests that the linkage disequilibrium and heterozygote deficit in intermediate sites is due in a large part to migration (Vines *et al.* 2003).

This conclusion leaves open the question of how the sites of high allele frequency are themselves maintained. The role of selection and habitat preference on the composition of these sites has not yet been studied. Although the experimental methods are difficult to achieve, determination of the relative contribution of migration and selection in maintaining the high allele frequencies in such sites would be an interesting goal of future research in *Bombina* hybrid zones.

5.6.3 Prezygotic isolation through habitat preference

The conclusion that there is no habitat assortment in by mating adults within a local area implies that this causes no prezygotic isolation of the taxa on a local scale. However on a broader scale there are still differences between local areas in the mean allele frequencies of their occupants. This local variation may generate prezygotic isolation relative to global panmixis across the hybrid zone by maintaining local areas with divergent allele frequencies. However such a pattern of mating would not explain the observed levels of linkage disequilibrium and heterozygote deficit seen within adult populations which would still have to result from selection during the juvenile stages or from immigration from areas of differing allele frequency.

From these results it remains unclear precisely how this hybrid population is maintained as no mechanisms of reproductive isolation have been demonstrated to be effective barriers to gene flow in themselves (selection, assortative mating or habitat preference). The reproductive isolation appears not to be strong enough to maintain the genetic composition of these populations in their present state without an effect of selection. Vines *et al.* (2003) estimated that strong selection on several loci is needed to maintain the hybrid zone. A further theoretical analysis is needed to determine whether weak selection, assortative mating and habitat preference in combination are capable of generating significant barriers to gene flow.

Chapter 6. Conclusions

6.1 Introduction

In a hybrid zone neutral divergence between populations will tend to quickly break down unless there are processes to counteract this. Reproductive isolation is the key factor that maintains differences in neutral loci and quantitative traits between hybridising taxa. The structure of genotypes throughout a hybrid zone is a result of the action of reproductive isolation mechanisms (selection, habitat assortment, non-random mating), patterns of migration and interbreeding. This thesis has considered the modes of reproductive isolation between the toads *Bombina bombina* and *Bombina variegata* in a hybrid zone found at Apahida, in North-West Romania. In this chapter I will summarise the results of the observations and experiments described in previous chapters and relate these results in the broader context of the study of different hybrid zones in *Bombina* and other species and for the study of the speciation process generally. I conclude by suggesting avenues of further research suggested by the results presented which could illuminate some of the outstanding questions.

6.2 Results summary

6.2.1 Chapter 1

In chapter one I outlined the theory and empirical evidence for the processes that underlie the speciation process and how the study of the generation of reproductive isolation between populations is the most important aspect of speciation research. One

tool by which we may access the process of reproductive isolation is by examination of diverged populations that have met in secondary contact and exhibit incomplete reproductive isolation, i.e. hybrid zones. These allow us to identify the mechanisms of reproductive isolation in detail in a natural setting. Such experiments might also be possible in a species currently undergoing divergence, but as such a stage would presumably be fleeting, hybrid zones will often offer far better opportunity for such studies.

6.2.2 Chapter 2

In chapter two I outline a number of techniques used in the field and the laboratory enabling the quantification of variation in the Apahida hybrid zone and used extensively throughout all later analyses.

6.2.3 Chapter 3

In chapter three I described the results of a mark-release-recapture experiment carried out on adult *Bombina*. This experiment was used to measure the within and between year dispersal distances of adult *Bombina,* the size and turnover rates of the populations of aquatic sites and to evaluate the role of habitat preference in the adult choice of habitat.

This revealed that, in comparison with the results of a similar experiment carried out on the *Bombina* hybrid zone at Pešćenica, Croatia (MacCallum *et al.* 1998), within and between year dispersal distances are an order of magnitude smaller. However it is apparent that this need not necessarily reflect a intrinsic tendency for toads in Apahida to disperse over smaller distances as the arrangement of aquatic sites in this hybrid zone is

such that sites are infrequently separated by distances greater than around 100m or less than a few kilometres.

Such an arrangement of sites would be likely to have great significance on the structure of the hybrid zone by greatly reducing the rate of gene flow between sub-populations. There is evidence that this structure is partly due to the fact that sampling occurred in years with drought conditions, greatly reducing the number and spread of available aquatic habitat, which need not be the case in wetter years. Also the presence of animals with alleles of only one or other taxon in sites in which they are extremely unlikely to be found there unless they arrived by immigration (Vines *et al.* 2003 and unpublished data), and for whom the nearest likely sites of origin are somewhat distant, indicates that either migration in wetter years occurs over greater distances or land without standing water is not a strong barrier to longer range migration.

Finally the presence of an active habitat preference in the site choice of recaptured animals was tested. This gave some striking results. The association of the mean allele frequency of the occupants of sites and the habitat of that site, measured over a number of ecological parameters, correspond closely in this sample and indeed in adults generally, even when measured over small distances (Vines *et al.* 2003). It is considered extremely unlikely that association of genotype and habitat is created by environmental selection, and experiments on tadpoles give no clear evidence of environmental selection at this life stage (Köhler 2003). However, in the individuals that will later be recaptured in different sites, this relationship is not found.

What is more, this relationship is not created by their choice of different sites. This seems to indicate that at least a proportion of the animals that move site have a much weaker habitat preference than those that do not move site. The strength of habitat preference was estimated with a model that also accounts for the average dispersal distance, the site availability and the range over which habitat may be detected. If habitat

preference was estimated under the assumption that philopatry is a habitat choice in itself, then habitat preference was estimated to be rather strong, but if philopatry is considered not to reflect a habitat choice at all, then there was estimated to be no habitat preference.

To untangle the role of environmental selection and habitat preference more information is needed about the behaviours of these species. As selection is presumed to be most effective at the embryonic stage, more knowledge is required about the relationship of natal site and adult site. The habitat preference itself is strongest in individuals at their first capture of each field season therefore further knowledge is required of philopatry during the rest of the year, and particularly during the presumed abandonment of sites during the Winter and reoccupation in the Spring. Further experiments are need to clarify the role of environmental selection at the tadpole stage.

6.2.4 Chapter 4

In chapter four I examined the distribution of three quantitative traits across the Apahida hybrid zone and the relationship of these traits. This allows us to make inferences about the strength of environmental and intrinsic selection and the relative importance of selection and mixing in creating the observed populations.

On average the three quantitative traits (egg size, leg length and colouration) all closely followed each other and the frequency of marker alleles across the whole hybrid zone and also closely followed changes in site habitat. Linkage disequilibrium between quantitative trait loci, inferred from between trait covariances, peaks in sites of intermediate allele frequency and has a pattern and maximum strength comparable to that observed in previous studies in the Pešćenica hybrid zone (Nürnberger *et al.* 1995, Kruuk 1997). Together with the strong concordance between traits this strongly suggests

that this results largely from mixing in the hybrid sites between individuals more characteristic of the parental species, rather than from alternative means such as spatially or ecologically varying trait optima.

However the presence of linkage disequilibrium in sites at the ends of the habitat spectrum and the excess of linkage disequilibrium between one pair of traits in sites with puddle-like habitats seems to indicate that there is environmental selection acting in these habitats favouring pure species phenotypes and also that there is some degree of immigration into these sites.

6.2.5 Chapter 5

In chapter 5 I quantify the extent to which the observed association of adult genotype and habitat could have resulted from the choice of mating site made by parents. The parents that laid eggs within particular sites seem to differ in genetic composition from populations drawn at random from all local sites in only a few cases. As such, the association of habitat and genotype in both parents populations and in their offspring are weaker than in the adult population generally.

Furthermore there is a striking reduction in the observed degree of heterozygote deficit and linkage disequilibrium. If these are being maintained in the hybrid zone, then this clearly indicates one or both of two processes; either selection acting against hybrids across and within genomes or the immigration at high rates of individuals from sites with high frequencies of the species-characteristic allele types. It is considered that the latter is much more likely. Experiments have so far shown no evidence of intrinsic or extrinsic selection in intermediate sites (Köhler 2003) that could account for the regeneration of levels of heterozygote deficit or linkage disequilibrium, whereas there is

clear evidence of immigration into intermediate sites of individuals carrying a set of allele derived from one species at marker loci (Vines 2002).

6.3 Significance for *Bombina* hybrid zones

The structure of the Apahida hybrid zone differs greatly from that of other *Bombina* hybrid zones that have been studied in detail, at Przemysl, Krakow, Stryj and Pešćenica. These hybrid zones are characterised by steep clines in allele frequency and other traits between two divergent populations (Szymura and Barton 1991, Yankchukov *et al.* 2003, MacCallum 1994) , whereas in the Apahida hybrid zone clines are absent and rather the only observed structure is in the form of an association between genotypes and habitat types (Vines *et al.* 2003). This immediately poses the question of why this hybrid zone differs so drastically from the others. One clear difference is that in Apahida, as opposed to the hybrid zones mentioned above, there are no obvious large local populations containing only one of the species. There are two main hypotheses how this could result in a mosaic structure. The first is that the reproductive isolation between the species in this hybrid zone is so strong that pure populations are maintained in sympatry in alternate habitats, although hybridisation may occur in some sites. Secondly this situation could be rather recent and divergence between the species is in rapid decline, although some geographic structure in genotypes is maintained by residual habitat preferences or ecological selection.

In the Apahida hybrid zone a wide range of hybrid genotypes are found in sites of intermediate allele frequencies (Vines 2002) and both heterozygote deficit and linkage disequilibrium between unlinked marker loci are strong in intermediate sites (Vines *et al.* 2003). The latter are at least partially due to an excess of pure *Bombina bombina* (Vines *et al.* 2003) and *Bombina variegata* (unpublished results) in sites of intermediate

genotype. This clearly indicates that there is a strong degree of mixing of immigrants from sites in which such genotypes are more likely to arise.

Mixing of pure populations is consistent with both hypotheses for the origins of the mosaic hybrid zone, although if the second hypothesis holds, then it means that this hybrid zone must be very recent (~5 generations old, Vines *et al.* 2003). Assuming this is not the case, to understand the hybrid zone it is necessary to identify and quantify the nature of reproductive isolation. If selection against hybrid genotypes alone occurs to maintain the neutral divergence in the hybrid zone, then if migration from *Bombina bombina* populations to *B. variegata* is one-way, under a simple selection scheme, selection strength of 1.7 favouring pure species genotypes at 20 loci could maintain the neutral divergence back the previous ice-age (Vines *et al.* 2003). However selection this strong should be immediately apparent and experimental evidence shows no greater selection on hybrid tadpoles (Köhler 2003) than pure, although there is a suggestion of some weak selection against *Bombina bombina* alleles in a *Bombina variegata* background (Köhler 2003). Another plausible mechanism of reproductive isolation, through assortative mating, was found to be entirely absent (Vines 2002).

The results presented in this thesis go some way to clarifying which modes of reproductive isolation are important in this hybrid zone. Habitat preferences are a likely means by which habitat-genotype association are created, as it is known that the pure species tend to occupy different habitat types and in the Pešćenica hybrid zone active habitat preference has been demonstrated (MacCallum *et al.* 1998). There is a strong association of genotype and habitat in Apahida and furthermore the association of genotype and habitat is maintained even over small scales, strongly suggesting that habitat preference is the cause (Vines *et al.* 2003). In fact over small scales habitat preference has been estimated to be twice as strong in Apahida as Pešćenica (Vines *et al.* 2003) but this is complicated by the clinal structure in the latter hybrid zone. The high rates of turnover between sites within and between years in Apahida gives further

indirect evidence that the habitat-genotype association results from habitat preference (or else the association would be broken down by migration) but within a single season movements between sites do not show any habitat bias with genotype (Chapter 3).

Furthermore the parental aggregations show no habitat-genotype association (chapter 5) indicating that the choice of mating site is much more random with respect to habitat than the adult site choice generally and therefore does not contribute greatly to the association in the next generation. The resulting offspring show greatly reduced heterozygote deficit, linkage disequilibrium and no habitat-genotype association (chapter 5). The heterozygote deficit and linkage disequilibrium may be replenished by the adult stage by either selection against hybrids or immigration but the habitat-genotype association must be recreated by habitat preference.

Although it was suggested above that migration was one-way from *Bombina bombina*-like populations to *Bombina variegata*, in fact there is migration the other way too, leading to linkage disequilibrium in sites with habitats characteristic of both species (chapter 4). The variation in linkage disequilibrium between different traits pairs in these sites does suggest that in these sites alone there is selection for the pure species phenotypes in some trait combinations (chapter 4). Assortative mating has not been tested in these sites but this could also act to reduce the rate of introgression in sites at the extremes of the occupied habitat spectrum, effectively maintaining pools of pure individuals in these sites.

6.4 Significance for speciation research

Mosaic hybrid zones with this broad overlap of populations appear to provide a suitable model of the later stages of sympatric speciation as they contain two diverged population living in the same area but separated by a different choice of habitat. As such this hybrid

zone illustrates some of the problems that will frequently act to prevent speciation in sympatry.

Via (2001) lists conditions that would appear to promote sympatric divergence. Hybridising *Bombina* in Apahida appear to fulfil many of the criteria, possessing amongst others strong habitat choice and mating in their host environments. Evidence suggests that other criteria may also be fulfilled in *Bombina* such as the existence of trade-offs between traits adapted to alternate habitats and genetically based habitat choice. However the results presented in this thesis demonstrate that several of the potential mechanisms of reproductive isolation fail to prevent interbreeding. In sites of intermediate allele frequency, assortative mating is completely absent (Vines 2002). Surprisingly, this analysis also reveals that a strong association of genotype and habitat, in all probability created through a strong and active habitat preference, fails to create assortment of genotypes of the mating populations with sites and hence does not create reproductive isolation (chapter 5). What appears to be key to maintaining this hybrid zone is selection.

Although there is some suggestive evidence of selection for pure species phenotypes at the opposite ends of the habitat spectrum (chapter 4, B. Nürnberger, S. Nell and F. Zajitschek, unpublished results), it also seems probable that selection against hybrids *per se* is a potent factor conserving the integrity of the parental species. Intrinsic selection against hybrids is a key factor in the structuring of clines in other *Bombina* hybrid zones (e.g. Szymura and Barton 1991, MacCallum 1994, Yanchukov *et al.* 2003) . Evidence for this in the Apahida hybrid zone has been inconclusive, largely due to experimental constraints (Köhler 2003) but there is no reason to believe that these populations of *Bombina* differ from others in not showing reduced fitness of hybrids.

The problem this poses for sympatric speciation is that populations differences of this magnitude are the result of around 2-7 million years of divergence in allopatry (Szymura

1993). Populations diverging in sympatry will at first have no such divergence. Incompatibilities of the Dobzhansky-Muller type arise by gradual fixation in allopatric populations, despite incompatibility with other parts of the genome revealed when populations are brought together again. Such a mechanism requires a low rate of gene flow between populations, as on population admixture the neutral alleles that could later form incompatibilities acquire a selective disadvantage and are therefore unlikely to rise to fixation. In demonstrating that sympatric speciation is occurring, it is therefore critical to conclusively demonstrate reproductive isolation rather that features commonly associated as these may exist even where there is some gene flow. Without a drastic reduction in gene flow between the diverging subpopulations it is improbable that complete isolation will ever occur.

6.5 Future directions for the study of *Bombina*

In the discussion sections of each chapter suggestions have been made as to potential future work that would elucidate the processes underlying hybridisation in *Bombina.* These fall into three categories; the better description of the genetic architecture of species differences, a better understanding of how and when selection acts and a general better knowledge of *Bombina* lifestyle and ecology.

There is already some work illuminating the genetic architecture of *Bombina* species differences (Nürnberger *et al.* 2003) and this work is being continued to locate and identify loci controlling important distinguishing features of the species (B. Nürnberger, pers. comm.). Hopefully this work will allows us to better see how the underlying genetics structures the hybrid zone. It would be particularly interesting to identify loci associated with habitat preference to determine if the preference is largely genetically determined and if the apparently random habitat choice of many individuals is indicative of the breakdown of the genetic control of habitat preference, recombination between

habitat preference loci and markers or actually a more general absence of habitat preferences.

The measurement of patterns of selection has produced some interesting and unexpected results (Köhler 2003). However selection is always difficult to measure accurately, particularly in the wild. Small sample sizes, extreme rates of mortality during the experiments and difficulties in maintaining adequate experimental conditions make these results more difficult to interpret. As a result the effects of selection need further examination in Apahida. Also experiments have concentrated on the larval stages. Although this is likely to be the time of life when selection is at its most potent, as there will be many critical and sensitive developmental pathways that may be disrupted, amphibians have a later stage when major rearrangements of body structure occur, during the (as yet unexamined) later stages of metamorphosis. It would be interesting to test the role of selection at this stage also.

The role of selection in the natural environment has necessarily been limited to the smaller and hence more accessible sites (Köhler 2003). It would be extremely interesting to test the patterns of selection across a wider spectrum of sites. The evidence from chapter 4 seems to suggest that selection is acting primarily in the extremes of the habitat range and the patterns of environmental selection by water temperature appears to conform with this conclusion (B. Nürnberger, S. Nell and F. Zajitschek, unpublished results). Unfortunately such sites offer great practical difficulties for experimenters and such experiments might only be possible in semi-natural experimental enclosures.

Finally, it has become apparent at several stages that observations have been made of the hybrid zone that support several possible conclusions and that it would be possible to distinguish between them but for a lack of information about rather basic aspects of *Bombina* lifestyle and ecology. For instance, the behaviours of toadlets after metamorphosis and before they are encountered again as young adults is completely

unknown. It is very important to determine to what extent toadlets disperse from their natal habitat during this phase of life. Such information could allow the inference of the degree to which habitat preference and selection structure the adult populations, without additional experiments on these processes themselves. Also information is missing on the extent to which *Bombina* use their local environment, which would give vital information about habitat choice and environmental selection. Other missing facts regard the dynamics of populations and behaviour and ecology of individuals outside of the three month spring field seasons, particularly the behaviours and mortality rates associated with overwintering, and the age structure of the populations (giving some indication of generation time).

The accumulation of further information about these aspects of *Bombina* life would complement the already in depth knowledge of the *Bombina* hybrid system and contribute greatly to the providing an extremely full description of its underlying processes. Such a picture provides a valuable asset to evolutionary biology both for the study of hybrid zones and of reproductive isolation during speciation.

6.6 Summary

This thesis has described the nature of several modes reproductive isolation maintaining the hybrid zone between *Bombina bombina* and *Bombina variegata* in Apahida, Romania. The results obtained show that although habitat preference is strong and has a major effect in creating the previously observed association of genotype and habitat in adults, this has a relatively minor effect in maintaining the divergence of neutral traits in the population generally. Instead it seems that there is an important role of selection in maintaining populations that are similar to the pure species within the hybrid zone and that these sites play a key role in maintaining the hybrid zone structure.

References

Arditi R. and Dacorogna B.1988. Optimal foraging on arbitrary food distributions and the definition of habitat patches. Am Nat 131:837-846.

Arnold M.L. and Hodges S.A. 1995. Are natural hybrids fit or unfit relative to their parents. Trends Ecol. Evol. 10:67-71.

Arnold ML, Kentner EK, Johnston JA, et al. 2001. Natural hybridisation and fitness. Taxon 50:93-104.

Arnold, S. J., and Wade, M. J.. 1984a. On the measurement of natural and sexual selection: applications. Evolution 8:720–734.

Arnold, S. J., and Wade, M. J.. 1984b. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

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

Babik, W., Szymura, J. M. and Rafinksi, J. 2003. Nuclear markers, mitochondrial DNA and male secondary sexual traits variation in a newt hybrid zone (*Triturus vulgaris* × *T. montandoni*). Mol. Ecol. 12:1913–1930.

Bailey, R.I., Lineham, M.E., Thomas, C.D. and Butlin, R.K. 2003. Measuring dispersal and detecting departures from a random walk model in a grasshopper hybrid zone. Ecological Entomology 28:129–138.

Bailey, R.I., Thomas, C.D. and Butlin, R.K. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic hybrid zone between *Chorthippus brunneus* and *C-jacobsi* (Orthoptera : Acrididae). J. Evol. Biol. 17:108-119

Bajger J. 1980. Diversity of defensive responses in populations of fire toads (*Bòmbina-bombina* and *Bombina-variegata*). Herpetologica 36:133-137.

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

Barandun, J. 1995. Reproductive ecology of *Bombina variegata* (Amphibia). Ph.D. diss. Univ. of Zurich, Switzerland.

Barraclough T.G. and Nee S. 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16 :391-399.

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.

Barraclough, T. G., Vogler, A. P. and Harvey, P. H. 1998. Revealing the factors that promote speciation. Phil. Trans. R. Soc. Lond. B. 353:241-249.

Barraclough T.G. and Vogler A.P. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. Am. Nat. 155:419-434.

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. 2000. Estimating multilocus linkage disequilibria. Heredity 84:373-389.

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

Barton, N. H., and Gale, K. S. 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 Hewitt, G.M. 1981. The genetic basis of hybrid inviability between two chromosomal races of the grasshopper *Podisma pedestris*. Evolution 35:1008-1018.

