Barriers to gene flow:
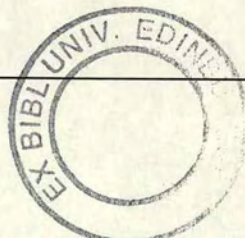# a *Bombina* (fire-bellied toad) hybrid zone and multilocus cline theory

Loeske E. B. Kruuk

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

1997

585

ABSTRACT

1. In a stable hybrid zone between two taxa, natural selection creates a barrier to gene flow which counterbalances the homogenising effects of dispersal and interbreeding. Studies of hybrid zones can contribute to our understanding of the nature of reproductive isolation between two populations, and hence of speciation mechanisms. For example, reproductive isolation may be generated by natural selection acting against alleles in the wrong environment or against alleles in the wrong genetic background. In this thesis, I consider these issues with (i) a study of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* and (ii) development of analytical and simulation models of geographical variation maintained by clines at multiple loci.

2. In a transect across the *Bombina* hybrid zone in Croatia, the pattern of change in three phenotypic traits (leg length, belly pattern and egg size) corresponds to the stepped clines observed in previous studies of diagnostic allozyme loci. There is close concordance between the mean values of the traits and between estimates of linkage disequilibrium calculated from associations between alternative phenotypic and genetic variables. Clines in allozyme frequency and spot score are wider in males than in females, suggesting differential dispersal patterns.

3. There is direct evidence for hybrid dysfunction in the field: in samples collected from *Bombina* breeding sites, embryonic mortality, larval mortality and frequency of larval developmental abnormalities all increase significantly towards the centre of the hybrid zone. However, a cohort analysis of adult toads within a central region shows no evidence of differential mortality with respect to genotype.

4. *Bombina bombina* and *B. variegata* typically use different breeding habitats, with *B. variegata* showing strong avoidance of the semi-permanent ponds favoured by *B. bombina.* The abundance of aquatic predators is shown to differ between the alternative breeding habitats preferred by either species. Behavioural experiments show that the feeding strategy of *B. bombina* larvae reduces their risk of predation, relative to that of *B. variegata* larvae, thus demonstrating an adaptive advantage to the adults' habitat preference. As the availability of either habitat type changes across the zone, these adaptations imply that differential adaptation across an environmental gradient is also creating a barrier to gene flow between the taxa.

5. An analytical model of multilocus clines maintained by differential adaptation to alternative environments is developed, considering in particular the effect of selection on neutral markers. The dynamics are qualitatively similar to those of previous models of heterozygote disadvantage. Computer simulations are used to test the analytical predictions, and the restrictions imposed by assumptions of weak selection. These show that while cline shape can accurately estimate parameters such as fitness in hybrid populations, others, such as the number of genes under selection, may be less robust.

6. The effect of a habitat preference in a hybrid zone such as *Bombina* is explored with a simulation model. Pooling across habitats, overall statistical associations increase with preference strength, illustrating the effect of the habitat preference in maintaining the integrity of the parental genomes. However, the resulting magnitude of linkage disequilibrium and deviations from Hardy-Weinberg proportion *within* habitats vary non-monotonically with the strength of the preference. The resulting shape of the zone will be largely dependent on the underlying distribution of habitat availability.

7. The *Bombina* hybrid zone is therefore maintained by both endogenous and exogenous selection. Although theoretical models show that the dynamics of either regime are similar, the results imply that divergence during allopatry has been driven by adaptation to alternative environments, rather than solely non-adaptive factors. The resulting reproductive isolation has therefore been, at least partially, determined by ecological factors, and environmentally-mediated factors such as a habitat preference will generate further barriers to gene flow between two populations.

DECLARATION

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

Loeske E.B. Kruuk

Acknowledgements

Table of contents

Chapter 1

Speciation, hybrid zones and *Bombina*

"The view commonly entertained by naturalists is that species, when intercrossed, have been specially endowed with sterility, in order to prevent their confusion."

(Darwin 1872, p 304)

Studies of hybrid zones shed light on the process of speciation. This thesis describes inferences from empirical work on a natural hybrid zone and theoretical work on models of geographic variation. In this introductory chapter I discuss the relevance of the study of speciation in evolutionary biology, and outline its major theories and contentions (section 1.1). In section 1.2, I consider the potential for research on hybrid zones to further our understanding of the process of speciation. Sections 1.3 and 1.4 describe, respectively, important characteristics and the theoretical representation of hybrid zones; other factors which may affect hybrid zone dynamics are considered in section 1.5. Finally, section 1.6 contains an introduction to the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata* in eastern Europe.

1.1 Speciation

1.1 (i) Evolutionary biology and the study of speciation

Evolutionary biology is the study of organic diversity. If the first forms of life were few and simple, how has current variety been generated? Species diversity is a principal component of this variety, and yet the process by which new species are formed is not understood. Questions concerning the origin of species have been debated since the publication in 1859 of Darwin's landmark text which, whilst affecting human thought as radically as any work might ever hope to, did not fully address the subject of its title. The energy expended on the debate implies the absence of a simple answer (such as that given by the theory of natural selection, which elegantly explains a multitude of other biological processes). Consensus shifts between alternative arguments, and issues which were presumed to have been resolved are frequently re-opened and re-debated. Direct tests of hypotheses and predictions are rendered impossible by the time-scales involved: since speciation events are rarely observed, historical processes must either be inferred indirectly from extant patterns or investigated through theoretical models, and in the inter-

pretation of patterns and the formulation of models, dissent is generated. The issue is further confused by a plethora of opinions on the exact definition a species. Species are distinct from each other, but distinctness, separation, isolation and divergence are continuous measures, whereas the number of species is discrete. The difficulty inherent in mapping a continuous space to a discrete space creates a conceptual impasse, resulting in a diversity of species definitions reflective of the diverse priorities of their authors.

Any consideration of biological diversity therefore needs to consider speciation, but any consideration of speciation should be aware of these difficulties. The lack of a comprehensive simple explanation is no reason to abandon research into an area, but rather calls for a careful approach. In this thesis, I consider what light a study of the genetics of populations can shed on the process of speciation. First, a species definition is necessary; I will not discuss the possible alternatives in detail here, but simply adopt that which seems most suitable for investigating the creation of 'species diversity'.

1.1 (ii) Definition of species

Within evolutionary biology, Mayr's Biological Species Concept has proved the hardiest of proposed definitions. Mayr defines species as "groups of actually or potentially interbreeding populations, which are reproductively isolated from other such groups" (Mayr 1942). However, the BSC suffers serious short-comings for a comprehensive species definition, of which a lack of relevance to all asexual taxa is the most severe. In addition, when there is complete geographic isolation between two populations, the extent of reproductive isolation can only be inferred and the resulting classification will be imperfect (Cracraft 1989). Thirdly, limited gene exchange with other "species" is an accepted fact for the majority of plant taxa (Barrett 1989; Grant 1981) and a significant proportion of animal taxa (Grant & Grant 1992; Templeton 1989).

What are the alternatives? Some definitions emphasise ecological aspects, defining a species by the ecological "niche" (after Hutchinson 1965) to which it is adapted; others invoke the (nebulous) concept of a population being on a unique evolutionary pathway; others again combine various aspects of these possibilities; see Otte and Endler (1989) for a full representation of the variety. However, whatever the "sexual hang-ups" (Templeton 1989) of the biological species concept (too little sex amongst asexual or selfing taxa, too much amongst interbreeding taxa), classifying organisms by their gene pools creates a useful conceptual link between systematics and population genetics (Ridley 1993). Whilst its drawbacks should be kept in mind, Mayr's definition sets unambiguous criteria with which to formulate testable hypotheses. It therefore provides population geneticists with practicable means for investigating speciation in sexually reproducing organisms.

1.1 (iii) Modes of divergence

Understanding biological diversity and speciation requires knowledge of the mechanisms which can generate divergence in morphology, ecology, behaviour and genetics between two populations. Divergence is the first component of speciation, with reproductive isolation accumulating as a secondary by-product. Various factors may result in divergence between two gene pools:

(1) Varying selective conditions: if the environmental conditions encountered by either population differ, natural (or sexual) selection will favour different characters in the respective populations. This constitutes adaptive divergence, driven by the pressure of natural selection.

(2) The hazards of random genetic drift will result in differences between gene pools: drift is a possible mechanism by which one population might cross from one peak on the adaptive landscape to another (Wright 1932; 1988).

(3) Alternative genetic mutations will inevitably accumulate in isolated populations over time, also generating differences between the gene pools.

Note that these last two mechanisms are stochastic and non-adaptive. The geographic relationship between the populations in questions will be crucial, so divergence can also be classified by the geographical circumstances:

(a) Allopatric divergence

Allopatric divergence occurs if two populations are geographically isolated before any differentiation. The concept of a population's range being split by a physical barrier, leaving two populations to follow their respective evolutionary trajectories, is widely accepted. Mayr (1942; 1963) advocated allopatric speciation as the exclusive mode of speciation. His forceful opposition to alternative theories reflected and enhanced a general consensus of belief, prevalent over decades, that a physical barrier to gene flow is a pre-requisite for divergence. Alternative modes (described below) are now more widely accepted, but allopatric divergence is still generally assumed to be the most frequent cause of speciation (Lynch 1989; Ridley 1993).

Any or all of the three mechanisms listed above (selection in alternative environments, random genetic drift, mutation accumulation) may generate divergence in allopatric popu-lations. The effects of the last two stochastic forces will depend on the number of indi-viduals in either population. Distinctions therefore need to be made between the scenario in which the ancestral population is divided into two comparably large populations and that in which a small population at the edge of the current range is separated off. (Note however that both empirical evidence and theoretical justification for founder event speciation from a small population are weak (Barton & Charlesworth 1984; Barton & Turelli 1989).) Note also that any of these mechanisms might result in exploitation of different ecological niches within the populations' respective ranges, a point which is relevant to the work presented in this thesis.

(b) Parapatric divergence

Parapatric divergence occurs as populations evolve while adapting to a gradient in the biotic or abiotic environment; the populations' ranges are contiguous, with no physical impediment to gene flow across the zone of contact. Selection is therefore the primary force fuelling parapatric divergence, acting, for example, as a population's range expands into a new region or environment. Random drift and the accumulation of mutations may also be invoked (Rouhani & Barton 1987; Wright 1941). Theoretical models have illustrated the possibility for differentiation along an environmental gradient to result in a stable cline (Endler 1977; Slatkin 1973). However, as the incidence of a cline in a natural population could also be the result of secondary contact between two populations that have previously diverged in allopatry (see below), the frequency of parapatric speciation events is debatable.

(c) Sympatric divergence

The final option is divergence in sympatry, which does not require any spatial separation of the two populations. As with parapatric divergence, theoretical models have demonstrated that a stable polymorphism can be maintained despite contact between individuals. These models were initiated with the work of Maynard Smith (1962, 1966), and have since become increasingly sophisticated and decreasingly restrictive in the criteria required for sympatric divergence. Common to all is the requirement for disruptive, frequency- or density-dependent natural selection, resulting in a shift in host or habitat association due to partitioning of an essential yet limiting resource (e.g. Bush & Diehl 1982; Diehl & Bush 1989; Rausher 1984; Rice 1984, 1987; Tauber & Tauber 1977; see section 6.1 for a more detailed consideration of their characteristics). Sympatric speciation therefore invokes only selection-driven divergence, through adaptation to and preference for alternative ecological niches. Its plausibility is being increasingly appreciated, particularly with respect to phytophagous or parasitic invertebrates mating within a preferred habitat:

proponents (e.g. Bush 1975, 1994) have argued its importance as forcefully as ever Mayr presented his opinions on allopatric speciation.

1.1 (iv) *Accumulation of reproductive isolation*

Divergence will therefore split an ancestral population into populations "located at different equilibria under selection" (Barton & Charlesworth 1984), and in doing so will generate reproductive isolation. The nature of the isolation, or of the barrier to gene flow, is more amenable to study than the historical process of divergence. More specifically, we can consider whether reproductive isolation, or the existence of barriers to gene flow, is due to (i) *endogenous* factors (after Moore & Price 1993), whereby genetic incompatibilities render hybrid offspring inviable and/or infertile or (ii) *exogenous* factors, with alleles or sets of alleles characteristic of each population conferring increased fitness in alternative environments or ecological niches? Note that the former would be a plausible outcome of any form of divergence, whereas the latter implies that divergence has been driven by selection pressures. The accumulation of sufficient reproductive isolation entails speciation. The relative import of endogenous vs. exogenous factors therefore determines whether species numbers are ultimately limited by rates of genetic alteration or by the diversity of available ecological niches, and is the principal motivation of the work presented here. I concentrate on the case where divergence is known to have occurred in allopatry, and ask what form of barrier to gene flow such divergence can generate. In the following section, I outline the reasons why such questions can be addressed through studies of hybrid zones.

1.2 Why study hybrid zones?

1.2 (i) Barriers to gene flow

The nature of the reproductive isolation which has accumulated during allopatry can be discerned if the geographic isolation between the populations is removed: the relative efficacy of possible barriers to gene flow will then be put to the test. Fortunately, such "experiments" (after Hewitt 1988) are provided naturally, when previously-allopatric populations expand their ranges and come into contact once more. If the reproductive isolation is negligible, the populations will merge back into one. At the other extreme, complete reproductive isolation would result in the parapatric (i.e. adjacent) or sympatric (i.e. overlapping) distribution of the two populations. Which of these occurs depends on the nature and strength of the barriers to gene flow which have evolved in allopatry. Most fruitful for study are intermediate cases when reproductive isolation is incomplete but the merging effects of interbreeding are counterbalanced by other factors pushing populations towards either parapatry or sympatry. The outcome will then constitute a hybrid zone, indicative of an incomplete barrier to gene flow whose semi-permeable nature expose its structure to investigation (Barton & Hewitt 1981b; 1985).

1.2 (ii) Definitions of hybrid zones

Hybrid zones have been reported in a range of taxa, including mammals, birds, reptiles, fish, insects, flowering plants and trees (Hewitt 1993b). Opinions on the exact definition of a hybrid zone are (nearly) as numerous as those on species, but the following definition of Harrison's (1990, p72) seems entirely comprehensive:

> "Hybrid zones are interactions between genetically distinct groups of individuals resulting in at least some offspring of mixed ancestry. Pure populations of the two genetically distinct groups are found outside of the zone of interaction."

Harrison's definition forces no requirement on either the origin or maintenance of the hybrid zone, nor of the taxonomic relationship between the two groups. In order to address the questions outlined above, I consider here one type of hybrid zone: that formed on secondary contact between two populations whose divergence in allopatry has been such as to merit species status. However, hybrid zones need not be the result of secondary contact. Similar patterns of transition between one form and another could arise *in situ* in direct response to spatially varying selection pressures or environmental gradients, by the equivalent processes invoked in theories of parapatric speciation. The clines, or gradients in traits, resulting in hybrid zones formed through this primary intergradation will be indistinguishable from those formed on secondary contact. Direct inference of past processes from present patterns is therefore not feasible (Endler 1977), so indirect arguments must be relied upon to ascertain their origins. The coincidence of clines in many genetic and phenotypic characters at the same place frequently suggests secondary contact (Barton & Hewitt 1985; Hewitt 1989); in many cases, range expansion and contractions can be explained by climatic factors such as ice ages (Hewitt 1993b; 1996). (It is not clear whether the relative paucity of tropical hybrid zones is relevant here, or merely due to a sampling problem.) The arguments are linked to the allopatric/non-allopatric speciation debate, with the leaning towards predominantly secondary contact explanations implicitly supporting the notion that geographic isolation has been a prerequisite for differentiation.

A second important characteristic of hybrid zones is the role played by dispersal in their maintenance. In dispersal-independent clines, trait values track gradients in selection coefficients: clines then represent balanced polymorphisms with spatially-varying equilibria. These include cases of "bounded hybrid superiority", in which hybrid individuals are most fit within a restricted area (Moore 1977). However, dispersal will only be negligible when clines are very broad relative to the average dispersal range of an

individual. The frequently-observed phenomenon of hybrid zones forming a long, narrow strip between two populations' ranges argues for a balance being maintained between dispersal, mixing genes from different populations, and natural selection in some way preventing the "interactions" (from Harrison's definition above) spreading outwith the narrow strip. I concentrate here on the case when secondary contact between two populations results in a stable balance of this form, but it should be noted that dispersal-dependent clines will also form during the decay of an initially steep gradient in a neutral character, or with the wave of advance of an advantageous allele or set of alleles through a population (Fisher 1937).

1.3 Tension zone and ecotone models

If a stable hybrid zone is maintained by a balance between natural selection, acting through some form of barrier to gene flow, and the homogenising effects of dispersal, what form might these barriers take? There are two predominant modes of selection, outlined below: the existence of others such as frequency-dependent selection (for example for warning patterns in *Heliconius* butterflies (Mallet 1986; Mallet *et al.* 1990)) is acknowledged, but examples are sparse.

Differentiation can be maintained by adaptation to alternative environments or by genomic incompatibilities. In the first, there is spatial variation in environmental conditions, the most simple case being when parental types have higher fitness on their own side of the environmental discontinuity (Endler 1977; Haldane 1948). I refer to this as the ecotone model. In the second, hybrids have reduced fitness (either through heterozygote disadvantage or epistatic interactions) and the environment is homogeneous. This scenario can be termed a tension zone, because of its tendency to minimise its length (Key 1968).

Both cases can be described with simple population genetics models. The relative effects of either type of selection are inextricably linked to the importance of exogenous (mediated by environmental interactions) or endogenous (mediated by within-genome interactions) factors in maintaining taxonomic diversity. Unfortunately, single locus mathematical models suggest that clines in traits generated by either type of selection are indistinguishable in shape (Barton & Gale 1993). In addition, the position of a hybrid zone at an environmental transition is not sufficient evidence for the ecotone model: a tension zone will move until trapped by a density trough or physical barrier, or until the relative fitnesses of the parental types balance, both of which may occur at an environmental transition (Barton & Hewitt 1985). Alternative approaches to interpretation of the observed patterns are therefore required.

In reviews of more than 100 case studies, Barton and Hewitt (1981b; 1985) concluded that "hybrid zones are primarily determined by hybrid dysfunction". Their arguments were largely indirect, and again invoked the frequently-observed concordance and coincidence of clines in a range of traits: "if clines at each locus or for each phenotype were maintained in direct response to the environment, one would not expect them to change in the same way or at the same place: in contrast, almost all hybrid zones consist of a cluster of parallel clines, often involving characters with no obvious functional relation" (from Barton & Gale, 1993; p14). This implies that reproductive isolation requires an accumulation of genetic incompatibilities, rather than adaptation to alternative ecological conditions, and several studies have clearly demonstrated hybrid unfitness in hybrid zones in a range of taxa (to name but a few, *Iris* (Arnold & Bennett 1993), *Mus* (Sage *et al.* 1986), *Podisma* (Barton 1980, Barton & Hewitt 1981a), *Caledia* (Shaw *et al.* 1993), *Pseudophryne* (Woodruff 1979)). However, the issue is controversial. Various cases of well-studied hybrid zones invoke exogenous selection: for example, that of the northern flicker (*Colaptes auratus*) in North America (Moore & Price 1993); of *Heliconius* butter-

flies in South America (Jiggins *et al.* 1996); the mosaic distribution of two species of field crickets *Gryllus*, again in North America (Rand & Harrison 1989), but differential adaptation across an environmental gradient has been unequivocally demonstrated in few animal hybrid zones (Harrison 1990). I return to these issues at subsequent points in this thesis.

1.4 Inferences from patterns

Studies of hybrid zones shed light on more than just the issue of the type of natural selection maintaining a barrier to gene flow: they illustrate the consequences of one genome being challenged by another, and thus allow inferences concerning the dynamics of multiple loci and the effects of interactions within and between genomes.

1.4 (i) Associations between loci

Interbreeding entails recombination and segregation, eroding the integrity of the parental genomes. This can be quantified by measuring the statistical association between alleles of the same ancestry, either as the excess of loci at which both alleles have come from the same parental population, or as the excess of gametes in which alleles at different loci have come from the same parental population. If the loci in question are bi-allelic, with the alleles diagnostic for either population, these two measures are simply the heterozygote deficit (F_{is}) and the linkage disequilibrium (D). F_{is} will be generated by dispersal of individuals between sites of different gene frequencies, but will be removed by random mating within sites. Linkage disequilibrium is also generated by dispersal into the hybrid zone, but is more persistent: each round of random mating only reduces D by a factor of (1-r), where r is the recombination rate. Whilst teasing apart the effects of these factors may be complicated, the magnitude of the associations directly reflects the extent to which alleles from one population remain together despite mixing with another population.

A mathematical model of dispersal counterbalanced by selection on a single locus, either against hybrids or through differential adaptation across an environmental step, generates a tanh (or sigmoidal) curve (Bazykin 1969; Endler 1977; Haldane 1948). These clines may become increasingly stepped in the centre under the influence of selection acting on many loci. Using a model of interbreeding between populations fixed for alternative alleles across multiple loci, Barton has shown how the shape of and association between clines can give estimates of a range of parameters, including the effective selection acting on individual loci, mean fitness in hybrid populations, the strength of the barrier to gene flow presented by this selection and even the number of loci involved (see Barton & Gale 1993 and references therein). Thus in addition to allowing inferences about the form of natural selection acting to maintain differentiation, analysis of hybrid zones can afford quantitative estimates concerning the magnitude of reproductive isolation. The estimate of the number of genes in particular generates widespread interest, and the studies where this has been applied (hybrid zones in the alpine grasshopper *Podisma* (Barton & Hewitt 1981a) and in the fire-bellied toads *Bombina* (Szymura & Barton 1991)) are frequently cited as evidence for the polygenic nature of reproductive isolation. I explore these issues further in the theoretical analyses described in Chapters 5 and 6.

1.5 Other factors affecting hybrid zones

The above issues have formed the backbone for the proliferation of studies of hybrid zones since the mid-1970's. Whilst clearly highlighting fundamental issues, the two models outlined above are highly simplistic, for example in their assumption of random mating between individuals. What might be the effect of assortative mating, relative to the factors discussed above, in preventing gene flow between two populations? In the extreme, will it lead to the third of the three possible outcomes of secondary contact

mentioned above: the two populations remaining distinct, but with sympatric distributions (rather than the "blurred" parapatry of the above models)?

Several studies have documented non-random mating within hybrid zones. The hybrid zone between two subspecies of the European meadow grasshopper, *Chorthippus parallelus parallelus* and *C. p. erythropus* runs through the Pyrenees, approximately following the border between France and Spain; significant positive assortative mating is facilitated by differences in male calling song and courtship song (Butlin & Hewitt 1985; Ritchie *et al.* 1989). In the region of contact between two species of *Heliconius* butterflies (*H. erato* and *H. himera*), hybrids are found at low frequency and marked genetic differences between the parental taxa are maintained in sympatry. This is explained by strong positive assortative mating and (inferred) adaptations across an ecological transition, with no evidence of hybrid inviability and fertility (Jiggins *et al.* 1996; McMillan *et al.* 1997). Non-random mating is also reported in the hybridising field crickets *Gryllus pennsylvanicus* and *G. firmus* in the eastern United States: although differences in the effectiveness of sperm transfer or sperm competition no doubt play a significant role, the magnitude of the deficit in hybrid offspring is such as to require assortative mate choice (Harrison & Rand 1989).

The above examples involve preference for conspecific mates. Non-random mating may also result as a by-product of differences in use of ecological habitats, if mating occurs within habitat type. The preference of different races of the apple maggot fly (*Rhagoletis pomonella*) for different host plants forces assortative mating, as adults tend to rest, and hence presumably to mate, on their native fruit; limited gene flow is however maintained between the races (Feder *et al.* 1990; 1994). A second example of a habitat preference which is maintained despite interbreeding is found in the hybrid zone between the fire-

bellied toads *Bombina bombina* and *B. variegata*; as the subject of the work presented in this thesis, I describe the system in detail below.

Finally, note that habitat association need not imply assortative mating. For example, the hybrid zone between chromosomal races of the lizard *Sceloporus grammicus* in central Mexico is characterised by a mosaic of local patches, and a significant association between the environment of the patch and the karyotype of the individuals found therein. Arguments of scale suggest that the pattern can be explained by strong selection in relation to habitat and against hybrids rather than a habitat preference or assortative mating (Sites *et al.* 1995). Similarly, Hagen (1990) reports differences in performance on alternative host plants between two hybridising subspecies of the eastern tiger swallowtail butterfly *Papilio glaucus*, but no evidence of assortative mating or preference in oviposition site.

These examples illustrate, firstly, the range of possibilities for other factors to affect the dynamics of hybrid zones, and, secondly, the need for careful interpretation of the observed patterns.

1.6 The *Bombina* hybrid zone

I address some of the issues outlined above with a study of the hybrid zone between the fire-bellied toad, *Bombina bombina,* and the yellow-bellied toad *B. variegata*. This section contains a brief description of the taxa, followed by a review of the results of date from previous studies of the hybrid zone.

1.6 (i) The genus Bombina

The fire-bellied toads *Bombina* are one of four genera in the family Discoglossidae. Discoglossids form a small family whose first appearance in the fossil history, approximately 160 Myr ago in Spain, dates back to the Late Jurassic period making it the second oldest family of frogs (Anura) after the Leiopelmatidae of New Zealand (Zug 1993). It is also the only predominantly European family of amphibians or reptiles. As the name suggests, discoglossids are characterised by a disc-shaped tongue; all species are small (for example, adult *Bombina* are usually less than 5 cm long), predominantly aquatic and breed in water. Apart from these characteristics, there is little in common between the genus *Bombina* and the other genera in the family. *Bombina* (Figure 1.6.1) are primarily diurnal and, although dark and camouflaged from above, easily detected due to frequently high activity levels and a tendency to aggregate. They possess a warty, glandular skin, the secretions from which are distasteful to predators; if molested, this unsuitability is advertised by the unken-reflex, an aposematic arching display which reveals the brightly-coloured ventral surfaces (Bajger 1980).

1.6 (ii) Past and present distributions of European Bombina

Of the four species within the genus, two are found in Europe: the fire-bellied toad *Bombina bombina* (L. 1761) and the yellow-bellied toad *Bombina variegata* (L. 1758). The species' ranges are parapatric (see Figure 1.6.2). *B. bombina*'s range extends over

Figure 1.6.1 Adult *Bombina bombina* (fire-bellied toad) male.
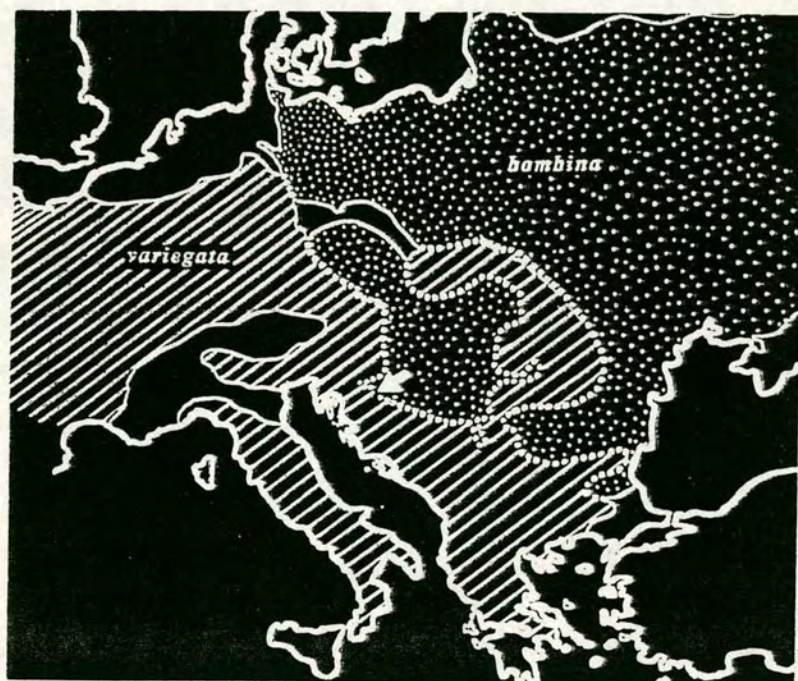
Figure 1.6.2 Distribution of *Bombina bombina* and *B. variegata* in Europe. The location of the transect across the hybrid zone in Croatia is marked with an arrow.

much of eastern Europe, reaching as far west as Denmark and Sweden (although probably now extinct from the latter) in the north and to Slovenia and Croatia in the south. It is generally found in lowland areas. *Bombina variegata* occurs throughout most of central and southern Europe; it is divided into three subspecies, with populations in the southern Balkans defined as *B. v. scabra*, those in Italy as *B. v. pachypus*, and all others as *B. v. variegata* (Mertens & Wermuth 1960). The ranges of the fire-bellied toads must have contracted and expanded periodically following ice-sheet movement during the Pleistocene glaciations (Arntzen 1978). During periods of glaciation, the respective sub-groups were presumably restricted to various refugia: southern Italy for *B. v. pachypus*, the southern Balkans for *B. v. scabus,* the north-western part of the Balkan peninsula for *B. v. variegata*, and the plains bordering the Black Sea for *B. bombina* (Szymura 1993). In post-glacial range expansions, *B. bombina* appears to have expanded at the expense of *B. variegata*, colonising the plains of central and eastern Europe, with *B. variegata* restricted to higher altitudes, sometimes as enclaves within large regions of *B. bombina* (Arntzen 1978). The current distributions of the two species are therefore parapatric, overlapping slightly at the altitudinal transitions at which they meet. Anthropogenic disturbance has excluded both species from certain areas in their respective ranges, but the lowland *B. bombina* has perhaps suffered most from development and land drainage (J. Piálek, pers. comm.).

1.6 (ii) Morphological characteristics

Bombina bombina and *B. variegata* differ in a variety of characteristics, reflecting alternative lifestyles and environments. Most striking, and most well known, is the difference in belly pattern: the ventral patterning is the primary characteristic both identifying the genus and separating the two taxa. *B, bombina* (the fire-bellied toad) has a predominantly black ventral surface, with numerous small white spots and, more importantly, several distinct red-orange or red spots, whereas that of *B. variegata* (the yellow-bellied toad) is usually

of a paler grey background covered largely with bright yellow areas of colour; see Figure 1.6.3.

Within their respective areas, *B. bombina* breeds in semi-permanent ponds, small pools, marshy areas, drainage ditches and even wheel ruts and ephemeral puddles, whereas *B. variegata* prefers only the last two of the list, namely temporary water-bodies (Barandun 1995; Lörcher 1969; Madej 1973). Thus *B. bombina* generally occurs on warmer, agricultural flood-plains, frequently in more permanent sites, whereas *B. variegata* is found in temporary sites at higher altitudes.

An amphibian lifestyle in the respective environments requires different adaptations. Firstly, egg size and development rate is commonly associated with both latitude and altitude, presumably because of the inverse correlation of either with temperature: amphibians generally lay larger eggs at cooler temperatures (Beachy 1993; Berven 1982; Gollmann & Gollmann 1994; Ruibal 1955). Correspondingly, *B. variegata*, existing at consistently higher altitudes, lays larger eggs (Nürnberger *et al.* 1995; Rafinska 1991).

Ephemeral puddles, the breeding sites typical of *B. variegata*, either dry up or become overgrown quickly. This necessitates, firstly, frequent migration of the adults in search of new sites. A more terrestrial life-style generates selection for thick skin as protection against desiccation, a skeletal morphology suitable for migration (Czopkowa & Czopek 1955; Michalowksi 1961; Nürnberger *et al.* 1995) and, presumably, an effective aposematic colour pattern. Secondly, the temporary nature of puddles imposes strong selection on larvae to reach metamorphosis before the sites dry out. This is facilitated by (i) hatching from larger eggs; (ii) faster larval growth and development rates (Gilchrist 1993; Nürnberger *et al.* 1995; Rafinska 1991).
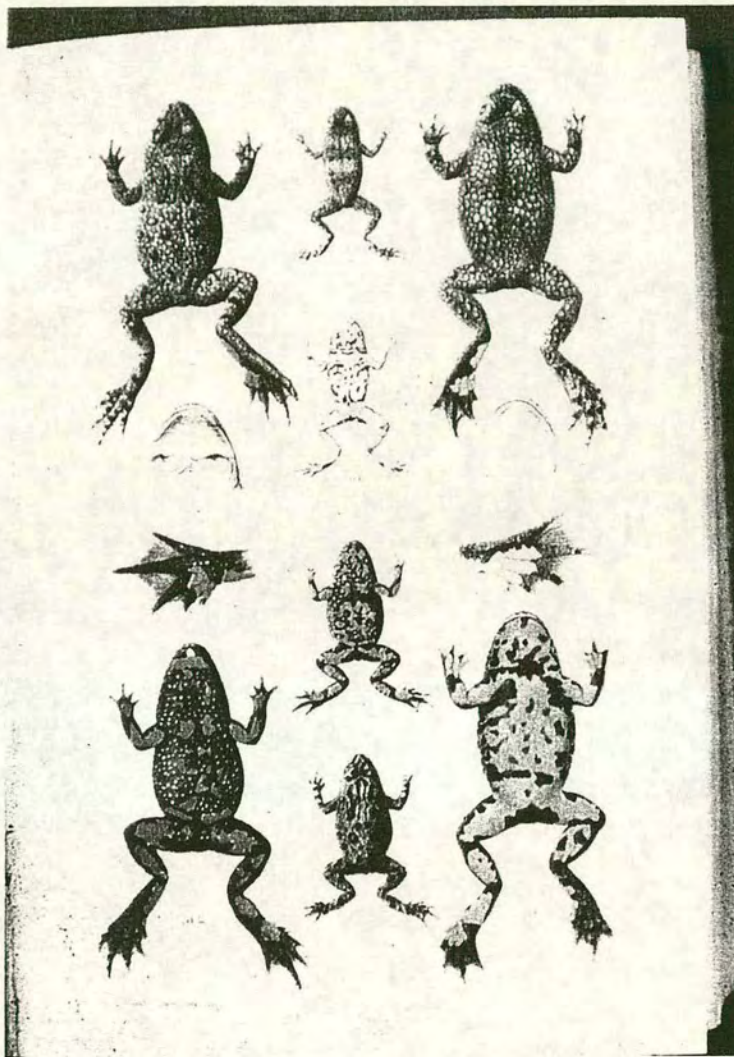
Figure 1.6.3 The yellow-bellied toad *Bombina variegata* and the fire-bellied toad *B. bombina.* Note the difference in body proportions in addition to that in ventral pattern.

Male *B. bombina* possess internal vocal sacs, whereas *B. variegata* do not. Their mating calls vary correspondingly: these are described, helpfully, by amphibian field guides as a "musical, although mournful, 'oop... oop... oop'" (*B. bombina*) as compared to *B. variegata*'s "rather musical 'poop.... poop... poop....'", brighter and faster than the call of *B. bombina*" (Arnold *et al.* 1978). More specifically, the call of *B. bombina* is deeper than and generates more than five times the sound pressure of that of *B. variegata* (Lörcher 1969; Sanderson *et al.* 1992). Large choruses of male *B. bombina* produce a wonderful resonating sound which can be heard from distances of several 100 metres or even kilometres (pers. obs.). It is not known whether the difference in mating call is adaptive to different acoustic characteristics of the respective breeding sites, or whether other factors have constrained the evolution of vocal sacs in *B. variegata*.

1.6 (iii) Genetic characteristics

Data on the genetics of the two species have been collected from more than 70 sites throughout the ranges of *B. bombina* and *B. variegata* in Europe (reviewed in Szymura 1993). Studies have predominantly involved allozyme loci (e.g. Szymura 1983), but mitochondrial DNA (Szymura 1988; Szymura *et al.* 1985) and albumin (Maxson & Szymura 1984) have also been analysed. Estimates of Nei's genetic distance (Nei 1972) from 29 enzyme loci give values ranging from 0.37 and 0.59 between *B. bombina* and the different subgroups of *B. variegata* (Szymura 1993); divergence in mt-DNA sequence is also high, at between 5.6% and 7% (Szymura *et al.*, unpubl. data). A molecular clock argument based on these data suggests a Pliocene split between the taxa, between 2 and 7 million years ago (the argument for a more recent, Pleistocene divergence (Arntzen 1978) is not based on molecular data).

1.6 (iv) The Bombina *hybrid zone and its origins*

Hybridisation between *B. bombina* and *B. variegata* had long been suspected from the occurrence of individuals of intermediate morphology in areas of contact between the respective distributions (Horbulewicz 1927; Karaman 1922; Méhely 1892); Figure 1.6.4 shows a range of intermediate phenotypes found in belly patterns. With the advent of molecular techniques, the hybrid status of these populations was fully confirmed. Evidence also emerged of introgression over long distances into either species' range (Szymura 1976a, b). Despite the many morphological, life history and biochemical differences separating the two species, interbreeding was detected in all regions of contact which have been studied, revealing a hybrid zone which extends for thousands of kilometres around the altitudinal contours of Eastern Europe (Arntzen 1996; Gollmann 1984, 1996; Gollmann *et al.* 1988; Szymura 1976a, 1976b; Szymura & Barton 1986, 1991). Its stability (at least over the last century) is evident from comparisons of the earlier morphological data with more recent observations.

I outline below results from detailed studies of three transects across the hybrid zone: two in southern Poland, and one in central Croatia, marked on Figure 1.6.2. After common usage, I refer to the two taxa as separate species throughout, despite the contradiction posed by hybridising populations under the biological species concept. Phrases such as "hybridisation in *Bombina*" or "the *Bombina* hybrid zone" refer to the hybrid zone in question, between *Bombina bombina* and *B. variegata*.
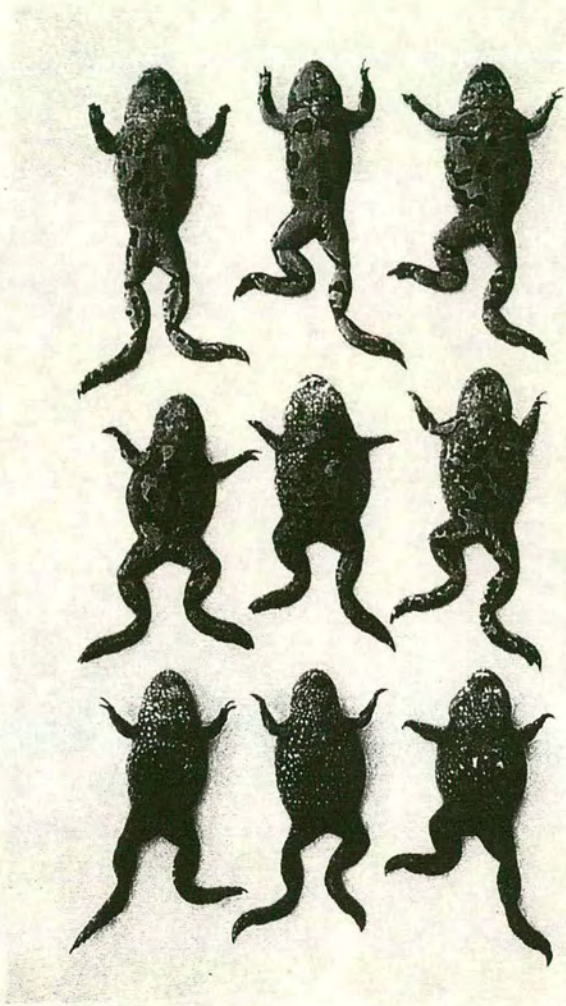
Figure 1.6.4 Hybrid *Bombina* individuals (middle row), shown between pure *B. variegata* (top row) and pure *B. bombina*.

1.6 (v) Inferences from transects across the hybrid zone in Poland

Szymura and Barton 1986, 1991

Detailed studies on transects at Cracow and Przemyśl in southern Poland involved analysis of six enzyme loci; these unlinked loci are diagnostic for *B. bombina* and *B. variegata* and parental alleles are equally functional in F1 hybrids (Szymura & Farana 1978).

Across both transects, clines in allele frequency coincided closely with each other and with morphological (Horbulewicz 1933; Michalowksi 1958) and mitochondrial DNA (Szymura *et al.* 1985) clines; the widths in either location were also similar (maximum likelihood estimates of 6.15 km in Cracow and 6.05 km in Przemyśl). There was no evidence of deviations from Hardy-Weinberg equilibrium within samples from breeding sites, implying random mating across the zone. The clines showed a sharp central step and shallow tails of introgression on either side. This central step is presumably generated by linkage disequilibria between many loci: the maximum value for D between the unlinked allozyme loci was estimated to be 0.055 (from maximum likelihood estimates; 2-unit support limits 0.0375-0.0725). Its effect as a barrier to the flow of a neutral allele was calculated to be equivalent to 51 (22-81) km of unimpeded habitat, and implies that hybrid fitness is 58% (54-68%) that of fitness in the pure population.

The close coincidence of a range of morphological features and genetic markers, and observation of abnormalities in hybrids, were taken as evidence that the hybrid zone is maintained by selection acting against individuals of mixed ancestry (section 1.3), and therefore fits the requirements of a tension zone model. Under a model of selection acting (multiplicatively) against heterozygotes, further calculations on cline shape imply a total of 55 (22-81) loci under selection. (See Chapter Five for theoretical derivations of these estimates.)

MacCallum 1994

MacCallum (1994; MacCallum *et al.* 1997) used a similar analysis to map the transition between *B. bombina* and *B. variegata* across the hybrid zone near Pešćenica, Croatia. The clines shared many features of those in Poland, showing a transition over similar scales with a sharp central step in frequency bounded by shallow tails. However, there were also important differences. Most significantly, populations in Croatia were not in Hardy-Weinberg equilibrium: heterozygote deficit increased significantly towards the centre of the zone, reaching a maximum value of 0.26. Linkage disequilibrium was also higher (even having taken into account the inflation of association due to heterozygote deficits): the maximum value observed was 0.139, compared to 0.055 and 0.043 at Przemyśl and Cracow respectively.

These differences are presumably due to another distinguishing characteristic of the southern populations: mating within the hybrid zone is not random because of an association between type of breeding habitat and genotype. MacCallum (1994) measured a number of ecological characteristics of the alternative breeding habitats; I describe these further in Chapter 4, which considers differential adaptation to the respective ecological conditions. Mark-recapture studies show extensive movement of individuals, both underlining the importance of dispersal in the hybrid zone and implying that the observed association between adult genotype and habitat type is due to an active behavioural preference rather than differential selection on breeding adults. Although this habitat preference introduces problems for the traditional interpretation of clinal patterns, it provides an opportunity to explore the effect of environmental heterogeneity on interactions between divergent populations.

Nürnberger *et al.* (1995) performed a large scale breeding experiment using toads collected from across the Pešćenica transect in Croatia. They scored characteristics in four morphological traits in adults: belly pattern, skin thickness, mating call and skeletal proportions. Crosses were formed either within populations or of F1's (pairing pure individuals of either species): egg size, development time, larval survival and metamorph survival were then scored in the offspring. Many of their results are relevant to the issues addressed in Chapters 2 and 3 of this thesis, and are therefore discussed further in the respective sections. However, summarising briefly, adult traits showed coincident and concordant changes, but clines in egg size and development time were shifted in alternate directions, presumably as a result of variation in environmentally-determined selection pressures. Estimates of linkage disequilibrium calculated from covariances between traits showed large central values, although not as great as those calculated from the data on allozyme frequencies.

1.7 Aims of thesis and chapter outline

Recent studies of the hybrid zone between the fire-bellied toad, *Bombina bombina*, and the yellow-bellied toad, *B. variegata*, in Croatia have implicated both environmentally-mediated selection and assortative mating, generated by a preference for breeding habitats, in its dynamics. This is significant for a hybrid zone which had previously appeared, to all intents and purposes, to fit the predictions of a tension zone model. I explore these issues further in this thesis, taking advantage of the detailed knowledge of the Croatian transect accumulated in the previous studies.

Chapters 2, 3 and 4 describe work on the *Bombina* hybrid zone. In Chapter 2, I consider clines in phenotypic traits across the Croatian transect, testing for any indirect evidence of the selection mechanisms operating. Chapter 3 considers direct evidence for endogenous selection at each stage of the amphibian life cycle, and tests for the principal requirement of a tension zone model: reduced fitness in hybrids. Chapter 4 is concerned with exogenous selection, and tests the prediction that a habitat preference which is maintained despite interbreeding should confer a selective advantage.

Chapters 5 and 6 describe theoretical studies. In Chapter 5, an analytical model of multi-locus clines maintained by differential adaptation across an environmental transition is developed, and its predictions compared with those of a simulation model. Chapter 6 describes an extension of the simulation model to incorporate a habitat preference, and explores the transition between parapatric distributions joined by clines and sympatric distribution.

Chapter 2

Clines in quantitative traits in a *Bombina* hybrid zone

2.1 Introduction

2.1 (i) Patterns of geographic variation

Since Darwin and Wallace, patterns of geographic variation have been used to make deductions about historical evolutionary events. This chapter describes the patterns observed in a transect across the *Bombina* hybrid zone, and the inferences they afford our understanding of the system. It also contains a description of the study site and sampling methods, relevant to all three *Bombina* chapters.

The hybrid zone between *B. bombina* (the fire-bellied toad) and *B. variegata* (the yellow-bellied toad) occurs at the ecotone created by an altitudinal transition: *B. variegata* is found in ephemeral water-bodies in hillier regions, in forested, cooler sites, whereas *B. bombina* is found on the warmer, agricultural flood-plains, generally in more permanent water bodies. As outlined in section 1.6, the respective distributions and life-styles of the two species can probably account for observed differences in a range of traits: skeletal proportions, skin thickness, warning coloration, egg size, larval development rate, male mating call, mating behaviour. These have all been documented in a long history of both field and laboratory studies of *Bombina* (see section 1.6 for references). Any hybridisation between the two species is maybe even surprising, given the extent of differentiation. However, given its occurrence, can we use patterns in these traits to make inferences about the way in which natural selection maintains the distinctions?

2.1 (ii) Inferences from clines

Analysis of the patterns of transition, or clines, in numerous hybrid zones has yielded insight into genomic interactions and selective forces. For example, clines (or gradients) in a variety of characters frequently show coincident patterns of change (Barton & Hewitt 1985; Harrison 1990). *In situ* selection on phenotypic traits provides a plausible

explanation for coincident morphological changes. However, the frequently-observed coincidence of changes in components of cytoplasmic and nuclear DNA (including supposedly neutral markers) as well as in phenotypic traits argues strongly for secondary contact between two already-diverged populations (see section 1.2). Secondly, the coincidence of clines in different traits has been one of the key arguments for the relative predominance of hybrid dysfunction in the maintenance of hybrid zones: clines maintained in direct response to the environment should vary in position and width as the environment varies, and therefore should not cluster together. Finally, concordance (or similar width) of clines implies selection of roughly equivalent magnitude acting on the respective traits. In these arguments, patterns of geographic variation are being used to infer both the origin and the current dynamics of the hybrid zone. I discuss these inferences further in section 2.6, but consider here particular cases which have proved interesting exceptions to the rule.

2.1 (iii) Differences between position and width of clines in hybrid zones

In the hybrid zone between a northern and southern race of the grasshopper *Caledia captiva*, in eastern Australia, pericentric chromosomal arrangements and C-band markers show a sharp transition over a distance of approximately 1km (Shaw *et al.* 1993). In contrast, allozymes, ribosomal DNA and mitochondrial DNA characteristic of the southern race are found hundreds of kilometres to the north, within the range of the northern race. The latter, presumably neutral, markers are thought to be a residue left behind on the contraction of the southern race caused by reduced rainfall over the past 8000 years.

Similar patterns are observed in two hybridising species of European newts, *Triturus marmoratus* and *T. cristatus*, where agricultural practices have altered distribution: higher levels of introgression are observed in areas once occupied by one species but which the

other has now superseded (Arntzen & Wallis 1991). In the hybrid zone between two subspecies of the European meadow grasshopper, *Chorthippus parallelus parallelus* and *C. p. erythropus* in the Pyrenees, there is significant variation in the shape of clines in morphological, electrophoretic, behavioural and chromosomal characters (Butlin 1989; Butlin *et al.* 1991; Hewitt 1993a). Variation in width can be explained by differences in the strength of selection acting against heterozygotes for some of the traits, but in the case of testes dysfunction there is evidence for displaced clines, allowing amelioration of hybrid unfitness (Virdee & Hewitt 1994). These examples illustrate how lack of concordance (different widths) or coincidence (different position) in clines can suggest either past events or present selection mechanisms.

2.1 (iv) Clines in the Bombina hybrid zone

In contrast to the above examples, transects across the *Bombina* hybrid zone in Poland revealed high coincidence between clines in a suite of characters, ranging from mt-DNA and allozymes to characteristics of the male mating call (Sanderson *et al.* 1992; Szymura & Barton 1986, 1991: see section 1.6). This was taken as confirmation both that hybrid dysfunction was the primary factor maintaining the hybrid zone and that none of the traits measured were under strong selection. However, there is evidence that the dynamics of the hybrid zone in southern populations may differ from those of the Polish transects (MacCallum 1994; MacCallum *et al.* 1997; Nürnberger *et al.* 1995).

In their laboratory breeding experiment using 450 toads from the Croatian transect, Nürnberger *et al.* (1995) found no differences between clines in allozyme markers, skeletal proportions, skin thickness, belly patterning and mating call, but displacement (in alternate directions) of clines in egg size and larval development time. The displacement of the cline in larval development time might be explained in terms of environmental features of the area around Pešćenica. The transect is characterised by a staggered change

in various factors, with forest stretching out into the flat, low-lying arable areas; water body permanence is most critical in the puddles of this central area, generating strong selection for minimal larval periods. This contrasts with at least one of the Polish transects: at Cracow, forest cover and relief effectively change simultaneously (Arntzen 1996). If the hybrid zone is maintained by differential selection in relation to environment, different traits should be affected by different environmental factors, and adaptation might be manifested in disconcordant clines such as those shown in the breeding experiment.

In this chapter, I consider data from 2700 adult toads measured in the field in Croatia, for which there is information on genotype at four diagnostic allozyme markers, relative leg length and belly pattern, and data on egg size measured in the field. It would seem justifiable to expect selection of different magnitude on these four traits; the larger sample sizes should provide the resolution necessary to tease apart any effects of varying selection pressures or environmental heterogeneity.

2.1 (v) Multilocus associations: linkage disequilibrium

Geographic variation in chosen traits can be described in various ways. The most simple is to map the change in means and variances of traits measured in populations across the transect. However, associations between traits are also informative because they represent the underlying associations between sets of loci, or linkage disequilibrium (disregarding the possible effects of pleiotropy). Measures of linkage disequilibrium from data on different traits should therefore reflect the homogeneity of changes in genome composition across the cline. Linkage disequilibrium can be calculated from covariances in quantitative traits; this allows inferences about genetic parameters to be made solely from phenotypic data. This method was used by Nürnberger *et al.* (1995), but generated discrepancies: estimates of linkage disequilibrium from phenotypic traits were half those

estimated from genetic data. Analysis of a larger data set, allowing finer-scale classification of hybrid classes, may resolve these issues.

2.1 (vi) Aims

This chapter therefore has the following aims:

(1) To describe the transect across the *Bombina* hybrid zone at Pešćenica, in Croatia, and the methods of sampling and measurement of toads.

(2) To compare the changes in quantitative traits across the hybrid zone, and test the prediction that either variation in selection pressures or adaptation to different environmental features should result in clines of dissimilar width or position.

(3) To describe methods of estimating underlying statistical associations between loci from phenotypic data.

In section 2.2, I describe the general methods of data collection; the following sections deal with, respectively, changes in the mean, variance and covariance of all the traits. Derivations of the statistics used are presented in the respective sections.

2.2 Methods

2.2 (i) Study site

The study site is located 20 km south-east of Zagreb, Croatia, centred on the village of Peščenica (45°36'N, 16°10'E); see Figure 2.2.1. The hybrid zone occurs at an altitudinal transition between the arable flood plains of the river Sava (to the north-east) and low, forested hills (to the south-west). In this area, the distance across the zone between pure *B. bombina* and pure *B. variegata* populations is approximately 10 km (MacCallum 1994) and the altitude changes from ≈100m to 280m. *Bombina* are found in a variety of breeding sites within the hybrid zone, ranging from the ox-bow of a river, artificial ponds, drainage ditches, small pools in marshes or swampy areas to numerous temporary puddles: Figure 2.2.2 shows a semi-permanent pond (typical of *B. bombina*) and a puddle formed in a wheel-rut (typical of *B. variegata*). The centre of the hybrid zone runs through an area of forest stretching out from the hills onto the lowlands, consisting predominantly of *Quercus robur* (pedunculate or summer oak), *Q. petraea* (sessile or winter oak) and *Carpinus betulus* (hornbeam), changing to *Fagus sylvatica* (beech) in the hilly regions. The forest is managed with low-intensity selective logging. Whilst this management reduces the incidence of natural depressions formed by old trees falling over, logging tractors create deep wheel-ruts in the clay-like soil; the majority of all puddles used by *Bombina* as breeding sites therefore appear to be in wheel-ruts.

The data presented in this chapter were collected over four field seasons: 1991 and 1992 (MacCallum, Barton and Nürnberger), 1994 and 1995 (Kruuk, Gilchrist and Piálek). All field seasons coincided approximately with the *Bombina* breeding period from late April until early July, although variability in weather conditions between years had a large effect on the timing of reproduction. Over the four years, a total of 197 sites were sampled, many in successive years; almost all sites were visited in later years, so the

Figure 2.2.1 SPOT satellite image of the area around the village of Pešćenica (marked by red cross). Dark areas represent forest, green or white represent arable land. Note the increased relief to the south-west, and the river Sava running across the north-east corner.

Figure 2.2.1 Overlay The overlay shows the frequency of *B. variegata* alleles (in black) at four diagnostic enzyme loci. Note that some sites are found outside the region covered by the satellite image.

Figure 2.2.2
Breeding sites of *Bombina* in
Croatia: (top) semi-permanent
pond, typical of *B. bombina*;
(below) temporary puddle formed
in wheel rut, typical of *B.
variegata* (spot the toad).

subsequent absence of samples implies that no *Bombina* were found at the site. In the case of large water bodies, or collections of puddles over an extended area, a site was always specified to cover an area less than 100 m in radius; in practice, most sites covered significantly smaller areas. MacCallum (1994) describes a full ecological analysis of the sites sampled in 1991 and 1992. Sites were numbered sequentially by the order in which they were found, prefixed by the year of first sampling: for example, the 113 sites first sampled in 1991 were labelled 1001-1113, the 65 sites first sampled in 1992 were labelled 2114-2179, and so forth. Site number therefore conveys no spatial information. Sites were located on a SPOT satellite image of 10m resolution, from which co-ordinate readings were taken.

2.2 (ii) Sample collection

Toads collected from a site were brought back to the field laboratory in Pešćenica village. Each individual was given a unique identification number consisting firstly of one digit for the year in which it was sampled (1 for 1991 up to 5 for 1995), followed by three digits referring to the site (001 up to 208), and finally an individual suffix: for example, the first toad sampled at site 001 in 1991 was defined as 1001.01, the fourth toad sampled from site 011 in 1992 was defined as 2011.04 and so forth. Toads were anaesthetised in 0.2% MS222 (3-aminobenzoic acid ethyl ester, Sigma); the anaesthetic lasts for approximately 15 minutes, during which time an individual's belly pattern was photographed, body measurements and the connectedness spot-score (described below) taken, sex determined from the presence (males) or absence of nuptial pads on the fore limbs and a toe clip taken. Toads were returned to their original sites within two days; toe clips were frozen in liquid nitrogen and returned to Edinburgh. The number of individuals measured at a site varied between traits; for example, sample sizes for the more-easily collected phenotypic data are larger than for the genetic data. "Population

mean" refers to the mean value of the trait among all individuals measured at that site across all years.

2.2 (iii) Genetic analysis: the hybrid index

Tissue samples (toe clips) were scored for genetic variation at four diagnostic allozyme loci: adenylate kinase (Ak), malate dehydrogenase (Mdh-1); lactate dehydrogenase (Ldh-1), and isocitrate dehydrogenase (Idh-1). MacCallum scored genotypes of 1354 individuals sampled from 173 breeding sites in 1991 and 1992 (MacCallum 1994), providing a high-resolution map of the change in genotype distribution across the hybrid zone. As some of the procedures described in the subsequent chapters involved 20 breeding-sites not sampled by MacCallum, I scored genotypes on a further 131 individuals to classify these sites. All four allozyme markers are known to be unlinked (Szymura & Farana 1978) and are assumed to be selectively neutral. Staining was performed using agar overlays (Shaw & Prasad 1970); detailed protocols and recipes are given by Szymura (Szymura 1976a,b; Szymura 1983). The hybrid index H for an individual is defined as the number of *B. variegata* alleles scored out of the total possible 8 (so $0 \leq H \leq 8$). An individual's gene frequency p is the proportion of *B. variegata* alleles (so p=H/8 if all 4 loci have been successfully scored); $\bar{p}$ refers to the population (or site) mean. The overlay on Figure 2.2.1 shows the location and $\bar{p}$ value for each of the 193 sites.

2.2 (iv) Spot score

The ventral pattern of *B. bombina* is typically characterised by small, unconnected red spots, whereas that of *B. variegata* contains larger, frequently connected, yellow spots (see Figure 1.6.4). Analysis of digitised images of photographs of 400 toads indicated that, whilst characteristics such as the amount of colour or its hue are certainly diagnostic, a "spot score" (derived by J. Szymura (Szymura and Barton 1991)) based on the number of connections between a specified set of colour areas is subject to less random noise

and, more importantly, most highly correlated with an individual's genotype (Kruuk and Baird, unpubl. data). The spot score is derived by summing over ten possible connections. On the resulting scale from 0 to 10, *B. bombina* scores low values and *B. variegata* high values, although individuals at either extreme are rare: in particular, individuals from pure *B. variegata* populations are more likely to have spot scores of 7 or 8 than 9 or 10.

2.2 (v) Femur length

Femur and snout-vent length were measured on anaesthetised toads, to 0.1 mm with a vernier calliper. Femur length was taken as the vent-knee distance in 1991-2 measurements (C. MacCallum, pers. comm.) and as half the between-knee distance for 1994-5 measurements; both were taken with the toad placed face-down with legs positioned so that the femur was perpendicular to the spine. The difference in techniques does not introduce any bias in measurements. To remove correlations with body size, femur length was divided by snout-vent length; the analysis uses a logarithmic transformation of this (dimensionless) ratio. The slope of a linear regression of log femur length against log body size did not differ significantly from unity; the magnitude of the femur ratio is therefore not confused by any allometric relationships within skeletal proportions (as in Nürnberger *et al.* 1995).

2.2 (vi) Egg size

Eggs which had been laid naturally in the field were collected from twenty-one sites throughout the 1994 and 1995 breeding seasons. To maximise the number of families measured, collections were made on different dates and from batches several metres apart. These were transported to the field laboratory in small glass phials, and kept cool during transport to slow development, as measurements of size require an egg to be at pre-gastrula stages (Gosner 1960); prior to this stage changes in size are negligible

(Nürnberger *et al.* 1995). The diameter of each egg was measured using a dissecting microscope fitted with a graticule, giving resolution to the nearest 0.03 mm. Diameters were converted to estimates of volume (under the assumption that eggs were spherical), and the mean of five eggs for each batch calculated.

Sites were chosen to form an approximate transect across the hybrid zone. The number of egg batches sampled and the mean allozyme frequency in adults at each site are listed in Table 2.2.1.

Table 2.2.1 Sample sizes (N) and adult gene frequency $\bar{p}$ at egg collection sites.

Site	(N)	$\bar{p}$	Site	(N)	$\bar{p}$	Site	(N)	$\bar{p}$	Site	(N)	$\bar{p}$
5208	(18)	0.021	1103	(12)	0.227	5203	(20)	0.578	1099	(5)	0.786
5205	(3)	0.031	1011	(11)	0.402	1113	(3)	0.695	4156	(10)	0.852
1039	(10)	0.063	4183	(2)	0.403	1001	(20)	0.701	2165	(12)	0.921
5207	(2)	0.139	2150	(2)	0.406	1021	(10)	0.719	2138	(4)	0.932
1064	(4)	0.223	4187	(16)	0.519	4180	(7)	0.755	5209	(8)	0.964

Species identification

Identification of amphibian eggs to species or even genera is a non-trivial task, particularly at the early stages of development. Within the Croatian study site, *Hyla arborea* (tree frog) and six *Rana* (green and brown frog) species are common; *Bufo bufo* (common toad), *Bufo viridis* (green toad) and *Pelobates fuscus* (common spadefoot toad) are also present. Differentiation of *Bombina* eggs from those of the other species present is straight-forward for most cases, although not as clear-cut as amphibian field guides suggest (Arnold *et al.* 1978). Problems were often caused by *Hyla* eggs, for which the difference in size, compactness of cluster or delineation of individual jelly capsules from

Bombina may not be obvious, particularly in the absence of comparisons. With practice (by the end of a field season!) *Hyla* eggs could be reliably identified by the above characteristics. However all batches were reared in the laboratory to stages at which tadpoles could be identified with complete confidence: *Hyla* tadpoles are readily distinguished by their widely-spaced eyes. The colour and size of *Rana* eggs are similar to *Bombina*, but numbers per egg mass are considerably greater. *Rana* tadpoles at early stages are not so distinct from *Bombina* as *Hyla* are. Potential confusion of tadpoles can always be avoided by checking the position of the spiracle (using a blade of grass for small tadpoles): *Bombina* is one of the few genera with a ventral, rather than side, spiracle.

2.3 Changes in mean of quantitative traits

2.3 (i) Methods

For an initial comparison of quantitative trait values, sites are divided into three broad categories according to their mean gene frequency: *B. bombina* for populations where $\bar{p}<0.2$, hybrid for $0.2\leq\bar{p}\leq0.8$ and *B. variegata* for $\bar{p}>0.8$. Populations are pooled within these categories, and the significance of differences tested for using a non-parametric Kruskal-Wallis test for spot score and F-tests for log femur ratio and egg volume.

If clines in different traits are either non-concordant (different widths) or non-coincident (different centres), plotting one against the other will give a non-linear relationship; see Figure 2.3.1. This indirect approach avoids problems associated with any spatial analysis, such as variation in cline width or between alternative environment types.

2.3.1 (a)
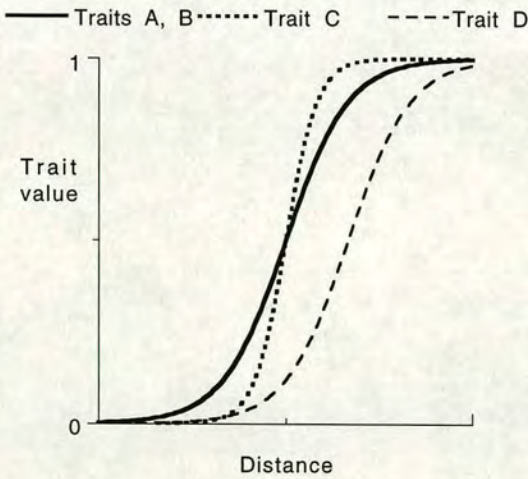
2.3.1 (b)
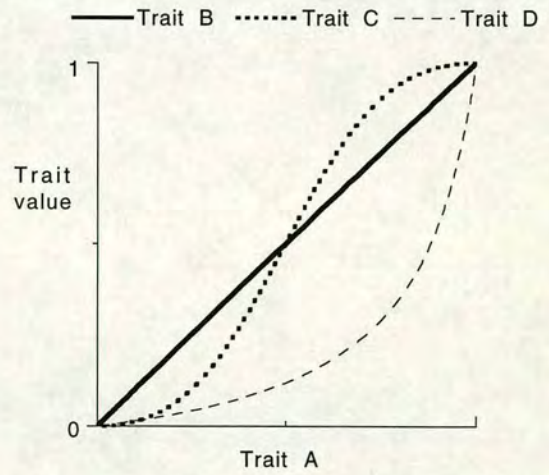
Figure 2.3.1. Plot of the relationship between clines in four traits, two of which (A and B) are concordant and coincident, one of which is narrower (C) and one of which is displaced (D). **(a)** Four clines plotted against distance across transect; **(b)** Clines B, C and D plotted against A. A narrower (or wider) cline implies a significant cubic term in a regression of one against the other; a displaced cline implies a significant quadratic term.

A least-squares regression of mean trait value on $\bar{p}$ is fitted, with sites weighted by sample size; the F-ratio for the sum of squares accounted for by a quadratic or cubic term indicates the significance of any non-linearity, and hence of any non-coincidence or non-concordance (respectively) between clines. The total change (Δ) in the value of a trait across the cline is calculated by extrapolation of the regression fit to $\bar{p}=0$ and $\bar{p}=1$. Changes in variance across the transect are similarly described with least-squares regression models.

2.3 (ii) Results

Table 2.3.1 shows the mean values in the three genotype classes; differences between classes are significant for all three traits.

Changes in population means with respect to $\bar{p}$ and a least-squares regression are given in Figure 2.3.2. For spot score, the relationship is significantly non-linear (quadratic term: $F_{1,141}=4.5196$, $p=0.035$), though, as Figure 2.3.2(a) shows, not greatly different qualitatively from a linear relationship: the maximum difference between the two regressions being only 0.213, at $p=0.5$. (A cubic term is not significant: $F_{1,140}=0.0137$, $p=0.907$.) For both femur ratio and egg size, the quadratic term is not statistically significant (femur ratio: $F_{1,136}=2.445$, $p=0.120$; egg volume: $F_{1,19}=0.003$, $p=0.958$). There is good correspondence between $\bar{p}$ and both mean spot score ($R^2=0.92$) and egg volume ($R^2=0.84$), but relatively more noise in the relationship between $\bar{p}$ and log femur ratio ($R^2=0.30$).

Table 2.3.1 Means, standard deviations (SD) and samples sizes (N) for quantitative traits with sites classified as *B. bombina* ($\bar{p}<0.2$), hybrid or *B. variegata* ($\bar{p}>0.8$). P-value for tests of difference between three classes calculated using [1]Kruskal-Wallis or [2]F-tests. Change (Δ) in mean value across cline is calculated from the regression models given in Figure 2.3.2.

	B. bombina	Hybrid	*B. variegata*	p-value	Δ
	Mean (SD) N	Mean (SD) N	Mean (SD) N		
Spot score	1.79 (1.52) 1015	4.62 (2.39) 830	6.70 (1.62) 568	<.001[1]	5.94
Log femur ratio	-1.20 (0.15) 947	-1.16 (0.15) 759	-1.09 (0.11) 507	<.001[2]	0.046
Egg volume (mm^3)	2.79 (0.54) 33	4.56 (2.76) 115	5.89 (3.38) 34	<.001[2]	3.66

2.3.2 (a) Spot score
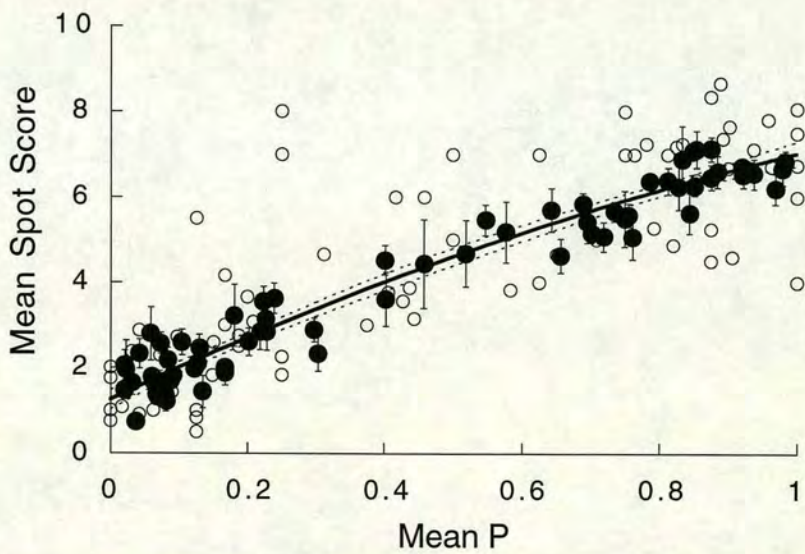
Least-squares regression: *mean spot score* $= 1.257 + 7.613\bar{p} - 1.834\bar{p}^2$, $R^2 = 0.924$.

Figure 2.3.2 Population means against $\bar{p}$ for **(a)** spot score; **(b)** log femur ratio **(c)** egg volume; dotted lines give 95% confidence intervals for regression prediction (see text of regression statistics). In (a) and (b), open circles indicate sites with data on fewer than 10 individuals, closed circles indicate sites with data on 10 or more individuals.