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

Beebee, T.J.C. 1996. Ecology and conservation of amphibians. Chapman Hall. London.

Bernstein, C., Kacelnik, A. and Krebs, J. R. 1991. Individual decisions and the distribution of predators in a patchy environment. II. The influence of travel costs and structure of the environment. J. Animal. Ecol. 60:205-255.

Bierne N. ,David, P.,Boudry, P. and Bonhomme, F. 2002. Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. Evolution, 56:292–298

Bierne, N. Bonhomme F. and David, P. 2003a. Habitat preference and the marine-speciation paradox. Proc. R. Soc. Lond. B 270:1399–1406.

Bierne, N. Borsa , Daguin, C., Jollivet, D. Viard, F. Bonhomme, F. and David, P. 2003b. Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M. galloprovincialis Molecular Ecology 12:447–461.

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

Bradshaw, Jr. H. D., Otto, K. G., Frewen,B E. McKay J. K. and Schemske, D. W. 1998. Quantitative trait loci affecting differences in Floral Morphology Between Two Species of Monkeyflower (*Mimulus*). Genetics 149: 367–382.

Bridle, J.R. and Butlin, R.K. 2002. Mating signal variation and bimodality in a mosaic hybrid zone between *Chorthippus* grasshopper species. Evolution, 56:1184–1198.

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

Bridle, J.R., Vass-de-Zomba, J. and Butlin, R.K. 2001b. Fine-scale genetic and ecological variation in a *Chorthippus* grasshopper hybrid zone. Ecological Entomology 27:499-504.

Buerkle, C. A., Morris, R. J., Asmussen, M. A. and Rieseberg, L. H. 2000. The likelihood of homoploid hybrid speciation. Heredity 84:441-451.

Burke, J. M., Carney, S. E. and Arnold, M. L. 1998. Hybrid fitness in the Louisiana irises: Analysis of parental and F-1 performance/ Evolution 52:7-43

Bush, G. L. 1994. Sympatric speciation in animals - new wine in old bottles. Trends. Ecol. Evol. 9:285-288.

Butlin R. 1987. Speciation by reinforcement. Trends Ecol. Evol. 2 :8-13.

Butlin, R.K. 1995. Reinforcement – An idea evolving. Trends Ecol. Evol. 10:432-434.

Butlin R.K, Beaumont M. and Hewitt G.M. Selection for assortative mating between parapatric subspecies of grasshopper. Anim. Behav. 43:1045-1047.

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. 1985. Unification of speciation theory in plants and animals. Syst Bot 10:380-390.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in Drosophila. Evolution 43:362-381.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in Drosophila" revisited. Evolution 51:295–303.

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

Cracraft, J. 1989. Speciation and its Ontology. Pp 180-207. In D. Otte and J. Endler (eds.). Speciation and its Consequences. Sinauer Associates Inc. Sunderland Massachusetts

Cruzan, M. B. and Arnold, M. L. 1994. Assortative mating and natural selection in an Iris hybrid zone. Evolution 48:1946-1958.

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.

Darwin, C. 1859. On the origin of species. John Murray. London.

Dasmahaptra, K.K., Blum, M.J., Aiello, A., Hackwell, S., Davies, N., Bermingham, E.P. and Mallet, J. 2002. Inferences from a rapidly moving hybrid zone. Evolution, 56:741–753.

Day, T. 2000. Sexual selection and the evolution of costly female preferences: spatial effects. Evolution, 54:715–730.

Dieckmann, U. and Doebeli, M. 1999. On the origin of species by sympatric speciation. Nature 400:354-357.

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. 1995. Biology of amphibians. McGraw-Hill, New York, NY.

Emelianov I, Dres M, Baltensweiler W, Mallet J. 2001. Host-induced assortative mating in host races of the larch budmoth. Evolution 55 :2002-2010.

Emms, S. K., Arnold, M. L. 1997. The effect of habitat on parental and hybrid fitness: Transplant experiments with Louisiana irises. Evolution 51:1112-1119

Emms, S. K., Hodges, S. A. and Arnold, M. L.. 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.. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

Falconer, D.S and Mackay, T.H. 1995. Introduction to Quantitative Genetics. 3rd Edition. Longman. New York.

Feder J.L., Chilcote C.A. and Bush G.L. 1988. Genetic Differentiation between sympatric host races of the apple maggot fly *Rhagoletis-pomonella*. Nature 336 :61-64.

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

Fritz R.S., Hochwender C.G., Brunsfeld S.J., and Roche B.M. 2003. Genetic architecture of susceptibility to herbivores in hybrid willows. J Evol. Biol. 16:1115-1126

Gardner, K. Buerkle, A. Whitton, J. and Rieseberg, L. 2000. Inferring epistasis in wild sunflower hybrid zones. In Wolf, J. B., Brodie, E. D. and Wade, M. J. (eds.) Epistasis and the Evolutioary Process. Oxford University Press, Oxford.

Gavrilets, S. 2003. Models of speciation: What have we learned in 40 years? Evolution, 57:2197–2215.

Gee, J.M. 2003. How a hybrid zone is maintained: Behavioral mechanisms of interbreeding between California and Gambel's Quail (*Callipepla californica* and *C. gambelii*). Evolution, 57:2407–2415.

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

Geyer, L.B. and Palumbi, S.R. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution, 57:1049–1060.

Gleason J.M. and Ritchie M.G. 1998. Evolution of courtship song and reproductive isolation in the Drosophila willistoni species complex: Do sexual signals diverge the most quickly? Evolution 52:1493-1500.

Gollman, G. 1984. Allozymic and morpholgical variation in the hybrid zone between *Bombina bombina* and *Bombina variegata.* (anura: Discoglossidae) in Northeastern Austria. Z. Zool. Syst. Evol. Forsch. 22:51-64.

Gollmann, G. 1986. Genetic analysis of *Bombina* hybrids from eastern Slovakia. Stud. Herpetol. (Prague) 1986:121-126.

Good, T. P., Ellis, J. C., Annett, C. A. and Pierotti, R. 2000. Bounded hybrid superiority in an avian hybrid zone: Effects of mate, diet, and habitat choice. Evolution, 54:1774–1783.

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

Graham J.H., Freeman D.C and Mcarthur E.D. 1992. Narrow hybrid zone between 2 subspecies of big sagebrush (*Artemisia-tridentata*, Asteraceae) .2. selection gradients and hybrid fitness. Am. J. Bot. 82:709-716.

Grant, B. R. and Grant, P.R. 1997. Evolution of Darwin's finches caused by a rare climatic event. Proc. Roy. Soc. Lond. B 251:111-117.

Grant V. 1971. Plant speciation. Columbia University. New York.

Haldane J.B.S. 1922. Sex-ratio and unisexual sterility in hybrid animals. J. Genetics 12, 101-109.

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

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

Harrison, R. G. 1993. 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.

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

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

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.

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

Hodges, S.A. and Arnold, M.L. 1995. Are natural hybrids fit or unfit relative to their parents? Trends Eco. Evol. 10:67-71.

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

Holenweg AK, Reyer H-U. 2000. Hibernation behavior of *Rana lessonae* and *R-esculenta* in their natural habitat. Oecologia 123 :41-47.

Horbulewicz, L. 1933. Die Verbreitung der Bombinator Arten im Bereiche der Bezirke Dobromil-Przemysl-Jaroslaw (in Polish). Kosmos (Lwow). A58:209-223.

Howard, D. J. and Gregory, P. G. 1993. Post insemination signalling systems and reinforcement. Phil. Trans. R. Soc. Lond. B. 340:231-236.

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.

Hunt, W. G. and Selander, R. K. 1973. Biochemical genetics of hybridisation in European house mice. Heredity 31:11-33.

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.

J.B.S. Haldane (1922) Sex-ratio and unisexual sterility in hybrid animals. J. Genetics 12, 101-109.

Jiggins C.D., Mallet J. 2000. Bimodal hybrid zones and speciation. Trends. Ecol. Evol. 15:250-255.

Jiggins C.D., McMillan W.O., Neukirchen W., Mallet J. 1996. What can hybrid zones tell us about speciation? The case of Heliconius erato and H-himera (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 59:221-242.

Jiggins C.D., Naisbit R.E., Coe RL, and Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry Nature 411:302-305.

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.

Jolly, G. M. 1965. Explicit estimates from capture-recpature data with both death and immigration - stochastic model. Biometrika 52:225-247.

Kawecki T.J., Barton N.H. and Fry J.D. 1997.Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. J. Evol. Biol. 10:407-429.

Kendall, M. 1980. Multivariate Analysis. 2nd ed. Griffin, Lond.

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

Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A., Gibert, P. and Beerli. P. 2001. The Strength of Phenotypic Selection in Natural Populations. The American Naturalist. 157(3): 245-261.

Kirkpatrick, M. and Ravigné, V. 2002. Speciation by Natural and Sexual Selection: Models and Experiments. Am. Nat. S22–S35.

Kocher, T. D. and Sage, R. D. 1986. Further genetic analyses of a hybrid zone between Leopard frogs (Rana pipiens complex) in Central Texas. Evolution 40:21-33.

Koenig, W.D., Van Vuren, D. and Hooge, P.N. 1996. Detectability, philopatry, and the distribution of dispersal distances in vertebrates. Trends Ecol. Evol. 11:514-517.

Köehler, S. 2003. Mechanisms for partial reproductive isolation in a Bombina hybrid zone in Romania. Ph. D. diss., Ludwig-Maximilians-Universitat Munchen.

Kondrashov, A. S. and Kondrashov, F. A. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400: 351-354.

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. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences of the USA 78:3721–3725.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37: 1210–1226.

Lexer, C., Welch, M. E., Raymond, O. and Rieseberg, L. H. 2003. The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): Selection on transgressive characters in a novel hybrid environment. Evolution, 57:1989–2000.

Li, W.H. and Nei, M. 1974. Stable linkage disequilibria without epistasis in subdivided populations. Theor. Pop. Biol. 6:173-183.

Li, Z., Pinson, S. R. M., Park, W. D., Paterson, A. H. and Stansel, J. W. 1997. Epistasis for three grain yield components in Rice (*Oryza sativa* L.). Genetics 145:453-465.

Liou, L. W. and Price, T. D. 1994. Speciation by reinforcement of premating isolation. Evolution 48:1451-1459.

Lörcher, K. 1965. Vergleichende bio-akustische Untersuchungen an der Rot- und Gelbbauchunke, *Bombina bombina* und *B. v. variegata*. Oecologia 3:84-124.

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

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., Nurnberger B. and Barton N.H. 1995. Experimental-Evidence For Habitat Dependent Selection In A *Bombina* Hybrid Zone. Proc. Roy. Soc. Lond. B. 260;257-264.

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.

Mackay, T. F. C. 1985. A quantitative genetic analysis of fitness and its components in *Drosophila melanogaster*. Genetics 111:351-74.

Macnair, M. R. and Cumbes, Q. J. 1989. The genetic architecture of interspecific variation in *Mimulus*. Genetics 122:211-222.

Madej, Z. 1965. Variations in sacral region of spine in *Bombina bombina* (Linnaaeus 1761) and *Bombina variegata* (Linnaeus 1758) (salientia discoglossidae). Acta Biol. Cracov Zoo. 8:85-& 1965.

Madej, Z. 1973. Ecology of European fire-bellied toads (*Bombina*, oken, 1816) [Polish with English summary].

Mallet J. 1986. Hybrid zones of heliconius butterflies in panama and the stability and movement of warning color clines. Heredity 56: 191-202.

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

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

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

Mangel, M and Hilborn, R. 1997. The ecological detective.: confronting models with data. Princeton University Press. Princeton, NJ.

Masterson, J. 1994. Stomatal size in fossil plants - evidence for polyploidy in majority of angiosperms. Science 264:421-424.

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

Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. pp 381–424 in M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds. Species: the units of biodiversity. Chapman & Hall, London.

Maynard Smith, J. 1966. Sympatric speciation. Am. Nat. 100:637–650.

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

Mayr, E. 1963. Animal species and evolution. Harvard University Press, Cambridge Mass.

Michalowski, J. 1958. Geographische Verbreittung der Unken (*Bombina* Oken) zwischen den Flussen Weichsel, Skawa und Raba (Woidwodschaft Krakow). [In Polish with Russian and German summaries]. Acta Zool. Cracow. 3:247-283.

Michalowski, J. 1961. Studies of species characteristics in *Bombina bombina.* (L.) and *Bombina variegata* (L.). Applying the L:T indicator to the classifying purposes. Acta Zool. Cracow. 6:51-59.

Michalowski, 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.

Miyatake, T., Matsumoto, A., Matsuyama, T., Ueda, H,, Toyosato, T. and Tanimura, T. 2002. The period gene and allochronic reproductive isolation in *Bactrocera cucurbitae.* Proc. R. Soc. Lond. B 269:2467–2472.

Mooers A.O., Rundle H.D., Whitlock M.C. 1999. The effects of selection and bottlenecks on male mating success in peripheral isolates. Am. Nat. 153:437-444.

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

Moore, W. S. and P, J. T. 1993. Nature of selection in the Northern Flicker hybrid zone and its implications for speciation theory. pp 196-225 in R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University

Mousseau T.A., Howard D.J. Genetic variation in cricket calling song across a hybrid zone between two sibling species. Evolution. 52:1104-1110

Muller, H. J. 1942. Isolating mechanisms, evolution and temperature. Biol. Symp. 6:71-125.

Murray, Jr, B. G. 1967. Dispersal in vertebrates. Ecology 48:975-978.

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. Evolution, 57:447–459.

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

Nichols, R.A. and Hewitt, G.M. 1986. Population structure and the shape of of a chromosomal cline between two races of *Podisma pedestris* (Orthoptera:Acriridae). Biol. J. Linn. Soc. 29: 301-316.

Nichols, R.A. and Hewitt, G.M. 1988. Genetical and ecological differentiation across a hybrid zone. Ecol. Entomol. 13:39-49.

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

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

Nurnberger, B. Hofman, S, Forg-Brey, B., Praetzell, G., Maclean, A., Szymura, J. M., Abbott, C. M. and Barton, N. H. 2002. A linkage map for the hybridising toads *Bombina bombina* and *B. variegata* (Anura: Discoglossidae. Heredity 91:136–142

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

Nürnberger, B.N., Barton, N. H., Kruuk, L.E.B., and Vines, T.H. Mating patterns in a *Bombina* hybrid zone: inferences from adult and full sib genotypes. Evolution in Press.

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 Turelli M. 2001. The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities Evolution 55:1085-1094.

Orr H.A., Irving S. Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. Genetics 158:1089-1100.

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

Palestrini C. and Rolando A. Differential calls by Carrion and Hooded Crows (*Corvus corone corone* and *C-c-cornix*) in the Alpine hybrid zone. Bird Study 43: 364-370.

Palopi, M. F. and Wu, C-I. 1994. Genetics of hybrid male sterility between *Drosophila* Sibling Species: A complex web of pistasis is revealed in interspecific studies. Genetics 138:329-341.

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

Paterson, H. E. H. 1982. Perspective on speciation by reinforcement. S. Afr. J. Sci. 74:369-371.

Pearson, S. F. and Rowher, S. 2000. Asymmetries in male agression across an avian hybrid zone. Behav. Ecol. 11:93-101.

Prout,T. 1973. Appendix to Mitton, J. B. and Kroehn, R. K. 1973. Populations genetics of marine pelecypods. Genetics 73:483-496.

Quinn, T.P., Unwin, M.J. and Kinnison, M.T. 2000. Evolution of temporal isolation in the wild: Genetic divergence and the timing of migration and breeding by introduced Chinook salmon populations. Evolution 54:1372–1385

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.

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

Reyer H-U., Niederer B. and Hettyey A. 2003. Variation in fertilisation abilities between hemiclonal hybrid and sexual parental males of sympatric water frogs (*Rana lessonae*, R-*esculenta, R-ridibunda*). Behav. Ecol. Sociobiol. 54+274-284.

Reyer, H.-U., G. Frei, and C. Som. 2000. Cryptic female choice: frogs reduce clutch size when amplexed by an undesired male. Proc. Roy. Soc. Lond. B 266:2101-2107.

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

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., Sinervo B., Linder C.R., Ungerer, M.C. and Arias D.M. 1996. Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. Science 272:741-745.

Rieseberg L.H., Soltis D.E. amd Palmer J.D. 1988. A molecular reexamination of introgression between *Helianthus-annuus* and *Helianthus-bolanderi* (Compositae). Evolution 42 :227-238.

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.

Risch M. and Andersen L. Non-random mating of crows (*Corvus corone*) in a hybrid zone of Carrion Crow (*C-c-corone*) and Hooded Crow (*C-c-cornix*). J. Ornithol 139 :173-177.

Ritchie M.G, Butlin R.K. amd Hewitt G.M. 1989. Assortative mating across a hybrid zone in *Chorthippus-Parallelus* (Orthoptera, Acrididae). J. Evol. Biol. 2:339-352.

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, 56:2296–2312.

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.

Sætre G.P., Post E. and Kral M. 1999. Can environmental fluctuation prevent competitive exclusion in sympatric flycatchers? Proc. Roy. Soc. Lond. B. 266:1247-1251.

Saino N. and Villa S. 1992. Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. AUK 109 :543-555.

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.

Schliewen U.K., Tautz D. and Paabo S. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids Nature 368: 629-632.

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

Seber, G. A. F. 1965. A note of the multiple-recapture census. Biometrika 52:249.

Seber, G. A. F. 1982. The estimation of animal abunance and related parameters. Griffin, London.

Secondi, J., Bretagnolle, V., Compagnon, C. and Faivre, B. 2003. Species-specific song convergence in a moving hybrid zone between two passerines. Biol. J. Linn. Soc. 80:507–517.

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

Sinsch, U. 1991. The orientation behavior of amphibians. Herpetol J. 1:541-544.

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.

Skellam, J. G. 1951. Random dispersal in theoretical populations. Biometrika 38:196-218.

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

Slatkin, M. 1985. Gene flow in natural populations. Ann. Rev. Ecol. Syst. 16:393-430.

Smadja, C., Catalan, J. and Ganem, G. 2004. Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. J. Evol. Biol. 17:165–176

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd Ed. W.H. Freeman and Co., New York, NY.

Soltis, D. E. and Soltis, P. S. 1999. Polyploidy: recurrent formation and genome evolution. Trends. Ecol. Evol. 14:348-352.

Southwood, T. R. E. and Henderson, 1998. Ecological methods. Ed. Chapman, London.

Stebbins, G. L. 1950. Variation and evolution in plants. 2nd ed. Chapman, London.

Szymura, J. M. 1976. New data on the hybrid zone between *Bombina bombina* and *Bombina variegata* (Anura:Discoglossidae). Bull. Acad. Polon. Sci. Cl. II 24:355-363.

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. 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 Farana, I. 1978. Inheritance and linkage analysis of five enzyme loci in interspecific hybrids of toadlets, genus *Bombina*. Biochem Genet. 16: 307-319.

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.

Szymura, J. M., Spolsky, C. and Uzzell, T. 1985. Concordant change in mitochondrial and nuclear genes in a hybrid zone between two frog species (genus *Bombina*). Experienta 41:1469-1470.

Tilley, S. G., Verrell, P. A. and Arnold, S. J. 1990. Correspondence between sexual isolation and allozyme differentiation: A test in the salamander *Desmognathus ochrophaeus*. Proc. Natl. Acad. Sci. USA. 87:2715-2719.

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

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

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., Hawthorne D.J. 2002. The genetic architecture of ecological specialization: Correlated gene effects on host use and habitat choice in pea aphids. Am. Nat. 159: S76-S88.

Vines, T.H. 2003. Migration, habitat choice and assortative mating in a *Bombina* hybrid zone. Ph. D. diss., University of Edinburgh.

Vines, T.H. and Barton, N.H. 2003. A new approach to detecting mixed families. Mol. Ecol. 12 :1999-2002.

Vines T.H., Koher S.C., Thiel M., Ghira I., Sands, T.R., MacCallum, C.J., Barton, N.H. and Nurnberger, B.N. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata.* Evolution. 57:1876–1888.

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

Weis A.E. and Abrahamson W.G. 1986. Evolution of host-plant manipulation by gall makers - ecological and genetic-factors in the *Solidago-eurosta* system. Am. Nat. 127:681-695.

Westerbergh, A., Doebley, J. 2002. Morphological traits defining species differences in wild relatives of maize are controlled by multiple quantitative trait loci. Evolution 56: 273-283.

Wiens, J. A. 2001. The landscape context of dispersal. Pp 96-109. In Clobert, J., danchin, E., Dhondt, A. A. and Nichols, J. D. (eds) Dispersal. Oxford University Press. Oxford.

Wiens, J. J. 2004. What Is Speciation and How Should We Study It? Am. Nat. 163:914–923.

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

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Yanchukov, A.W., Hofman S., Szymura, J.M., Morozov-Leonov, S.Yu, Nürnberger, B. and Barton, N.H. The genetic structure of the clinal hybrid zone between *Bombina bombina* and *Bombina variegata* (Anura, Bombinatoridae) near Stryj in western Ukraine. Submitted to Evolution.

Appendix I - Habitat Data

Habitat data from all Apahida sites sampled in 2000, 2001 and 2002. The second column gives the habitat score, H. The next columns give the ecological variables from which H is calculated: the site width, depth, amount of emergent and submerged vegetation, and the proportions of bank vegetation in height classes.

Site	H	Width (m)	Depth (m)	% em	% sub	Bank veg <15	Bank veg 15-50 cm	Bank veg >50cm
85	0.14	40	1.7	60	20	0	10	90
200.3	0.43	2.7	2	10	5	40	10	30
200.4	0.52	3.2	0.7	5	5	30	10	0
200.5	0.47	3	1.2	10	5	10	20	10
200.6	0.42	2	1.7	20	20	20	40	40
200.7	0.39	4	1.5	20	20	30	20	10
200.8	0.43	4	0.95	20	30	70	20	10
200.9	0.54	3.1	0.6	0	20	25	25	0
200.10	0.43	4.4	1.2	10	20	40	20	10
244	0.63	1.5	0.35	5	0	100	0	0
245	0.45	3	0.4	50	20	100	0	0
246	0.5	1.5	0.4	50	5	20	0	10
247	0.71	3	0.1	5	0	70	0	30
248	0.65	3.6	0.1	10	40	100	0	0
249	0.52	1.5	1.2	5	5	80	0	0
250	0.69	1.2	0.15	15	5	80	10	0
251	0.55	5.2	0.12	30	60	80	10	10
252	0.73	1.5	0.05	35	0	100	0	0
253	0.61	1.3	0.2	10	80	50	30	0
254	0.78	0.4	0.17	5	0	0	90	10
255	0.52	1.2	0.2	60	40	15	45	40
256	0.56	1.2	0.25	50	5	80	10	10
257	0.68	0.9	0.2	15	10	70	30	0
258	0.41	8	0.45	40	20	80	0	0
259	0.85	0.2	0.1	10	0	80	10	0
260	0.76	0.6	0.15	5	5	60	20	20
261	0.87	0.2	0.1	0	0	20	40	40
262	0.68	0.2	0.2	50	5	80	20	0
263	0.74	0.6	0.2	5	0	40	0	0
264	0.76	0.4	0.2	5	0	80	0	0
265	0.9	0.4	0.05	0	0	10	0	0
266	0.61	3.6	0.1	40	0	50	0	30
267	0.79	0.6	0.1	10	0	30	20	10
268	0.39	7.6	0.4	45	40	10	80	10
269	0.83	0.4	0.1	5	0	90	10	0
270	0.42	8	0.3	50	20	70	10	0
271	0.82	0.4	0.1	5	5	30	10	0
272	0.71	0.3	0.4	5	5	10	80	0
273	0.74	0.3	0.1	30	50	90	10	0
274	0.76	0.5	0.2	0	0	60	0	0

Appendix I - Habitat Data

Site	H	Width (m)	Depth (m)	% em	% sub	Bank veg <15	Bank veg 15-50 cm	Bank veg >50cm
275	0.82	0.3	0.1	10	10	60	30	5
276	0.5	0.6	0.4	20	10	70	10	0
277	0.77	1	0.1	5	0	50	40	0
279	0.81	0.5	0.1	5	0	0	10	0
280	0.77	0.7	0.1	10	10	10	0	30
281	0.85	0.2	0.12	0	5	50	0	0
282	0.35	15	0.5	50	10	20	60	20
283	0.65	0.5	0.12	50	30	5	55	0
284	0.7	0.5	0.4	0	0	5	75	0
285	0.68	1	0.3	0	5	0	5	0
286	0.6	2.5	0.4	0	5	0	0	2
287	0.82	0.5	0.1	0	5	70	0	0
288	0.77	0.5	0.15	5	0	20	60	20
289	0.52	2.5	0.6	20	0	10	20	60
290	0.36	6	0.5	50	60	0	30	70
291	0.76	0.5	0.15	5	20	0	45	0
292	0.62	5	0.2	0	5	30	50	20
293	0.19	70	2	20	40	85	5	5
294	0.09	50	2	60	50	95	0	5
295	0.64	0.3	0.2	50	30	0	50	50
296	0.6	2	0.12	40	30	85	15	0
297	0.58	1.5	0.4	5	50	20	50	20
298	0.9	0.2	0.05	10	20	10	40	50
299	0.82	0.4	0.1	5	0	30	60	0
300	0.73	1.5	0.1	5	20	10	90	0
301	0.69	1.5	0.2	0	10	60	30	0
302	0.7	0.25	0.15	20	80	50	45	0
303	0.69	0.5	0.4	0	10	40	30	30
304	0.66	1.3	0.3	5	0	50	30	5
305	0.37	6.8	2	10	10	40	0	5
306	0.86	0.2	0.1	5	0	40	50	10
307	0.86	0.2	0.1	5	0	90	0	0
315	0.33	1.8	0.25	90	5	5	5	90
316	0.35	5.6	0.23	80	5	100	0	0
317	0.67	0.4	0.13	1	0	50	0	50
318	0.63	0.6	0.26	2	0	90	10	0
319	0.41	1.5	0.15	80	0	100	0	0
320	0.69	0.4	0.09	2	0	100	0	0
321	0.33	4.1	0.28	85	5	100	0	0
323	0.56	12	0.2	3	5	100	0	0
324								
325	0.31	10	0.13	90	10	100	0	0

Appendix I - Habitat Data

Site	H	Width (m)	Depth (m)	% em	% sub	Bank veg <15	Bank veg 15-50 cm	Bank veg >50cm
327	0.57	0.26	0.1	50	0	100	0	0
328	0.61	1.3	0.24	5	0	100	0	0
329	0.59	0.13	0.08	50	0	98	2	0
330	0.51	16	0.4	10	5	100	0	0
331	0.8	0.1	0	0	100	0	0	0
332	5.1	0.26	5	3	100	0	0	0
333	60	1.2	85	5	20	70	10	10
334	5	0.23	50	10	100	0	0	0
335	45	0.35	70	0	40	60	0	0
336	1.1	0.43	25	30	100	0	0	0
337	0.87	0.05	4	0	100	0	0	0
338	1.2	0.19	30	30	40	0	60	60
339	1.05	0.1	10	0	50	50	0	0
341	5	0.1	0	0	100	0	0	0
342	2.1	0.27	80	0	90	10	0	0
343	5.2	0.26	5	3	100	0	0	0
344	0.25	0.08	50	0	100	0	0	0
345	2.2	0.12	25	10	100	0	0	0
346	1.9	0.15	2	2	95	5	0	0
347	0.8	0.13	2	2	95	5	0	0
348	1.2	0.09	10	5	60	30	10	0
371	0.25	0.13	5	0	50	50	0	0
372	0.95	0.36	0	2	100	0	0	0
373	0.5	0.1	15	10	0	100	0	0
374	2	0.27	80	10	100	0	0	0
375	5	0.28	50	5	10	30	60	60
376	2.4	0.13	10	0	10	20	70	70
377	4.9	0.24	45	55	2	98	0	0
378	3.6	0.31	1	10	30	70	0	0
379	2	0.11	60	0	0	100	0	0
380	1.4	0.28	2	1	70	30	0	0
381	0.33	0.13	80	10	1	50	50	50
382	0.12	0.08	0	0	100	0	0	0
383	0.6	0.01	60	0	60	40	0	0
384	2.5	0	0	0	0	0	0	0
385	5	0.1	0	1	80	20	0	0
386	0.25	0.1	0	0	0	10	0	0
387	0.25	0.1	0	0	0	10	0	0
388	2.6	0.5	0	0	30	70	0	0
389	0.9	0.2	60	10	10	60	30	30
390	2	0.28	60	10	10	60	30	30
391	0.25	0.1	0	0	0	10	0	0
392	0.35	0.14	1	1	100	0	0	0
393	0.25	0.1	5	1	80	20	0	0

Appendix I - Habitat Data

Site	H	Width (m)	Depth (m)	% em	% sub	Bank veg <15	Bank veg 15-50 cm	Bank veg >50cm
394	2.9	0.8	80	10	10	30	60	60
395	5.8	1.7	15	20	100	0	0	0
396	3.4	0.4	0	0	0	0	100	100
397	1.9	0.35	15	15	10	60	30	30
398	3.8	0.2	0	0	100	0	0	0
399	0.25	0.35	0	0	90	5	5	5
400	0.23	0.08	20	0	10	80	10	10
401	0.45	0.25	5	0	0	85	15	15
402	0.55	0.15	100	1	0	0	100	100
403		1.2	60	10	80	20	0	0
404		1.6	0.55	35	10	35	65	5
408	0.55	1.6	1.6	0	0	100	0	0

Appendix II – Adult genotype data

This appendix shows the marker genotype for adults from 2000, 2001 and 2002. The columns show the site of capture, year of capture, individual no. and genotypes at four loci, 7.4, 12.19, 24.11 and 24.11 and the mean *variegata* allele frequency at all four loci. A genotype 0 indicates a *bombina* homozygote, 0.5 a heterozygote and 1 a *variegata* homozygote.

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
3	2000	1	0.5	0.5	1	1	0.75
3	2000	2	1	1	1	0.5	0.875
3	2000	4	1	0.5	1	1	0.875
3	2000	5	1	1	1	1	1.
3	2000	6	0.5	-1	0.5	0.5	0.5
3	2000	7	1	1	0	1	0.75
3	2000	8	1	1	1	1	1.
3	2000	9	1	-1	0.5	1	0.833
3	2000	10	1	1	0.5	1	0.875
3	2000	11	1	1	0.5	1	0.875
3	2000	12	0	-1	0	0.5	0.166
3	2000	13	1	0.5	0.5	1	0.75
4	2000	1	1	1	1	1	1.
4	2000	2	1	1	1	0.5	0.875
4	2000	3	-1	0.5	-1	1	0.75
4	2000	4	1	0.5	1	1	0.875
4	2000	5	1	0.5	1	1	0.875
4	2000	6	0	1	0.5	0.5	0.5
4	2000	7	1	0.5	0	1	0.625
4	2000	8	1	0.5	0.5	1	0.75
4	2000	9	0.5	1	0.5	1	0.75
4	2000	10	0.5	-1	0	0.5	0.333
4	2000	11	0.5	0.5	1	1	0.75
4	2000	12	0	0.5	1	0.5	0.5
5	2000	1	1	1	1	1	1.
5	2000	2	1	1	1	0.5	0.875
5	2000	3	1	0.5	1	1	0.875
5	2000	4	0.5	1	0.5	0.5	0.625
5	2000	5	0	0.5	0.5	0.5	0.375
5	2000	6	0.5	1	0.5	0.5	0.625
5	2000	7	1	0.5	0.5	1	0.75
5	2000	8	1	1	0.5	1	0.875
6	2000	1	1	-1	0.5	1	0.833
6	2000	2	1	1	0.5	1	0.875
6	2000	3	1	0.5	0.5	1	0.75
6	2000	4	1	1	1	0	0.75
6	2000	5	1	0.5	1	1	0.875
6	2000	6	1	0.5	1	1	0.875
6	2000	7	1	1	0.5	0.5	0.75
7	2000	1	1	1	0.5	0.5	0.75
7	2000	2	0.5	0.5	0.5	0.5	0.5
7	2000	3	1	1	0.5	1	0.875
7	2000	4	0	1	0.5	0	0.375
7	2000	5	0	0	0	0	0.
7	2000	6	1	0	0	0.5	0.375
8	2000	1	0	1	1	1	0.75
8	2000	2	0	0.5	0.5	1	0.5
8	2000	3	0	0	0	0	0.
8	2000	4	1	1	0.5	0.5	0.75
8	2000	5	-1	0	0	0	0.
8	2000	7	1	1	0.5	1	0.875
8	2000	8	1	0	1	0.5	0.625
8	2000	9	1	0.5	0	0	0.375
8	2000	10	1	1	0.5	0.5	0.75
9	2000	1	1	0.5	1	0.5	0.75
9	2000	2	1	1	0	1	0.75
9	2000	3	1	0.5	0	1	0.625
9	2000	4	1	1	0.5	1	0.875
9	2000	5	1	0.5	1	1	0.875
9	2000	6	0.5	1	0.5	1	0.75
10	2000	1	1	0.5	1	1	0.875
10	2000	2	0.5	0.5	0	0	0.25
10	2000	3	1	1	1	1	1.
10	2000	4	1	-1	1	0.5	0.833
10	2000	5	0.5	-1	1	1	0.833
244	2000	1	0	0.5	1	1	0.625
244	2000	2	0.5	0.5	0.5	0.5	0.5
244	2000	3	1	1	0.5	1	0.875
244	2000	4	1	1	0.5	1	0.875
244	2000	5	1	0.5	1	1	0.875
244	2000	6	0.5	0.5	1	1	0.75
245	2000	1	0.5	0.5	0	0	0.25
245	2000	2	1	1	0.5	1	0.875
245	2000	3	1	-1	0.5	1	0.833
245	2000	4	0	0	0.5	0	0.125
245	2000	5	0.5	-1	1	1	0.833
245	2000	6	0	0.5	0	0.5	0.25
245	2000	8	1	1	1	1	1.
246	2000	1	0.5	0.5	0.5	0.5	0.5
246	2000	2	0	0.5	0	0	0.125
246	2000	3	1	-1	1	0.5	0.833
246	2000	4	0.5	0	0	0	0.125
246	2000	5	0	0	0	0	0.
247	2000	1	0.5	1	0.5	1	0.75
247	2000	2	0.5	0	0	0.5	0.25
247	2000	3	0.5	0.5	1	0.5	0.625
247	2000	4	1	1	1	1	1.
247	2000	5	1	-1	1	0.5	0.833
247	2000	6	-1	0.5	0.5	0.5	0.5
247	2000	7	1	0.5	0	0.5	0.5
247	2000	8	0.5	0.5	0.5	0	0.375
247	2000	9	1	0.5	0.5	-1	0.666
247	2000	10	0.5	1	0.5	0.5	0.625
247	2000	11	1	1	1	1	1.
247	2000	12	0	0.5	0.5	0.5	0.375

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
248	2000	1	1	0.5	1	0.5	0.75
248	2000	2	0	0.5	1	0.5	0.5
248	2000	3	1	1	1	1	1.
248	2000	4	1	1	1	0.5	0.875
248	2000	5	0.5	-1	1	0.5	0.666
248	2000	6	1	-1	0.5	1	0.833
248	2000	7	1	1	0.5	1	0.875
248	2000	8	0.5	0	0	0	0.125
248	2000	9	0.5	0.5	0.5	0.5	0.5
248	2000	10	1	0.5	0.5	0.5	0.625
249	2000	1	0	0	0	0	0.
249	2000	2	0.5	0.5	1	1	0.75
249	2000	3	1	1	1	0.5	0.875
249	2000	4	0.5	0.5	0.5	1	0.625
249	2000	5	0	-1	1	0.5	0.5
249	2000	6	1	-1	1	1	1.
249	2000	7	0.5	-1	0.5	0.5	0.5
249	2000	8	0.5	1	0.5	0	0.5
250	2000	1	0	0.5	0.5	0.5	0.375
250	2000	2	1	-1	1	1	1.
250	2000	3	1	1	1	1	1.
250	2000	4	0.5	1	1	0.5	0.75
250	2000	5	0.5	1	-1	0.5	0.666
250	2000	6	1	-1	0.5	0.5	0.666
250	2000	7	1	-1	1	1	1.
250	2000	8	1	1	1	0.5	0.875
250	2000	9	0.5	-1	0.5	1	0.666
250	2000	10	1	-1	1	0.5	0.833
250	2000	11	0.5	1	1	0.5	0.75
250	2000	12	0.5	0	0.5	0.5	0.375
251	2000	1	1	1	0.5	1	0.875
251	2000	2	0	0.5	0	0.5	0.25
251	2000	3	0.5	0.5	0.5	0.5	0.5
251	2000	4	0.5	0.5	1	1	0.75
251	2000	5	0	1	0.5	0	0.375
251	2000	6	0	0	0	0	0.
251	2000	7	0	0	0	0	0.
251	2000	8	0.5	0	0.5	0	0.25
251	2000	9	0	0.5	0	0	0.125
251	2000	10	0.5	0.5	0.5	1	0.625
251	2000	11	0.5	1	1	0.5	0.75
251	2000	12	1	0.5	0.5	1	0.75
252	2000	1	0.5	0.5	0.5	0	0.375
252	2000	2	1	1	0.5	1	0.875
252	2000	3	1	1	0	1	0.75
252	2000	4	1	1	1	1	1.
252	2000	5	1	1	1	0.5	0.875
252	2000	6	0.5	1	1	1	0.875
252	2000	7	1	1	0.5	1	0.875
253	2000	1	0.5	0.5	0.5	0.5	0.5
253	2000	2	1	0	0.5	0	0.375
253	2000	3	1	-1	1	1	1.

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
253	2000	4	0.5	1	0.5	1	0.75
253	2000	5	0	0.5	0	0.5	0.25
253	2000	6	0.5	0	0.5	0	0.25
253	2000	7	1	0.5	1	0.5	0.75
253	2000	7	1	-1	0.5	1	0.833
253	2000	8	1	0.5	1	1	0.875
254	2000	1	0.5	1	0.5	1	0.75
254	2000	2	1	0.5	1	0.5	0.75
255	2000	1	1	1	1	1	1.
256	2000	5	1	0	1	0.5	0.625
256	2000	6	1	1	1	1	1.
256	2000	7	1	1	0	1	0.75
256	2000	12	0	0	0.5	0.5	0.25
256	2000	13	1	1	1	1	1.
256	2000	18	1	1	0.5	0.5	0.75
256	2000	19	1	1	0	1	0.75
256	2000	20	1	1	1	1	1.
256	2000	21	0.5	0	0.5	0.5	0.375
256	2000	22	1	1	0.5	0.5	0.75
256	2000	23	1	0.5	1	1	0.875
256	2000	24	0.5	1	1	1	0.875
257	2000	1	0	0.5	0.5	0	0.25
257	2000	2	1	1	0.5	0.5	0.75
257	2000	3	0	0	0.5	0	0.125
257	2000	4	0.5	0	0.5	0.5	0.375
257	2000	5	0.5	0.5	0.5	0	0.375
257	2000	6	1	0.5	0	0	0.375
257	2000	7	0.5	1	1	0.5	0.75
257	2000	8	0.5	1	1	0.5	0.75
257	2000	9	1	0.5	1	1	0.875
257	2000	10	1	0.5	0.5	1	0.75
257	2000	11	1	1	1	1	1.
257	2000	12	1	1	0.5	0.5	0.75
257	2000	13	0.5	0.5	0.5	0.5	0.5
257	2000	14	1	-1	1	0.5	0.833
257	2000	15	0	0.5	0	0.5	0.25
257	2000	16	0.5	1	0.5	0.5	0.625
257	2000	17	0	1	0.5	0.5	0.5
257	2000	18	1	1	1	1	1.
257	2000	19	0.5	1	0.5	1	0.75
257	2000	20	1	-1	0.5	0	0.5
257	2000	21	1	1	0.5	0.5	0.75
257	2000	22	1	1	0.5	0.5	0.75
257	2000	23	1	1	1	1	1.
257	2000	24	1	1	1	1	1.
257	2000	25	1	1	1	0	0.75
257	2000	26	1	1	1	1	1.
257	2000	27	0.5	0.5	0	0	0.25
257	2000	28	1	-1	1	1	1.
258	2000	1	1	1	1	0.5	0.875
258	2000	3	0	0	0	0	0.
258	2000	4	0	0	0	0	0.