2.3.2 (b) Log femur ratio
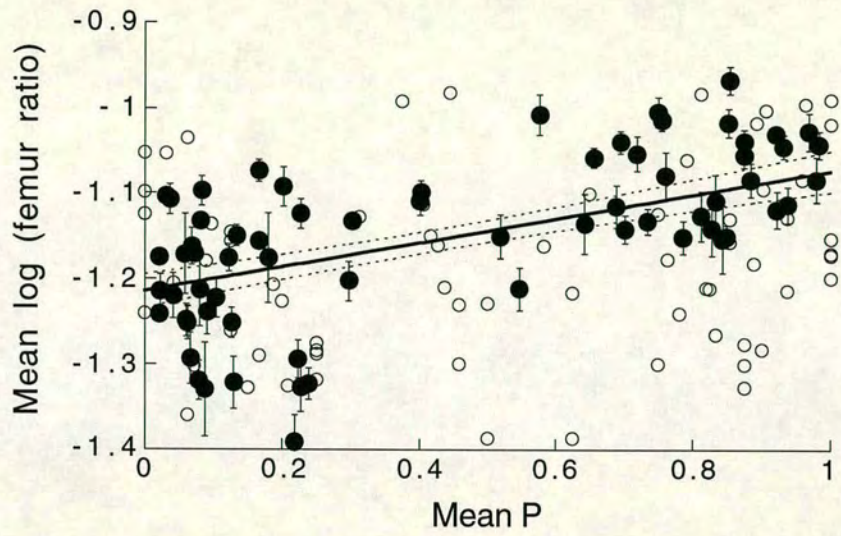
Least-squares regression: *Mean log femur ratio* $= -1.215 + 0.141\overline{p}, \quad R^2 = 0.304$.

2.3.2 (c) Egg volume
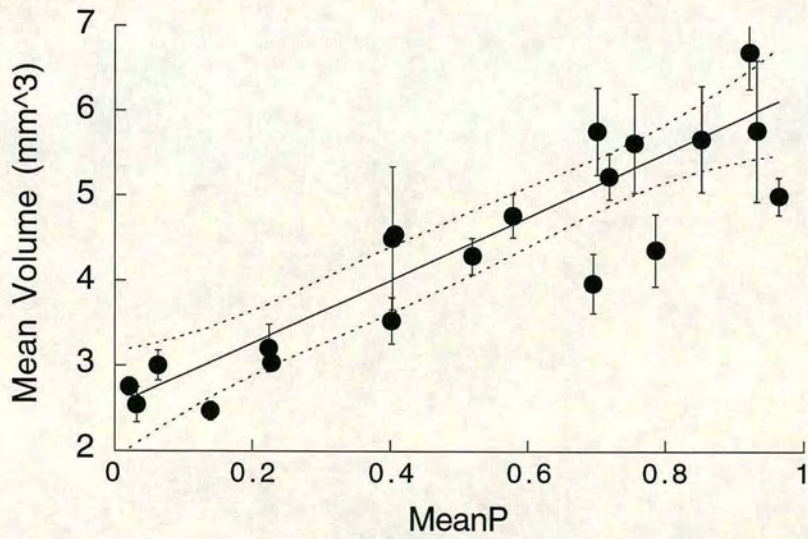
Least-squares regression: *Mean egg volume* $(mm^3) = 2.527 + 3.712\overline{p}, \quad R^2 = 0.842$.

Figure 2.3.2 See previous page for figure legend.

Data on both sexes were combined in the above comparisons. In Figure 2.3.3, the mean allozyme frequency (at the 4 diagnostic loci) in females at each site is plotted against the mean allozyme frequency in males. Although no pattern is obvious from the plot, the cubic term in a (weighted) linear regression model is significant; see Table 2.3.2 for details of the model. As comparison of Figure 2.3.3 with Figure 2.3.1(b) indicates, this implies that the cline in female gene frequency is narrower than the cline in male gene frequency. The spot score shows a similar pattern: the cubic term is significant in a regression of female spot score against mean male spot score ($F_{1,134}=6.468$, $p=0.012$). However the cubic term is not significant in the equivalent regression on log femur ratio ($F_{1,134}=1.024$, $p=0.313$).

Table 2.3.2 Details of cubic regression of mean allozyme frequency in females on mean allozyme frequency in males.

Term	df	Sequential SS	F-ratio (df)	p-value
$\bar{p}$	1	232.51	846.03	<0.0001
$\bar{p}^2$	1	0.10	0.38	0.5401
$\bar{p}^3$	1	4.33	15.73	0.0001
Error	101	27.76		
Total	104	264.69		
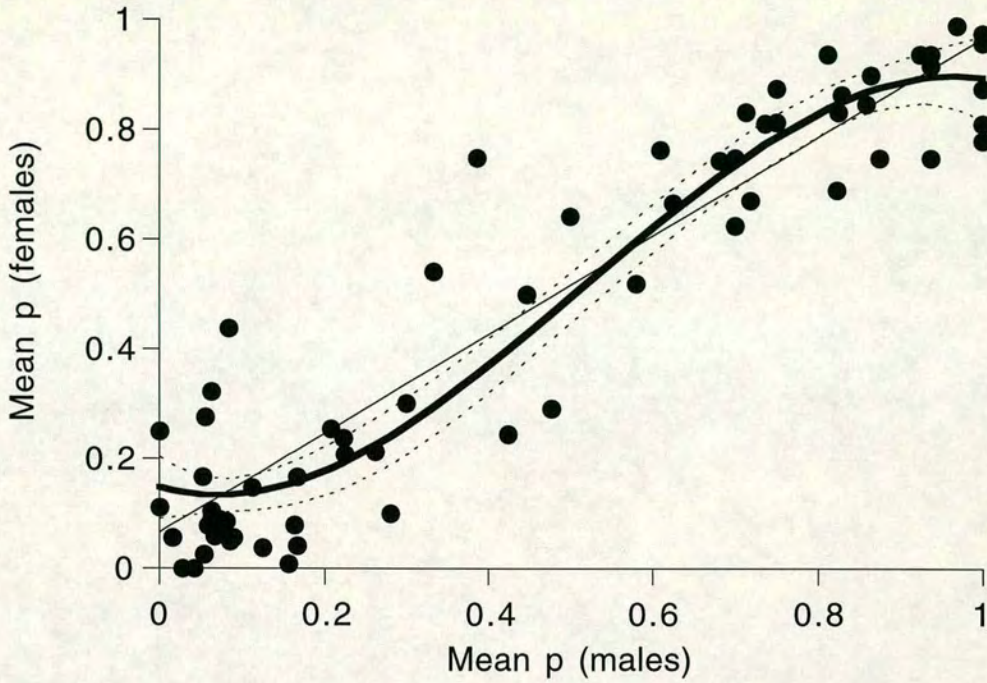
Figure 2.3.3 Comparison of mean allozyme frequency in females with mean allozyme frequency in males. Only sites at which more than 10 individuals were sampled are shown, but the regression is calculated on all sites: the heavy solid line gives a cubic regression (see Table 2.3.2 for details); the dotted lines give 95% confidence intervals for the cubic fit; the light solid line gives a linear regression.

2.4 Changes in variance of quantitative traits

Figure 2.4.1 shows the magnitude of within-population variance in spot score, log femur ratio and egg volume, plotted against $\bar{p}$, and the respective least-squares regressions. The three traits each change in a different way across the cline: for spot score, variance increases in the centre; for femur ratio, variance is greatest on the *B. bombina* side; for egg volume, on the *B. variegata* side.

2.4.1 (a) Spot score
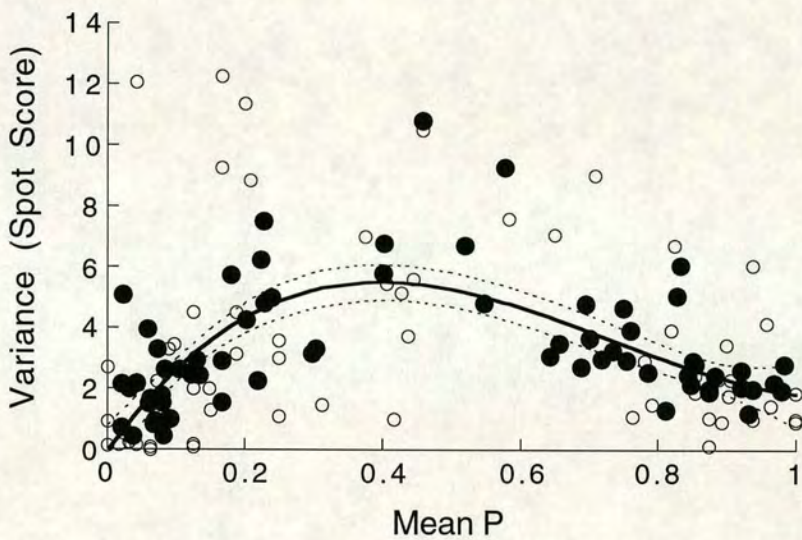
$$Spot\ score\ variance = -0.07 + 31.92\bar{p} - 55.04\bar{p}^2 + 25.02\bar{p}^3, \quad R^2 = 0.407$$

Figure 2.4.1 Within-population variance against $\bar{p}$ for **(a)** spot score; **(b)** log femur ratio **(c)** egg volume (mm³); dotted lines give 95% confidence intervals for the prediction. In (a) and (b), open symbols indicate sites with data on fewer than 10 individuals and closed symbols indicate sites with data on 10 or more individuals. (2.4.1(b) & (c) overleaf.)

2.4.1 (b) Log femur ratio
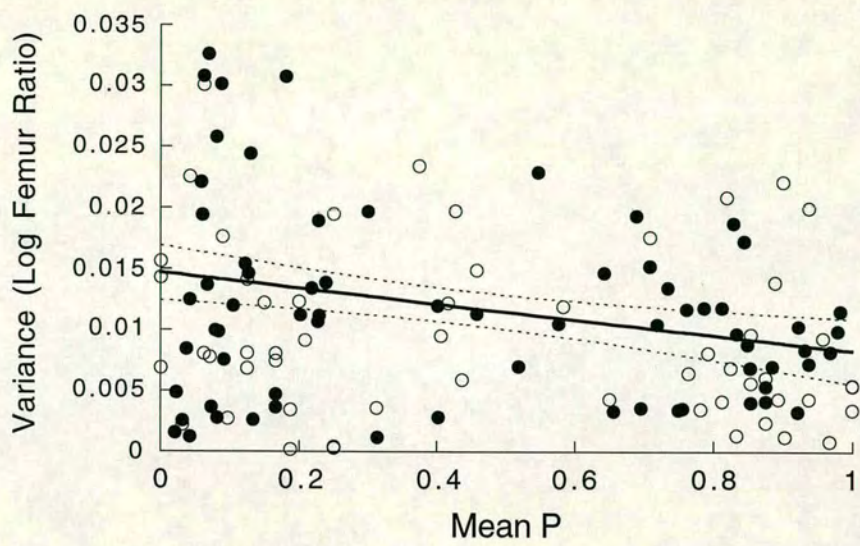
Log femur ratio variance $= 0.0147 - 0.0064\bar{p}, \quad R^2 = 0.076$

2.4.1 (c) Egg volume
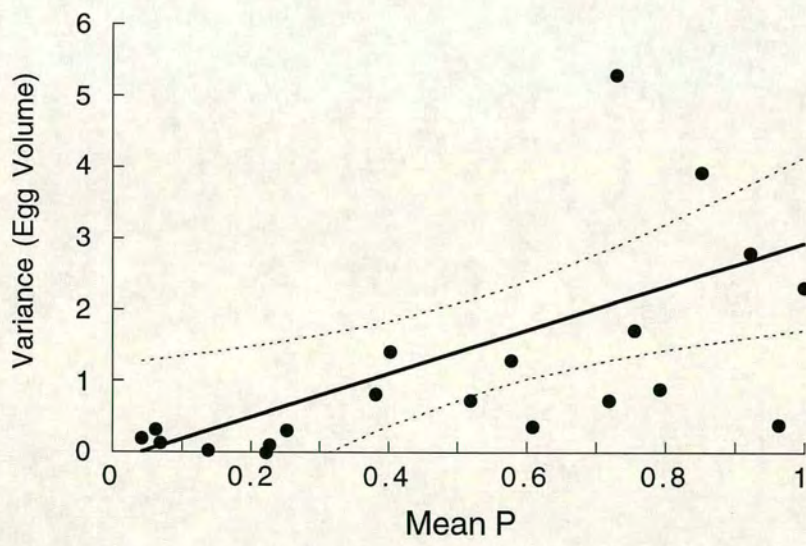
Egg volume variance $= -0.118 + 3.050\bar{p}, \quad R^2 = 0.325$

Figure 2.4.1 See previous page for figure legend.

2.5 Changes in linkage disequilibrium

I consider in this section various approaches to estimating the statistical associations between sets of loci. These rely on the correspondence between linkage disequilibrium and covariances of genetic values, allowing linkage disequilibrium D to be estimated from phenotypic data. As outlined in section 1.4, linkage disequilibrium is a highly informative statistic, reflecting the extent to which alleles from one population remain together despite interbreeding between populations. However, the rationale behind the analysis presented here is ultimately the same as in the preceding sections: are there differences between the estimates made using data on alternative traits which might suggest variation in the intensity or the spatial distribution of the selection pressures to which each is subject?

2.5 (i) Methods

In all the estimates of linkage disequilibrium described below, populations are pooled within eight classes according to their $\bar{p}$ values, as defined in Table 2.5.1. (Note that these differ slightly from the classes used by MacCallum (1994).)

Table 2.5.1 Classification of sites according to $\bar{p}$.

Class	Range of $\bar{p}$	No. of sites	Class	Range of $\bar{p}$	No. of sites
1	0.00 - 0.10	32	5	0.50 - 0.65	14
2	0.10 - 0.20	20	6	0.65 - 0.80	14
3	0.20 - 0.35	13	7	0.80 - 0.90	22
4	0.35 - 0.50	12	8	0.90 - 1.00	20

(a) Definition of pairwise linkage disequilibrium D

The pairwise linkage disequilibrium D between two loci is defined as the excess in the observed frequency of coupling gametes above that expected due to random association (e.g. Hartl & Clark 1989). If two loci are each segregating for two alleles at respective

frequencies p_1 and p_2, and a coupling gamete containing both alleles has frequency P_{11}, then $D = P_{11} - p_1 * p_2$.

Linkage disequilibrium is generated by the mixing of different gene pools through dispersal and by selection for co-adapted gene complexes; it is broken down by recombination between genomes. A mathematical description of these dynamics is given in Chapter Five; I outline here the qualitative arguments. The effect of selection on disequilibrium is probably small relative to the effect of mixing generating it and recombination removing it (Barton 1983). Because the magnitude of D generated by mixing two populations is proportional to the difference between their respective gene frequencies, linkage disequilibrium will be a function of the slopes of the clines in gene frequencies. It will therefore reach maximum values in the centre of the cline, at the steepest point. Using a similar argument as above, any non-concordance or non-coincidence in different clines will be reflected as differences in the pattern of change of D across the zone.

(b) Estimating D from covariance of allelic values

The pairwise linkage disequilibrium D is equal to the covariance in state between the two loci[1] (Hartl & Clark 1989). This can be used to derive an estimate for average pairwise linkage disequilibrium $\overline{D}$ using the variance of the hybrid index. If data from L loci are summed to give an index $H = \sum_{i=1}^{L} x_i$, where x_i is the number of alleles (0, 1 or 2) of one type (e.g. *B. variegata*) at a diploid locus, then, under the assumption (for the moment) of Hardy-Weinberg equilibrium, the mean $\overline{D}$ of the linkage disequilibrium between all possible pairs of loci can be calculated from the following:

---

[1] Defining the allelic states at the two loci as the random variables X and Y respectively:

$covar(X, Y) = E\left((X - \overline{X})(Y - \overline{Y})\right)$

$= (0 - p_1)(0 - p_2) freq00 + (0 - p_1)(1 - p_2) freq01 + (1 - p_1)(0 - p_2) freq10 + (1 - p_1)(1 - p_2) freq11$

$= p_1 (p_2 freq00 - (1 - p_2) freq01) - (1 - p_1)(p_2 freq10 - (1 - p_2) freq11)$

$= p_1 D - (1 - p_1) D \qquad since\ D = freq11 - p_1 p_2 = -freq01 + (1 - p_1) p_2\ etc.$

$= D$

$$var(H) = \sum_{i,j=1}^{L} covar(x_i, x_j)$$

$$= \sum_{i=1}^{L} var(x_i) + \sum_{i \neq j} covar(x_i, x_j)$$

$$= 2\sum_{i=1}^{L} p_i q_i + 2\sum_{i \neq j} D_{ij}$$

$$= 2\sum_{i=1}^{L} p_i q_i + 2L(L-1)\overline{D} \qquad (2.5.1)$$

where p_i is the allele frequency at the ith locus, ($q_i = 1-p_i$) and D_{ij} is the pairwise linkage disequilibrium between locus i and locus j (Barton & Gale 1993). This method provides an unbiased estimate of satisfactory resolution, although the accuracy is not quite as high (C. MacCallum, unpublished data) as that of estimates made using a more complicated maximum likelihood method (Edwards 1972; Hill 1974).

(c) Estimating D from covariances of quantitative traits

The equivalence between covariances and linkage disequilibrium can also be exploited to give estimates of the genetic linkage disequilibrium from data on morphological traits (Barton & Gale 1993; Nürnberger *et al.* 1995). Consider two quantitative traits Z and Z' defined by the effects of, respectively, L and L' loci. Under the assumptions of additive gene action, the absence of pleiotropy (so the two sets of loci are disjunct) and independent environmental effects E and E', the observed phenotypes will be:

$$Z = \sum_{i=1}^{L} \alpha_i x_i + E, \quad Z' = \sum_{i=1}^{L'} \alpha_i' x_i' + E', \qquad (2.5.2)$$

where x_i is, as before, the number of alleles of one type at the ith locus and α_i is the average effect of one allele at that locus. The change in the mean of a particular trait across the zone is:

$$\Delta Z = \sum_{i=1}^{L} \alpha_i (2\Delta p_i)$$

$$= 2 \sum_{i=1}^{L} \alpha_i \qquad \text{if underlying alleles are fixed}$$

(2.5.3)

Under the above assumptions, the covariance between Z and Z' must be genetic and hence due to the linkage disequilibria. The pairwise linkage disequilibrium D_{ij} between loci i and j can be described by the mean value $\overline{D}$; assuming that this average is across loci that are fixed on either side of the zone gives:

$$cov(Z, Z') = \sum_{i=1}^{L} \sum_{i=1}^{L'} \alpha_i \alpha_j \, cov(x_i, x_j)$$

$$= 2 \sum_{i=1}^{L} \sum_{i=1}^{L'} \alpha_i \alpha_j D_{ij}$$

$$= \tfrac{1}{2} \Delta Z \, \Delta Z' \overline{D}$$

(2.5.4).

This gives a formula for estimating $\overline{D}$ from data on any two suitable quantitative traits (Nürnberger *et al.* 1995):

$$\overline{D} = \frac{2 * cov(Z, Z')}{\Delta Z \, \Delta Z'}$$

(2.5.5)

In the calculations below, a mean value $\overline{D}$ is estimated for each of the eight classes. To prevent covariances being inflated by between-site variation within each class, population means are subtracted from individuals' trait values before analysis. Treating hybrid index, spot score and log femur ratio as three quantitative traits gives three different pairwise combinations from which to estimate $\overline{D}$; note that the spot score and log femur ratio combination affords an estimate of linkage disequilibrium entirely from phenotypic data.

(d) Effect of non-Hardy-Weinberg proportions

The above derivations were all made under the assumption of Hardy-Weinberg equilibrium. However, correlations between allelic state at two homologous genes (within locus associations, or heterozygote deficit F_{is}) will also inflate the total covariance. Estimates

of D using the variance in the hybrid index (equation 2.5.1) or the covariance in quantitative traits (equation 2.5.5) will therefore be biased by any degree of non-random gametic assortment. I refer to such estimates as F_{is}-inflated linkage disequilibrium, or D_F. Chapter 6 considers explicitly the effects of F_{is} on the observed covariances. For the purposes of this section, all D_F estimates should have been affected to the same degree, and so whilst comparison with absolute measures of D estimated from other methods (such as maximum likelihood estimates) are not valid, D_F can still be used to give a comprehensive measure of overall statistical associations.

(e) Statistical confidence for linkage disequilibrium

Measures of statistical confidence for covariance estimates are not straightforward. Any n*n variance-covariance matrix with a given number of degrees of freedom follows a Wishart distribution (using the Multivariate Statistics routine in Mathematica 3.0 (Wolfram 1996)). Given the variances and covariance between two variables (i.e. a 2*2 variance-covariance matrix) and the degrees of freedom on which they are based, a likelihood function can be generated by integrating the Wishart distribution over the range of possible variance values, and thus defining a support curve. Support limits for the covariance are then given by the range of values falling within two units of the maximum likelihood (N. Barton, pers. comm.).

2.5 (ii) Results

Figure 2.5.1 shows the change in D_F across the transect; values (and support limits) are given in Table 2.5.2. All the values show a significant increase in the centre of the zone, implying that dispersal of relatively intact gene combinations (and possibly selection, but see section 5.4) into the hybrid zone is maintaining statistical associations across the genome despite the shuffling effects of recombination. Although the value estimated for D_F from the covariance between spot score and log femur ratio in the central class is

higher than those estimated from other combinations of the traits (between which there is a notable correspondence), it does not lie outside any of the support limits. There is therefore no evidence of any difference in the patterns observed in any of the trait combinations.
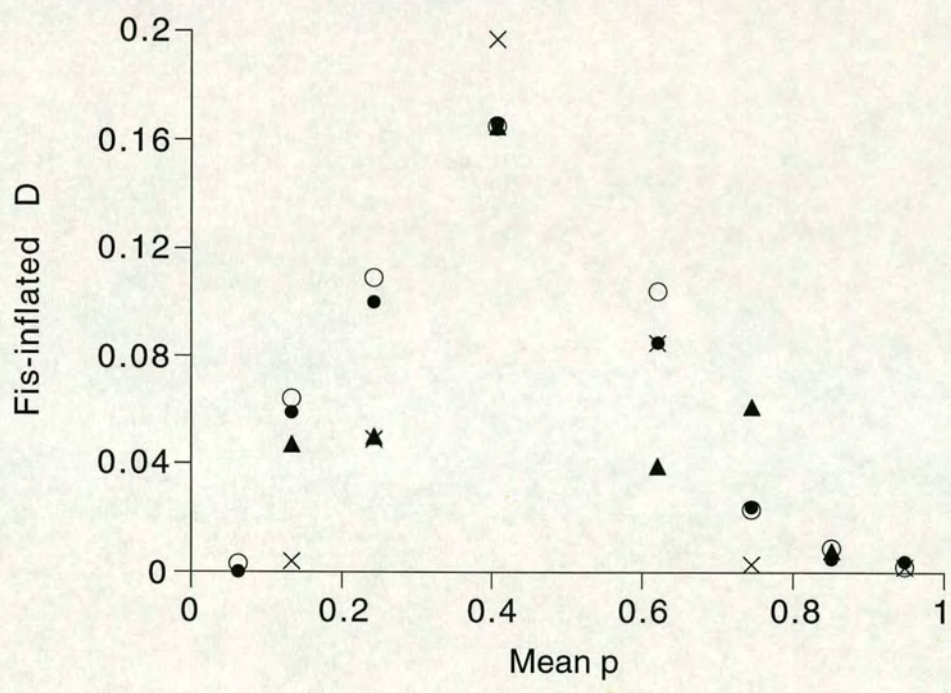
Figure 2.5.1 Fis-inflated linkage disequilibrium D_F calculated from four sources:
(i) variance in hybrid index [open circles];
(ii) covariance between hybrid index and spot score [closed circles];
(iii) covariance between hybrid index and log femur ratio [triangles];
(iv) covariance between spot score and log femur ratio [crosses].

Table 2.5.2 Fis-inflated linkage disequilibrium D_F calculated from four sources:

(i) variance in hybrid index (H);

(ii) covariance between hybrid index and spot score (H*S);

(iii) covariance between hybrid index and log femur ratio (H*F);

(iv) covariance between spot score and log femur ratio (S*F).

Mean p is the average of all the individuals in a class. N1 = sample sizes for (i)-(iii); N2 = sample sizes for (iv).

(Values in central class are given in bold for ease of comparison.)

Mean p	N1	(i) D_F from var. H	(ii) D_F from covar. H * S	(iii) D_F from covar. H * F	N2	(iv) D_F from covar. S * F
0.063	362	0.003	0.000 (-0.004, 0.004)	-0.002 (-0.017, 0.013)	606	-0.002 (-0.030, 0.027)
0.133	123	0.064	0.059 (0.039, 0.082)	0.047 (-0.002, 0.098)	288	0.004 (-0.041, 0.049)
0.243	160	0.109	0.100 (0.071, 0.133)	0.050 (-0.020, 0.123)	213	0.049 (-0.026, 0.127)
0.407	73	**0.165**	**0.166** (0.106, 0.241)	**0.165** (0.074, 0.273)	131	**0.197** (0.105, 0.302)
0.621	120	0.104	0.085 (0.054, 0.121)	0.039 (-0.031, 0.112)	158	0.085 (0.005, 0.169)
0.745	138	0.023	0.024 (0.007, 0.042)	0.061 (0.013, 0.113)	243	0.003 (-0.050, 0.057)
0.851	136	0.009	0.005 (-0.006, 0.016)	0.008 (-0.019, 0.035)	263	-0.001 (-0.040, 0.038)
0.948	115	0.002	0.004 (-0.002, 0.011)	-0.003 (-0.018, 0.013)	241	0.002 (-0.033, 0.038)

2.6 Discussion

2.6 (i) Concordance and coincidence

The results presented here show no evidence of differences in the width or position of any of the clines in adult traits measured (I discuss the egg size cline further below). This confirms that even if there is a differential gradient in selection coefficients acting on larval traits (as suggested by the displaced clines shown in the laboratory breeding experiment (Nürnberger *et al.* 1995); see also Chapters 3 and 4), adult characteristics remain closely associated. Despite the high resolution afforded by the large sample sizes, there is therefore no indication that the staggered nature of the change in environmental factors at Pešćenica is affecting the relative position of the clines in the adult traits considered here. The close coincidence might suggest that the dynamics of the hybrid zone are primarily determined by selection against hybrids (see section 1.3). However the strength of selection required to generate displaced clines in response to staggered environmental variables might be relatively large, given that the high values of linkage disequilibrium observed in the centre of the zone will serve to pull all clines closer together. This is a subject that requires theoretical investigation.

The equivalent width of clines in all four traits is also remarkable, given that each presumably conveys different adaptive significance: a naive expectation would be that allozyme frequency would have least effect on fitness, with spot score, leg length and egg size under increasingly strong (potential) selection pressures. There are theoretical arguments which suggest that changes in the strength of stabilising selection on quantitative traits (towards a spatially varying optimum, between the extremes of very weak or very strong values) may have surprisingly small effects on cline width (Nürnberger *et al.* 1995), although it is not clear how this result fits with known examples of variation in cline width (e.g. *Chorthippus*; Hewitt 1993a).

In conclusion, the null hypothesis that should be accepted for these results is that none of the adult traits considered are under sufficiently different selection pressures to generate differences in their cline shape.

2.6 (ii) Difference between clines in male and female traits

The comparison of clines in male and female traits reveals the surprising result that the transition from one type to another is wider in males than in females. The result is significant for both allozyme frequency and spot score. As any correspondence between sex and character value should be removed on reproduction, this suggests two possible explanations:

(1) Differential selection on adults after migration would generate the result, but seems unlikely given the strength of selection that would be required to create such an effect in one generation. A plausible selective mechanism is also not obvious but the possibility cannot be discounted. If males simply had greater longevity, this would equate to the second possibility:

(2) The same effect could be generated by high and differential dispersal rates, with males dispersing further than females. This seems more probable, and corresponds with the values of linkage disequilibrium observed in the centre of the zone being generally high.

I suggest that the lack of a comparable result for the femur ratio measurement is due to the considerably greater noise in its cline (Figure 2.3.2(b)) rather than evidence of differential selection on the respective traits.

2.6 (iii) Egg size

The cline in egg size does not show the displacement in the direction of *B. variegata* observed in the laboratory breeding experiment (Nürnberger *et al.* 1995). The discrepancy in the clines lies in the hybrid sites to the *B. bombina* side of the centre ($0.2 < \bar{p} < 0.4$): in the laboratory, females from these sites produced eggs no larger than

those from females from pure *B. bombina* sites, but in the field results presented here eggs from these sites were larger than those from the pure sites. The obvious explanation for this is differential expression of a phenotypically plastic trait, with the results from the breeding experiment illustrating the underlying genetic variation unconfounded by the environmental effects encountered in the field. Environmental effects were apparent in egg size being consistently larger in the field than in the laboratory (but see Chapter 3 for consideration of the effects of hormone injection). Greater variation in egg size in the field relative to that observed in the laboratory (B. Nürnberger, pers. comm.) also suggests the existence of considerable environmental variance which has been suppressed in the laboratory. However, two other points need to be considered in relation to this issue. Firstly, the laboratory result is itself surprising in the egg size cline being displaced in a different direction from that of developmental time, despite the usual correlation observed in amphibians between the two characters (Kaplan & Cooper 1984; Petranka *et al.* 1987b; Rafinska 1991; Travis *et al.* 1987). Secondly, there is a notable absence of any obvious environmental difference between the sites in question that would generate lower temperatures (and hence encourage larger eggs) in the hybrid sites relative to the pure sites, although a human perspective on conditions in a puddle is undoubtedly different from a toad's.

2.6 (iv) *Linkage disequilibrium*

The values for (F_{is}-inflated) linkage disequilibrium calculated from covariances between quantitative traits are not greatly different from the values calculated from the variance in the hybrid index. This suggests that the discrepancies observed between these measures in a previous study (Nürnberger *et al.* 1995) were due to smaller sample sizes. In addition, if the data presented here are analysed in the identical manner, using three rather than eight classes, values for the central hybrid class are considerably lower (data not shown).

Hybrid zones are frequently characterised by high values of linkage disequilibrium: see, in addition to *Bombina*, the studies of *Heliconius* butterflies (Mallet *et al.* 1990), of the leopard frog *Rana pipiens* (Kocher & Sage 1986), of the grasshopper *Caledia captiva* (Shaw *et al.* 1993) and of *Gryllus* crickets (Rand & Harrison 1989). As outlined in section 2.5(i), the most obvious explanation is the mixing of populations of individuals carrying different sets of alleles. The magnitude of the linkage disequilibrium will be determined by the extent of the mixing, or the rates of dispersal shown by individuals. High values such as those observed here imply considerable migration; this is confirmed by direct observations of movement from recaptures of marked *Bombina*[2].

2.6 (v) Methodological relevance

Natural hybridisation events were initially identified from morphological traits, using the assumption that hybrids are intermediate between parental species. The recent wide-spread availability of genetic markers has facilitated more accurate classifications from genetic data: for example, in a comparison of morphological and genetic traits of hybrid *Hyla* (treefrogs), Lamb and Avise (1987) found that use of morphological data alone would have misclassified more than 40% of known hybrids as being of one or other parental species. However, analysis of geographic variation in many populations is notably different from the task of classifying individuals. The results presented here, considering means across populations, suggest that analysis of spatial variation can be as powerful when (more easily scored) phenotypic traits are considered as genetic data. However the calculation of linkage disequilibrium used here relies on assumptions such as the absence of pleiotropy or independent environmental effects[3]. Hybrid index, spot score and the log femur ratio could be expected to satisfy these requirements, but the method should only be applied with caution.

---

[2]*Bombina* belly patterns are unique to each individual; recaptures can be recognised by toe-clips, and identified by their spotting pattern. Recapture data are not presented here.

[3] The assumption of additivity is less restrictive, as the relationship between D and covariances can be shown to hold approximately for alternative gene actions (Nürnberger *et al* 1995).

Chapter 3

Evidence for hybrid dysfunction

3.1 Introduction

The clines described in Chapter 2 are maintained by natural selection counterbalancing the homogenising effects of dispersal. This natural selection must be acting at one or more of the different stages of the amphibian life cycle: egg, larvae or adult. It may also act in different ways, such as against hybrids or through differential adaptation across an environmental gradient. In this chapter I consider the first of these possibilities, and describe tests for evidence of selection against hybrids at different stages of the life cycle.

3.1 (i) Importance of selection against hybrids

Hybrid fitness and its role in interactions between divergent (or diverging) populations are crucial issues both for studies of natural hybridisation and in the wider arena of speciation theory. The reasons for its importance are twofold: (1) if novel genetic combinations can result in hybrid superiority, hybridisation may lead to the founding of new evolutionary lineages; (2) if hybrid unfitness is the primary factor responsible for differentiation between populations, it is the rate of evolution of postzygotic rather than prezygotic isolation which will determine species diversity. Note that these two scenarios are not exclusive: some hybrid gene combinations could convey superior fitness despite reduced fitness in the majority maintaining reproductive isolation between parental populations. The second reason motivates many studies of hybrid zones, and is relevant to this study, but the first reason is a reminder of the potential for alternative consequences of hybridisation events (e.g. Charlesworth 1995; Rieseberg *et al.* 1995; Rieseberg & Wendel 1993); see also Grant (1981).

In a survey of nearly 150 case studies, Barton and Hewitt (1985) concluded that most hybrid zones are maintained by "some sort of hybrid unfitness". Frequent correlations with an environmental gradient can be explained by the fact that these "tension zones"

(Key 1968) will stabilise at regions with low population density or barriers to dispersal, but environmental factors are not significant in their actual maintenance. However, this conclusion rests heavily on the indirect argument that hybrid unfitness is the most parsimonious explanation for the frequently-observed coincidence of clines (see section 1.3). Direct evidence is not abundant (Harrison 1993): in the 28 well-studied hybrid zones listed by Harrison (1990), evidence for reduced fitness in "individuals of mixed ancestry" was available in only 11. Arnold and Hodges (1995) review data from 19 detailed studies which aim specifically to ascertain relative hybrid fitness, involving a range of both organisms and fitness measures. This survey contains examples of relative hybrid fitness being less than, equivalent to or even greater than that of parental taxa, and therefore it too does not reach a conclusion of uniform hybrid unfitness. As Harrison (1990) emphasises, a null result in a test for reduced hybrid fitness is difficult to interpret, particularly given either the artificial constraints of laboratory studies or the inherent difficulties of measuring fitness in the field; in addition, null results are less likely to be published. Thus there does not appear to be a general rule regarding hybrid fitness, and the debate continues over the relative importance of hybrid breakdown vs. environmental adaptations in preventing gene flow between two populations.

3.1 (ii) Hybrid dysfunction in Bombina

Molecular information dates divergence between *B. bombina* and *B. variegata* at an estimated 2-7 Myr ago (Szymura 1993). Incompatibility between genomes which have been diverging for so long, and which produce markedly different phenotypes, might not be unexpected. This is confirmed by data collected on Polish populations revealing hybrid dysfunction in the form of reduced embryonic viability (Koteja 1984), abnormal tadpole mouthparts (Czaja 1980) and skeletal malformations (Madej 1965).

The data on embryonic viability (Koteja 1984) are the most frequently cited evidence for hybrid dysfunction in *Bombina* (e.g. Szymura and Barton 1986, 1991). Figure 3.1.1(a) is after Szymura and Barton (1986), and shows increased mortality in the periods between egg fertilisation and gastrulation, and between fertilisation and independent feeding in central populations. In Figures 3.1.1(b) and (c), I present the same data, but as individual points for two disjunct time periods: between fertilisation and gastrulation, and between gastrulation and independent feeding (note the time periods in Figure 3.1.1(a) are overlapping). Although the two sites with entirely pure allozyme frequencies ($\bar{p}=0$ or $\bar{p}=1$) unarguably had the lowest failure rates, large increases in the mean towards the centre of the zone are primarily due to the two data points. Secondly, there is no evidence of increased mortality during the second period.

Bombina is represented in Arnold and Hodge's (1995) survey by Nürnberger *et al.*'s (1995) large scale laboratory breeding experiment using animals from Croatian populations. Offspring from crosses between individuals from sites in the centre of the hybrid zone showed equivalent viability to those from crosses within pure populations, although F1 offspring showed reduced survival rates. A difference between Poland and Croatia in the relative genetic incompatibility of the hybridising populations is not implausible, as the two transects involve different subgroups (though not different subspecies: section 1.6) of *B. variegata*, and stronger selection in Poland would fit with the narrower cline width (Szymura 1993) than that observed in Croatia (MacCallum 1994). It nevertheless seems unlikely that hybrids from one place would not show reduced fitness if hybrid dysfunction is supposedly strong enough to maintain a stable zone at another point. Data from the breeding experiment (Nürnberger *et al.* 1995) are potentially confounded by laboratory-genotype interactions: *B. variegata* consistently fared better than *B. bombina* in the laboratory conditions in Edinburgh (B. Nürnberger, pers. comm.). Such effects may have introduced too much noise to detect underlying differences in fitness.
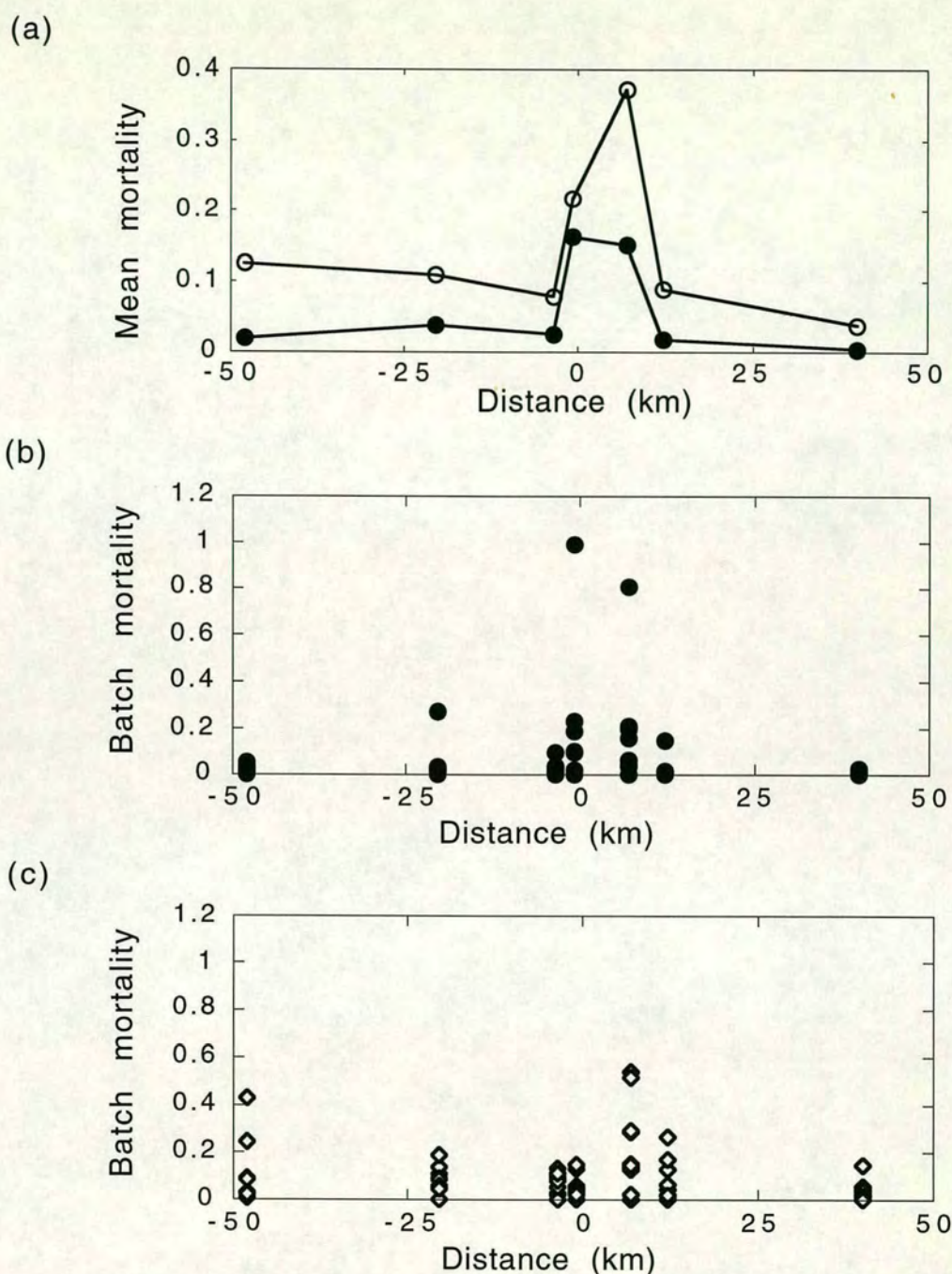
Figure 3.1.1 Mortality in crosses from transect around Kopanka, Poland; data from Koteja (1984). **(a)** Population mean mortality: between fertilisation and gastrulation (closed circles), and between fertilisation and time of independent feeding (open circles), after Szymura & Barton (1986); **(b)** Mortality between fertilisation and gastrulation, for each batch; **(c)** Mortality between gastrulation and independent feeding for each batch. X-axis is distance in km across Cracow transect, from Table 1 in Szymura and Barton (1991). *Bombina bombina* populations are located to the left of the scale.

The degree of hybrid dysfunction in *Bombina* is therefore uncertain, particularly for the southern populations, and the majority of evidence to date has stemmed from laboratory crosses. I consider here two alternative approaches to fitness measurement, one measuring performance of naturally-produced offspring under laboratory conditions, the other tracking performance entirely in the field. Fitness may vary between hybrid classes, and the detailed knowledge of the Pešćenica transect allows high resolution in relating fitness to hybrid class.

3.1 (iii) Embryonic mortality as fitness measure

Even within intraspecific crosses in amphibians, interactions between maternal and zygote genomes have been shown to contribute to a large proportion of embryonic mortality (Gurdon 1977). Dissecting the quantitative genetic components of variation in fitness traits in *Hyla crucifer*, Travis *et al.* (1987) found nearly all embryonic mortality occurred between gastrulation and neuralation, at a stage of "extensive interaction between the newly unlocked nuclear genome and cytoplasmically borne elements that initiate and regulate gene transcription in the nucleus". These effects will presumably be magnified by the degree of divergence between the parental genomes. Embryonic mortality is therefore an obvious candidate fitness measure. The *Bombina* studies described above measured it, but both involved crosses where ovulation was hormone-induced. This will presumably introduce considerable noise into measures of embryonic viability, as injected females will ovulate even if their ova are not fully yolked. The use of eggs laid naturally in breeding sites would circumvent this problem: mean hatching success in pure populations (defined as those with mean allozyme frequency less than 0.2 or greater than 0.8) was 0.486 in the breeding experiment on Croatian animals (Nürnberger *et al.* 1995) compared to 0.991 for naturally-laid eggs (this study). Subsequent development from the latter can then be followed for further comparisons.

3.1 (iv) Adult survival as fitness measure

There have been no estimates of hybrid fitness in *Bombina* under natural conditions. Any selective disadvantage (or advantage (Moore 1977)) of hybrids will ultimately only be relevant in the context of specific environments, for example in their response to predators, competitors or pathogens (Harrison 1990). Comparing relative viabilities in the field requires a cohort analysis, in which the same population is sampled more than once during the course of the life cycle and the relative proportions of different genotypes compared. Such studies are most easily conducted either on univoltine organisms (e.g. Howard *et al.* 1993: *Allonemobius* crickets) or stages of the life cycle in which individual age is easily recognised (e.g. Kocher and Sage, 1993: tadpole to juvenile stage in *Rana* leopard frogs). However, they are also feasible in taxa where the age structure of the population can be inferred despite overlapping generations (e.g. Bert & Arnold 1995: *Mercenaria* fresh-water clams). The latter approach is possible with *Bombina*.

Data have been collected from the Pešćenica transect over four field seasons: 1991, 1992, 1994 and 1995, and adult *Bombina* are known to live for several years (pers. obs.; J. Szymura, pers. comm.). Amphibian body length usually increases with age (Duellman & Trueb 1985; Zug 1993), providing an approximate indication of the relative age distribution in a population and hence allowing a cohort of adults to be identified (see section 3.4). Any differential mortality should be reflected by changes in the composition of the cohort. Various parameters are estimated, both to maximise the chances of detecting an effect and as a comparison of the power of alternative approaches.

The above discussion has been confined to the possibilities of selection against hybrids. Adult survival will also be affected by environmentally-determined fitness values (see section 3.6 for discussion of the problems, sometimes unappreciated, this may cause a

cohort analysis). However, as outlined in section 3.4, the two forms of selection will have different effects on genotype distributions, and hence may be distinguished.

3.1 (v) Aims

A tension zone model requires reduced hybrid fitness, for which evidence in the *Bombina* hybrid zone is equivocal. Specific predictions can be made (and tested) from this requirement, with respect to the following parameters:

(i) embryonic mortality: eggs from hybrid populations should suffer higher mortality;

(ii) larval development: larvae from hybrid populations should show increased frequency of developmental abnormalities;

(iii) adult survival: the proportion of hybrid individuals in central populations should decrease over time, resulting in a corresponding increase in the magnitude of linkage disequilibrium.

3.2 Embryonic viability & larval development: Methods

3.2 (i) Egg collection

Naturally-laid eggs were collected from the field on a total of 27 sampling days between 28 April and 3 June 1995. Methods and locations of egg collection are described in section 2.2(vi); allele frequencies in adult *Bombina* at four diagnostic enzyme loci are known for each of the 19 sites sampled. For each batch, Gosner developmental stage (Gosner 1960) was determined using a dissecting microscope, and any non-fertilised eggs removed; whenever possible, 12 eggs from each batch were considered, but some batches contained smaller numbers.

3.2 (ii) Rearing scheme

The 12 eggs from each batch were divided into groups of 3; these groups were placed in plastic tumblers, in which the bottom consisted of nylon mesh. The tumblers were placed in groups of nine in 6-litre plastic boxes filled with non-chlorinated water, such that each contained approximately 250 ml. This system was intended to minimise the disturbance caused by water changes.

Larvae were then reared in the same tumblers on hatching, with dead eggs or tadpoles being removed. Tadpoles were fed daily on boiled, powdered nettle leaves. The position of the plastic boxes was rotated frequently to avoiding confounding effects of temperature variation within the room. Whilst individual rearing tumblers would clearly have been preferable, space was a limiting factor in the field "laboratory", necessitating the regime of 3 per cup. In total, 1821 tadpoles from 167 batches were reared.

3.2 (iii) Measuring viability

Embryonic mortality was described by hatching failure: the proportion of the total number of eggs grown which failed to reach hatching.

The rearing experiment was ended on 24 June 1995, three weeks after the last batch of eggs had been collected. All comparisons are therefore made on observations from the first three weeks of development of each batch. Within this period, the frequency of apparently abnormal individuals was recorded: developmental abnormalities included kinked tails, bloated abdominal regions, growth of body but not of head region and (in one case) an inexplicable failure to grow coupled with complete lack of pigment. Most abnormalities resulted in death, either during the three weeks' monitoring period or subsequently.

Rearing 3 tadpoles per tumbler creates competition between individuals, and therefore limits the parameters available for valid comparisons between genotypes: any effect of genotype on survivorship and growth would be partially determined by a genotype-competition interaction. There is reason to expect *B. variegata* tadpoles to be stronger competitors and better adapted to high densities than *B. bombina* (see section 4.4), but such factors deserve their own full-scale investigation. I present data on survivorship, but these caveats should be borne in mind; growth rates are not considered in this study. (Note that Rafinska (1991) compares growth and development rates of *B. bombina* and *B. variegata* larvae reared until metamorphosis in batches of 10 per 2.5-litre box, inevitably introducing an effect of competition.)

3.2 (iv) Statistics

As detailed in the methods of Chapter 2, egg sampling was designed to minimise the probability of sampling an individual toad's clutch more than once. This was still a possibility; to avoid any bias in statistics or pseudo-replication, analyses of hatching failure rate, frequency of larval abnormalities and larval mortality to 3 weeks are mainly conducted on population means rather than values for individual clutches. As the number of sites sampled is not sufficient to detect any asymmetry across the zone, population means are related to their distance from the closest edge, using a "folded" index p' defined as:

$$p' = \begin{cases} \overline{p} & if\ \overline{p} \leq 0.5 \\ 1 - \overline{p} & if\ \overline{p} > 0.5 \end{cases}$$

Proportional data (for example, the proportion of eggs failing to hatch) are arc-sin transformed (Sokal & Rohlf 1981) before fitting a least-squares linear regression model; in all regression models, the data are weighted by the site sample size and the normality of residuals is checked using the Shapiro-Wilks test. Correlations between the three variables are tested for with Spearman's (non-parametric) coefficient of rank correlation (Sokal & Rohlf 1981): population means are subtracted from values for individual

clutches to remove the effects of between-site variation. All statistical analysis was carried out using JMP Version 3.1.5.

3.3 Embryonic viability & larval development: Results

3.3 (i) Embryonic mortality

Overall hatching success was high, with 96.7% of all eggs hatching, and no hatching failure in 82% of the 167 egg batches monitored. Figure 3.3.1(a) shows the mean hatching failure rate for each site plotted against the folded index p'. The slope of a weighted linear regression of (arc-sin transformed) mean failure rate on p' is significantly greater than zero: $ArcSin\left(\sqrt{Failure}\right) = 0.006 + 0.582p'$; $F_{1,17}$=22.557; p<0.001. (Note that the slope is significant even after removal of the batches from 5011, the high outlier: $ArcSin\left(\sqrt{Failure}\right) = 0.022 + 0.444p'$; $F_{1,16}$=19.268; p<0.001). Hatching failure rate therefore increases significantly towards the centre of the zone. Individual failure rates for each clutch are given in Figure 3.3.1(b).

(a)
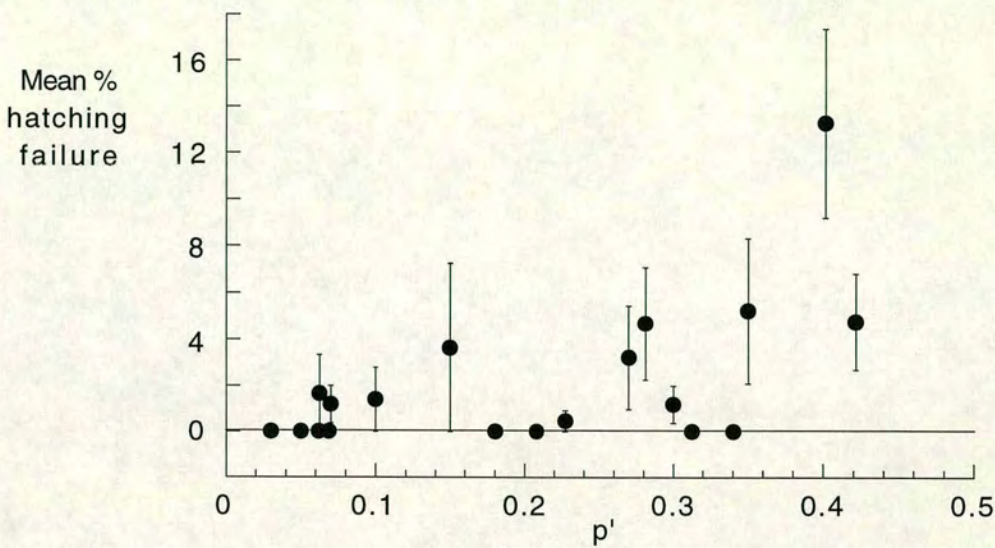
(b)
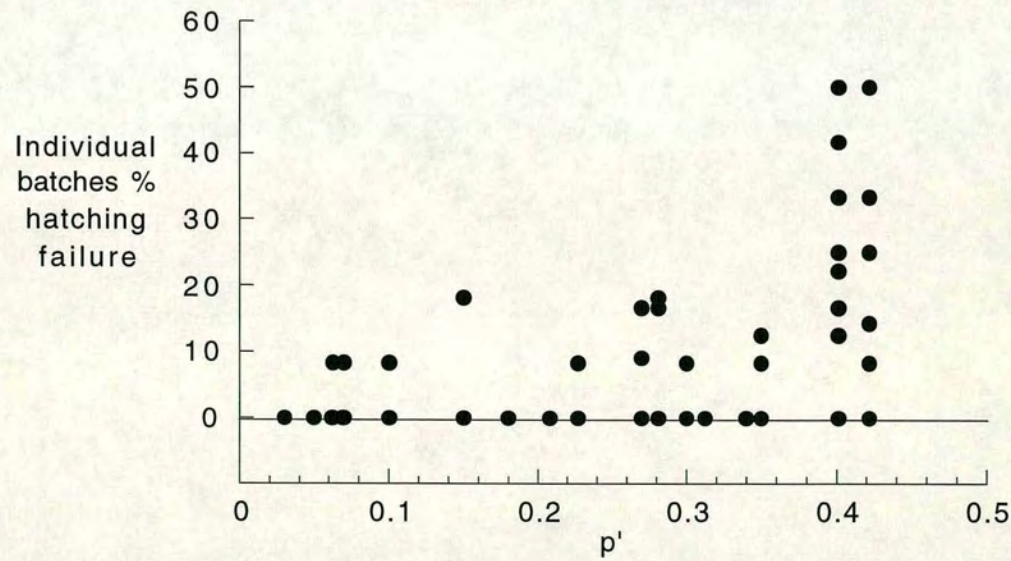
Figure 3.3.1. (**a**) Mean hatching failure rate for each site; bars represent one standard error of the mean. (**b**) Individual hatching failure rate for each clutch. (Folded index $p' = \bar{p}$ if $\bar{p} \leq 0.5$, $(1-\bar{p})$ if $\bar{p} > 0.5$.)

3.3 (ii) Larval developmental abnormalities

Larval development can be considered similarly. The frequency of developmental problems was low, with only 41 out of 1821 tadpoles reared appearing abnormal. Figure 3.3.2 shows the mean proportion per site of abnormalities in batches; again, the highest value is at site 5011. As with embryonic mortality, the slope of a linear regression of mean proportion of abnormalities against p' is significantly positive: $ArcSin\left(\sqrt{PropAbnormalities}\right) = 0.052 + 0.322p'$; $F_{1,17}=5.629$; p=0.0297. However this significance is dependent on site 5011, without which the regression becomes: $ArcSin\left(\sqrt{PropAbnormalities}\right) = 0.071 + 0.153p'$; $F_{1,16}=2.290$; p=0.150.

The unusually high frequency of abnormal tadpoles in 5011 might be due to a single parental pair producing a batch (or batches) with extremely low fitness; the mean might therefore be biased upwards by repeated sampling of these batches. Figure 3.3.3 shows the proportion of abnormalities for individual batches by collection date; although one batch has an exceptionally high frequency of abnormalities (only 7 of 12 eggs hatched, of which 6 were visibly abnormal; the 7th grew very little in 3 weeks) there is no evidence of problematic batches being confined to one sampling date, nor to clusters of dates. We do not have strict evidence on laying patterns and mating behaviour in *Bombina*, so the possibility that the batches are from the same parental pair breeding repeatedly across the sampling period cannot be entirely excluded. However, detailed recapture studies have shown a rapid and continual turnover of animals present at a breeding site (MacCallum *et al.* 1997; pers. obs.), suggesting that the above possibility is unlikely. This implies frequently low fitness in offspring from the site.
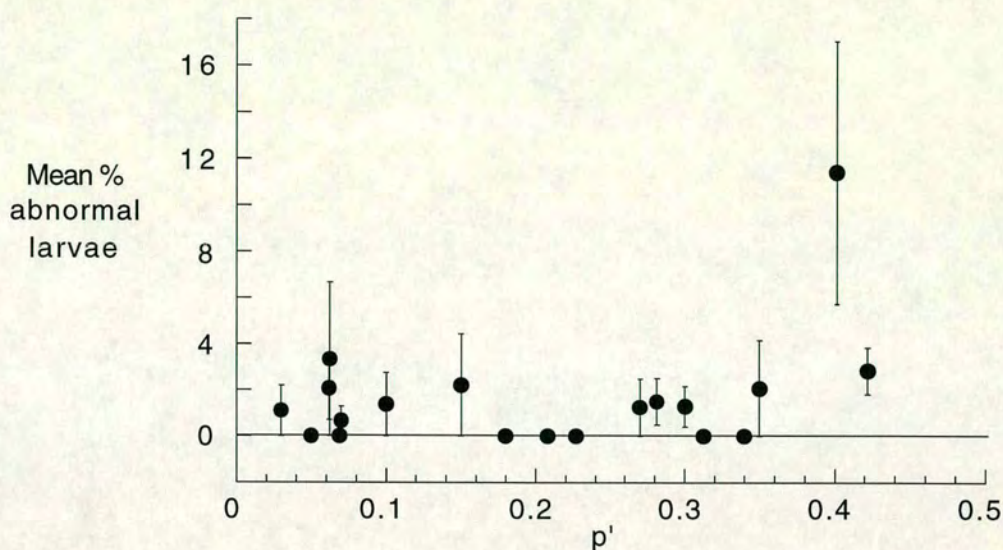
Figure 3.3.2. Mean proportion of abnormal larvae per batch for each site; bars represent one standard error of the mean. (Folded index $p'=\bar{p}$ if $\bar{p} \leq 0.5$, $(1-\bar{p})$ if $\bar{p} > 0.5$.)
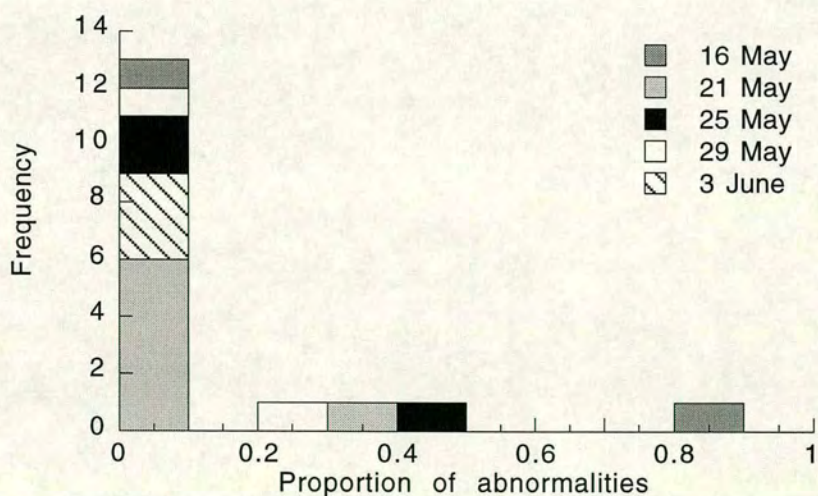
Figure 3.3.3. Proportion of abnormal individuals in batches sampled from site 5011 on separate sampling dates.

3.3 (iii) Larval survival rates

I present here data on survival rates to three weeks, although the results should be interpreted with the caution given the caveats mentioned in section 3.2. Once again, the slope of a (weighted) linear regression of arc-sin mean mortality per batch against p' is significantly positive: $ArcSin\left(\sqrt{PropMortality}\right) = 0.058 + 0.772p'$; $F_{1,17}=20.722$; $p<0.001$. Data are presented in Figure 3.3.4; the significance of the relationship is clearly not dependent on one data point, as in the previous section. In total, 1688 larvae survived from the 1821 hatchlings, giving a total survival rate of 92.7%, compared to 73.0% from the laboratory breeding experiment (Nürnberger *et al.* 1995; excluding F1 crosses).
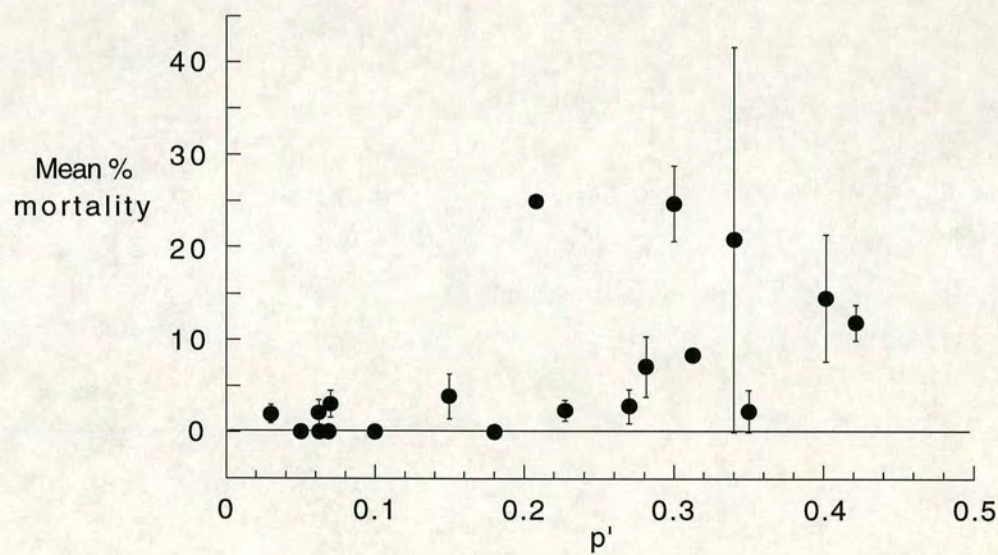
Figure 3.3.4 Mean % mortality across batches collected at each site; bars represent one standard error of the mean. (Folded index p'=$\overline{p}$ if $\overline{p}$ ≤0.5 and 1-$\overline{p}$ if $\overline{p}$ >0.5.)

3.3 (iv) Correlations between fitness measures

The above three variables (hatching failure, frequency of larval abnormalities and larval mortality) are clearly not unrelated. Spearman's coefficient of rank correlation, τ, is used to test for interdependence between the values observed for each batch (corrected for population means). All three pairwise combinations are significant; see Table 3.3.1.

Table 3.3.1 Spearman's coefficient of rank correlation τ and corresponding p-value for pairwise correlations between three fitness traits.

	Hatching failure	Abnormality frequency
Abnormality frequency	$\tau = 0.282,\ p<0.001$	
Mortality	$\tau = 0.214,\ p=0.005$	$\tau = 0.322,\ p<0.001$

3.4 Adult cohort analysis: Methods

3.4 (i) Defining cohorts

Sampling of adult toads at breeding sites is described in Chapter 2; allozyme frequency, spot score, femur length and snout-vent (trunk) length were determined by the methods outlined in the same section. The last of these, trunk length, increases with age, and so can be used as an approximate indicator of the distribution of age classes. Figure 3.4.1 shows the distribution of trunk lengths among all toads measured in each field season.

Toads first measured in 1991 show a unimodal distribution of sizes (Figure 3.4.1); I define these animals as one cohort. Recaptures have shown that *Bombina* survive several years in the field (pers. obs. and J. Szymura, pers. comm.), and the distributions from the subsequent years show a steady increase in the size of individuals measured; I therefore make the assumption that the respective peaks represent the same cohort. The distributions also reveal the advent (tentatively in 1994, firmly established in 1995) of a second cohort; samples were taken from breeding sites, so the distribution fits the observation of a highly successful breeding season in 1991 (C. MacCallum, pers. comm.) and juveniles not appearing at breeding sites during their first years. I consider here the first cohort, and test for evidence of differential survival between these individuals. The distributions are not discrete, so cut-off values must be decided upon to define the respective cohorts; the resulting classification will inevitably be approximate. Thus in the following analysis I define the first cohort as: all 1991 individuals; 1992 individuals with trunk length greater than 32 mm; 1994 individuals with trunk length greater than 39 mm; 1995 individuals with trunk length greater than 41 mm. There was no indication of correlation between genotype and trunk length for 1991 individuals, so defining a cohort by size will not have any biasing effect.
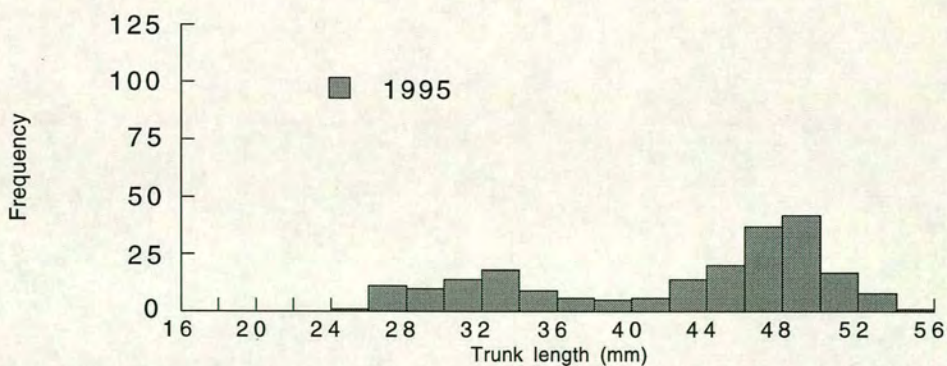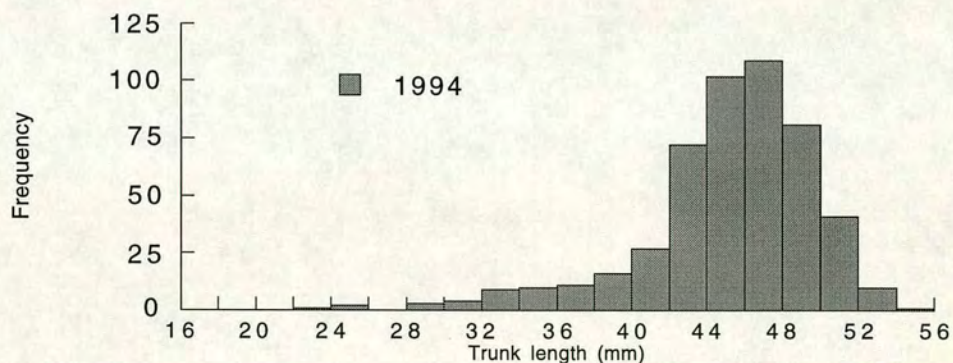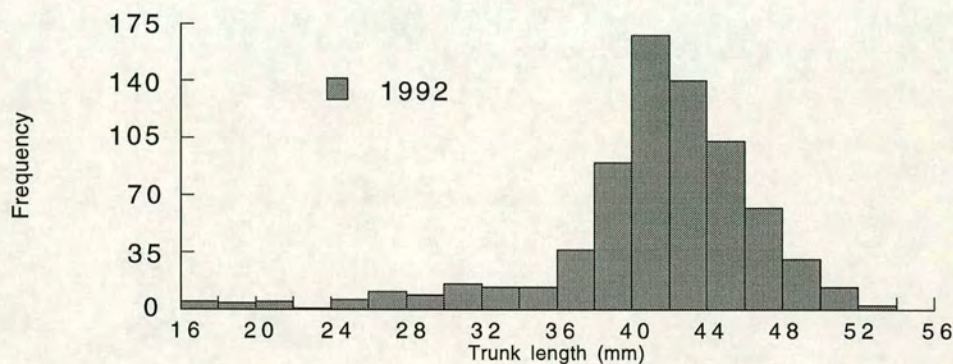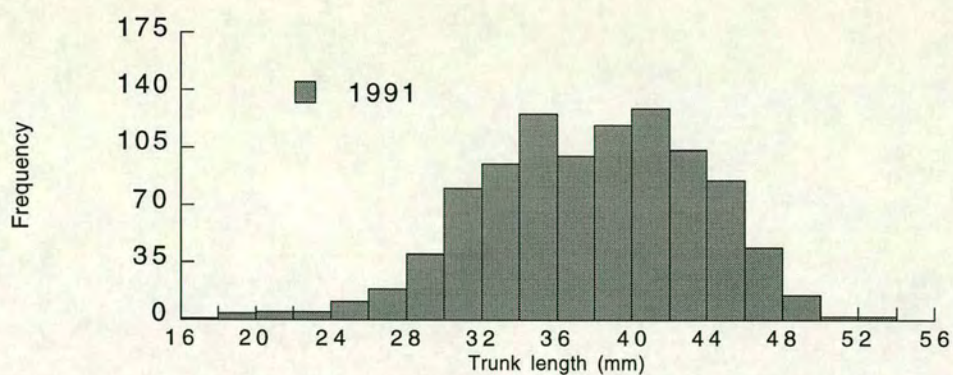
Figure 3.4.1 Distribution of trunk lengths across sampling years. Note different scales on y-axis between years.

Comparisons of survival rates between genotypes are not valid if confounded by geographic or habitat differences: for example, higher mortality in populations at the centre of the hybrid zone relative to those at the edges could be due to a lower carrying capacity of the environment rather than any genetic factors. I therefore consider hybrid populations in a single central region of forest (see Figure 3.4.2) covering an area of approximately 2 km x 4 km immediately to the *B. bombina* side of the centre of the zone. The mean frequency of *B. variegata* alleles observed in sites within this region ranges from 0.00 to 0.64, with an average value of 0.30. Figure 3.4.3 shows the distribution of (a) individual genotypes (b) individual spot scores for all toads sampled in the region in 1991, illustrating the range of character types present.