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
258	2000	5	0.5	0.5	0	1	0.5
258	2000	6	0	0	0	0	0.
258	2000	7	0	0	0	0	0.
258	2000	8	1	1	1	0.5	0.875
258	2000	9	1	0.5	1	0.5	0.75
258	2000	10	0.5	0.5	0.5	0.5	0.5
258	2000	11	1	1	1	1	1.
258	2000	12	1	-1	0.5	0.5	0.666
258	2000	13	1	0.5	1	1	0.875
258	2000	14	0.5	1	1	0.5	0.75
258	2000	15	1	0.5	1	0.5	0.75
258	2000	16	1	0	1	1	0.75
258	2000	17	0.5	0.5	1	0	0.5
258	2000	18	1	1	0.5	0.5	0.75
258	2000	19	0	0	0	0	0.
258	2000	20	1	-1	0.5	1	0.833
258	2000	21	0.5	0.5	0	0.5	0.375
258	2000	22	0	0.5	0.5	0.5	0.375
258	2000	23	1	-1	0.5	1	0.833
258	2000	24	0.5	0.5	0.5	1	0.625
258	2000	25	0.5	0.5	0.5	0.5	0.5
258	2000	26	1	1	0.5	1	0.875
258	2000	27	0.5	0	0.5	0.5	0.375
258	2000	28	0.5	0	0	0	0.125
258	2000	29	1	1	1	0.5	0.875
258	2000	30	1	-1	0.5	1	0.833
258	2000	31	1	0.5	1	1	0.875
258	2000	32	0	0	0.5	0.5	0.25
259	2000	1	1	1	1	0.5	0.875
259	2000	2	1	0.5	0.5	1	0.75
259	2000	3	1	0.5	1	1	0.875
259	2000	4	0.5	0.5	0.5	1	0.625
259	2000	5	0	0.5	0.5	0.5	0.375
259	2000	6	0.5	0.5	0.5	0.5	0.5
259	2000	7	1	1	1	1	1.
259	2000	8	1	1	1	1	1.
259	2000	9	0.5	-1	0.5	0.5	0.5
259	2000	10	0	1	1	1	0.75
260	2000	1	0	0	0	0.5	0.125
260	2000	2	1	1	1	0.5	0.875
260	2000	3	1	1	0.5	0.5	0.75
260	2000	4	1	1	1	1	1.
260	2000	5	0.5	1	1	1	0.875
260	2000	6	1	0.5	0.5	0.5	0.625
260	2000	7	0.5	-1	0	0.5	0.333
260	2000	8	1	1	1	1	1.
260	2000	9	1	1	0.5	-1	0.833
260	2000	10	0.5	1	0.5	1	0.75
260	2000	11	1	1	1	1	1.
260	2000	12	1	-1	0.5	1	0.833
261	2000	1	1	1	0.5	1	0.875
261	2000	2	1	1	0.5	1	0.875

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
261	2000	3	1	1	1	1	1.
261	2000	4	1	1	0.5	1	0.875
261	2000	5	1	1	1	1	1.
261	2000	6	1	0	1	1	0.75
261	2000	7	0.5	1	1	1	0.875
262	2000	1	1	1	0.5	1	0.875
262	2000	2	1	1	1	0.5	0.875
262	2000	3	1	1	1	0.5	0.875
262	2000	4	1	1	1	1	1.
262	2000	5	1	0.5	1	0.5	0.75
262	2000	6	0.5	1	1	0.5	0.75
262	2000	7	1	-1	0	1	0.666
262	2000	8	1	0.5	1	1	0.875
262	2000	9	1	1	1	1	1.
262	2000	10	1	1	0	1	0.75
262	2000	11	0.5	0.5	0.5	1	0.625
262	2000	12	0.5	-1	1	1	0.833
263	2000	1	0.5	-1	0.5	1	0.666
263	2000	2	0	-1	0.5	1	0.5
263	2000	3	1	1	0.5	1	0.875
263	2000	6	1	1	0.5	1	0.875
263	2000	8	0.5	1	0.5	0.5	0.625
263	2000	9	1	1	1	1	1.
263	2000	9	1	0.5	0.5	1	0.75
263	2000	10	0.5	1	0.5	0.5	0.625
263	2000	11	1	0.5	1	1	0.875
263	2000	12	1	1	1	0.5	0.875
264	2000	1	0	1	0.5	1	0.625
264	2000	2	1	0.5	0.5	0.5	0.625
264	2000	3	1	1	1	0.5	0.875
264	2000	4	0.5	1	0.5	1	0.75
264	2000	5	1	0.5	0.5	1	0.75
264	2000	6	1	0	0.5	0.5	0.5
264	2000	7	1	1	0.5	1	0.875
264	2000	8	0.5	1	1	1	0.875
264	2000	9	1	1	1	1	1.
264	2000	10	1	0.5	1	1	0.875
264	2000	11	0	-1	0.5	0.5	0.333
264	2000	12	1	1	1	1	1.
265	2000	1	1	0.5	1	1	0.875
265	2000	2	1	1	0.5	1	0.875
265	2000	3	1	1	1	1	1.
265	2000	4	1	0.5	1	1	0.875
265	2000	5	1	1	1	1	1.
265	2000	5	1	1	1	1	1.
265	2000	6	1	1	0.5	1	0.875
265	2000	7	1	1	1	1	1.
265	2000	8	1	1	1	1	1.
265	2000	9	1	1	1	1	1.
265	2000	10	1	1	1	1	1.
265	2000	11	1	1	1	1	1.
265	2000	12	1	1	1	1	1.

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
266	2000	1	1	0.5	1	1	0.875
266	2000	2	1	0.5	1	1	0.875
266	2000	3	1	0.5	1	1	0.875
266	2000	4	1	1	0.5	1	0.875
266	2000	5	1	1	1	1	1.
266	2000	6	1	1	1	0.5	0.875
266	2000	7	1	1	1	1	1.
266	2000	8	1	1	1	1	1.
266	2000	9	0.5	1	1	1	0.875
266	2000	10	1	1	1	0.5	0.875
266	2000	11	1	1	1	1	1.
267	2000	1	1	1	1	1	1.
267	2000	2	0.5	-1	0.5	1	0.666
267	2000	3	1	1	1	1	1.
267	2000	4	1	0.5	0.5	1	0.75
267	2000	5	1	0	1	1	0.75
267	2000	6	1	1	1	1	1.
267	2000	7	1	1	0.5	0.5	0.75
267	2000	8	1	1	0.5	1	0.875
267	2000	9	1	1	1	1	1.
267	2000	10	1	1	1	1	1.
267	2000	11	1	-1	1	0.5	0.833
267	2000	12	1	1	1	1	1.
268	2000	1	1	1	1	1	1.
268	2000	2	1	1	1	1	1.
268	2000	3	1	1	0.5	0	0.625
268	2000	4	1	0	0.5	1	0.625
268	2000	5	1	1	1	1	1.
268	2000	6	0.5	1	1	1	0.875
268	2000	7	1	1	1	0.5	0.875
268	2000	8	1	1	1	1	1.
268	2000	9	1	1	1	1	1.
268	2000	10	1	0	1	1	0.75
268	2000	11	1	0	1	0.5	0.625
268	2000	12	1	-1	0	1	0.666
270	2000	1	1	1	1	0.5	0.875
270	2000	2	1	1	1	1	1.
270	2000	3	1	1	0.5	1	0.875
270	2000	4	0.5	1	0.5	1	0.75
270	2000	5	0.5	1	1	1	0.875
270	2000	6	1	0.5	0.5	0.5	0.625
270	2000	7	0.5	1	0.5	1	0.75
270	2000	8	1	1	1	1	1.
270	2000	9	1	0.5	1	0.5	0.75
270	2000	10	1	1	0	1	0.75
270	2000	11	1	1	0.5	1	0.875
270	2000	12	1	1	1	0.5	0.875
271	2000	1	1	1	0.5	1	0.875
271	2000	2	1	0.5	1	0.5	0.75
271	2000	3	0.5	1	0.5	1	0.75
271	2000	4	1	1	1	1	1.
271	2000	5	1	0.5	0.5	1	0.75

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
271	2000	6	1	1	1	1	1.
271	2000	7	1	1	0.5	0.5	0.75
271	2000	8	1	1	1	1	1.
272	2000	1	0.5	1	0.5	1	0.75
272	2000	2	1	1	1	1	1.
272	2000	3	1	1	1	0.5	0.875
272	2000	4	1	1	1	0.5	0.875
272	2000	5	0.5	0.5	0.5	1	0.625
272	2000	6	1	1	1	1	1.
272	2000	8	1	1	1	1	1.
272	2000	9	1	0.5	1	1	0.875
272	2000	10	0.5	-1	1	0.5	0.666
272	2000	11	1	1	1	1	1.
273	2000	1	1	1	0.5	1	0.875
273	2000	2	1	0.5	1	1	0.875
273	2000	3	1	1	1	0.5	0.875
273	2000	4	1	1	1	1	1.
273	2000	5	1	1	1	1	1.
273	2000	6	1	1	1	0.5	0.875
274	2000	1	1	0.5	1	1	0.875
274	2000	2	0.5	-1	0.5	1	0.666
274	2000	3	1	1	1	1	1.
274	2000	4	1	1	1	1	1.
274	2000	5	1	1	1	1	1.
274	2000	6	1	1	0.5	1	0.875
274	2000	7	1	1	1	1	1.
274	2000	8	1	1	0.5	0.5	0.75
274	2000	9	1	0.5	1	1	0.875
274	2000	10	1	1	1	1	1.
274	2000	11	1	1	0	1	0.75
274	2000	12	1	0.5	1	0.5	0.75
275	2000	1	1	-1	1	1	1.
275	2000	2	1	1	1	1	1.
275	2000	3	1	1	0.5	1	0.875
275	2000	4	0.5	1	0.5	1	0.75
275	2000	5	1	1	1	1	1.
275	2000	6	0.5	1	0.5	1	0.75
275	2000	7	1	1	1	1	1.
275	2000	8	1	1	1	1	1.
275	2000	9	1	1	0.5	1	0.875
275	2000	10	1	1	1	1	1.
275	2000	11	1	0.5	1	1	0.875
275	2000	12	1	1	1	1	1.
276	2000	1	0.5	0.5	0	1	0.5
276	2000	2	0.5	0.5	0.5	0.5	0.5
276	2000	3	0.5	0	1	0	0.375
276	2000	4	1	0.5	1	1	0.875
276	2000	5	0.5	0	0.5	0.5	0.375
276	2000	6	1	0.5	1	0.5	0.75
276	2000	7	1	1	1	0.5	0.875
276	2000	8	1	1	0	0.5	0.625
276	2000	9	0.5	0	0	0.5	0.25

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
276	2000	10	0.5	1	0	0	0.375
276	2000	11	0.5	1	1	0.5	0.75
277	2000	1	1	1	1	0.5	0.875
277	2000	3	0	0.5	0.5	0	0.25
277	2000	3	1	0.5	0.5	0.5	0.625
277	2000	4	1	0.5	1	1	0.875
277	2000	5	0.5	1	0	1	0.625
277	2000	6	1	0.5	1	1	0.875
277	2000	7	-1	1	0.5	1	0.833
279	2000	1	0.5	1	1	0.5	0.75
279	2000	3	1	1	1	1	1.
279	2000	4	1	0.5	0.5	0.5	0.625
279	2000	5	1	-1	1	0.5	0.833
279	2000	6	0.5	-1	1	0.5	0.666
280	2000	1	0	0	0.5	1	0.375
280	2000	2	-1	1	1	1	1.
280	2000	3	1	-1	1	1	1.
280	2000	4	0.5	1	1	1	0.875
280	2000	5	1	1	1	1	1.
280	2000	6	1	1	1	1	1.
280	2000	7	1	1	1	1	1.
280	2000	8	1	0.5	1	1	0.875
280	2000	9	1	1	0.5	1	0.875
280	2000	10	0.5	0.5	0.5	1	0.625
280	2000	11	1	1	0.5	1	0.875
280	2000	1	0.5	0.5	0.5	1	0.625
281	2000	2	1	0.5	0	1	0.625
281	2000	3	1	1	-1	1	1.
281	2000	4	0.5	-1	0.5	1	0.666
281	2000	5	-1	1	1	1	1.
281	2000	6	0.5	1	1	1	0.875
281	2000	7	1	1	1	1	1.
281	2000	8	-1	1	1	0.5	0.833
281	2000	9	0.5	1	0.5	1	0.75
281	2000	10	-1	1	0.5	1	0.833
281	2000	11	1	1	1	1	1.
281	2000	12	1	1	0.5	1	0.875
282	2000	1	0	0	0	0	0.
282	2000	2	0	0	0	0	0.
282	2000	3	0	0.5	0.5	1	0.5
282	2000	7	0.5	0.5	0	1	0.5
282	2000	8	0	0	0	0	0.
282	2000	9	0.5	0	0.5	1	0.5
282	2000	10	1	0.5	1	0.5	0.75
282	2000	11	0	0.5	0.5	1	0.5
282	2000	12	0	0.5	0	0.5	0.25
283	2000	1	0	1	0.5	1	0.625
283	2000	2	0	1	0.5	0	0.375
283	2000	3	0.5	1	0	0.5	0.5
283	2000	4	0	1	0.5	0	0.375
283	2000	5	1	0.5	0.5	0.5	0.625
283	2000	6	0	1	0.5	0.5	0.5

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
283	2000	7	0	0	0.5	0.5	0.25
284	2000	1	0	1	1	1	0.75
284	2000	2	0	0.5	0	1	0.375
284	2000	3	0.5	1	0.5	1	0.75
284	2000	4	1	1	1	1	1.
284	2000	5	0	1	0.5	1	0.625
284	2000	6	1	1	0	0	0.5
284	2000	7	1	0	0.5	-1	0.5
285	2000	1	1	1	0	0.5	0.625
285	2000	2	0	1	0.5	1	0.625
285	2000	3	0.5	1	0.5	0.5	0.625
285	2000	4	0.5	0	0.5	0	0.25
285	2000	5	0	1	0.5	1	0.625
285	2000	6	1	0.5	1	1	0.875
285	2000	7	0	1	0.5	1	0.625
285	2000	8	1	1	0.5	1	0.875
285	2000	9	1	1	0.5	0.5	0.75
285	2000	10	0.5	1	0.5	1	0.75
285	2000	11	1	1	0.5	0.5	0.75
285	2000	12	1	1	0.5	1	0.875
285	2000	13	0.5	1	0.5	1	0.75
285	2000	14	1	-1	0.5	0.5	0.666
285	2000	15	1	1	1	0.5	0.875
285	2000	16	1	1	0.5	1	0.875
285	2000	17	1	1	0.5	0.5	0.75
285	2000	18	1	0.5	0.5	1	0.75
285	2000	19	0.5	1	1	1	0.875
285	2000	20	0.5	0.5	1	-1	0.666
285	2000	21	1	1	0.5	1	0.875
285	2000	22	0.5	1	1	1	0.875
285	2000	23	0	0.5	1	0.5	0.5
285	2000	24	1	1	0.5	1	0.875
285	2000	25	1	0.5	0	0	0.375
285	2000	26	1	1	1	1	1.
285	2000	27	1	1	1	1	1.
285	2000	28	1	0	0.5	0.5	0.5
285	2000	29	0	1	0.5	0.5	0.5
285	2000	30	1	1	0.5	1	0.875
285	2000	31	0.5	0.5	0.5	1	0.625
285	2000	32	1	0.5	0.5	1	0.75
285	2000	33	1	0.5	1	1	0.875
285	2000	34	0.5	1	1	1	0.875
286	2000	1	0	0	0.5	1	0.375
286	2000	2	0	1	0	0.5	0.375
286	2000	3	0	1	0.5	-1	0.5
286	2000	4	1	1	0	0.5	0.625
286	2000	5	1	1	0.5	1	0.875
286	2000	6	0.5	1	0.5	0.5	0.625
286	2000	7	1	1	0	0.5	0.625
286	2000	8	1	1	0.5	0.5	0.75
286	2000	9	0.5	1	0.5	0.5	0.625
286	2000	10	0	0.5	0	1	0.375

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
286	2000	11	1	1	0.5	0	0.625
286	2000	12	1	0.5	1	0.5	0.75
286	2000	13	-1	0.5	0.5	1	0.666
286	2000	14	1	0.5	1	0.5	0.75
286	2000	15	0	1	0	1	0.5
286	2000	16	1	1	0.5	1	0.875
286	2000	17	1	0.5	0.5	0.5	0.625
286	2000	18	1	1	0.5	0.5	0.75
287	2000	1	0.5	0.5	0	1	0.5
287	2000	2	1	1	0.5	0.5	0.75
287	2000	3	0.5	0.5	1	1	0.75
287	2000	4	1	0.5	0.5	0.5	0.625
287	2000	5	1	1	0.5	1	0.875
287	2000	6	0	0.5	0.5	0.5	0.375
287	2000	7	1	0	0.5	0	0.375
287	2000	8	0.5	1	0.5	0.5	0.625
287	2000	9	0.5	0.5	0.5	1	0.625
287	2000	10	0.5	0.5	0.5	1	0.625
288	2000	1	0.5	0.5	1	0.5	0.625
288	2000	2	0.5	1	1	1	0.875
288	2000	3	0.5	1	1	1	0.875
288	2000	4	1	1	1	0	0.75
288	2000	5	0.5	0	1	1	0.625
288	2000	6	0.5	1	1	0.5	0.75
288	2000	7	1	1	0.5	0.5	0.75
288	2000	8	0.5	1	1	1	0.875
288	2000	9	0.5	0	0.5	1	0.5
289	2000	1	0.5	1	0	0.5	0.5
289	2000	2	0	1	0.5	0	0.375
289	2000	3	0.5	0.5	0.5	1	0.625
289	2000	4	0.5	0.5	1	0.5	0.625
289	2000	5	0.5	1	0.5	1	0.75
289	2000	6	0.5	0.5	0	0.5	0.375
289	2000	7	1	1	0.5	0	0.625
289	2000	8	0	0.5	0.5	1	0.5
289	2000	9	0.5	1	1	1	0.875
290	2000	1	0.5	1	1	0.5	0.75
290	2000	2	0	0.5	1	0.5	0.5
290	2000	3	0.5	0	0.5	0.5	0.375
290	2000	4	0	0	0	0	0.
290	2000	5	0	0.5	0	1	0.375
290	2000	6	1	1	1	0.5	0.875
290	2000	7	0	0.5	0.5	1	0.5
290	2000	8	1	1	0	0.5	0.625
290	2000	9	0.5	0.5	0.5	1	0.625
291	2000	1	1	0.5	0.5	0.5	0.625
291	2000	2	0.5	1	0.5	0.5	0.625
291	2000	3	1	0.5	0.5	1	0.75
291	2000	4	1	1	0.5	1	0.875
291	2000	5	1	0.5	0.5	1	0.75
291	2000	6	0.5	1	1	1	0.875
291	2000	7	0.5	0.5	0.5	1	0.625

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
291	2000	8	1	0.5	0.5	0.5	0.625
291	2000	9	0.5	1	0.5	0	0.5
292	2000	1	0.5	0	1	1	0.625
292	2000	2	1	0.5	0.5	0.5	0.625
292	2000	3	0.5	0.5	1	0.5	0.625
292	2000	5	1	0.5	0.5	1	0.75
292	2000	6	0	0	0	0	0.
292	2000	7	0.5	1	0.5	1	0.75
292	2000	8	1	0.5	0	1	0.625
292	2000	9	1	0.5	0	0.5	0.5
292	2000	10	0	0	0	0	0.
292	2000	11	1	0	0.5	0.5	0.5
292	2000	12	1	0.5	1	0.5	0.75
292	2000	13	0	0	0	0	0.
292	2000	14	0	0	0	0	0.
292	2000	15	0	0	0	0	0.
293	2000	1	0	0	0	0	0.
293	2000	2	0	0	0	0	0.
293	2000	3	0	0	0	0	0.
293	2000	4	0.5	0	0	0	0.125
293	2000	5	0	0	0	0	0.
293	2000	7	0	0	0	0	0.
293	2000	8	1	0	0	0	0.25
293	2000	9	0.5	0.5	0	0	0.25
293	2000	10	0	0	0	0	0.
293	2000	11	0	0	0	0	0.
293	2000	12	0	0	0	0	0.
293	2000	13	0	0	0	0	0.
293	2000	14	0	0	0	0.5	0.125
293	2000	15	0.5	0	0	0	0.125
294	2000	1	0.5	0.5	0.5	0.5	0.5
294	2000	2	0	0	0.5	0	0.125
294	2000	3	0.5	0.5	1	0.5	0.625
294	2000	4	1	0.5	1	0.5	0.75
294	2000	5	0.5	1	1	1	0.875
295	2000	1	0	0.5	0.5	0.5	0.375
295	2000	2	0	0.5	0.5	1	0.5
295	2000	3	0	0.5	0	0.5	0.25
295	2000	4	1	0.5	1	1	0.875
295	2000	5	0.5	1	1	1	0.875
295	2000	6	0.5	0.5	1	0.5	0.625
295	2000	7	0.5	1	1	0.5	0.75
295	2000	8	1	1	0	1	0.75
295	2000	9	0.5	1	1	1	0.875
295	2000	10	1	1	1	0.5	0.875
295	2000	11	0	1	1	1	0.75
295	2000	12	0	1	1	1	0.75
296	2000	1	0.5	0	1	1	0.625
296	2000	2	0	0.5	1	1	0.625
296	2000	3	0	0.5	1	0.5	0.5
296	2000	5	0	1	1	1	0.75
296	2000	6	0.5	1	0.5	0.5	0.625

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
296	2000	7	0.5	0	0.5	1	0.5
296	2000	8	0.5	1	1	1	0.875
296	2000	9	0.5	1	1	0.5	0.75
296	2000	10	1	1	1	1	1.
296	2000	11	1	0	1	0.5	0.625
296	2000	12	1	0.5	1	0.5	0.75
297	2000	1	1	0.5	0	1	0.625
297	2000	2	1	0.5	1	1	0.875
297	2000	3	0	0	1	1	0.5
297	2000	4	0	0	1	1	0.5
297	2000	5	0	0.5	0.5	0.5	0.375
297	2000	6	1	1	1	0.5	0.875
297	2000	7	1	0.5	1	1	0.875
297	2000	8	0.5	0.5	1	0.5	0.625
297	2000	9	0.5	1	0	0.5	0.5
297	2000	10	0	0	0	0.5	0.125
297	2000	11	0.5	1	1	0.5	0.75
297	2000	12	0.5	1	0	1	0.625
297	2000	13	1	1	0	0.5	0.625
298	2000	1	1	1	1	0.5	0.875
298	2000	2	0.5	1	0.5	0.5	0.625
298	2000	3	1	0	0.5	0.5	0.5
298	2000	4	0	1	1	1	0.75
298	2000	5	0.5	1	1	1	0.875
298	2000	6	0.5	1	0	0.5	0.5
298	2000	7	0.5	0.5	0.5	0.5	0.5
298	2000	8	0.5	0.5	1	0.5	0.625
298	2000	9	1	1	1	0.5	0.875
298	2000	10	1	1	0	0.5	0.625
298	2000	11	1	0.5	0.5	0.5	0.625
298	2000	12	1	1	1	1	1.
298	2000	13	0.5	1	0.5	0	0.5
298	2000	14	0	0.5	0	0.5	0.25
298	2000	15	0	1	0.5	0.5	0.5
299	2000	1	1	0	0.5	1	0.625
299	2000	2	1	0	0.5	0.5	0.5
299	2000	3	0.5	0.5	0.5	0.5	0.5
299	2000	4	0.5	0.5	0.5	1	0.625
299	2000	5	0.5	0.5	1	0	0.5
299	2000	6	0.5	0.5	0	0	0.25
299	2000	7	1	1	0	1	0.75
299	2000	8	0.5	0.5	0.5	0.5	0.5
299	2000	9	1	0.5	0.5	1	0.75
299	2000	10	0	1	0	0	0.25
299	2000	11	1	1	0.5	0	0.625
299	2000	12	1	0.5	0.5	0	0.5
299	2000	13	1	0	1	0	0.5
300	2000	1	1	0.5	1	1	0.875
300	2000	2	0.5	0.5	0.5	0	0.375
300	2000	3	0.5	0.5	1	0	0.5
300	2000	4	0.5	0.5	1	0.5	0.625
300	2000	5	1	1	1	1	1.

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
300	2000	6	1	0	0.5	0.5	0.5
300	2000	7	0.5	0.5	0.5	0.5	0.5
300	2000	8	1	0.5	1	0.5	0.75
300	2000	9	1	1	1	0.5	0.875
300	2000	10	1	0.5	1	0	0.625
300	2000	11	0	1	1	1	0.75
300	2000	12	1	1	1	0	0.75
300	2000	13	1	0.5	1	1	0.875
300	2000	14	0.5	1	1	1	0.875
300	2000	15	0	0.5	1	0.5	0.5
301	2000	1	0.5	1	1	1	0.875
301	2000	2	0.5	1	1	1	0.875
301	2000	3	1	0	0.5	0.5	0.5
301	2000	4	0.5	0.5	0	1	0.5
301	2000	5	1	1	1	0.5	0.875
301	2000	6	0.5	1	1	1	0.875
301	2000	7	0.5	0	0.5	1	0.5
301	2000	8	0.5	0	1	1	0.625
301	2000	9	0.5	0.5	0.5	1	0.625
301	2000	10	0.5	0.5	1	0.5	0.625
301	2000	11	0.5	0.5	1	1	0.75
301	2000	12	0.5	0.5	1	0.5	0.625
301	2000	13	1	1	1	0.5	0.875
301	2000	14	1	0.5	1	1	0.875
301	2000	15	0.5	0	1	1	0.625
302	2000	1	1	0.5	1	0.5	0.75
302	2000	2	0.5	1	0.5	1	0.75
302	2000	3	1	0.5	1	1	0.875
302	2000	4	0.5	1	0	0.5	0.5
302	2000	5	1	0.5	0.5	0.5	0.625
302	2000	6	0.5	1	1	0	0.625
302	2000	7	0.5	1	0.5	0	0.5
302	2000	8	0.5	1	0.5	0.5	0.625
302	2000	9	0.5	0.5	0.5	0.5	0.5
302	2000	10	0.5	1	0.5	0	0.5
302	2000	11	1	1	0	1	0.75
302	2000	12	0	1	0.5	0.5	0.5
302	2000	13	0	0.5	0.5	1	0.5
302	2000	14	1	0	1	0	0.5
302	2000	15	0	0	1	1	0.5
303	2000	1	1	0.5	0	1	0.625
303	2000	2	1	1	0.5	0.5	0.75
303	2000	3	1	1	1	0.5	0.875
303	2000	4	1	0.5	0.5	1	0.75
303	2000	5	0	1	1	1	0.75
303	2000	6	1	0.5	0	1	0.625
303	2000	7	1	0.5	0.5	1	0.75
303	2000	8	1	1	0	1	0.75
303	2000	9	1	-1	0.5	1	0.833
304	2000	1	1	0.5	0.5	1	0.75
304	2000	2	0.5	-1	0.5	1	0.666
304	2000	3	1	0.5	1	0.5	0.75

Appendix II – Adult genotype data

Site	year	Ind	7.4	12.19	24.11	24.12	Allele freq.
304	2000	4	1	0	1	0.5	0.625
304	2000	5	0.5	0.5	0.5	1	0.625
304	2000	6	1	0	0	0	0.25
304	2000	7	1	0	0.5	1	0.625
304	2000	8	1	-1	1	0.5	0.833
304	2000	9	1	0	0.5	1	0.625
304	2000	10	0.5	1	0.5	0.5	0.625
305	2000	1	1	1	1	0.5	0.875
305	2000	2	0.5	0	0.5	0.5	0.375
305	2000	3	0.5	0.5	0.5	0.5	0.5
305	2000	4	1	0.5	1	0.5	0.75
305	2000	5	1	0	1	1	0.75
305	2000	6	1	0	1	1	0.75
305	2000	7	0.5	1	0.5	0.5	0.625
305	2000	8	1	1	0.5	1	0.875
305	2000	9	0.5	1	0.5	1	0.75
306	2000	1	1	1	1	1	1.
306	2000	2	1	1	1	0.5	0.875
306	2000	3	0.5	1	1	1	0.875
306	2000	4	1	0	0.5	1	0.625
306	2000	5	1	1	0.5	1	0.875
306	2000	6	1	0.5	0.5	0.5	0.625
306	2000	7	1	1	1	1	1.
306	2000	8	0.5	0.5	0	0.5	0.375
306	2000	9	0.5	1	0.5	1	0.75
306	2000	10	1	1	0	1	0.75
306	2000	11	1	1	1	1	1.
306	2000	12	0	1	1	1	0.75
307	2000	1	0	-1	0	1	0.333
307	2000	2	1	1	0	0.5	0.625
307	2000	3	0	0	0	0.5	0.125
307	2000	4	0	1	0	0.5	0.375
307	2000	5	0.5	0.5	0.5	0.5	0.5
307	2000	6	0	1	0.5	1	0.625
307	2000	7	0.5	1	1	0.5	0.75
307	2000	8	1	1	0	1	0.75
85	2001	1	0	1	0	0	0.25
85	2001	2	0	-1	0	0	0.
85	2001	3	0.5	1	0.5	0.5	0.625
85	2001	4	0	0.5	0	0	0.125
85	2001	5	0.5	0	0	0.5	0.25
85	2001	6	0.5	0	0.5	0.5	0.375
85	2001	7	0	0	0.5	0	0.125
85	2001	8	0.5	-1	0	0.5	0.333
85	2001	9	0	0	-1	0.5	0.166
85	2001	10	0.5	0	0	0	0.125
85	2001	11	1	0	1	1	0.75
85	2001	12	0	0	0	0	0.
258	2001	1	0.5	0.5	-1	1	0.666
258	2001	2	0.5	0.5	0.5	-1	0.5
258	2001	3	0.5	0.5	0.5	1	0.625
258	2001	4	1	1	0.5	-1	0.833
258	2001	5	1	0.5	0.5	0.5	0.625
258	2001	6	0.5	1	0.5	0.5	0.625
258	2001	7	0.5	0.5	0.5	-1	0.5
258	2001	8	1	0.5	1	1	0.875
258	2001	9	0	0.5	0.5	0.5	0.375
258	2001	11	1	0	1	1	0.75
258	2001	12	0.5	0	0	0	0.125
258	2001	13	1	1	0.5	0	0.625
271	2001	1	1	1	0.5	-1	0.833
271	2001	2	1	-1	1	-1	1.
271	2001	3	0.5	0.5	1	1	0.75
271	2001	4	1	1	1	1	1.
282	2001	1	0	0	-1	0.5	0.166
282	2001	2	0	0	0.5	0.5	0.25
282	2001	3	0	0.5	0.5	0.5	0.375
282	2001	4	0	0	0.5	0	0.125
282	2001	5	0.5	1	0.5	0.5	0.625
282	2001	6	0.5	0.5	0.5	0.5	0.5
282	2001	7	0	1	0.5	0.5	0.5
282	2001	8	0	-1	0.5	0.5	0.333
282	2001	9	0.5	0.5	0.5	0.5	0.5
282	2001	10	0	0.5	0.5	0.5	0.375
282	2001	11	0.5	1	0	0.5	0.5
282	2001	12	0	0	0	0.5	0.125
290	2001	1	1	0.5	0.5	0.5	0.625
290	2001	2	0	1	0	-1	0.333
290	2001	3	0.5	0	0	0	0.125
290	2001	4	1	1	1	1	1.
290	2001	5	0.5	0	0	-1	0.166
290	2001	6	1	-1	0.5	-1	0.75
290	2001	7	0.5	0	0	-1	0.166
290	2001	8	1	-1	0.5	-1	0.75
290	2001	9	0	0	0	-1	0.
290	2001	10	1	0.5	0	0.5	0.5
290	2001	11	0.5	0.5	-1	0.5	0.5
290	2001	12	0.5	1	0	1	0.625
315	2001	1	0	0	0.5	0.5	0.25
315	2001	2	0.5	0	1	0	0.375
315	2001	3	-1	0.5	1	0.5	0.666
315	2001	4	1	0	1	0	0.5
315	2001	5	0.5	1	1	1	0.875
315	2001	6	1	0.5	1	0	0.625
315	2001	7	1	1	1	1	1.
315	2001	8	1	-1	1	1	1.
315	2001	9	1	0.5	1	1	0.875
315	2001	10	0	0.5	1	0.5	0.5
317	2001	1	1	0.5	1	1	0.875
317	2001	2	1	1	1	0.5	0.875
317	2001	3	1	1	1	1	1.
317	2001	4	0	1	1	0	0.5
317	2001	5	0.5	1	0.5	0.5	0.625
317	2001	6	0.5	1	1	1	0.875
317	2001	7	1	1	0.5	-1	0.833
317	2001	8	0.5	1	0.5	-1	0.666

Appendix II – Adult genotype data

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
317	9	2001	0.5	1	0	-1	0.5
317	10	2001	0.5	1	0.5	1	0.75
317	11	2001	1	1	0	1	0.75
317	12	2001	1	1	1	1	1.
318	1	2001	0.5	1	0.5	1	0.75
318	2	2001	1	0.5	0.5	1	0.75
318	3	2001	1	1	-1	0.5	0.833
318	4	2001	0.5	0.5	-1	0.5	0.5
318	5	2001	1	1	1	1	1.
318	6	2001	1	1	1	1	1.
318	7	2001	0.5	1	0.5	1	0.75
318	8	2001	1	1	0.5	1	0.875
318	9	2001	1	1	0.5	1	0.875
318	10	2001	0.5	0.5	0.5	1	0.625
318	11	2001	1	1	1	1	1.
321	1	2001	1	0	0.5	1	0.625
321	2	2001	0	0	1	1	0.5
321	3	2001	0.5	-1	0	0.5	0.333
321	4	2001	0	0	1	0.5	0.375
321	5	2001	0.5	-1	0.5	0.5	0.5
321	6	2001	1	1	0.5	0.5	0.75
321	7	2001	0	1	0.5	0.5	0.5
321	8	2001	1	1	0	0.5	0.625
321	9	2001	-1	1	1	1	1.
321	10	2001	0	1	1	-1	0.666
327	1	2001	0.5	1	1	1	0.875
327	2	2001	0.5	0.5	1	1	0.75
327	3	2001	1	1	1	1	1.
327	4	2001	1	1	1	0	0.75
327	5	2001	1	0.5	1	-1	0.833
327	6	2001	1	1	1	-1	1.
330	1	2001	1	0.5	1	0.5	0.75
330	2	2001	0.5	1	0	1	0.625
330	3	2001	1	0.5	1	1	0.875
330	4	2001	0	0.5	0	0	0.125
330	5	2001	-1	-1	1	0	0.5
330	6	2001	0.5	1	1	0	0.625
330	7	2001	1	0.5	1	0	0.625
330	8	2001	1	1	0.5	-1	0.833
330	9	2001	1	-1	1	-1	1.
330	10	2001	1	0	0.5	1	0.625
330	11	2001	0	0.5	0.5	0	0.25
330	12	2001	0	0.5	1	0.5	0.5
333	1	2001	-1	-1	1	0.5	0.75
333	2	2001	1	0	0.5	1	0.625
333	3	2001	0	0	1	0.5	0.375
333	4	2001	0	0	0.5	-1	0.166
333	5	2001	0	0	0	0	0.
333	6	2001	0	0	0	0	0.
333	7	2001	0	0	0.5	0	0.125
334	1	2001	0.5	1	0	1	0.625
334	2	2001	1	1	0.5	1	0.875

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
334	3	2001	1	1	0	1	0.75
334	4	2001	1	0.5	0	0.5	0.5
334	5	2001	1	1	0	1	0.75
334	6	2001	0.5	0.5	0	1	0.5
334	7	2001	1	1	0	1	0.75
334	8	2001	0.5	0	0	1	0.375
334	9	2001	1	1	0	1	0.75
334	10	2001	1	1	0	-1	0.666
335	1	2001	0.5	0	0	-1	0.166
335	2	2001	0.5	0.5	1	0	0.5
335	3	2001	0.5	0.5	-1	-1	0.5
335	4	2001	0.5	0	0.5	0.5	0.375
335	5	2001	1	0.5	0.5	0.5	0.625
335	6	2001	1	0	1	1	0.75
335	7	2001	0.5	0.5	1	0.5	0.625
335	8	2001	0.5	0.5	1	0.5	0.625
335	9	2001	0.5	0	1	0.5	0.5
335	10	2001	-1	0.5	1	0.5	0.666
342	1	2001	0	1	0.5	0.5	0.5
342	2	2001	1	1	0.5	1	0.875
342	3	2001	1	-1	1	1	1.
342	4	2001	0.5	1	0.5	0.5	0.625
342	5	2001	1	-1	0.5	-1	0.75
342	6	2001	0.5	1	1	1	0.875
342	7	2001	0.5	1	0.5	0.5	0.625
342	8	2001	1	1	0.5	0.5	0.75
342	9	2001	1	1	1	0.5	0.875
342	10	2001	1	0.5	1	-1	0.833
344	1	2001	1	1	0.5	0	0.625
344	2	2001	1	-1	1	1	1.
344	3	2001	0	1	1	1	0.75
344	4	2001	1	1	0.5	-1	0.833
344	5	2001	0.5	0	0.5	-1	0.333
344	6	2001	0.5	0.5	1	1	0.75
344	7	2001	1	0.5	1	1	0.875
344	8	2001	0.5	-1	1	-1	0.75
345	1	2001	0	0.5	1	-1	0.5
345	2	2001	0.5	0.5	1	1	0.75
345	3	2001	0.5	1	1	-1	0.833
345	4	2001	1	0.5	0.5	1	0.75
345	5	2001	-1	-1	1	0.5	0.75
345	6	2001	1	-1	0.5	0.5	0.666
345	8	2001	0	0.5	-1	0.5	0.333
347	1	2001	0.5	0.5	1	0	0.5
347	2	2001	0.5	0.5	1	0	0.5
372	1	2001	0.5	0.5	1	1	0.75
372	2	2001	0	0.5	0.5	1	0.5
372	3	2001	0.5	0.5	0	1	0.5
372	4	2001	0.5	1	1	-1	0.833
372	5	2001	1	0.5	1	1	0.875
372	6	2001	1	1	1	0.5	0.875
372	7	2001	1	1	0.5	1	0.875

Appendix II – Adult genotype data

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
372	8	2001	1	1	0.5	-1	0.833
372	9	2001	1	1	1	-1	1.
372	10	2001	0.5	0	0.5	0.5	0.375
374	1	2001	0.5	0.5	0.5	0.5	0.5
374	2	2001	0.5	0	1	0.5	0.5
374	3	2001	1	1	1	1	1.
374	4	2001	1	1	1	1	1.
374	5	2001	0	1	0.5	1	0.625
374	6	2001	1	1	1	-1	1.
374	7	2001	1	1	1	1	1.
374	8	2001	0.5	1	0.5	0.5	0.625
377	1	2001	1	0.5	1	-1	0.833
377	2	2001	0.5	0.5	1	0.5	0.625
377	3	2001	-1	1	-1	1	1.
377	4	2001	1	1	-1	1	1.
377	5	2001	-1	0.5	1	-1	0.75
377	6	2001	0.5	0	0.5	-1	0.333
377	7	2001	1	1	0	-1	0.666
377	8	2001	0.5	0.5	-1	0.5	0.5
379	1	2001	0	0.5	0.5	-1	0.333
379	2	2001	1	0.5	-1	-1	0.75
379	3	2001	1	1	0	1	0.75
379	4	2001	0.5	1	1	1	0.875
379	5	2001	1	1	-1	1	1.
379	7	2001	0.5	1	0.5	-1	0.666
379	8	2001	0.5	0	0.5	1	0.5
379	9	2001	0.5	1	-1	0.5	0.666
380	1	2001	1	1	0.5	0.5	0.75
380	2	2001	0.5	1	0	0.5	0.5
380	3	2001	-1	1	0	1	0.666
380	4	2001	1	1	1	1	1.
380	5	2001	1	1	0.5	0.5	0.75
380	6	2001	0.5	0.5	1	1	0.75
385	1	2001	1	1	0	0.5	0.625
385	2	2001	0.5	1	0	0.5	0.5
385	3	2001	1	1	-1	1	1.
385	4	2001	1	1	-1	0	0.666
385	5	2001	0	0.5	1	0.5	0.5
388	1	2001	1	1	1	-1	1.
388	2	2001	1	0.5	0.5	1	0.75
388	3	2001	1	1	0.5	1	0.875
388	4	2001	0.5	1	1	0.5	0.75
388	5	2001	-1	-1	0.5	0.5	0.5
388	6	2001	-1	-1	-1	1	1.
389	1	2001	-1	-1	0.5	1	0.75
389	2	2001	-1	1	-1	-1	1.
389	3	2001	1	0.5	0	-1	0.5
389	4	2001	1	-1	0	-1	0.5
389	5	2001	0	0.5	0	0.5	0.25
389	6	2001	0	0.5	-1	0	0.166
389	7	2001	1	1	0	1	0.75
396	1	2001	-1	0	0.5	-1	0.25

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
396	2	2001	1	1	0.5	1	0.875
396	3	2001	1	1	1	1	1.
396	4	2001	0	1	1	-1	0.666
396	5	2001	0.5	0.5	-1	0.5	0.5
396	6	2001	0.5	-1	1	0.5	0.666
396	7	2001	0.5	-1	0.5	0.5	0.5
396	8	2001	0.5	1	0.5	0.5	0.625
397	1	2001	1	1	1	-1	1.
397	2	2001	0.5	1	0.5	-1	0.666
397	3	2001	1	0.5	0	-1	0.5
397	4	2001	0.5	0.5	0	1	0.5
397	5	2001	0.5	0	0	0	0.125
397	6	2001	1	0.5	0	0.5	0.5
397	7	2001	1	1	0.5	1	0.875
397	8	2001	1	0.5	0	0.5	0.5
3	2	2001	0.5	0.5	-1	-1	0.5
3	11	2001	0.5	0.5	-1	0.5	0.5
3	14	2001	0.5	0.5	-1	-1	0.5
3	15	2001	0.5	0	0.5	-1	0.333
3	26	2001	0.5	0.5	-1	0.5	0.5
4	2	2001	0.5	0	0.5	0.5	0.375
4	4	2001	0.5	0.5	-1	-1	0.5
4	7	2001	0.5	0.5	0	0.5	0.375
4	8	2001	0.5	0.5	-1	-1	0.5
4	14	2001	0.5	0	-1	0.5	0.333
4	20	2001	0.5	0	0.5	0.5	0.375
5	1	2001	0.5	0.5	0.5	0.5	0.5
5	3	2001	0	0.5	0	0.5	0.25
5	5	2001	0	0.5	-1	0	0.166
6	1	2001	0.5	0.5	0.5	0.5	0.5
6	3	2001	0.5	0.5	-1	0.5	0.5
6	4	2001	0.5	0.5	-1	0.5	0.5
6	5	2001	0.5	0.5	0.5	0	0.375
6	6	2001	-1	0.5	-1	-1	0.5
7	1	2001	0	0.5	0	0	0.125
7	2	2001	0.5	0.5	0.5	0.5	0.5
7	4	2001	0.5	0.5	-1	0.5	0.5
7	5	2001	0.5	0.5	0.5	0.5	0.5
8	3	2001	0.5	0.5	0.5	0.5	0.5
8	4	2001	-1	0.5	0.5	0.5	0.5
8	5	2001	0.5	0.5	0.5	0.5	0.5
9	1	2001	0.5	0.5	-1	-1	0.5
9	5	2001	0.5	0.5	-1	-1	0.5
9	22	2001	0.5	0.5	0.5	0.5	0.5
9	25	2001	0.5	0	-1	0.5	0.333
10	2	2001	0.5	0.5	-1	0.5	0.5
10	7	2001	0	0.5	-1	0	0.166
11	3	2001	0.5	0.5	0	0.5	0.375
11	4	2001	0.5	-1	-1	0.5	0.5
11	8	2001	0.5	-1	0.5	0.5	0.5
11	10	2001	0.5	-1	0.5	0.5	0.5
11	12	2001	0.5	-1	0.5	0.5	0.5