Even within this central region of the hybrid zone, there is an association between type of breeding habitat and the genotype of the toads found there, with more *B. variegata*-like hybrids being found in puddles (MacCallum *et al.* 1997; see Chapter 4). For this reason, samples from each habitat had to be analysed separately, as the proportion of either habitat type sampled changed between years: for example, a disproportionate number of individuals sampled from ponds in a given year would, by the nature of the habitat preference, bias the overall gene frequency for that year. As the region contained only two ponds, analysis was confined to puddles alone. The sites used are listed in Appendix 3.1; these were visited frequently in all four field seasons. This results in the following sample sizes, for the number of individuals assigned to the first cohort and sampled within the specified region: 1991, N=71; 1992, N=43; 1994, N=24; 1995, N=45.
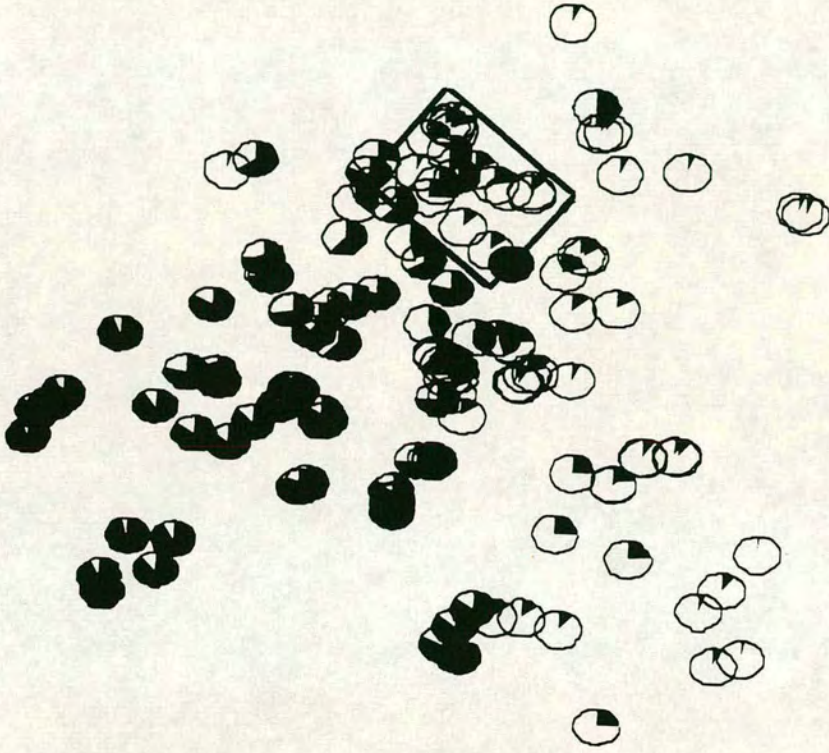
Figure 3.4.2 Map of allozyme frequency across the hybrid zone; each pie shows the mean proportion of *B. variegata* alleles (in black) in individuals sampled at that site. The box shows the sites used for the cohort analysis (compare with Figure 2.2.1, p37).
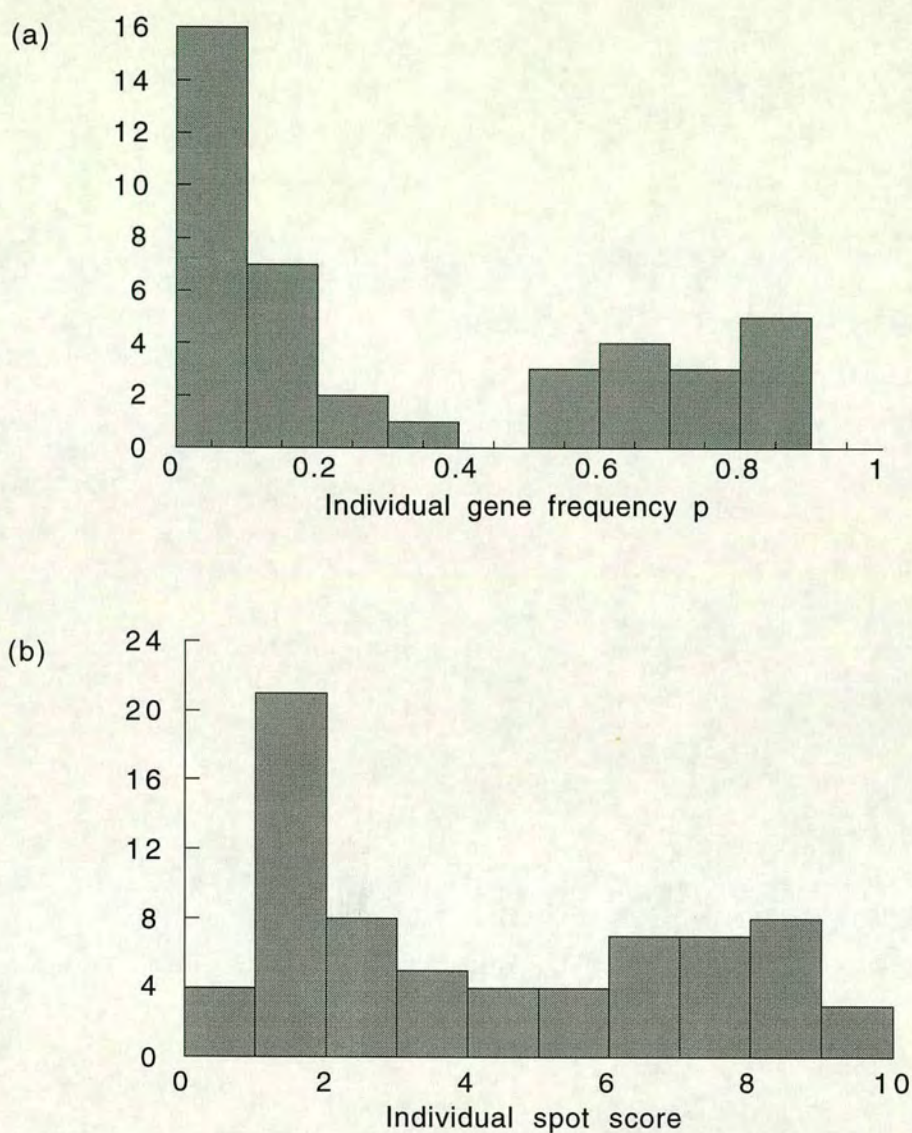
Figure 3.4.3. Distribution of **(a)** individual gene frequencies at four allozyme loci; **(b)** spot scores of toads found in 1991 in the area of forest shown in Figure 3.4.2.

3.4 (ii) Statistics

(a) Genotype distributions 1991-2

Suitable sample sizes for analysis of genetic data were only available for the 1991 and 1992 collections (1991 N=44; 1992 N=38 for the adult cohort; MacCallum 1994). Between these two years, I compare: (i) mean frequency of *B. variegata* alleles; (ii) maximum likelihood estimates of heterozygote deficits F_{is}; (iii) maximum likelihood estimates of linkage disequilibrium. All maximum likelihood estimates are made using the software package Analyse (Barton & Baird 1997). In addition, genotypic data can be combined with phenotypic information on spot score as in Chapter 2, to give higher resolution in determining individual ancestry. I define individuals as being "pure" *B. bombina* if they have no *B. variegata* allozyme alleles and a spot score of less than 2, and as "hybrid" otherwise (because all sites are to the *B. bombina* side of the centre of the hybrid zone, the chances of an individual being of pure *B. variegata* ancestry are minimal, whatever its spot score or p value). The proportion of hybrids in the cohort can then be compared between years. Fis-inflated linkage disequilibrium D_F is calculated from the covariance between hybrid index and either spot score or femur ratio (see section 2.5); this is also compared between years. Support limits for covariances are generated from the Wishart distribution for a variance-covariance matrix (using the Multivariate Statistics routine of Mathematica Version 3.0 (Wolfram 1996)).

The relatively small sample sizes for analyses involving genetic data do not permit correction for any spatial variation within the region. Samples are therefore treated as being from one population, which necessarily reduces resolution; it is unlikely to bias results as there was no consistent difference in sampling locations between years. In addition, frequent turnover of site availability (through dessication and then creation of new puddles) necessitated high levels of migration, as apparent from individual recaptures.

(b) Phenotype distributions 1991-5

Individual spot scores and femur lengths were recorded for individuals caught in all years, so comparisons between all four sampling years can be performed using phenotypic data. I consider the mean spot score, and the F_{is}-inflated linkage disequilibrium (D_F) estimated firstly from the covariance between spot score and femur (see section 2.5) and secondly from an additional method which uses the variance in spot score (described in Appendix 3.2). Note that, as discussed in Chapter 2, with data on phenotypes alone there is no means of separating F_{is} from linkage disequilibrium D, but that D_F provides an aggregate measure of the total association within and between loci. The large sample sizes afford a more sophisticated analysis, in which spatial variation in population type (i.e. populations becoming more *B. bombina*-like towards the edge of the forest) can be taken into consideration: the linear model comparing mean spot score between years therefore incorporates site co-ordinates, and population means are subtracted from values before covariances are calculated.

Given the initial distribution of the character indices (Figure 3.4.3), predictions can be made for the effects of alternative selection regimes:

(i) Selection against hybrids would reduce the frequency of central classes. This would lower the mean slightly, given that the distribution is initially skewed in favour of *B. bombina*. More pronounced would be the effect on the variance and linkage disequilibrium, both of which would be increased.

(ii) The predictions for environmental selection are difficult because although the region is lowland (hence typical for *B. bombina* populations), the sites under consideration are temporary puddles within forest. *Bombina variegata,* with its adaptations to a more terrestrial lifestyle, should therefore be better adapted to such an environment. Selection for *B. variegata*-like genotypes would increase the mean. The change in both the variance

and linkage disequilibrium will depend on the magnitude of the effects involved, but will be considerably weaker than the change due to selection against hybrids.

The contrasting predictions for changes in the mean under the alternative selection regimes are therefore sufficient to distinguish them. Note also that while the alternative selection regimes are not exclusive, one predicts a strong effect on the mean, the other a strong effect on the variance. Observation of significant change in both variables would therefore suggest the action of both. Any observed changes can then be interpreted within the framework of these predictions.

3.5 Adult cohort analysis: Results

3.5 (i) Genotype distributions 1991-2

Table 3.5.1 contains comparisons of the mean gene frequency and maximum likelihood estimates of heterozygote deficit and linkage disequilibrium calculated from data on genotype distributions in 1991 and 1992. None of these show any evidence of significant changes between years. The average pairwise linkage disequilibrium observed in 1992 is at the upper bound of the support limits for the 1991 value, potentially indicating an increase. However, maximum likelihood estimates for average values of D are calculated by summing the log-likelihood contributed by each pairwise association, thereby (falsely) assuming independence of associations. This assumption of greater information than is actually contained in the data will result in support curves that are slightly too narrow (N. Barton, pers.comm.).

Table 3.5.1 also contains inferences from combining genotypic and phenotypic information. There is a significant decrease in the proportion of "hybrids" (defined as any

individual with a non-zero p value and spot score greater than or equal to 2). The F_{is}-inflated linkage disequilibrium, D_F, shows no evidence of change.

3.5 (ii) Phenotype distributions 1991-5

Using a linear model of Mean Spot=X+Y+YearEffect, where X and Y are spatial co-ordinates of sites, there is no evidence of a change in mean spot score across years: $F_{3,181}=1.316$; p=0.271. Table 3.5.2 contains comparisons of the mean spot score and F_{is}-inflated linkage disequilibrium calculated from data on phenotype distributions in all four field seasons. None of these show any evidence of change between years.

Table 3.5.1 Comparison of adult cohort in 1991 and 1992: genetic parameters. Maximum likelihood (ML) values for heterozygote deficit (F_{is}) and linkage disequilibrium (D) are averages across all loci; support limits are given in brackets. F_{is}-inflated linkage disequilibrium D_F is calculated from covariances between characters named. See Methods for details of statistical calculations. [a] Normal approximation to Wilcoxon rank sums test. [b] Chi-squared test, df=1.

Statistic	1991 (N=44)	1992 (N=38)	Test of difference
(1) Genetic data			
Mean p (± s.err.)	0.368 (± 0.056)	0.324 (± 0.052)	Z=-0.668, p=0.504 [a]
ML F_{is}	0.293 (0.105, 0.472)	0.205 (0.028, 0.3817)	
ML D	0.109 (0.080, 0.115)	0.115 (0.103, 0.117)	
(2) Genetic and phenotypic data			
% "hybrids"	81.82	60.53	$\chi^2_{(1)}=4.613$, **p=0.032** [b]
Enzyme*spot D_F	0.185 (-0.112, 0.512)	0.186 (-0.133, 0.540)	
Enzyme*femur D_F	0.231 (-0.064, 0.564)	0.234 (-0.083, 0.595)	

Table 3.5.2 Comparison of phenotypic parameters for adult cohort from 1991 to 1995: sample size N, means and standard deviations (sd). S*S D_F and S*F D_F are the F_{is}-inflated linkage disequilibrium calculated using spot*spot covariance and spot*log femur ratio covariance respectively. Support limits for D_F values are given in brackets. See text (section 3.4 (ii)) for details of statistical calculations.

	1991	1992	1994	1995
N	71	43	24	45
Mean (sd) spot	3.91 (2.88)	3.31 (2.64)	4.52 (2.33)	4.03 (2.61)
S*S D_F (support limits)	0.198 (-0.120,0.364)	0.206 (-0.182,0.444)	0.167 (-0.311,0.534)	0.183 (-0.190,0.421)
S*F D_F (support limits)	0.191 (-0.152.0.331)	0.203 (-0.215,0.410)	0.114 (-0.364,0.480)	0.113 (-0.250,0.358)

3.6 Discussion

3.6 (i) Summary of results

The rearing experiment showed significant increases in mortality at both egg and larval stage, and in the incidence of developmental abnormalities, in samples from hybrid populations. The results are not independent, as all three variables were significantly correlated; this implies the existence of fundamental genomic problems in hybrid individuals which may manifest themselves in effects on either embryonic viability, developmental stability, larval survival, or other fitness components.

Despite considering a variety of measures, there was almost no evidence of differential survival in the adult cohort. The proportion of "pure" *B. bombina* individuals showed a significant increase between 1991 and 1992; the lack of a corresponding reduction in the mean suggests a compensatory change at the upper end of the distribution, namely

through an increase in the frequency of more *B. variegata*-like individuals relative to more intermediate genotypes. This could be interpreted as evidence of selection against intermediate genotypes. However it is the only significant result, and should be treated with caution given the multiple comparisons employed in the analysis; although support limits on linkage disequilibria calculated from covariances are consistently broad, there is not even an apparent trend with time for any of the phenotypic parameters.

3.6 (ii) Null result of the cohort analysis

Is the absence of any change within the cohort a reflection of negligible differential selection at the adult stage, or of inadequate resolution in the test statistics considered? Firstly, in the theoretical models outlined in Chapters 5 and 6, I make the explicit assumption that selection is too weak to have a noticeable effect on linkage disequilibrium. However, testing for such an effect is not in contradiction to this; the precise assumption of the models are that the effects of selection are weak *relative to* the effects of recombination, which will act in the opposite direction. Comparisons between years of a cohort of individuals does not, by definition, incorporate any effects of recombination. Nevertheless, if selection is acting, it is too weak to have been discerned with this data set. Genetic data on samples from all four years might provide the necessary resolution; Figure 3.4.3 underlines the superior discerning power of data on an individual's genotype compared to spot score alone.

Secondly, any results will unavoidably be confounded by the effects of migration into and out of the region. How strong this effect will be is not certain. The number of toads assigned to this first cohort declined markedly over years, despite comparable sampling efforts, suggesting that there is not continual influx from source populations elsewhere. In addition, the region of forest is bounded on the *B. bombina* side by the river Odra (see Figure 3.4.1), which will inevitably impede dispersal to some degree, and no sites were

found on the other side of the river. However, there are no environmental factors preventing migration from the *B. variegata* side. Dispersal is clearly an important factor in the *Bombina* hybrid zone (Chapter 2; MacCallum 1994), and the only unequivocal test for differential adult survival would require fencing. Data on individual recaptures within the region would provide an alternative unconfounded measure, but are not sufficiently numerous (only 22 individuals over the four field seasons).

3.6 (iii) Alternative fitness measures

Given the inherent problems of a cohort study, should fitness measures be restricted to more straightforward methods? Fitness measurement is open to a wide range of interpretations, of which it is only feasible to consider a subset in any one study. For example, within a laboratory rearing experiment such as this one, larval growth rate would be another obvious parameter for consideration (although, for the reasons outlined in section 3.2(iii), not under the experimental design employed here); other studies have shown that hybrid individuals may be most critically affected at time of metamorphosis (Woodruff 1979; Kocher & Sage 1986 (although see below); Nürnberger *et al.* 1995). Finally, the above arguments have all been concerned with measurement of viability, whereas interbreeding may have similar detrimental effects on fertility.

However, fitness is always conditional on environment. The most pertinent estimates of selective effects are therefore those derived from field data; they are also the most difficult to obtain. A test for differential adult survival would ideally consist of a transplant experiment in which fitness is compared in the different environments (e.g. Malhotra & Thorpe 1991). This is logistically unfeasible in the *Bombina* case, given the relatively low recapture rates of adults; the unavoidable perturbation of genotype distributions in the centre of the study site would also be problematic. Similarly, a pilot transplant experiment of tadpoles during the 1994 field season yielded negligible data due to heavy predation

rates (and loss of experimental units due to various environmental factors, namely dessication, flooding and wayward tractors). However, with a more robust experimental design and sufficient resolution of molecular markers, tracking survival in a tadpole cohort should be feasible, and would provide an elegant test of the relative importance of alternative forms of selection acting at the larval stage.

3.6 (iv) Other cohort studies

For comparison, I consider here three other cohort studies, all of which aimed to ascertain hybrid fitness. Despite highly suitable conditions for such analyses, demonstrating selection acting against hybrids proved difficult in all.

(a) Kocher & Sage 1986

In a study of the hybrid zone between two species of leopard frogs (*Rana berlandieri* and *R. utricularia*) in Texas, Kocher and Sage (1986) report differential mortality in a cohort sampled first as mid-stage larvae and then, three months later, as either final stage larvae or recently metamorphosed juveniles. The frequency of intermediate genotypes decreased and the magnitude of pairwise linkage disequilibrium increased between the two sampling periods, although no measure of statistical significance is given at any stage of the analyses. The argument for increased hybrid mortality is based on the absence of individuals from two central hybrid classes amongst juveniles, despite the fact that the combined frequency of these two classes was only 0.055 in the earlier sample. There is a significant difference in gene frequency between mid and late stage larvae, but not between late stage larvae and juveniles, a possible indication of selection for *R. berlandieri*-like genotypes (ie. differential environmental adaptation); however, the authors do not mention this.

(b) Howard, Waring, Tibbets, & Gregory 1993

A mosaic hybrid zone between the ground crickets *Allonemobius fasciatus* and *A. socius* occurs in regions (in the United States) where the life cycles of both species is univoltine, facilitating a straightforward cohort analysis. Howard *et al.* (1993) showed generally higher survival of one parental type relative to the other, with hybrids being intermediate (although results were not entirely uniform between either sites or sampling stages). This is also consistent with a simple ecotone model: it would imply that all five of their sampling sites were unfortunately situated on the same side of the environmental gradient, but is not, as the authors contend, a "rarely considered" phenomenon. The importance of adaptation to different environments in determining the mosaic distribution of the two species is fully recognised, so selection favouring one type over another is perhaps not surprising, but the authors conclude .

(c) Bert & Arnold 1995

Bert and Arnold (1995) outline clearly the respective predictions of a tension zone or an ecotone model for a hybrid zone, and test these with data from the hard-clam (genus *Mercenaria*) hybrid zone in a lagoon in Florida. A large-scale survey of clams from different regions of the lagoon was performed; individuals were aged and genotyped at four semi-diagnostic loci, facilitating a "static" cohort analysis of differences in the genetic make-up of different age classes. Evidence for selection against hybrids is claimed through the existence of heterozygote deficit, but the data presented do not agree with the authors' assertion that F_{is} increases across age classes, and are more suggestive of non-random mating. Linkage disequilibrium does increase across the first half of the age classes, but the proportion of hybrids in samples increases over the same period.

Evidence of spatial variation in selection, and hence support for an ecotone model, is claimed from the increase in heterozygote deficit and linkage disequilibrium and the

deficit of recombinant genotypes in more hybrid populations. Such an increase may simply have been generated by the fact that, of the seven regions sampled, data from three different hybrid populations were pooled (generating the Wahlund effect and increased D). More importantly, there is no evidence of spatial variation *with respect to genotype*, which distinguishes the ecotone model from that of a tension zone (see section 1.3). Finally, although multiple statistical comparisons are corrected for in a standard way, the fact that, for example, the various pairwise linkage disequilibrium estimates are not independent data points is not considered. Regardless of the validity of the statistics, the static cohort analysis has limitations, but the authors discount the possibility of temporal variation in dispersal, and do not mention the possibilities of temporal variation in selection or mating patterns.

Thus, in all three studies, interpretation of the results is not straightforward; in the first and the third, the conclusions are perhaps more confident than suggested by a closer analysis of the data presented. I draw attention to these problems by way of illustration of the potential pitfalls of a cohort analysis; however, given its superior power in resolving the question of relative fitnesses, increased use of the approach will hopefully result in the necessary fine-tuning.

3.6 (v) Potential for reinforcement?

The existence of postzygotic isolation generates selection for prezygotic isolation (Dobzhansky 1940; Mayr 1963). However, the evidence for any corresponding strengthening of prezygotic barriers to gene exchange, or reinforcement, in natural systems is equivocal (Howard 1993). There are some serious theoretical difficulties with the concept (Butlin 1989); of direct relevance here is that simulation studies have shown its maintenance to be difficult in the face of gene flow (Sanderson 1989). There is no indication of reproductive character replacement, a key prediction of the reinforcement

hypothesis, in studies of the male mating call in *Bombina* in Poland (Sanderson *et al.* 1992). However, levels of heterozygote deficit observed in hybrid populations in the *Bombina* transect (MacCallum 1994) are too great to be explained either by mixing of populations from different habitats or by selection against intermediate genotypes, suggesting the existence of at least some positive assortative mating. Reproductive character displacement has not been tested for in the Croatian transect. Taking the argument to its extreme, the habitat preference of the Croatian hybrid zone is a prezygotic barrier which has the potential to be strengthened due to its effect of reducing the frequency of unproductive interbreeding. Demonstrating character displacement in such a trait would nevertheless be extremely difficult given the confounding effects of habitat availability.

3.6 (vi) Implications for tension zone model

Nürnberger *et al.* (1995) demonstrated higher fitness in crosses from hybrid parents than in F1 crosses. The fitness reductions shown here for hybrid populations indicate the potential for extremely strong selection against F1's. The likelihood of a naturally-occurring F1 cross is obviously minimal, given the distance between pure populations, but it is the magnitude of the potential fitness reduction which ultimately represents the degree of reproductive isolation generated by genome divergence. In conclusion, the results of this study suggest that, however strong the effects of adaptation to alternative altitudes, terrestrial environment or aquatic breeding habitat (see, for example, Chapter Four), genomic incompatibilities must have major effects on fitness and hence on the overall dynamics of the *Bombina* hybrid zone.

Appendix 3.1 List of sites used for cohort analysis

1010	1018	1038	1058	1106	2117
1013	1019	1049	1103	1110	2129
1016	1020	1057	1105	2012	4183

Appendix 3.2 Estimating D from covariance of spot values

A second estimate of linkage disequilibrium from phenotypic data alone can be gained through an alternative interpretation of the spot score. The sum of the seven most nearly diagnostic[1] spot connections is a character index equivalent to the hybrid index defined by the genetic markers. If each connection s_i is thought of as comparable to a haploid locus, contributing 0 or 1 to the sum, then the methods for estimating linkage disequilibrium outlined in section 2.5(ii) are applicable. Having made necessary adjustments of factors of two to allow for the "haploid" state of connections, the equivalent of equation 2.5.1 is:

$$var(S) = \sum_{i=1}^{7} p_i q_i - \sum_{i \neq j} cov(s_i, s_j)$$

However, since the spot score is not completely diagnostic, Δs_i must also be considered, so $\overline{D}_F$ is calculated as:

$$\overline{D}_F = \overline{cov\,(s_i, s_j)} \Big/ \overline{\Delta s_i}^2$$

This provides a second means of estimating $\overline{D}_F$ from phenotypic data; but note that the treatment of spot connections as haploid "loci" is at best a tenuous approximation. The resolution of the statistic may not be great.

---

[1] Seven most diagnostic spot connections: Sternum, Sternum-Clavicle, Sternum-Visceral, Visceral, Visceral-Pelvic, Pelvic, Pelvic-Femural.

Chapter 4

Adaptation to alternative breeding habitats: the role of predation

4.1 Introduction

Ecological differences between the breeding habitats preferred by either *Bombina* species generate selection for different behavioural strategies in larvae. I demonstrate differences in the predation risk faced by a tadpole in either of the two habitat types, pond and puddle, and the existence of genetically-determined behavioural adaptations to these differences which afford a selective advantage to the adults' habitat preference.

4.1 (i) Habitat preference

Bombina bombina and *B. variegata* differ in their preferred breeding habitats: *B. bombina* breeds mainly in semi-permanent ponds whereas *B. variegata* is a characteristic puddle breeder (Arntzen 1978; Barandun 1995; Madej 1973). Although interbreeding will continually break down original gene combinations, with the genes determining habitat choice recombining away from other loci of the same parental ancestry, a preference for alternative habitats is maintained within the hybrid zone. This generates an association between habitat type and the genotype of individuals therein (Bugter *et al.* 1995; MacCallum 1994). For example, at the centre of the Croatian study site, the frequency of individuals with more than half *B. variegata* alleles (at the 4 diagnostic allozyme loci: p>0.5) was 9% in two large ponds (sites 1055 and 4185), compared with a frequency of 61% in immediately-adjacent puddles formed in wheel-ruts (1054 and 1056). The preference is not absolute, and appears to be asymmetric: individuals of a range of genotypes are found in puddles, whereas *B variegata* and *B. variegata*-like hybrids are almost never found in ponds. Figure 4.1.1 shows the change in mean gene frequency ($\bar{p}$) in ponds and puddles sampled across the Pešćenica transect. Frequent desiccation of puddles forces migration in search of new sites, and yet the association remains consistent despite the availability of both habitat types within individual toads' dispersal ranges.
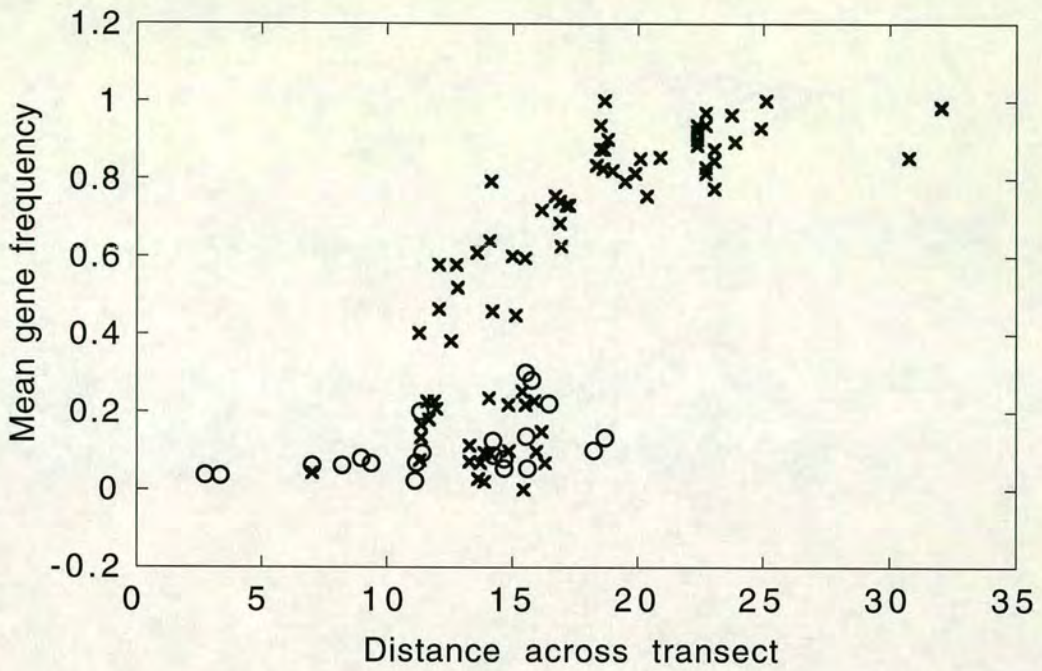
Figure 4.1.1 Mean frequency ($\bar{p}$) of *B. variegata* alleles in ponds (circles) and puddles (crosses) for sites from which 5 or more toads were collected. Distance across transect is measured in km, in north-east to south-west direction.

The relative frequency of either habitat type changes across the zone (Figure 4.1.1), with few puddles on the *B. bombina* side and no ponds on the *B. variegata* side. Representation of a given pond or puddle is obviously dependent on it containing *Bombina*, and by the above arguments there could be ponds on the *B. variegata* side which the toads avoid entirely and which therefore do not feature in the data set. However, in four seasons of extensive field work on the transect, with the aid of a high-resolution satellite image, no such ponds were discovered. In any case, I shall argue that it is the declining availability of suitable puddles which is relevant to the points made in this chapter, and which is only to be expected given that the *B. bombina* side of the zone

is arable lowland: lack of forest cover, lower precipitation and warmer temperatures minimise the likelihood of a puddle lasting for the duration of egg and larval development.

To be maintained in areas of interbreeding, a habitat preference must have adaptive advantages (Rice & Hostert 1993). From an ecological perspective, these advantages would explain the occupation of different environmental niches; from a population genetics perspective, they reduce the potential for hybridisation and so constitute a barrier to gene flow. In Chapter Six, I explore the theoretical implications of a habitat preference in a hybrid zone; in this chapter, I address the question of what are the adaptive advantages of such a preference in the *Bombina* system?

4.1 (ii) Habitat ecology

The different breeding habitats favoured by the two *Bombina* species can be classified by their permanence: the relative longevity of ponds creates a very different aquatic environment from ephemeral puddles. MacCallum (1994) measured a range of ecological characteristics of *Bombina* breeding sites in the Pešćenica area, and concluded that ponds and puddles were best discriminated by the amount of aquatic vegetation and size, with ponds containing greater amounts of both submergent and emergent vegetation, and being generally wider and deeper. In addition to the paucity of flora, puddles will presumably contain only the most opportunistic fauna. In particular, the abundance of potential aquatic predators, both vertebrate and invertebrate, on amphibian eggs and larvae should be higher in ponds. Previous studies have demonstrated the importance of interspecific interactions such as predation on larvae in determining species composition in anuran guilds (Cortwright & Nelson 1990; Morin 1983, 1986; Wilbur *et al.* 1983; Woodward 1983). Distributions of taxa along various environmental gradients can frequently be determined in terms of adaptation to respective ecological conditions;

predation therefore has the potential to define differential selection pressures within the *Bombina* system.

4.1 (iii) Aims

In this chapter, I test the hypothesis that, firstly, predation pressure in ponds is higher than in puddles and, secondly, that *B. variegata*'s avoidance of ponds is associated with inferior adaptation of its tadpoles to a predator-rich environment. In such an environment, any trait which reduces the risk of mortality from predation should be strongly favoured. Movement, in addition to increasing encounter rate, is a cue to many aquatic predators which detect their prey using visual or mechanosensory reception (Richards & Bull 1990). A tadpole moving around foraging will therefore be at greater risk of predation than a less active one (Werner & Anholt 1993). Differences in microhabitat use (e.g. Lawler 1989), morphology and palatability (e.g. Werner & McPeek 1994) will also determine the relative vulnerability of both species to predation. The above hypothesis generates the following three predictions:

(1) the abundance of potential predators on tadpoles is higher in ponds than in puddles during the *Bombina* larval period;

(2) the behaviour of *B. bombina* tadpoles makes them less vulnerable than *B. variegata* tadpoles to a predator: specifically, *B. bombina* tadpoles are less active;

(3) *B. bombina* tadpoles suffer lower mortality rates from predators than *B. variegata*.

4.2 Materials and methods

4.2 (i) Ecological surveys

Five breeding sites of each habitat type, distributed across the hybrid zone, were surveyed. Predator abundance was compared using a catch-per-unit-effort approach (Southwood 1978): at two locations in each site, three sweeps with a metal sieve (25cm diameter) were made. Sweeps were taken immediately below the water surface. This allowed consistency in sampling depth between shallow wheel-ruts and other sites, and preliminary samples with bottle traps indicated that, in deeper ponds, *Bombina* larvas were found in higher strata. All fauna caught were identified at least to the level of family. The abundance per sample of those known to be major tadpole predators (Fitter & Manuel 1995) and shown in pilot trials to attack *Bombina* larvae was then compared.

4.2 (ii) Behavioural experiments

Bombina variegata eggs were collected from breeding sites on one side of the hybrid zone, *B. bombina* eggs from the other side. Within a site, egg batches were taken from as widespread an area as possible, in order to maximise the number of families sampled, and not all clusters were taken. Eggs were reared in laboratory aquaria; after hatching, larvae were provided with abundant food in the form of powdered nettle leaves. Eggs were collected between 24-28 May 1995, and experiments conducted between 9-16 June 1995, by which time tadpoles had reached Gosner developmental stages 26-28 (Gosner 1960). Within a trial, tadpoles were matched on developmental stage and, as far as possible, on size.

The most common species of aquatic predators, as determined by the ecological surveys, were used for the experimental trials. These were: final instar *Aeshna* (hawker dragonfly) nymphs; final and penultimate instar *Dytiscus* sp. (great diving beetle) larvae; *Triturus*

dobrogicus (Danube crested newt) and *T. vulgaris* (smooth newt) adults. These were fed on tadpoles of other anuran species, *Rana* (brown and green frogs)and *Hyla* (treefrog), until 24 hours before an experiment.

Fresh non-chlorinated water at 20°C was used for each trial, thus preventing possible carry-over of semiochemical cues (Petranka *et al.* 1987a). Tadpoles were transferred to opaque plastic containers containing water and abundant food (nettle powder) 1 hour prior to an experiment, during which acclimatisation period container lids were left on.

(i) Experiment I : Activity levels

Activity levels of the two species before and after the introduction of a predator were compared. For each trial, five tadpoles (at Gosner developmental stage 27-28) of the same species were placed together in a 2-litre plastic container. Behaviour was scored by an observer standing motionless approximately 1 metre away: observations were dictated. An "activity index" was calculated as follows. For each group of five tadpoles in a container, the number which showed any movement during a 5-second observation periods was noted; twelve such observations were then taken at 1 minute intervals, and the activity level defined as the mean of the series of observations. This gave a value on a continuous scale between 0 and 5; note that no assumption was required of independence of either individuals' behaviour or of behaviour across time points.

After the first twelve observations, a single predator was added to each container. *Triturus vulgaris* were used in 20 trials, and *Dytiscus* larvae in another 20, with the two predator types assigned equally to the two tadpole species, creating a balanced design. Tadpole behaviour was measured as a function of a perceived predation threat, and not of actual predation events. To facilitate this, *T. vulgaris* adults of a size too small to damage the tadpoles were used, and the mandibles of each *Dytiscus* larva were temporarily taped

down. A further twelve observations were made immediately after the predator's introduction, and a second activity index for the group calculated as before.

Trials were conducted at 0900-1100 and at 1700-1900 hours; a time factor is therefore included in the analysis to allow for changing behavioural patterns through the day. A total of 40 trials were run, 20 for each species; each tadpole was only used once. Data were analysed using analysis of variance.

(ii) Experiment II : Survival rates

The relative vulnerability of *B. bombina* and *B. variegata* larvae to predation was compared using preference trials with various predators. For each trial, ten *B. bombina* and ten *B. variegata* tadpoles were placed together in a 6-litre container, with a moderate density of aquatic plants. Containers were covered throughout to minimise external disturbance. After the one-hour acclimatisation period, a single predator was introduced into each container. The number of trials run with each predator type were: *Dytiscus* larvae, 16; *Aeshna* nymphs, 19; *Triturus dobrogicus*, 2; *Triturus vulgaris,* 5. Each predator was used in only one trial.

Predators were left to feed until approximately half the tadpoles present had been eaten; this was determined by brief scans of the containers at 2 hour intervals, and so was not always exact. Average time taken was 13.4 ($\pm$ 9.9 s.d.) hours. The number of surviving tadpoles of either species was then recorded. Tadpoles of *B. bombina* and *B. variegata* are distinguishable by their stripe patterns, with *B. bombina*'s longitudinal dorsal stripes being more sharply defined than those of *B. variegata*; the latter is generally darker. Predator feeding bias was quantified using Manly's preference index:

$$i = \frac{ln(p_v)}{ln(p_v p_b)},$$

where p_v is the proportion of *B. variegata* surviving out of the initial 10, p_b the proportion of *B. bombina* (Cheeson 1978; Manly 1974). The index allows for the effect of prey depletion on availability. Values range from 0 to 1, with 0.5 representing random selection of prey and 1 representing only *B. variegata* being taken. Preference indices were analysed separately for two stage groups: tadpoles at Gosner stage 26 ('hatchlings'), and those at Gosner stage 27-28.

4.3 Results

4.3 (i) Ecological surveys

The mean abundances of different predator categories are given in Table 4.3.1. In all four categories a null hypothesis of equal abundance in ponds and puddles can be rejected. The abundance of (i) newts (*Triturus* sp.), (ii) dragonfly (suborder Anisoptera) and damselfly (suborder Zygoptera) nymphs and (iii) diving beetle (family Dytiscidae) adults or larvae is significantly greater in ponds. Salamander (*Salamandra salamandra*) larvae were found exclusively but only occasionally in puddles. The density of predators in pond samples was therefore substantially higher than in puddle samples.

Table 4.3.1. Abundance of predator categories in pond and puddle samples: means ± standard errors of abundance of each category in sieve-sweep samples. Categories are as follows: 1. Newt adults: *Triturus dobrogicus; T. alpestris; T. vulgaris*; 2. Dragonfly (Anisoptera) and damselfly (Zygoptera) nymphs; 3. Diving beetle (Dytiscidae) adults and larvae; 4. Fire salamander (*Salamandra salamandra*) larvae.

predator category	ponds (N=5)	puddles (N=5)	G-test of equal distribution; df=1
1. newts	4.83 (± 2.41)	1.33 (± 0.76)	12.6; p<0.01
2. dragonfly / damselfly nymphs	16.5 (± 7.07)	0	137.0; p<0.01
3. diving beetles	12.0 (± 3.50)	0.17 (± 0.17)	90.6; p<0.01
4. salamander larvae	0	0.50 (± 0.22)	4.16; p<0.01

4.3 (ii) Experiment I : Activity levels

There were no tadpole groups for which zero activity was recorded, either before or after the introduction of a predator. Movement was strongly associated with feeding. Mean activity indices for each species in the absence and then presence of a predator are presented in Figure 4.3.1; Table 4.3.2 contains an analysis of variance of the data.
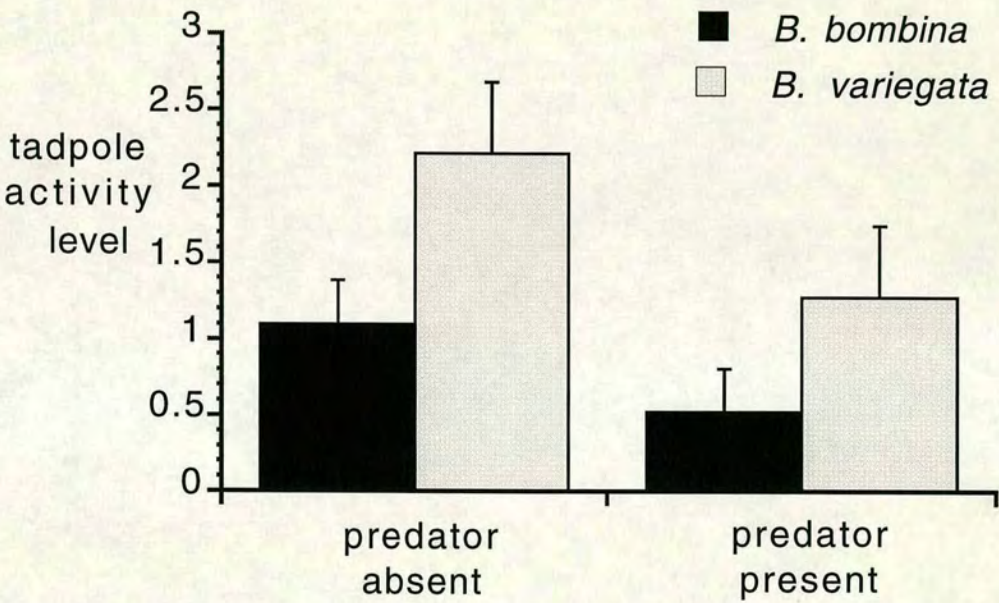
Figure 4.3.1. Mean activity indices for *B. bombina* and *B. variegata* tadpoles before and after the introduction of a predator; bars represent standard errors. See text for derivation of indices, and Table 4.3.2 for analysis of variance of activity level.

Table 4.3.2. Analysis of variance and parameter estimates of tadpole activity levels. Species term compares *B. variegata* to *B. bombina*; presence of predator term includes both predator types. Time effect compares evening activity level with that of morning; this is nested within species, to give a parameter estimate for either species. Repeated measures on each tadpole group (one before the introduction of a predator, one after) allow comparison between groups, nested within time and species.

source	df	sequential SS	F-ratio	p-value
species	1	17.500	91.047	**< 0.001**
presence of predator	1	11.438	59.509	**< 0.001**
species * predator presence	1	0.657	3.418	0.0723
time[species]	2	5.030	13.085	**< 0.001**
group[time[species]]	36	15.593	2.253	**0.0075**
predator * time[species]	2	0.107	0.2682	0.766
error	36	7.197		
total	79	57.522		

term	parameter estimate (± std error)
intercept	1.024 (± 0.329)
species	2.848 (± 0.450)
presence of predator	-0.575 (± 0.139)
species * predator presence	-0.363 (± 0.196)
B. bombina: time effect	0.125 (± 0.438)
B. variegata: time effect	-1.958 (± 0.438)

The difference in the overall activity levels of the two species was highly significant: *B. variegata* tadpoles were consistently more active than *B. bombina*, with the parameter estimate indicating an average increase of 2.85 units. After the introduction of a predator, activity was significantly lower in both species: the overall effect was to lower the activity index by 0.58. There was no evidence of a difference between species in response to predator introduction. Time of day did not affect the activity level for *B. bombina*, but reduced that of *B. variegata* by nearly 2 units, despite controlled water temperatures of 20°C. There was also significant variation between groups of tadpoles within species and time of day.

Activity level in the presence of a predator can be expressed as a proportion of the activity level before the introduction. *B. bombina*'s activity was reduced to an average of 55.12 (± 6.77 s.e.)% of the original level; *B. variegata*'s to 61.04 (± 8.36)%. An analysis of variance of log-transformed percentages showed no difference between species (F-ratio=0.23; df=1,36; p=0.64), nor between predator types (F-ratio=1.38; df=1,36; p=0.25). There was therefore no evidence that *B. bombina* reduce their activity more than *B. variegata* or vice versa, nor that the effect of a *Dytiscus* larva was different from that of a *Triturus vulgaris* adult.

4.3 (iii) Experiment II : Survival rates

The Manly preference indices for the two developmental stage groups are presented in Figure 4.3.2. Amongst the hatchling size class (Gosner stage 26), prey depletion was apparently random with respect to species (mean index = 0.496 ± 0.057 s.e.; t-statistic = -0.078; n = 19; p=0.93). However, at the later stages (Gosner 27-28) *B. variegata* were more vulnerable than *B. bombina*: the mean preference index was significantly greater than 0.5 (mean index = 0.612 ± 0.045; t-statistic = 2.49; n=23; p=0.02). Analysis of variance of the preference index (*Triturus* sp., *Dytiscus* larvae and *Aeshna* nymphs) showed no evidence of a difference between the three predator types (*Triturus* sp., *Dytiscus* larvae and *Aeshna* nymphs: F-ratio=2.03; df=2,36; p= 0.146) nor of an interaction between predator type and developmental class (F-ratio=0.065; df=2,36; p=0.937).

4.3.2. (a) Gosner stage 26; N=19
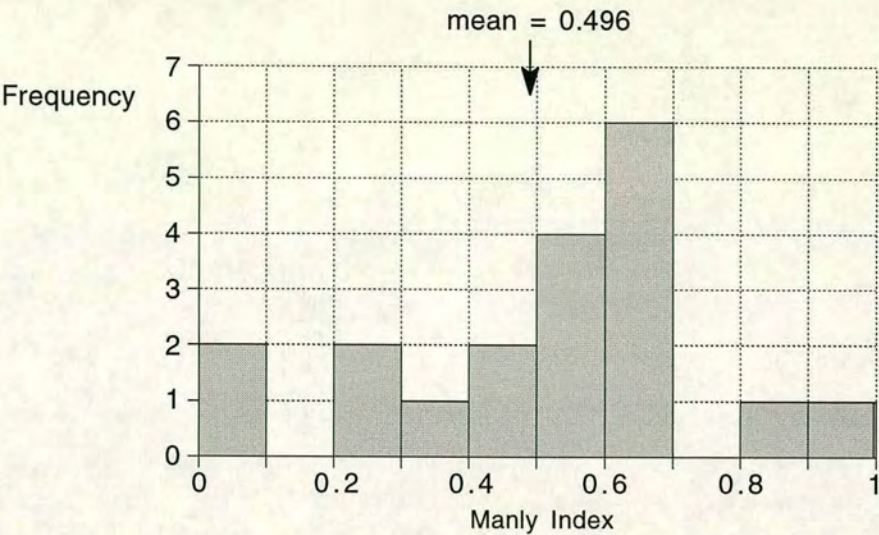
4.3.2. (b) Gosner stage 27-28; N=23
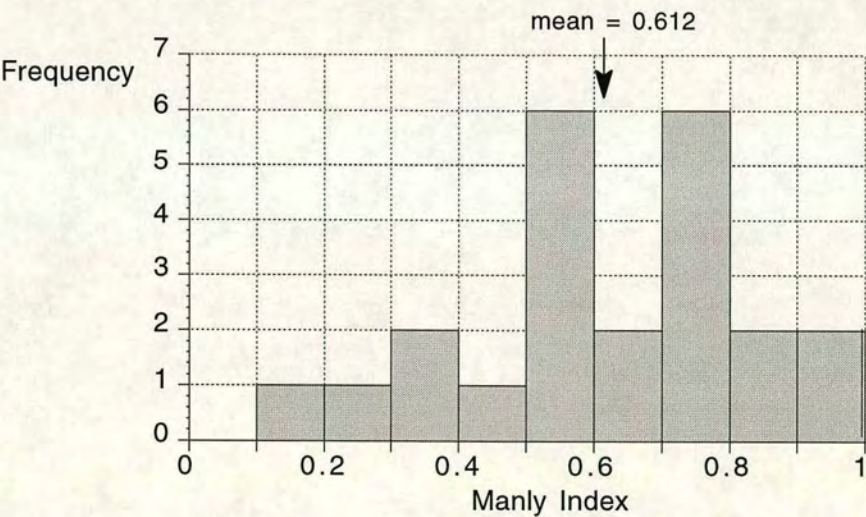
Figure 4.3.2. Distribution of Manly preference index (see text for derivation) from predator-preference trials, for two categories of larval developmental stage.

4.4 Discussion

4.4 (i) Summary of results

The results confirm the prediction that, during the *Bombina* larval period, predator density was higher in the semi-permanent ponds in which the fire-bellied toad *B. bombina* breeds than in the temporary puddles which are typical habitat of the yellow-bellied toad, *B. variegata*. In experimental trials, *B. bombina* tadpoles differed from *B. variegata* in their activity levels: *B. bombina* were consistently less active and spent less time moving around feeding than *B. variegata*. Both species responded to the disturbance induced by introduction of a predator by reducing activity, but *B. variegata* still remained more than twice as active. Once past the hatchling stage, *B. variegata* tadpoles suffered higher levels of mortality than *B. bombina* in predator choice experiments. The availability of hatchlings at Gosner stage 26 was restricted, so they were used for Experiment II only. Whilst a similar comparison of their activity levels would provide a more complete picture, tadpoles at this stage are largely inactive (Lawler, 1989; pers. obs.), so it is unlikely that their inclusion would affect the conclusions.

4.4 (ii) Implications for fitness

Do the behavioural differences demonstrated imply adaptation to different breeding habitats? Development under the risk of predation creates a trade-off between resource-acquisition and predator avoidance (Skelly 1995; Werner & Anholt 1993; Wilbur & Fauth 1990). Low activity rate will reduce vulnerability, but will also decrease foraging rate and hence growth and development rate. Conversely, higher activity levels facilitate faster development rates and so a shorter time to reaching either a size refugia (e.g. Richards & Bull 1990) or metamorphosis: the average larval period for *B. variegata* is 87% that for *B. bombina* (Nürnberger *et al.* 1995). This implies less time at risk, a point which seems to be frequently overlooked in literature on similar studies. The benefits of higher activity

might therefore outweigh the cost of increased vulnerability in a predator-rich environment. However, physiological constraints should impose diminishing returns on increases in development rate with time active, whereas the probability of encounter with a predator will increase linearly with time active. The incremental benefit to development rate of increased activity would be lower than the incremental cost in terms of predation risk (Werner & Anholt 1993). A less active species will therefore be better adapted to an environment in which predation is a regulating factor, in spite of the longer larval period incurred. The results imply that response to the growth-mortality trade-off differs between *B. bombina* and *B. variegata* tadpoles as predicted by the ecology of their usual breeding habitats. The genetically-determined behavioural patterns demonstrated here should serve to increase *B. bombina*'s fitness relative to *B. variegata* in a semi-permanent pond. Ideally, this should be confirmed with measurements of differential survival of tadpoles in the field, using molecular markers to compare changes in the frequency of alternative genotypes in a cohort of tadpoles (but see section 3.6 for mention of the difficulties encountered in a pilot experiment on tracking larval survival).

4.4 (iii) Implications for Bombina distribution

Differences in tadpole behaviour and susceptibility to predation have been found in other taxa and, similarly, have been invoked to explain species' distributions along environmental gradients or occupation of alternative ecological niches (Chovanec 1992; Lawler 1989; Morin 1983; Skelly 1995; Werner & McPeek 1994; Woodward 1983). The data are consistent with the hypothesis that predator-rich environments should contain less active species. I do not however know of any other study where the contrast has been drawn between two regularly interbreeding species. Traits characteristic of either *Bombina* species are in strong statistical association (linkage disequilibrium) within hybrid populations (MacCallum 1994; Chapter 2, this thesis); selection on differences at the larval stage will therefore cause a correlated advantage of habitat preference in adults.

Simulation models confirm that a selectively-advantageous habitat preference, by forcing assortative mating, will play an important role in maintaining the integrity of the genomes of two hybridising taxa (Chapter Six).

4.4 (iv) Other factors affecting fitness in relation to habitat

The results presented here demonstrate differential adaptation to a predator-rich pond environment. There may also be differential adaptation to the puddle habitat: *B. bombina*'s slower development rate (Nürnberger *et al.* 1995) will reduce its fitness relative to *B. variegata* in temporary water bodies where dessication is often the main source of mortality (e.g. Skelly 1995). Food and space are presumably more limited in puddles, suggesting that competition will be the regulating ecological factor (Begon *et al.* 1990). Under such conditions, faster-developing, more active tadpoles have a competitive advantage (Morin 1983; Petranka & Sih 1986; Scott 1990; Semlitsch & Caldwell 1982; Travis 1980; Werner 1992), suggesting a fitness advantage for *B. variegata* in puddles. Whilst this might suggest interesting implications for the relative importance of predation and competition in structuring amphibian guilds, considerably more work would be required on either aspect before any such conclusions could be drawn. In particular, the relative performance of either species in puddles at the cooler temperatures of *B. variegata*'s range needs to be established.

4.4 (v) Implications for the ecotone model

As outlined in section 4.1, the availability of either habitat type changes across the transect of the hybrid zone at Peščenica. The frequency of puddles declines drastically on the *B. bombina* side of the zone, in the warmer, arable lowlands, with the majority of breeding sites being semi-permanent ponds or drainage ditches. The results of this chapter illustrate that *B. variegata* is less well adapted than *B. bombina* to the ecological conditions of such breeding habitats, and so provide the first direct evidence of

differential adaptation (although only in one direction) across the environmental gradient. In addition to the endogenous selection acting against hybrids (Chapter Three and references therein), exogenous selection must also be contributing to the maintenance of the *Bombina* hybrid zone. The following chapter uses theoretical models to compare the dynamics of either regime.

Chapter 5

Multilocus clines

5.1 Introduction

Evolutionary biology is concerned with diversity, from the nucleotide variation within individual genomes to the macro-evolutionary changes revealed by palaeontology. An important aspect of biological diversity is the maintenance, by a barrier to gene exchange, of divergence between two populations. When the populations in question classified as different subspecies or species, such barriers to gene exchange are responsible for maintaining taxonomic diversity. Their structure is best understood in cases where the barrier is not completely effective, in hybrid zones in which the homogenising effect of migration and recombination is counterbalanced by natural selection acting to maintain the integrity of the parental genomes, resulting in a set of clines (Barton & Hewitt 1985). This intermediate stage - a system which endures despite having neither achieved nor abandoned reproductive isolation - provides an indirect means of observing mechanisms capable of preventing gene flow, and hence of elucidating the components of speciation; hybrid zones offer a "window on the evolutionary process" (Harrison 1990).

Hybrid zones have been described in a wide range of organisms, from grasshoppers (e.g. *Chorthippus*, Hewitt 1993a) to deer (e.g. *Cervus*, Abernethy 1994). With hundreds of case studies documented in the literature, it is important to be able to relate descriptions to a theoretical framework, to facilitate comparisons and, hopefully, highlight parameters of importance. Theoretical descriptions of a cline (or gradient in a single trait) date back to Fisher (1937) and Haldane (1948), with significant recent contributions by Bazykin (1969), Slatkin (1973), Nagylaki (1975; 1976a) and Endler (1977). However, earlier analysis was restricted to the fate of a single locus (or at most two loci (Slatkin 1975)). In a series of papers, Barton has extended these models to a multilocus system (Barton 1983, 1986; Barton & Bengtsson 1986). This renders plausible their application to the study of hybrid zones in which differences between the two populations may be expected

across a large proportion of the genome. On a theoretical level, the models offer insight into the effects of associations between loci, and the counterbalance between natural selection acting to maintain the integrity of two parental genomes and recombination acting to break it down. On a practical level, measurable features such as the shape of clines in genetic markers or the statistical associations between them can be used to estimate the strength of the barrier to gene flow, the strength of natural selection acting, the dispersal rates of the organisms and even the number of genes under selection. These methods have been applied in several well-studied hybrid zone (*Bombina*: Szymura & Barton 1986, 1991; *Podisma*: Barton 1980, Barton & Hewitt 1981a, Jackson 1992; *Heliconius*: Mallet *et al.* 1990; *Sceloporus*: Sites *et al.* 1995).

I present here a comprehensive review of Barton's multilocus cline theory and the derivations of the calculations mentioned above. (In particular, the derivation of the number of genes under selection has not been published in explicit form.) At various stages, assumptions or approximations are required to keep the analysis tractable: for example, (i) populations are infinite and all processes are deterministic; (ii) the strengths of both selection and associations between loci are assumed to be relatively weak. The assumptions are individually justifiable, but their cumulative effect may be unreliable. Computer simulations have been used to test the robustness of some of the predictions (Barton & Gale 1993; Jackson 1992); with the greater computing capacity now available, I develop Barton and Gale's model to test the complete set of predictions over a range of parameters.

Consider a hybrid zone maintained by a balance between migration, removing differentiation, and natural selection, maintaining it. As discussed in earlier chapters, various alternative modes of natural selection are possible. The primary distinction lies between selection maintaining different alleles either in different environments or in different genetic

backgrounds: respectively, (i) "exogenous" selection, with fitness defined in relation to the environment, and a geographic selection gradient determining the relative fitness of genotypes and (ii) "endogenous" selection, with fitness defined by within-genome interactions such as heterozygote disadvantage or epistasis, and being independent of environment.

The alternative modes have very different implications for both the mechanics of selection and the ultimate fate of a hybrid zone, yet the shapes of the clines they produce are effectively indistinguishable, at least for single locus models. Their relative frequencies have therefore been the subject of considerable debate in the hybrid zone literature. Reviews of more than 100 case studies (Barton & Hewitt 1981b, 1985) concluded that most hybrid zones are maintained by the latter, specifically selection against hybrids independent of the environment, or "tension zones" (Key 1968). With this justification, Barton's multilocus models were developed using models of heterozygote disadvantage; simulations showed that the effects of epistatic selection were qualitatively similar (Barton & Gale 1993; Jackson 1992). However, there is also substantial evidence for exogenous selection in hybrid zones (Harrison 1990; see Chapter One). This chapter develops an analogous treatment to Barton's earlier model under the alternative selection regime, specifically adaptation of the two populations to different environments. Important aspects of the two models are compared, and the robustness of the predictions mentioned above to alternative selection regimes regimes tested with simulations.

I consider first a cline in a single selected locus, then clines in multiple selected loci and finally clines in neutral loci linked to others under selection. Within each section, the relevant theory is developed, its predictions compared with results from the computer simulations and inferences from either discussed. At the end of the chapter, Appendix 5.1 lists the mathematical symbols used; details of the simulation model, statistical method-

ology and tests of the program are presented in Appendices 5.2, 5.3 and 5.4 respectively; Appendix 5.5 outlines Barton's multilocus heterozygote disadvantage model; and a full listing of the program code is given in Appendix I at the end of the thesis.

5.2 Single locus clines

Consider the most simple case of a cline in a single locus which is segregating for two alleles. Two populations fixed for alternate alleles are in contact, and interbreed at their interface. As in many biological models, the random movement of individuals is represented by a deterministic reaction-diffusion process (Murray 1989; Roughgarden 1979). (Fisher and Haldane's use of the diffusion approximation to describe gene flow is rigorously derived by Nagylaki (1975).) The rate of change at any point in space is proportional to the gradient in curvature at that point:

$$\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2},$$

(5.2.1)

where p is the frequency of one of the alleles ($0 \le p \le 1$), x is the geographic distance along the cline, σ^2 is the variance in distance moved per generation ($\sigma^2 > 0$), and both time and distance are measured as a continuous variables (Haldane 1948; Nagylaki 1975).

If the allele in question is selectively neutral, the outcome will be entirely determined by equation 5.2.1, and the initial sharp step in gene frequency will decay with time until all spatial differentiation is lost. Alternatively, selection may act to prevent this decay, resulting in a stable cline maintained by the counterbalancing forces of migration and selection. If selection is weak, the change in gene frequency in one generation due to selection (from Wright 1931) will be small, allowing a continuous time approximation for the effect of selection. Assuming also that there is no interaction between selection and dispersal, the rate of change in gene frequency will then be determined by the additive effects of these factors, and is given by:

$$\frac{\partial p}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} + \frac{1}{2}\frac{\partial \log(\overline{W}(p))}{\partial p}pq, \qquad (q = 1-p). \qquad (5.2.2)$$

$\overline{W}(p)$ is the mean fitness in a population at frequency p, and defines the type of selection acting. As outlined in the introduction, natural selection could act in a variety of ways to stabilise a hybrid zone. Reduced fitness of hybrid offspring, independent of their external environment, will result in a "tension zone" (Barton & Hewitt 1985; Key 1968); in a single locus model, this is depicted by defining heterozygote fitness as 1-s, relative to a fitness of 1 of either homozygote (Bazykin 1969, 1973). The analysis of multilocus clines to date (Barton 1983, 1986; Barton & Bengtsson 1986; Barton & Gale 1993) has concentrated on the extension of this definition to many genes. I consider here an alternative mode of natural selection: adaptation of either population to alternative environments, producing a cline the position and width of which will be determined by the change in environmental conditions.

Consider a stepped environment, with different fitness regimes on either side of a step at x=0 (after Endler 1977; Haldane 1948; May *et al.* 1975; Slatkin 1973). When x<0, the homozygote for the allele at frequency p has fitness $(1-s)^2$, the heterozygote has fitness 1-s and the homozygote for the alternative allele has fitness 1. The situation is reversed for x≥0, with relative fitnesses given by $1:1-s:(1-s)^2$. I will refer throughout to this type of selection regime as environmental selection. Equation 5.2.2 then becomes:

$$\frac{\partial p}{\partial t} = \begin{cases} \dfrac{\sigma^2}{2}\dfrac{\partial^2 p}{\partial x^2} - \dfrac{s}{\overline{W}(p)}pq, & x < 0; \\[3mm] \dfrac{\sigma^2}{2}\dfrac{\partial^2 p}{\partial x^2} + \dfrac{s}{\overline{W}(p)}pq, & x \geq 0. \end{cases} \qquad (5.2.3)$$

At equilibrium, $\partial p/\partial t = 0$; the assumption of weak selection gives $\overline{W}(p) \approx 1$ and the fact that $\partial(\partial p/\partial x)^2/\partial p = 2\partial^2 p/\partial x^2$ gives a first order differential equation of $(\partial p/\partial x)^2$ with respect to p. Integrating this, and incorporating the boundary condition of zero gradient at p=0 gives, for x<0:

$$\left(\frac{\partial p}{\partial x}\right)^2 = \frac{2s}{3\sigma^2}\left(3p^2 - 2p^3\right) \qquad\qquad (5.2.4)$$

This can be integrated a second time; defining p(0)=0.5 gives:

$$p(x) = \frac{3}{2} - \frac{3}{2}\tanh\left[\sqrt{arctanh(2/3)} - \sqrt{\frac{s}{2\sigma^2}}\,x\right]^2 \qquad x < 0 \qquad (5.2.5a)$$

Similarly:

$$p(x) = -\frac{1}{2} + \frac{3}{2}\tanh\left[\sqrt{arctanh(2/3)} + \sqrt{\frac{s}{2\sigma^2}}\,x\right]^2 \qquad x \geq 0 \qquad (5.2.5b)$$

The two solutions give an explicit description of gene frequency as a function of distance across the cline. Figure 5.2.1 shows clines resulting from alternative selection strengths s; increasing the dispersal rate σ has a similar widening effect on cline width. The model of heterozygote disadvantage at a single locus also contains a tanh curve in its equilibrium solution (Barton 1979a; Bazykin 1969), though the cline shapes differ in the tails. In both cases, the quantitative nature of the resulting cline depends on the ratio $\sigma/\sqrt{s}$, the "characteristic length" of the cline.
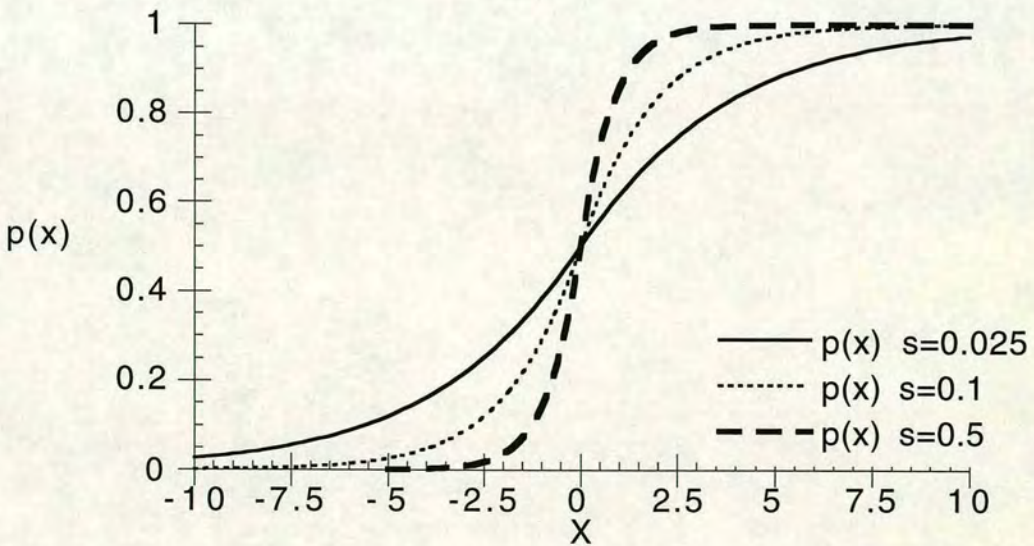
Figure 5.2.1 Shape of single locus cline with environmental selection (equation 5.2.5), for selection strengths s=0.025, s=0.1 and s=0.5. Migration rate $\sigma^2 = 0.5$.

5.3 Selection acting on multiple loci

The preceding section described the formation of a cline at a single locus under selection. In a hybrid zone, parental populations will probably differ across numerous loci, more than one locus will be affected by selection and other loci again will be selectively neutral. Most forms of interactions between these loci will render a single locus description inadequate. What do we expect if the above analysis is extended to many loci? I outline below the main arguments from Barton and Bengtsson's analysis (1986; see also Barton 1983, 1986) of clines at multiple loci.

At one extreme, the presence of other loci under selection will not affect the fate of a particular allele; at the other, the locus in question will behave as if subject to the total selection pressure acting across the genome. We therefore need a measure of how "congealed" (Turner 1967) the genome is under the given selection pressures. Barton's (1983) analytical results show that the behaviour of such a system will be defined by the relative strengths of the selection acting on a single locus, s, and its rate of recombination with adjacent loci, r, described by their ratio $\theta=s/r$, and by the number of loci in question. The analysis that follows makes the assumption that this selection is weak, and in particular that the ratio θ, the "coupling coefficient", is small.

Interactions will generate statistical associations between loci, so a list of the gene frequencies at each locus will not fully describe the situation. Gametic frequencies would give a complete description of the system, but quickly become unwieldy if more than three loci are considered. An alternative approach is to track individual gene frequencies and the statistical associations between alleles at different loci; the two approaches are interchangeable. With the assumption that selection is weak but acts across many loci, associations higher than pairwise or, at most, three-way become negligible, so the latter

approach is considerably more tractable (Barton 1983). The effect of selection on gene combinations, or epistasis, will be of order s^2 so should also be unimportant under a weak selection approximation. The single locus theory outlined above required the assumption that fitness be independent of the genotype frequency. Here, in addition, I consider the case of selective effects being approximately equal across loci, and (at a later stage) approximate the probability of recombination between any pair of loci by the harmonic mean across all pairs.

The frequency p of an allele at a locus under selection will be affected by four factors: (i) migration; (ii) selection acting directly on that locus; (iii) selection acting on gene combinations containing the locus in question (epistasis); (iv) selection acting directly on other loci (or other combinations of loci) with which the locus is associated. The first two are described as for a single locus model (equation 5.2.2); I do not consider epistasis here[1]. To quantify the effect of the selection acting on other individual genes, consider first two loci. The degree to which selection at one locus affects the dynamics of the other will clearly depend on how closely associated the two are. If two loci are segregating with alleles P_1/Q_1 and P_2/Q_2 respectively, their "association" can be formally quantified by the pairwise linkage disequilibrium D, defined as the excess of coupling gametes over that expected under random association:

$$D = \text{frequency}(P_1P_2 \text{ gametes}) - (\text{freq}(P_1) * \text{freq}(P_2)) \qquad (5.3.1)$$

(This can be shown to equal to the covariance in allelic states; see Chapter 2. The terminology is misleading, suggesting both physical linkage and a state of instability, neither of which is relevant.)

_____

[1]In any case, a cline maintained by epistatic selection against hybrids has very similar dynamics to one maintained by heterozygote disadvantage (Jackson 1992; Mallet and Barton 1989).

Define p_1=freq(P_1), p_2=freq(P_2), and PP=freq(P_1P_2) (so q_1=freq(Q_1), PQ=freq(P_1Q_2), QP=freq(Q_1P_2) etc.; note that PP+PQ=p_1). As before, selection on the first locus will have the following effect on its own gene frequency:

$$\frac{\partial p_1}{\partial t} = \frac{\partial \log \overline{W}}{\partial p_1} \frac{p_1 q_1}{2} \qquad (5.3.2)$$

What are the effects of this selection on the gamete frequencies? Considering selection on the first locus alone implies that, *within genetic backgrounds of equivalent fitness, frequencies at the second locus do not change.* The proportion of coupling gametes P_1P_2 amongst all gametes carrying the P_1 allele, PP/(PP+PQ), is therefore constant:

$$\partial(PP/p_1)/\partial t = 0$$

$$\Rightarrow \frac{1}{p_1}\frac{\partial PP}{\partial t} - \frac{PP}{p_1^2}\frac{\partial p_1}{\partial t} = 0$$

$$\Rightarrow \frac{\partial PP}{\partial t} = \left(\frac{PP}{p_1}\right)\frac{\partial p_1}{\partial t} \qquad (5.3.3)$$

$$= (p_1 p_2 + D)\frac{q_1}{2}\frac{\partial \log \overline{W}}{\partial p_1} \qquad since\ PP = (p_1 p_2 + D)$$

The frequency QP of the repulsion gamete Q_1P_2 behaves similarly; since p_2=PP+QP, this gives the overall effect of selection on the first locus on the gene frequency at a second locus:

$$\frac{\partial p_2}{\partial t} = \frac{\partial PP}{\partial t} + \frac{\partial QP}{\partial t}$$

$$= \frac{1}{2}\frac{\partial \log \overline{W}}{\partial p_1}(q_1(p_1 p_2 + D) - p_1(q_1 p_2 - D)) \qquad (5.3.4)$$

$$= \frac{1}{2}\frac{\partial \log \overline{W}}{\partial p_1}D$$

(Note that this is directly analogous to equation 5.3.2, but with the variance in allelic states within a locus replaced by the covariance between loci.)

This can be extended to a system involving L selected genes, where the allele frequency p_i at any one locus is also affected by selection on the other L-1 loci (each with

frequencies p_j). Assume that, with weak coupling θ, higher order terms (for example, three-way or more linkage disequilibria) need not be considered; this gives:

$$\frac{\partial p_i}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 p_i}{\partial x^2} + \frac{p_i q_i}{2}\frac{\partial \log \overline{W}}{\partial p_i} + \frac{1}{2}\sum_{j\neq i, j=1}^{L}\frac{\partial \log \overline{W}}{\partial p_j}D_{ij}$$ (5.3.5)

The summation is over all the other loci under selection and D_{ij} is the pairwise linkage disequilibrium between the locus in question and locus j.

Selection is assumed to act multiplicatively, although with weak selection the distinction from additivity is negligible; selective effects are assumed to be approximately equal across loci. Environmental selection in a multilocus system can then be described by defining an individual's fitness as:

$$W = \begin{cases}(1-s)^k \approx e^{-ks} & x < 0 \\ (1-s)^{2L-k} \approx e^{-(2L-k)s} & x \geq 0\end{cases}$$ (5.3.6),

where k is the number of alleles originating in the population to the right of the environmental step; the exponential term assumes both that s is small and L is large. To order s^2, the mean fitness in a population is given as:

$$\overline{W}(\underline{p}) = \begin{cases}e^{-2\sum p_i s} & x < 0 \\ e^{-2\sum(1-p_i)s} & x \geq 0\end{cases}$$ (5.3.7),

where $\underline{p} = (p_1, p_2 \cdots p_L)$. Incorporating this in equation 5.3.5 at equilibrium gives:

$$0 = \begin{cases}\dfrac{\sigma^2}{2}\dfrac{\partial^2 p_i}{\partial x^2} - p_i q_i\, s - s\sum_{i\neq j}D_{ij} & x < 0 \\[4mm] \dfrac{\sigma^2}{2}\dfrac{\partial^2 p_i}{\partial x^2} + p_i q_i\, s + s\sum_{i\neq j}D_{ij} & x \geq 0\end{cases}$$ (5.3.8)

The equilibrium state of the multilocus system is thus described in terms of the allele frequencies and pairwise disequilibria.

Simulation model

The simplicity of equation 5.3.8 is achieved via ever-thickening layers of assumptions introduced at each stage of the analysis. Although these are essentially all variations on the weak selection approximation, their combined effect may render the entire description invalid. A check on whether a system will indeed follow these patterns is therefore necessary, and can be performed with a simulation model. I simulate a one-dimensional chain of demes with nearest-neighbour migration; diploid individuals, described by multilocus genotypes, follow a life cycle of migration, reproduction and then selection. Selection is soft, as in the Levene (1953) model: population regulation occurs within each deme, with each deme producing a constant number of individuals. Both time and space are therefore discrete variables, in contrast to their continuous form in the analytical equations. Details of the model are given in Appendix 5.2; Appendix 5.4 summarises checks that the program is performing as expected. Throughout the following sections of the analysis, I use the simulation to check broadly that the predictions of the analytical theory are sensible and, secondly, to test for the critical parameter values at which certain assumptions break down.

5.4 Describing interactions between genes: linkage disequilibria

I outlined above the expected dynamics of allele frequency at a given locus; in this section, I use a similar approach to describe the linkage disequilibrium. As before, let p_1 and p_2 be the gene frequencies at the two loci, PP be the frequency of the coupling gamete, and let D ($=PP-p_1*p_2$) describe the pairwise linkage disequilibrium between the loci. The diffusion approximation can be used again to describe the effect of migration on gamete frequencies to give, for example:

$$\frac{\partial PP}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 PP}{\partial x^2} \tag{5.4.1}$$

The effect of migration on D is therefore described by:

$$
\begin{aligned}
\frac{\partial D}{\partial t} &= \frac{\partial\left(PP-p_1*p_2\right)}{\partial t}\\[2mm]
&= \frac{\partial P_{11}}{\partial t} - p_1\frac{\partial p_2}{\partial t} - \frac{\partial p_1}{\partial t}p_2\\[2mm]
&= \frac{\sigma^2}{2}\left(\frac{\partial^2(D+p_1p_2)}{\partial x^2} - p_1\frac{\partial^2 p_2}{\partial x^2} - \frac{\partial^2 p_1}{\partial x^2}p_2\right)\\[2mm]
&= \frac{\sigma^2}{2}\left(\frac{\partial^2 D}{\partial x^2} + p_1\frac{\partial^2 p_2}{\partial x^2} + 2\frac{\partial p_1}{\partial x}\frac{\partial p_2}{\partial x} + \frac{\partial^2 p_1}{\partial x^2}p_2 - p_1\frac{\partial^2 p_2}{\partial x^2} - \frac{\partial^2 p_1}{\partial x^2}p_2\right)\\[2mm]
&= \frac{\sigma^2}{2}\frac{\partial^2 D}{\partial x^2} + \sigma^2\frac{\partial p_1}{\partial x}\frac{\partial p_2}{\partial x}
\end{aligned}
\tag{5.4.2}
$$

In this last line, the first term describes the change in D as it diffuses out from a maximum value. It will be of order sD (Barton 1983); as demonstrated below, D is itself of order s, so for weak selection the first term is assumed to be negligible relative to the second term. The second term, the product of gene frequency gradients, is the generation of linkage disequilibrium through the mixing of populations with different frequencies (Li & Nei 1974).

In discrete time, each generation of random mating will reduce D by a factor of $(1-r)$, where r is the recombination rate between the two loci. If the recombination rate is low, this can be approximated to continuous time:

$$\frac{\partial D}{\partial t} = -rD \tag{5.4.3}$$

Disequilibria may also be generated by selection for "co-adapted" gene complexes, epistasis, but it is assumed that this effect will be weak compared to the effects of migration between populations at different gene frequencies and recombination (Barton 1983). The dynamics of D can therefore be approximately described by:

$$\frac{\partial D}{\partial t} = -rD + \sigma^2 \frac{\partial p_1}{\partial x} \frac{\partial p_2}{\partial x} \tag{5.4.4}$$

Define the width of a cline as the inverse of its maximum gradient. If the clines are concordant (same width), the equilibrium solution is given by:

$$D = \frac{\sigma^2}{r}\left(\frac{\partial p_1}{\partial x}\right)\left(\frac{\partial p_2}{\partial x}\right)$$

$$\Rightarrow \quad D_{centre} = \frac{\sigma^2}{rw^2} \qquad (w = width) \tag{5.4.5}$$

Note that with the assumption of weak selection, the disequilibrium is determined solely by the relative magnitudes of recombination and migration, and so is not explicitly affected by the manner in which selection acts.

Exchange between three demes

The simulation model can be used to test equation 5.4.5 and the effect of the many assumptions involved in its derivation. Populations are distributed over discrete demes, rather than through continuous space, so I first outline an analogous analysis for a discrete deme case, showing that the prediction for the discrete model corresponds directly to that for the continuous case.

Consider three demes, in which gene frequency is lowest in the first (deme 0), intermediate in the central (deme 1) and highest in the third (deme 2). A proportion m of the

diploid individuals in the central deme migrate each generation; these are replaced by m immigrants, half from each of the neighbouring demes. Let p_{1x}, p_{2x} and D_x be, respectively, the frequencies of the two loci and the linkage disequilibrium in deme x, PP etc. be the gametic frequencies in the *central* deme, and r be the recombination rate between the two loci. After migration, the gene and gamete frequencies are given by:

$$p'_{11} = (1-m)p_{11} + \frac{m}{2}p_{10} + \frac{m}{2}p_{12};$$

$$p'_{21} = (1-m)p_{21} + \frac{m}{2}p_{20} + \frac{m}{2}p_{22}; \tag{5.4.6}$$

$$PP' = (1-m)(p_{11}p_{21} + D_1) + \frac{m}{2}(p_{10}p_{20} + D_0) + \frac{m}{2}(p_{12}p_{22} + D_2)$$

In the production of gametes at meiosis, recombination can only break down the linkage disequilibrium within the pool of non-migrants and the two pools of migrants separately. The frequency of the coupling gametes PP in the gamete pool immediately after meiosis is therefore:

$$freq\,PP'' = (1-m)(p_{11}p_{21} + (1-r)D_1)$$
$$+ \frac{m}{2}(p_{10}p_{20} + (1-r)D_0) + \frac{m}{2}(p_{12}p_{22} + (1-r)D_2) \tag{5.4.7}$$

Reproduction does not alter gene frequencies, and the effect of selection on disequilibrium is assumed to be negligible, so the new linkage disequilibrium becomes:

$$D''_1 = freqPP'' - p'_{11}p'_{21}$$

After simplification, this reduces to:

$$D''_1 = (1-r)D_1 + \frac{m}{2}(1-r)(D_0 + D_2 - 2D_1)$$
$$+ \frac{m}{2}(1-m)(p_{10} - p_{11})(p_{20} - p_{21}) + \frac{m}{2}(1-m)(p_{12} - p_{11})(p_{22} - p_{21}) \tag{5.4.8}$$
$$+ \frac{m^2}{4}(p_{12} - p_{10})(p_{22} - p_{20})$$

The first term describes the reduction in disequilibrium due to recombination, the next the averaging of disequilibria from the three demes due to migration; the last three terms are the linkage disequilibrium generated by the three pairwise combinations of demes. (If the

two neighbouring demes are identical, so $p_{10} = p_{12}$ etc., equation 5.4.8 is identical to that for exchange between 2 demes (Barton & Gale 1993; Li & Nei 1974).)

Equation 5.4.8 can be solved for a "quasi-equilibrium" (Nagylaki 1976b), in which a balance has been reached between the loss due to recombination and the generation by migration. This assumes that differences between the linkage disequilibria in each deme change less than differences in the gametic frequencies, so that $D_0 \approx D_1 \approx D_2$; secondly, the system is taken to be symmetric, so that $\Delta p_1 = p_{12} - p_{11} = p_{11} - p_{10}$ etc.. Equation 5.4.8 then reduces to:

$$D_1 = (1-r)D_1 + m\,\Delta p_1\,\Delta p_2$$
$$= \frac{m\,\Delta p_1\,\Delta p_2}{r} \tag{5.4.9}$$

Using this result to estimate linkage disequilibrium at the centre of a chain of demes, where Δp_i represents the gradient, and hence the inverse of the cline width w, gives:

$$D = \frac{m}{rw^2} \tag{5.4.10}$$

As m is equal to σ^2, the variance in parent-offspring difference, this is exactly equivalent to the prediction for the continuous case (equation 5.4.5).

Note that the above prediction is a factor of (1-m) less than that expected for exchange between 2 demes (Barton & Gale 1993; Li & Nei 1974). This is because migrants enter a deme from two sides, so the genetic make-up of the demes on either sides of a central deme is important. If the neighbouring demes are identical, this collapses to the two deme case, but in a chain of demes model of a cline they will differ in opposite directions, and so the linkage disequilibria generated by exchange between three demes will be greater. The relationship applies to values measured after recombination. If measurements are made after migration, mixing will have increased D by the amount $m\Delta p_1\Delta p_2$ (from equation 5.4.9), and so the QLE approximation will be greater by a factor of (1+r).

The assumptions of quasi-equilibrium, of convergence of D between demes and of entirely negligible selection are equivalent to the weak selection assumption of the continuous model. Simulations can be used to checked how restrictive these assumptions are, and also, for example, whether recombination or migration rates affect the predictions. A simulation model will also incorporate the possible effects of random genetic drift, which can also generate linkage disequilibria; this effect will be of order $(1/N)$, where N is the number of individuals in a deme (Hill & Robertson 1966).

Simulation results

Figure 5.4.1 shows the comparison between the maximum values of linkage disequilibria observed in simulation runs and the prediction $m/(\bar{r}w^2)$ (from equation 5.4.10), where $\bar{r}$ is the harmonic mean rate of recombination between all pairs of loci. See Appendix 5.3 for the method of calculation of average linkage disequilibrium in the simulations. Note that, due to the method, both the estimated and predicted values may be greater than 0.25, the maximum possible value of D; results within these regions are therefore invalid. Results are presented for environmental selection ((a)→(c)) and heterozygote disadvantage ((d)→(f)), and for various linkage scenarios. For unlinked and loosely linked loci, the simulation results fit extremely well to the predicted values for both fitness regimes, despite the assumption of low recombination rates required for equation 5.4.3. As Figure 5.4.1 shows, the correspondence appears to hold even for strong selection (for example, where fitness in the centre of the zone is reduced by up to 80%). However, the predictions break down for tightly linked loci, presumably because the assumption that s/r is small is being violated. Note also that the slightly lower migration rates in the simulations presented in Figures 5.4.1(a) and 5.4.1(d), or the number of loci used, do not affect the correspondence. To summarise, except where recombination between adjacent loci is very low, linkage disequilibrium is determined primarily by dispersal and recombination rates.

(a) Environmental: unlinked
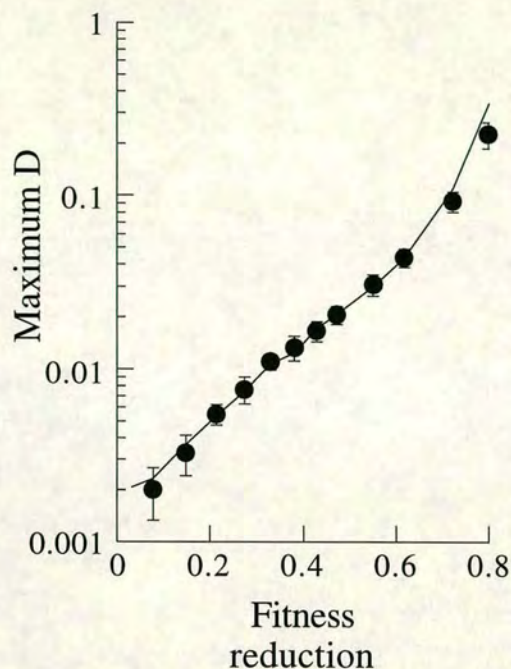
(b) Environmental: loose linkage
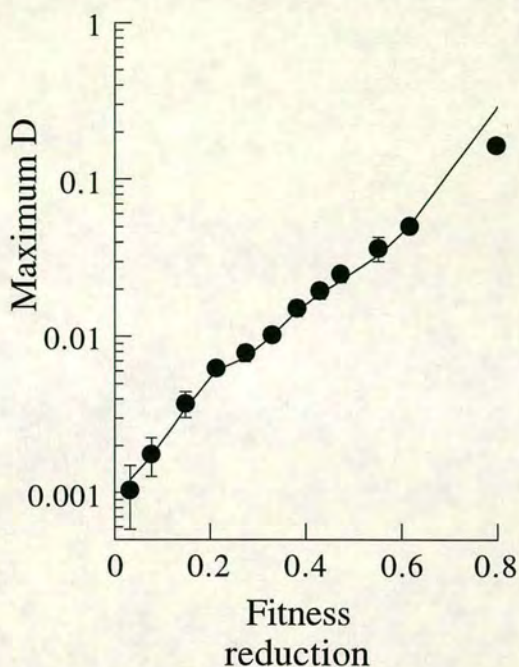
(c) Environmental: tight linkage
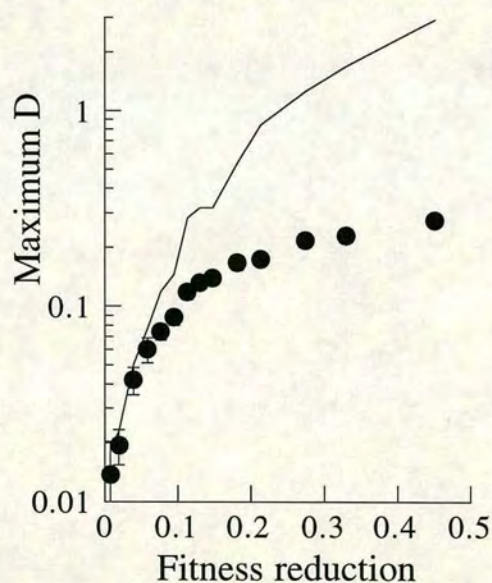
Figure 5.4.1. Maximum linkage disequilibrium D against maximum fitness reduction ($=1-exp(-Ls)$): filled circles = simulation results, with 95% confidence limits; line = prediction from observed width. **(a)** Environmental selection acting on 32 unlinked loci, with σ^2 =0.2, recombination rate between adjacent loci r=0.5; **(b)** Environmental selection on 32 loosely linked loci, with σ^2 =0.5, r=0.2; **(c)** Environmental selection on 8 tightly linked loci, with σ^2 =0.5, r=0.02. (Figure 5.4.1 continued overleaf.)

(d) Heterozygote dis: unlinked
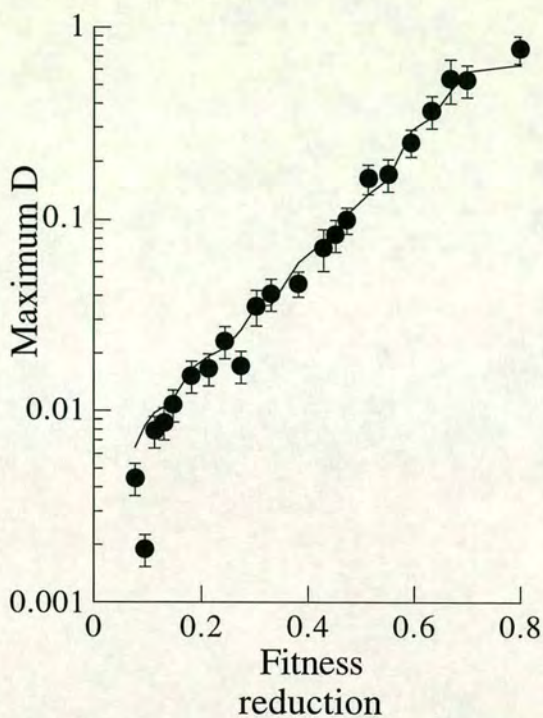
(e) Heterozygote dis: loose linkage
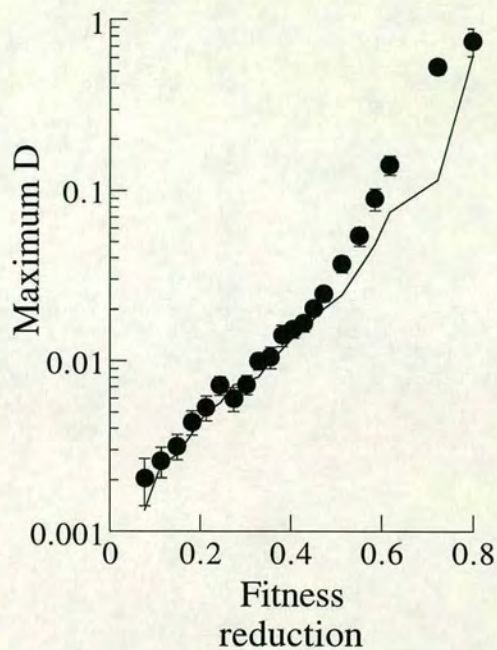
(f) Heterozygote dis: tight linkage
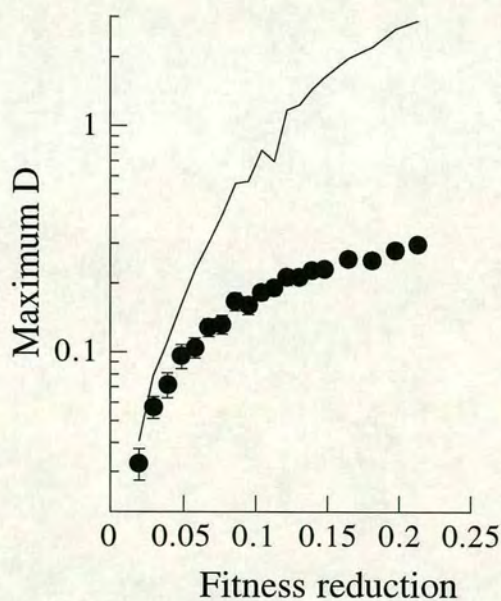
Figure 5.4.1. cont. (d) Heterozygote disadvantage acting on 8 unlinked loci, with $\sigma^2=0.2$, r=0.5; **(e)** Heterozygote disadvantage acting on 32 loosely loci, with $\sigma^2=0.5$, r=0.2; **(f)** Heterozygote disadvantage acting on 8 tightly linked loci, with $\sigma^2=0.5$, r=0.01.

5.5 Consequences of interactions between genes

With the validity of the above description of linkage disequilibrium (equation 5.4.5) confirmed by the simulation results, equation 5.3.5 can be simplified to give, at equilibrium:

$$0 = \frac{\sigma^2}{2} \frac{\partial^2 p_i}{\partial x^2} + \frac{p_i q_i}{2} \frac{\partial \log \overline{W}}{\partial p_i} + \frac{1}{2} \sum_{j \neq i, j=1}^{L} \frac{\partial \log \overline{W}}{\partial p_j} \left(\frac{\sigma^2}{r_{ij}} \frac{\partial p_i}{\partial x} \frac{\partial p_j}{\partial x} \right) \tag{5.5.1}$$

We can now consider the extent to which a genome behaves as a set of independent loci or, at the other extreme, as a single unit under selection. The analysis presented below considers environmental selection, developed from the equivalent treatment of heterozygote disadvantage in Barton (1983) and Barton and Bengtsson (1986), which is summarised in Appendix 5.5.

Mean fitness under environmental selection is given by equation 5.3.7. I consider only the case where x<0 in the following analysis; for x≥0 simply replace p by q throughout (by definition, the two solutions will coincide at p=0.5). Approximating p_i by the arithmetic mean, p, (an assumption of concordance) and r_i by the harmonic mean recombination rate, $\bar{r}$, equation 5.5.1 becomes:

$$0 = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} - spq - \sigma^2 \frac{Ls}{\bar{r}} \left(\frac{\partial p}{\partial x} \right)^2 \tag{5.5.2}$$

Define a "summed coupling" coefficient, $\varphi = Ls/\bar{r}$. Using the same equivalence of $(2 \partial^2 p / \partial x^2) = \partial (\partial p / \partial x)^2 / \partial p$ as in section 5.4 gives a first order differential equation in $(\partial p / \partial x)^2$, which can be integrated to give:

$$\left(\frac{\partial p}{\partial x} \right)^2 = \frac{s}{8\sigma^2} \frac{\left(1 - e^{4p\varphi} (1 - 2\varphi) + 2\varphi(p-q) - 8\varphi^2 pq \right)}{\varphi^3} \tag{5.5.3}$$

(using the boundary condition that the gradient is zero at the edges). The analogous relationship for heterozygote disadvantage is given in Appendix 5.5 (equation A5.5.4).

What do these equations reveal about the shape of clines generated by either selection regime? Integrating the resulting formulae for $\partial x/\partial p$ (using the numerical integration routines of Mathematica 3.0 (Wolfram 1996) and ignoring a common factor of $\sqrt{s/(8\sigma^2\varphi^2)}$) gives the relationship between p and x in terms of φ, allowing a direct comparison of cline shape. This is shown in Figure 5.5.1, which plots logit[2] gene frequency against distance.

On a logit scale, a tanh curve such as that derived for the single locus model (equation 5.2.5) will be approximately linear (see, for example, Figure 2.1 in Barton and Gale (1993)); by contrast, the multilocus clines for both regimes show a stepped pattern on the logit scale (Figure 5.5.1). The step is due to the linkage disequilibria in the centre increasing the effective selection pressures experienced by each locus; I consider this effect further in section 5.7. For the moment, note the close correspondence between the two selection regimes in the central region, but not in the tails where the clines for heterozygote disadvantage fall away more steeply than those for environmental selection. (This can be explained by thinking of the selection acting on a single locus. Fitness was defined so as to give equivalent mean values in the centre for a given selection strength, but this implies that at the edges, where introgressing alleles are rare and therefore likely to be found in heterozygous loci, the average selection experienced *by each locus* is twice as great under heterozygote disadvantage as under environmental selection.) However, Figure 5.5.1 confirms that the difference between cline shape under either selection regime is slight and certainly unlikely to be detected in the field given the usual lack of resolution from samples at the edges of a hybrid zone.

_____

[2]The logit transformation is given by logit(p)=ln(p/(1-p)).
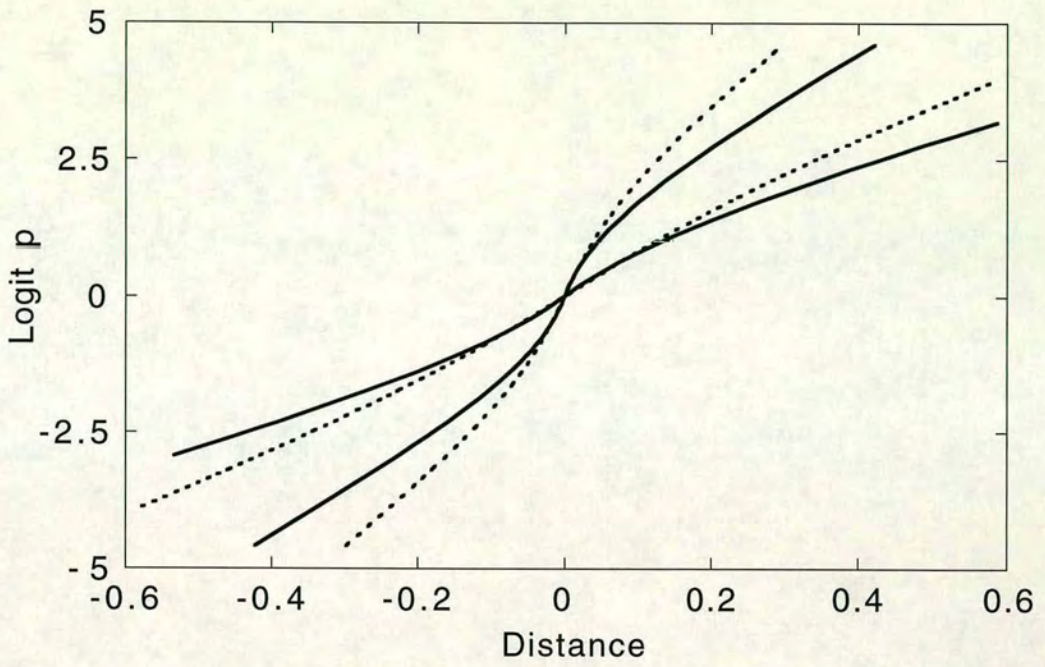
Figure 5.5.1 Logit gene frequency plotted against distance for environmental selection [solid line] and heterozygote disadvantage [dotted line]. Shallow clines represent $\varphi = 0.5$, steeper clines represent $\varphi = 2$. From numerical integration of equations 5.5.3 and A5.5.4; the common factor of $\sqrt{s/(8\sigma^2\varphi^2)}$ is omitted from both equations.

The description of the gradient in equation 5.5.3 can now be used to calculate the width of the cline:

$$width = \left(\frac{\partial p}{\partial x}\right)^{-1}\Bigg|_{p=0.5}$$

$$= \sqrt{\frac{8\sigma^2}{s}}\sqrt{\frac{\varphi^3}{(1-e^{2\varphi}(1-2\varphi)-2\varphi^2)}} \tag{5.5.4}$$

The equivalent equation for a multilocus cline maintained by heterozygote disadvantage is given by: :

$$width = \sqrt{\frac{8\sigma^2}{s}}\sqrt{\frac{\varphi^2}{e^{2\varphi}-2\varphi-1}} \tag{5.5.5}$$

from Barton & Bengtsson (1986: see Appendix 5.5).

Simulation results

Figure 5.5.2(a) and Figure 5.5.2(b) compare the width of the clines observed in simulation runs to this prediction, for environmental selection acting on unlinked and loosely linked loci respectively. There is clearly little correspondence between the two values: the observed width is (almost) always greater than the prediction. However, simulation values fit extremely well to the relevant predictions for heterozygote disadvantage (outlined in Appendix 5.5) for unlinked loci, as shown in Figure 5.5.2(c), and for a range of values when loci are unlinked (Figure 5.5.2(d)).

(a) Environmental: unlinked
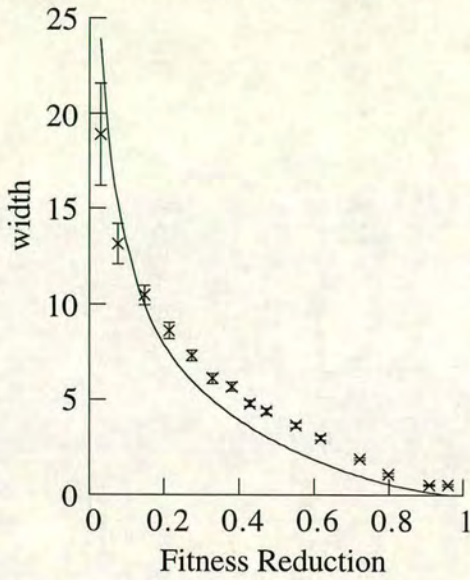
(b) Environmental: loose linkage
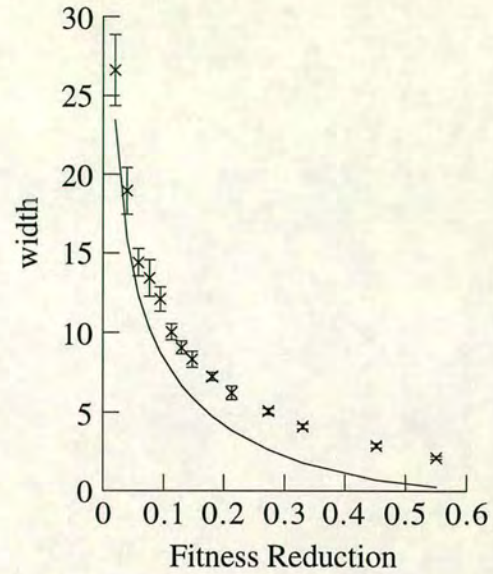
(c) Heterozygote dis: unlinked loci
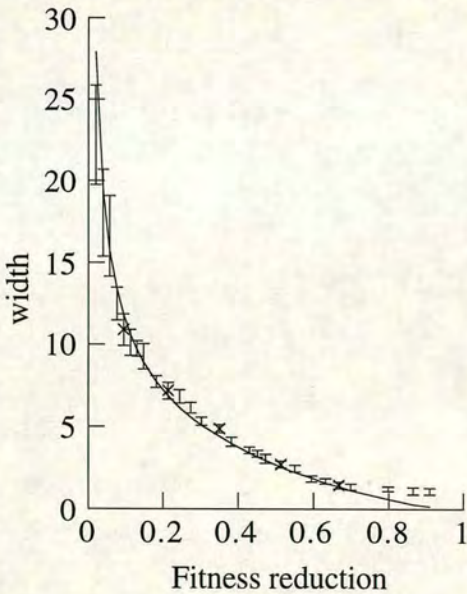
(d) Heterozygote dis: loose linkage
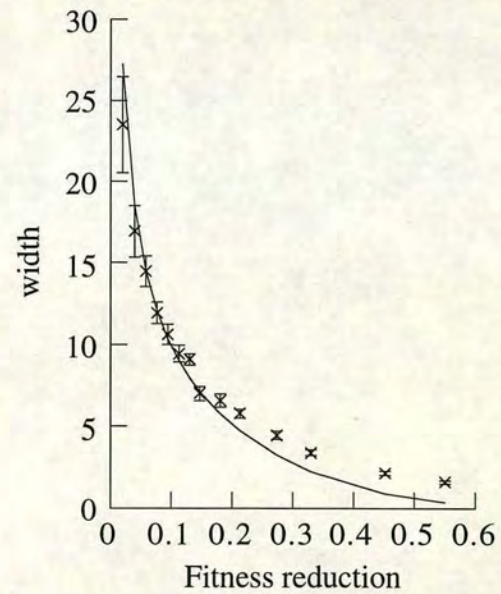
Figure 5.5.2. Cline width against maximum fitness reduction $(=1-exp(-Ls)$, for both selection regimes): crosses = simulation results, with 95% confidence limits; line = predicted width (see text). **(a)** Environmental selection acting on 32 unlinked loci, with $\sigma^2=0.2$, recombination rate between adjacent loci r=0.5; **(b)** Environmental selection on 8 loosely linked loci, with $\sigma^2=0.5$, r=0.1. **(c)** Heterozygote disadvantage acting on 8 unlinked loci, with $\sigma^2=0.2$, r=0.5; **(d)** Heterozygote disadvantage acting on 8 loosely loci, with $\sigma^2=0.5$, r=0.1.

What causes the discrepancy between the values observed in the simulations and those predicted by the analytical theory? Note that the simulation runs were the same as those described in the previous section, where the linkage disequilibrium observed corresponded to that predicted from the observed width (Figure 5.4.2), so the width measurement should be reliable. The key to the discrepancy lies in the difference between the two selection regimes. Width is defined as the inverse of the maximum gradient, at $p=0.5$; this gradient is found by regression of the (logit-transformed) gene frequencies within a central region. Since drift introduces random fluctuations between demes, this central region must necessarily contain several demes; in addition, when clines are narrow, no demes have $p{\approx}0.5$. Figure 5.5.3 shows how the gradient of the cline in gene frequency changes for either selection regime. It is clear that estimating the *maximum* gradient from regression over a central interval will inevitably underestimate the value for environmental selection, but should be more accurate for heterozygote disadvantage.
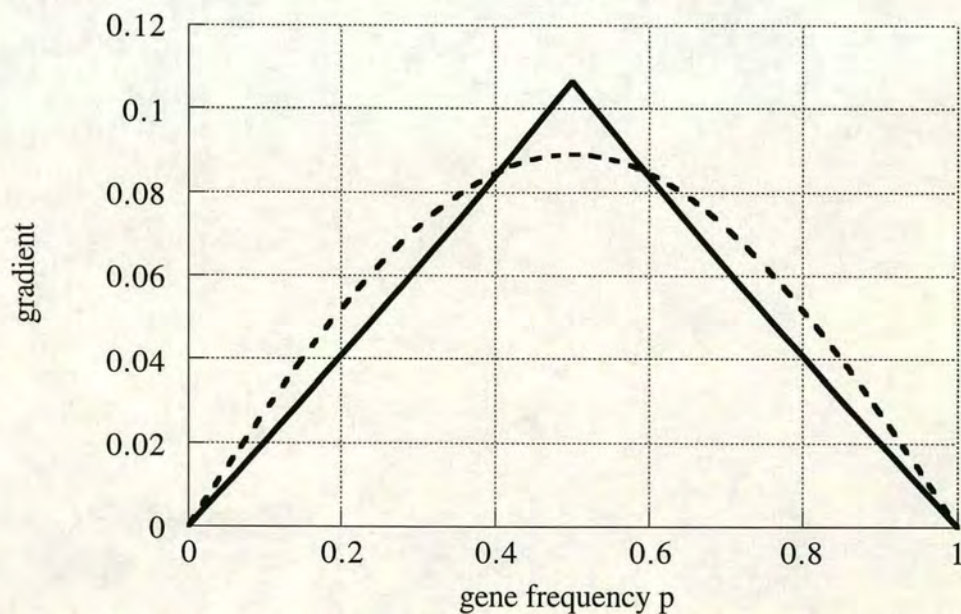
Figure 5.5.3. Gradient (dp/dx) against gene frequency (p) for environmental selection [solid line] and heterozygote disadvantage [dotted line], for 32 unlinked (r=0.5) loci with selection per locus s=0.01 and migration rate $\sigma^2=0.5$.

To confirm that this is indeed the source of the problem, Figure 5.5.4 compares the simulation estimates of gradient in the centre with the predicted gradient at the point at which p=0.4 or, by symmetry, p=0.6; this is the average value of p in the region over which the simulation regression is taken. The observed values fit the predictions, for both types of selection, except at high selection values.

(a) Environmental

(b) Heterozygote disadvantage
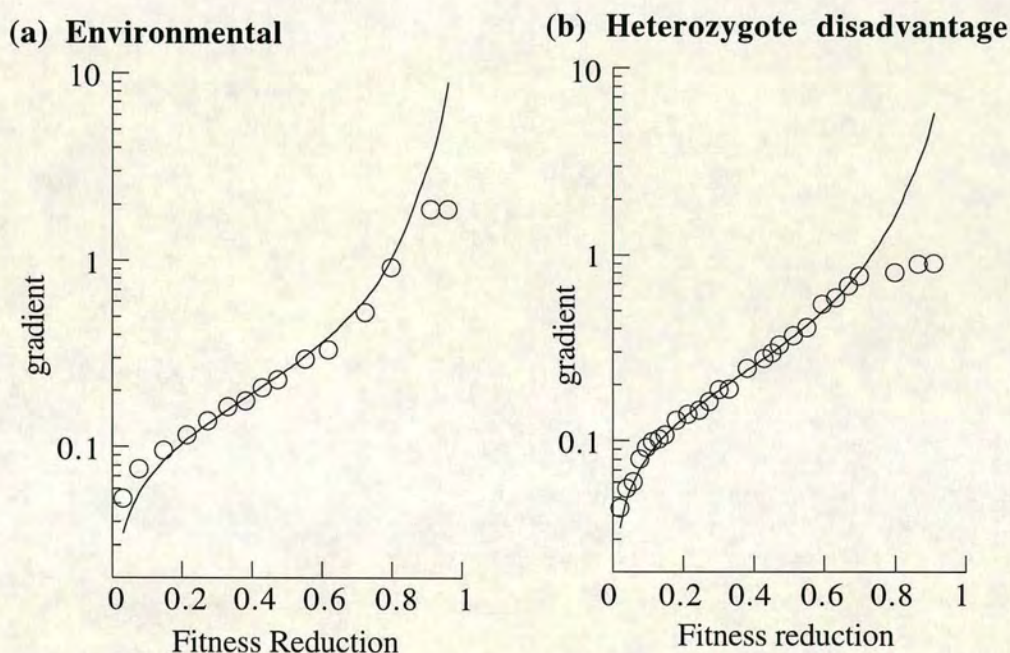
Figure 5.5.4 Predicted gradient of cline at p=0.4 (or p=0.6) compared with central gradient estimated in simulation runs for **(a)** environmental selection on unlinked loci (as for Figure 5.5.2(a)) and **(b)** heterozygote disadvantage (as for Figure 5.5.2(c)).

Comparison with single locus clines

We can now consider the extent to which a genome behaves as a set of independent loci or, at the other extreme, as a single unit under selection. From the single locus model of environmental selection (equation 5.2.7), the width of a cline determined solely by selection at one locus will be $\sqrt{3\sigma^2/s}$; by extension, if a set of L loci behave as a single, unbreakable unit (i.e. recombination is zero) the cline in the frequency of each will have width $\sqrt{3\sigma^2/Ls}$. The analogous values for the heterozygote disadvantage model described here are $\sqrt{4\sigma^2/s}$ and $\sqrt{4\sigma^2/Ls}$. In Figure 5.5.5, I show how clines maintained by selection on several loci show a transition between these two states. As selection (and hence the coupling coefficient φ) increases, the average width in a multilocus system changes smoothly from being equivalent to a single locus cline to that expected if the whole genome acted together: this is apparent in both the analytical predictions and the observed simulated values (although the two do not correspond exactly for environmental selection, for the reasons outlined above). Note that as fitness decreases, the prediction itself becomes invalid as it extends outwith the domain bounded by the two extreme cases.
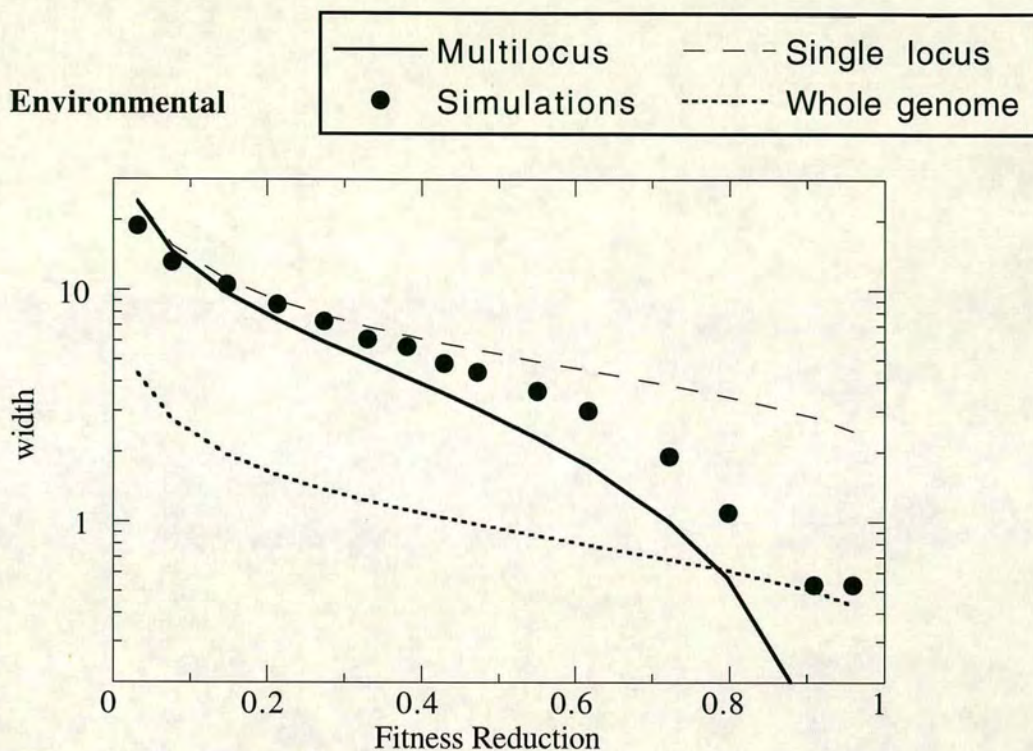
Figure 5.5.5 Observed width for clines in unlinked loci maintained by **(a)** environmental selection (parameters as 5.5.2(a)) **(b)** heterozygote disadvantage (parameters as 5.5.2(c)), compared to prediction for cline in a single locus, under multilocus selection, and if whole genome acts as unit of selection. See text for details.

Effective selection pressure

Equations 5.5.4 and A5.5.5 therefore describe the transition to an increasingly "congealed" genome: as coupling increases, each locus is increasingly affected by selection acting elsewhere in the genome. Define s* as the *effective* selection pressure which would have to act on a single allele to give a cline of the observed width (so s≤s*≤Ls). The ratio s*/Ls gives the proportion of the selection on the whole genome acting on a single locus.

This concept can be used to compare the behaviour of environmental selection with heterozygote disadvantage: Figure 5.5.3 depicts the ratio s*/Ls for the two selection regimes. The relationship becomes meaningless for high values of φ, with values greater than 1 implying that the effective selection strength is greater than the total amount acting; in these regions, the weak selection assumptions are breaking down. More interestingly, the behaviour of the system is remarkably similar for the two different regimes. Thus despite the alternative mechanisms by which selection acts in either case and despite the slightly different spatial scales involved (environmental selection produces a single locus cline of width $\sqrt{3\sigma^2/s*}$ compared to one of $\sqrt{4\sigma^2/s*}$ for heterozygote disadvantage), the degree of interaction within the genome is highly conserved. This is important with respect to the sensitivity of such analyses to the selection regime assumed.
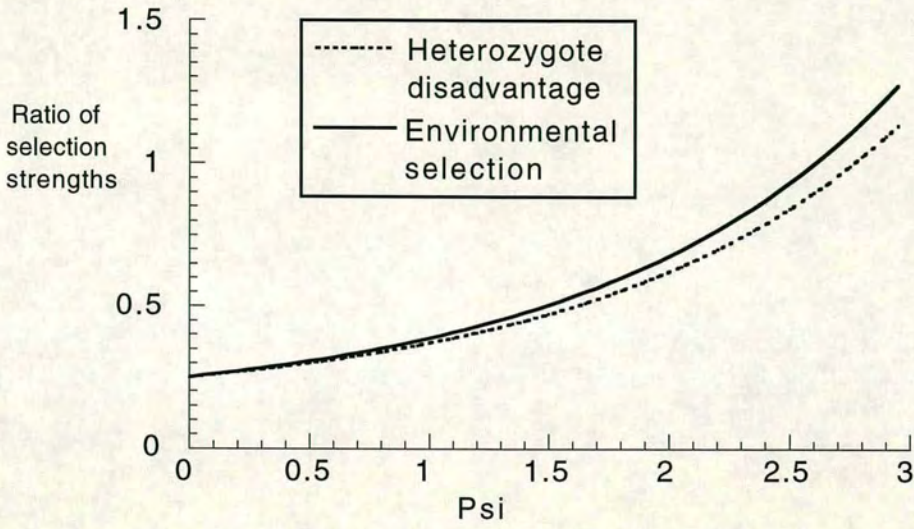
Figure 5.5.3 Ratio of effective selection strength s* to total selection acting, Ls, for both heterozygote disadvantage and environmental selection, for 16 loci, plotted against increasing values of psi, $\varphi = Ls/r$.

5.6 The effect on neutral markers of associated selected loci

The previous sections have shown how the fate of any locus will be affected by selection at other loci. This is particularly useful for inferences about selection across a genome from patterns in genetic markers, such as microsatellites, RAPD's or RFLP's, which are themselves under very weak or no selection. If an effectively-neutral marker entering a population is associated with an unfavoured allele at a separate locus, it will be at a disadvantage itself and its flow into the new population will be impeded. The selection at the other locus therefore generates a barrier to the flow of the marker: the stronger and more widespread the total selection acting, the greater the latter's effect on the resulting cline in the marker's frequency. As shown by Barton and Bengtsson (1986), there exists a conveniently general relationship between the shape of selectively neutral clines and the reduction in mean fitness due to natural selection. From this relationship, information about the strength of the barrier to gene flow (and hence the degree of isolation) between the two populations can be inferred, and even the total number of loci under selection and average selection strengths can be estimated. I outline the derivations of these relationships in this section, again using the simulation model to test the reliability of the theoretical predictions.

Assume that the reproductive isolation between the populations in question is based on many loci (probably a safe assumption for a secondary contact hybrid zone). The chance of a neutral marker being entirely unassociated with any locus under selection is therefore negligible. Consider the frequency of an allele at a neutral locus, defined as u to distinguish it from frequencies of selected loci, and let r_i be the recombination rate between this neutral locus and the i-th selected locus. From equation 5.5.1, the equilibrium solution for the frequency will be:

$$0 = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \frac{1}{2} \sum_{i=1}^{L} \frac{\partial \log \overline{W}}{\partial p_i} \left(\frac{\sigma^2}{r_i} \frac{\partial u}{\partial x} \frac{\partial p_i}{\partial x} \right) \tag{5.6.1}$$

Further assumptions are required at this stage: firstly, that fitness is not explicitly dependent on x (i.e. that within the region x<0, the relationship between gene frequency p and fitness does not change with x) and secondly that r_i can be represented by the harmonic mean recombination rate between neutral and selected loci, $\overline{r_{ns}}$. The equivalence $d\log \overline{W}/dx = \sum_{i=1}^{L} (\partial \log \overline{W}/\partial p_i)(\partial p_i/\partial x)$ can then be used to simplify equation 5.6.1:

$$0 = \frac{\sigma^2}{2} \left(\frac{\partial^2 u}{\partial x^2} + \frac{1}{\overline{r_{ns}}} \frac{\partial \log \overline{W}}{\partial x} \frac{\partial u}{\partial x} \right)$$
$$\Rightarrow 0 = \frac{\partial}{\partial x} \left[\frac{\partial u}{\partial x} \overline{W(x)}^{1/\overline{r_{ns}}} \right] \tag{5.6.2}$$

Integrating gives a relationship between the slope of the neutral marker and the mean fitness in a population at any point across the cline:

$$\left(\frac{\partial u}{\partial x} \right) = \left(\frac{\partial u}{\partial x} \right)_{edge} \left(\frac{\overline{W}_{edge}}{\overline{W}(x)} \right)^{1/\overline{r_{ns}}} \tag{5.6.3}$$

The gradient $\partial u/\partial x$ can be measured both at the edge and centre of the cline; secondly, at the edge, the mean fitness should be 1. Incorporating these facts gives a prediction for the mean fitness at the centre, which by definition will be its minimum value:

$$\overline{W}_{centre}^{1/\overline{r_{ns}}} = \frac{(\partial u/\partial x)_{edge}}{(\partial u/\partial x)_{centre}} \tag{5.6.4}$$

This derivation did not involve a specific selection function; the relationship therefore does not require assumptions about the way in which selection is acting, which makes it useful for estimating the amount by which fitness at the centre of the zone is reduced.

Equation 5.6.4 was tested using simulations with environmental selection. The results are shown in Figure 5.6.1, and show a fairly good agreement within the central range of values considered. The data are noisy because of the difficulty in estimating the low values of the gradient at the edges.

(a)
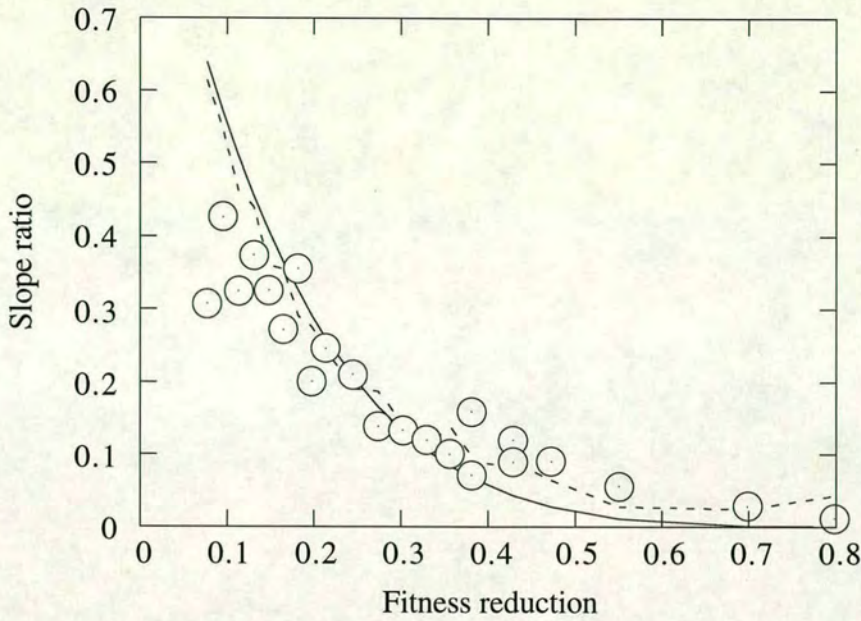
(b)
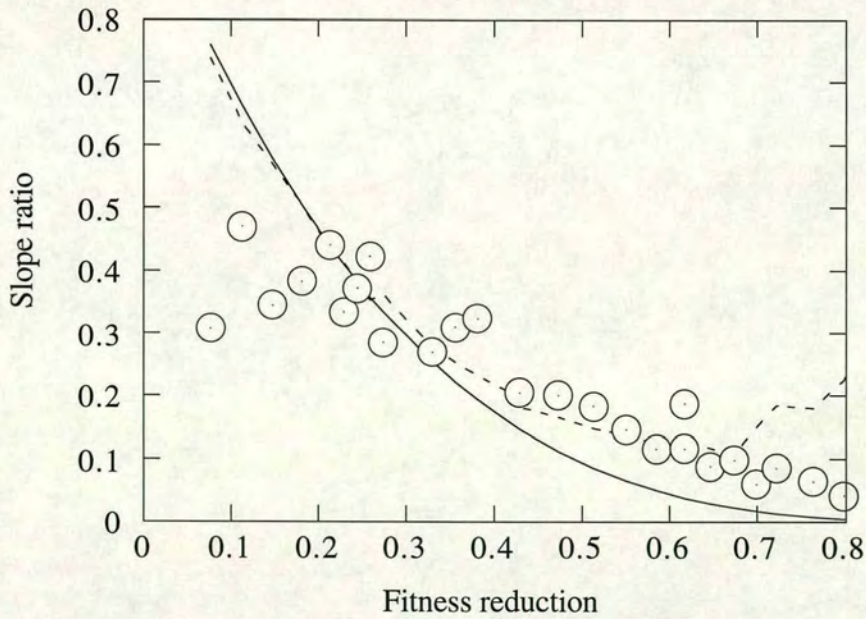
Figure 5.6.1 Ratio of slopes at edge and centre in simulation of environmental selection for neutral loci compared to $\overline{W}_{min}{}^{\wedge}(1/\overline{r_{ns}})$, where $\overline{W}_{min}$ is defined as either the observed minimum fitness in the simulation [dotted line] or the predicted minimum fitness $exp(-Ls)$ [solid line]. Simulations involved 16 neutral loci interspaced between 16 selected loci. **(a)** $\overline{r_{ns}}$=0.18; **(b)** $\overline{r_{ns}}$=0.30.

5.7 Barrier strength and the number of genes under selection

Defining the magnitude of the barrier to gene flow

Figure 5.7.1 shows simulation data on changes in the mean fitness and the corresponding frequencies of neutral and selected loci across a cline: both clines show sharp steps in the region of minimum fitness. This is because, within the central region, linkage disequilibrium increases the effective selection on all loci above its value at the edges. The effect on clines in neutral loci is the same as that induced by a physical barrier to migration (e.g. *Podisma*: Jackson 1992), and can therefore be thought of as a genetic barrier (see also Bengtsson 1985). In both cases, a stronger barrier will generate a larger step relative to the cline gradient. Formally (after Nagylaki 1976a), the strength B of the barrier can be quantified as:

$$B = \frac{\Delta u}{\left(\partial u / \partial x\right)_{edge}}$$

(5.7.1),

where Δu is the magnitude of the step in frequency of a neutral marker. If the spatial dimension is continuous, B has the dimensions of distance (Barton & Bengtsson 1986), and can be thought of as the distance of unimpeded habitat which would present the same obstacle. Introgression of neutral alleles into the alternative population will be delayed by $(B/\sigma)^2$ generations (Barton 1979b; Piálek & Barton 1997). If the barrier is strong, this may be an evolutionary significant time period, but for simplicity I consider only analysis of the system once it has reached equilibrium.
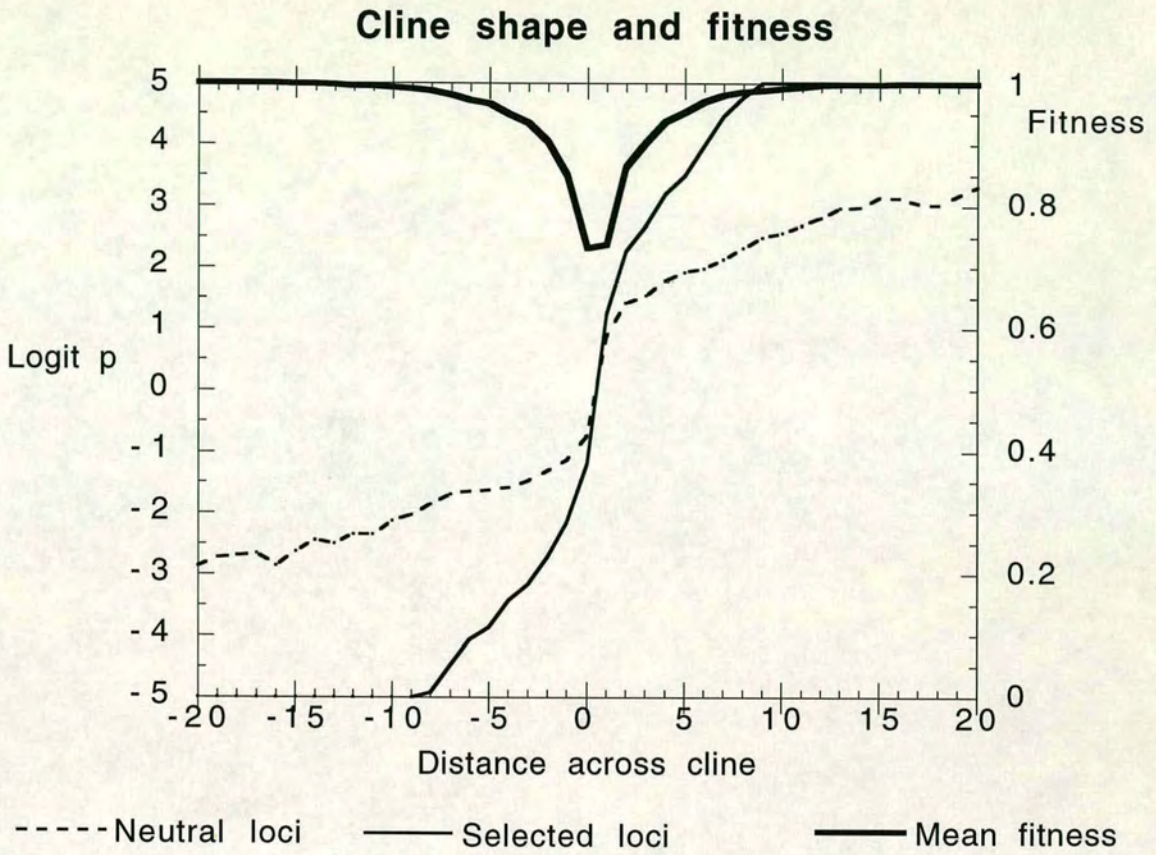
Figure 5.7.1. Data from simulation of clines maintained by environmental selection, showing change in mean frequency of 10 neutral loci and 60 selected loci across the cline: selection per locus $s=0.02$; $\bar{r}=0.20$; $\sigma^2=0.5$. Frequencies are measured 1000 generations after initial contact.

The magnitude of the step, Δu, can be determined by comparison of the value of u expected in the absence of a barrier to that observed in the presence of one; see Figure 5.7.2. Consider a point x_1 somewhere to the right of the centre, where distance is measured from the left-hand edge, and assume that, outside the region of reduced fitness, neutral clines can be approximated by a linear function with gradient u'_{edge}. In the absence of a barrier the gene frequency at x_1 would be:

$$u(x_1)_{without\ barrier} = u'_{edge}\, x_1 \qquad (\because u(0) = 0) \qquad (5.7.3)$$

With a barrier, integrating equation 5.6.3 (and using the condition $\overline{W}_{edge} = 1$) gives:

$$u(x_1)_{with\ barrier} = u'_{edge} \int_0^{x_1} \overline{W}(x)^{-1/r}\, dx \qquad (5.7.4)$$

This defines the step:

$$\begin{aligned}
\Delta u &= u(x_1)_{with\ barrier} - u(x_1)_{without\ barrier} \\[2mm]
&= u'_{edge}\left(\int_0^{x_1} \overline{W}(x)^{-1/r}\, dx \right) - u'_{edge} x_1 \\[2mm]
&= u'_{edge} \int_0^{x_1} \left(\overline{W}(x)^{-1/r} - 1 \right) dx \\[2mm]
&= u'_{edge} \int_0^{\infty} \left(\overline{W}(x)^{-1/r} - 1 \right)\, dx
\end{aligned} \qquad (5.7.5)$$
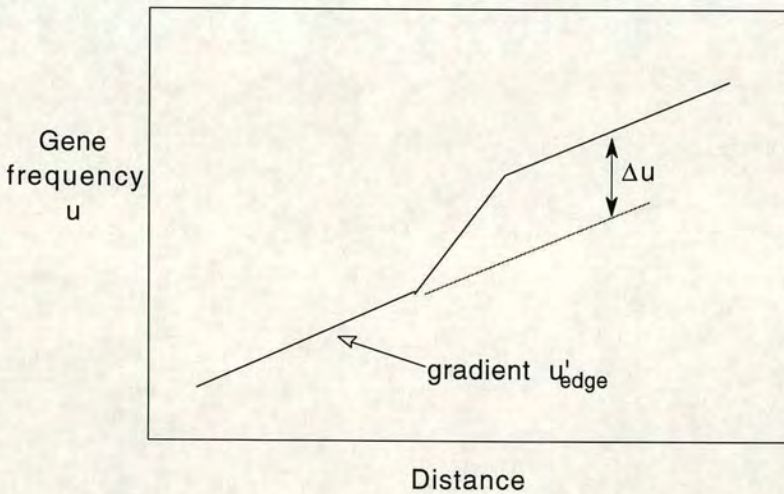
Figure 5.7.2 Definition of Δu as difference between observed frequency u and that expected in the absence of a barrier.

The last line depends on the fact that to the right of the value x_1, fitness is 1 and so the integrand will always be 0. This formula for Δu can be substituted into equation 5.7.1:

$$B = \int_0^\infty (\overline{W(x)}^{-1/\overline{r}} - 1)\,dx = \int_0^1 \frac{(\overline{W(p)}^{-1/\overline{r}} - 1)}{p'}\,dp \qquad (5.7.6)$$

Determining the number of genes under selection

The above relationships reduce to the following two equations:

$$\overline{W}(0.5)^{1/\overline{r}} = u'_{edge}/u'_{centre} \qquad \text{(from 5.6.4)} \qquad (5.7.7a)$$

$$\int_0^1 \frac{(\overline{W}(p)^{-1/\overline{r}} - 1)}{p'}\,dp = \Delta u/u'_{edge} \qquad \text{(from 5.7.1 and 5.7.6)} \qquad (5.7.7b)$$

At this stage a selection regime must be specified (thus defining $\overline{W}$ and p' in terms of $\varphi\,(= Ls/\overline{r})$ and s). The observed measurements of cline shape (Δu, u'$_{edge}$ and u'$_{centre}$) and an estimated value for $\overline{r}$ can then be substituted into 5.7.7, giving two equations in the only remaining unknown parameters, L and s. The equations can be solved simultaneously to give a full description of the system. I illustrate the method using the formulae for environmental selection.

Under environmental selection, $\overline{W}(p) = exp(-2\,Lps)$ for p<0.5 etc. and p' is as defined in equation 5.5.3. Equation 5.7.7 then becomes:

$$e^{-\varphi} = u'_{edge}\big/u'_{centre} \qquad (5.7.8a)$$

$$2\sqrt{\frac{8\sigma^2\varphi^3}{s}} \int_0^{0.5} \frac{e^{2\varphi p} - 1}{\sqrt{1 - (1 - 2\varphi)e^{4\varphi p} + 2\varphi(p - q) - 8\varphi^2 pq}}\,dp = \Delta u\big/u'_{edge} \qquad (5.7.8b)$$

Equation 5.7.8a can be solved explicitly for φ; 5.7.8b must then be solved by numerical integration (using Mathematica 3.0 (Wolfram 1996)) to give s. Combining these with an estimate of the average recombination rate $\overline{r}$, the number L of loci under selection follows from the definition $\varphi = Ls/\overline{r}$.

Simulation results

In Figure 5.7.3, the observed barrier strength ($B = \Delta u / u'_{edge}$) is compared to its predicted value as given by the left-hand side of equation 5.7.8b. Observed values of B are consistently higher than the prediction. The data used are from the same simulation runs as Figure 5.6.1, which showed a general agreement between the theoretical predictions and the simulation results. There are two possible reasons for the discrepancy. Firstly, the algorithm used in the program to calculate B may require modification. Because of the agreement in Figure 5.6.1, it is assumed that the calculation of the slope at the edge is accurate; however, the simulation model differs most radically from the theoretical representation in this area, so it is not clear whether the discrepancy could lie either in its calculation or in that of Δu. Secondly, the approximation in Equation 5.7.3 of clines to a linear function in the absence of a barrier may be dubious; it is suspicious that the predicted value for the barrier strength does not increase with increasing selection strengths. These aspects require further work. As implied by the lack of correspondence between the observed barrier strength and the theoretical predictions, any estimates of the number of loci under selection using these simulation data will be meaningless and are not presented.
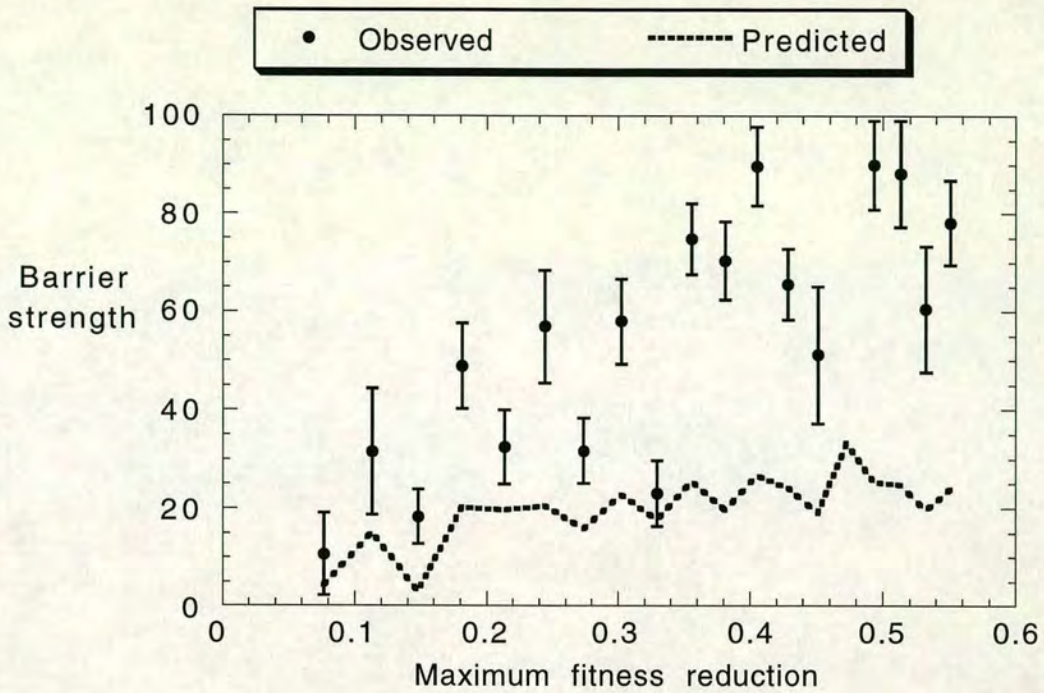
Figure 5.7.3 Barrier strength for environmental selection: observed value of $\Delta u/u'_{edge}$ and predicted value from equation 5.7.8b. Simulations involved 20 neutral loci spaced between 60 selected loci, with $\bar{r}=0.18$ and $\sigma^2=0.5$.

Comparison of the two selection regimes

Prior to the calculation of the number of genes under selection, estimating parameters of interest (for example, the mean fitness in central populations) from the cline shape did not require any selection function to be specified. Assuming for the moment that the discrepancies observed above are due to a problem with the statistical algorithms of the simulation model, which when modified would confirm the equivalence in equation 5.7.7b, how robust would predictions be if the wrong fitness function were specified?

Consider the integrand of equation 5.7.7b, which defines the barrier strength. Figure 5.7.5 compares the magnitude of this for environmental selection vs. heterozygote disad-

vantage (using the formula given in Appendix 5.5 for the latter). The difference between

the alternative regimes appears considerable, and suggests that even were the above

method reliable, it is not robust to misassumptions about the mechanism of selection. The

curves in Figure 5.7.5 are derived from selection strengths which define identical mean

fitness in a central population, but the integral under the heterozygote disadvantage curve

is larger. Even without differences in the respective height of the curves, their different

shapes will introduce discrepancy. This contrasts with the earlier analysis of the propor-

tion of the genome acting together (section 5.4), which suggested that dynamics under

the alternative selection regimes were similar. However, numerical integration shows that

the factor difference in size of the respective integrals (for a given number of loci and

selection strength) will always lie between 1.28 and 1.41 for all φ, which would not

affect estimates by an order of magnitude.
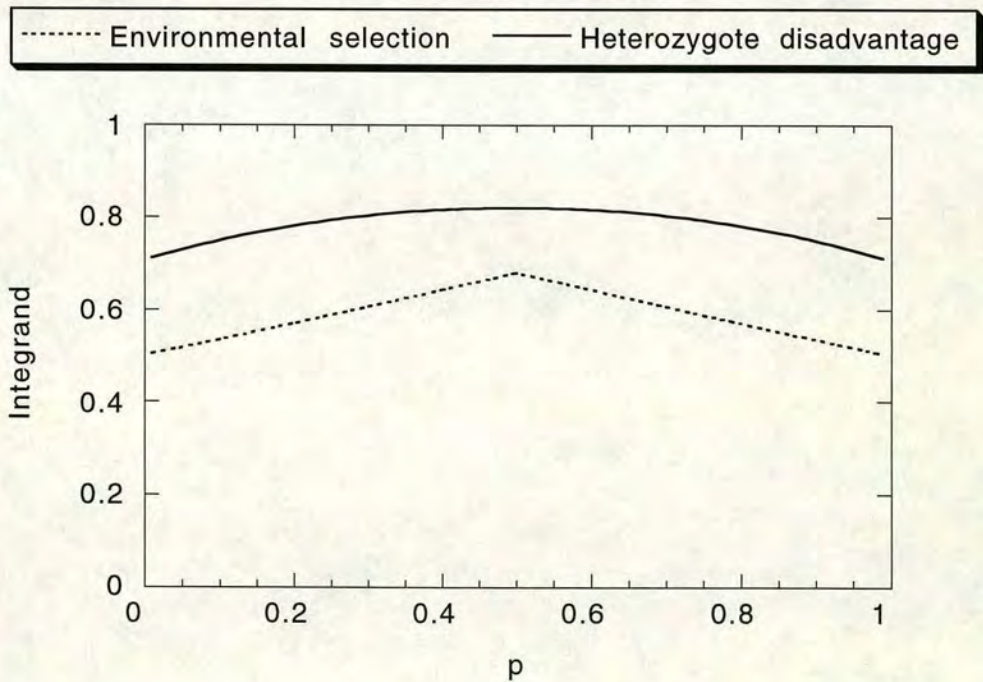
Figure 5.7.5 Integrand defining barrier strength in equation 5.7.7b, for environmental
selection (equation 5.7.8b) and heterozygote disadvantage (appendix 5: equations A5.1
and A5.3); $\varphi=1$, and the common factor $\sqrt{8\sigma^2/s}$ is excluded.