Appendix II – Adult genotype data

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
11	14	2001	0.5	0.5	0	0.5	0.375
11	16	2001	0.5	0	-1	0.5	0.333
13	1	2001	0.5	0.5	0.5	0.5	0.5
13	2	2001	0.5	0.5	0.5	0.5	0.5
13	3	2001	0.5	0.5	0.5	0.5	0.5
13	5	2001	0.5	0.5	0.5	0.5	0.5
14	1	2001	0.5	0	-1	0.5	0.333
258	15	2001	0.5	-1	0.5	0.5	0.5
258	16	2001	0	0.5	0.5	0.5	0.375
258	17	2001	0.5	0.5	0.5	0.5	0.5
258	22	2001	0.5	0.5	0	0.5	0.375
258	27	2001	0.5	0.5	0	0.5	0.375
258	32	2001	0.5	0.5	0.5	0.5	0.5
258	33	2001	-1	0.5	-1	-1	0.5
258	44	2001	0.5	-1	0.5	0.5	0.5
267	1	2001	0.5	-1	-1	-1	0.5
267	2	2001	0.5	-1	-1	-1	0.5
271	8	2001	-1	-1	0.5	0.5	0.5
275	8	2001	-1	0.5	-1	-1	0.5
275	18	2001	0	-1	-1	-1	0.
289	2	2001	0.5	0	0.5	0.5	0.375
290	13	2001	0.5	-1	-1	-1	0.5
290	16	2001	0	0	0	0	0.
290	18	2001	0	0	0.5	0.5	0.25
290	20	2001	0	-1	-1	0.5	0.25
290	21	2001	-1	0	0	0.5	0.166
290	22	2001	0.5	0	0.5	0.5	0.375
290	26	2001	0	0.5	0.5	0.5	0.375
290	38	2001	0.5	-1	0.5	0.5	0.5
290	45	2001	0.5	-1	0.5	0.5	0.5
292	21	2001	0.5	-1	-1	-1	0.5
315	14	2001	0.5	0.5	-1	0.5	0.5
315	15	2001	0.5	0.5	-1	0.5	0.5
315	16	2001	0.5	0.5	-1	0.5	0.5
315	17	2001	0.5	0.5	0	0.5	0.375
315	18	2001	0.5	0.5	0.5	0	0.375
315	21	2001	0.5	0.5	0.5	0.5	0.5
315	24	2001	0.5	0.5	0.5	0.5	0.5
315	25	2001	0	0.5	-1	0.5	0.333
315	26	2001	0.5	0.5	0.5	0.5	0.5
315	28	2001	0.5	0.5	-1	0.5	0.5
315	31	2001	0	0.5	0	0	0.125
315	33	2001	0	0.5	0.5	0.5	0.375
315	36	2001	0.5	0.5	-1	0.5	0.5
315	41	2001	0.5	0.5	0.5	0.5	0.5
315	42	2001	0.5	0.5	0.5	0.5	0.5
316	2	2001	0.5	0.5	0.5	0.5	0.5
316	3	2001	0.5	0.5	0.5	0.5	0.5
316	4	2001	0.5	0.5	0.5	0.5	0.5
316	6	2001	0.5	0.5	-1	-1	0.5
316	27	2001	0.5	0.5	-1	-1	0.5
316	32	2001	0	0.5	-1	-1	0.25

Site	Ind	Year	7.4	12.19	24.11	24.12	Allele freq.
316	33	2001	0.5	0.5	-1	-1	0.5
316	34	2001	0.5	-1	-1	-1	0.5
317	14	2001	0.5	0.5	0.5	0.5	0.5
317	16	2001	-1	0.5	0.5	0.5	0.5
317	20	2001	0.5	0.5	0.5	0.5	0.5
321	13	2001	0.5	0.5	0.5	0.5	0.5
321	16	2001	0.5	0.5	0.5	0.5	0.5
321	21	2001	0.5	0.5	0	0.5	0.375
321	23	2001	0	0	0.5	0.5	0.25
331	3	2001	0.5	0.5	0.5	0	0.375
332	3	2001	0.5	-1	0.5	0	0.333
332	5	2001	0.5	0.5	0.5	0.5	0.5
332	20	2001	0.5	0.5	0.5	0.5	0.5
332	24	2001	-1	0.5	-1	-1	0.5
332	39	2001	-1	0.5	-1	-1	0.5
332	52	2001	-1	0.5	-1	-1	0.5
336	3	2001	0	0.5	0.5	0.5	0.375
336	14	2001	0.5	0.5	0.5	0.5	0.5
336	16	2001	0.5	0.5	0.5	0.5	0.5
336	17	2001	0.5	0	0.5	0.5	0.375
336	22	2001	0.5	0.5	-1	-1	0.5
336	23	2001	0.5	0.5	0.5	0.5	0.5
336	29	2001	0.5	0.5	-1	-1	0.5
337	2	2001	0.5	-1	0.5	0.5	0.5
337	3	2001	0.5	0.5	0.5	0.5	0.5
337	12	2001	0.5	0.5	0.5	0.5	0.5
337	13	2001	-1	-1	0.5	0.5	0.5
337	24	2001	0.5	-1	0.5	0.5	0.5
337	39	2001	0.5	-1	0.5	0.5	0.5
337	52	2001	0.5	-1	0	0.5	0.333
342	26	2001	0.5	0.5	0	0	0.25
343	2	2001	-1	0.5	-1	0.5	0.5
343	3	2001	0.5	0.5	-1	-1	0.5
343	4	2001	0.5	0.5	-1	-1	0.5
372	11	2001	0.5	0.5	-1	-1	0.5
372	12	2001	0.5	0.5	0.5	0.5	0.5
372	14	2001	-1	-1	0.5	0.5	0.5
372	15	2001	0.5	0.5	-1	-1	0.5
372	16	2001	0	0	0	0.5	0.125
373	1	2001	0.5	0.5	0.5	0.5	0.5
373	3	2001	0.5	0.5	0.5	0.5	0.5
373	4	2001	0.5	0.5	0.5	0.5	0.5
373	6	2001	0.5	0.5	0.5	0.5	0.5
374	10	2001	0	0.5	0.5	-1	0.333
374	11	2001	0.5	0.5	0.5	0.5	0.5
374	12	2001	0.5	-1	0.5	0.5	0.5
374	14	2001	0.5	0.5	-1	0.5	0.5
374	15	2001	0.5	0.5	0.5	0.5	0.5
377	13	2001	0.5	0.5	-1	-1	0.5
377	14	2001	0.5	0.5	-1	-1	0.5
384	1	2001	0.5	0.5	-1	-1	0.5

Appendix III Egg genotypes.

The genotypes are labeled by the number of *variegata* alleles at each locus. -1 indicates a missing genotype.

Site	Egg batch	Number	7.4	12.19	24.11	24.12
256	1	3	1	1	-1	-1
256	1	4	2	1	-1	-1
256	1	5	2	1	-1	-1
256	1	6	2	1	2	1
256	1	7	1	1	2	1
256	1	8	1	2	2	2
256	1	9	1	2	2	2
256	1	10	1	2	2	2
256	1	11	2	1	2	2
256	1	12	1	2	2	1
256	1	13	1	2	2	2
256	2	1	2	2	-1	-1
256	2	2	1	2	-1	-1
256	2	3	1	2	-1	-1
256	2	4	1	2	2	1
256	3	5	2	2	2	2
256	3	8	1	2	2	2
256	3	9	1	2	2	2
256	3	10	1	2	2	2
256	3	11	1	2	2	2
256	2	12	2	2	2	2
256	3	13	1	2	2	2
256	4	1	-1	-1	-1	-1
256	4	2	-1	-1	-1	-1
256	4	3	-1	-1	-1	-1
256	4	4	2	1	2	1
256	4	5	1	2	2	1
256	4	6	2	2	2	1
256	4	7	1	2	2	2
256	4	8	2	1	2	1
256	4	9	2	2	2	2
256	5	10	1	1	2	2
256	4	11	1	2	2	1
256	4	12	1	2	2	2
256	6	1	-1	-1	-1	-1
256	6	2	-1	-1	-1	-1
256	6	3	-1	-1	-1	-1
256	6	5	-1	-1	-1	-1
256	6	6	-1	-1	-1	-1
256	6	7	-1	-1	-1	-1
256	6	8	2	1	2	2
256	6	9	2	1	1	1
256	6	10	2	1	1	1
256	6	11	2	1	2	1
256	6	12	2	1	2	1
256	5	1	2	1	2	2
256	4	2	2	1	2	2
256	4	3	2	1	2	2
256	4	4	2	1	2	2
256	4	7	2	1	2	2
256	4	8	2	1	2	2
256	4	9	2	1	2	2
256	4	10	2	1	2	2
256	5	11	2	1	2	2
256	6	1	-1	-1	-1	-1
256	6	2	2	1	2	1
256	6	3	2	2	2	1
256	6	4	-1	-1	-1	-1
256	6	5	-1	-1	-1	-1
256	6	6	-1	-1	-1	-1
256	6	7	-1	-1	-1	-1
256	6	8	2	1	2	2
256	6	10	2	1	-1	-1
271	1	1	2	2	2	2
271	1	2	2	2	2	2
271	1	3	2	2	2	2
271	1	4	2	2	2	2
271	1	5	2	2	1	2
271	1	6	2	2	1	2
271	1	7	2	2	2	2
271	1	8	2	2	2	2
271	1	9	2	2	2	2
271	1	10	2	2	1	2
271	2	1	2	2	2	1
271	2	2	2	2	2	1
271	2	3	2	2	2	1
271	2	4	2	2	2	1
271	2	5	2	2	2	0
271	2	6	2	2	2	1
271	2	7	2	2	2	1
271	2	8	2	2	2	1
271	2	9	2	2	2	2
271	2	10	2	2	2	2
271	2	11	2	2	2	1
271	3	1	2	2	1	0
271	3	2	2	1	1	1
271	3	3	1	2	1	0
271	3	4	0	1	1	0
271	3	5	2	2	1	0
271	3	6	2	2	1	0
271	3	7	1	1	1	0
271	3	8	1	1	1	0
272	1	1	1	1	2	2

Appendix III – Egg genotypes

Site	Egg batch	Number	7.4	12.19	21.11	24.12	Site	Egg batch	Number	7.4	12.19	21.11	24.12
272	1	2	1	2	2	2	274	3	2	2	1	2	2
272	1	3	2	2	2	2	274	3	3	2	1	2	2
272	1	4	2	2	2	2	274	3	4	2	1	2	2
272	1	5	1	2	2	2	274	3	5	-1	-1	-1	-1
272	1	6	1	1	2	2	274	3	6	-1	-1	-1	-1
272	1	7	1	1	2	2	274	3	7	2	1	2	2
272	1	8	2	2	2	2	274	3	8	2	1	2	2
272	1	9	1	2	2	2	274	3	9	2	1	2	2
272	1	10	1	1	2	2	274	1	1	2	2	2	1
272	1	11	1	2	2	2	274	1	2	1	1	2	1
272	1	12	1	2	2	2	274	1	3	1	2	2	1
272	1	13	2	1	2	2	274	1	4	2	1	2	2
272	2	1	-1	-1	2	2	274	1	5	2	1	2	2
272	2	2	-1	-1	-1	-1	274	1	6	2	2	1	2
272	2	3	2	2	2	2	274	2	7	-1	1	1	2
272	2	4	2	2	2	2	274	1	8	1	2	2	1
272	2	5	2	2	2	2	274	4	1	2	1	1	2
272	2	6	2	2	2	1	274	4	2	2	1	2	2
272	2	7	2	2	2	1	274	4	3	2	1	2	2
272	2	8	2	2	2	2	274	4	4	1	1	2	1
272	2	9	2	2	2	1	274	4	5	1	1	2	2
272	2	10	2	-1	2	1	274	4	6	2	1	-1	1
272	2	11	2	2	2	2	274	5	3	2	2	2	2
272	2	12	2	2	2	1	274	5	5	2	2	2	2
272	2	13	2	2	2	2	274	5	7	2	2	2	2
274	1	1	-1	1	2	1	274	5	8	2	2	2	2
274	1	2	2	2	2	1	274	6	2	2	-1	2	2
274	1	3	-1	-1	-1	-1	274	7	3	2	-1	-1	2
274	1	4	2	2	-1	-1	274	7	5	-1	2	2	2
274	2	5	2	2	1	-1	274	7	6	2	2	2	2
274	1	6	1	2	-1	-1	274	7	10	2	-1	2	2
274	1	7	2	2	-1	-1	276	1	1	1	0	2	2
274	1	8	2	1	-1	1	276	1	2	1	-1	2	0
274	1	9	2	1	1	1	276	1	3	1	0	2	0
274	1	10	1	2	1	2	276	1	4	0	0	2	0
274	3	1	2	1	2	2	276	1	5	1	1	1	1
274	3	2	2	1	2	2	276	1	6	1	1	2	0
274	3	3	2	1	2	2	276	1	7	1	0	1	1
274	3	4	2	1	2	2	276	1	8	0	0	2	1
274	3	5	2	1	2	2	276	2	9	2	0	2	1
274	3	6	2	1	2	2	276	2	10	2	0	2	1
274	3	7	2	1	2	2	276	1	11	2	0	2	1
274	3	8	2	1	2	2	276	1	12	2	1	2	0
274	3	9	2	1	2	2	276	2	13	2	0	2	1
274	3	10	2	1	-1	2	276	3	1	0	0	-1	2
274	3	11	2	1	2	2	276	3	2	0	0	1	2
274	3	12	2	1	-1	2	276	3	3	0	0	-1	2
274	3	13	2	1	2	2	276	3	4	0	0	2	2
274	3	1	2	1	2	2	276	3	5	0	0	2	2

Appendix III – Egg genotypes

Site	Egg batch	Number	7.4	12.19	24.11	24.12
276	3	6	0	0	2	2
276	3	7	0	0	-1	2
276	3	8	0	0	-1	2
276	4	5	2	0	-1	2
276	4	6	2	-1	-1	2
276	5	7	-1	1	-1	2
276	5	8	1	1	-1	2
276	5	9	0	1	-1	2
276	5	10	1	1	-1	1
276	5	11	-1	-1	-1	-1
276	5	12	-1	1	-1	1
282	1	1	1	0	1	0
282	1	2	1	2	1	1
282	1	3	1	0	1	1
282	1	4	1	2	1	0
317	1	1	2	2	2	1
317	1	2	2	2	2	1
317	1	3	2	2	2	1
317	1	4	2	-1	2	2
317	1	5	2	2	2	1
317	1	6	2	1	2	2
317	1	7	2	2	2	1
317	1	8	2	2	2	2
317	1	9	2	2	2	2
317	1	10	2	2	2	1
317	1	11	2	2	2	2
317	1	1	2	1	2	1
317	1	2	2	1	2	2
317	1	3	2	1	2	1
317	1	4	2	2	2	1
317	1	5	2	1	2	2
317	1	6	2	1	2	1
317	1	7	2	2	2	2
317	1	8	2	1	2	2
317	1	9	2	1	2	1
317	1	10	2	1	2	2
317	1	11	2	2	-1	2
317	2	1	2	2	1	2
317	2	2	2	2	1	2
317	2	3	2	1	2	2
317	5	4	2	1	2	2
317	2	5	2	0	2	2
317	3	1	2	1	-1	2
317	3	2	2	2	2	-1
317	3	3	2	1	2	2
317	3	4	2	2	2	2
317	3	5	2	1	2	2
317	3	6	2	1	2	2
317	3	7	2	2	2	2

Site	Egg batch	Number	7.4	12.19	24.11	24.12
317	3	8	2	2	2	2
317	3	9	2	1	2	2
317	3	10	2	1	2	2
317	4	1	2	2	1	0
317	4	2	2	1	1	1
317	4	3	1	2	1	0
317	4	4	0	1	1	0
317	4	5	2	2	1	0
317	4	6	2	2	1	0
317	4	7	1	1	1	0
317	4	8	1	1	1	0
318	3	1	2	2	2	2
318	3	2	2	2	2	2
318	3	3	2	1	1	2
318	1	1	1	1	2	2
318	1	2	1	1	2	2
318	1	3	1	1	2	2
318	1	4	1	1	2	2
318	1	5	1	1	2	2
318	1	6	2	1	2	2
318	1	7	1	1	2	2
318	1	8	1	1	2	2
318	2	1	1	1	1	0
318	2	2	1	2	1	1
318	2	3	1	1	1	1
318	2	4	1	2	1	0
318	2	6	1	2	-1	-1
318	2	7	1	2	-1	-1
318	2	9	1	2	-1	-1
318	2	10	1	2	1	0
330	1	1	0	2	1	1
330	1	2	0	2	1	2
330	2	3	-1	-1	-1	-1
330	2	4	0	2	1	1
330	2	5	0	2	1	1
330	1	6	0	2	1	1
330	1	7	0	1	1	2
330	2	8	0	2	1	2
330	1	9	0	2	1	0
330	3	1	1	2	-1	-1
330	3	2	1	1	-1	0
330	3	4	2	1	0	-1
330	3	6	2	1	0	2
330	3	7	2	1	0	2
330	4	1	1	2	0	1
330	5	2	1	2	1	1
330	4	3	-1	2	1	1
330	4	1	1	2	1	2
330	4	2	1	2	1	1

Appendix III – Egg genotypes

Site	Egg batch	Number	7.4	12.19	24.11	24.12	Site	Egg batch	Number	7.4	12.19	24.11	24.12
330	4	3	1	2	1	2	260	10	1	2	1	-1	1
330	4	6	1	2	2	1	260	10	2	1	2	1	2
330	4	7	1	2	0	2	260	10	3	2	0	2	1
330	4	8	1	2	1	1	260	10	4	2	1	0	1
330	4	9	1	2	1	1	260	1	1	1	0	2	1
330	4	10	0	2	0	2	260	1	2	2	0	0	2
330	3	1	2	2	1	2	260	1	3	2	0	2	1
330	3	2	2	2	0	2	260	1	4	-1	2	2	1
330	3	3	2	1	1	1	260	2	1	1	1	1	0
330	3	4	2	2	0	1	260	2	2	1	2	2	2
330	3	5	2	2	2	0	260	2	3	1	1	2	-1
330	3	6	2	1	2	0	260	23	4	-1	-1	1	0
330	3	7	1	2	1	1	260	2	4	1	2	2	2
330	3	8	2	1	2	0	260	3	1	2	2	1	-1
330	3	9	2	1	1	2	260	3	2	2	2	1	-1
330	3	10	2	1	1	0	260	3	3	-1	-1	1	1
330	3	11	2	2	0	1	260	3	4	-1	-1	0	1
330	3	12	2	1	0	0	260	3	5	2	1	-1	-1
330	3	13	2	2	0	2	260	3	6	1	2	-1	-1
330	6	1	1	2	1	2	260	4	1	1	1	2	1
330	6	2	1	1	1	2	260	4	2	1	1	2	1
330	7	3	2	2	2	0	260	4	3	2	1	1	1
330	6	4	0	1	1	1	260	4	4	2	0	2	1
330	9	1	2	2	1	2	260	6	2	2	1	1	1
330	9	2	2	1	1	0	260	6	3	2	1	1	1
330	9	3	2	2	1	2	260	6	4	2	1	0	2
330	9	4	2	2	-1	2	260	6	5	2	0	2	1
330	7	1	1	2	-1	2	260	9	4	1	1	-1	-1
330	8	2	2	2	-1	2	260	9	5	2	0	0	1
330	8	3	2	2	2	-1	260	9	6	2	1	1	0
330	8	4	0	2	2	1	260	9	8	2	0	2	0
330	8	5	-1	2	2	-1	267	10	1	2	-1	2	1
330	8	6	-1	2	2	-1	267	10	2	2	-1	2	1
330	8	7	0	2	2	-1	267	10	3	2	-1	2	1
330	8	8	1	2	1	-1	267	10	4	2	2	1	2
330	7	9	1	-1	1	2	267	1	1	2	2	2	1
330	8	10	0	2	2	1	267	1	2	2	-1	2	1
330	8	11	1	2	1	1	267	1	3	2	-1	2	2
330	7	12	1	2	2	2	267	2	1	2	1	2	1
330	8	13	0	2	1	2	267	2	2	2	2	2	0
330	11	1	1	2	1	2	267	2	3	1	1	2	1
330	11	2	-1	2	1	2	267	3	1	2	-1	2	1
330	11	3	1	2	1	1	267	3	2	2	-1	2	1
330	11	4	1	2	1	2	267	3	3	2	2	2	1
330	10	1	2	1	-1	2	267	5	1	2	2	2	1
330	10	2	2	2	2	1	267	5	2	2	-1	2	1
330	10	3	2	2	-1	0	267	5	3	-1	-1	2	1
330	10	4	2	2	2	0	267	5	5	2	2	-1	-1
330	10	5	2	2	2	1	267	9	1	-1	-1	0	2