5.8 Discussion

The model developed in this chapter is analogous to that created by Barton (1983, 1986; Barton & Bengtsson 1986) to describe clines maintained by a balance between the homogenising effects of dispersal and recombination and selection acting to maintain the integrity of the parental genomes. However, whereas Barton considered selection acting against hybrids (in the form of heterozygote disadvantage), the selection regime invoked here defines differential adaptation across an environmental transition, with different alleles favoured on either side of a central point. The analytical results confirm the conclusion of previous single (or two) locus models, that the clines produced by either selection regime are indistinguishable in shape. The theoretical representations all require an assumption that selection is a relatively weak force. This may seem unduly restrictive, particularly in the application of the theory to data from natural populations, but note, for example, that if an allele crossing into the other environment type reduces an individual's fitness by only 1%, the resulting cline in its frequency will be equal to only 5.48 times individual average dispersal distance. Because linkage disequilibrium increases the effective selection on each locus, this effect is even more pronounced when more loci are considered. Comparisons of the analytical predictions with results from simulations also imply that the weak selection approximation holds for an apparently quite wide range of fitness values (e.g. Figure 5.4.1).

Barton's multilocus representation differs from single locus models in a number of ways which afford inferences into the magnitude of reproductive isolation between two populations. The *Bombina* hybrid zone has been thoroughly analysed in this way: see section 1.6. Firstly, under the weak selection approximation, associations between loci (linkage disequilibrium) will be generated by the mixing of populations with different gene frequencies. As shown by equation 5.4.5, knowledge of linkage disequilibrium and

cline width can give an indirect estimate of average dispersal rates. Secondly, the ratio of the slope of the cline at the edges and at the centre reflects the mean fitness in central hybrid populations: if there is a substantial reduction in fitness, the cline will be much steeper at the centre than the edges, and so forth. This estimate requires no assumption about the ways in which selection acts, and the analytical predictions fit (reasonably) well with simulation results.

Linkage disequilibrium between loci increases the effective selection pressure experienced by each locus, generating a central step in gene frequency (see Figure 5.5.1) not realised in single locus models. The step allows an estimate of the magnitude of the barrier impeding flow of a neutral marker created by selection at other loci, and hence an estimate of the number of selected loci creating the effect. This is an important result, giving an indication of the number of loci responsible for reproductive isolation and hence shedding light on the genetics of speciation. Although the calculations require a fitness regime to be specified, the above results imply that the difference between estimates derived under either selection regime is not substantial. Unfortunately, at this stage of the analysis, the correspondence between the predictions and the simulation results breaks down. Because the issue concerns one of the most important inferences from analysis of hybrid zones, it is hoped that further work will clarify the source of the discrepancy.

A final point should be made with respect to the assumption of the system reaching an equilibrium state. In an alternative approach to analysing the consequences of interbreeding, Baird (1995) uses Fisher's (1953) junction theory to track the fate of introgressing blocks of genetic material entering a population. Baird's simulation models show that the approach to an analytic equilibrium is slow, of the order of thousands of generations. The results from the simulations presented here fit more closely with the equilibrium predictions within a shorter time span, and were also checked to ensure that

statistics were measured after a time point at which the cline had stabilised. The difference between these conclusions is probably due to the difference in the rates of recombination modelled. Baird considers the fate of a single chromosome of infinitely many loci, implying that recombination will be much lower than in the model chosen here, where recombination rates were chosen to represent either unlinked or loosely linked genes. Further comparison of the two systems would be interesting.

Appendix 5.1 Definition of mathematical symbols

Symbol	Definition	Range
x	geographic distance	$(-\infty < x < \infty)$
p_i	frequency of allele at ith (selected) locus	$(0 \leq p_i \leq 1)$
q_i	(1-p)	
$p_i{'}$	$\partial p_i / \partial x$	
u	frequency of neutral marker	$(0 \leq u \leq 1)$
v	(1-u)	
σ^2	variance in parent-offspring distance	$(0 \leq \sigma^2)$
s	selection acting per allele or locus	$(0 \leq s)$
r	recombination rate between two loci	$(0 < r \leq 0.5)$
$\bar{r}$	harmonic mean recombination rate	$(0 < \bar{r} \leq 0.5)$
$\overline{W(x)}$	mean fitness in deme at position x	$(0 \leq \overline{W(x)} \leq 1)$
N	number of individuals per deme or population	$(N=2,3,....)$
L	number of loci	$(L=1,2,....)$
D	pairwise linkage disequilibrium	$(0 \leq D \leq 0.25)$
θ	s/r	
φ or psi	$Ls / \bar{r}$	

Appendix 5.2 Simulation model

The simulation model is based on a program written by Gale and Barton (Barton & Gale 1993; Jackson 1992), to whom I am grateful for providing me with their program code. A one-dimensional stepping stone model is used to simulate contact between two populations initially fixed for two alternative alleles at each of L loci. The loci may be linked (so recombination between adjacent loci occurs with probability r<0.5) or unlinked (r=0.5), and each gene may be either neutral or under selection. The respective number and positions of neutral and selected loci can be varied.

The population is distributed along a chain of demes, each of which contains N diploid individuals. Migration occurs between adjacent demes, with a fixed proportion m of individuals moving per generation, half in either direction; migrants are chosen at random. With this model of nearest-neighbour migration, the migration rate m is identical to the variance of distance moved per generation, σ^2.

Migration is followed by reproduction. Recombination during gamete production occurs at a rate r between adjacent loci, with chiasma positions randomly determined. Within each deme, individuals then mate randomly, and each pair produces K offspring. Selection, of specified strength, determines offspring viability; selection regimes are defined below. Selection is soft (Levene 1953), so that a total of N offspring are chosen from the offspring pool of N*K/2 individuals. These form the new adult population, which migrate, reproduce and so forth. Unless otherwise stated, N=20 and each adult pair produces K=2 offspring in the simulation runs presented here.

End demes remain fixed in gene frequency, to simulate an infinite pure population on either side of the cline. The total number of demes used in each simulation was varied

according to the overall strength of selection involved (and hence the expected width of the cline), to ensure that end effects will not disrupt the dynamics. Statistics are calculated after migration and reproduction; this is analogous to sampling offspring before they leave the sites in which they were born. Unless otherwise stated, statistics are calculated at 100 generation intervals between generations 500 and 3000 of each run; results were checked to insure that (i) the system had reached equilibrium; (ii) these intervals were long enough to avoid autocorrelation, and hence to allow values to be treated as independent data points. The 95% confidence intervals presented in most graphs are taken as twice the standard error of a statistic. Appendix 5.3 describes the statistical calculations; see Appendix 5.4 for tests that the component procedures are performing as expected.

Selection regimes

In the L selected loci of each individual, let the number of alleles originating from the population to the right of the environmental step be i, and the number of heterozygous loci be h. Individual fitnesses for the two possible selection regimes considered are then defined as:

Environmental selection:

$$W = \begin{cases} e^{-is} & x < 0, \qquad i = 0\ldots 2L; \\ e^{-(2L-i)s} & x \geq 0. \end{cases}$$

Heterozygote disadvantage: $$W = e^{-2hs} \qquad h = 0\ldots L.$$

Note that the selection factor in heterozygote disadvantage is greater by a factor of 2 than the function used by Barton (1983; also Bazykin 1969). This is so that, for given values of s and L, the average fitness in a population at the centre is equal for the two selection regimes (e^{-Ls}).

Appendix 5.3 Statistical methods for simulations

A5.3.1 Cline gradient and width

A single locus cline should roughly follow a tanh curve between the points at which
$\ln(p/q)=-2$ and $\ln(p/q)=2$; over this region, the logit-transformed frequency will therefore
be close to linear, allowing its gradient to be calculated by regression. The gradient
calculated from the logit transformation will be exactly four times the gradient of the non-
transformed frequency at the centre ($p=0.5$). The width of a cline at a single locus is
defined as the inverse of its maximum gradient. In the simulation runs, the width of each
cline is calculated; the widths of all clines are then averaged to give a mean value. (Note
that this gives a different estimate of width than estimating the maximum gradient of the
cline in mean gene frequency; the latter will considerably underestimate the rate of change
when, due to random drift, clines are not centred at the same point.) All statistics are
calculated separately for neutral or selected loci.

A5.3.2 Linkage disequilibrium

An estimate of the average pairwise linkage disequilibria of the individuals in each deme
can be calculated using the variance in a hybrid index (Barton & Gale 1993; see section
2.5). Define the hybrid index H of an individual as the number of type 1 alleles it
possesses out of the total possible 2L: $H = \sum_{i=1}^{L} x_i$, where $x_i=0$, 1 or 2. The variance in
H will then be given by:

$$Var(H) = \sum_{i,j=1}^{L} cov\left(x_i, x_j\right) = 2\left(\sum_{i=1}^{L} p_i q_i + \sum_{i \neq j} D_{i,j}\right)$$

$$\Rightarrow \overline{D} = \frac{1}{2n(n-1)}\left(Var(H) - 2\sum_{i=1}^{L} p_i q_i\right)$$

(A5.3.1)

The variance in H will be inflated by any heterozygote deficit. This is because the
possibility of autozygosity causes correlations between homologous genes at one locus in

addition to the correlations between loci measured by the linkage disequilibrium (Nürnberger *et al.* 1995); see Chapter Two. In this chapter, statistics are measured immediately after random mating, so gene frequencies are in Hardy-Weinberg proportions and values of linkage disequilibrium estimated using this method will not be inflated. The effect of non-random mating is considered in Chapter Six.

The maximum linkage disequilibrium, at the centre of the cline, is estimated from a linear regression. This uses the fact that the gradient of a tanh curve is proportional to pq, so D should be proportional to its square; a regression of D on $(pq)^2$ can therefore be used to give a value of D at p=0.5, and gives a better estimate than relying on the maximum D or the value nearest the centre. However, with increasingly strong selection pressures, cline shape shows an increasing departure from the standard tanh curve, nullifying the above argument, and nonsensical values of D (namely, greater than 0.25) may be returned. The prediction $m/\bar{r}w^2$ is calculated using mean value of width estimated as described above.

A5.3.3 Barrier strength and ratio of cline slopes

The mean fitness in central populations is defined by the ratio of gradients at the edge and centre of the cline respectively (equation 5.6.4). Gradients at the edge are found by a linear regression of logit-transformed gene frequencies over the region between fixation and the point at which fitness is reduced by 5% of its total fall; assuming symmetry, the average of both sides is used. The central gradient is found by regression over 10 demes centred on the deme in which fitness is minimum. The step in gene frequency, required to quantify the strength of the barrier to gene flow (equation 5.7.1) is estimated by extrapolation of the linear regressions at the edges (see Figure 5.7.2).

Appendix 5.4 Checking the simulations

The simulation model was tested at each stage to check that it was performing as expected; I present here two examples of tests. Firstly, to check the behaviour of the stochastic processes, I consider the decay of linkage disequilibrium D with time: in the absence of selection and migration, recombination and random drift reduce D in a diploid by a factor of (1-r-1/2N) (Hill & Robertson 1966) with each generation, where r is the recombination rate and N is the number of individuals per deme. Figure A5.4.1 compares this prediction for the decay in D with the mean observed in simulations of 100 populations (demes start with 50% of individuals fixed for one allele type, 50% for the other, generating maximum possible values D). The effect of the random number generator was also followed by checking that genotype frequencies were in Hardy-Weinberg proportions after mating, and that, with no selection or migration, the probability of fixation of a deme equalled the starting gene frequency.

The analytical theory assume continuous time, with the effect of the various processes (migration, reproduction, selection) occurring simultaneously, and continuous spatial dimensions, with migration represented by a diffusion process; population size is effectively infinite. In contrast, the simulation uses discrete generations, with the respective processes occurring in a prescribed order, and the population distributed over discrete demes, linked by nearest-neighbour migration. The equivalence of the two systems needs to be established before the simulations can be used to test more elaborate theory; I do so by testing the firmest of the analytical predictions.

Under environmental selection, from equation 5.2.4, the predicted width of a single locus cline will be $\sqrt{3\sigma^2/s}$. In a chain of demes, the variance in distance moved per generation, σ^2, is given by the migration rate m. Figure A5.4.2 compares the width

observed in simulations with those predicted; there is a reasonably good fit to the theory, although the observed values are consistently greater than the predictions (I return to this point in section 5.5). We can therefore be confident that the model is providing a good approximation to a diffusion process in continuous time and space, and even appears, at this stage, robust to some violation of the weak selection assumption.
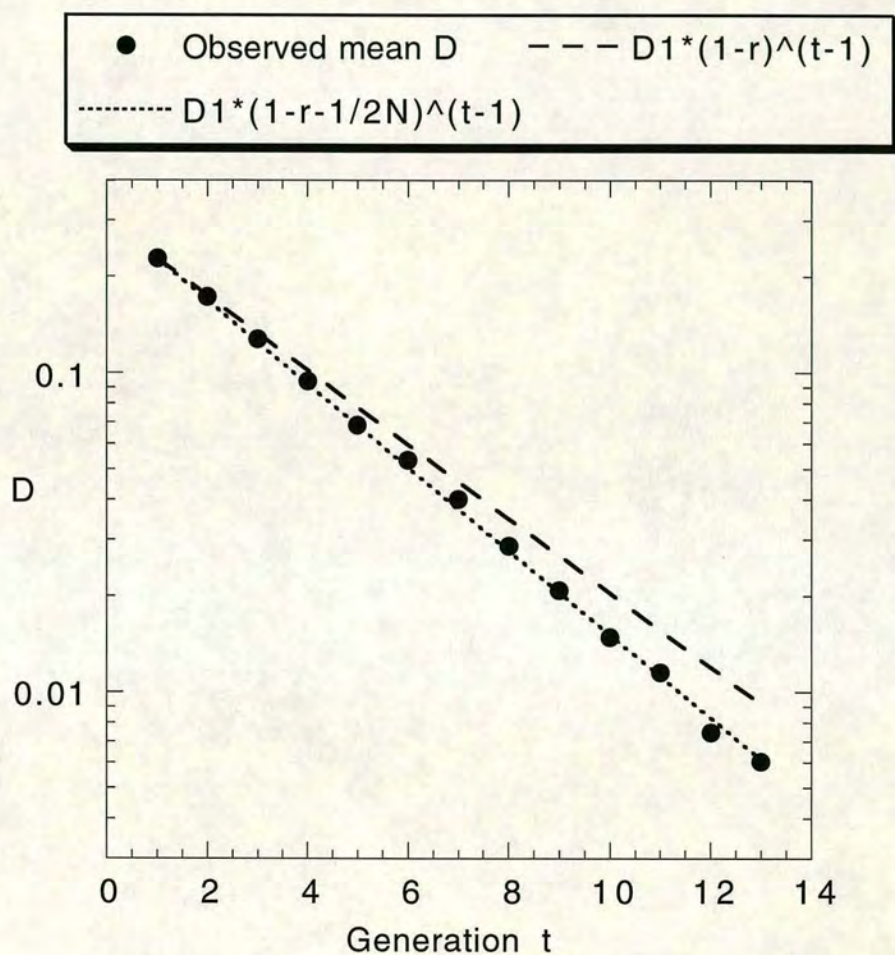
Figure A5.4.1 Decay of linkage disequilibrium between two neutral loci with recombination rate r=0.25. Mean D is the arithmetic mean of values observed in 100 demes, each containing N=40 individuals (no migration between demes); D1 is the value observed in the first generation. Dashed line: prediction without correcting for the effect of random drift; dotted line: predictions after correcting for drift observations show close correspondence to the former.

Width of single locus cline:
environmental selection
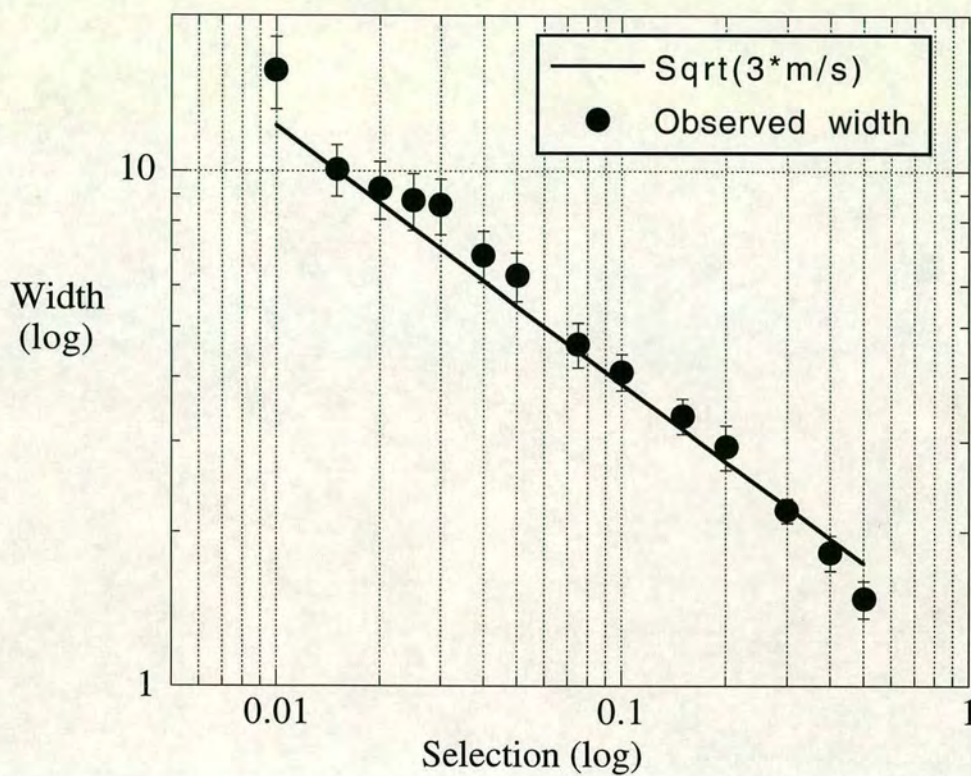
Figure A5.4.2 Comparison of observed width (with 95% confidence limits) of cline in single locus, under environmental selection (equation 5.2.5), with prediction of width =sqrt(3*m/s). Migration rate m=0.5; number of diploid individuals per deme N= 40.

Appendix 5.5 Multilocus analysis of heterozygote disadvantage

I present here the treatment of a multilocus cline as described in Barton & Bengtsson (1986; see also Barton 1983, 1986), under endogenous selection generated by heterozygote disadvantage. The notation used here differs from theirs as follows: L (rather than n) represents the number of loci (to distinguish from population size); selection of strength 2s here is equivalent to their strength s (so that for given values of s and L, the mean fitness in a central population is equal to that under a model of environmental selection; see Appendix 5.2). Under heterozygote disadvantage:

$$\overline{W} = exp(-4\Sigma p_i q_i\ s) \quad \Rightarrow \quad \partial \log \overline{W}/\partial p_i = 4(p_i - q_i)s \qquad (A5.5.1)$$

Assuming concordance, using the harmonic mean recombination rate and defining $\varphi = Ls/\overline{r}$, equation 5.5.1 becomes:

$$0 = \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2} + 2pq(p-q)s + 2\sigma^2\varphi(p-q)\left(\frac{\partial p}{\partial x}\right)^2 \qquad (A5.5.2)$$

Substituting as before (section 5.5) gives a differential equation which can then be solved with an integrating factor:

$$\frac{\partial\left[(\partial p/\partial x)^2 e^{-8\varphi pq}\right]}{\partial p} = \frac{8s}{\sigma^2}\int -pq(p-q)e^{-8\varphi pq}dp \qquad (A5.5.3)$$

$$\Rightarrow \quad \left(\frac{\partial p}{\partial x}\right)^2 = \frac{s}{8\sigma^2}\frac{1}{\varphi^2}\left(e^{8\varphi pq} - 8\varphi pq - 1\right) \qquad (A5.5.4)$$

Figure 5.5.1 compares the solution for this equation with that of the analogous equation for environmental selection, obtained by numerical integration of $\partial x/\partial p$. (Equation A5.5.4 is equivalent to A21 in Barton & Bengtsson (1986). Although the factor of 2 difference in s between the respective notations introduces a discrepancy, note that there is also a misprint in the original paper, with a factor of $2\theta^2$ missing from the denominator.) The maximum gradient is then given by:

$$width = 1/max.\ gradient = \sqrt{\frac{8\sigma^2}{s}}\frac{\varphi}{\sqrt{e^{2\varphi} - 2 - 1}} \qquad (A5.5.5)$$

Chapter 6

Modelling a habitat preference in a hybrid zone

6.1 Introduction

There has been much debate over the prevalence of sympatric speciation, and the question of whether geographic isolation is a prerequisite for reproductive isolation. The theoretical models demonstrating its plausibility generally rely on some behavioural mechanism generating assortative mating; if this is associated with exploitation of distinct niches, reproductive isolation will develop. In this chapter, I consider the implications of such behavioural mechanisms for a hybrid zone, and ask whether the linkage disequilibrium typically found in hybrid zones relaxes the usually strict criteria for the maintenance of a balanced polymorphism.

6.1 (i) Sympatric speciation and the role of habitat preference

In an extension of Levene's (1953) soft selection model, Maynard Smith (1962; 1966) demonstrated that density-dependent regulation could result in a balanced polymorphism at loci conferring adaptation to alternative sympatric habitats. If individuals demonstrate positive assortative mating, linkage disequilibrium between the genes comprising the selected polymorphism and the genes determining habitat preference will be built up and sympatric divergence will ensue. The conditions of selection, gene flow and linkage under which this can occur are fairly restricted because, as Felsenstein (1981) illustrates, recombination continually breaks down the requisite associations. However, these models assume that regulation in either habitat is not associated with breeding habits. Subsequent theoretical models have invoked the frequently-observed phenomenon of mating between individuals utilising the same habitat: a habitat preference therefore defines both fitness and assortative mating (Bush & Diehl 1982; Diehl & Bush 1989; Rausher 1984; Rice 1984, 1987; Tauber & Tauber 1977). Under these conditions, appreciable divergence can be generated between the populations using alternative habitats without requiring implausible criteria. Evidence for genetically-based differences in

6. MODELLING A HABITAT PREFERENCE

preference for, and fitness in, different habitats is accumulating, particularly in phytophagous insects (e.g. reviews in Diehl & Bush 1984; Mopper 1997; Tauber & Tauber 1989).

6.1 (ii) Hybrid zones and the role of habitat preference

These sympatric speciation models have considered the effect of preference for alternative breeding habitats in generating reproductive isolation within a population. I consider here an alternative scenario: secondary contact between two populations maintained at different equilibria (c.f. Barton & Charlesworth 1984) by natural selection, whose individuals will interbreed but have a genetically-determined preference for alternative mating sites. In section 1.2, I outlined three possible consequences of secondary contact between two divergent populations: (i) the two populations merge back into one; (ii) the two populations remain distinct, with parapatric distributions; (iii) the two populations remain distinct, with overlapping, sympatric distributions. The clines considered in Chapter 5 were maintained at a balance between (i) and (ii), but the availability of alternative habitats, the existence of disruptive selection on loci affecting fitness and the possibility of assortative mating makes (iii) a further option.

In terms of the parameters considered in Chapter 5, what will be the net effect of alternative breeding habitats and a behavioural preference on the dynamics of a hybrid zone? Gene flow will presumably be reduced because of fewer opportunities for interbreeding. However, a preference could facilitate broader sympatry, and might therefore increase the chance of an allele crossing into the alternative genetic background. If gene frequencies track habitat availability, and the latter follows a wider transition than a cline maintained in the face of selection alone, clines will be widened. However, mixing between habitats at different gene frequencies will generate linkage disequilibria, which should increase the effective selection acting on an allele and therefore narrow a cline.

166

The outcome is therefore not obvious, but will clearly be affected by three factors:

(i) The availability of either habitat type across the hybrid zone. This may appear trivial, but the environmental structure of the hybrid zone will have a large affect on its dynamics: with equal availability of either type of habitat, the effect of a habitat preference will differ from that with a sharp transition from one type to the other. (See section 4.4 for a discussion of the implications of habitat availability for the *Bombina* hybrid zone.)

(ii) The selection mechanism involved. As in the previous chapter, the distinction between endogenous (acting uniformly against hybrids) and exogenous (favouring alternative alleles in particular environments; after Moore and Price 1993) selection needs to be made. Bearing in mind Rice and Hostert's conclusion (1993) that a habitat preference can only be maintained in the face of interbreeding if it confers a selective advantage, selection must act either against hybrids or in relation to habitat type. I restrict the analysis presented here to the former; see section 6.6 for a discussion of the implications of the latter.

(iii) The strength of the preference: the effects of absolute preference and random mating are intuitive, but the consequences of intermediate values need to be explored.

The effects of these factors are not independent, nor, in some cases, are there unequivocal predictions for even the most simple scenarios. In this chapter, I aim to establish their qualitative effects and to explore the interdependence between them. Analytical predictions for such a system are complex; a simulation model is therefore a useful starting point from which to explore the possible implications.

6.1 (iii) Relevance to the Bombina *hybrid zone*

Despite the theoretical appeal of an alternative consideration of habitat preference, the questions posed above were initially motivated by the need to explain data from the *Bombina* hybrid zone. MacCallum (1994) revealed an apparent paradox in the comparison of the transect across the hybrid zone in Croatia with the transects in Poland

(Szymura & Barton 1986, 1991). Linkage disequilibrium values were twice as high as in Poland, implying twice as much effective selection acting on the allozyme markers (see section 5.5). However the Croatian cline is shallower, with a smaller central step in gene frequency than in Poland, implying that the mean fitness in central populations is not as low (section 5.7). Data from the Polish transects do not suggest any evidence of habitat differentiation, and MacCallum argues that the discrepancy must arise because of the habitat preference, which would also explain the deviations from Hardy-Weinberg proportions observed in the Croatian transect but not in the Polish transects.

In this chapter, I check that a habitat preference can indeed generate such effects, and, if so, under what conditions. Although the strength of preference appears to differ between *B. bombina* and *B. variegata* (see section 4.1 for details) for simplicity I confine the following analysis to a case of equal strength of preference in the parental taxa.

Finally, in the analysis of quantitative traits described in Chapter Two, the effect of non-random mating on linkage disequilibrium was incorporated in the measurements of linkage disequilibrium presented. I return to this issue here, and consider a method developed by N. Barton to separate the effects of within and between genome associations.

6.1 (iv) Aims

The results presented here therefore aim to address the following questions:

1. (a) What is the effect of the nature of change of habitat availability under increasingly strong habitat preference? In particular, what values of selection and preference are required for the cline in gene frequency to track changing habitat availability?

2. What is the effect of the habitat preference in maintaining the integrity of either genome? This can be measured by considering all individuals at a given point of the cline, pooled across both habitat types, and measuring the magnitude of heterozygote deficit and linkage disequilibria amongst them. Intuitive predictions are that for a given selection strength, overall heterozygote deficit and linkage disequilibrium should increase with habitat preference strength.

3. Could a habitat preference generate effects similar to the differences observed between the clines across the *Bombina* hybrid zone in Poland (Szymura & Barton 1986, 1991) and those in Croatia (MacCallum 1994)? In particular, can a habitat preference generate heterozygote deficits, increase linkage disequilibrium in populations measured within alternative habitats and make a cline shallower?

Section 6.2 outlines the statistical implications of non-random mating, and describes the incorporation of a genetically-determined habitat preference into the simulation model used in Chapter Five. Section 6.3 illustrates the resulting distribution of the habitat preference loci and the genes under selection. The remaining sections deal with particular parameters used to describe cline shape. In 6.4, I consider the effect of the underlying environmental structure (changing habitat availability) on the resulting width of the cline. Sections 6.5 and 6.6 deal, respectively, with measurements of heterozygote deficit and linkage disequilibrium.

6.2 Methods

6.2 (i) Statistics

Mixing populations at different gene frequencies generates heterozygote deficits (the Wahlund effect), in a manner analogous to the generation of linkage disequilibrium. However, linkage disequilibria are only broken down by a factor of (1-r) with each round of random mating, whereas Hardy-Weinberg proportions are immediately restored by random mating. In the previous chapter, statistics were measured after reproduction and selection, but before migration, so the only source of deviations from Hardy-Weinberg proportions would have been the effect of selection, which is assumed to be weak. However, in studies such as that of *Bombina*, statistics are measured on adults in their chosen breeding habitats. Thus in addition to the baseline deficit generated by selection and dispersal across a transect, the heterozygote deficit (F_{is}) will be inflated by any degree of habitat preference.

As outlined in section 2.5, the linkage disequilibrium D is equivalent to the covariance in state between alleles *within* a haplotype. However, haplotype frequencies cannot be measured explicitly, and so estimates are made from data on diploid genotypes. As a result, any covariance *between* genomes will inflate the total covariance. In a previous study (Nürnberger *et al.* 1995), the effect of non-random mating was assumed to be an increase of D by a factor of $(1+F_{is})$. However, further analytical work on the effect of a habitat preference in a hybrid zone has shown that the covariances within and between genomes are additive (N. Barton, pers. comm.). The F_{is}-inflated linkage disequilibrium, D_F, estimated from the variance in diplotype is therefore given by:

$$D_F = D + F_{is} * pq \qquad (6.2.1)$$

In the following sections, reference to linkage disequilibrium implies D, calculated as $(D_F - F_{is} * pq)$, under the assumption that maximum values for both D_F and F_{is} are reached in central demes at which p=0.5.

As mentioned in section 6.1, measures of D and F_{is} across both habitats indicate the extent to which the integrity of the parental genomes is maintained despite interbreeding. However, a natural hybrid zone will not be arranged in a crocodile of habitat pairs, which can conveniently be paired off to give these pooled estimates. Statistics calculated from field data will therefore be calculated for each sample from each breeding habitat, so it is necessary to consider the effect of a habitat preference on these. Under random mating, the values should be no different from those measured across both habitats, except for the sampling effects of smaller population size. At the other extreme, absolute habitat preference (and hence no inbreeding) would generate maximum values for the across-habitat statistics and zero values for the within-habitat statistics. The relationship between the two forms of statistics at intermediate strengths of habitat preference is less predictable. The effect was apparent in Sites _et al._'s (1995) analysis of data on hybridising races of _Sceloporus grammicus_ lizards, whereby changes in statistical estimates as samples were pooled over an increasingly wide area revealed a non-random association of karyotype with environment type. Both types of statistics are therefore considered.

6.2 (ii) Simulation model

The simulation model used in this chapter is based on that described in Chapter 5 (see Appendix 5.2 for details). A hybrid zone is represented by a chain of demes, but each deme is now split into two alternative breeding habitats. After migration between demes, adults chose one or other habitat within a deme. An individual's choice is defined by the following three factors:

(i) The relative abundance or availability of either habitat at different points. This can be varied, ranging from a scenario of equal availability to a step function, with intermediate stages represented by increasingly-steep tanh curves.

(ii) The individual's genotype at a set of habitat preference loci: an individual whose alleles are all of one type will have the strongest preference for one habitat, if of the other, preference will be for the other habitat. Total habitat preference is determined by the additive effects of the habitat preference loci, which are interspersed between neutral and selected loci; their number and recombination rate are given for each set of runs presented.

(iii) The overall strength of the habitat preference. This parameter ("hps") tunes the degree of preference: for example, low values represent a scenario in which not even "pure" individuals show a very strong preference, but with high values pure individuals will always choose their respective habitat type, unless its availability is negligible, and individuals of intermediate genotypes will show a marked preference.

Defining the number of habitat preference loci as nh, and the number of type 1 alleles at the total nh preference loci, the probability of choosing habitat type 1 is specified by:

$$Prob(1) = \frac{x}{x + (1-x)Exp[-hps*(i-nh)/nh]}, \quad \begin{array}{l} 0 \leq x \leq 1; \\ 0 \leq i \leq 2nh; \\ -\infty < hps < \infty. \end{array} \quad (6.2.2)$$

A function of this form allows strong choice even in hybrid genotypes (except, trivially, for individuals in which i=nh), if the habitat preference strength is set to high values. Figure 6.2.1 shows this probability (as a function of habitat availability) for the 5 trait values possible if habitat preference is controlled by 2 loci, for two strengths of habitat preference strength. All of the above factors are assumed to have symmetric effects.

6.2.1 (a)
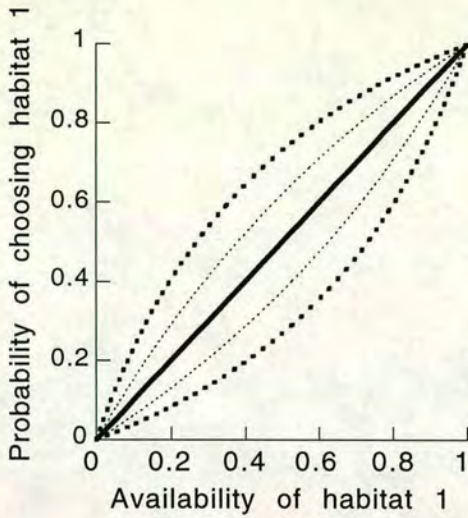
6.2.1 (b)
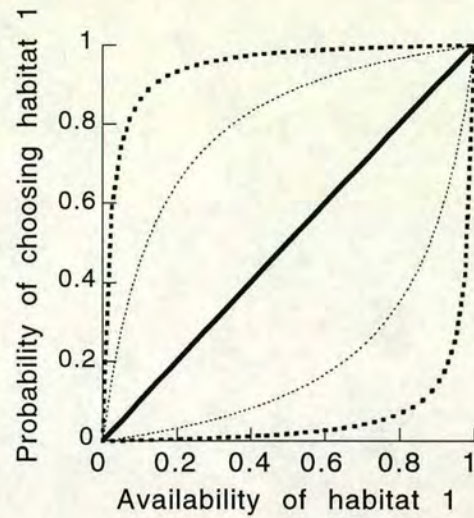
Figure 6.2.1 Probability of an individual choosing habitat 1 given habitat preference strength (a) hps=1.5 (b) hps=6. Different lines represent the preference of 5 different genotypes, as defined by the number (i) of type 1 alleles at 2 habitat preference loci: i=0 [lower dotted line]; i=1 [lower broken line]; i=2, i.e. no preference [solid line]; i=3 [upper broken line]; i=4 [upper dotted line].

Individuals then mate randomly within each habitat. Selection acts on the offspring, before migration; in this case, selection favours alternative alleles in either habitat (instead of on either side of the central point, as with the environmental selection model used in Chapter 5). As before, selection is soft, but with the number of individuals produced *by each habitat* within a deme held constant (and defined as the habitat availability).

Statistics are measured on adults in their chosen breeding habitats. I estimate heterozygote deficit and linkage disequilibrium (i) across all individuals in a given deme (i.e. for the two habitat types pooled) and (ii) within each habitat. In all the simulations presented in this chapter, statistics are measured every 100 generations between 500 and 2400 generations, the number of individuals per deme is set as N=40, the migration rate is

m=0.5 and the total number of demes considered is 100. When cline width is referred to, this is defined (as before) as the inverse of the maximum gradient, and is measured in numbers of demes. Error bars in all figures represent 95% confidence intervals, defined as twice the standard error of the mean. The different types of loci are always evenly interspersed between each other.

6.3 Effect of habitat preference on cline shape

The results presented in this chapter use various (i) widths of the habitat availability cline; (ii) number of loci under selection or determining habitat preference; (iii) linkage relationships between loci. Each combination will have different implications for the overall dynamics, but I start here with an illustration of the general effect of a habitat preference.

6.3 (i) Width of a cline in multiple loci

Consider the scenario in which the availability of either habitat type changes in a smooth transition, forming a cline of width set at 50 demes (Figure 6.3.1(a)). Fitness acts in relation to habitat type, and both fitness and habitat preference are polygenic traits defined by unlinked loci. As shown in Figure 6.3.1(b), a weak habitat preference will have little effect on the distribution of the other loci. Figure 6.3.1(c) represents an intermediate preference strength, as apparent from the increased difference in the frequency of the habitat preference loci in either habitat. The cline in selected loci at this time point (t=2000 generations) appears similar to that for the weaker preference, but the average width across all time points was significantly wider (38.09 demes (± 1.92 standard error), estimated every 100 generations between t=500 and t=2400) than in the weak preference cline (27.64 (± 2.34) demes). This suggests that with increasing habitat preference, the

shape of the cline increasingly reflects the underlying distribution of habitat availability, an effect which I consider further in section 6.4.

In Figure 6.3.1(d), the habitat preference is nearly strong enough for the populations in either habitat to be fixed for both sets of loci. Further simulations confirmed that a marginally stronger preference resulted in complete sympatry, as habitat choice was absolute and so no interbreeding occurred. Thus there is a threshold strength of preference below which clines differ little from the patterns expected in a homogenous environment, and above which the two populations can effectively exist in sympatry with little hybridisation. In all cases, the clines in the preference loci seem to level off in the centre, thereby precluding linkage disequilibrium being generated by mixing between demes at different gene frequencies. It is only when the preference becomes substantial that the difference in gene frequency between habitats is sufficiently large to generate associations between preference and selected loci on mixing. Thus despite the complete initial association between alleles conferring greater fitness in and preference for a given habitat, the preference must be strong to maintain an effect in the face of recombination.

(a)
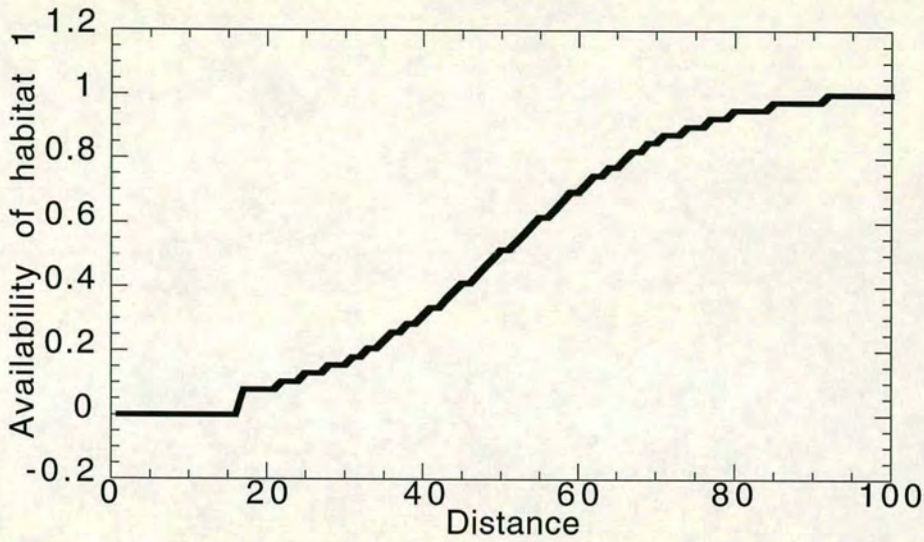
(b) Weak preference
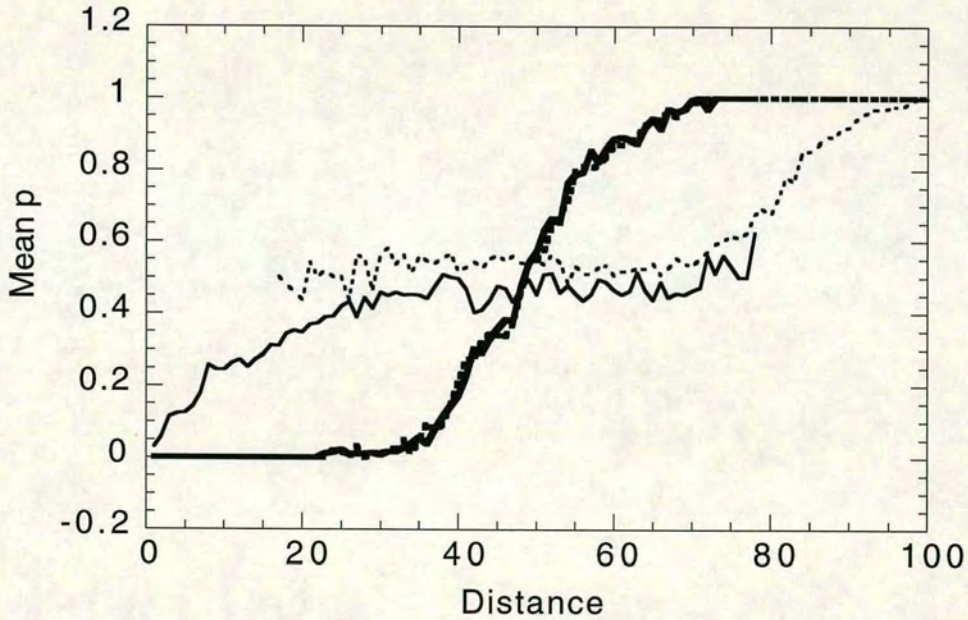
Figure 6.3.1 (a) Underlying availability of habitat type 1. **(b)-(d)** Mean gene frequency at time t=2000 of 8 selected loci [bold lines] and 8 habitat preference loci [fine lines] amongst individuals in habitat 0 [unbroken lines] and habitat 1[dotted lines]. All loci are unlinked (r=0.5), and selection per allele in the wrong habitat is s=0.01. Three strengths of habitat preference are: (b) hps=6; (c) hps=12; (d) hps=15.

(c) Intermediate preference
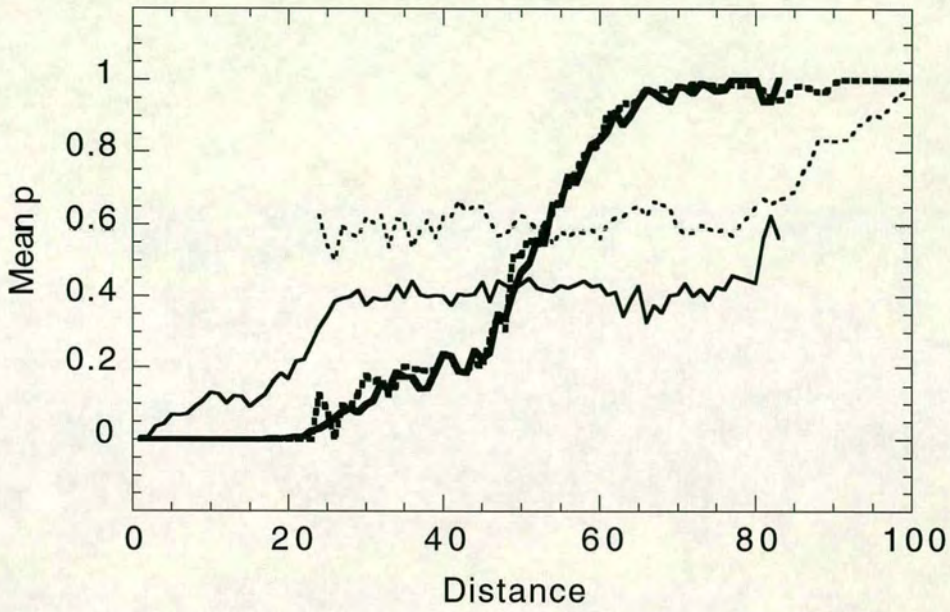
(d) Strong preference
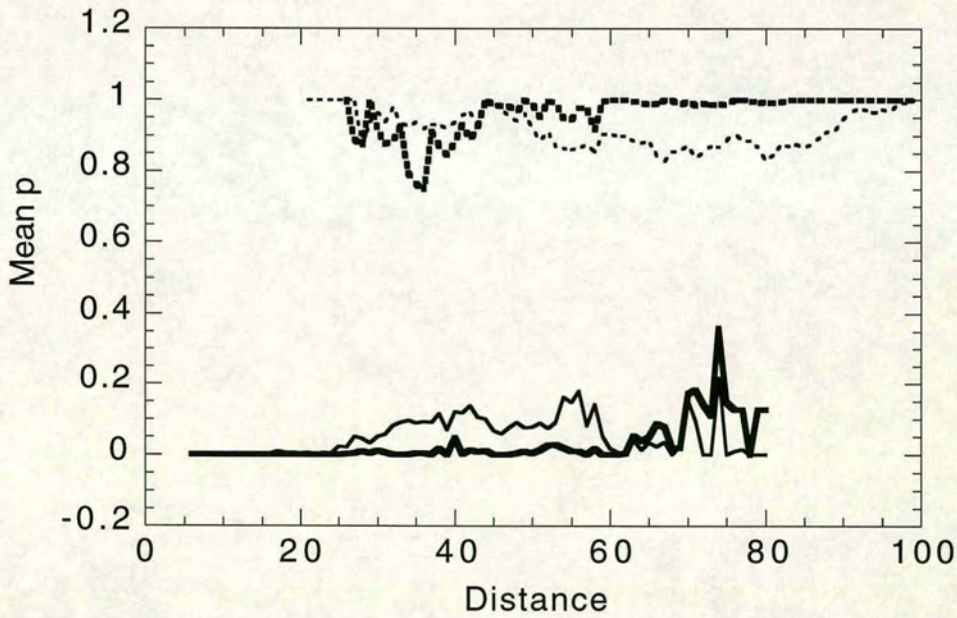
Figure 6.3.1 See previous page for figure legend.

6.3 (ii) Width of a cline in a single locus

The previous results give an impression of the dynamics of a set of unlinked loci: clines will only be affected if habitat preference is sufficiently strong. This is presumably because of the destructive effects of recombination. To illustrate the consequences of the habitat preference alone, I consider the effect of the width of the habitat availability cline on the shape of a cline in a single selected locus tightly linked to a single habitat preference locus. Figure 6.3.2 shows the width of the clines in gene frequency observed in runs where the underlying habitat availability changes over a narrow, an intermediate and a wide region respectively.

As the preference strength increases, the clines widen initially and then tend towards the width of the underlying cline in habitat availability. The exact reason for this broadening and then narrowing is not obvious. It presumably reflects the response to an initial slackening of fitness reduction as the habitat preference guides alleles towards the habitat in which they are fittest. With stronger preference strengths, the loci must then increasingly track the change in habitat availability. Non-monotonic effects appear again in section 6.5.

(Note that such tight linkage implies the two loci are on the same chromosome. As discussed in section 5.8, Baird (1995) has shown that if loci are tightly linked on a chromosome, the approach to an equilibrium state will be slow. Although there was no apparent change in the widths of the clines described here during the sampling period, the results should be interpreted with the caveat that the equilibrium they claim to represent may not actually have been reached.)
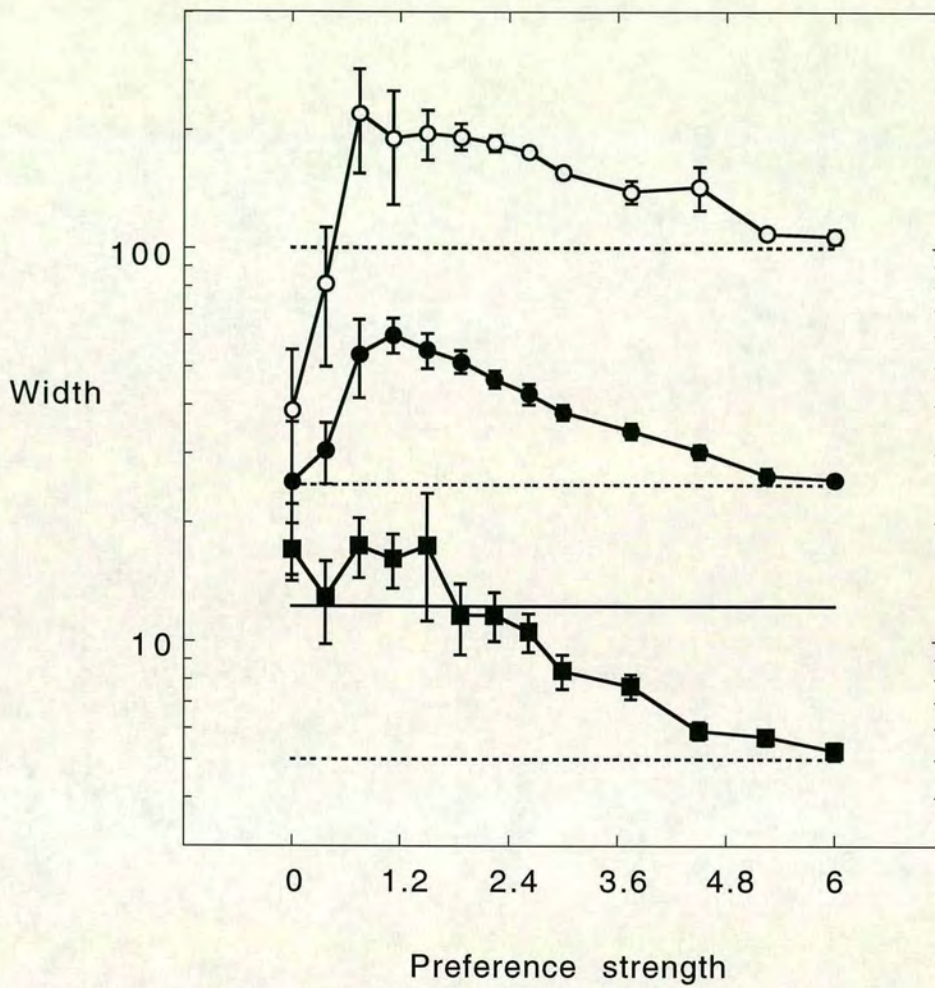
Figure 6.3.2 Width in cline of a single selected locus closely linked (r=0.001) to a single locus affecting habitat preference. Selection is habitat dependent; s=0.01 and migration rate m=0.5, so the predicted width of a single locus cline with step environment = 12.247 [solid straight line]. Underlying cline in habitat availability has width 5 [filled boxes], width 25 [filled circles] or width 100 [open circles]. See section 6.2 for other simulation parameters.

6.4 Maintaining the integrity of the parental genomes

The extent to which an allele remains associated with other alleles from the same parental population can be quantified by the heterozygote deficit (F_{is}) and the linkage disequilibrium (D) (section 6.2). The respective estimates are estimated for all individuals within a deme, pooled from both habitats. Figures 6.4.1 and 6.4.2 show the behaviour of, respectively, F_{is} and D of the selected loci under increasingly strong habitat preference, for two possible selection strengths.

As implied by the earlier consideration of cline shape (Figure 6.3.1), at a threshold magnitude of habitat preference, the dynamics show a transition from a system in which the selected loci appear unaffected by the preference loci to one in which maximum values of association are observed, implying complete assortative mating. The transition is rapid, with few values of preference strength resulting in the intermediate stages.

The behaviour of F_{is} and D is (predictably) qualitatively similar. There are however some inexplicable discrepancies in the value of linkage disequilibrium: instead of levelling off at a maximum possible value of 0.25, it reaches values around 0.3. Furthermore, baseline levels of F_{is} are positive, whereas those of D centre around 0. This is presumably because of an error introduced in applying equation 6.2.1 to the maximum values of F_{is} and D observed across the cline, and my assumption that both will have been reached in a central deme at which p=0.5. Analytical expectations for the effect of a habitat preference on these moments are currently being developed by N. Barton, and it is hoped that the combination of theoretical predictions and simulation results will provide a clearer understanding of the system.

(a) r=0.1
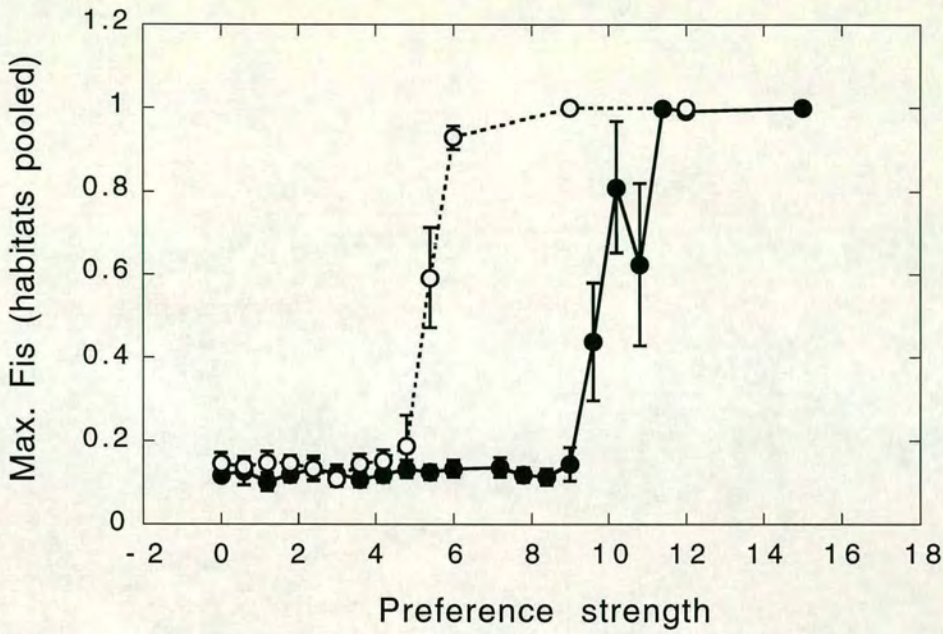
(b) r=0.5
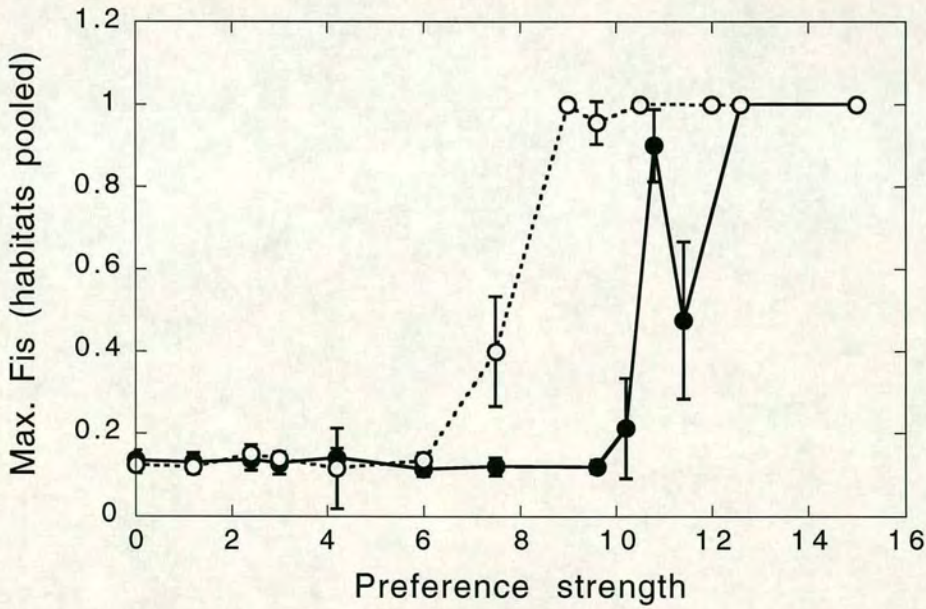
Figure 6.4.1 Mean maximum heterozygote deficit (F_{is}) between 8 selected loci in each deme, pooled across habitats. Number of habitat preference loci=8. Recombination rate between adjacent loci (**a**) r=0.1; (**b**) r=0.5. Selection strength per allele s=0.1 [open circles] or s=0.01 [filled circles].

(a) r=0.1
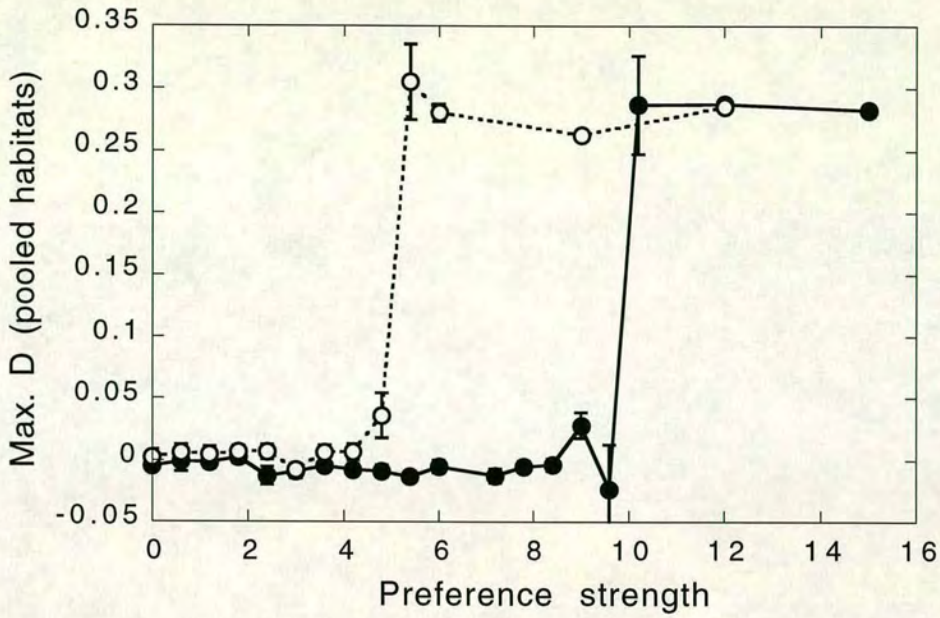
(b) r=0.5
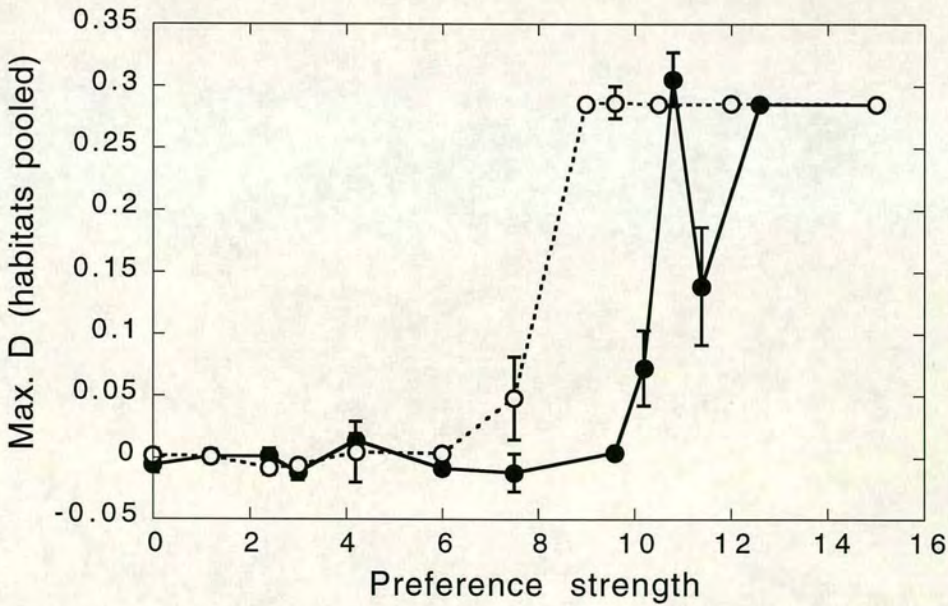
Figure 6.4.2 Mean maximum linkage disequilibrium (D) between 8 selected loci in each deme, pooled across habitats. Number of habitat preference loci=8. Recombination rate between adjacent loci (a) r=0.1; (b) r=0.5. Selection strength per allele s=0.1 [open circles] or s=0.01 [filled circles].

6.5 Associations within habitats

The statistics presented in the previous section give an impression of the overall dynamics of the system. What happens to the values of heterozygote deficit and linkage disequilibrium within habitats? Because of the discrepancy in the linkage disequilibrium calculations, I restrict the analysis in this section to the heterozygote deficit, F_{is}. Figure 6.5.1 depicts the change in Fis under increasingly strong habitat preference, again considering both loosely linked and unlinked loci.

Interestingly, for the stronger selection strength only, associations within habitat initially increase at the point at which the across-habitat values change, and then decrease. At this stage, the habitat preference must be strong enough to generate substantial levels of assortative mating, but not so strong that individuals do not still make "mistakes". Within-habitat values will be at their maximum at this stage, when the populations consist of relatively intact genotypes which are still coming into contact with each other. The increased differences between gene frequencies in either habitats increases the Wahlund effect, and hence the degree of association, resulting in a positive feedback similar to that observed in the stepped clines of the previous chapter. The difference between the two selection strengths is surprising, but (both here and in section 6.4) qualitative differences between the different recombination rates were negligible.

(a) r=0.1
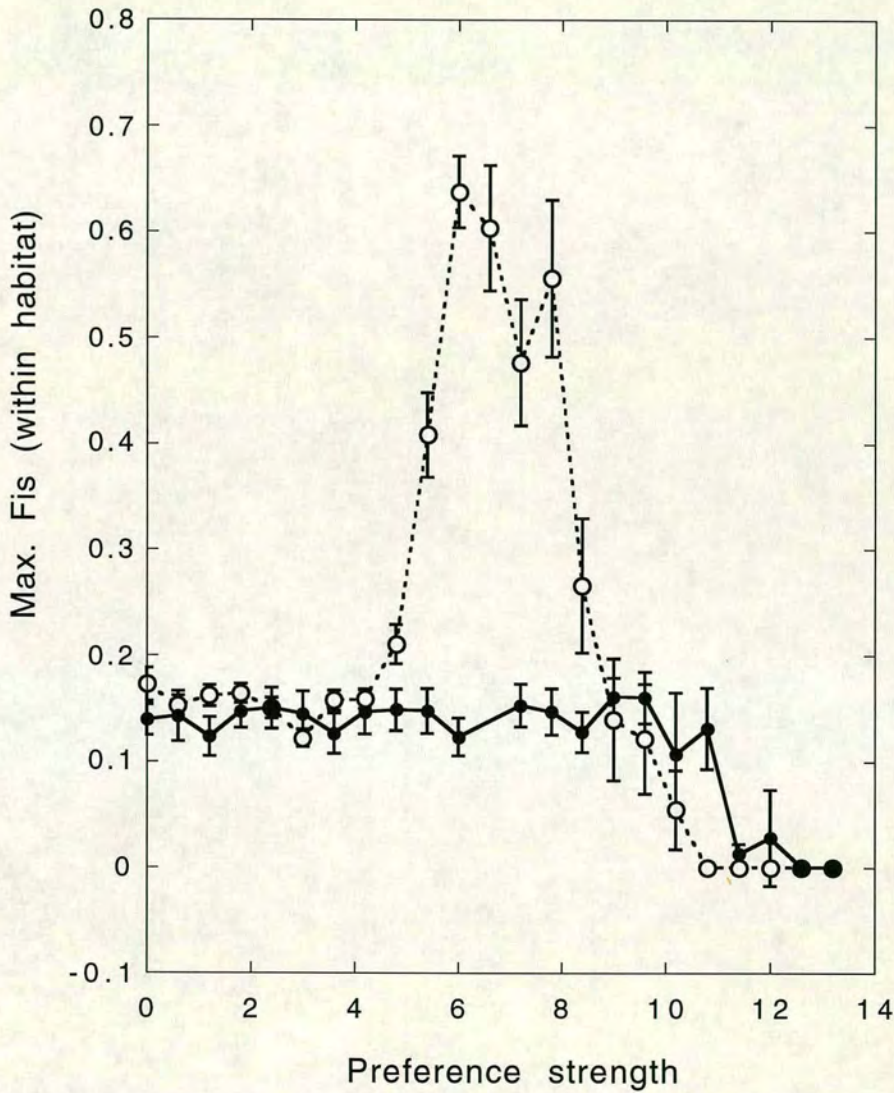
Figure 6.5.1 Maximum average heterozygote deficit (F_{is}) between 8 selected loci in each deme, pooled across habitats. Number of habitat preference loci=8. Recombination rate between adjacent loci **(a)** r=0.1; **(b)** r=0.5 (overleaf). Selection strength per allele s=0.1 [open circles] or s=0.01 [filled circles].

(b) r=0.5
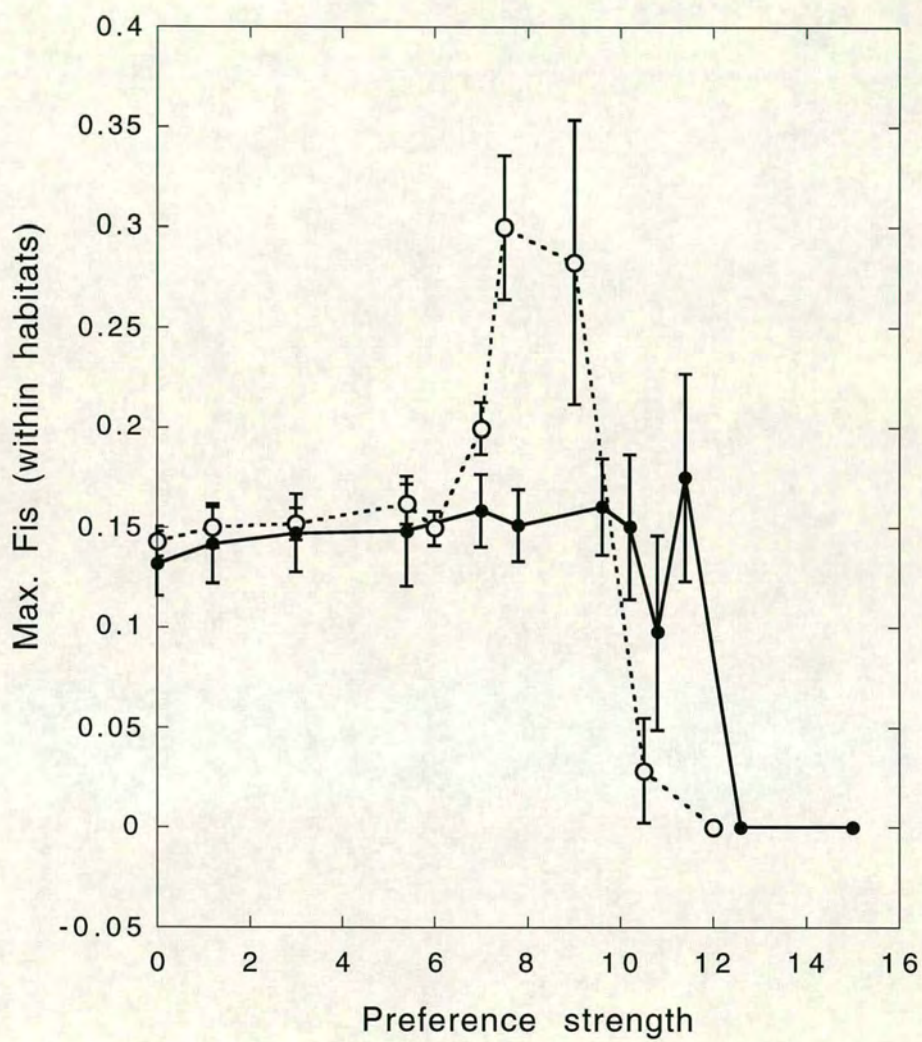
Figure 6.5.1 (b) See previous page for figure legend.

6.6 Discussion

The results fit with the intuitive predictions that extremely strong habitat preference will generate complete association in samples pooled across habitat types and weak association within habitat types. Secondly, recombination prevents weak preferences from having any effect on the dynamics. However, the results for intermediate values of habitat preference strength are less predictable. Non-monotonic effects are observed in two instances, and the system seems highly sensitive to the strength of selection but not to the strength of recombination. The transition from baseline levels of total association (pooled across habitats) is rapid in all cases.

These results suggest numerous avenues for further research. Firstly, a wider range of recombination and selection strengths need to be tested on the systems described here. The consequences of selection against heterozygotes, rather than in relation to habitat type, are worth consideration. Speculatively, might sympatry be a more absolute state than parapatry? The clines described in Chapter 5, maintained by a dispersal-selection balance, represent blurred parapatry, with intermediate levels of genomic associations. By contrast, the dynamics of the system explored here show a rapid transition from baseline to maximum association. Blurred sympatry may not be an option.

The resulting patterns in neutral markers also need to be established; as outlined in section 6.1, one of the primary motivations for this study was to explain data on allozyme frequency in the *Bombina* hybrid zone. Assuming for the moment that the effect of the habitat preference will be similar to that on the selected loci, the results presented here imply that a habitat preference can indeed, as expected, alter the shape of clines and increase the magnitude of statistical associations. Whether it will only do so, as implied by Figure 6.5.1, for a narrow window of preference strength and relatively strong selection strengths requires further investigation.

Chapter 7
Concluding remarks

"Amongst the Batrachian reptiles, I found only one little toad, which was most singular from its colour. If we imagine that... it had been allowed to crawl over a board, freshly painted with the brightest vermilion, so as to colour the soles of its feet and parts of its stomach, a good idea of its appearance will be gained. If it had been an unnamed species, surely it ought to have been called *Diabolicus*, for it is a fit toad to preach in the ear of Eve." (Darwin 1839, p99) (describing *Phryniscus nigricans*)

This thesis describes empirical and theoretical work on hybrid zones. In this chapter, I summarise the principal results of the thesis in terms of their implications for the arguments raised in Chapter One, discuss the implications of the theoretical work for interpretation of the *Bombina* system and suggest areas for future research.

In Chapter One, I outlined how divergence between two allopatric populations might occur as a result of selective forces, of random drift or of the accumulation of different mutations. If the divergence is sufficient to generate reproductive isolation, speciation is said to have occurred. Speciation is a process which has been variously referred to as "utterly mysterious" (Bateson 1922) or even the "mystery of mysteries" (Darwin 1872). However, some insight into its mechanics can be gained if there is secondary contact between the allopatric populations which results in some interbreeding and hence the formation of a hybrid zone: partial reproductive isolation can be investigated.

In the study of a transect across a *Bombina* hybrid zone in Croatia presented here, I show direct evidence for reduced fitness both of individuals of hybrid ancestry and of individuals of one parental type in the other's environment. This has not, to my knowledge, been shown directly in any other hybrid zone. Reduced hybrid fitness constitutes an endogenous barrier to gene flow, mediated by within-genome interactions,

which may be the result of either adaptive or non-adaptive divergence. Reduced fitness of one parental type in the other's environment implies that the divergence during allopatry has been driven (at least partially) by selective forces, with the two populations accumulating adaptations to alternative environments. Note that the various forms of divergence need not have occurred simultaneously; knowledge of the respective ages of, for example, the differences in behaviour, morphology or ecology would shed interesting light on the dynamics of divergence, but would be extremely difficult to obtain.

Analytical models of clines maintained by selection acting on multiple loci (Chapter 5) confirm previous single locus results that the clines generated by either type of selection are indistinguishable in shape. It would seem reasonable to argue that the combined action of the two (as in the *Bombina* hybrid zone) should also produce similar dynamics, but the point would benefit from further theoretical work. In any case, estimates of the magnitude of the fitness reduction in hybrid populations do not require any assumption about the type of selection acting, and the distribution of its effects across the genome (namely the effective selection acting on a locus as a proportion of the total selection) show relatively little difference between the two selection regimes.

A more fundamental question clearly concerns the relative magnitudes of the effect of the two types of selection. The concordance and coincidence of clines in various traits is remarkable, given the apparent environmental heterogeneity of the Pešćenica transect. By the reasoning outlined in Chapter One (from Barton & Hewitt 1985), this would suggest the predominance of hybrid dysfunction in determining the dynamics of the hybrid zone. However, the difference in the strength of selection on alternative traits or in the position of alternative environmental transitions that would be required to alter the relative width or position in clines is also a subject that requires investigation through simulation models. Secondly, the exogenous selection demonstrated here was of a peculiar nature:

behavioural adaptations at one stage of the life cycle to a type of breeding habitat, the availability of which is negligible to one side of the hybrid zone (Chapter Four). If selection on these traits was the principal mechanism maintaining the hybrid zone, and linkage disequilibrium maintained their association with all other traits, the hybrid zone would be centred on the point of transition of habitat availability.

The adaptation to breeding habitats does more than imply that the *Bombina* hybrid zone is maintained by environmental factors as well as hybrid dysfunction. It is associated with a behavioural preference in the adults for breeding habitat type. Given a sufficiently strong preference, and suitable availability of both habitat types, secondary contact between two populations which have diverged in allopatry could plausibly result in their sympatric distribution. The simulations results described in Chapter Six confirm this supposition, but show that the strength of either the habitat preference or the selection defining differential adaptation to breeding habitat must be considerable for the system to behave any differently from simple dispersal-selection clines. This is surprising given the strong linkage disequilibrium (a key requirement in models of sympatric speciation) implicit in dispersal of pure individuals from either side of the zone. The simulation results also demonstrated the dependence of cline width on the underlying availability of either habitat type for stronger preference strengths. For the *Bombina* system, this suggests that measurements of the width, and hence estimates of individual dispersal distance (using also observations of linkage disequilibrium) and effective selection acting, should be treated with caution. Finally, note that any factor generating assortative mating will only be maintained in the face of interbreeding if it confers a selective advantage (Rice & Hostert 1993); this can be confirmed with simulations for the multilocus clines considered here (data not shown). The habitat preference in the *Bombina* system will be doubly advantageous: larvae will develop in the habitat to which they are best suited, and the resulting assortative mating will reduce the frequency of unfit hybrids produced.

At several stages throughout this thesis, I have drawn attention to areas requiring further investigation. Of these, studies of *Bombina* larval ecology, and in particular of the effects of temperature and competition, would fill the current gap in our understanding of the relative fitness of either taxa in the puddle habitat. Such studies would ideally be carried out under field conditions, although a laboratory experiment would be considerably more practicable. As emphasised in Chapter 3, a cohort study would provide a comprehensive indication of the relative fitness of different genotypes: the larval stage appears to be a crucial one at which fitness differences might be manifest. Finally, the role of sexual selection in the *Bombina* hybrid zone is a relatively unexplored area, which again might have significant implications on the possibilities of interbreeding. It is not known whether females distinguish between the different calls of either species, nor what part intrasexual competition plays. The theoretical analysis indicated the existence of similar uncharted territories, the investigation of which would enhance our understanding of the genetic implications of interbreeding between divergent taxa.

This thesis has described various forms of barriers to gene flow. The results suggest that several factors interact to maintain the differentiation between the fire-bellied toad *Bombina bombina* and the yellow-bellied toad *B. variegata*, and imply that at least part of their divergence during allopatry was driven by natural selection. Gene flow between the populations will be impeded by the reduced fitness of hybrids, by differential adaptation to alternative environments and, concurrent with and maintained by these factors, by the assortative mating generated by a habitat preference.

References

Abernethy, K. 1994 The establishment of a hybrid zone between red and sika deer (Genus *Cervus*). *Molecular Ecology* **3**, 551-562.

Arnold, E. N., Burton, J. A. & Ovenden, D. W. 1978 *Reptiles and amphibians of Britain and Europe*. Collins Field Guide. London: Collins.

Arnold, M. L. & Bennett, B. D. 1993 Natural hybridization in Louisiana irises: genetic variation and ecological determinants. In *Hybrid zones and the Evolutionary Process*. (ed. R. G. Harrison), pp. 115-139. Oxford University Press.

Arnold, M. L. & Hodges, S. A. 1995 Are natural hybrids fit or unfit relative to parents? *Trends in Ecology and Evolution* **10**, 67-71.

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

Arntzen, J. W. 1996 Parameters of ecology and scale integrate the gradient and mosaic models of hybrid zone structure in *Bombina* toads and *Triturus* newts. *Israel Journal of Zoology* **42**, 111-119.

Arntzen, J. W. & Wallis, G. P. 1991 Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution* **45**, 805-826.

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

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

Barandun, J. 1995 Reproductive ecology of *Bombina variegata* (*Amphibia*). PhD thesis: University of Zürich.

Barrett, S. C. H. 1989 Mating system evolution and speciation in heterostylous plants. In *Speciation and its consequences.* (ed. D. Otte & J. A. Endler). pp. 257-283. Sunderland, Massachusetts: Sinauer Associates, Inc.

Barton, N. H. 1979a The dynamics of hybrid zones. *Heredity* **43**, 341-359.

Barton, N. H. 1979b Gene flow past a cline. *Heredity* **43**, 333-339.

Barton, N. H. 1980 The fitness of hybrids between two chromosomal races of the grasshopper *Podisma pedestris. Heredity* **45**, 47-59.

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. & Baird, S. J. E. 1997 Analyse: maximum likelihood analysis of geographic variation and hybrid zones. Macintosh version available from S. J. E. Baird (http://helios.bto.ed.ac.uk/evolgen/Analyse/Analyse.html).

Barton, N. H. & Bengtsson, B. O. 1986 The barrier to genetic exchange between hybridising populations. *Heredity* **56**, 357-376.

Barton, N. H. & Charlesworth, B. 1984 Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133-164.

Barton, N. H. & Gale, K. S. 1993 Genetic analysis of hybrid zones. In *Hybrid Zones and the Evolutionary Process.* (ed. R. G. Harrison), pp. 13-45. Oxford: Oxford University Press.

Barton, N. H. & Hewitt, G. M. 1981a The genetic basis of hybrid inviability between two chromosomal races of the grasshopper *Podisma pedestris. Heredity* **47**, 367-383.

Barton, N. H. & Hewitt, G. M. 1981b Hybrid zones and speciation. In *Evolution and speciation* (ed. W. R. Atchley & D. S. Woodruff), pp. 109-145. Cambridge: Cambridge University Press.

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

Barton, N. H. & Turelli, M. 1989 Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* **23**, 337-370.

Bateson, W. 1922 Evolutionary faith and modern doubts. *Science* **55**, 55-61.

Bazykin, A. D. 1969 Hypothetical mechanism of speciation. *Evolution* **23**, 685-687.

Bazykin, A. D. 1973 Population genetics of stabilising and disruptive selection. II. Systems of adjacent populations and populations within a continuous area. *Genetika* **9**, 156-166.

Beachy, C. K. 1993 Differences in variation in egg size for several species of salamanders (Amphibia: Caudata) that use different larval environments. *Brimleyana* **18**, 71-82.

Begon, M., Harper, J. L. & Townsend, C. R. 1990 *Ecology: Individuals, Populations, and Communities.* Cambridge, Massachusetts: Blackwell Scientific Publications.

Bengtsson, B. O. 1985 The flow of genes through a genetic barrier. In *Essays in honour of John Maynard Smith.* (ed. P. J. Greenwood, P. H. Harvey & M. Slatkin). pp. 31-42. Cambridge: Cambridge University Press.

Bert, T. M. & Arnold, W. S. 1995 An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution* **49**, 276-289.

Berven, K. A. 1982 The genetic basis of altitudinal variation in the wood frog *Rana sylvatica*. I. An experimental analysis of life history traits. *Evolution* **36**, 962-983.

Berven, K. A. 1987 The heritable basis of variation in larval developmental patterns within populations of the wood frog (*Rana sylvatica*). *Evolution* **41**, 1088-1097.

Berven, K. A. & Chadra, B. G. 1988 The relationship among egg size, density and food level on larval development in the wood frog (*Rana sylvatica*). *Oecologia* **75**, 67-72.

Boulenger, G. A. 1886 On two European species of *Bombinator. Proceedings of the Zoological Society of London.* 449-501.

Bugter, R., MacCallum, C. J., Arntzen, J. W. & Barton, N. H. 1995 Analysis of the terrestrial habitat component in a *Bombina* hybrid zone using remote sensing. In *8th Meeting Societas Europaea Herpetelogica*. Bonn.

Bush, G. L. 1975 Modes of animal speciation. *Annual Review of Ecology and Systematics* **6**, 339-364.

Bush, G. L. 1994 Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology and Evolution* **9**, 285-288.

Bush, G. L. & Diehl, S. R. 1982 Host shifts, genetic models of sympatric speciation and the origin of parasitic insect species. In *5th International Symposium on Insect-Plant Relationship*, pp. 297-305. Wageningen.

Butlin, R. 1989 Reinforcement of premating isolation. In *Speciation and its Consequences* (ed. D. Otte & J. A. Endler), pp. 158-179: Sinauer Associates, Inc.

Butlin, R. K. & Hewitt, G. M. 1985 A hybrid zone between *Chorthippus parallelus parallelus* and *C. p. erythropus* (Orthoptera: Acrididae): morphological and electrophoretic characters. *Biological Journal of the Linnean Society* **26**, 269-285.

Butlin, R. K., Ritchie, M. G. & Hewitt, G. M. 1991 Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera; Acrididae). *Philosophical Transactions of the Royal Society of London B* **334**, 297-308.

Charlesworth, D. 1995 Evolution under the microscope. *Current Biology* **5**, 835-836.

Cheeson, J. 1978 Measuring preference in selective predation. *Ecology* **59**, 211-215.

Chovanec, A. 1992 The influence of tadpole swimming behaviour on predation by dragonfly nymphs. *Amphibia-Reptilia* **13**, 341-349.

Cortwright, S. A. & Nelson, C. E. 1990 An examination of multiple factors affecting community structure in an aquatic amphibian community. *Oecologia* **83**, 123-131.

Cracraft, J. 1989 Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In *Speciation and its consequences*. (ed. D. Otte & J. A. Endler), pp. 28-59. Sunderland, Massachusetts: Sinauer Associates, Inc.

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): MSc thesis; Jagiellonian University, Cracow, Poland.

Czopkowa, G. & Czopek, J. 1955 The vascularization of the respiratory surfaces in *Bombina variegata*. *Bulletin de l'Academie Polonaise des Sciences,* Cl.II **3**, 313-318.

Darwin, C. R. 1839 *The Voyage of a Naturalist round the World in H.M.S. "Beagle".* London: George Routledge & Sons, Ltd.

Darwin, C. R. 1872 *The Origin of Species (6th edition).* London: Oxford University Press. (First edition 1859)

Diehl, S. R. & Bush, G. L. 1984 An evolutionary and applied perspective of insect biotypes. *Annual Review of Entomology* **29**, 471-504.

Diehl, S. R. & Bush, G. L. 1989 The role of habitat preference in adaptation and speciation. In *Speciation and its Consequences* (ed. D. Otte & J. A. Endler), pp. 345-354: Sinauer Associates, Inc.

Dobzhansky, T. 1940 Speciation as a stage in evolutionary divergence. *American Naturalist* **74**, 312-321.

Dobzhansky, T. 1970 *Genetics of the Evolutionary Process.* New York: Columbia University Press.

Duellman, W. E. & Trueb, L. 1985 *Biology of Amphibians.* New York: McGraw-Hill.

Edwards, A. W. F. 1972 *Likelihood.* Cambridge University Press.

Endler, J. A. 1977 *Geographic variation, speciation and clines.* Princeton, New Jersey: Princeton University Press.

Feder, J. L., Chilcote, C. A. & Bush, G. L. 1990 Regional, local and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. *Evolution* **44**, 595-608.

Feder, J. L., Chilcote, C. A. & Bush, G. L. 1994 Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Science* **91**, 7990-7994.

Felsenstein, J. 1975 Genetic drift in clines which are maintained by migration and natural selection. *Genetics* **81**, 191-207.

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

Fisher, R. A. 1937 The wave of advance of advantageous genes. *Ann. Eugenics* **7**, 355-369.

Fisher, R. A. 1953 A fuller theory of junctions in inbreeding. *Heredity* **8**, 187-197.

Fitter, R. & Manuel, R. 1995 *Lakes, rivers, streams & ponds of Britain & north-west Europe.* Collins Photo Guide: HarperCollins.

Gilchrist, J. 1993 *Bombina* larval development and temperature. Honours thesis: University of Edinburgh.

Gollmann, B. & Gollmann, G. 1994 Life history variation across a hybrid zone in *Geocrinia*: embryonic development and larval growth. *Acta Oecologica* **15**, 247 - 259.

Gollmann, G. 1984 Allozymic and morphological variation in the hybrid zone between *Bombina bombina* and *Bombina variegata* (Anura, Discoglossidae) in north-eastern Austria. *Z. zool. Systematk u. Evolutionsforschung* **22**, 51-64.

Gollmann, G. 1996 Structure and dynamics of a hybrid zone in *Bombina* (Amphibia: Anura: Discoglossidae). *Israel Journal of Zoology* **42**, 121-133.

Gollmann, G., Roth, P. & Hodl, W. 1988 Hybridisation between the fire-bellied toads *Bombina bombina* and *B. variegata* in the karst regions of Slovakia and Hungary: morphological and allozyme evidence. *Journal of Evolutionary Biology* **1**, 3-14.

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

Grant, P. R. & Grant, B. R. 1992 Hybridisation of bird species. *Science* **256**.

Grant, V. 1981 *Plant speciation.* New York: Columbia University Press.

Gurdon, J. B. 1977 Egg cytoplasm and gene control in development. *Proceedings of the Royal Society B.* *19*8, 211-247.

Hagen, R. H. 1990 Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). *Evolution* **44**, 1914-1930.

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

Harrison, R. G. 1990 Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology* **7**, 69-128.

Harrison, R. G. 1993 Hybrids and hybrid zones: historical perspective. In *Hybrid zones and the Evolutionary Process*. (ed. R. G. Harrison), pp. 3-12. Oxford University Press.

Harrison, R. G. (ed.) 1993 *Hybrid zones and the evolutionary process*. Oxford: Oxford University Press.

Harrison, R. G. & Rand, D. M. 1989 Mosaic hybrid zones and the nature of species boundaries. In *Speciation and its consequences*. (ed. D. Otte & J. A. Endler). pp. 111-133. Sunderland, Massachusetts: Sinauer Associates, Inc.

Hartl, D. L. & Clark, A. G. 1989 *Principles of Population Genetics (2nd edition)*: Sunderland, Massachusetts: Sinauer.

Hewitt, G. M. 1988 Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology and Evolution* **3**, 158-167.

Hewitt, G. M. 1989 The subdivision of species by hybrid zones. In *Speciation and its consequences*. (ed. D. Otte & J. A. Endler). pp. 85-110. Sunderland, Massachusetts: Sinauer.

Hewitt, G. M. 1993a After the ice: *Parallelus* meets *Erythropus* in the Pyrenees. In *Hybrid zones and the evolutionary process* (ed. R. G. Harrison), pp. 140-164. Oxford: Oxford University Press.

Hewitt, G. M. 1993b Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. In *Evolutionary patterns and processes*, pp. 98-123. London: Linnean Society.

Hewitt, G. M. 1996 Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

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

Hill, W. G. & Robertson, A. 1966 The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269-294.

Horbulewicz, L. 1927 Die Verbreitung der *Bombinator*- und *Triton*-Arten im Bereiche der Bezirke Sambor, Drohobycz, Stryj (Kleinpolen). *Bull. Int. Acad. Polon. Sci. Lett. (Série B, Sciences naturelles)* **1**, 87-111.

Horbulewicz, L. 1933 Die geographische Verbreitung der *Bombinator* Arten im Bereiche der Bezirke Dobromil-Przemyśl-Jaroslaw. [In Polish with German summary]. *Kosmos (Lwów) A* **58**, 209-223.

Howard, D. J., Waring, G. L., Tibbets, C. A. & Gregory, P. G. 1993 Survival of hybrids in a mosaic hybrid zone. *Evolution* **47**, 789-900.

Howard, D. J. 1993 Reinforcement: origins, dynamics and fate of an evolutionary hypothesis. In *Hybrid Zones and the Evolutionary Process.* (ed. R. G. Harrison). pp. 46-69. Oxford: Oxford University Press.

Hutchinson, G. E. 1965 The niche: an abstractly inhabited hypervolume. In *The Ecological Theatre and the Evolutionary Play*, pp. 26-78. New Haven: Yale University Press.

Jackson, K. S. 1992 The population dynamics of a hybrid zone in the alpine grasshopper *Podisma pedestris*: an ecological and genetical investigation. PhD thesis: University College, London.

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). *Biological Journal of the Linnean Society* **59**, 221-242.

Kaplan, R. H. & Cooper, W. S. 1984 The evolution of developmental plasticity in reproductive characteristics: an application of the "adaptive coin-flipping" principle. *American Naturalist* **123**, 393-410.

Karaman, S. 1922 Über unsere Bombinatorarten. *Glasnik Kroat. Naturwiss. Ges. Zagreb* **34**, 63-70.

Key, K. H. 1968 The concept of stasipatric speciation. *Systematic Zoology* **17**, 14-22.

Kocher, T. D. & Sage, R. D. 1986 Further genetic analyses of a hybrid zone between leopard frogs (*Rana pipiens* complex) in central Texas. *Evolution* **40**, 21-33.

Koteja, E. 1984 Mortality of embryos of the fire-bellied toad (*Bombina bombina*), yellow-bellied toad (*Bombina variegata*) and their hybrids. M.S. thesis: Zakald Anatomii Porownawczej Inst. Zoologii, Uniw. Jagiellonski, Krakow.

Lamb, T. & Avise, J. C. 1987 Morphological variabaility in genetically defined categories of anuran hybrids. *Evolution* **41**, 157-165.

Lawler, S. P. 1989 Behavioural responses to predators and predation risk in four species of larval anurans. *Animal Behaviour* **38**, 1039-1047.

Levene, H. 1953 Genetic equilibrium when more than one niche is available. *American Naturalist* **87**, 331-333.

Li, W. H. & Nei, M. 1974 Stable linkage disequilibrium without epistasis in subdivided populations. *Theoretical Population Biology* **6**, 173-183.

Lörcher, K. 1969 Vergleichende bio-akustische Untersuchungen and der Rot- und Gelbbauchunke, *Bombina bombina* (L.) und *B. v. variegata* (L.). *Oecologia (Berl.)* **3**, 84-124.

Lynch, J. D. 1989 The gauge of speciation: on the frequencies of modes of speciation. In *Speciation and its consequences.* (ed. D. Otte & J. A. Endler), pp. 527-553. Sunderland, Massachusetts: Sinauer Associates, Inc.

MacCallum, C. 1994 Adaptation and habitat preference in a hybrid zone between *Bombina bombina* and *Bombina variegata*: PhD thesis, University of Edinburgh.

MacCallum, C. J. M., Nürnberger, B. D., Barton, N. H. & Szymura, J. M. 1997 Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution*, in press.

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

Madej, Z. 1973 Ecology of European bellied toads (*Bombina* Oken, 1816) [in Polish with English summary]. *Przeglad Zoologiczny* **17**, 200-204.

Malhotra, A. & Thorpe, R. S. 1991 Experimental detection of rapid evolutionary response in natural lizard populations. *Nature* **353**, 347-348.

Mallet, J. L. B. 1986 Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. *Heredity* **56**, 191-202.

Mallet, J. L. B. & Barton, N. H. 1989 Inference from clines stabilized by frequency-dependent selection. *Genetics* **122**, 967-976.

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

Manly, B. F. J. 1974 A model for certain types of selection experiments. *Biometrics* **30**, 281-294.

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

May, R. M., Endler, J. A. & McMurtrie, R. E. 1975 Gene frequency clines in the presence of selection opposed by gene flow. *American Naturalist* **109**, 659-676.

Maynard Smith, J. 1962 Disruptive selection, polymorphism and sympatric speciation. *Nature* **195**, 60-2.

Maynard Smith, J. 1966 Sympatric speciation. *American Naturalist* **100**, 637-650.

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

Mayr, E. 1963 *Animal Species and Evolution.* Cambridge, Massachusetts: Belknap Press, Harvard University Press.

McMillan, W. O., Jiggins, C. D. & Mallet, J. 1997 What initiates speciation in passion-vine butterflies? submitted to *Proceedings of the National Academy of Science (USA)*.

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

Mertens, R. & Wermuth, H. 1960 *Die Amphibien and Reptilian Europas.* Frankfurt: W. Kramer.

Michalowksi, 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 Zoological Cracow* **3**, 247-283.

Michalowksi, J. 1961 Studies of species characteristics in *Bombina bombina* (L.) and *B. variegata* (L.). Applying the L:T indicator to the classifying purposes. *Acta Zoological Cracow* **6**, 51-59.

Moore, W. S. 1977 An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52**, 263-277.

Moore, W. S. & Price, J. T. 1993 Nature of selection in the Northern Flicker Hybrid Zone and its Implications for Speciation Theory. In *Hybrid Zones and the Evolutionary Process* (ed. R. G. Harrison), pp. 196-225. Oxford: Oxford University Press.

Mopper, S. 1997 Adaptive genetic structure in phytophagous insect populations. *Trends in Ecology and Evolution* **11**, 235-238.

Morin, P. J. 1983 Predation, competition and the composition of larval anuran guilds. *Ecological Monographs* **53**, 119-138.

Morin, P. J. 1986 Interactions between intraspecific competition and predation in an amphibian predator-prey system. *Ecology* **67**, 713-720.

Murphy, E.E.P. 1995 Cultivation of *Brassica oleracea* under increased temperature and reduced light conditions. Woodburn Press, Edinburgh.

Murray, J. D. 1989 *Mathematical Biology.* Springer-Verlag.

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

Nagylaki, T. 1976a Clines with variable migration. *Genetics* **83**, 867-886.

Nagylaki, T. 1976b The evolution of one and two locus systems. *Genetics* **83**, 583-600.

Nei, M. 1972 Genetic distance between populations. *American Naturalist* **106**, 283-292.

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

Otte, D. & Endler, J. A. 1989 *Speciation and its consequences*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Petranka, J. W. & Sih, A. 1986 Environmental instability, competition, density-dependent growth and survivorship of a stream-dwelling salamander. *Ecology* **67**, 729-736.

Petranka, J. W., Kats, L. B. & Sih, A. 1987a Predator-prey interactions among fish and larval amphibians: use of chemical cues to detect predatory fish. *Animal Behaviour* **35**, 420-425.

Petranka, J. W., Sih, A., Kats, L. B. & Holomuzki, J. R. 1987b Stream drift, size-specific predation, and the evolution of ovum size in an amphibian. *Oecologia* **71**, 624-630.

Piálek, J. & Barton, N. H. 1997 The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics* **145**, 493-504.

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. *Biological Journal of the Linnean Society* **43**, 197-210.

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

Rausher, M. D. 1984 The evolution of habitat preference in subdivided populations. *Evolution* **38**, 596-608.

Rice, W. R. 1984 Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution* **38**, 1251-1260.

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

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

Richards, S. J. & Bull, C. M. 1990 Size-limited predation on tadpoles of three Australian frogs. *Copeia* **4**, 1041-1046.

Ridley, M. 1993 *Evolution*. Oxford: Blackwell Science.

Rieseberg, L. H., Linder, C. R. & Seiler, G. J. 1995 Chromosomal and genic barriers to introgression in *Helianthus. Genetics* **141**, 1163-1171.

Rieseberg, L. H. & Wendel, J. F. 1993 Introgression and its consequences in plants. In *Hybrid Zones and the Evolutionary Process* (ed. R. G. Harrison), pp. 70-114. Oxford: Oxford University Press

Ritchie, M. G., Butlin, R. K. & Hewitt, G. M. 1989 Assortative mating across a hybrid zone in *Chorthippus parallelus* (Orthoptera: Acrididae). *Journal of Evolutionary Biology* **2**, 339-352.

Roughgarden, J. 1979 *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. New York: Macmillan Publishing.

Rouhani, S. & Barton, N. H. 1987 The probability of peak shifts in a founder population. *Journal of Theoretical Biology*. **126**, 51-62.

Ruibal, R. 1955 A study of altitudinal races in *Rana pipiens. Evolution* **9**, 322-338.

Sage, R. D., Heyneman, D., Lim, K. C. & Wilson, A. C. 1986 Wormy mice in a hybrid zone. *Nature* **324**, 60-63.

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

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

Scott, D. E. 1990 Effects of larval density in *Ambystoma opacum* - an experiment in large-scale field enclosures. *Ecology* **71**, 296-306.

Semlitsch, R. D. & Caldwell, J. P. 1982 Effects of density on growth, metamorphosis, and survivorship in tadpoles of *Scaphiopus holbrooki. Ecology* **63**, 905-911.

Shaw, C. R. & Prasad, R. 1970 Starch gel electrophoresis of enzymes - a compilation of recipes. *Biochemical Genetics* **4**, 297-320.

Shaw, D. D., Marchant, A. D., Contreras, N., Arnold, M. L., Groeters, F. & Kohlmann, B. 1993 Genomic and environmental determinants of a narrow hybrid zone: cause or coincidence? In *Hybrid zones and the evolutionary process*. (ed. R. G. Harrison). pp. 165-195. Oxford: Oxford University Press.

Sites, J. W., Barton, N. H. & Reed, K. M. 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.

Skelly, D. K. 1995 A behavioural trade-off and its consequences for the distribution of *Pseudacris* treefrog larvae. *Ecology* **76**, 150-164.

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

Slatkin, M. 1975 Gene flow and selection in a two-locus system. *Genetics* **81**, 209-222.

Slatkin, M. & Maruyama, T. 1975 Genetic drift in a cline. *Genetics* **81**, 209-222.

Sokal, R. R. & Rohlf, F. J. 1981 *Biometry. 2nd. ed.* New York: W. H. Freeman.

Southwood, T. R. E. 1978 *Ecological methods*. London: Chapman & Hall.

Szymura, J. M. 1976a Hybridization between discoglossid toads *Bombina bombina* and *Bombina variegata* in southern Poland as revealed by the electrophoretic technique. *Z.zool.Syst. Evolut. forsch.* **14**, 227-236.

Szymura, J. M. 1976b New data on the hybrid zone between *Bombina bombina* and *Bombina variegata* (Anura, Discoglossidae). *Bulletin de l'Academie Polonaise des Sciences, Cl.II* **24**, 355-363.

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

Szymura, J. M. 1988 Regional differentiation and hybrid zones between the fire-bellied toads *Bombina bombina* (L.) and *B. variegata* (L.) in Europe. [In Polish] In *Rozprawy Habilitacyjne No. 147*. Cracow: Uniwersytet Jagiellonski.

Werner, E. E. 1992 Competitive interactions between wood frog and northern leopard frog larvae - the influence of size and activity. *Copeia* **1**, 26-35.

Werner, E. E. & Anholt, B. R. 1993 Ecological consequences of the trade-off between growth and mortality rates mediated by foraging activity. *American Naturalist* **142**, 242-272.

Werner, E. E. & McPeek, M. A. 1994 Direct and indirect effects of predators on two anuran species along an environmental gradient. *Ecology* **75**, 1368-1382.

Wilbur, H. M. & Fauth, J. E. 1990 Experimental aquatic food webs: interactions between two predators and two prey. *American Naturalist* **135**, 176-204.

Wilbur, H. M., Morin, P. J. & Harris, R. N. 1983 Salamander predation and the structure of experimental communities: anuran responses. *Ecology* **64**, 1423-1429.

Wolfram, S. 1996 Mathematica Version 3.0 - a system for doing mathematics by computer. Redwood City, California: Addison-Wesley.

Woodruff, D. S. 1979 Postmating reproductive isolation in *Pseudophryne* and the evolutionary significance of hybrid zones. *Science* **203**, 561-563.

Woodward, B. D. 1983 Predator-prey interactions and breeding-pond use of temporary-pond species in a desert anuran community. *Ecology* **64**, 1549-1555.

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

Wright, S. 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics.* **1**, 356-366.

Wright, S. 1941 On the probability of fixation of reciprocal translocations. *American Naturalist* **75**, 513-522.

Wright, S. 1988 Surfaces of selective value revisited. *American Naturalist* **131**, 115-123.

Zug, G. R. 1993 *Herpetology.* London: Academic Press, Inc.

Appendix I Multilocus program code

Program code written in Pascal and run using Metrowerks Code Warrior version 1.7.2. See Chapter 5 for description of model. Original program written by Kathy Gale and Nick Barton; see Jackson (1992) and Barton and Gale (1993) for further details.

```
program multilocus;

  uses        {global constants and types are in housekeeping}
    memory,housekeeping, utilities, graphics, functions, graphs, Events, quickdraw,
    random,  newhabstats, multilocus_procs, PasLibIntf, QuickdrawText,
    Fonts, Windows, Menus, SegLoad, OSUtils,
    TextEdit, Dialogs, Events,StandardFile,files,resources,PasLibintf,Types;

  var
    pop, newpop: popptrtype;
    habpop: habpopptrtype;
    kidsdmpop, dmpop, newdmpop, intermed_dmpop: kidsdmpopptrtype;
    s, rec, mutn, mig, migbarr, hwidth, beta, prop:real;
    habprefst, habslope, Dopt, hab_cline_width: real;
    ninit, nfix, g, count, replicate,nreps: integer;
    demeno: demep1type;
    d: demep1type; i, sum: indtype;
    l:locitype; btt: bittype;
    out_file, bf, out_file, dump_file: Text;
    out_fname, list_fname, fname_root, batch_fname, output_fname: string;
    dump_fname,ans, ans2, ans3: string;
    batch, dumpflag, dipflag, detailflag, just_reps: boolean;
    INstringSource:(afile,keyboard); {from utilities.p}
    INstringfile,INstringRec:text;

{********************************************************************}

  procedure askstuff; {inputting parameters manually}
    var
      ib: blocktype;
      gg: genetype;
    begin
      writeln('The program runs several replicates (nreps); each is run for tmax generations.');
      writeln('Do you want just the means of each replication saved?');
      repeat readln(ans) until (ans = 'y') or (ans = 'n');
      just_reps := (ans = 'y');
      writeln('Statistics are recorded every dt gens, after an initial delay of twarm gens');
      writeln('nreps, tmax, dt, twarm ? ');
      repeat  readln(nreps, tmax, dt, twarm)
      until (((tmax-twarm) div dt)<maxsamples) and (nreps>0) and (tmax>0) and (dt>0);
      writeln('Diploid individuals migrate between demes, and then choose breeding habitat');
      writeln('Selection acts on offspring viability within habitat:');
      writeln('mating is random within habitat, producing NK diploid individuals per adult pair.');
      writeln('# demes (<=',maxdemes:3,'; # diploid inds/deme (4<=',maxinds:3,';multiple of 2) ?');
      readln(ndemes, ninds);
      writeln('# of offspring per pair (1<=',maxkids:4,',)');
      readln(nkids);
      writeln('Would you like 2 habitat types?');        {if NO, all inds will be assigned to hab 1}
      repeat  readln(ans)    until (ans = 'y') or (ans = 'n');
      two_habitats := (ans = 'y');
      if two_habitats then
        begin
          writeln('What width would you like the cline in habitat availability to have?');
```

```
          writeln('Enter 999 for equal availibity');
          readln(hab_cline_width);
          if (hab_cline_width = 999) then  habslope := 0
          else
             if (hab_cline_width=0) then    habslope := 999
             else  habslope := 1 / hab_cline_width;
          writeln('Strength of habitat preference? (0<=hps<=1)');
          readln(habprefst);
       end;
  writeln('# of genes ( <= ', maxgenes : 3, ' )? NB. These can be neutral if required');
  repeat  readln(ngenes)     until (ngenes>0) and (ngenes<=maxgenes);
  nblocks := 1 + ((ngenes - 1) div 16);
  nhabblocks := 1 + ((nhabgenes - 1) div 16);
  for ib := 1 to nblocks do    allOne[ib] := -1;          {in binary, this is all ones}
  for ib := 1 to nblocks do    allZero[ib] := 0;          {all zeroes}
  if two_habitats then
     writeln('Would you like some genes to be selected or to affect habitat preference? ')
  else   writeln('Would you like some genes to be selected? ');
  repeat  readln(ans)    until (ans = 'y') or (ans = 'n');
  nsel := 0; nhabgenes := 0; nneut:=0;                 {Default is that all are neutral}
  seln_mask := allZero; hab_mask := allZero;
  nlociclasses:=0;
  if (ans = 'y') then
     begin
        if two_habitats then writeln('For each gene: n for neutral, s for selected, h for hab pref')
        else   writeln('For each gene, type n for neutral, s for selected');
        for gg := 0 to ngenes - 1 do
           begin
              write('Gene ', gg : 4, '? ');
              repeat  readln(ans)     until ((ans = 's') or (ans = 'n') or (ans = 'h'));
              if ans = 's' then
                 begin
                    nsel := nsel + 1; putgene(gg, 1, seln_mask);
                    if (nsel=1) then nlociclasses:=nlociclasses+1;
                 end;
              if ans = 'h' then
                 begin
                    nhabgenes := nhabgenes + 1; putgene(gg, 1, hab_mask);
                    if (nhabgenes=1) then nlociclasses:=nlociclasses+1;
                 end;
              if ans = 'n' then
                 begin
                    nneut:=nneut+1; if (nneut=1) then nlociclasses:=nlociclasses+1;
                 end;
           end;
     end
  else
     begin
        nlociclasses:=1; nneut:=ngenes;
     end;
  if (nneut <> (ngenes - nsel - nhabgenes)) then scream;

  writeln('The genes are on a chromosome, with recombination rate r between each.
  writeln('Recombination, mutation ? ');
  repeat readln(rec, mutn) until (rec>0) and (mutn>=0);
  writeln('nmig individuals migrate (half in each direction); at the centre, there is a barrier,');
  writeln('across which nmigbarr migrate. If there is no barrier, set nmig=nmigbarr');
  writeln('nmig (a multiple of 2),nmigbarr (a multiple of 2) ');
  repeat readln(nmig, nmigbarr) until (nmig>=0) and (nmigbarr>=0);

  repeat
     writeln('Gaussian epistasis ("e"), power epistasis ("pe"),');
     writeln('geographic ( "g ") , habitat dependent ****( "b")***** ');
     writeln('heterozygote disadv ( " h " ) , stabilising seln ( " s " ) ');
     writeln(' OR both het dis and geographic ("t") ?');
     readln(ans);
  until ((ans = 'h') or (ans = 's')) or ((ans = 'e') or (ans = 'pe')) or ((ans = 'g')
```

```pascal
                     or ((ans = 'b') and two_habitats));              {NB. CodeWarrior needs millions of brackets
    epistasis := (ans = 'e');                                            in multiple conditions}
    epistasisp := (ans = 'pe'); hetdis := (ans = 'h'); stabilising := (ans = 's'); geographic := (ans = 'g');
    habitat_dependent := (ans = 'b');
    if epistasis then
       begin
          writeln('s (epistatic),  1/e gene freq halfwidth for epistasis (try 0.25) ? ');
          readln(s, hwidth)
       end;
    if epistasisp then
       begin
          writeln('s (epistatic),  beta ? '); readln(s, beta)
       end;
    if hetdis then
       begin
          writeln('s (against each heterozygous locus) ? '); readln(s)
       end;
    if geographic then
       begin
          writeln('s (against each allele on wrong side of zone) ?'); readln(s)
       end;
    if habitat_dependent then
       begin
          writeln('s (against each allele in the wrong habitat) ?'); readln(s)
       end;
    if stabilising then
       begin
          writeln('s, Dopt (strength of stabilising seln, & change in optimum:  W = exp(-s(z-Dopt/2)^2/2)) ? ');
          readln(s, Dopt);
          repeat
             writeln('There are several ways of initialising the population: type 1, 2, 3 or 4');
             writeln('1:  all loci in a deme at the same frequency,so the mean is at the optimum');
             writeln('2:  all loci fixed; all demes the same, and the mean at the midpoint');
             writeln('3:  all loci fixed; mean at optimum, & as far as possible, same across the centre');
             writeln(' 4:  all loci fixed; mean at optimum, but different states across the centre');
             readln(initmethod)
          until (initmethod >= 1) and (initmethod <= 4);
       end;
    if epistasis or hetdis or epistasisp or geographic or habitat_dependent then
       begin
          repeat
             writeln('There are several possible starting frequencies: type 1, 2, 3 or 4');
             writeln('1:  a step; all loci in demes 0 to nmidleft are set to one, in demes nmidright to ndemes +1
                        to one; fixed ends');
             writeln('2:  proportion p of all individuals in each deme have all genes in state "1" , the rest have
                        genes in state "0"');
             writeln('3:   initial individual gene frequencies in each deme set at p=0.5 , on average');
             writeln('4:   to simulate a cline : note this option not yet coded! ');
             writeln('5:   Start with one individual in each deme fixed for "1", all others fixed for zero');
             readln(initfrequencies)
          until (initfrequencies >= 1) and (initfrequencies <= 5);
       end;
    if (initfrequencies = 2) or (initfrequencies = 3) then
       begin
          writeln('Proportion p (0<=p<=1)?');
          repeat  readln(prop)   until (prop >= 0) and (prop <= 1)
       end;
    writeln('Seed for random numbers ? ');
    readln(seed);

    if not just_reps then
    begin
       writeln('Write out zbar, Vgenic etc. for each deme (y) in addition to Dmax and width?');
       readln(ans); detailflag := (ans = 'y');
    end
    else detailflag:=false;

    writeln('Dump complete haplotypes (may use a lot of disc space) ? ');
```

```pascal
      readln(ans);
      dumpflag := (ans = 'y');

      if ((tmax - twarm) div dt > maxsamples) then
         begin
            writeln('TOO MANY SAMPLES (max=', maxsamples : 3, ')');
            writeln('Currently, tmax=', tmax : 3, ', dt=', dt : 3, ', twarm=', twarm : 3);
            writeln('Input new tmax, dt, twarm, such that (tmax-twarm) div dt <', maxsamples : 3);
            readln(tmax, dt, twarm);
         end;
      mig := nmig / ninds;

end;
{***************************** end  procedure  askstuff ****************************************}

procedure askstuff_batch; {reading in parameters from batch file}
   var
      ib: blocktype; gg: genetype; bt: bittype; genes: string; int1,int2,int3,int4:Longint;
begin
   readln(bf);  readln(bf, ans); ans := copy(ans, 1, 1); just_reps := (ans = 'y');
   readln(bf);  readln(bf, int1 {nreps}, int2 {tmax}, int3 {dt}, int4 {twarm});
   {range checking - it won't scream if outside limits}
   if not (int2 > 0) and not (int3 > 0) and not (int4 >= 0) then
      begin
         writeln('Error in time values; input Y to stop'); readln(tmax);
      end;
   nreps:=int1; tmax:=int2; dt:=int3; twarm:=int4;
   readln(bf); readln(bf, int1 {ndemes}, int2 {ninds});
   readln(bf); readln(bf,int3 {nkids});
   if not (int1 > 0) or not (int2 > 0) or not (int3>0) or not (int2<=maxinds) or not (int1<=maxdemes) then
      begin
         writeln('Error in ndemes; input Y to stop'); readln(tmax);
      end;
   ndemes:=int1;        {'Would you like 2 habitat types?'}
   ninds:=int2; nkids:=int3;
   readln(bf); readln(bf, ans); ans := copy(ans, 1, 1);
   if not (ans = 'y') and not (ans = 'n') then
      begin
         writeln('Error in 2 hab types?; input Y to stop'); readln(tmax);
      end;
   two_habitats := (ans = 'y');
   if two_habitats then
      begin
         readln(bf); readln(bf, hab_cline_width);
         if not (hab_cline_width >= 0) then
            begin
               writeln('Error in 2 hab types?; input Y to stop'); readln(tmax);
            end;
         if (hab_cline_width = 999) then  habslope := 0
         else
            if (hab_cline_width=0) then    habslope := 999
            else   habslope := 1 / hab_cline_width;
         readln(bf); readln(bf, habprefst);
         if not (habprefst >= 0) then
            begin
               writeln('Error in habprefstrength; input Y to stop'); readln(tmax);
            end;
      end
   else
      begin
         readln(bf); readln(bf); readln(bf); readln(bf);
      end;
   readln(bf); readln(bf, ngenes);
   if not (ngenes > 0) then
      begin
         writeln('Error in ngenes; input Y to stop'); readln(tmax);
      end;
   nblocks := 1 + ((ngenes - 1) div 16);
```

```pascal
nhabblocks := 1 + ((nhabgenes - 1) div 16);
for ib := 1 to nblocks do    allOne[ib] := -1;    {in binary, this is all ones}
for ib := 1 to nblocks do    allZero[ib] := 0;    {all zeroes}
readln(bf);                              {'Would you like some of the genes to be selected}
readln(bf, ans);                         {or to determine habitat preference ? '}
if not (ans = 'y') and not (ans = 'n') then
  begin
    writeln('Error in gene type; input Y to stop'); readln(tmax);
  end;
nsel := 0; nhabgenes := 0; nneut := 0; nlociclasses:=0; seln_mask := allZero; hab_mask := allZero;
if copy(ans, 1, 1) = 'y' then
  begin
    readln(bf, genes);
    for gg := 0 to ngenes - 1 do
      begin        {0 for neutral, 1 for selected genes, 2 for habitat preference}
        read(bf, i);
        if copy(genes, (gg + 1), 1) = 'n' then
          begin
            nneut := nneut + 1; if (nneut=1) then nlociclasses:=nlociclasses+1;
          end;
        if copy(genes, (gg + 1), 1) = 's' then
          begin
            nsel := nsel + 1; putgene(gg, 1, seln_mask);
            if (nsel=1) then nlociclasses:=nlociclasses+1;
          end;
        if copy(genes, (gg + 1), 1) = 'h' then
          begin
            nhabgenes := nhabgenes + 1; putgene(gg, 1, hab_mask);
            if (nhabgenes=1) then nlociclasses:=nlociclasses+1;
          end;
      end;
  end
else
  begin
    nlociclasses:=1; nneut:=ngenes; readln(bf);
  end;
if (nneut <> (ngenes - nsel - nhabgenes)) then scream;

readln(bf); readln(bf, rec, mutn);               {Recombination, mutation}
if not (rec > 0) or not (mutn >= 0) or not (nmig >= 0) or not (nmigbarr >= 0) then
  begin
    writeln('Error in rec/mig; input Y to stop');eadln(tmax);
  end;
readln(bf); readln(bf, nmig, nmigbarr);        {Migration rates}

{'Would you like epistasis ("e"), heterozygote disadv ("h"), geographic("g"), habt_dependt
("hd"),stabilising seln ("s") ?'}
readln(bf); readln(bf, ans); ans := copy(ans, 1, 1);
epistasis := (ans = 'e'); epistasisp := (ans = 'p'); hetdis := (ans = 'h');
stabilising := (ans = 's'); geographic := (ans = 'g'); habitat_dependent := (ans = 'b');
if not (epistasis or epistasisp) or (hetdis or stabilising) or (geographic or habitat_dependent) then
  begin
    writeln('Error in selection type; input Y to stop'); readln(tmax);
  end;
readln(bf);
if epistasis then   readln(bf, s, hwidth);
if epistasisp then    readln(bf, s, beta);
if hetdis or geographic or habitat_dependent  then   readln(bf, s);
readln(bf);
if stabilising then
  begin
    readln(bf, s, Dopt); readln(bf, initmethod);        {'initialisation method: 1 to 4'}
  end;
if hetdis or epistasis or epistasisp or geographic or habitat_dependent  then
    readln(bf, initfrequencies);
if (initfrequencies = 2) or (initfrequencies = 3) then
  begin
    readln(bf); readln(bf, prop);
```

```pascal
            end;
        if not ((initfrequencies > 0) and (initfrequencies <= 5)) then
            begin
                writeln('Error in initfrequencies; input Y to stop'); readln(tmax);
            end;

        readln(bf); readln(bf, seed);                {random seed}

        readln(bf);          {('Write out zbar, Vgenic etc. for each deme (y) as well as Dmax and width?');}
        readln(bf, ans); detailflag := (copy(ans, 1, 1) = 'y');
        readln(bf); readln(bf, ans); dumpflag := (copy(ans, 1, 1) = 'y');
        readln(bf); readln(bf, ans); dipflag := (copy(ans, 1, 1) = 'y');

        if ((tmax - twarm) div dt > maxsamples) then
            begin
                writeln('TOO MANY SAMPLES (max=', maxsamples : 3, ')');
                writeln('Currently, tmax=', tmax : 3, ', dt=', dt : 3, ', twarm=', twarm : 3);
                writeln('Input new tmax, dt, twarm, such that (tmax-twarm) div dt <', maxsamples : 3);
                readln(tmax, dt, twarm);
            end;
        mig := nmig / ninds;
    end;
{*************************** end  askstuff_batch **************************************}

    procedure set_listfile (var fl: text); {set up output file}
    begin
        writeln(fl, 'Simulating a multilocus cline (diploid individuals): ', date, ' ', time, ' Seed=', seed);
        writeln(fl, ' tmax=', tmax : 4, ' dt=', dt : 4, ' ndemes=', ndemes : 4, ' ninds =', ninds : 4);
        writeln(fl, ' ngenes =', ngenes : 4, ' nselected=', nsel : 4, 'nhabpref=', nhabgenes : 4);
        writeln(fl, ' nmig =', nmig:4, ' nmigbarr=', nmigbarr:4, ' mig =', mig:8:6, ' nkids =', nkids:4);
        if two_habitats then
            writeln(fl, 'habslope=', habslope : 8 : 4, ' habclinewidth=', hab_cline_width : 8 : 4,
            'habprefst=', habprefst : 8 : 4);
        if initfrequencies = 1 then    writeln(fl, 'Start with a step');
        if initfrequencies = 2 then    writeln(fl, 'Start with proportion', prop : 8 : 6, ' individuals fixed at
            all loci for "1", rest for "0", in all demes');
        if initfrequencies = 3 then
            writeln(fl, 'On average start with individual gene frequencies in each deme at p=', prop : 3);
        if initfrequencies = 4 then    writeln(fl, 'Start with a cline');
        if initfrequencies = 5 then
            writeln(fl, 'Start with one individual in each deme fixed for "1" all others fixed for zero');
        if epistasis then   writeln(fl, 'Epistasis: s =', s : 8 : 6, ' halfwidth for epistasis=', hwidth : 8 : 6);
        if epistasisp then    writeln(fl, 'Epistasis: s =', s : 8 : 6, ' beta=', beta : 8 : 6);
        if hetdis then    writeln(fl, 'Heterozygote disadvantage: s =', s : 8 : 6);
        if geographic then   writeln(fl, 'Geographic seltn gradient: s=', s : 8 : 6);
        if habitat_dependent then writeln(fl, 'Habitat dependent: s=', s : 8 : 6);
        if stabilising then writeln(fl, 'Stabilising selection: s =',s:8:6,'change in optimum:', Dopt:8: 2);
        if initmethod = 1 then
            writeln(fl, ' Start with all loci in a deme at the same frequency; mean at the local optimum');
        if initmethod = 2 then
            writeln(fl, 'Start with all loci fixed at 0 or 1; all demes the same,  mean at overall midpoint');
        if initmethod = 3 then writeln(fl, ' Start with all loci fixed at 0 or 1; mean is at the local optimum,
                but as far aspossible, loci are in the same state on either side');
        if initmethod = 4 then writeln(fl, ' Start with all loci fixed at 0 or 1; mean is at the local optimum,
                and loci are assigned 0 or 1 independently on either side');
        writeln(fl, ' rec =', rec : 8 : 6, ' mutn =', mutn : 8 : 6, ' seed =', seed : 4);
        writeln(fl, 'Output file corresponding to the above parameters has root: ', fname_root);
        writeln(fl, 'Batch file was: ', batch_fname);
        writeln(fl, 'Number of replicates = ',nreps:4);
        writeln(fl, 'From 3/10/96, predD is calculated using cline slope, not width');
    end;

{*************************** end  setlist_file **************************************}


    procedure update_stats;    {updates the list of statistics:collected every dt generations}
{Statistics are calculated separately for each of the 3 locus types: 1=neutral; 2=selected; 3=habitat
preference : (i) separately for each habitat, or (ii) across both habitats:
```