Appendix III – Egg genotypes

Site	Egg batch	Number	7.4	12.19	24.11	24.12
267	9	2	2	-1	2	2
267	9	3	2	0	2	2
267	9	4	2	0	2	2
267	9	5	2	-1	2	0
268	10	4	-1	-1	2	0
268	2	1	2	2	1	2
268	2	2	1	2	1	2
268	2	3	-1	-1	1	1
268	3	1	2	1	2	1
268	3	2	2	1	2	0
268	3	3	2	2	1	0
268	4	1	2	2	1	0
268	4	2	2	2	1	0
268	4	3	2	1	1	0
268	5	1	1	-1	2	1
268	5	2	2	2	1	1
268	5	3	2	1	1	1
268	7	1	1	-1	2	2
268	7	10	2	2	2	0
268	7	2	-1	-1	2	2
268	7	3	-1	-1	2	1
268	7	4	-1	-1	2	1
268	7	8	1	2	2	0
268	7	9	1	2	2	0
268	8	1	0	0	2	-1
268	8	2	0	-1	1	-1
268	8	3	0	-1	2	1
268	8	4	0	0	0	-1
275	1	1	2	-1	2	2
275	1	2	2	-1	2	2
275	1	3	2	2	2	2
275	2	1	2	-1	2	2
275	2	2	2	-1	2	1
275	2	3	2	2	2	-1
275	3	1	2	-1	2	1
275	3	2	2	-1	2	-1
275	3	3	2	2	2	2
275	4	1	2	2	2	2
275	4	2	2	2	2	2
275	4	3	2	2	2	2
275	5	1	2	2	2	2
275	5	2	2	2	2	2
275	5	3	2	1	2	2
275	6	1	2	1	1	1
275	6	2	2	2	2	1
275	6	3	2	-1	0	2
275	8	4	-1	-1	2	0
278	1	1	2	1	1	2
278	1	2	1	0	1	2
278	1	3	1	0	1	1
278	2	1	1	1	1	1
278	2	2	2	1	1	2
278	2	3	1	2	0	1
278	3	1	2	1	0	2
278	3	2	2	1	1	2
278	3	3	1	2	1	0
278	4	1	2	1	2	2
278	4	2	1	0	1	2
278	4	3	1	2	1	2
278	5	1	2	2	1	2
278	5	3	1	1	1	2
278	5	5	-1	-1	2	0
278	5	6	2	2	1	2
278	6	1	2	2	1	2
278	6	2	1	1	0	0
278	6	3	2	2	2	0
285	1	1	1	1	1	2
285	1	10	-1	-1	2	0
285	1	11	-1	-1	2	0
285	11	1	1	2	0	2
285	11	10	2	1	0	1
285	11	11	2	1	0	2
285	11	2	2	1	0	0
285	11	3	1	1	0	1
285	11	4	2	1	0	1
285	11	5	1	1	0	1
285	11	6	1	1	0	0
285	1	2	1	1	2	1
285	1	3	0	1	2	2
285	13	11	2	2	2	1
285	13	12	2	1	1	1
285	13	13	2	1	2	2
285	13	14	2	1	2	2
285	13	15	-1	-1	0	1
285	13	3	2	1	2	0
285	13	4	2	1	2	1
285	13	5	2	1	1	1
285	1	4	1	0	2	0
285	1	5	0	2	2	0
285	1	6	0	2	2	0
285	1	7	1	-1	-1	-1
285	1	8	0	1	0	0
285	1	9	1	2	0	0
285	3	2	2	2	1	2
285	3	3	2	2	1	2
285	3	4	2	1	1	2
285	3	5	2	-1	1	2
285	3	8	2	2	1	2

Appendix III – Egg genotypes

285	3	9	2	-1	1	2	287	5	3	-1	-1	-1	0
285	4	1	2	1	1	2	287	6	1	-1	-1	-1	0
285	4	2	2	1	2	2	287	6	2	-1	-1	1	0
285	4	3	2	2	2	0	287	6	3	-1	-1	1	0
285	4	4	2	2	2	1	287	7	1	-1	-1	1	0
285	4	5	2	1	2	1	287	7	2	-1	-1	1	1
285	4	6	2	1	2	1	287	7	3	-1	-1	0	0
285	4	7	2	2	2	1	321	21	1	0	1	2	1
285	4	8	1	-1	2	2	321	21	2	2	1	0	0
285	5	1	1	-1	1	1	321	22	1	2	1	-1	0
285	5	2	1	1	2	1	321	22	2	0	2	0	1
285	5	3	1	1	1	1	321	22	3	1	2	0	2
285	5	4	1	1	2	1	321	23	1	2	1	0	1
285	5	5	1	2	1	1	321	23	2	2	1	1	1
285	5	6	1	1	2	1	321	23	3	1	1	0	1
285	5	7	-1	-1	1	0	321	24	1	1	1	0	1
285	5	8	1	2	1	2	321	24	2	1	2	-1	-1
285	5	9	1	2	1	1	321	24	3	1	1	-1	-1
285	6	3	-1	-1	2	0	321	25	1	2	0	-1	-1
285	8	1	2	-1	1	2	321	25	2	1	1	-1	-1
285	8	3	-1	-1	1	1	321	25	3	1	1	-1	-1
285	8	4	1	-1	0	2	321	26	1	2	2	-1	-1
285	8	5	1	-1	0	0	321	26	2	1	-1	-1	-1
285	8	5	ii	1	-1	1	321	26	3	1	1	-1	-1
285	8	6	1	0	2	1	321	7	3	-1	-1	2	0
285	8	9	1	1	1	2	372	1	1	2	1	2	0
285	9	1	2	1	0	0	372	1	2	1	2	0	0
285	9	10	2	-1	1	2	372	1	3	2	1	1	0
285	9	11	2	2	1	2	372	2	1	2	2	2	0
285	9	12	2	1	1	1	372	2	2	1	2	0	0
285	9	13	1	-1	1	1	372	2	3	1	2	1	0
285	9	2	2	2	1	2	372	23	3	-1	-1	-1	0
285	9	3	1	1	0	0	372	3	1	1	2	2	1
285	9	8	1	1	0	1	372	3	2	0	2	2	0
285	9	9	1	1	0	1	372	3	3	2	2	2	2
287	1	1	2	1	0	2	372	4	1	1	2	2	0
287	1	2	2	2	1	2	372	4	2	1	2	1	0
287	1	3	2	1	0	2	372	4	3	1	2	2	1
287	13	15	-1	-1	2	0	372	5	1	1	2	1	1
287	2	1	2	1	-1	-1	372	5	2	1	2	1	1
287	2	2	2	1	-1	-1	372	5	3	1	2	2	0
287	2	3	2	0	-1	-1	372	5	5	-1	-1	1	1
287	3	1	0	2	0	0	372	6	1	1	2	2	1
287	3	2	0	0	0	1	372	6	2	1	2	2	1
287	3	3	0	2	0	2	372	6	3	2	2	1	1
287	4	1	2	1	1	0	372	8	1	-1	-1	0	2
287	4	2	2	0	2	1	5	10	1	1	2	2	2
287	4	3	-1	-1	1	0	5	10	2	2	2	1	1
287	5	1	-1	-1	-1	0	5	10	3	1	2	1	1
287	5	2	-1	-1	-1	0	5	1	1	-1	-1	-1	-1

Appendix III – Egg genotypes

5	1	2	2	0	-1	-1	85	5	5	2	-1	1	0
5	1	3	-1	2	-1	-1	85	6	1	1	2	-1	-1
5	4	1	-1	2	1	1	85	6	2	1	2	-1	-1
5	4	2	-1	2	-1	2	85	6	3	1	-1	0	1
5	4	3	-1	2	-1	-1	85	6	4	2	1	1	2
5	4	4	2	2	0	1	85	6	5	0	2	1	0
5	4	5	2	2	-1	-1	85	6	6	-1	-1	1	0
5	4	6	1	2	0	1	4	1	1	2	2	2	2
5	5	1	-1	-1	-1	0	4	1	2	2	2	2	2
5	5	2	-1	-1	-1	1	4	1	3	2	2	2	2
5	5	3	-1	-1	-1	1	4	1	4	2	2	2	2
5	6	1	1	2	1	-1	4	1	5	2	2	2	2
5	6	2	1	2	2	-1	4	1	6	2	1	2	2
5	6	3	1	2	2	-1	4	1	7	2	2	2	2
5	7	1	1	2	2	0	4	1	8	2	2	2	2
5	7	2	2	2	2	2	4	1	9	2	1	2	2
5	7	3	2	2	2	1	4	1	10	2	1	2	2
5	8	1	1	2	1	1	4	2	1	0	2	2	2
5	8	2	1	2	2	0	4	2	2	1	2	2	2
5	8	3	1	2	2	1	4	2	3	0	2	2	2
85	1	1	0	0	0	2	4	2	4	1	2	2	2
85	1	2	0	1	0	-1	4	2	5	1	2	2	2
85	1	5	0	-1	-1	0	4	2	6	1	2	2	2
85	1	6	0	0	-1	0	4	2	7	1	2	2	2
85	19	1	2	0	0	0	4	2	8	1	2	2	2
85	19	2	1	0	1	0	4	2	9	-1	2	2	2
85	19	3	1	1	1	2	4	2	10	1	2	2	2
85	19	4	0	0	1	0	4	2	11	1	2	2	2
85	19	5	1	0	0	0	4	2	12	1	2	2	2
85	19	6	1	1	1	0	4	2	13	0	2	2	2
85	2	10	1	2	2	-1	4	2	14	0	2	2	2
85	21	1	0	1	0	0	4	2	15	1	2	2	2
85	2	12	-1	2	2	-1	4	2	16	1	2	2	2
85	21	2	2	-1	0	0	4	2	17	0	2	2	2
85	21	3	1	1	0	1	4	2	18	1	2	2	2
85	21	4	-1	-1	0	0	4	4	1	1	2	2	2
85	23	1	1	0	0	1	4	4	2	1	-1	2	2
85	23	2	1	1	0	2	4	4	3	0	2	2	2
85	23	3	0	0	0	1	4	4	4	2	2	2	2
85	23	4	0	1	0	1	4	4	6	2	2	2	2
85	2	6	2	2	2	0	4	4	7	1	2	2	2
85	2	9	-1	-1	2	1	4	4	9	2	2	1	2
85	4	3	-1	-1	0	0	4	4	10	2	2	1	1
85	4	4	-1	-1	1	0	4	4	11	0	2	2	1
85	4	5	-1	-1	2	0	4	4	12	2	2	1	1
85	4	6	-1	-1	2	0	4	4	13	0	2	2	2
85	5	1	1	2	-1	-1	4	4	14	1	2	2	1
85	5	2	1	2	0	-1	4	4	15	2	2	2	1
85	5	3	2	-1	0	2	4	4	16	2	2	1	2
85	5	4	2	-1	1	2	4	5	1	2	2	2	2

Appendix III – Egg genotypes

4	5	2	2	2	2	2	315	6	2	2	0	2	1
4	5	3	1	2	1	2	315	6	3	2	0	2	1
4	5	4	2	2	1	2	315	6	4	2	2	2	2
4	5	5	1	2	1	2	315	6	5	2	2	2	1
4	5	6	2	2	2	2	315	6	6	2	1	2	1
4	5	7	2	2	2	2	315	6	7	2	1	2	1
4	5	8	2	2	2	2	315	6	8	2	1	2	2
4	5	9	1	2	2	2	315	7	1	2	1	1	2
4	5	10	2	2	2	2	315	7	2	2	1	2	2
4	5	11	2	2	1	2	315	7	3	2	1	2	2
4	5	12	1	2	2	2	315	7	4	2	1	2	2
4	5	13	2	2	1	2	315	7	5	2	1	2	2
4	5	14	1	2	1	2	315	8	1	1	1	2	2
4	5	15	-1	2	-1	2	315	8	2	2	1	2	1
4	6	1	2	-1	-1	2	315	8	3	2	1	2	1
4	6	2	2	2	-1	2	315	8	4	2	1	2	2
4	6	3	2	2	2	2	315	8	5	1	2	2	2
4	6	4	2	2	2	2	315	8	6	2	1	1	2
4	7	1	2	2	1	2	315	8	7	1	1	2	2
4	7	2	1	2	0	2	315	8	8	1	1	2	1
4	7	3	1	2	1	2	315	8	9	1	2	-1	-1
4	7	5	2	2	0	2	257	1	2	1	1	1	2
4	7	6	2	2	1	2	257	1	3	1	2	1	2
4	7	7	2	2	1	2	257	1	4	2	1	2	2
4	7	8	2	2	0	2	257	1	5	1	2	1	2
4	7	9	2	2	1	2	257	1	6	2	2	2	2
4	7	10	1	2	1	2	257	1	7	2	2	2	2
4	8	1	1	2	0	1	257	1	8	2	1	1	2
4	8	2	2	2	1	1	257	1	9	2	1	2	2
4	8	3	2	2	0	1	257	1	10	1	1	1	2
4	8	4	1	2	1	1	257	1	11	2	2	2	2
4	8	5	1	2	1	2	257	1	12	1	2	2	-1
4	9	1	2	2	2	2	257	1	13	1	1	1	2
4	9	2	1	2	2	2	257	1	14	2	1	1	2
4	9	3	2	2	2	2	257	1	15	2	2	2	2
315	3	1	1	0	2	1	257	2	1	2	2	2	1
315	3	2	2	1	2	2	257	2	2	2	1	2	1
315	3	3	2	1	2	2	257	2	3	2	2	2	0
315	3	4	1	2	2	2	257	2	4	2	0	2	0
315	3	5	1	1	2	2	257	2	5	2	1	2	0
315	3	6	2	2	2	2	257	2	6	2	1	2	2
315	4	1	2	2	2	0	257	2	7	2	1	2	1
315	4	3	2	2	2	1	257	2	8	2	2	2	0
315	4	4	-1	-1	2	-1	257	2	9	2	1	2	0
315	4	6	-1	2	2	-1	257	2	10	2	1	2	1
315	4	7	-1	2	-1	-1	257	2	11	2	0	2	2
315	5	1	2	2	1	2	257	2	12	2	1	2	0
315	5	2	2	2	1	2	257	2	13	2	1	2	0
315	5	3	2	2	0	2	257	3	1	1	-1	2	1
315	6	1	2	0	2	1	257	3	2	1	2	2	2

Appendix III – Egg genotypes

257	3	3	1	2	2	1	257	9	11	2	2	1	2
257	4	1	2	1	1	1	257	10	1	2	1	2	2
257	4	2	2	1	1	2	257	10	2	2	1	1	2
257	4	3	2	1	2	2	257	10	3	2	1	0	2
257	4	5	2	-1	2	1	257	10	4	2	1	1	2
257	4	7	2	1	2	1	257	10	5	2	-1	1	2
257	4	8	2	2	1	1	257	10	6	2	1	1	2
257	4	13	2	2	1	1	257	10	7	2	1	1	2
257	4	14	2	2	1	2	257	10	8	2	1	1	2
257	4	15	2	2	2	2	257	10	9	2	1	2	2
257	4	17	2	2	1	2	257	10	10	2	1	1	2
257	5	1	1	0	1	2	257	10	11	2	1	1	2
257	5	2	1	0	2	2	257	11	1	2	1	1	2
257	5	3	2	0	2	2	257	11	2	2	1	0	2
257	5	4	1	1	2	2	257	11	3	2	1	0	2
257	5	5	2	2	1	2	257	11	4	2	1	1	2
257	5	6	1	2	1	2	257	11	5	2	1	1	2
257	5	7	2	0	1	2	257	11	6	2	1	1	2
257	6	1	2	1	1	2	257	11	7	2	1	1	2
257	6	2	2	2	1	1	257	11	8	2	1	2	2
257	6	3	2	2	1	2	257	11	9	2	1	1	2
257	6	4	2	1	1	1	257	11	10	2	1	1	2
257	6	5	2	2	1	2	257	11	11	2	1	1	2
257	6	6	2	2	1	2	257	12	1	2	2	1	2
257	6	7	2	1	1	1	257	12	2	2	2	-1	2
257	7	1	2	2	1	2	257	12	3	2	2	1	2
257	7	2	2	2	2	2	257	12	4	2	2	2	2
257	7	3	2	2	2	2	257	12	5	1	2	1	2
257	7	4	2	2	1	2	257	12	6	2	2	2	2
257	8	1	1	1	2	2	257	12	7	2	2	1	2
257	8	2	0	2	2	2	257	12	8	1	2	1	2
257	8	3	0	2	2	2	257	13	1	1	2	1	0
257	8	4	0	2	1	2	257	13	2	1	2	1	0
257	8	5	0	1	1	2	257	13	3	1	2	1	1
257	8	6	2	2	1	2	257	13	4	1	2	1	0
257	8	7	1	2	1	2	257	13	5	2	2	1	1
257	8	8	1	2	1	2	257	13	6	-1	2	2	1
257	8	9	1	2	2	2	257	13	7	1	2	0	0
257	8	10	1	1	2	2	257	13	9	1	2	2	1
257	8	11	1	-1	2	2	257	13	10	2	2	0	0
257	9	1	0	2	1	2	257	13	11	1	-1	0	1
257	9	2	2	2	2	2	257	14	1	2	2	1	1
257	9	3	2	2	2	2	257	14	2	2	2	1	1
257	9	4	2	2	1	2	257	14	3	2	2	1	2
257	9	5	2	2	1	2	257	14	4	2	2	1	2
257	9	6	2	2	2	2	257	14	5	2	2	1	2
257	9	7	2	2	2	2	257	14	6	2	2	1	2
257	9	8	1	2	1	2	257	14	7	2	2	1	1
257	9	9	2	2	2	2	257	14	8	2	2	1	1
257	9	10	2	2	2	2	257	14	9	2	2	1	2

Appendix III – Egg genotypes

257	14	10	2	2	1	2	257	19	5	2	2	1	1
257	14	11	2	2	1	2	257	19	6	2	2	1	0
257	14	12	2	2	1	1	257	19	8	2	2	2	2
257	14	13	2	2	1	1	257	19	9	2	2	1	0
257	15	1	2	1	2	2	257	19	10	2	-1	2	-1
257	15	2	2	1	2	2	257	20	1	2	2	1	2
257	15	3	2	1	2	2	257	20	2	2	1	2	2
257	15	4	2	1	2	2	257	20	3	2	1	1	2
257	15	5	2	1	2	2	257	20	4	2	1	2	2
257	15	6	2	2	2	2	257	20	5	2	1	1	2
257	15	7	2	1	2	2	257	20	7	2	1	2	2
257	15	8	2	1	2	2	257	20	8	2	2	1	2
257	15	9	2	2	2	2	257	20	9	2	2	1	2
257	15	10	2	1	2	2	257	20	10	2	1	2	2
257	15	11	2	2	2	2	257	20	11	2	2	2	2
257	15	12	2	2	2	2	257	20	12	2	2	1	2
257	16	1	1	2	1	2	257	21	1	2	2	1	2
257	16	2	1	2	1	1	257	21	2	2	2	2	2
257	16	3	1	1	1	2	257	21	3	2	2	2	1
257	16	4	1	1	1	1	257	21	4	2	2	2	1
257	16	5	1	1	1	1	257	21	5	2	2	1	1
257	16	6	1	1	1	2	257	21	6	2	2	2	1
257	16	7	1	2	1	1	257	21	7	2	2	1	1
257	16	8	1	1	1	1	257	21	8	2	2	1	2
257	16	9	1	2	1	2	257	21	9	2	2	2	2
257	16	10	1	1	1	1	257	21	10	2	2	2	2
257	16	11	1	0	1	1	257	21	11	2	2	2	1
257	16	12	1	1	1	1	257	21	12	2	2	2	2
257	16	13	1	2	1	1	257	21	13	2	2	1	2
257	17	7	2	2	1	1	257	22	1	1	0	1	1
257	17	8	2	2	1	1	257	22	2	1	1	1	0
257	17	9	2	2	1	1	257	22	3	1	1	1	1
257	18	1	2	1	2	1	257	22	4	1	2	1	1
257	18	2	2	1	2	1	257	22	5	1	1	1	2
257	18	3	2	0	0	1	257	22	6	1	2	1	1
257	18	4	1	2	2	0	257	22	7	1	0	1	1
257	18	5	2	2	1	1	257	22	8	1	1	1	0
257	18	6	2	0	1	1	257	23	1	2	2	-1	0
257	18	7	2	1	1	1	257	23	2	2	2	1	0
257	18	8	2	1	2	1	257	23	3	2	2	0	0
257	18	9	1	1	1	1	257	23	4	2	2	2	1
257	18	10	2	1	2	1	257	23	5	2	2	1	1
257	18	11	2	1	0	1	257	23	6	2	2	1	1
257	18	12	1	1	1	2	257	23	7	2	2	0	1
257	18	13	2	2	1	1	257	23	8	2	2	0	1
257	18	14	2	1	1	1	257	23	9	2	2	1	0
257	19	1	2	2	2	2	257	23	10	2	2	1	1
257	19	2	2	2	0	1	257	23	11	2	2	0	0
257	19	3	2	2	1	1	257	23	12	2	2	1	1
257	19	4	2	2	1	1	257	23	13	2	2	1	0