the same routines are used for (i) and (ii), by passing a dummy habpop array defining all
individuals to be in the same habitat for (ii). }

```
    var
      d, deme_mid,deme_deltap_max_: demep1type;
      hab: bittype; loci:locitype; g: genetype; wb2h, vw2h, rloci, deltap_max_: real;

begin
  for loci:=1 to 3 do
    begin
      for d := 0 to ndemes + 1 do
        begin   {NB. dummyhabpop is all 1's}
          zbar_2h[loci]^[d] := getzbar(pop,dummyhabpop, dummyhabpop_size, d, 1, loci);
        Vgenic_2h[loci]^[d] := genic_var(pop, dummyhabpop, dummyhabpop_size, d, 1, loci);
          Vtotal_2h[loci]^[d] := total_var(pop,dummyhabpop, dummyhabpop_size, d, 1,loci);
          Fis_2h[loci]^[d] := heterozyg_deficit(pop, dummyhabpop, dummyhabpop_size, d, 1, loci);
          Vdiseq_2h[loci]^[d] := Vtotal_2h[loci]^[d] - Vgenic_2h[loci]^[d];
          for hab := 0 to 1 do
            if habpop_size^[d, hab] > 0 then
              begin
                sample_count[d,hab]:=sample_count[d,hab]+1;
                zbar[loci]^[d, hab] := getzbar(pop, habpop, habpop_size, d, hab, loci);
                Vgenic[loci]^[d, hab] := genic_var(pop, habpop, habpop_size, d, hab, loci);
                Vtotal[loci]^[d, hab] := total_var(pop, habpop, habpop_size, d, hab, loci);
                Fis[loci]^[d, hab] := heterozyg_deficit(pop, habpop, habpop_size, d, hab, loci);
                Vdiseq[loci]^[d, hab] := Vtotal[loci]^[d, hab] - Vgenic[loci]^[d, hab];
              end
            else
              begin   {999 is a missing value, representing no individuals in that habitat}
                zbar[loci]^[d, hab] := 999; Vgenic[loci]^[d, hab] := 999; Fis[loci]^[d, hab] := 999;
                Vdiseq[loci]^[d, hab] := 999; Vtotal[loci]^[d, hab] := 999;
              end;   {end habitat loop}
          if (two_habitats and (habpop_size^[d, 1] > 0)) and (habpop_size^[d, 0] > 0) then
            delta_p[loci]^[d] := (zbar[loci]^[d, 1] - zbar[loci]^[d, 0])/(2*nloci[loci])
            else delta_p[loci]^[d] :=999;
        end;      {deme loop}
      width_2h[loci]^[sample] := get_width_2h(zbar_2h[loci], sample, nloci[loci]);
      slope_2h[loci]^[sample] := slope_2hL(pop, loci, 1.38629);
      width2_2h[loci]^[sample] := NEWwidth_2hL(pop, loci, 1.38629);
      predD_2h[loci]^[sample] := mig*slope_2h[loci]^[sample]*slope_2h[loci]^[sample]/all_rec[loci];
      rloci:=all_rec[loci];
      predD_2h[loci]^[sample] := predD(slope_2h[loci], sample, rloci, mig);
      Dmax_2h[loci]^[sample] := get_Dmax_2h(Vdiseq_2h[loci], nloci[loci]);
      for hab := 0 to 1 do
        if two_habitats or (hab = 1) then
          begin
            Dmax_reg[loci]^[sample, hab] :=
              get_Dmax_reg(Vgenic[loci], Vdiseq[loci], sample, hab, nloci[loci]);
            Dmax[loci]^[sample, hab] := get_Dmax(Vdiseq[loci], hab, nloci[loci]);
            Dmean[loci]^[sample,hab]:= get_Dmean(Vdiseq[loci], sample, hab, nloci[loci]);
            Fis_max[loci]^[sample, hab] := get_max(Fis[loci], hab);
          end;
      if two_habitats then
        begin
          get_deltap_max(delta_p[loci],deltap_max_,deme_deltap_max_);
          deltap_max[loci]^[sample] := deltap_max_;
        end;
    end; {locus type loop}
  if nloci[selected] > 0 then      {Get fitness in each deme}
    begin
      for d := 0 to ndemes + 1 do
        begin
          find_wbar(pop, habpop, habpop_size, d, 1, true, wb2h, vw2h); wbar_2h^[d] := wb2h;
          wbar_2hMEAN^[d] := wbar_2hMEAN^[d]+wb2h; varw_2h^[d] := vw2h;
        end;
    end;

  if (nloci[selected] > 0) then            {Get minimum fitness}
```

```
                    begin
                        minimum_fitness(wbarmin_, wbarreg_,zbar_2h[selected], wbar_2h, sample,
                            nloci[selected],deme_mid);
                        wbarmin^[sample] := wbarmin_; wbarreg^[sample] := wbarreg_;
                    end;

                for loci:=1 to 3 do if nloci[loci]>0 then        {Get max Fis. Only want to max from central region.....}
                    begin
                        for hab:=0 to 1 do
                            begin {if no obvious central region, just take mid-point deme}
                            if (two_habitats and (hab_cline_width=999)) or (nloci[selected]=0) or (wbarmin_=1) then
                                Fis_central_max[loci]^[sample,hab]:=get_max_Fis_centre(Fis[loci],hab,round(ndemes/2))
                            else {o'wise define centre as area of maximum fitness reduction - allows clines to have drifted}
                                Fis_central_max[loci]^[sample,hab]:=get_max_Fis_centre(Fis[loci],hab,deme_mid);
                            Fis_deltap_max[loci]^[sample,hab]:=Fis[loci]^[deme_deltap_max_,hab];
                        end;
                    Fis_central_max_2h[loci]^[sample] := get_max_Fis_centre_2h(Fis_2h[loci],deme_mid);   {rerere}
                        deltap_centre[loci]^[sample]:=delta_p[loci]^[deme_mid];
                    end;
                end;

{*********************** end procedure update_stats **************************}


    procedure update_barrier;   {barrier calculations are tedious, so kept separate}
        var
            mean_slope_:real;
        begin
            calculate_barrier(pop,step_neut_, Step2_neut_,slope_left_neut_, slope_mid_neut_,
slope_right_neut_, NEWslope_left_neut_, NEWslope_right_neut_, zbar_2h[neutral], wbar_2h, sample,
nloci[neutral]);
            step_neut^[sample] := step_neut_;
            if width_2h[neutral]^[sample] <> 999 then
                begin
                    slope_edge^[sample]:=(slope_left_neut_ + slope_right_neut_)/2;
                    NEWslope_edge^[sample]:=(NEWslope_left_neut_ + NEWslope_right_neut_)/2;
                    slope_ratio^[sample] :=slope_edge^[sample] / slope_mid_neut_;
                    NEWslope_ratio^[sample] :=NEWslope_edge^[sample] / slope_mid_neut_;
                    if slope_edge^[sample]<>0 then barrier^[sample]:=step_neut_/slope_edge^[sample]
                    else barrier^[sample]:=999;
                    if NEWslope_edge^[sample]<>0 then
                        Barrier2^[sample]:=Step2_neut_/NEWslope_edge^[sample]
                    else Barrier2^[sample]:=999;
                    if slope_mid_neut_<>0 then central_d5_slope^[sample] := slope_mid_neut_;
                end
            else
                begin
                    slope_edge^[sample] := 999; slope_ratio^[sample] := 999; central_d5_slope^[sample] := 999;
                    NEWslope_edge^[sample] := 999; NEWslope_ratio^[sample] := 999; barrier^[sample]:=999;
                    Barrier2^[sample]:=999;
                end;
        end;
{*********************** end procedure update_barrier **************************}


    procedure update_drawing; {graphics procedures listed by Jackson (1992)}
    begin
        useDrawWindow;
        clear_Display;
        draw_parameters(hwidth, beta, s, rec, mutn, habslope, habprefst);
        draw_zbar_habs(habpop_size);
        draw_wbar;
        draw_pind(pop, dummyhabpop, dummyhabpop_size, 1);
        draw_Vd_habs(habpop_size, s, mutn);
        draw_Vg_habs(habpop_size, s, mutn);
    end;
{*********************** end procedure update_drawing **************************}
```

```pascal
procedure write_QTfull (dipflg: boolean);
{using same procedure as OLD write_full, but only writing out first 10 neutral markers}
    var
        d: demep1type; g: genetype; ind1, ind2: indtype; dipl: diplotype;
    begin
      writeln(dump_file, 'Generation ', t : 8);
      if nneut > 0 then
        begin
          writeln(dump_file);
          if dipflag then    writeln(dump_file, 'Generating diploid data');
          writeln(dump_file, 'Data on ', nneut : 4, ' neutral loci');
          write(dump_file, 'Deme', chr(9), ' ind', chr(9),'Maternal',chr(9));
          for g := 0 to ngenes - 1 do
            if (getgene(g, seln_mask) = 0) and (getgene(g, hab_mask) = 0) then
              write(dump_file, 'gene', g : 3, chr(9));
          write(dump_file,'Paternal',chr(9));
          for g := 0 to ngenes - 1 do
            if (getgene(g, seln_mask) = 0) and (getgene(g, hab_mask) = 0) then
              write(dump_file, 'gene', g : 3, chr(9));
          writeln(dump_file);
          for d := 0 to ndemes + 1 do
            begin
              for ind1 := 1 to ninds do
              begin
                write(dump_file, d : 4, chr(9), ind1 : 4, chr(9),chr(9));
                dipl := pop^[d, ind1];
                for g := 0 to ngenes - 1 do
                  if ((getgene(g, seln_mask) = 0) and (getgene(g, hab_mask) = 0)) then
                    write(dump_file, getgene(g, dipl[0]) : 2, chr(9));
                write(dump_file, chr(9), chr(9));
                for g := 0 to ngenes - 1 do
                  if ((getgene(g, seln_mask) = 0) and (getgene(g, hab_mask) = 0)) then
                    write(dump_file, getgene(g,dipl[1]) : 2, chr(9));
                writeln(dump_file);
              end;
              writeln(dump_file)
            end;
          writeln(dump_file)
        end;
    end;
{********************************** end procedure writeQT_full ********************}


procedure write_output (full: boolean);
{Writes out a list of statistics to output file;  if full is set, lists all the individual logit transformed allele
frequencies. If sample = 0, these data give overall means}
{Note: update-stats must already have been run, to set up the data}
    var
        d, start_deme, end_deme, mid_deme: demep1type; g: genetype; hab: bittype;
        wb, vw, sdw_2h: real; loci:locitype; wanted:boolean;

    procedure write_output_sample_zero;    {Sample 0 contains all the means}
      var
        g: genetype; hab: bittype; wb, vw, sdw_2h, sdpD_2h, sdd, sdf, sdstep, sdStep2, rb: real;
        sdNw_2h, sdpD_2h: real; sdsp_2h, sdslrr, sdslrl: real; l:locitype; samp:integer;
        sddm, sdfisc, sdfism, sdfisdp: bitrealstype;
        sdNEWsl, sdsl, sdslr, sdbr, sdNEWbr, sddp_max, sddp_c:real;

    procedure write_output_sample_zero_barrier;
        begin
          writeln(out_file, 'Barrier:');
          calculate_barrier(step_neut_, Step2_neut_, slope_left_neut_, slope_mid_neut_,
    slope_right_neut_, NEWslope_left_neut_, NEWslope_right_neut_,
          zbar_2h[neutral], wbar_2h, sample, nloci[neutral]);
          sdstep := sqrt(variance1(step_neut, nsamples));
          write(out_file, ' step ', step_neut^[0] :10:6, ' sd ', sdstep :10:6);
          writeln(out_file, ' 95% limits +/- ', 2 * sdstep / sqrt(nsamples):10:6);
```

```pascal
            sdstep := sqrt(variance1(Step2_neut, nsamples));
            write(out_file, ' Step2 ', Step2_neut^[0] :10:6, ' sd ', sdstep :10:6);
            writeln(out_file, ' 95% limits +/- ', 2 * sdstep / sqrt(nsamples) :10:6);
            writeln(out_file);
            sdsl:=sqrt(variance1(slope_edge, nsamples));
            sdNEWsl:=sqrt(variance1(NEWslope_edge, nsamples));
            writeln(out_file,'Edge slope: ',slope_edge^[0]:10:6,' 95% ci',
                2*sdsl/sqrt(nsamples):10:6);
            writeln(out_file,'NEW Slope at edge: ',NEWslope_edge^[0]:10:6,' 95% ci',
                 2 * sdNEWsl/sqrt(nsamples):10:6);
            writeln(out_file);

            writeln(out_file,'Barrier=delta u/slope edge: ');
            sdbr:=sqrt(variance1(barrier, nsamples));
            sdNEWbr:=sqrt(variance1(Barrier2, nsamples));
            writeln(out_file,'Barrier:',barrier^[0]:10:6,' 95% ci',2 * sdbr/sqrt(nsamples):10:6);
            writeln(out_file,'Barrier2:',Barrier2^[0]:10:6,' 95% ci', 2 * sdNEWbr/sqrt(nsamples):10:6);
            writeln(out_file);

            writeln(out_file, 'All slope ratios calculated using central slope, from regression over
                       5 central demes in calculate_barrier');
            sdsl:=sqrt(variance1(central_d5_slope,nsamples));
            writeln(out_file,'Central slope',central_d5_slope^[0]:10:6,' sd ',sdsl:10:6,' 95% ci ',
                       2*sdsl/sqrt(nsamples):10:6);
            writeln(out_file);

            sdslr := sqrt(variance1(slope_ratio, nsamples));
            sdNEWslr := sqrt(variance1(NEWslope_ratio, nsamples));
            write(out_file, 'Slope ratio ', slope_ratio^[0]:10:6, ' sd ', sdslr:10:6);
            writeln(out_file, ' 95% limits +/- ', 2 * sdslr / sqrt(nsamples) :10:6);
            write(out_file, 'NEW Slope ratio ', NEWslope_ratio^[0]:10:6, ' sd ', sdNEWslr:10:6);
            writeln(out_file, ' 95% limits +/- ', 2 * sdNEWslr / sqrt(nsamples) :10:6);
        end;
{********************************************************}

    procedure write_output_sample_zero_minw;
        var
          sdwbarmin, sdwbarreg,rb:real;
        begin
          rb := rbar(seln_mask, hab_mask, rec);
          sdwbarmin:=sqrt(variance1(wbarmin, nsamples));
          sdwbarreg:=sqrt(variance1(wbarreg, nsamples));
          write(out_file, 'Mean minimum fitness:', wbarmin^[0]:8:6, ' 95% limits +/- ',
            2*sdwbarmin/sqrt(nsamples):10:6);
          writeln(out_file, '; by reg:', wbarreg^[0]:8:6, ' 95% limits +/- ',
            2*sdwbarreg/sqrt(nsamples));
          writeln(out_file, 'Harmonic mean r is ', rb:8:6, ';  expected slope ratio: wbarmin=',
            power(wbarmin^[0],1/rb):10:6, '; wbarreg=', power(wbarreg^[0],1/rb):10:6);
          writeln(out_file);
        end;

{********************************************************}

    begin
      writeln(out_file);
      writeln(out_file, 'Overall means after ', t : 8, ' generations: samples every ', dt : 4, ' gens,
        starting at ', twarm : 6);
      writeln(out_file);
      if detailflag then   writeln(out_file, 'Linkage disequilibrium:')
      else   writeln(out_file, 'Means and 95% confidence intervals');
      for l:=1 to 3 do
       if nloci[l] > 0 then
         begin
           for hab := 0 to 1 do
             if ((hab = 0) and two_habitats) or (hab = 1) then
               begin
                  sdd := sqrt(variance1h(Dmax_reg[l], hab, nsamples));
                  sddm[hab] := sqrt(variance1h(Dmax[l], hab, nsamples));
```

```pascal
                    sdfism[hab] := sqrt(variance1h(Fis_max[l], hab, nsamples));
                    sdfisc[hab] := sqrt(variance1h(Fis_central_max[l], hab, nsamples));
                    sdfisdp[hab] := sqrt(variance1h(Fis_deltap_max[l], hab, nsamples));
                    if true {detailflag} then
                      begin
                        case l of
                          1:writeln(out_file, 'Data on ',nloci[l]:4, ' neutral loci in habitat ', hab, ':');
                          2:writeln(out_file, 'Data on ',nloci[l]:4, ' selected loci in habitat ', hab, ':');
                          3:writeln(out_file, 'Data on ',nloci[l]:4, ' hab pref loci in habitat ', hab, ':');
                        end;
                        write(out_file,'Fis_max ',Fis_max[l]^[0, hab]:10:6, ' sd ', sdfism[hab] :10:6);
                        writeln(out_file,'95%limits+/-',2*sdfism[hab]/sqrt(nsamples):10:6);
                        write(out_file,'Fis_central_max',Fis_central_max[l]^[0,hab]:10:6,
                                'sd',sdfisc[hab]:10:6);
                        writeln(out_file,'95%limits+/-',2*sdfisc[hab]/sqrt(nsamples):10:6);
                        write(out_file,'Fis_deltap_max',Fis_deltap_max[l]^[0,hab]:10:6,
                                'sd',sdfisdp[hab]:10:6);
                        writeln(out_file,'95%limits+/-',2*sdfisdp[hab]/sqrt(nsamples):10:6);
                        writeln(out_file);
                        if nloci[l]>1 then
                          begin
                            write(out_file,'Dmax_reg',Dmax_reg[l]^[0,hab]:10:6,'sd',sdd:10:6);
                            writeln(out_file,'95%limits+/-',2*sdd/sqrt(nsamples):10:6);
                            write(out_file,'Dmax',Dmax[l]^[0,hab]:10:6,'sd',sddm[hab]:10:6);
                            writeln(out_file,'95%limits+/-',2*sddm[hab]/sqrt(nsamples):10:6);
                          end;
                        writeln(out_file);
                      end
                    else
                      begin
                        writeln(out_file,'Loci type ',l:2,': Fismaxcentre, Fisdeltapmax,Dreg, Dmax,
                           slope, w, predD');
                        write(out_file,Fis_central_max[l]^[0, hab]:10:6,chr(9),2 * sdfisc[hab] /
                            sqrt(nsamples) :10:6,chr(9));
                        write(out_file, Fis_deltap_max[l]^[0, hab] :10:6, 2 * sdfisdp[hab] /
                           sqrt(nsamples) :10:6, nsamples : 6, ' samples');
                        write(out_file, Dmax_reg[l]^[0, hab]:10:6,chr(9), 2*sdd/sqrt(nsamples):10:6);
                        writeln(out_file, Dmax[l]^[0, hab]:10:6,chr(9),  2*sddm[hab]/sqrt(nsamples):10:6);
                      end;
                  end;
                if two_habitats then
                  begin
                    writeln(out_file,'Average of two habitats, assigning equal weight:');
                    write(out_file, ' Fis_max ', (Fis_max[l]^[0, 0]+Fis_max[l]^[0,1])/2 :10:6);
                    writeln(out_file,'95%ci',2*sqrt((sqr(sdfism[0])+sqr(sdfism[1]))/2)/sqrt(2*nsamples):10:6);
                    write(out_file, ' Fis_central_max ',
                        (Fis_central_max[l]^[0, 0]+Fis_central_max[l]^[0,1])/2 :10:6);
                    writeln(out_file,'95%ci',2*sqrt((sqr(sdfisc[0]) + sqr(sdfisc[1])) / 2)/sqrt(*nsamples) :10:6);
                    write(out_file, ' Fis_deltap_max ',
                        (Fis_deltap_max[l]^[0, 0]+Fis_deltap_max[l]^[0,1])/2 :10:6);
                    writeln(out_file, ' 95%ci ',
                        2 * sqrt((sqr(sdfisdp[0]) + sqr(sdfisdp[1])) / 2) / sqrt(2 * nsamples) :10:6);
                    if nloci[l]>1 then
                      begin
                        write(out_file, ' Dmax ', (Dmax[l]^[0,0]+Dmax[l]^[0,1])/2:10:6);
                        writeln(out_file, ' 95%ci ',
                            2 * sqrt((sqr(sddm[0]) + sqr(sddm[1])) / 2) / sqrt(2 * nsamples) :10:6);
                      end;
                  end;
                writeln(out_file);
          sdw_2h := sqrt(variance1(width_2h[l], nsamples));
          sdsp_2h := sqrt(variance1(slope_2h[l], nsamples));
          sdNw_2h := sqrt(variance1(width_2h[l], nsamples));
          sdpD_2h := sqrt(variance1(predD_2h[l], nsamples));
          sddp_max := sqrt(variance1(deltap_max[l],nsamples));
          sddp_c := sqrt(variance1(deltap_centre[l],nsamples));
          sdd:=sqrt(variance1(Dmax_2h[l],nsamples));
          sdf:=sqrt(variance1(Fis_central_max_2h[l],nsamples));
```

```pascal
              write(out_file, ' width ', width_2h[l]^[0] :10:6, '  sd  ', sdw_2h :10:6);
              writeln(out_file, '  95% limits +/- ', 2 * sdw_2h / sqrt(nsamples) :10:6);
              write(out_file, ' Slope ', slope_2h[l]^[0] :10:6, '  sd  ', sdsp_2h :10:6);
              writeln(out_file, '  95% limits +/- ', 2 * sdsp_2h / sqrt(nsamples) :10:6);
              write(out_file, ' NEWwidth ', width_2h[l]^[0] :10:6, '  sd ', sdNw_2h :10:6);
              writeln(out_file, '  95% limits +/- ', 2 * sdNw_2h / sqrt(nsamples) :10:6);
              write(out_file, ' Pred D ', predD_2h[l]^[0] :10:6, '  sd  ', sdpD_2h :10:6);
              writeln(out_file, '  95% limits +/- ',  2 * sdpD_2h / sqrt(nsamples) :10:6);
              writeln(out_file);
              writeln(out_file, 'Dmax across both habitats ',Dmax_2h[l]^[0]:10:6,' 95%ci ',
                        2*sdd/sqrt(nsamples):10:6);
              writeln(out_file, 'Fis_central_max_2h across both habitats ',
                        Fis_central_max_2h[l]^[0]:10:6, ' 95%ci ', 2*sdf/sqrt(nsamples):10:6);
              writeln(out_file);
              if two_habitats then
                begin
                  writeln(out_file, ' Max. difference between habitats ', deltap_max[l]^[0]:10:6,
                            chr(9),' 95% ci: ',chr(9), 2*sddp_max/sqrt(nsamples):10:6);
                  writeln(out_file, ' Difference at centre ',deltap_centre[l]^[0]:10:6, chr(9),'
            95% ci: ',chr(9), 2*sddp_c/sqrt(nsamples):10:6);
                end;
              writeln(out_file);
            end; {all loci types}
          writeln(out_file);
          if (nsel>0) then write_output_sample_zero_minw;
          if (nsel > 0) and (nneut > 0) then write_output_sample_zero_barrier;
          writeln(out_file,'Overall harmonic mean recomb rate=',all_rec[all]:8:6);
          if nneut>0 then writeln(out_file,'Mean r betw. neutral loci=',all_rec[neutral]:8:6);
          if nsel>0 then writeln(out_file,'Mean r between selected loci=',all_rec[selected]:8:6);
          if nhabgenes>0 then writeln(out_file,'Mean r betw. habpref loci=',all_rec[habpref]:8:6);
          writeln(out_file,'Number of loci: neutral, selected = ',nneut:4,chr(9),nsel:4);
          writeln(out_file,nloci[neutral]:4,chr(9),nloci[selected]:4,'Nloci  classes ',nlociclasses:4);
        end;

{XXXXXXXXXXXXXXXXXXXXXXXXXXwrite_output_sample_zero_means xxxxxxxxxxxxxxxxxxxxx}


      begin
        if sample=0 then      write_output_sample_zero;        {i.e. at end, having calculated means}
        if sample=1 then
          begin
            writeln(out_file,'NB. Ratio now calculated using slope estimate, not width! 1/11/96');
            writeln(out_file, 'REPLICATE ', replicate : 8);
            writeln(out_file);
          end;
        mid_deme:=round(ndemes/2);  {No point writing out stats for demes which are at fixation...}
        start_deme:=0;    {.... so find start point and end point}
        end_deme:=ndemes+1;
        if sample >0 then
          begin
            repeat start_deme:=start_deme+1
            until (zbar_all^[start_deme+2]>0) or (start_deme=mid_deme-2);
            repeat end_deme:=end_deme-1
            until (zbar_all^[end_deme-2]<2*nloci[0]) or (end_deme=mid_deme+2);
          end;
        for loci:=1 to 3 do
          if (nloci[loci]>0) then
          begin
            if ((sample=1) or detailflag) then
            case loci of
              1:writeln(out_file, 'Data on ', nloci[1] : 4, ' neutral loci');
              2:writeln(out_file, 'Data on ', nloci[2] : 4, ' selected loci');
              3:writeln(out_file, 'Data on ', nloci[3] : 4, ' hab pref loci');
            end;
        if ((t>(tmax-dt)) and (sample>1)) or (detailflag or (sample=0)) then
            begin
              if two_habitats then
                  write(out_file,'Time',chr(9),'Deme',chr(9), 'Habitat', chr(9), 'Hab-pop size', chr(9))
```

```
            else   write(out_file, 'Time', chr(9), 'Deme', chr(9));
        if full then
            for g := 0 to ngenes - 1 do
                begin
                    wanted:=((loci=selected) and (getgene(g,seln_mask)=1))
                        or ((loci=habpref) and (getgene(g,hab_mask)=1)) or ((loci=neutral) and
                            (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
                    if wanted then   write(out_file, ' gene', g : 3, chr(9));
                end;
        if two_habitats then
            if true or (loci=selected) then
                writeln(out_file, ' mean p', chr(9), ' logit p', chr(9), ' Vgenic ', chr(9), ' Vdiseq ',
                    chr(9), ' Fis', chr(9), ' Wbar ', chr(9), '  varW ', chr(9), ' Habavail', chr(9),
                    'p_hab1-p_hab0', chr(9), 'Vdiseq_2h', chr(9), 'Fis_2h')
            else  writeln(out_file, ' mean p', chr(9), ' logit p', chr(9), ' Vgenic ', chr(9), '
                    Vdiseq ',      chr(9), ' Fis', chr(9),' Habavail', chr(9), 'Zhab1-Zhab0')
        else
            if true or (loci=selected) then
                writeln(out_file, ' mean p', chr(9), ' logit p', chr(9), ' Vgenic ', chr(9), ' Vdiseq ',
                    chr(9), ' Fis', chr(9), ' Wbar ', chr(9), '  varW ')
            else
            writeln(out_file, ' mean p', chr(9), ' logit p', chr(9), ' Vgenic ', chr(9), ' Vdiseq ',
                    chr(9), ' Fis');
        for hab := 0 to 1 do
            if ((hab = 0) and two_habitats) or (hab = 1) then
                begin
                    for d := start_deme to end_deme do
                        begin
                            if two_habitats then
                                write(out_file, t : 4, chr(9), d : 4, chr(9), hab : 4, chr(9),
                                    habpop_size^[d,hab], chr(9))
                            else
                                write(out_file, t : 4, chr(9), d : 4, chr(9));
                            if full then
                                for g := 0 to ngenes - 1 do
                                    begin
                                        wanted:=((loci=selected) and (getgene(g,seln_mask)=1))
                                            or ((loci=habpref) and (getgene(g,hab_mask)=1)) or ((loci=neutral)
                                            and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
                                        if wanted then
                                            write(out_file, getgenefreq(pop,habpop,habpop_size,d,hab,g));
                                    end;
                            write(out_file, zbar[loci]^[d, hab]/ (2 * nloci[loci]) : 8 : 3, chr(9));
                            write(out_file, logit(zbar[loci]^[d, hab] / (2 * nloci[loci]), 5) : 8 : 4, chr(9));
                            write(out_file, Vgenic[loci]^[d, hab] : 8 : 3, chr(9));
                            write(out_file, Vdiseq[loci]^[d, hab] : 8 : 3, chr(9));
                            write(out_file, Fis[loci]^[d, hab] : 8 : 3, chr(9));
                            if true or (loci=selected) then
                                begin
                                    write(out_file, wbar_2h^[d] : 8 : 5, chr(9));
                                    write(out_file, varw_2h^[d] : 8 : 5, chr(9));
                                end;
                            if two_habitats then
                                begin
                                    write(out_file, habitat_availability^[d, hab], chr(9));
                                    write(out_file,(zbar[loci]^[d,1]/(2*nloci[loci])-zbar[loci]^[d,0]/(2*nloci[loci])));
                                    write(out_file, Vdiseq_2h[loci]^[d]:10:6, chr(9),  Fis_2h[loci]^[d]:10:6);
                                end;
                            writeln(out_file);
                        end;
                end;
    end;
    if ((sample=1) or detailflag) then
        if two_habitats then
        writeln(out_file, 'time; 1/slope; NEWw; deltap_max, deltap_centre; both_hab
Fis_central_max; hab0 Fis_max, Fis_central_max, Fis_deltap_max; hab1 Fis_max, Fis_central_max,
Fis_deltap_max; both_hab Dmax; hab0:Dmax_reg,Dmax; hab1:Dmax_reg,Dmax; predD(NEWw_p8):')
        else
```

```
              writeln(out_file, 'time; width; 1/slope, NEWw, Fis_central_max, Fis_max, Dmax_reg,
Dmax,{Dmean}, predD(NEWw):');
          if two_habitats then
            begin
              write(out_file, t, chr(9));
              write(out_file, (1/slope_2h[loci]^[sample]):10:6, chr(9));
              write(out_file, width_2h[loci]^[sample]:10:6, chr(9));
              write(out_file, deltap_max[loci]^[sample]:10:6, chr(9));
              write(out_file, deltap_centre[loci]^[sample]:10:6, chr(9));
              write(out_file, Fis_central_max_2h[loci]^[sample]:10:6,chr(9));
              write(out_file, Fis_max[loci]^[sample,0]:10:6, chr(9));
              write(out_file, Fis_central_max[loci]^[sample,0]:10:6, chr(9));
              write(out_file, Fis_deltap_max[loci]^[sample,0]:10:6, chr(9));
              write(out_file, Fis_max[loci]^[sample,1]:10:6, chr(9));
              write(out_file, Fis_central_max[loci]^[sample,1]:10:6, chr(9));
              write(out_file, Fis_deltap_max[loci]^[sample,1]:10:6, chr(9));
              if (nloci[loci]>1) then
                begin
                  write(out_file, Dmax_2h[loci]^[sample]:10:6,chr(9));
                  write(out_file, Dmax_reg[loci]^[sample,0]:10:6, chr(9),Dmax[loci]^[sample,0]:10:6);
                  write(out_file, Dmax_reg[loci]^[sample,1]:10:6, chr(9), Dmax[loci]^[sample,1]:10:6);
                  writeln(out_file, predD_2h[loci]^[sample]:10:6, chr(9))
                end
              else writeln(out_file);
            end
          else
            begin
              write(out_file, t, chr(9), width_2h[loci]^[sample]:10:6, chr(9));
              write(out_file, (1/slope_2h[loci]^[sample]):10:6, chr(9));
              write(out_file, width_2h[loci]^[sample]:10:6, chr(9));
              write(out_file, Fis_central_max[loci]^[sample,1]:10:6, chr(9));
              write(out_file, Fis_max[loci]^[sample,1]:10:6, chr(9));
              if (nloci[loci]>1) then
                begin
                  write(out_file, Dmax_reg[loci]^[sample,1]:10:6, chr(9));
                  write(out_file, Dmax[loci]^[sample,1]:10:6, chr(9));
                  writeln(out_file, predD_2h[loci]^[sample]:10:6, chr(9))
                end
              else writeln(out_file);
            end
      end; {end of all loci types loop}

    if (nloci[neutral] > 0) and (nloci[selected] > 0) then
      begin
        update_barrier;
        write(out_file, 'Step size, Step2,edge slope,  Edge2 slope,central slope);
        writeln(out_file,'slope ratio,NEWslope ratio, barrier,Barrier2,wbar_min, wbar_reg);
        write(out_file,step_neut^[sample]:10:6,chr(9),Step2_neut^[sample]:10:6, chr(9));
        write(out_file, slope_edge^[sample]:10:6, chr(9), NEWslope_edge^[sample]:10:6,
              chr(9), central_d5_slope^[sample]:10:6, chr(9));
        write(out_file, slope_ratio^[sample]:10:6, chr(9), NEWslope_ratio^[sample]:10:6);
        write(out_file, barrier^[sample]:10:6, chr(9), Barrier2^[sample]:10:6, chr(9));
        writeln(out_file, wbarmin^[sample] :10:6, chr(9), wbarreg^[sample] :10:6);
      end;
    writeln(out_file);
  end;
{*************************** end  procedure  write_output  ***************************}

  procedure set_sample_zero;    {Sets data for the fictional sample zero to the overall mean}
    var
      d: demep1type; i: indtype; g: genetype; hab: bittype; l:locitype;
  begin
    for d:=0 to ndemes+1 do
      begin
        for hab:=0 to 1 do sample_count[d,hab]:=round(sample_count[d,hab]/nlociclasses);
        if (sample_count[d,0]>nsamples) or (sample_count[d,1]>nsamples) then scream;
      end;
    for hab := 0 to 1 do
```

```
              begin
                for I:=1 to 3 do
                  if nloci[I] > 0 then
                    begin
                      Dmax_reg[I]^[0, hab] := mean1h(Dmax_reg[I], hab, nsamples);
                      Dmax[I]^[0, hab] := mean1h(Dmax[I], hab, nsamples);
                      Fis_max[I]^[0, hab] := mean1h(Fis_max[I], hab, nsamples);
                      Fis_central_max[I]^[0, hab] := mean1h(Fis_central_max[I], hab, nsamples);
                      Fis_deltap_max[I]^[0, hab] := mean1h(Fis_deltap_max[I], hab, nsamples);
                    end;
                end;   {habitat loop}
              if nloci[selected]>0 then
                begin
                  wbarmin^[0] := mean1(wbarmin, nsamples); wbarreg^[0] := mean1(wbarreg, nsamples);
                  for d:=0 to ndemes+1 do
                    begin
                      wbar_2h^[d] := wbar_2hMEAN^[d]/nsamples; varw_2h^[d] := varw_2hMEAN^[d]/nsamples;
                    end;
                end;
              for I:=1 to 3 do
                if nloci[I]>0 then
                  begin
                    width_2h[I]^[0] := mean1(width_2h[I], nsamples);
                    width_2h[I]^[0] := mean1(width_2h[I], nsamples);
                    slope_2h[I]^[0] := mean1(slope_2h[I], nsamples);
                    predD_2h[I]^[0] := mean1(predD_2h[I], nsamples);
                    if two_habitats then
                      begin
                        deltap_max[I]^[0] := mean1(deltap_max[I], nsamples);
                        deltap_centre[I]^[0]:=mean1(deltap_centre[I], nsamples);
                      end;
                    Dmax_2h[I]^[0]:=mean1(Dmax_2h[I],nsamples);
                    Fis_central_max_2h[I]^[0]:=mean1(Fis_central_max_2h[I],nsamples);
                  end;
              if (nloci[selected] > 0) and (nloci[neutral] > 0) then
                begin
                  step_neut^[0] := mean1(step_neut, nsamples);
                  Step2_neut^[0] := mean1(Step2_neut, nsamples);
                  slope_edge^[0] := mean1(slope_edge, nsamples);
                  NEWslope_edge^[0] := mean1(NEWslope_edge, nsamples);
                  central_d5_slope^[0] := mean1(central_d5_slope, nsamples);
                  slope_ratio^[0] := mean1(slope_ratio, nsamples);
                  NEWslope_ratio^[0] := mean1(NEWslope_ratio, nsamples);
                  barrier^[0]:=mean1(barrier,nsamples);
                  Barrier2^[0]:=mean1(Barrier2,nsamples);
                end;
            end;
{*********************************************}


{@ @ @ @ @ @ @ @   BEGIN MAIN PROGRAM  @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @}

begin

{ initialise; 14/8/96 - putting initialise procs directly into program seems to placate Code Warrior}
  InitGraf(@qd.thePort);
  InitFonts; InitWindows; InitMenus;TEInit; InitDialogs(NIL); InitCursor;
  PLHeapInit(preference_heap_minimum, 0,NIL,false,true); MoreMasters;
  check_memory('procedure initialise'); INstringSource:=keyboard;
  rewrite(INstringRec,'INStringRec');       {end of initialise stuff}

  writeln('Heap is ', Memavail : 10);
  new(pop); new(newpop); new(intermed_dmpop); new(habpop); new(dummyhabpop); new(habpop_size);
  new(dummyhabpop_size); new(cumwptr); new(habitat_availability); new(zbar_all); new(wbar_2h);
  new(varw_2h); new(wbar_2hMEAN); new(countbitstable);
  for I:=1 to 3 do
    begin
```

```pascal
        new(zbar[l]); new(zbar_2h[l]); new(delta_p[l]); new(Vgenic[l]); new(Vgenic_2h[l]);
        new(Vdiseq[l]); new(Vdiseq_2h[l]); new(Vtotal_2h[l]); new(Vtotal[l]); new(Fis[l]); new(Fis_2h[l]);
        new(Fis_max[l]); new(Fis_central_max[l]); new(Fis_central_max_2h[l]); new(Fis_deltap_max[l]);
        new(deltap_max[l]); new(deltap_centre[l]); new(Dmax[l]); new(Dmax_2h[l]); new(Dmean[l]);
        new(Dmax_reg[l]); new(predD_2h[l]); new(width_2h[l]); new(slope_2h[l]); new(width_2h[l]);
      end;
  new(step_neut); new(Step2_neut); new(slope_edge); new(NEWslope_edge); new(central_d5_slope);
  new(slope_ratio); new(NEWslope_ratio); new(barrier); new(Barrier2); new(wbarmin); new(wbarreg);
  new(fastrandom); new(kfarleft_all); new(kleft_all); new(kfarright_all); new(kright_all);

  countnot := 0; countdone := 0; rand_count := 0; nlociclasses:=0;

  writeln('Heap is ', Memavail : 10);

  write('Do you want to read parameters from a file ? ');
  readln(ans); batch := (ans = 'y');
  if batch then
    begin
      Get_OldFileName(batch_fname, 'Parameter file ? ');
      reset(bf, batch_fname); readln(bf, fname_root);
    end
  else
    Get_NewFileName(fname_root, 'File for output and pictures ?');
  while not (fname_root = 'END') do
    begin
      if batch then askstuff_batch else askstuff;

      nfix := 0; pict_fname := concat(fname_root, '.pic');
      dump_fname := concat(fname_root, '.dump');
      output_fname := concat(fname_root, '.@');
      rewrite(out_file, output_fname);
      set_listfile(out_file);

      for replicate:=1 to nreps do
      begin
       seed:=seed+4;
       initialise_vars;
       init_viewport;
       set_lastmask; set_countbitstable; set_fitness_table(s, hwidth, beta, Dopt);
       set_ntable; init_popns(pop, newpop,intermed_dmpop, prop); swappop(pop, newpop);
{habpop is initialised so that all individuals are in habitat 1 -
which can be left as such if only 1 habitat}
       initrep_hab(habpop, dummyhabpop, habitat_availability, habpop_size, dummyhabpop_size);
       init_rec_rates(rec,all_rec);
       for d:=0 to ndemes+1 do
         begin
           old_zbar_all^[d]:=0;  zbar_all^[d]:=getzbar(pop, dummyhabpop, dummyhabpop_size, d, 1, 0);
         end;

{Define habitat availability across cline}
       if two_habitats then
         begin
           get_habitat_availability(habitat_availability, habslope);
           if initfrequencies = 1 then        {ie. stepped cline - so want to fix end demes}
             init_end_demes(habpop, habpop_size, habitat_availability);
         end;
       sample:=0;
       set_viewport(0.05, 0.95, 0.125, 0.7);

       repeat
       t := t + 1;
{Get total gene freq in each deme, so that don't have to bother with reproduction if a deme is at fixation,
nor with migration if a deme and its neighbours are at fixation}
         swapzbar(zbar_all,old_zbar_all)  ;
         for d:=0 to ndemes+1 do
           if fixation_surround(old_zbar_all,d) then zbar_all^[d]:=old_zbar_all^[d]
           else zbar_all^[d]:=getzbar(pop, dummyhabpop, dummyhabpop_size, d, 1, 0);
```

```
            migrate(pop, newpop, zbar_all);   swappop(pop, newpop);

        if two_habitats then
          choose_habitat(pop, habpop, habpop_size, habitat_availability, habprefst);

{OPTION A: STATISTICS AFTER MIGRATION, BEFORE REPRODUCTION}
        if (t mod dt = 0) then
          writeln('Generation ', t : 3);
        if (t >= twarm) and (t mod dt = 0) then
          begin
            sample := sample + 1;
            update_stats;
            update_drawing;
            if not just_reps then
              begin
                if sample=1 then writeln(out_file, 'Sampling before migration and reproduction');
                write_output(false);            {write data in output graph format}
                plflush(out_file);
              end;
          end;


        for d := 0 to ndemes+1 do
            if fixation(zbar_all,d) then
              newpop^[d]:=pop^[d]
            else
              begin
{kidsdmpop are arrays of individuals in one deme. For reproduction and selection to be
interchangeable, each must be passed the same type of variable}

                dmpop:=kidsdmpopptrtype(@pop^[d,1]);
                newdmpop:=kidsdmpopptrtype(@newpop^[d,1]);

{OPTION 1:   REPRODUCTION THEN SELECTION  }
                if (t >= twarm) and (t mod dt = 0) and (sample =1) and (d=1) then
                    writeln(out_file, 'Reproduction then selection');
                reproduce(dmpop, intermed_dmpop, habpop, habitat_availability, habpop_size, rec,
                    mutn, d);

                sum:=0;
                for i:=1 to ninds do sum:=sum+habpop^[d,i];
                if (sum<>habpop_size^[d,1]) then scream;
                viability_selection(intermed_dmpop, newdmpop, habpop, habitat_availability,
                    habpop_size, d, s);

{OPTION 2:   SELECTION THEN REPRODUCTION}
{                if (t >= twarm) and (t mod dt = 0) and (sample=1) and (d=1) then
                    writeln(out_file, 'Selection then reproduction');
                viability_selection(s, dmpop, intermed_dmpop, habpop, habitat_availability,
                    habpop_size, d, freq_00, freq_01, freq_11);
                reproduce(intermed_dmpop, newdmpop, habpop, habitat_availability, habpop_size, rec,
                    mutn, d);
}

              end;
          swappop(pop, newpop);

{OPTION B: STATISTICS AFTER REPRODUCTION, BEFORE MIGRATION
        if (t mod dt = 0) then
          writeln('Generation ', t : 3);
        if (t >= twarm) and (t mod dt = 0) then
          begin
            sample := sample + 1;
            update_stats;
            update_drawing;
            if not just_reps then
              begin
                if sample=1 then writeln(out_file, 'Sampling after migration and reproduction');
```

```
                        write_output(false);           {write data in output graph format}
{                       plflush(out_file);
                    end;
             end}


        until t = tmax;
        nsamples := sample;
        set_sample_zero;
        sample := 0;{this is a trick to draw the overall means, which are stored in element 0}
        update_drawing;
        if just_reps then
          write_output_justreps
        else
          write_output(false);
        plflush(out_file);
           if dumpflag then
             begin
                rewrite(dump_file, dump_fname);
                write_QTfull(dipflag);   {write data on diplotypes}
                close(dump_file);
             end; {beware: writing out individual freqs gives too much output with many genes}

        sample:=1;
        update_stats;
        write_output(false);
      end;      {end of replicate loop}


       close(out_file);
       if batch then
         readln(bf, fname_root)
       else
         fname_root := 'END';
   end;

  useDrawWindow;
  quit_Display;

  dispose(pop); dispose(newpop); dispose(intermed_dmpop); dispose(habpop); dispose(dummyhabpop);
  dispose(habpop_size); dispose(dummyhabpop_size); dispose(cumwptr); dispose(habitat_availability);
  dispose(nptr); dispose(zbar_all); dispose(old_zbar_all);
  for l:=1 to 3 do
    begin
      dispose(zbar[l]); dispose(zbar_2h[l]); dispose(delta_p[l]); dispose(Vgenic[l]);
      dispose(Vgenic_2h[l]); dispose(Vdiseq[l]); dispose(Vdiseq_2h[l]); dispose(Vtotal[l]);
      dispose(Vtotal_2h[l]); dispose(Fis[l]); dispose(Fis_2h[l]); dispose(Fis_max[l]);
      dispose(Fis_central_max[l]); dispose(Fis_central_max_2h[l]); dispose(Fis_deltap_max[l]);
      dispose(deltap_max[l]); dispose(deltap_centre[l]); dispose(Dmax[l]); dispose(Dmax_2h[l]);
      dispose(Dmean[l]); dispose(Dmax_reg[l]); dispose(predD_2h[l]); dispose(width_2h[l]);
      dispose(slope_2h[l]); dispose(width_2h[l]);
    end;

  dispose(wbar_2h); dispose(wbar_2hMEAN); dispose(varw_2h); dispose(countbitstable);
  dispose(step_neut); dispose(Step2_neut); dispose(slope_edge); dispose(NEWslope_edge);
  dispose(central_d5_slope); dispose(slope_ratio); dispose(NEWslope_ratio); dispose(barrier);
  dispose(Barrier2); dispose(wbarmin); dispose(wbarreg); dispose(fastrandom); dispose(kfarleft_all);
  dispose(kleft_all); dispose(kfarright_all); dispose(kright_all);

end.
```

KEY PROCEDURES
unit multilocus_L4_procs;

interface

  uses

```
    quickdraw, housekeeping;

    procedure migrate (var pp, newpp: popptrtype; zb_all:dmptrtype);
    procedure get_habitat_availability (var hab_avail: dmhabsizeptrtype; hs: real);
    procedure choose_habitat (var pp: popptrtype; var habpp: habpopptrtype;
                              var habp_size, hab_av: dmhabsizeptrtype; hps: real);
    procedure reproduce ( var dp, new_dp:kidsdmpopptrtype; var habpp:habpopptrtype;
                          var hab_av, hp_size: dmhabsizeptrtype; rc, mu: real;  dm:demep1type);
    procedure viability_selection ( var dp, newdp:kidsdmpopptrtype; var habpp: habpopptrtype;
                          var hb_avail, hp_size: dmhabsizeptrtype; dm:demep1type; sel: real);
    procedure set_fitness_table (sel, hwidth, beta, Dopt: real);
    procedure init_popns (var pp, newpp: popptrtype; var dmpp:kidsdmpopptrtype; prop: real);
    procedure initialise_habitats (var habp, dummyh: habpopptrtype; var habt_av, habp_size,
                          dummyhabp_size: dmhabsizeptrtype);
    procedure write_array (var out_file: text; x: dtdmptrtype; tt, nf, nd: integer);
    procedure  initialise_vars;
    procedure init_end_demes (var hp: habpopptrtype; var hp_size, h_avail: dmhabsizeptrtype);

implementation

uses
    ToolUtils, utilities, functions,random, newhabstats;

    procedure migrate (var pp, newpp: popptrtype; zb_all:dmptrtype);
      var
        d: demetype;i, k, kb: indtype;

begin
    k := round(nmig / 2);
    kb := round(nmigbarr / 2);
    if ndemes = 2 then
      begin
        transfer(pp, newpp, k, kb, k, kb, 1, ninds);
        transfer(pp, newpp, kb, k, kb, k, 2, ninds);
        transfer(pp, newpp, 0, k, 0, k, 0, ninds); { commenting out migration into end demes}
        transfer(pp, newpp, k, 0, k, 0, 3, ninds);   { }
      end;
    if ndemes = 3 then
      begin
        transfer(pp, newpp, k, kb, k, kb, 1, ninds);
        transfer(pp, newpp, kb, k, kb, k, 2, ninds);
        transfer(pp, newpp, k, k, k, k, 3, ninds);
         transfer(pp, newpp, 0, k, 0, k, 0, ninds);{}
         transfer(pp, newpp, k, 0, k, 0, 4, ninds);{}
      end;
    if ndemes > 3 then
      begin
        if nmidleft - 1 > 1 then
          for d := 1 to nmidleft - 1 do
            if (fixation_surround(zb_all,d)) then newpp^[d]:=pp^[d]
            else  transfer(pp, newpp, k, k, k, k, d, ninds);

        if nmidright + 1 < ndemes then
          for d := nmidright + 1 to ndemes do
            if (fixation_surround(zb_all,d)) then  newpp^[d]:=pp^[d]
            else transfer(pp, newpp, k, k, k, k, d, ninds);

        transfer(pp, newpp, kb, k, kb, k, nmidright, ninds);
        transfer(pp, newpp, k, kb, k, kb, nmidleft, ninds);
      { transfer(pp, newpp, 0, k, 0, k, 0, ninds);{migration into end demes commented out}
      { transfer(pp, newpp, k, 0, k, 0, ndemes + 1, ninds);{}
      end;

end; { @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ procedure migrate
@ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ }
```

```pascal
procedure transfer (var pp, newpp: popptrtype; mfromleft, mfromright, mtoleft, mtoright:
                    indtype; demeno: demep1type; ninds: integer);
{necessary for mating scheme that adults are in a random order, otherwise migrants will }
{be next to migrants, and so mate with them - therefore need to shuffle them}
{NB. This does not matter if m=0 though, because selection will pick them at random}
    var
        offset,counter: integer; i,rn,newplace: indtype; taken:booleanvectype;

begin

    for i:=1 to ninds do taken[i]:=false;

    if mfromleft + mfromright <> mtoleft + mtoright then
        writeln('ERROR in transfer: asymmetric migration');

    if mfromleft > 0 then
        for i := 1 to mfromleft do
            begin
                repeat rn := round(rand(ninds)+0.5) until taken[rn] = false;
                newpp^[demeno, rn]:= pp^[demeno - 1, i];
                taken[rn]:=true;
            end;

    if mfromright > 0 then
        for i := ninds downto ninds - mfromright + 1 do
            begin
                repeat rn := round(rand(ninds)+0.5) until taken[rn] = false;
                newpp^[demeno, rn]:= pp^[demeno + 1, i];
                taken[rn]:=true;
            end;

    offset := mtoright - mfromleft;
    newplace:=1;
    for i := mfromleft + 1 to ninds - mfromright do
        begin
            while (taken[newplace]=true) do newplace:=newplace+1;
            newpp^[demeno,newplace]:=pp^[demeno,(i+offset)];
            taken[newplace]:=true;
        end;

end;

{ @ @ @ @ @ @ @ @ @ @ @ @ @ @ procedure transfer @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ }


procedure set_fitness_table (sel, hwidth, beta, Dopt: real);
    var
        i, imax: integer; s_epi, c_epi: real;

    function fitness (n: integer; s, hw, bta, zopt: real; nwrong:integer): real;
{if gausian epistasis, power epistasis or stabilising, then n is countones;}
{if hetdis, n is counthets; if geographic, n is count_number_on_wrong_side; if habitat,
n=number_wrong_hab}
{These definitions ensure that with heterozygote disadvantage or epistasis, the least fit}
{individual has fitness exp(-n s). s has a different interpretation with stabilising selection}
    begin
        if epistasis then fitness := c_epi * exp(-s_epi * exp(-sqr(n / (2 * nsel) - 0.5) / (2 * sqr(hw))));
        if epistasisp then fitness := exp(-s * nsel * power(4 * (n / (2 * nsel)) * (1 - (n / (2 * nsel))), bta));
        if stabilising then fitness := exp(-s * sqr(n - zopt) / 2);
        if hetdis then fitness := exp(-2 * s * n); {used to be: fitness := exp(-s * n);
            ...changing this 1.3.97 so that fitness in centre is equivalent
            under both hetdis and geographic selection = exp(-Ls), where L is no. of loci}
        if geographic or habitat_dependent then fitness := exp(-s * nwrong);
        end;
{ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ function fitness @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @}

begin
```

```pascal
        if epistasis then
          begin
            s_epi := nsel * sel / (1 - exp(-1 / (8 * sqr(hwidth))));
            c_epi := exp(s_epi * exp(-1 / (8 * sqr(hwidth))));
            imax := 2 * nsel
          end;
        if hetdis then imax := nsel;
        if epistasisp or stabilising or geographic or habitat_dependent then imax := 2 * nsel;
           {left of centre, or hab0 under hab_dep selt'n}
        for i := 0 to imax do fitness_table[i, 0] := fitness(i, sel, hwidth, beta, nsel-Dopt/2,i);
           {right of centre, or hab1 }
        for i := 0 to imax do fitness_table[i, 1] := fitness(i, sel, hwidth, beta, nsel+Dopt/2, (imax - i));
    end;
{@ @ @ @ @ @ @ @ @ @ @ @ @  procedure set_fitness_table @ @ @ @ @ @ @ @ @ @ @ @ @ @ @}



    procedure set_cumwptr (seln: real; var dmpp: kidsdmpopptrtype;  var hp: habpopptrtype;
                nfams, nkds: indtype;  dm: demep1type);
{returns cumulative fitness table for all offspring}
{Individuals are in one array, with their habitat in a corresponding array.
CumW is therefore defined by two arrays of the same length as the individuals' array:
 for each habitat, individuals in that habitat have their fitness defined by set_fitness_table }
      var
        k: kidtype; family_id: indtype; habitat: bittype; w: real; count: integer;
        ft_index: bittype; hom00,hom11,het, first_allele_1, second_allele_1:boolean;

    begin
      if dm < nmidright then ft_index := 0 else ft_index := 1;
      for habitat := 0 to 1 do
        if two_habitats or (habitat=1) then
        begin
          count := 0;
          cumwptr^[habitat, count] := 0;     {initialising cumulative fitness probability table}
          for k := 1 to (nfams*nkids) do
            begin
              family_id := (k - 1) div nkids + 1;
                if (hp^[dm, family_id])= habitat then
                  begin
                    if hetdis then w := fitness_table[counthets(dmpp^[k]), ft_index];
                    if epistasis then w := fitness_table[countones(dmpp^[k]), ft_index];
                    if epistasisp then w := fitness_table[countones(dmpp^[k]), ft_index];
                    if stabilising then w := fitness_table[countones(dmpp^[k]), ft_index];
                    if geographic then w := fitness_table[countones(dmpp^[k]), ft_index];
                    if habitat_dependent then w := fitness_table[countones(dmpp^[k]), habitat];
                  end
                else
                  w := 0; {fitness is defined separately for either habitat type}
                count := count + 1;
                cumwptr^[habitat, count] := cumwptr^[habitat, count - 1] + w
            end; {loop through all offspring of all individuals}
        end; {loop through both habitat types}
    end;
{&&&&&&&&&&&&&&&&&&&& end procedure set_cumwtpr &&&&&&&&&&&&&&&&&&&&&}


    procedure get_habitat_availability (var hab_avail: dmhabsizeptrtype; hs: real);
{If habitat availability changes across zone, amount of hab 1 is given by u,
where u/(1-u)=(x/(1-x))^habslope. }
{If habslope=0, hab1availability=0.5, by definition of the power function}
      var
        x, proph1a: real;
        dm: demep1type;

      function tanh (x: real): real;
      begin
        tanh := (exp(x) - exp(-x)) / (exp(x) + exp(-x))
      end;
```

```pascal
begin
  for dm := 0 to ndemes + 1 do
    if (hs<999) then
      begin
        x := dm / (ndemes + 1);
        proph1a := (1 + tanh(2 * (ndemes+1) * hs * (x - 0.5))) / 2;
        hab_avail^[dm, 1] := round(proph1a * ninds);
        hab_avail^[dm, 0] := ninds - hab_avail^[dm, 1];
      end
    else
      if (dm<nmidright) then
        begin
          hab_avail^[dm, 0] := ninds;
          hab_avail^[dm, 1] := 0;
        end
      else
        begin
          hab_avail^[dm, 0] := 0;
          hab_avail^[dm, 1] := ninds;
        end;
end;


{ @ @ @ @ @ @ @ @ @ @ @ @ procedure get_habavail @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @}


procedure choose_habitat (var pp: popptrtype; var habpp: habpopptrtype; var habp_size, hab_av:
dmhabsizeptrtype; hps: real);
{for one deme, assigns 1 to individual's corresponding position in hab array if random < habpref_function
for that individual, where each individual's habpref_function is defined by: the global strength of the
habitat pref (hps); its genotype at the hab pref loci; and habitat availability at that point in the cline.... see
Mathematica notebook for picture .... no.chose1 is no. of individuals who chose hab 1}

  var
    i: indtype; whichhab: bittype; dm: demep1type;

  function habpref_function (hps: real; h_av: indtype; hab_ones: integer): real;
  begin
{If hg=counthab1s(i)=nhab_alleles/2=nhabgenes, individual i has no preference; if counthabgen(i)=0,
strongest pref for hab 0; if =1, for hab 1 (h/(ninds-h)) is the logistic transform for the habitat }
{availability (h_av) across the cline, 0<=ha<=1 hps tunes the overall strength of the habitat preference -
so if set to 0, habpref is just  hab availability across the zone - similarly if there are no habitat pref genes}
    if nhabgenes>0 then
      habpref_function := h_av / ((ninds - h_av) *Exp(-hps*(hab_ones-nhabgenes)/nhabgenes) + h_av)
    else habpref_function := h_av/ninds;
  end;


begin
  for dm := 1 to ndemes do
    begin
      habp_size^[dm, 1] := 0;
      for i := 1 to ninds do
        begin
          if rand_highres(1)>habpref_function(hps, hab_av^[dm, 1], counthab1s(pp^[dm, i])) then
            whichhab := 0
          else whichhab := 1;
          habpp^[dm, i] := whichhab;
          habp_size^[dm, 1] := habp_size^[dm, 1] + whichhab;
        end;     {loop through individuals}
      habp_size^[dm, 0] := ninds - habp_size^[dm, 1];
    end;
end;
{ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ procedure choose_habitat @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ }


procedure reproduce ( var dp, new_dp:kidsdmpopptrtype; var habpp:habpopptrtype;
              var hab_av, hp_size: dmhabsizeptrtype; rc, mu: real; dm:demep1type);
```

```
{dp & new_dp are the adults and offspring in either habitat, habpp describes their respective habitats}
{ nhab1 is nchose, nnewhp is havail}
      var
         i, ii, rn, kid: indtype;  j, habitat: bittype; zygote: diplotype; rec_masktable: rec_masktabletype;

{Individual are paired up at random in the same habitat. Recombination and mutation happens within each
individual, and then one haplotype from each goes to form the new diploid individual.}

      function make_gamete (parent: indtype): haplotype;
         var
         r_mask, m_mask: haplotype; blk: integer; dummy: haplotype;

      begin {procedure make_gamete}
         set_mask(r_mask, rec_masktable, rc);      {sets up recombination mask using table already created}
         for blk := 1 to nblocks do
            begin                                          {get one new gamete from recombination}
               dummy[blk] := BitOr(BitAnd(r_mask[blk], dp^[parent, 0, blk]), BitAnd(BitNot(r_mask[blk]),
dp^[parent,1, blk]));
            end;   { end recombination; block loop}
         if mu > 0 then                                     {mutate if necessary}
            begin
               m_mask := mutation_mask(mu);
               for blk := 1 to nblocks do dummy[blk] := BitXor(dummy[blk], m_mask[blk])
            end; {end if mu>0}
         make_gamete := dummy;
      end;    { ------------------------------ function make_gamete ------------------------------ }


      function sonnyjim_random (i: indtype): diplotype;    {randomly pairs up parents}
         var
            rn, mum, dad: indtype;  x:integer;

      begin     {function sonnyjim}
         repeat  rn := round(rand(ninds)+0.5)   until (habpp^[dm,rn] = habpp^[dm,i]);
         mum:=rn;
         repeat  rn := round(rand(ninds)+0.5)   until (habpp^[dm,rn] = habpp^[dm,i]);
         dad:=rn;
         sonnyjim_random[0] := make_gamete(mum);              {parents undergo meiosis}
         sonnyjim_random[1] := make_gamete(dad);
      end;
{*************************** function  sonnyjim_random  ********************************}


{Reproduction within each habitat.... }
{Reproduction can only occur in a particular habitat if it contains 2+ individuals - so need also to}
{deal with case in which all reproduction is in 1 habitat. If reproduction in both habitats, }
{assign offspring from hab0 to random position in newpop and then fill in the gaps with offspring}
{ from hab1 - thus mixing the offspring from the 2 habitats before migration.}
{So, 3 cases:
{I. 2+ individuals in each habitat.}
{II. 0 or 1 individual in one of the habitats, so reproductn entirely in other.}
{III. Only considering 1 habitat}
      begin                             {procedure reproduce}
         set_recmasktable(rec_masktable);
         if two_habitats then
            if (hp_size^[dm, 0] > 1) and (hp_size^[dm, 1] > 1) then        {2 or more inds in both habitats}
{CASE I. 2+ individuals in each habt. Reproductn in both habs; habsize same after selection}
               for i := 1 to ninds do for kid:=1 to nkids do new_dp^[(i-1)*nkids + kid]:=sonnyjim_random(i);
            else
{CASE II. Two habitat types, but only enough individuals for reproductn in one}
               begin
                  if hp_size^[dm,0] <= 1 then      {ie. all ninds offspring must be in habitat 1}
                     habitat:=1
                  else habitat:=0;
                  for i := 1 to hp_size^[dm,habitat] do            {Doesn't matter what happens in kidsdmpp}
                     for kid:=1 to nkids do                   {corresponding to the empty habitat (poss 1 ind)}
                        if habpp^[dm,i]=habitat then            {as these will be ignored during seln}
                           new_dp^[(i-1)*nkids + kid]:=sonnyjim_random(i);
```

```
                 end
        else
{CASE III:  only 1 habitat..... }
        for i := 1 to ninds do for kid:=1 to nkids do new_dp^[(i-1)*nkids + kid]:=sonnyjim_random(i);
    end;


    { @ @ @ @ @ @ @ @ @ @ @ @ @ procedure reproduce @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @}


    procedure viability_selection (  var dp, newdp:kidsdmpopptrtype;  var habpp: habpopptrtype;
                    var hb_avail, hp_size: dmhabsizeptrtype; dm:demep1type; sel: real);
{Combine new popn of newhp[i] offspring in either habitat into one jumbled-up deme popn,}
{ unless one habitat has not produced any offspring ie. nhp[i]<2}
{habpp_size contains the number of offspring that will be picked from that habitat}
        var
            i, j, rn,offspring: indtype; habitat:bittype; taken:booleanvectype;

    function chosen_kid (habt: bittype): diplotype;
    {Chooses 1 offspring from popn by finding position of random var in cumulative fitness table }
        var
            rnd: real; kid_id: kidtype; family_id: indtype;
        begin
            repeat          {cumwptr^[pairid-1]<rand<=cumwptr^[pairid]..... }
                rnd := rand_highres(cumwptr^[habt, (ninds*nkids)]);
                kid_id := search(cumwptr, rnd, (ninds*nkids), habt);
                family_id := (kid_id - 1) div nkids + 1;
            until habpp^[dm, family_id]= habt;          {just checking}
            chosen_kid:=dp^[kid_id];
        end;  { ------------------------------- procedure chosen_kid ------------------------------- }


begin
    set_cumwptr(sel, dp, habpp, ninds, nkids, dm);    {Set up cum. fitnesses table for this deme}
    if (hp_size^[dm,0] > 1) and (hp_size^[dm,1] > 1) then
{CASE I}  begin
        for i:=1 to ninds do         {pick_offspring is random, but only within habitat..}
            taken[i]:=false;         {.... therefore need to jumble up habitats}
        for habitat:=0 to 1 do
            for i:=1 to hb_avail^[dm,habitat] do
                begin
                    repeat rn := round(rand(ninds - 1)) + 1 until taken[rn] = false;
                    newdp^[rn]:=chosen_kid(habitat);
                    taken[rn] := true;
                end;
        hp_size^[dm]:=hb_avail^[dm];
        end          {if hp_size[0],[1]>1}
    else
{Case II & III}
        begin
            if (hp_size^[dm,0] <=1) then habitat:=1 else habitat:=0;       {all ninds offspring must be in habitat 1}
            for i := 1 to ninds do newdp^[i]:=chosen_kid(habitat);
        end;
    end;


    { @ @ @ @ @ @ @ @ @ @ @ @ @ procedure viability_selection @ @ @ @ @ @ @ @ @ @ @ @ }


    procedure init_popns (  var pp, newpp: popptrtype;  var dmpp:kidsdmpopptrtype; prop: real);
        var
            rvalue, exfreq, beta, x, tanh, middeme: real; i: indtype; j: bittype; ib: blocktype; m_mask: haplotype;
            g: genetype; d, demeno: integer; plo, phi, exp: real; nhi, nlo, nmid: integer;

begin
    t := 0;      {a bit roundabout, but safest to set haplotypes to zero first}
    for d := 0 to ndemes + 1 do
        for i := 1 to ninds do
            for j := 0 to 1 do
```

```
                begin
                    pp^[d, i, j] := AllZero; newpp^[d, i, j] := AllOne;
                end;            {loop through haplotype, individuals,demes}
    for i:=1 to (ninds*nkids) do
        begin
            dmpp^[i,0]:=allZero; dmpp^[i,1]:=allZero;
        end;


    if stabilising then
        begin
            plo := 0.5 - Dopt / (4 * ngenes);       {lower eqbrm under stabilising selection}
            plo := min(max(plo, 0), 1);
            phi := 0.5 + Dopt / (4 * ngenes);       {upper eqbrm under stabilising selection}
            phi := min(max(phi, 0), 1);
            if initmethod = 1 then               {set all loci to phi or plo}
                begin   {mutation_mask returns a mask with 1 at probability plo or phi}
                    for d := 0 to nmidleft do
                        for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, mutation_mask(plo), d, i, j); {set to 1 or 0 with prob. plo}
                    for d := nmidright to ndemes + 1 do
                        for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, mutation_mask(phi), d, i, j)   {set to 1 or 0 with prob. plo}
                end;                                            {******** if initmethod=1 ******** }
            if initmethod = 2 then                              {set  ngenes*p loci to 1, rest to 0}
                begin
                    nmid := round(ngenes * (plo + phi) / 2);
                    m_mask := allZero;
                    if nmid > 0 then
                        for g := 0 to nmid - 1 do putgene(g, 1, m_mask);
                    for d := 0 to nmidleft do
                        for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, m_mask, d, i, j); {genes set to 1 or 0 with probability p}
                    for d := nmidright to ndemes + 1 do
                        for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, m_mask, d, i, j) {genes set to 1 or 0 with probability p}
                end;
            if initmethod = 3 then              {set  ngenes*plo loci to 1, rest to 0}
                begin
                    nlo := round(ngenes * plo);
                    m_mask := allZero;
                    if nlo > 0 then
                        for g := 0 to nlo - 1 do putgene(g, 1, m_mask);
                    for d := 0 to nmidleft do for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, m_mask, d, i, j); {genes set to 1 or 0 with probability plo}
                    nhi := round(ngenes * phi);
                    m_mask := allZero;
                    if nhi > 0 then
                        for g := 0 to nhi - 1 do putgene(g, 1, m_mask);
                    for d := nmidright to ndemes + 1 do
                        for i := 1 to ninds do
                            for j := 0 to 1 do puthap(newpp, m_mask, d, i, j) {genes set to 1 or 0 with probability phi}
                end;
            if initmethod = 4 then
                {set  ngenes*plo loci to 1, rest to 0; choose loci independently on either side}
                begin
                    m_mask := mutation_mask(plo);
                    for d := 0 to nmidleft do
                        for i := 1 to ninds do
                            for j := 0 to 1 do
                                puthap(newpp, m_mask, d, i, j); {genes set to 1 or 0 with probability plo}
                    m_mask := mutation_mask(phi);
                    for d := nmidright to ndemes + 1 do
                        for i := 1 to ninds do
                            for j := 0 to 1 do
                                puthap(newpp, m_mask, d, i, j) {genes set to 1 or 0 with probability phi}
                end;
        end     { ************* if stabilising *************** }
```

```
        else if initfrequencies = 1 then
          begin
            for d := 0 to nmidleft do
              for i := 1 to ninds do
                for j := 0 to 1 do
                  begin
                    if d = 0 then puthap(pp, allZero, d, i, j); {setting end demes - these will never change}
                    puthap(newpp, allZero, d, i, j); {all genes set to zero}
                  end;
            writeln;
            for d := nmidright to ndemes + 1 do
              for i := 1 to ninds do
                for j := 0 to 1 do
                  begin
                    if d = ndemes + 1 then puthap(pp, allOne, d, i, j);
                    puthap(newpp, allOne, d, i, j); {all genes set to one}
                  end;
          end   {********** initfrequencies=1 **********}

        else if initfrequencies = 2 then {proportion prop of all individuals in each deme set to 1}
          begin
            for d := 0 to ndemes + 1 do
              begin
                for i := 1 to round(ninds * (1 - prop)) do
                  for j := 0 to 1 do
                    puthap(newpp, allZero, d, i, j); {all genes set to zero}
                for i := round(ninds * (1 - prop)) + 1 to ninds do
                  for j := 0 to 1 do
                    puthap(newpp, allOne, d, i, j) {all genes set to one}
              end {loop through demes}
          end {***************** initfrequencies=2 *************}
        else if initfrequencies = 3 then {set all loci to prop}
          begin
{mutation_mask returns a mask with 1 at probability prop}
            for d := 0 to ndemes + 1 do
              for i := 1 to ninds do
                for j := 0 to 1 do puthap(newpp, mutation_mask(prop), d, i, j); {genes set to 1 or 0 with prob.
prop}
          end   {***************** initfrequencies=3 ************* }
        else if initfrequencies = 5 then
          begin
            for d := 0 to ndemes + 1 do
              begin
                for i := 2 to ninds do
                  for j := 0 to 1 do puthap(newpp, allZero, d, i, j); {all genes set to zero}
                puthap(newpp, allOne, d, 1, j) {when i = 1, all genes set to one}
              end;
          end;
      end;

  {@ @ @ @ @ @ @ @ @ @ @ procedure initrep @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @ }


  procedure initialise_vars;
    var
      d: demep1type; i, j: indtype; t: integer; h: bittype; l:locitype; g:genetype;
  begin
    nmidleft := (ndemes div 2); nmidright := (nmidleft + 1); randomise(seed);
    sample := 0; nloci[0]:=ngenes; nloci[1]:=nneut; nloci[2]:=nsel; nloci[3]:=nhabgenes;
    for d := 0 to ndemes + 1 do
      begin
        for l:=1 to 3 do
          begin
            zbar_2h[l]^[d]:=0; Vgenic_2h[l]^[d]:=0; Vtotal_2h[l]^[d]:=0;
            Vdiseq_2h[l]^[d]:=0; Fis_2h[l]^[d]:=0; elta_p[l]^[d]:=0;
          end;
```

```
          wbar_2h^[d] := 0.0; wbar_2hMEAN^[d] := 0.0; varw_2h^[d] := 0.0;
          for h := 0 to 1 do
            begin
             sample_count[d,h]:=0;
             for l:=1 to 3 do
               begin
                  zbar[l]^[d, h] := 0.0; Vgenic[l]^[d, h] := 0.0; Vtotal[l]^[d, h] := 0.0;
                  Vdiseq[l]^[d, h] := 0.0; Fis[l]^[d, h] := 0.0;
               end;
            end;
        end;
     for t := 0 to tmax div dt do
        begin
          for l:=1 to 3 do
            begin
               width_2h[l]^[t]:=0; NEWwidth_2h_p08[l]^[t]:=0; slope_2h_p08[l]^[t]:=0;
               predD_2h_p08[l]^[t]:=0; deltap_max[l]^[t]:=0; deltap_centre[l]^[t]:=0;
               Fis_central_max_2h[l]^[t]:=0; Dmax_2h[l]^[t]:=0;
            end;
          for h := 0 to 1 do
            for l:=1 to 3 do
              begin
                 Fis_max[l]^[t,h]:=0; Fis_central_max[l]^[t,h]:=0; Fis_deltap_max[l]^[t,h]:=0;
                 Dmax_reg[l]^[t, h] := 0; Dmax[l]^[t,h]:=0; Dmean[l]^[t,h]:=0;
              end;
          step_neut^[t] := 0; Step2_neut^[t] := 0; slope_edge^[t] := 0; NEWslope_edge^[t] := 0;
          central_d5_slope^[t] := 0; slope_ratio^[t] := 0; NEWslope_ratio^[t] := 0; barrier^[t]:=0;
          Barrier2^[t]:=0; wbarmin^[t] := 0; wbarreg^[t] := 0; kfarleft_all^[t]:=0; kleft_all^[t]:=0;
          kfarright_all^[t]:=0; kright_all^[t]:=0;
        end;
     nloci[0]:=ngenes; nloci[1]:=nneut; nloci[2]:=nsel; nloci[3]:=nhabgenes;
  end;
{*********************** procedure initialise_vars ************************}

  procedure initialise_habitats (var habp, dummyh: habpopptrtype; var habt_av, habp_size,
          dummyhabp_size: dmhabsizeptrtype);
     var     d: demep1type; i: indtype; j, h: bittype;
  begin
     for d := 0 to ndemes + 1 do  begin
        for i := 1 to ninds do    begin
            habp^[d, i] := 1;
            dummyh^[d, i] := 1;
        end;              {loop through haplotype, individuals}
        for h := 0 to 1 do    begin
            habt_av^[d, h] := h * ninds;
            habp_size^[d, h] := h * ninds;
            dummyhabp_size^[d, h] := h * ninds;  {in dummyhab, all inds are in hab 1}
        end;
     end;
  end;

  procedure init_end_demes (var hp: habpopptrtype; var hp_size, h_avail: dmhabsizeptrtype);
{Sets the numbers in habitats 0 /1 in the end demes to the habitat availability - }
{this can then stay constant, but is necessary for statistics}
     var     i: indtype; j: bittype; enddeme: demep1type;
  begin
     for j := 0 to 1 do
        begin
        if j = 0 then enddeme := 0 else enddeme := ndemes + 1;
        hp_size^[enddeme] := h_avail^[enddeme];
        for i := 1 to h_avail^[enddeme, 0] do   {not be executed if no individuals in habitat}
            puthab(hp, enddeme, 0, i);
        for i := h_avail^[enddeme, 0] + 1 to ninds do puthab(hp, enddeme, 1, i);
        end;
  end;

end.
```

```
unit housekeeping;

interface
  uses
    ordering;

  const
    logitmax = 5;{bounds on the logit function} Pi=3.141592654; all=0; {locus type indicators}
    neutral=1; selected=2; habpref=3;

    maxdemes = 210;  maxdemesp1 =211;
    maxinds = 40; maxmig = 20; maxkids = 2;
    maxkidstotal = 40;     {maxkids * maxinds}
    checkseln = false; checkmigrn = false;
    maxsamples = 30; maxt = 10000;
    maxblocks =7; maxgenes = 112;{maxblocks*16} ; maxgenes2 =224; {2*maxgenes}

  type
    bittype = 0..1;
    locitype=0..3;{0=all,1=neut,2=sel,3=hab}
    genetype = 0..maxgenes;
    demetype = 1..maxdemes;
    demep1type = 0..maxdemesp1;
    indtype = 0..maxinds;
    kidtype = 0..maxkidstotal;
    blocktype = 1..maxblocks;
    locustype = array[0..1] of bittype;
    nlocitype=array[0..3] of genetype;
    recratestype=array[0..3] of real;
    bitrealstype=array[0..1] of real;
    haplotype = array[1..maxblocks] of integer;
    diplotype = array[0..1] of haplotype;
    samplefreqtype = array[0..maxdemesp1,0..1] of integer;
    poptype = array[0..maxdemesp1, 1..maxinds] of diplotype;   {genotypes of individuals in demes}
    habpoptype = array[0..maxdemesp1, 1..maxinds] of bittype;    {habitats of individuals in demes}
    popptrtype = ^poptype; habpopptrtype = ^habpoptype;
    kidsdmpoptype = array[1..maxkidstotal] of diplotype; kidsdmpopptrtype = ^kidsdmpoptype;
    dttype = array[0..maxsamples] of real; dtptrtype = ^dttype; locidtptrtype=array[1..3] of dtptrtype;
    dmhabsizetype = array[0..maxdemesp1, 0..1] of indtype; dmhabsizeptrtype = ^dmhabsizetype;
    dmhabtype = array[0..maxdemesp1, 0..1] of real; dmhabptrtype = ^dmhabtype;
    locidmhabptrtype=array[1..3] of dmhabptrtype;
    dmchaintype = array[0..maxdemesp1] of real; dmptrtype = ^dmchaintype;
    locidmptrtype=array[1..3] of dmptrtype;
    dtdemetype = array[0..maxsamples, 0..maxdemesp1] of real; dtdmptrtype = ^dtdemetype;
    locidtdmptrtype=array[1..3] of dtdmptrtype; dthabtype = array[0..maxsamples, 0..1] of real;
    dthabptrtype = ^dthabtype; locidthabptrtype=array[1..3] of dthabptrtype;
    realp2vectype = array[0..1, 0..maxkidstotal] of real;realp2vecptrtype = ^realp2vectype;
    intpvectype = array[0..maxkidstotal] of integer;intpvecptrtype = ^intpvectype;
    realttype = array[1..maxt] of real;
    countbitstabletype = array[0..32767] of 0..15; countbitstableptrtype = ^countbitstabletype;
    rec_masktabletype = array[1..15] of integer; booleanvectype = array[1..maxkidstotal] of boolean;
    booleanvecptrtype = ^booleanvectype; {used when pairing parents & combining popns from 2 habs}
    altbooleanvectype = array[1..maxkidstotal] of bittype; {because new debugger sees booleans as
being 8 bits long}

  var
    t, dt, sample, nsamples, h1,seed,nlociclasses: integer;
    all_rec_rates, my_rec_rates:recratestype;
    ngenes, nsel, nneut, nhabgenes: genetype; nloci:nlocitype;
    sample_count:samplefreqtype; nblocks, nhabblocks: blocktype;
    ninds, nmig, nmigbarr, nkids: indtype; ndemes,nmidleft, nmidright: demep1type;
    zbar_all,old_zbar_all:dmptrtype;
    habitat_availability, habpop_size, newhabpop_size, dummyhabpop_size: dmhabsizeptrtype;
    fitness_table: array[0..maxgenes2, 0..1] of real;  {array holding fitnesses of inds with i hets, or i ones}
    countbitstable: countbitstableptrtype;   {array holding the number of bits in an integer from 0 to 32767}
```

```pascal
    lastmask: integer;
    cumwptr: realp2vecptrtype;
    dummyhabpop: habpopptrtype;
    zbar, Vdiseq, Vgenic, Vtotal, Fis: locidmhabptrtype;
    zbar_2h,delta_p,Vdiseq_2h, Vgenic_2h, Vtotal_2h, Fis_2h:locidmptrtype;
    wbar_2h, varw_2h: dmptrtype;
    wbar_2hMEAN, varw_2hMEAN: dmptrtype;
    Dmax,Dmax_reg,Dmean, Fis_max, Fis_central_max, Fis_deltap_max:locidthabptrtype;
    width_2h,slope_2h_p08,NEWwidth_2h_p08,predD_2h_p08,deltap_max,deltap_centre:locidtptrtype;
    Dmax_2h,Fis_central_max_2h:locidtptrtype;
    kfarleft_all,kleft_all,kfarright_all,kright_all:dtptrtype;
    step_neut_, slope_left_neut_, slope_right_neut_, slope_mid_neut_, wbarmin_, wbarreg_: real;
    slope_edge, NEWslope_edge, central_d5_slope, wbarreg, wbarmin: dtptrtype;
    step_neut, NEWstep_neut:dtptrtype;
    NEWstep_neut_, NEWslope_left_neut_, NEWslope_right_neut_: real;
    slope_ratio, NEWslope_ratio: dtptrtype;
    barrier,NEWbarrier:dtptrtype;
    nptr: intpvecptrtype;
    step, hetdis, epistasis, stabilising, epistasisp, geographic, hetdis_and_geo:boolean;
    habitat_dependent, two_habitats: boolean;
    initmethod, initfrequencies, twarm, tmax: integer;
    seln_mask,hab_mask: haplotype;{whether or not an allele is selected/determines habitat preference}
    allOne, allZero: haplotype; {haplotypes corresponding to all ones or all zeroes}
    dopt, zbar_d: real; fastrandom:permptrtype; countnot, countdone,rand_count: longint;

  function rand (range: real): real;
  function rand_highres (range: real): real;
  function gethap (var p: popptrtype; d: demep1type; i: indtype; j: bittype): haplotype;
  function getdip (var p: popptrtype; d: demep1type; i: indtype): diplotype;
  function getgene (g: genetype; h: haplotype): bittype;
  procedure puthap (var p: popptrtype; h: haplotype; d: demep1type; i: indtype; j: bittype);
  procedure putdip (var p: popptrtype; d: demep1type; dip: diplotype; i: indtype);
  procedure putgene (var g: genetype; x: bittype; var h: haplotype);
  procedure swappop (var p, np: popptrtype);
  procedure swaphab (var hp, nhp: habpopptrtype;var hps, nhps: dmhabptrtype);
  procedure swapzbar (var zb, old_zb: dmptrtype);
  function fixation (zb:dmptrtype; d: demep1type): boolean;
  function fixation_surround (zb:dmptrtype; d: demep1type): boolean;
  function counthets (d: diplotype): integer;
  function counthoms (d: diplotype): homfreqtype;
  function countones (d: diplotype): integer;
  function search (var t: realp2vecptrtype; x: real; nn: integer; h: bittype): integer;
  procedure set_countbitstable;
  procedure set_lastmask;
  function mutation_mask (mm: real): haplotype;
  function rec_mask (j: integer): integer;
  procedure set_recmasktable (var rm: rec_masktabletype);
  procedure set_mask (var mask: haplotype; rm: rec_masktabletype; r: real);
  procedure scream;

implementation

uses toolUtils,quickdraw;

  function rand (range: real): real;   {random number from 0 to range}
  begin
    rand := (0.50000762939 + random / 65536) * range; {works on the Mac}
  end;

  function rand_highres (range: real): real;  {random number from 0 -range; higher resolution}
    var
      tr: longint;
  begin
    tr := random + bsl(random, 16);  rand_highres := (0.5 + tr / 4294967296.0) * range;{works on the Mac}
  end;

  function gethap (var p: popptrtype; d: demep1type; i: indtype; j: bittype): haplotype;
{gets haplotype for deme d, individual i, genome j from the array p}
```

235

```
    begin
       gethap := p^[d, i, j]
    end;

    function getdip (var p: popptrtype; d: demep1type; i: indtype): diplotype;
{gets diplotype for deme d, individual i from the array p}
    begin
       getdip := p^[d, i]
    end;

    function getgene (g: genetype; h: haplotype): bittype;
       var
          k: integer;
    begin
       k := g div 16;
       if odd(bsr(h[k + 1], (g - k * 16))) then getgene := 1 else getgene := 0;
    end;

    procedure puthap (var p: popptrtype; h: haplotype; d: demep1type; i: indtype; j: bittype);
    begin
       p^[d, i, j] := h
    end;

    procedure putdip (var p: popptrtype; d: demep1type; dip: diplotype; i: indtype);
    begin
       p^[d, i] := dip
    end;

    procedure putgene (var g: genetype; x: bittype; var h: haplotype);
{This works. At first sight, one should be able to use bset and bclr. However, these require }
{var variables, and so don't work. }
       var    k: integer;
    begin    {need care here, because bsl returns a longint}
       k := 1 + g div 16;
       if x = 1 then    h[k] := BitOr(h[k], LoWrd(bsl(1, (g mod 16))))
       else    h[k] := BitAnd(h[k], BitNot(LoWrd(bsl(1, (g mod 16)))));
{ h is an array of integers, where each integer represents 0 to 15 genes}
    end;

    procedure swapzbar (var zb, old_zb: dmptrtype);
       var
          tptr: dmptrtype;
    begin
       tptr := old_zb; old_zb := zb; zb := tptr;
    end;

    procedure swappop (var p, np: popptrtype);
       var
          tptr: popptrtype;
    begin
       tptr := p; p := np; np := tptr;
    end;

    procedure swaphab (var hp, nhp: habpopptrtype;var hps, nhps: dmhabptrtype);
       var
          tptr1: habpopptrtype; tptr2: dmhabptrtype;
    begin
       tptr1 := hp; hp := nhp; nhp := tptr1;
       tptr2 := hps; hps := nhps; nhps := tptr2;
    end;

    function fixation (zb:dmptrtype; d: demep1type): boolean;
    begin
       if (zb^[d]=0) or (zb^[d]=2*ngenes) then fixation:=true else fixation:=false;
    end;

    function fixation_surround (zb:dmptrtype; d: demep1type): boolean;
    begin
```

```pascal
        if (((zb^[d-1]=0) and (zb^[d]=0) and (zb^[d+1]=0)) or ((zb^[d-1]=2*ngenes) and (zb^[d]=2*ngenes)
           and (zb^[d+1]=2*ngenes))) then fixation_surround:=true
        else fixation_surround:=false;
      end;


      function countupbits (i: integer): integer;{important not to have var here}
        var
          sum, g: integer;
      begin
        sum := 0;
        for g := 0 to 15 do
          begin
            if odd(i) then sum := sum + 1;
            i := bsr(i, 1);
          end;
        countupbits := sum
      end;


      procedure set_lastmask;
        var
          k, kmax: integer;

      begin
        lastmask := 1;
        kmax := ((ngenes - 1) - ((nblocks - 1) * 16));
        if kmax = 15 then lastmask := -1 {this avoids an obscure overflow problem}
        else for k := 0 to kmax - 1 do lastmask := 1 + bsl(lastmask, 1);
{lastmask is 00001111.., designed to mask superfluous bits in the last block}
      end;


      procedure set_countbitstable;
        var
          k: integer;
      begin
        writeln('Setting up countbitstable..');
        for k := 0 to 32767 do
          begin
            countbitstable^[k] := countupbits(k);
            if (k mod 1000) = 0 then writeln(k);
          end;
      end;


      function countbits (h: haplotype; seln_not_hab_flag:boolean): integer;
        var
          sum1, ic, imax: integer;
      begin
        sum1 := 0;
        for ic := 1 to nblocks do
          begin
{slightly simpler to do this, rather than use two loops}
            if ic = nblocks then h[ic] := BitAnd(h[ic], lastmask);{lastmask (set by set_lastmask)}
{                       ensures that the extra parts of h are zero}
            if seln_not_hab_flag then h[ic] := BitAnd(h[ic], seln_mask[ic])
            else h[ic] := BitAnd(h[ic], hab_mask[ic]);
            if h[ic] < 0 then sum1 := sum1 + countbitstable^[32768 + h[ic]] + 1
            else sum1 := sum1 + countbitstable^[h[ic]]
          end;
        countbits := sum1
      end;


      function counthets (d: diplotype): integer;
        var
          ic: integer; temp: haplotype;
      begin
        for ic := 1 to nblocks do temp[ic] := BitXor(d[0, ic], d[1, ic]);
        counthets := countbits(temp,true);
      end;
```

```pascal
function counthoms (d: diplotype): homfreqtype;
    var
        ic: integer; temp: haplotype;
    begin
    for ic := 1 to nblocks do temp[ic] := BitAnd(BitNot(d[0, ic]), BitNot(d[1, ic]));
    counthoms[0] := countbits(temp,true);          {returns the number of 00 homozygotes}
    for ic := 1 to nblocks do temp[ic] := BitAnd(d[0, ic], d[1, ic]);      {returns the number of 11 homozygs}
    counthoms[1] := countbits(temp,true);
    end;

function countones (d: diplotype): integer;
    begin
    countones := countbits(d[0],true) + countbits(d[1],true);
    end;
function counthabones (d: diplotype): integer;
    begin
    counthabones := countbits(d[0],false) + countbits(d[1],false)
    end;

function search (var t: realp2vecptrtype; x: real; nn: integer; h: bittype): integer;
{Looks for the position of x in a table of reals;t^[0]=0,t^[j]-t^[j-1]=w[j].}
{0<x<t^[imax]; returns j if t^[j-1]<x<=t^[j]}
    var
        i, imax, imin: integer;
    begin
    if nn <= 1 then search := 1
    else
        begin
        imax := nn; imin := 0;{set the initial interval; x must lie between t^[h,imax] and t^[h,imin]}
        repeat
            i := (imax + imin) div 2; if x > t^[h, i] then imin := i else imax := i
        until imax = imin + 1;
        if x = t^[h, imin] then search := imin else search := imax
        end
    end;  .

function mutation_mask (mm: real): haplotype;
{returns an integer which has 1's with probability mu}
    var
        ib: blocktype; h: haplotype;

    function one_mutation_mask (mu: real; ng: integer): integer;
{returns an integer which has 1's with probability mu}
        var
            temp: longint;
            g: integer;
        begin
        if rand_highres(1) < mu then temp := 1 else temp := 0;
        if ng > 1 then
            for g := 2 to ng do
                begin
                temp := bsl(temp, 1);
                if rand_highres(1) < mu then temp := temp + 1
                end;
        one_mutation_mask := LoWrd(temp)
        end; {********** function one_mutation_mask************}

    begin
    h[nblocks] := one_mutation_mask(mm, (ngenes - 1) mod 16 + 1);
    if nblocks > 1 then for ib := 1 to nblocks - 1 do h[ib] := one_mutation_mask(mm, 16);
    mutation_mask := h;
    end;    {********** function mutation_mask ************}

function rec_mask (j: integer): integer;
{rec_mask(j) returns a mask 000111.., with j 1s. j between 1 and 15....}
```

```
{... no longer a sub-function of set_recmasktable because choosehabitat calls it}
    var
      i2, nn: integer;
  begin
    nn := 1;
    if j > 1 then for i2 := 2 to j do nn := 1 + bsl(nn, 1);
    rec_mask := nn;
  end;

  procedure set_recmasktable (var rm: rec_masktabletype);
    var
      k: integer;
  begin
    for k := 1 to 15 do rm[k] := rec_mask(k);
  end; {*************** procedure set_recmasktable *****************}

  procedure set_mask (var mask: haplotype; rm: rec_masktabletype; r: real);
    var
      ib: blocktype;
      oldbit: boolean;

    function one_mask (ng: integer; rec: real): integer;
      var
        i1, nk: integer;
    begin
      nk := -1;
      for i1 := 1 to ng - 1 do
        if rand_highres(1)<rec then      {used to be: if John_Major then}
           begin
           nk := BitXor(nk, rm[i1]);
           countdone:=countdone+1;
          end
        else countnot:=countnot+1;
      one_mask := nk;
    end;

  begin {procedure set_mask}
{set up the basic mask}
    for ib := 1 to nblocks do
      begin
        if (r >= 0.5) or (ngenes = 1) then mask[ib] := random
        else {dealing with linkage}
          if ib = nblocks then mask[ib] := one_mask(((ngenes - 1) mod 16) + 1, r)
          else mask[ib] := one_mask(16, r);
      end;
{Deal with recombination events between the integers making up the haplotype}
    if random < 0 then mask[nblocks] := BitNot(mask[nblocks]);
    oldbit := odd(mask[nblocks]);
    if nblocks > 1 then
      for ib := nblocks - 1 downto 1 do
        if rand_highres(1) < r then {there is a recombination event}
          if oldbit then mask[ib] := BitNot(mask[ib])
        else
          if not oldbit then mask[ib] := BitNot(mask[ib])
        oldbit := odd(mask[ib]);
  end;
{******************** procedure set_mask*************}

procedure scream;
  var i:integer;

  begin
    writeln('scream');
    readln(i);
  end;


end.
```

```
unit newhabstats;
interface

  uses
    housekeeping,memory,types;

  procedure writetime;
  function getgenefreq (var pp: popptrtype; var hp: habpopptrtype;
     var habpop_size: dmhabsizeptrtype; d: demep1type; habitat: bittype; g: genetype): real;
  function getzbar (var pp: popptrtype;  var habpp: habpopptrtype;
     var habpop_size: dmhabsizeptrtype;  d: demep1type; habitat: bittype;  loci:locitype): real;
  function genic_var (var pp: popptrtype; var habpp: habpopptrtype;
     var habpop_size: dmhabsizeptrtype; d: demep1type; habitat: bittype; loci:locitype): real;
  function heterozyg_deficit (var pp: popptrtype; var habpp: habpopptrtype;
     var habpop_size: dmhabsizeptrtype; d: demep1type; habitat: bittype; loci:locitype): real;
  function total_var (var pp: popptrtype; var hp: habpopptrtype; var habpop_size: dmhabsizeptrtype;
     d: demep1type; habitat: bittype; loci:locitype): real;
  procedure find_wbar (var pp: popptrtype; var hp: habpopptrtype; var habpop_size:
     dmhabsizeptrtype; d: demep1type; habitat: bittype; twohab_flag: boolean; var wb, varw: real);
  function mean1 (var x: dtptrtype; nn: integer): real;
  function variance1 (var x: dtptrtype; nn: integer): real;
  function mean1h (var x: dthabptrtype; h: bittype; nn: integer): real;
  function variance1h (var x: dthabptrtype; h: bittype; nn: integer): real;
  function mean (var x: dtdmptrtype; d: demep1type; nn: integer): real;
  function variance (var x: dtdmptrtype; d: demep1type; nn: integer): real;
  function rbar (smk,hmk: haplotype; rec: real): real;  {harmonic mean recombination between }
  function rbar_all (rec: real): real;                              {neutral and selected loci}
  procedure regress (var xb, yb, b: real; x, y: dmchaintype; nn: integer; originflag: boolean);
  function get_Dmax_reg (var Vg, Vd: dmhabptrtype; ns: integer; h: bittype; ng: integer): real;
  function get_Dmax (var Vd: dmhabptrtype; h: bittype; ng: integer): real;
  function get_Dmax_2h (var Vd_2h: dmptrtype; ng: integer): real;
  function get_max (var A: dmhabptrtype; h: bittype): real;
  function get_max_2h (var A: dmptrtype): real;
  procedure get_deltap_max (var A: dmptrtype; var dp_max: real; var deme_dp_max: demep1type);
  function get_Dmean (var Vd: dmhabptrtype; ns: integer; h: bittype; ng: integer): real;
  function predD(var w2h:dtptrtype; ns: integer; r, mig: real): real;
  function slope_2hL(pp: popptrtype; loci:locitype; lmax: real): real;
  function NEWwidth_2hL(pp: popptrtype; loci:locitype; lmax: real): real;
  function get_width_2h (var zb: dmptrtype; ns, ng: integer): real;
  procedure minimum_fitness (var wbmin, wbreg: real; var zb: dmptrtype;  var wb: dmptrtype;
            ns,ng:integer; var central_deme:demep1type);
  procedure calculate_barrier (var step,NEWstep,  slope_left, slope_mid, slope_right,
          NEWslope_left, NEWslope_right:real; var zb: dmptrtype; var wb: dmptrtype; ns, ng: integer);
  procedure init_rec_rates(r:real; var all_rates:recratestype);
  function get_max_Fis_centre(var Fisloci: dmhabptrtype;  hab: bittype;
            centre_deme:demep1type):real;
  function get_max_Fis_centre_2h(var Fis_2hloci: dmptrtype;centre_deme:demep1type):real;


implementation

  uses
    functions,utilities;

  procedure writetime;
  begin
    writeln(time)
  end;


  function getgenefreq (var pp: popptrtype; var hp: habpopptrtype;
      var habpop_size: dmhabsizeptrtype; d: demep1type; habitat: bittype; g: genetype): real;
    var
      sum: real; i, count: indtype; j: bittype;
  begin
```

```pascal
          sum := 0; count := 0;
          if habpop_size^[d, habitat] > 0 then
            begin
              for i := 1 to ninds do
                if hp^[d, i] = habitat then
                  begin
                    for j := 0 to 1 do sum := sum + getgene(g, pp^[d, i, j]);
                    count := count + 1;
                  end;

              if count > 0 then getgenefreq := sum / (2 * count)
              else
                begin
                  writeln('ERROR: no individuals for getgenefreq. Input Y to stop');
                  readln(count);
                end;
            end
          else getgenefreq := 999              {which should have been avoided}
       end;


  function gethetfreq_gene (var pp: popptrtype; var hp: habpopptrtype; var habpop_size:
dmhabsizeptrtype; d: demep1type; habitat: bittype; g: genetype): real;
    var
      sum: real; i, count: indtype; j: bittype;
  begin
    sum := 0; count := 0;
    if habpop_size^[d, habitat] > 0 then
      begin
        for i := 1 to ninds do
          if hp^[d, i] = habitat then
            begin
              if (getgene(g, pp^[d,i,0]) + getgene(g, pp^[d,i,1])=1) then sum := sum + 1;
              count := count + 1;
            end;
        if count > 0 then gethetfreq_gene := sum / count
        else
          begin
            writeln('ERROR: no individuals for gethetfreq. Input Y to stop');
            readln(count);
          end;
      end
    else gethetfreq_gene := 999                {which should have been avoided}
  end;



  function getzbar (var pp: popptrtype;  var habpp: habpopptrtype;
      var habpop_size: dmhabsizeptrtype;  d: demep1type; habitat: bittype;  loci:locitype): real;
    var
      g: genetype; sum: real; wanted:boolean;
  begin
   if habpop_size^[d, habitat] > 0 then
    begin
     sum := 0;
     for g := 0 to ngenes - 1 do
       begin
         wanted:=(loci=all)  or  ((loci=selected) and (getgene(g,seln_mask)=1))
             or ((loci=habpref) and (getgene(g,hab_mask)=1))
             or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
         if wanted then sum := sum + getgenefreq(pp, habpp, habpop_size, d, habitat, g);
       end;
     getzbar := 2 * sum;
    end
   else getzbar:=999;
  end;
{****************** function geniczbar********************}

  function genic_var (var pp: popptrtype; var habpp: habpopptrtype;
    var habpop_size: dmhabsizeptrtype; d: demep1type; habitat: bittype; loci:locitype): real;
```

```
        var
          g: genetype; p, sum: real; wanted:boolean;
     begin
       sum := 0;
       for g := 0 to ngenes - 1 do
          begin
            wanted:=(loci=all)  or ((loci=selected) and (getgene(g,seln_mask)=1))
                or ((loci=habpref) and (getgene(g,hab_mask)=1))
                or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
            if wanted then
               begin
                 p := getgenefreq(pp, habpp, habpop_size, d, habitat, g);
                 sum := sum + 2 * p * (1 - p);
               end;
          end;
       genic_var := sum
     end;



    function heterozyg_deficit (var pp: popptrtype; var habpp: habpopptrtype; var habpop_size:
dmhabsizeptrtype; d: demep1type; habitat: bittype; loci:locitype): real;
       var
          g: genetype; p, heterozygosity, f, sum: real; wanted:boolean;
     begin
       sum := 0;
       for g := 0 to ngenes - 1 do
          begin
            wanted:=(loci=all)  or ((loci=selected) and (getgene(g,seln_mask)=1))
                or ((loci=habpref) and (getgene(g,hab_mask)=1))
                or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
            if wanted then
               begin
                 p := getgenefreq(pp, habpp, habpop_size, d, habitat, g);
                 heterozygosity := gethetfreq_gene(pp, habpp, habpop_size, d, habitat, g);
                 if ((p>0) and (p<1)) then f := 1 - heterozygosity / (2 * p * (1 - p))
                 else f :=0;
                 sum := sum + f;
               end;
          end;
       heterozyg_deficit := sum / nloci[loci];
     end;
{******************* function  heterozyg_deficit*********************}



    function total_var (var pp: popptrtype; var hp: habpopptrtype; var habpop_size: dmhabsizeptrtype;
        d: demep1type; habitat: bittype; loci:locitype): real;
       var
          k, count: indtype; g: genetype; j: bittype; p, sum, sum2, zind: real; wanted:boolean;
     begin
       sum := 0; sum2 := 0; count := 0;
       for k := 1 to ninds do
          if hp^[d,k] = habitat then
            begin
              zind := 0;
              for j := 0 to 1 do
                 for g := 0 to ngenes - 1 do
                    begin
                      wanted:=(loci=all) or ((loci=selected) and (getgene(g,seln_mask)=1))
                          or ((loci=habpref) and (getgene(g,hab_mask)=1))
                          or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
                      if wanted then zind := zind + getgene(g, pp^[d, k, j]);
                    end; { g loop through all genes}
              sum := sum + zind;
              sum2 := sum2 + sqr(zind);
              count := count + 1;
            end;              {conditional on being in correct habitat , k loop through all individuals }
```

```
          if count > 0 then total_var := (sum2 / count) - sqr(sum / count)
        end;
{*********************** function  total_var  *******************}


    procedure find_wbar (var pp: popptrtype; var hp: habpopptrtype; var habpop_size:
dmhabsizeptrtype; d: demep1type; habitat: bittype; twohab_flag: boolean; var wb, varw: real);
        var
          i, count: indtype; ft_index: bittype; w, smw, ssqw: real;
    begin
      if d < nmidright then ft_index := 0 else ft_index := 1;
      smw := 0; ssqw := 0; count := 0;
      if twohab_flag or (habpop_size^[d, habitat] > 0) then
        begin
          for i := 1 to ninds do
            if twohab_flag or (hp^[d,i] = habitat) then
              begin
                if hetdis then w := fitness_table[counthets(pp^[d,i]), ft_index];
                if epistasis then w := fitness_table[countones(pp^[d,i]), ft_index];
                if epistasisp then w := fitness_table[countones(pp^[d,i]), ft_index];
                if stabilising then w := fitness_table[countones(pp^[d,i]), ft_index];
                if geographic then w := fitness_table[countones(pp^[d,i]), ft_index];
                if habitat_dependent then w := fitness_table[countones(pp^[d,i]), hp^[d,i]];
                smw := smw + w;
                ssqw := ssqw + sqr(w);
                count := count + 1;
              end;
          if count > 0 then
            begin
              wb := smw / count; varw := (ssqw / count) - sqr(wb);
            end;
        end     {conditional on being 1+ individuals in habitat}
      else
        begin
          wb := 999; varw := 999;
        end;
    end;



{lots of fiddly means and variance functions for dealing with various types of data ararays -}
{all need to take account of potentially empty cells when nothing in habitat}

    function mean1 (var x: dtptrtype; nn: integer): real;
        var
          i, count: integer; sm: real;
    begin
      sm := 0; count := 0;
      for i := 1 to nn do
        if x^[i] <> 999 then
          begin
            sm := sm + x^[i]; count := count + 1;
          end;
      if count > 0 then mean1 := sm / count else mean1 := 0
    end;

    function mean1_dmchain (var x: dmchaintype; nn: integer): real;
        var
          i, count: integer; sm: real;
    begin
      sm := 0; count := 0;
      for i := 1 to nn do
        if x[i] <> 999 then
          begin
            sm := sm + x[i]; count := count + 1;
          end;
      if count > 0 then mean1_dmchain := sm / count else mean1_dmchain := 0
    end;

    function mean1_dmptr (var x: dmptrtype; nn: integer): real;
```

```
      var
         i, count: integer; sm: real;
   begin
      sm := 0; count := 0;
      for i := 1 to nn do
         if x^[i] <> 999 then
            begin
               sm := sm + x^[i]; count := count + 1;
            end;
      if count > 0 then mean1_dmptr := sm / count else mean1_dmptr := 0
   end;


   function variance1 (var x: dtptrtype; nn: integer): real;
      var
         i, count: integer; mn, sm: real;
   begin
      mn := mean1(x, nn); sm := 0; count := 0;
      for i := 1 to nn do
         if x^[i] <> 999 then
            begin
               sm := sm + sqr(x^[i] - mn); count := count + 1;
            end;
      if count > 1 then variance1 := sm / (count - 1) else variance1 := 0
   end;

   function mean1h (var x: dthabptrtype; h: bittype; nn: integer): real;
      var
         i, count: integer; sm: real;
   begin
      sm := 0; count := 0;
      for i := 1 to nn do
         if x^[i, h] <> 999 then
            begin
               sm := sm + x^[i, h]; count := count + 1;
            end;
      if count > 0 then mean1h := sm / count else mean1h := 0
   end;

   function variance1h (var x: dthabptrtype; h: bittype; nn: integer): real;
      var
         i, count: integer; mn, sm: real;
   begin
      mn := mean1h(x, h, nn);
      sm := 0; count := 0;
      for i := 1 to nn do
         if x^[i, h] <> 999 then
            begin
               sm := sm + sqr(x^[i, h] - mn); ount := count + 1;
            end;
      if count > 1 then variance1h := sm / (count - 1) else variance1h := 0
   end;


   function mean (var x: dtdmptrtype; d: demep1type; nn: integer): real;
      var
         i, count: integer; sm: real;
   begin
      sm := 0; count := 0;
      for i := 1 to nn do
         if x^[i, d] <> 999 then
            begin
               sm := sm + x^[i, d]; count := count + 1;
            end;
      if count > 0 then mean := sm / count else mean := 0
   end;

   function variance (var x: dtdmptrtype; d: demep1type; nn: integer): real;
```

```
    var
      i, count: integer;
      mn, sm: real;
  begin
    mn := mean(x, d, nn); sm := 0; count := 0;
    for i := 1 to nn do
      if x^[i, d] <> 999 then
        begin
          sm := sm + sqr(x^[i, d] - mn); count := count + 1;
        end;
    if count > 1 then variance := sm / (count - 1) else variance := 0
  end;



    function rbar (smk,hmk: haplotype; rec: real): real;
  {Calculates the harmonic mean recombination rate between neutral (non habt pref) and selected loci;}
  { smk is the "selection mask", which determines whether or not the gene is selected;}
  { hmk is the "hab pref mask", which determines whether or not the gene is a habitat preference gene;}
  { the function takes into account multiple crossovers, using Haldane's mapping function.}
      var
        nn, ns, nh, g, kn, ks: integer; sum: real;
  begin
  {count the number of neutral and selected loci;}
  { could use nneut and nsel, but safer to recalculate}
    nn := 0; ns := 0; nh := 0;
    for g := 0 to ngenes - 1 do
      begin
        if getgene(g, smk) = 1 then ns := ns + 1;
        if getgene(g, hmk) = 1 then nh := nh + 1;
        if (getgene(g, smk) = 0) and (getgene(g, hmk) = 0) then nn := nn + 1;
      end;
    if nh+ns+nn<>ngenes then scream;
    sum := 0;
    for kn := 0 to ngenes - 1 do {sum over all loci}
      if (getgene(kn, smk) = 0) and (getgene(kn,hmk)=0) then {if neutral}
        for ks := 0 to ngenes - 1 do {sum over all selected loci}
          if getgene(ks, smk) = 1 then {if selected}
            sum := sum + 2 / (1 - power((1 - 2 * rec), abs(kn - ks)));
    rbar := nn * ns / sum;
  end;



    function rbar_all (rec: real): real;
  {Calculates the harmonic mean recombination rate across all loci;}
  { the function takes into account multiple crossovers, using Haldane's mapping function.}
      var
        i, j: integer; sum: real;
  begin
    sum := 0;
    for i := 0 to ngenes - 2 do {sum over all loci}
      for j := i+1 to ngenes - 1 do {sum over all pairs}
        sum := sum + 2 / (1 - power((1 - 2 * rec), abs(j-i)));
    rbar_all := ngenes * (ngenes-1) / (2*sum);
  end;



    procedure init_rec_rates(r:real; var all_rates:recratestype);
  {calculates up harmonic mean recombination rates within sets of loci - need for predictions of D}

    var loci:locitype;

    function rbar_within_gene_type (rec: real; loci:locitype): real;
  {Calculates the harmonic mean recombination rate between all (say) neutral loci;}
  { the function takes into account multiple crossovers, using Haldane's mapping function.}
      var
```

```pascal
            g1,g2,count: integer;
            sum: real;

        function wanted(g:integer):boolean;
        begin
            wanted:=(loci=all)  or ((loci=selected) and (getgene(g,seln_mask)=1))
                or ((loci=habpref) and (getgene(g,hab_mask)=1))
                or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
        end;


    begin
        sum := 0; count:=0;
        for g1 := 0 to ngenes - 2 do {sum over all neutral pref loci}
            if wanted(g1) then
                for g2 := g1+1 to ngenes - 1 do {sum over all selected loci}
                    if wanted(g2) then
                        begin
                            sum := sum + 2 / (1 - power((1 - 2 * rec), abs(g2 - g1))); count:=count+1;
                        end;
        if ((loci=neutral) and (count<>nneut*(nneut-1)/2)) then scream;
        rbar_within_gene_type := count / sum;
    end;


    begin
        for loci:=0 to 3 do
            if nloci[loci]>0 then all_rates[loci]:=rbar_within_gene_type(r,loci) else all_rates[loci]:=0;
    end;

{*********** end procedure init_rec_rates ************ }



    procedure regress (var xb, yb, b: real; x, y: dmchaintype; nn: integer; originflag: boolean);
{finds the regression of y on x; if originflag, this is forced through the origin; }
{ otherwise, y = yb + b*(x-xb) - passing data arrays not pointers }
        var
            x0, y0, sxx, sxy: real; k: integer;
    begin
        sxx := 0; sxy := 0;
        xb := mean1_dmchain(x, nn); yb := mean1_dmchain(y, nn);
        if originflag then x0 := 0 else x0 := xb; if originflag then y0 := 0 else y0 := yb;
        for k := 1 to nn do
        if (x[k]<>999) then
            begin
                sxx := sxx + sqr(x[k] - x0);
                sxy := sxy + (x[k] - x0) * (y[k] - y0);
            end;
        if sxx > 0 then b := sxy / sxx
        else
            begin
                writeln('ERROR in regress: sxx=0');  b := 0
            end;
    end;

    function get_Dmax_reg (var Vg, Vd: dmhabptrtype; ns: integer; h: bittype; ng: integer): real;
        var
            k: integer; slope, pq2bar, Dbar: real; pq2, D: dmchaintype;
    begin
        if ng > 1 then
            begin
                for k := 1 to ndemes do
                if (Vg^[k, h]<>999) and (Vd^[k, h]<>999) then
                    begin
                        pq2[k] := sqr(Vg^[{ns,} k, h] / (2 * ng));
                        D[k] := Vd^[{ns, }k, h] / (2 * ng * (ng - 1));
                    end
                    else begin
                        pq2[k] := 999; D[k] := 999;
```

```pascal
              end;
          regress(pq2bar, Dbar, slope, pq2, D, ndemes, true);
          get_Dmax_reg := slope / 16
        end
    else  get_Dmax_reg := 0; {one or no genes}
  end;
```
{***}

{***}


```pascal
  function get_Dmax (var Vd: dmhabptrtype; h: bittype; ng: integer): real;
  {simply returns the highest value in Vd - this should be incorporated into get_max below,
  with the correction for ng made within main program....}
    var
      k,kmid: integer; dummy: real; D: dmchaintype;
  begin
    if ng > 1 then
      begin
        for k := 1 to ndemes do
        if (Vd^[k, h]<>999) then D[k] := Vd^[k, h] / (2 * ng * (ng - 1)) else D[k] := 999;
        kmid := 1; dummy := 0;
        for k := 1 to ndemes do
          if (D[k] > dummy) and (D[k] <> 999) then
            begin
              kmid := k; dummy := D[k]
            end;
        get_Dmax := dummy
      end
    else   get_Dmax := 0; {one or no genes}
  end;

  function get_Dmax_2h (var Vd_2h: dmptrtype; ng: integer): real;
  {simply returns the highest value in Vd - this should be incorporated into get_max below,
  with the correction for ng made within main program....}
    var
      k,kmid: integer; dummy: real; D: dmchaintype;
  begin
    if ng > 1 then
      begin
        for k := 1 to ndemes do
        if (Vd_2h^[k]<>999) then D[k] := Vd_2h^[k] / (2 * ng * (ng - 1)) else D[k] := 999
        kmid := 1;  dummy := 0;
        for k := 1 to ndemes do
          if (D[k] > dummy) and (D[k] <> 999) then
            begin
              kmid := k;
              dummy := D[k]
            end;
        get_Dmax_2h := dummy
      end
    else
      get_Dmax_2h := 0; {one or no genes}
  end;

  function get_max (var A: dmhabptrtype; h: bittype): real;
  {simply returns the highest value in 2-D array}
    var
      k,kmid: integer; dummy: real;
    begin
      kmid := 1; dummy := 0;
      for k := 1 to ndemes do
        if (A^[k,h] > dummy) and (A^[k,h] <> 999) then
          begin
            kmid := k; dummy := A^[k,h]
          end;
      get_max := dummy
    end;
```

```
{*************************************************************}

    function get_max_2h (var A: dmptrtype): real;
    {simply returns the highest value in 1-D array}
       var
         k,kmid: integer; dummy: real;
       begin
         kmid := 1; dummy := 0;
         for k := 1 to ndemes do
            if (A^[k] > dummy) and (A^[k] <> 999) then
               begin
                  kmid := k;
                  dummy := A^[k]
               end;
         get_max_2h := dummy
       end;
{*************************************************************}

    procedure get_deltap_max (var A: dmptrtype; var dp_max: real; var deme_dp_max: demep1type);
    {simply returns the highest value in array}
       var
         k,kmax: integer; dummy: real;
       begin
         kmax := 1; dummy := 0;
         for k := 1 to ndemes do
            if (A^[k] > dummy) and (A^[k] <> 999) then
               begin
                  kmax := k; dummy := A^[k]
               end;
         deme_dp_max:=kmax; dp_max := dummy
       end;
{*************************************************************}


    function get_Dmean (var Vd: {dt}dmhabptrtype; ns: integer; h: bittype; ng: integer): real;
    {simply returns the highest value in Vd}
       var
         k,count: integer; sum: real; D: dmchaintype;
    begin
       if ng > 1 then
         begin
            for k := 1 to ndemes do
            if (Vd^[k, h]<>999) then D[k] := Vd^[k, h] / (2 * ng * (ng - 1)) else D[k] := 999;
            sum:=0; count:=0;
            for k := 1 to ndemes do
               if (D[k] <> 999) then
                  begin
                     count := count+1; sum := sum+D[k]
                  end;
            if count>0 then get_Dmean := sum/count else get_Dmean:=999;
         end
       else   get_Dmean := 0; {one or no genes}
    end;
{*************************************************************}




    function predD(var slope2h: dtptrtype; ns: integer; r, mig: real):real;
{Prediction of D from recomb'n rate (harmonic mean or adjacent as required) & observed width}
    begin
       predD := mig * slope2h^[ns] * slope2h^[ns]/ r;
    end;
{*************************************************************}


    function NEWwidth_2hL(pp: popptrtype; loci:locitype; lmax: real): real;
{takes old width function, uses it to calculate the width of the cline for each locus separately
```

then takes the arithmetic mean}
```pascal
    var
        slp, sum: real; g, count: genetype; cline1: dmptrtype; wanted: boolean;

    function slope_2h_1gene (g: genetype): real;
{This function searches for points at which ln(p/q) is -2 and +2; (p=0.2,0.8) it then measures }
{the slope by calculating a regression over this range. The algorithm is not foolproof, & should be}
{checked against a few graphs. If the system does not span the required range, zero is returned.}
        var
            d, k, kmin, kmax: demep1type; deme, cline, lp: dmchaintype;
            slope, pcritmin, pcritmax, demebar, lpbar: real;

    begin
        for d:=0 to ndemes+1 do         {gives cline in one gene}
            cline[d]:=getgenefreq(pp, dummyhabpop, dummyhabpop_size, d, 1, g);
        pcritmax := 1 / (1 + exp(-lmax)); pcritmin := 1 / (1 + exp(lmax));
        kmin := ndemes + 1;
        repeat kmin := kmin - 1 until (kmin = 1) or (cline[kmin] < pcritmin);
        kmax := 0;
        repeat kmax := kmax + 1 until (kmax = ndemes) or (cline[kmax] > pcritmax);
        if (kmin = 1) or (kmax = ndemes) then slope_2h_1gene := 0
        else
            begin
                for k := kmin to kmax do
                    begin
                        lp[k - kmin + 1] := logit(cline[k], 5); deme[k - kmin + 1] := k
                    end;
                regress(demebar, lpbar, slope, deme, lp, kmax - kmin + 1, false);
                slope_2h_1gene:=slope;
            end; {else}
    end;

    begin
        sum:=0; count:=0;
        for g := 0 to ngenes - 1 do
            begin
                wanted:=(loci=all)  or ((loci=selected) and (getgene(g,seln_mask)=1))
                    or ((loci=habpref) and (getgene(g,hab_mask)=1))
                    or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
                if wanted then
                    begin
                        slp:=slope_2h_1gene(g);         {calculates slope of 1 gene cline}
                        if slp>0 then
                            begin
                                sum:=sum+(4/slp);  count:=count+1;
                            end;
                    end;
            end;
        if (count >0) and (sum>0) then NEWwidth_2hL:=sum/count
        else NEWwidth_2hL:=999;
    end;


    function slope_2hL(pp: popptrtype; loci:locitype; lmax: real): real;
{takes old width function, uses it to calculate the slope of the cline for each locus separately
then takes the arithmetic mean - the inverse of this can then be used to give the harmonic mean width}
    var
        slp, sum: real; g, count: genetype; cline1: dmptrtype; wanted: boolean;

    function slope_2h_1gene (g: genetype; lm: real): real;
{LK: 11.6.96: USING NB'S ALGORITHM FOR WIDTH; don't know if I'm optimising use of pointers}
{This function searches for points at which ln(p/q) is -2 and +2; (p=0.2,0.8) it then measures }
{the slope by calculating a regression over this range. The algorithm is not foolproof, & should be}
{checked against a few graphs. If the system does not span the required range, zero is returned.}
        var
            d, k, kmin, kmax: demep1type; deme, cline, lp: dmchaintype;
            slope, pcritmin, pcritmax, demebar, lpbar: real;
```

```pascal
    begin
      for d:=0 to ndemes+1 do        {gives cline in one gene}
        cline[d]:=getgenefreq(pp, dummyhabpop, dummyhabpop_size, d, 1, g);
      pcritmax := 1 / (1 + exp(-lm));
      pcritmin := 1 / (1 + exp(lm));
      kmin := ndemes + 1;
      repeat kmin := kmin - 1 until (kmin = 1) or (cline[kmin] < pcritmin);
      kmax := 0;
      repeat kmax := kmax + 1 until (kmax = ndemes) or (cline[kmax] > pcritmax);
      if (kmin = 1) or (kmax = ndemes) then  slope_2h_1gene := 0
      else
        begin
          for k := kmin to kmax do
            begin
              lp[k - kmin + 1] := logit(cline[k], 5); deme[k - kmin + 1] := k
            end;
          regress(demebar, lpbar, slope, deme, lp, kmax - kmin + 1, false);
          slope_2h_1gene:=slope;
        end; {else}
    end;

  begin
    sum:=0; count:=0;
    for g := 0 to ngenes - 1 do
      begin
        wanted:=(loci=all)  or ((loci=selected) and (getgene(g,seln_mask)=1))
            or ((loci=habpref) and (getgene(g,hab_mask)=1))
            or ((loci=neutral) and (getgene(g,seln_mask)=0) and (getgene(g,hab_mask)=0));
        if wanted then
          begin
            slp:=slope_2h_1gene(g,lmax);     {calculates slope of 1 gene cline}
            sum:=sum+slp/4;       {slope of logit is 4 times that of untransformed cline at centre}
            count:=count+1;
            if loci=neutral then ind_centre_slope_p8^[sample,g]:=slp;
          end;
      end;
    if (count >0) and (sum>0) then
      begin
        slope_2hL:=sum/count;
        if loci=neutral then ind_centre_slope_p8^[sample,0]:=sum/count;
      end
    else slope_2hL:=999;
  end;


  function get_width_2h (var zb: {dt}dmptrtype; ns, ng: integer): real;
{This function searches for points at which ln(p/q) is -2 and +2; (p=0.2,0.8) it then measures }
{the width by calculating a regression over this range. If the system does not span the required range,
zero is returned.}
    const
      lmax = 1.38629;
    var
      k, kmin, kmax: demep1type; deme, lp: dmchaintype; slope, zcritmin, zcritmax, demebar, lpbar: real;
  begin
    zcritmax := 2 * ng / (1 + exp(-lmax));
    zcritmin := 2 * ng / (1 + exp(lmax));
    kmin := ndemes + 1;
    repeat  kmin := kmin - 1 until (kmin = 1) or (zb^[kmin] < zcritmin);
    kmax := 0;
    repeat  kmax := kmax + 1 until (kmax = ndemes) or (zb^[kmax] > zcritmax);
    if (kmin = 1) or (kmax = ndemes) then  get_width_2h := 0
    else
      begin
        for k := kmin to kmax do
          begin
            lp[k - kmin + 1] := logit(zb^[k] / (2 * ng), 5); deme[k - kmin + 1] := k
          end;
        regress(demebar, lpbar, slope, deme, lp, kmax - kmin + 1, false);
```

```pascal
        if slope <> 0 then    get_width_2h := 4 / slope
        else   get_width_2h := 9999;
      end;
  end;
{*********************************************************}



  procedure minimum_fitness (var wbmin, wbreg: real; var zb: dmptrtype;  var wb: dmptrtype;
              ns,ng:integer; var central_deme:demep1type);
{taking the minimum fitness out of barrier, so that it does it when there are just selected loci}
const reduction = 0.05;
var
    k,kmid,kwreg_left,kwreg_right:demep1type; pq, wbtemp: dmchaintype;
    wbmin_dummy,pqbar, wbbar, sl: real;

begin
    wbmin_dummy:=1;  kmid:=round(ndemes/2);
    for k := 1 to ndemes do
      if wb^[k] < wbmin_dummy then
        begin
          kmid := k;  wbmin_dummy := wb^[k]
        end;
    central_deme := kmid;
    kwreg_left := kmid;
    repeat  kwreg_left := kwreg_left - 1
    until (kwreg_left = 0) or (wb^[kwreg_left]>(1 - 4*(1 - wbmin_dummy) * reduction));
{find right deme where fitness=1}
    kwreg_right := kmid;
    repeat  kwreg_right := kwreg_right + 1
    until (kwreg_right = ndemes+1) or (wb^[kwreg_right]>(1 - 4*(1 - wbmin_dummy) * reduction));
    for k := kwreg_left to kwreg_right do
      begin
        wbtemp[k - kwreg_left + 1] := wb^[ k];
        pq[k - kwreg_left + 1] := (zb^[k] / (2 * ng)) * (1 - zb^[k] / (2 * ng));
      end;
{find minimum mean fitness by regression on pq - over area where w<1}
    regress(pqbar, wbbar, sl, pq, wbtemp, (kwreg_right-kwreg_left+1), false);
    wbmin:=wbmin_dummy;
    wbreg := wbbar + (0.25 - pqbar) * sl;
end;

   procedure calculate_barrier (var step,NEWstep,slope_left, slope_mid, slope_right,
            NEWslope_left, NEWslope_right:real;var zb:dmptrtype;
            var wb:dmptrtype; ns, ng: integer);
{This calculates barrier strength, from the average shape of a set of neutral loci. The method is
complicated, and may not be entirely reliable. First, the mean fitness (wb) is searched to find its minimum
(wbmin near kmid). Then, the points at which mean fitness is reduced by 5% of its maximum drop are
calculated. Outside this region (kleft, kright), hitchhiking should be negligible. The slope of the clines to
left and right are calculated, using least-squares regression on logit transformed data. Only the }region
which is not fixed is used (ie, abs(logit) < 5. The slope at the centre is found by regression of logit
transformed data, either between kleft, kright, or over 6 demes, whichever is the smaller. The barrier
strengths for the two directions are step/slope_left and step/slope_right}

    const
      reduction=0.05;
    var
      k, kleft, kright, kfarleft, kfarright, kmid, kmidleft, kmidright: integer;
      kwred4_right,kwred4_left:integer; lp, deme, wbtemp: dmchaintype;
      zbmin, zbmax,wbmin: real; lpbar_left, demebar_left, sl_logit_left: real;
      xmid, xleft, xright, zleft, zright, lpbar_mid, demebar_mid, sl_logit_mid: real;
      lpbar_right, demebar_right, sl_logit_right: real;
    begin
      zbmin := 2 * ng / (1 + exp(5));        {set minimum and maximum zb}
      zbmax := 2 * ng / (1 + exp(-5));
      kmid := 1;                 {find position of minimum wbar}
      wbmin := 1;
      for k := 1 to ndemes do  wbtemp[k] := wb^[k]; {strictly no longer necessary!}
```

```
        for k := 1 to ndemes do
          if wbtemp[k] < wbmin then
            begin
              kmid := k; wbmin := wbtemp[k]
            end;
        kleft := kmid + 1;                {find left edge of region of reduced fitness}
        repeat  kleft := kleft - 1 until (kleft = 0) or (wbtemp[kleft] > (1 - (1 - wbmin) * reduction));
        if kleft = 0 then
          writeln('ERROR in calculate_barrier: can''t find left edge of low wb region ');
        kright := kmid - 1;                {find right edge of region of reduced fitness}
        repeat  kright := kright + 1 until (kright = ndemes + 1) or (wbtemp[kright] > (1 - (1 - wbmin) * reduction));
        if kright = ndemes + 1 then
          writeln('ERROR in calculate_barrier: can''t find right edge of low wb region ');
        kfarleft := kleft-3;                {find far left edge, where fixation is reached}
        repeat  kfarleft := kfarleft - 1 until (zb^[kfarleft] < zbmin) or (kfarleft = 0);
        kfarright := kright+3;                          {find far right edge, where fixation is reached}
        repeat  kfarright := kfarright + 1 until (zb^[kfarright] > zbmax) or (kfarright = ndemes+1);
        if (kmid - kleft) > 3 then  kmidleft := kmid - 3    {find leftmost point to count central regression}
        else  kmidleft := kleft;{find leftmost point to count central regression}
        if (kright - kmid) > 3 then   kmidright := kmid + 3  {find rightmost point for central regression}
        else  kmidright := kright;
        for k := kfarleft to kleft do        {set up leftmost regression}
          begin
            deme[k - kfarleft + 1] := k; lp[k - kfarleft + 1] := logit(zb^[k] / (2 * ng), 5);
          end;
        regress(demebar_left, lpbar_left, sl_logit_left, deme, lp, kleft - kfarleft + 1, false);
        for k := kmidleft to kmidright do        {set up middle regression}
          begin
            deme[k - kmidleft + 1] := k; lp[k - kmidleft + 1] := logit(zb^[ k] / (2 * ng), 5);
          end;
        regress(demebar_mid,lpbar_mid,sl_logit_mid,deme,lp,kmidright-kmidleft+1, false);
        for k := kright to kfarright do        {set up rightmost regression}
          begin
            deme[k - kright + 1] := k; lp[k - kright + 1] := logit(zb^[k] / (2 * ng), 5);
          end;
        regress(demebar_right,lpbar_right,sl_logit_right,deme,lp,kfarright- kright+1, false);
{looking for points where w=1-4*red - just to take MIDslope values from : another DINOSAUR}
        kwred4_left := kmid;
        repeat  kwred4_left := kwred4_left - 1
        until (kwred4_left = 0) or (wbtemp[kwred4_left]>(1 - 4*(1 - wbmin) * reduction));
        kwred4_right := kmid;                {find right deme where fitness=1}
        repeat  kwred4_right := kwred4_right + 1
        until (kwred4_right = ndemes+1) or (wbtemp[kwred4_right]>(1 - 4*(1 - wbmin) * reduction));
{these regressions are on a logit scale (z).  What is wanted are dp/dx, not dz/dx;}
{to find these, the relation dp/dx = pq dz/dx = exp(z)/sqr(1+exp(z)) dz/dx = pq dz/dx is used.}
{This means that the "centre" of the barrier must be defined. I take it to be the point in the }
{central regression at which p=0.5, z=0: an implicit assumption of symmetry here}
{A modification: slopes are calculated at the point where the regressions intersect}
{xmid is the point at which the central regression intersects the origin}
{27/8/96 - no: this leads to monmumental problems, because when the slopes are very similar, the point of
intersection can be miles away eg. -7095, which will be returned as 295.}
{27.8 slope variables take midpoints of regressions;NEWslope variables take startpoints of regressions
ie. where fitness reduced by certain amounts}
        slope_left := (exp(lpbar_left) / sqr(1 + exp(lpbar_left))) * sl_logit_left;
        slope_mid := sl_logit_mid / 4;
        slope_right := (exp(lpbar_right) / sqr(1 + exp(lpbar_right))) * sl_logit_right;
        step := 1 / (1 + exp(-lpbar_right)) - 1 / (1 + exp(-lpbar_left));
        NEWslope_left := (zb^[ns, }kleft] / (2 * ng)) * (1 - zb^[ns, }kleft] / (2 * ng)) * sl_logit_left;
        NEWslope_right := (zb^[{ns, }kright] / (2 * ng)) * (1 - zb^[{ns, }kright] / (2 * ng)) * sl_logit_right;
        NEWstep := (zb^[{ns, }kright]-zb^[{ns, }kleft]) / (2 * ng);
      end;

      function get_max_Fis_centre(var Fisloci: dmhabptrtype; hb: bittype; centre_deme:demep1type):real;
{Want max Fis, but only interested in values at the centre (quite often get haywire results at
edges because of small population sizes -> consider 10 central demes?}
        var
          d,start_deme,end_deme:demep1type; dummy:real;
      begin
```

```pascal
      if (centre_deme>=5) then start_deme:=centre_deme-5 else start_deme:=0;
      if (centre_deme<=(ndemes-5)) then end_deme:=centre_deme+5 else end_deme:=ndemes+1;
      dummy:=0;
      for d:=start_deme to end_deme do if Fisloci^[d,hb]>dummy then dummy:=Fisloci^[d,hb];
      get_max_Fis_centre:=dummy;
    end;


   function get_max_Fis_centre_2h(var Fis_2hloci: dmptrtype;centre_deme:demep1type):real;
{Want max Fis, but only interested in values at the centre (quite often get haywire results at
edges because of small population sizes -> consider 10 central demes?}
      var
         d,start_deme,end_deme:demep1type; dummy:real;
   begin
      if (centre_deme>=5) then start_deme:=centre_deme-5 else start_deme:=0;
      if (centre_deme<=(ndemes-5)) then end_deme:=centre_deme+5 else end_deme:=ndemes+1;
      dummy:=0;
      for d:=start_deme to end_deme do if (Fis_2hloci^[d]>dummy) then dummy:=Fis_2hloci^[d];
      get_max_Fis_centre_2h:=dummy;
    end;


   end.
```

Mechanisms maintaining species differentiation: predator-mediated selection in a *Bombina* hybrid zone

LOESKE E. B. KRUUK* AND JASON S. GILCHRIST

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK
(loeske@tattoo.ed.ac.uk)

SUMMARY

Mechanisms which prevent gene flow will maintain differentiation between species, and therefore contribute to biological diversity. We describe an experimental study of such mechanisms in a hybrid zone between the fire-bellied toad *Bombina bombina* and the yellow-bellied toad *B. variegata*. In this system, preference for different breeding habitats reduces the frequency of hybridization. A comparison of habitat ecology shows that the semi-permanent ponds in which *B. bombina* usually breeds have higher densities of aquatic predators than the temporary puddles typically used by *B. variegata*. We test for behavioural adaptations in tadpoles to these different levels of predation. *B. bombina* tadpoles are significantly less active than *B. variegata*, both before and after the introduction of a predator to an experimental arena; this reduces their vulnerability as many predators detect prey through movement. Behavioural differences translate into differential survival: *B. variegata* suffer higher predation rates in laboratory experiments with three main predator types (*Triturus* sp., *Dytiscus* larvae, *Aeshna* nymphs). This differential adaptation to predation will help maintain preference for alternative breeding habitats, and thus serve as a mechanism maintaining the distinctions between the two species.

INTRODUCTION

If two taxa interbreed naturally and produce viable, fertile hybrids, the persistence of numerous morphological, genetic and physiological differences between them requires explanation. Mechanisms which prevent the taxa from merging will contribute to the maintenance of biological diversity. These mechanisms may be genetic, environmental, or a combination of both; for example, adaptation of different genotypes to different ecological niches may play a significant role. We describe here an investigation of one such mechanism in two hybridizing taxa of fire-bellied toad (genus *Bombina* (Anura: Discoglossidae)).

The fire-bellied toad *Bombina bombina* is found in the lowlands of central and eastern Europe, whereas the yellow-bellied toad *Bombina variegata* occurs at higher altitudes of southern and western Europe; after common usage, we refer to the two as separate species. Divergence occurred during geographic isolation an estimated 2–7 million years ago. Following postglacial range expansion, their distributions now meet at altitudinal transitions, overlapping slightly (Arntzen 1978). Despite many morphological, life history and biochemical differences, the two species interbreed within the region of overlap and produce viable, fertile hybrids, resulting in a stable hybrid zone usually less than 10 km wide (Szymura 1993, and references herein).

The two species show a preference for different breeding habitats: *B. bombina* breeds mainly in semi-permanent ponds, whereas *B. variegata* is a characteristic puddle breeder (Madej 1973; Arntzen 1978; Barandun 1995). Although interbreeding will continually break down original gene combinations, preference for alternative habitats is maintained within the hybrid zone, generating an association between habitat type and the genotype of individuals found there (MacCallum 1994; Bugter *et al.* 1995). For example, at the centre of our study site in Croatia, the frequency of individuals with more than half *B. variegata* alleles (at four diagnostic allozyme loci) was 9% in two large ponds, compared with a frequency of 61% in immediately adjacent puddles formed in wheel-ruts (MacCallum 1994). The preference appears most strong amongst *B. variegata* and *B. variegata*-like hybrids. Frequent desiccation of puddles forces migration in search of new sites, and yet the association remains consistent despite the availability of both habitat types within individual toads' dispersal ranges. To be maintained in areas of interbreeding, the habitat preference must have adaptive advantages. From an ecological perspective, these advantages would explain the occupation of different environmental niches; from a population genetics perspective, they reduce the potential for hybridization and so constitute a barrier to gene flow.

The different breeding habitats favoured by the two *Bombina* species can be classified by their permanence. The relative longevity of ponds creates a very different aquatic environment from ephemeral puddles, which will necessarily contain little vegetation and only the most opportunistic fauna. In particular, the abundance of potential aquatic predators, both vertebrate and

---

* Author for correspondence.

invertebrate, on amphibian eggs and larvae should be higher in ponds. Previous studies have demonstrated the importance of interspecific interactions such as predation on larvae in determining species composition in anuran guilds (Morin 1983; Wilbur *et al.* 1983; Woodward 1983; Morin 1986; Cortwright & Nelson 1990). We therefore consider the potential role of predation in defining differential selection pressures within the *Bombina* system, and ask specifically whether predation could explain *B. variegata*'s avoidance of the pond habitat.

We test the hypothesis that predation pressure in ponds, *B. bombina*'s preferred habitat, is higher than in puddles, and that *B. variegata*'s avoidance of ponds is therefore associated with inferior adaptation of its tadpoles to a predator-rich environment. In such an environment, any trait which reduces the risk of mortality from predation should be strongly favoured. Movement, in addition to increasing encounter rate, is a cue to many aquatic predators which detect their prey using visual or mechanosensory reception (e.g. Richards & Bull 1990). A tadpole moving around foraging will therefore be at greater risk of predation than a less active one (Werner & Anholt 1993). Differences in microhabitat use, morphology and palatability will also determine the relative vulnerability of both species to predation. We consider three predictions from the above hypothesis:

(1) the abundance of predators is higher in ponds than in puddles during the *Bombina* larval period;

(2) the behaviour of *B. bombina* tadpoles makes them less vulnerable than *B. variegata* tadpoles to a predator: specifically, *B. bombina* tadpoles are less active;

(3) *B. bombina* tadpoles suffer lower mortality rates from predators than *B. variegata*.

2. MATERIALS AND METHODS

(a) Study site

The study site is located around the village of Pešćenica, 25 km south-east of Zagreb, Croatia, at an altitudinal transition between low, forested hills and arable flood plains of the River Sava. In this area, the distance across the hybrid zone between pure *B. bombina* and pure *B. variegata* populations is approximately 10 km (MacCallum 1994); altitude changes from *ca.* 100 m to 280 m. Both habitat types, ponds and puddles, are found within the hybrid zone; the majority of puddles are formed in tractor wheel-ruts.

(b) Ecological surveys

Five breeding sites of each habitat type, distributed across the hybrid zone, were surveyed. Predator abundance was compared using a catch-per-unit-effort approach: at two locations in each site, three sweeps with a metal sieve (25 cm diameter) were made. Sweeps were taken immediately below the water surface: preliminary samples with bottle traps indicated that, in deeper ponds, *Bombina* larvae were found in higher strata; depth was also constrained by the shallowness of wheel-ruts. All fauna caught were identified at least to the level of family; the abundance per sample of those known to be major tadpole predators (Fitter & Manuel 1995) and shown in pilot trials to attack *Bombina* larvae was then compared.

(c) Behavioural experiments

Bombina variegata eggs were collected from breeding sites on one side of the hybrid zone, *B. bombina* eggs from the other side. Within a site, egg batches were taken from as widespread an area as possible, in order to maximize the number of families sampled, and not all clusters were taken. Eggs were reared in laboratory aquaria; after hatching, larvae were provided with abundant food in the form of powdered nettle leaves. Eggs were collected between 24–28 May 1995, and experiments conducted between 9–16 June 1995, by which time tadpoles had reached Gosner developmental stages 26–28 (Gosner 1960). Within a trial, tadpoles were matched according to developmental stage and, as far as was possible, on size.

For predation trials we used the most common species of aquatic predators, as determined by the ecological surveys: final instar *Aeshna* (hawker dragonfly) nymphs, final and penultimate instar *Dytiscus* sp. (great diving beetle) larvae, *Triturus dobrogicus* (Danube crested newt) and *T. vulgaris* (smooth newt) adults. These were fed on tadpoles of other anuran species (*Rana* and *Hyla* sp.), until 24 h before an experiment.

Fresh non-chlorinated water at 20 °C was used for each trial, thus preventing possible carry-over of semiochemical cues (Petranka *et al.* 1987). Tadpoles were transferred to opaque plastic containers containing water and abundant food (nettle powder) 1 h prior to an experiment, during which acclimatization period container lids were left on.

(i) Experiment I: activity levels

Activity levels of the two species before and after the introduction of a predator were compared. For each trial, five tadpoles (at Gosner developmental stages 27–28) of the same species were placed together in a 2-litre plastic container. Behaviour was scored by an observer standing motionless approximately 1 m away; observations were dictated. We calculated an activity index of the mean number of tadpoles in the group of five which showed any movement during a series of 5 s observation periods taken at 1 min intervals; each series consisted of 12 observations. This gave a value on a continuous scale between 0 and 5; no assumption was required of independence of either individuals' behaviour or of behaviour across time points.

After the first 12 observations, a single predator was added to each container. *Triturus vulgaris* were used in 20 trials, and *Dytiscus* larvae in another 20, with the two predator types assigned equally to the two tadpole species. Tadpole behaviour was measured as a function of a perceived predation threat, and not of actual predation events. To facilitate this, *T. vulgaris* adults of a size too small to damage the tadpoles were used, and the mandibles of each *Dytiscus* larva were temporarily taped down. A further 12 observations were made immediately after the predator's introduction, and an activity index for the group calculated as before.

Trials were conducted at 0900–1100 and at 1700–1900 h. A total of 40 trials were run, 20 for each species; each tadpole was only used once.

(ii) Experiment II: survival rates

Vulnerability of the two species to predation was compared using predator preference trials. For each trial, ten *B. bombina* and ten *B. variegata* tadpoles were placed together in a 6-litre container, with a moderate density of aquatic plants. Containers were covered throughout to minimize external disturbance. After the acclimatization period, a single predator was introduced into each container. The number of

ials run with each predator type were: *Dytiscus* larvae, 16; [Ae]shna nymphs, 19; *Triturus dobrogicus*, 2; *Triturus vulgaris*, 5. [e]ach predator was used in only one trial.

Predators were left to feed until approximately half the [ta]dpoles present had been eaten; this was determined by [br]ief scans of the containers at 2 h intervals, and so was not [al]ways exact. Average time taken was 13.4 h (± 9.9 s.d.). [T]he number of surviving tadpoles of either species was then [re]corded; tadpoles of *B. bombina* and *B. variegata* are [di]stinguishable by their stripe patterns. Predator feeding bias [w]as quantified using Manly's preference index:

$$= \frac{\ln(p_v)}{\ln(p_v p_b)},$$

[w]here p_v is the proportion of *B. variegata* surviving out of the [in]itial ten and p_b is the proportion of *B. bombina* (Manly 1974; [Me]eson 1978). The index allows for the effect of prey [de]pletion on availability. Values range from 0 to 1, with 0.5 [re]presenting random selection of prey and 1 representing [on]ly *B. variegata* being taken. Preference indices were analysed [se]parately for two stage groups: tadpoles at Gosner stage 26 [('ha]tchlings'), and those at Gosner stages 27–28.

RESULTS

(a) Ecological surveys

The mean abundances of different predator [ca]tegories are given in table 1. In all four categories a

null hypothesis of equal abundance in ponds and puddles can be rejected. The abundance of (i) newts (*Triturus* sp.), (ii) dragonfly (Suborder: Anisoptera) and damselfly (Suborder: Zygoptera) nymphs and (iii) diving beetle (Family: Dytiscidae) adults or larvae is significantly greater in ponds. Salamander (*Salamandra salamandra*) larvae were found exclusively but only occasionally in puddles. The density of predators in pond samples was therefore substantially higher than in puddle samples.

(b) Experiment I: activity levels

There were no tadpole groups for which zero activity was recorded, either before or after the introduction of a predator. Movement was strongly associated with feeding. Mean activity indices for each species in the absence and then presence of a predator are presented in figure 1. Table 2 contains an analysis of variance of the data.

The difference in the overall activity levels of the two species was highly significant: *B. variegata* tadpoles were consistently more active than *B. bombina*, with the parameter estimate indicating an average increase of 2.85 units. After the introduction of a predator, activity was significantly lower in both species: the overall effect was to lower the activity index by 0.58. There

[Ta]ble 1. *Abundance of predator categories in pond and puddle samples*

[M]eans and standard errors of abundance of each category in sieve-sweep samples. Categories are as follows: 1. Newt adults: [*Tr*]*iturus dobrogicus*; *T. alpestris*; *T. vulgaris*; 2. Dragonfly (Anisoptera) and damselfly (Zygoptera) nymphs; 3. Diving beetle [(D]ytiscidae) adults and larvae; and 4. Fire salamander (*Salamandra salamandra*) larvae.)

[pre]dator category	ponds ($n = 5$)	puddles ($n = 5$)	G-test of equal distribution; d.f. $= 1$
newts	4.83 ± 2.41	1.33 ± 0.76	12.6; $p < 0.01$
dragonfly/damselfly nymphs	16.5 ± 7.07	0	137.0; $p < 0.01$
diving beetles	12.0 ± 3.50	0.17 ± 0.17	90.6; $p < 0.01$
salamander larvae	0	0.50 ± 0.22	4.16; $p < 0.01$

[T]able 2. *ANOVA and parameter estimates of tadpole activity levels*

[S]pecies term compares *B. variegata* with *B. bombina*; presence of predator term includes both predator types. Time of day effect [(m]orning or afternoon) is nested within species, to give a parameter estimate for either species. Repeated measures on each [ta]dpole group (one before the introduction of a predator, one after) allow comparison between groups, nested within time and [sp]ecies.)

[sou]rce	d.f.	sequential SS	F ratio	p value
[sp]ecies	1	17.500	91.047	< 0.001 ***
[pr]esence of predator	1	11.438	59.509	< 0.001 ***
[spe]cies * predator presence	1	0.657	3.418	0.0723
[tim]e[species]	2	5.030	13.085	< 0.001 ***
[gro]up[time (species)]	36	15.593	2.253	0.0075 **
[pre]dator * time[species]	2	0.107	0.2682	0.766
[err]or	36	7.197		
[tot]al	79	57.522		

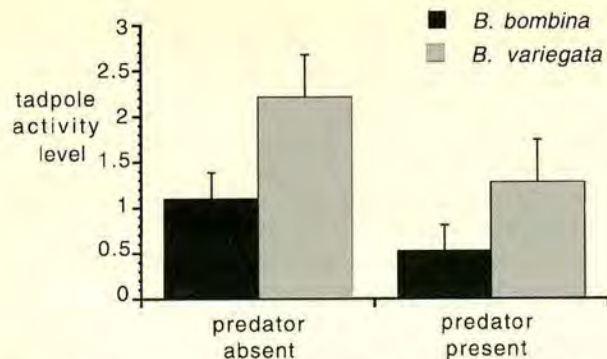| [ter]m	parameter estimate ($\pm$ s.e.)
[int]ercept	1.024 (± 0.329)
[spe]cies	2.848 (± 0.450)
[pre]sence of predator	-0.575 (± 0.139)
[spe]cies * predator presence	-0.363 (± 0.196)
[*B.*]*bombina*: time effect	0.125 (± 0.438)
[*B.*]*variegata*: time effect	-1.958 (± 0.438)

Figure 1. Mean activity indices (and s.e.) for *B. bombina* and *B. variegata* tadpoles before and after the introduction of a predator (see text for derivation of indices). Table 2 contains ANOVA of activity level.

was no evidence of a difference between species in response to predator introduction.

Time of day did not affect the activity level for *B. bombina*, but reduced that of *B. variegata* by nearly two units, despite controlled water temperatures. There was also significant variation between groups of tadpoles within species and time of day.

Activity level in the presence of a predator can be expressed as a proportion of the activity level before the introduction. *B. bombina*'s activity was reduced to an average of 55.12 % (± 6.77 s.e.) of the original level; *B. variegata*'s to 61.04 % (± 8.36). An analysis of variance of log-transformed percentages showed no difference between species (F ratio $= 0.23$; d.f. $= 1,36$; $p = 0.64$),
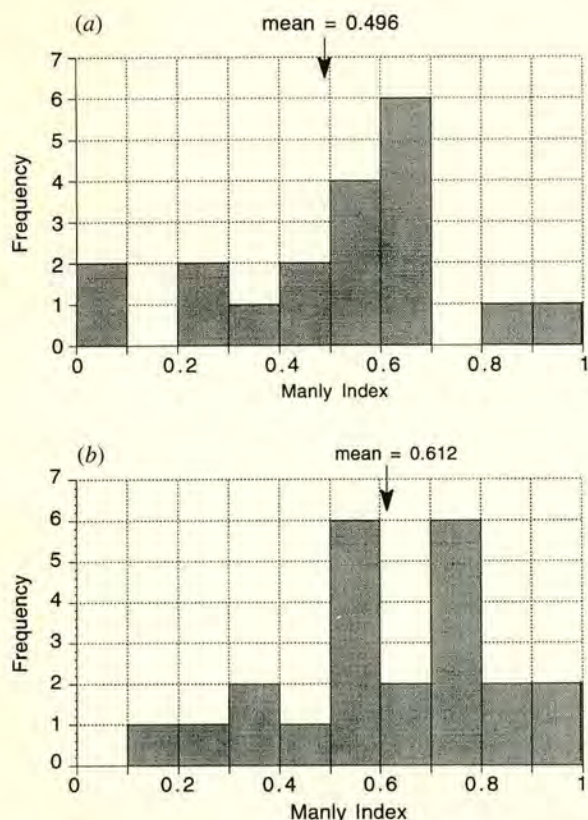

Figure 2. Distribution of Manly preference index (see text for derivation) from predator preference trials for two categories of larval developmental stage.

nor between predator types (F ratio $= 1.38$; d.f. $= 1,36$; $p = 0.25$). There was therefore no evidence that *B. bombina* reduce their activity more than *B. variegata* or vice versa, nor that the effect of a *Dytiscus* larvae was different from that of a *Triturus vulgaris* adult.

(c) Experiment II: survival rates

The Manly preference indices for the two developmental stage groups are presented in figure 2. Amongst the hatchling size class (Gosner stage 26), prey depletion was apparently random with respect to species (mean index $= 0.496 \pm 0.057$ s.e.; t-statistic $= -0.078$; $n = 19$; $p = 0.93$). However, at the later stages (Gosner stages 27–28) *B. variegata* were more vulnerable than *B. bombina*: the mean preference index was significantly greater than 0.5 (mean index $= 0.612 \pm 0.045$; t-statistic $= 2.49$; $n = 23$; $p = 0.02$). ANOVA of the preference index (*Triturus* sp., *Dytiscus* larvae and *Aeshna* nymphs) showed no evidence of a difference between the three predator types (*Triturus* sp., *Dytiscus* larvae and *Aeshna* nymphs: F ratio $= 2.03$; d.f. $= 2,36$; $p = 0.146$) nor of an interaction between predator type and developmental class (F ratio $= 0.065$; d.f. $= 2,36$; $p = 0.937$).

4. DISCUSSION

The results confirm the prediction that, during the *Bombina* larval period, predator density is higher in the semi-permanent ponds in which the fire-bellied toad *B. bombina* breeds than in the temporary puddles which are the typical habitat of the yellow-bellied toad, *B. variegata*. In experimental trials, *B. bombina* tadpoles differed from *B. variegata* in their activity levels: *B. bombina* were consistently less active and spent less time moving around feeding than *B. variegata*. Both species responded to the disturbance induced by introduction of a predator by reducing activity, but *B. variegata* still remained more than twice as active. Once past the hatchling stage, *B. variegata* tadpoles suffered higher levels of mortality than *B. bombina* in predator choice experiments. We were restricted by availability of hatchlings at Gosner stage 26, using them for Experiment II only. Whilst a similar comparison of their activity levels would provide a more complete picture, we know that at this stage tadpoles are largely inactive (Lawler 1989; personal observations), so it is unlikely that their inclusion would affect our conclusions.

Do the behavioural differences demonstrated imply adaptation to different breeding habitats? Development under the risk of predation creates a trade-off between resource-acquisition and predator avoidance (Wilbur & Fauth 1990; Werner & Anholt 1993; Skelly 1995). Low activity rate will reduce vulnerability, but will also decrease foraging rate and, therefore, growth and development rate. Conversely, higher activity levels facilitate faster development rates and so a shorter time to reaching either a size refugia (e.g. Richards & Bull 1990) or metamorphosis: the average larval period for *B. variegata* is 87 % of that for *B. bombina* (Nürnberger *et al.* 1995). This implies that *B. variegata* tadpoles are at risk for less time (a frequently

verlooked point). The benefit of higher activity might herefore outweigh the cost of increased vulnerability n a predator-rich environment. However, physio-ogical constraints should impose diminishing returns n increases in development rate with time active, vhereas the probability of an encounter with a redator will increase linearly with time active. The ncremental benefit to development rate of increased ctivity would be lower than the incremental cost in rms of predation rate (Werner & Anholt 1993). A less ctive species will therefore be better adapted to an nvironment in which predation is a regulating factor, n spite of the longer larval period incurred. Our results mply that the response to the growth–mortality trade-ff differs between *B. bombina* and *B. variegata* tadpoles s predicted by the ecology of their usual breeding abitats; the genetically-determined behavioural atterns demonstrated here will serve to increase *B. ombina*'s fitness relative to *B. variegata* in a semi-ermanent pond.

Differences in tadpole behaviour and susceptibility o predation have been found in other taxa and, milarly, have been invoked to explain species' istributions along environmental gradients or oc-upation of alternative ecological niches (Morin 1983; Voodward 1983; Lawler 1989; Chovanec 1992; Verner & McPeek 1994; Skelly 1995). The data are onsistent with the hypothesis that predator-rich nvironments should contain less active species. We do ot, however, know of any other study where the ontrast has been drawn between two regularly terbreeding species. Traits characteristic of either ombina species are in strong statistical association inkage disequilibrium) within hybrid populations MacCallum 1994); selection on differences at the rval stage will therefore cause a correlated advantage f habitat preference in adults. Simulation models onfirm that a selectively-advantageous habitat reference, by forcing assortative mating, will play an mportant role in maintaining the integrity of the enomes of two hybridizing taxa (L. Kruuk, in reparation).

Further modes of natural selection act to maintain e *Bombina* hybrid zone. There may be differential daptation to the puddle habitat: *B. bombina*'s slower evelopment rate (Nürnberger *et al.* 1995) will reduce fitness relative to *B. variegata* in temporary water odies where desiccation is often the main source of ortality (e.g. Skelly 1995). Hybrid unfitness will also eate a barrier to gene flow (Barton 1983, 1986) tween the two species. It is known that mor-ological abnormalities and embryonic mortality are gher in hybrids than in either parental form (Madej 65; Czaja 1980; Koteja 1984; all cited in Szymura & rton 1986; L. Kruuk & J. Gilchrist, unpublished ta). Finally, environmental factors such as the effect terrestrial habitat on adult survival require further vestigation, as does the role of competition between ults for mates and territories.

A combination of environmental and genetic factors erefore interact to maintain the genetic integrity of o populations despite their lack of reproductive lation. Differential adaptation of larvae to predation

pressures is one of a suite of traits that prevent the merging of *B. bombina* and *B. variegata*. We have shown that experimentation can give insight into the relevant interactions, and so clarify the mechanisms main-taining differentiation between species.

REFERENCES

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

Barandun, J. 1995 Reproductive ecology of *Bombina variegata* (*Amphibia*). Ph.D. thesis, University of Zürich.

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.

Bugter, R., MacCallum, C. J., Arntzen, J. W. & Barton, N. H. 1995 Analysis of the terrestrial habitat component in a *Bombina* hybrid zone using remote sensing. In *8th Meeting Societas Europaea Herpetologica*. Bonn.

Cheeson, J. 1978 Measuring preference in selective pre-dation. *Ecology* **59**, 211–215.

Chovanec, A. 1992 The influence of tadpole swimming behaviour on predation by dragonfly nymphs. *Amphibia–Reptilia* **13**, 341–349.

Cortwright, S. A. & Nelson, C. E. 1990 An examination of multiple factors affecting community structure in an aquatic amphibian community. *Oecologia* **83**, 123–131.

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, Poland.

Fitter, R. & Manuel, R. 1995 *Lakes, rivers, streams & ponds of Britain & north-west Europe.* Collins Photo Guide. London: HarperCollins.

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

Koteja, E. 1984 Mortality of embryos of the fire-bellied toad (*Bombina bombina*), yellow-bellied toad (*Bombina variegata*) and their hybrids. Praca magisterka, Zakald Anatomii Porownawczej Inst. Zoologii, Uniw. Jagiellonski, Krakow.

Lawler, S. P. 1989 Behavioural responses to predators and predation risk in four species of larval anurans. *Anim. Behav.* **38**, 1039–1047.

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

Madej, Z. 1965 Variations in the sacral region of the spine in *Bombina bombina* (Linnaeus, 1761) and *Bombina variegata* (Linnaeus, 1758) (Salientia, Discoglossidae). *Acta Biologica Cracov* **8**, 185–197.

Madej, Z. 1973 Ecology of European bellied toads (*Bombina* Oken, 1816). (In Polish with English summary.) *Przeglad Zoologiczny* **17**, 200–204.

Manly, B. F. J. 1974 A model for certain types of selection experiments. *Biometrics* **30**, 281–294.

Morin, P. J. 1983 Predation, competition and the com-position of larval anuran guilds. *Ecol. Monogr.* **53**, 119–138.

Morin, P. J. 1986 Interactions between intraspecific competition and predation in an amphibian predator–prey system. *Ecology* **67**, 713–720.

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

Petranka, J. W., Kats, L. B. & Sih, A. 1987 Predator–prey interactions among fish and larval amphibians: use of chemical cues to detect predatory fish. *Anim. Behav.* **35**, 420–425.

Richards, S. J. & Bull, C. M. 1990 Size-limited predation on tadpoles of three Australian frogs. *Copeia* **4**, 1041–1046.

Skelly, D. K. 1995 A behavioural trade-off and its consequences for the distribution of *Pseudacris* treefrog larvae. *Ecology* **76**, 150–164.

Szymura, J. M. 1993 Analysis of hybrid zones with *Bombina*. In *Hybrid zones and the evolutionary process*. (ed. R. G. Harrison), pp. 261–289. New York: Oxford University Press.

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

Werner, E. E. & Anholt, B. R. 1993 Ecological consequences of the trade-off between growth and mortality rates mediated by foraging activity. *Am. Nat.* **142**, 242–272.

Werner, E. E. & McPeek, M. A. 1994 Direct and indirect effects of predators on two anuran species along an environmental gradient. *Ecology* **75**, 1368–1382.

Wilbur, H. M. & Fauth, J. E. 1990 Experimental aquatic food webs: interactions between two predators and two prey. *Am. Nat.* **135**, 176–204.

Wilbur, H. M., Morin, P. J. & Harris, R. N. 1983 Salamander predation and the structure of experimental communities: anuran responses. *Ecology* **64**, 1423–1429.

Woodward, B. D. 1983 Predator–prey interactions and breeding-pond use of temporary-pond species in a desert anuran community. *Ecology* **64**, 1549–1555.