Appendix III – Egg genotypes

257	23	14	2	2	1	1	258	7	1	0	0	2	0
257	23	15	2	2	1	0	258	7	2	-1	0	1	-1
257	23	16	2	-1	1	-1	258	7	3	1	0	1	1
257	24	1	2	2	2	2	258	7	4	1	2	2	1
257	24	2	2	1	2	2	258	7	5	1	2	0	0
257	24	3	2	2	1	2	258	7	8	1	-1	2	2
257	24	4	2	-1	1	-1	258	8	2	1	2	0	1
257	24	5	2	1	2	2	258	8	3	1	2	0	1
257	24	6	2	2	1	2	290	1	1	2	1	0	2
257	24	8	2	2	2	2	290	1	2	1	1	0	2
257	24	9	2	2	2	2	290	1	3	1	2	0	2
257	24	10	2	2	2	0	290	1	4	2	2	1	2
257	25	1	2	1	1	1	290	1	5	2	2	2	2
257	25	2	2	1	0	1	290	1	6	2	2	2	2
257	25	3	2	1	0	1	290	1	7	1	-1	2	-1
257	25	4	2	1	2	1	290	1	8	2	-1	1	2
257	26	1	1	2	2	0	290	1	9	2	1	1	2
257	26	2	2	2	1	2	290	1	10	1	2	0	2
257	26	3	1	2	2	2	290	1	11	1	2	1	2
257	26	4	1	2	2	0	290	1	12	2	2	0	2
257	26	5	1	2	1	2	290	1	13	1	1	0	2
257	26	6	2	2	1	2	290	4	1	1	2	-1	-1
257	26	7	1	2	1	2	290	4	2	1	2	1	2
257	26	9	1	2	1	0	290	4	3	1	2	2	2
257	26	10	2	2	1	1	290	4	4	1	2	1	2
257	26	11	1	2	1	1	290	4	5	1	-1	0	1
257	26	12	1	2	1	1	290	4	6	1	2	1	2
258	3	1	1	1	0	1	290	4	7	2	-1	-1	-1
258	3	2	2	1	0	1	290	5	1	2	2	1	0
258	3	3	1	2	0	1	290	6	1	2	2	1	2
258	3	4	1	1	0	1	290	6	2	2	2	2	1
258	3	5	1	2	0	1	290	6	3	2	2	0	1
258	3	6	-1	-1	-1	2	290	7	1	0	2	2	2
258	3	7	1	2	-1	2	290	7	2	1	2	1	-1
258	3	8	1	1	-1	-1	290	7	3	2	2	0	1
258	3	9	1	0	0	2	290	7	4	1	2	1	2
258	4	1	0	2	1	1	290	7	5	1	-1	2	-1
258	4	2	0	1	1	2	290	7	6	0	-1	-1	-1
258	4	3	0	1	0	1	290	7	7	2	-1	-1	-1
258	4	6	1	-1	1	1	290	7	8	0	-1	2	1
258	4	7	1	2	1	0	290	8	1	2	2	1	1
258	4	8	1	1	0	1	290	8	2	2	2	2	2
258	4	9	1	2	2	1	290	8	3	2	2	2	2
258	4	10	1	-1	2	-1	290	8	4	1	-1	2	-1
258	5	1	2	1	1	2	290	8	5	2	2	2	2
258	5	2	2	1	1	1	290	9	1	1	2	2	1
258	5	3	2	1	1	1	290	9	2	1	2	2	1
258	5	4	2	0	2	1	290	9	3	1	-1	2	1
258	5	6	2	2	2	1	290	9	4	-1	-1	2	2
258	5	7	2	2	2	2	290	9	5	-1	-1	2	2

Appendix III – Egg genotypes

290	9	6	1	2	1	-1	290	13	5	2	2	1	2
290	9	7	2	2	1	2	290	13	6	2	2	1	2
290	9	8	0	2	2	-1	290	13	7	2	2	2	1
290	9	9	2	2	2	2	290	13	8	2	1	2	1
290	9	10	1	-1	-1	-1	290	13	9	2	2	2	2
290	9	11	1	2	1	2	290	13	10	2	2	2	2
290	9	12	1	2	1	-1	290	13	11	2	-1	1	-1
290	9	13	2	1	2	2	290	14	1	1	1	2	1
290	9	14	2	1	-1	1	290	14	2	1	1	2	2
290	9	15	1	2	2	2	290	14	3	2	1	2	1
290	9	16	2	1	2	1	290	14	4	2	-1	2	1
290	9	17	1	1	2	1	290	14	5	2	2	2	1
290	9	18	1	1	2	-1	290	14	6	1	2	2	2
290	9	19	1	2	2	2	290	14	7	1	2	2	2
290	9	20	1	2	2	-1	290	14	8	1	2	2	2
290	10	1	-1	1	1	0	290	14	9	2	2	2	2
290	10	2	-1	2	2	0	290	15	1	1	2	1	1
290	10	3	1	2	2	1	290	15	2	2	2	1	1
290	10	4	1	1	2	0	290	15	3	1	2	1	1
290	10	5	1	1	2	-1	290	15	4	2	-1	1	0
290	10	6	1	2	2	2	290	15	5	2	0	0	1
290	10	7	2	2	2	0	290	15	6	2	0	0	1
290	10	8	1	2	1	2	290	15	7	1	0	1	1
290	11	1	2	2	2	1	290	15	8	1	0	1	1
290	11	2	2	2	2	1	290	15	9	2	0	0	0
290	11	3	1	2	2	1	290	15	10	1	0	1	1
290	11	4	2	2	1	1	290	15	11	1	0	2	1
290	11	5	2	2	2	-1	290	15	12	2	0	2	0
290	11	6	2	1	2	2	290	15	13	2	0	2	0
290	11	7	1	1	2	2	290	15	14	1	0	1	2
290	11	8	1	1	2	2	290	16	1	2	2	1	2
290	11	9	1	1	2	2	290	16	2	1	2	1	2
290	11	10	2	2	2	2	290	16	3	-1	2	1	2
290	11	11	2	-1	2	1	290	16	4	2	0	1	2
290	11	12	2	2	2	1	290	16	5	1	0	1	2
290	11	13	1	2	1	1	290	16	6	2	1	1	-1
290	11	14	1	2	2	1	290	16	7	1	2	1	2
290	12	1	0	2	0	1	290	16	8	1	2	1	2
290	12	2	0	2	1	1	290	16	9	1	2	1	2
290	12	3	0	2	1	1	290	16	10	1	2	1	2
290	12	4	0	2	1	1	290	16	11	2	2	1	2
290	12	5	0	2	1	0	290	17	1	2	2	1	1
290	12	6	0	2	0	1	290	17	2	2	2	0	1
290	12	7	0	2	1	0	290	17	3	2	2	2	1
290	12	8	0	2	0	2	290	17	4	2	2	1	1
290	12	9	0	2	0	1	290	17	5	2	2	0	2
290	13	1	2	1	1	2	290	17	6	2	2	1	2
290	13	2	2	1	1	2	290	17	7	2	2	1	1
290	13	3	2	2	0	1	290	17	8	2	2	1	1
290	13	4	2	2	1	0	290	17	9	2	2	0	1

Appendix III – Egg genotypes

290	17	10	2	2	1	1
290	17	11	1	2	0	2
290	17	12	2	2	0	2
290	17	13	2	2	0	2
290	18	1	1	2	2	1
290	18	2	1	2	0	0
290	18	3	1	1	0	2
290	18	5	0	2	1	2
290	18	6	0	1	1	1
290	18	7	0	1	0	2
290	18	8	0	1	1	1
290	18	9	1	1	1	2
290	18	10	1	2	1	1
290	19	1	2	-1	0	1
290	19	2	2	-1	0	0
290	19	3	2	-1	1	0
290	19	4	1	2	1	0
290	19	5	1	2	1	0
290	19	6	2	2	1	0
290	19	7	1	2	0	0
290	19	8	1	2	0	1
290	19	9	0	2	0	1
290	19	10	0	2	1	1
290	20	1	2	2	1	0
290	20	2	2	2	0	0
290	20	3	2	2	1	1
290	20	4	2	2	2	1
290	20	5	2	2	0	0
290	20	6	2	2	2	0
290	20	7	2	1	0	1
290	20	8	2	2	2	2
290	20	9	2	1	1	0
290	20	10	2	2	1	0
290	21	1	2	1	2	1
290	21	2	2	1	1	1
290	21	3	2	2	2	2
290	21	4	2	2	-1	-1
290	21	5	2	2	1	2
290	21	6	2	2	1	1
290	21	7	2	2	1	2
290	21	8	2	2	2	2

Appendix IV – Mark-Recapture Data

Appenidix IV. Mark-release-recapture experiment. The following data show the animals that were caught on two or more occasions. Columns show first site, individual, no. recaps, year of first capture, and sites and days of subsequent captures

Site	Ind	N	Yr	Site	day	site	day
258	32	1	2000	258	–	258	24
4	23	1	2000	4	–	271	1
3	26	1	2000	3	–	4	35
257	11	1	2000	257	–	257	–
289	2	1	2000	289	–	290	25
285	2	1	2000	285	–	285	–
7	5	1	2000	7	–	6	–
6	3	1	2000	6	–	315	1
5	1	1	2000	5	–	6	27
8	4	1	2000	8	–	9	27
4	2	1	2000	4	–	4	35
259	8	1	2000	259	–	258	20
258	5	1	2000	258	–	258	15
247	4	1	2000	247	–	258	15
5	5	1	2000	5	–	4	35
5	5	2	2000	4	–	5	43
6	1	1	2000	6	–	11	34
6	1	2	2000	11	–	10	18
4	4	1	2000	4	–	4	18
4	4	2	2000	4	–	4	35
4	2	1	2000	4	–	315	8
4	2	2	2000	315	–	315	14
372	18	1	2001	372	36	372	43
290	34	1	2001	290	33	290	37
290	32	1	2001	290	33	290	37
373	20	1	2001	373	32	373	35
372	11	1	2001	372	32	372	35
10	7	1	2001	10	29	10	35
9	34	1	2001	9	28	9	35
9	33	1	2001	9	28	9	35
9	32	1	2001	9	28	9	35
9	27	1	2001	9	28	9	35
9	25	1	2001	9	28	9	35
9	22	1	2001	9	28	9	35
8	7	1	2001	8	28	8	35
6	5	1	2001	6	28	6	35
6	4	1	2001	6	28	3	35
377	14	1	2001	377	26	377	32
377	13	1	2001	377	26	377	32
377	2	1	2001	377	26	377	32
377	1	1	2001	377	26	377	32
330	52	1	2001	330	25	330	28
258	50	1	2001	258	25	258	24
321	23	1	2001	321	25	321	28
375	7	1	2001	375	25	375	31
374	15	1	2001	374	24	374	31
374	14	1	2001	374	24	374	31
Site	Ind	N	Yr	sit	day	site	day
374	12	1	2001	374	24	374	31
374	11	1	2001	374	24	374	31
374	10	1	2001	374	24	374	31
374	8	1	2001	374	24	374	31
372	7	1	2001	372	24	372	31
374	7	1	2001	374	24	374	31
373	6	1	2001	373	24	373	31
374	6	1	2001	374	24	374	31
374	5	1	2001	374	24	374	31
373	4	1	2001	373	24	373	31
374	4	1	2001	374	24	374	31
372	3	1	2001	372	24	372	31
372	3	1	2001	372	24	372	31
373	3	1	2001	373	24	373	31
374	3	1	2001	374	24	374	31
372	2	1	2001	372	24	372	31
372	2	1	2001	372	24	372	31
374	2	1	2001	374	24	374	31
258	45	1	2001	258	21	258	24
258	44	1	2001	258	21	258	24
330	39	1	2001	330	21	330	28
258	38	1	2001	258	21	258	24
342	30	1	2001	342	21	342	24
342	29	1	2001	342	21	342	24
342	27	1	2001	342	21	342	24
290	26	1	2001	290	21	290	28
342	26	1	2001	342	21	342	24
342	26	1	2001	342	21	342	24
342	25	1	2001	342	21	342	24
342	24	1	2001	342	21	342	24
342	23	1	2001	342	21	342	24
321	21	1	2001	321	21	321	24
342	21	1	2001	342	21	342	24
332	20	1	2001	343	21	343	24
342	19	1	2001	342	21	342	24
342	18	1	2001	342	21	342	24
342	17	1	2001	342	21	342	24
342	16	1	2001	342	21	342	24
342	15	1	2001	342	21	342	24
343	14	1	2001	343	21	343	24
342	13	1	2001	342	21	342	24
342	10	1	2001	342	21	342	24
342	9	1	2001	342	21	342	24
342	7	1	2001	342	21	342	24
342	6	1	2001	342	21	342	24
342	5	1	2001	342	21	342	24
343	4	1	2001	343	21	343	24
343	3	1	2001	343	21	343	24
342	2	1	2001	342	21	342	24
343	2	1	2001	343	21	343	24
342	1	1	2001	342	21	342	24
4	20	1	2001	4	19	5	35
317	20	1	2001	317	19	5	35
11	16	1	2001	11	19	11	27
317	16	1	2001	317	19	317	27
3	15	1	2001	3	19	10	27
11	14	1	2001	11	19	11	27

Appendix IV – Mark-Recapture Data

Site	Ind	N	Yr	sit	day	site	day
3	14	1	2001	3	19	3	35
4	14	1	2001	4	19	4	35
317	14	1	2001	317	19	317	27
11	12	1	2001	11	19	13	27
3	11	1	2001	3	19	3	35
318	7	1	2001	318	19	9	20
10	2	1	2001	10	19	10	27
336	29	1	2001	336	17	336	37
336	23	1	2001	336	17	336	32
336	22	1	2001	336	17	336	37
336	17	1	2001	336	17	336	32
336	16	1	2001	336	17	336	37
336	14	1	2001	336	17	336	32
336	3	1	2001	336	17	336	32
337	3	1	2001	337	17	336	32
337	2	1	2001	337	17	337	24
258	27	1	2001	258	16	258	24
258	22	1	2001	258	16	258	20
290	18	1	2001	290	16	290	20
315	42	1	2001	315	15	12	18
315	41	1	2001	315	15	10	27
315	33	1	2001	315	15	10	27
315	30	1	2001	315	15	12	18
335	10	1	2001	335	15	12	18
4	8	1	2001	4	15	3	35
13	3	1	2001	13	15	14	27
330	24	1	2001	330	13	330	20
330	6	1	2001	330	13	332	20
332	5	1	2001	332	13	332	20
330	4	1	2001	330	13	330	20
332	3	1	2001	332	13	332	20
258	17	1	2001	258	12	258	20
258	17	1	2001	258	12	258	15
258	16	1	2001	258	12	258	15
290	10	1	2001	290	12	290	15
315	28	1	2001	315	9	8	35
315	26	1	2001	315	9	318	25
315	24	1	2001	315	9	315	14
315	21	1	2001	315	9	10	35
315	19	1	2001	315	9	315	14
315	18	1	2001	315	9	10	27
315	17	1	2001	315	9	10	35
4	7	1	2001	4	9	4	35
290	3	1	2001	290	9	290	25
290	2	1	2001	290	9	290	25
258	11	1	2001	258	8	258	20
258	6	1	2001	258	8	258	15
258	6	1	2001	258	8	258	11
258	2	1	2001	258	8	258	20
321	13	1	2001	321	3	321	12
321	2	1	2001	321	3	321	12
315	15	1	2001	315	2	317	27
315	14	1	2001	315	2	315	8
317	11	1	2001	317	2	317	18
315	10	1	2001	315	2	7	34
317	9	1	2001	317	2	5	35
11	8	1	2001	11	2	392	35
317	7	1	2001	317	2	317	18
315	6	1	2001	315	2	13	27
316	6	1	2001	316	2	271	35
9	5	1	2001	9	2	9	27
11	4	1	2001	11	2	9	18
7	4	1	2001	7	2	317	18
315	4	1	2001	315	2	4	14
11	3	1	2001	11	2	10	27
8	3	1	2001	8	2	8	35
315	3	1	2001	315	2	4	35
316	3	1	2001	316	2	271	35
13	2	1	2001	13	2	13	27
3	2	1	2001	3	2	3	35
271	2	1	2001	271	2	3	35
315	2	1	2001	315	2	7	34
316	2	1	2001	316	2	4	18
318	2	1	2001	318	2	5	14
13	1	1	2001	13	2	13	43
14	1	1	2001	14	2	7	34
7	1	1	2001	7	2	315	8
7	1	1	2001	7	2	8	35
9	1	1	2001	9	2	10	27
315	1	1	2001	315	2	8	35
317	1	1	2001	317	2	4	18
372	16	1	2001	372	32	372	35
372	16	2	2001	372	35	372	43
372	15	1	2001	372	32	372	35
372	15	2	2001	372	35	372	43
372	12	1	2001	372	32	372	35
372	12	2	2001	372	35	372	43
372	10	1	2001	372	32	372	35
372	10	2	2001	372	35	372	43
372	7	1	2001	372	24	372	31
372	7	2	2001	372	31	372	35
372	6	1	2001	372	24	372	31
372	6	2	2001	372	31	372	35
372	6	1	2001	372	24	372	31
372	6	2	2001	372	31	372	43
373	1	1	2001	373	24	373	31
373	1	2	2001	373	31	373	35
290	26	1	2001	290	21	290	24
290	26	2	2001	290	24	290	32
13	5	1	2001	13	19	13	27
13	5	2	2001	13	27	8	35
8	5	1	2001	8	19	315	8
8	5	2	2001	315	8	13	50
321	16	1	2001	321	17	321	20
321	16	2	2001	321	20	321	28
290	21	1	2001	290	16	290	20
290	21	2	2001	290	20	290	24
290	20	1	2001	290	16	290	20
290	20	2	2001	290	20	290	24
290	15	1	2001	290	16	290	20
290	15	2	2001	290	20	290	24
11	10	1	2001	11	15	11	18

Appendix IV – Mark-Recapture Data

Site	Ind	N	Yr	sit	day	site	day
11	10	2	2001	11	18	11	27
258	15	1	2001	258	12	258	15
258	15	2	2001	258	15	258	20
290	13	1	2001	290	12	290	15
290	13	2	2001	290	15	290	20
290	9	1	2001	290	12	290	15
290	9	2	2001	290	15	290	24
315	25	1	2001	315	9	4	14
315	25	2	2001	4	14	392	35
290	1	1	2001	290	9	290	20
290	1	2	2001	290	20	290	25
258	10	1	2001	258	8	258	20
258	10	2	2001	258	20	258	24
258	7	1	2001	258	8	258	11
258	7	2	2001	258	11	258	15
258	4	1	2001	258	8	258	11
258	4	2	2001	258	11	258	24
4	4	1	2001	4	2	4	18
4	4	2	2001	4	18	317	27
316	4	1	2001	316	2	3	18
316	4	2	2001	3	18	3	35
7	2	1	2001	7	2	5	18
7	2	2	2001	5	18	6	35
290	22	1	2001	290	16	290	24
290	22	2	2001	290	24	290	32
290	22	3	2001	290	32	290	37

Appendix V. - Variations in laboratory methods

The genetic analysis of samples described in this thesis was performed largely by Sonja Köehler, Tim Vines and M. Thiel at LMU Munich. Genetic analyses by myself were carried out in both University of Edinburgh and LMU Munich. Generally the protocols and reagents used in both locations are identical, except where equipment differed. Although there has been no systematic study, it is not suspected that conducting genetic lab work in two locations biases the resulting genotypes.

The methods used at LMU Munich are described by Nürnberger *et al.* (2003), Köehler (2003) and Vines (2002). As there are only minor differences between these protocols and those previous described in this thesis, I briefly described the differences below.

SSCP genotyping

PCR reactions for SSCP genotyping were carried out in two location, but using identical reagents and protocols. The electrophoresis of PCR products was carried out entirely at LMU Munich and all protocols are therefore identical.

Microsatellite genotyping

The PCR reagents and protocols for microsatellite analysis were identical with one exception. In LMU Munich the primer used was an infra-red fluorescent supplied by Amersham Pharmacia Biotech.

Amplification products of microsatellites were mixed 1:1 with an electrophoresis buffer consisting of 95% formamide, 10 mM EDTA, 0.025% bromphenole blue and 0.05 µM xylene cyanol, denatured at 95°C, shock cooled on wet ice. Electrophoresis was carried out on an ALF express TM sequencer (Amersham Pharmacia Biotech) with denaturing gels 0.5cm thick with 6% acylamide gels (Long Ranger, BMA), 6M urea (Roth) and a gel buffer of 1 x T at pH 7.5 and room temperature. Electrophoresis was performed at 55°C and 55.5 V for 8 hours with an electrode buffer of 0.5 x TBE. The length polymorphisms were analysed with the program Fragment Manager 1.2 (Amersham Pharmacia Biotech).