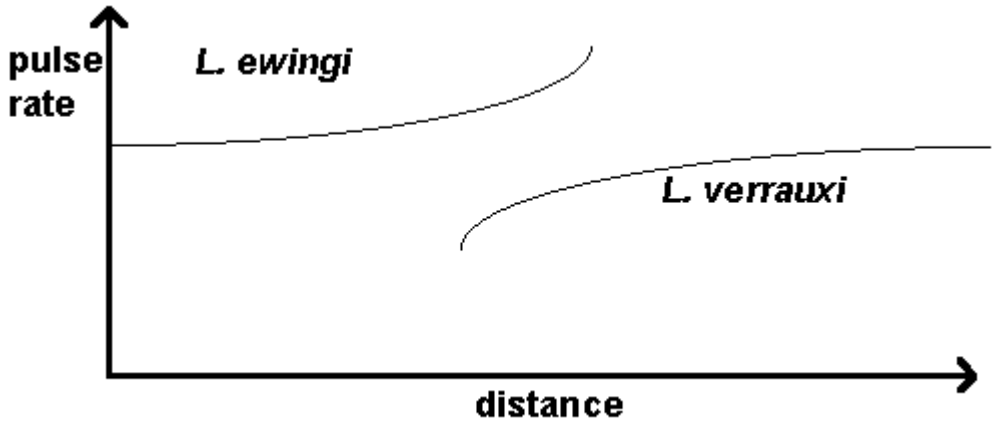
BIOL B242 Evolutionary and Population Genetics
Exam 2006
Model Answers

2. Write an essay on reinforcement and its relationship to sympatric speciation.

Reinforcement

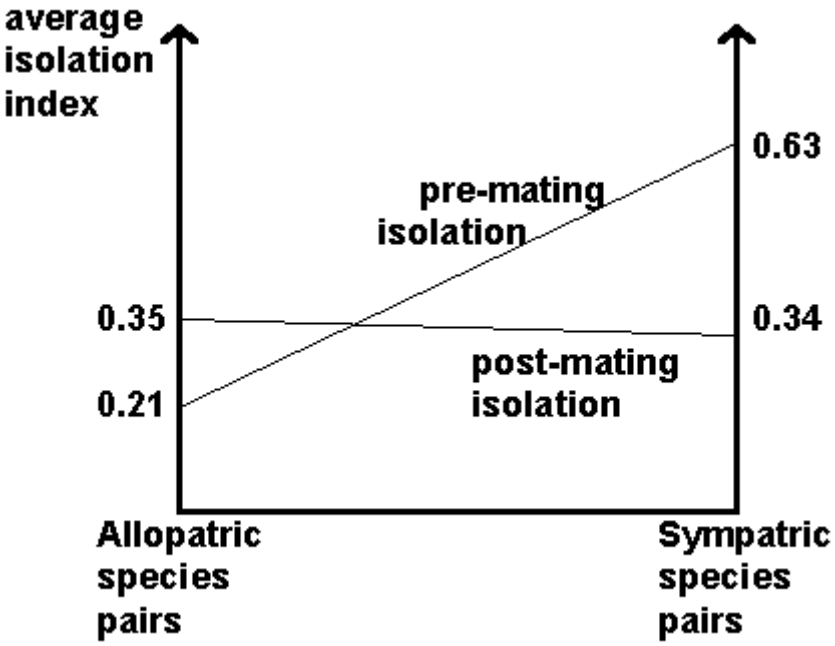
Suppose some adaptation has led to divergence which causes a reduction in fitness of hybrids. As we have seen from our discussion of intrinsic and extrinsic selection pressures, there are many ways in which this could happen. Now the two forms may either meet in *secondary contact* (if they were previously allopatric), or they may already be in contact (if divergence were parapatric or sympatric).

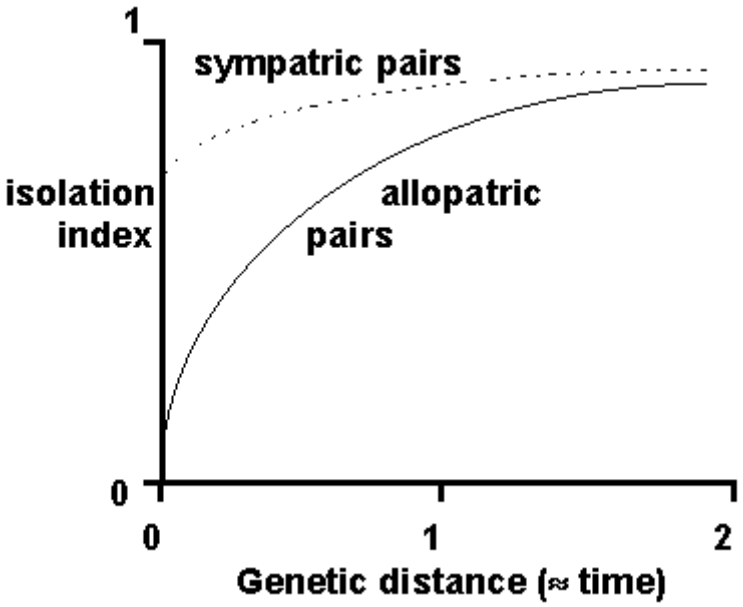
Regardless of how the contact happens, random mating between the divergent forms now creates unfit hybrids. Dobzhansky in 1940 postulated that, when the offspring produced will be of low fitness, hybridization should be opposed by natural selection: *assortative mating* is selected to reduce the time and effort put into mating and rearing of offspring. Essentially, this is the evolution *adaptive mate recognition*, and is now (perhaps confusingly) called *reinforcement*. *Reinforcement* is so called because populations can *reinforce* post-mating barriers by evolving pre-mating barriers. You will also recognize it as a kind of disruptive selection on mate choice, or a *good genes* mechanism, such as we have already dicussed in sexual selection. In Dobzhansky's view, reinforcement was able to take over after the evolution of some post-mating barriers, leading to a completion of speciation by the evolution of pre-mating barriers.



Evidence for reinforcement. It used to be thought that a good example of this was given by the Australian tree-frogs *Litoria ewingi* and *Litoria verrauxi.* The two frogs are largely allopatric, but they overlap in a parapatric contact zone. Hybrids are inviable. The pulse rate of the male call is important in mate recognition by females of these species. In allopatry, outside the contact zone, both species sing very similar songs. But inside the contact zone, pulse rates diverge, so that *L. ewingi* sings at a higher than normal rate, and *L. verrauxi* sings at a lower rate. Tape playback experiments have shown that the females preferentially choose the appropriate pulse rate of their own species.

In *Drosophila*, too, there is evidence for reinforcement. Coyne and Orr surveyed post-mating isolation in 171 pairs of sympatric and allopatric closely related divergent forms, usually species but sometimes subspecies. Post-mating isolation was measured crudely as the proportion of crosses in which hybrids were either sterile or inviable in crosses between two species A and B. An isolation index of 1.0 would indicate a complete lack of hybrid adult progeny. Pre-mating isolation between pairs was measured similarly, as the proportion of enclosures of males and females of opposite species that produced mating.





Allopatric pairs of *Drosophila* species had about the same average levels of postmating isolation (0.35), as sympatric pairs of *Drosophila* (0.34), and also similar average *genetic distances* between them. But average pre-mating isolation is much higher in sympatric species pairs (0.63) than in allopatric species pairs (0.21). This confirms a prediction of reinforcement: when exposed to one other, two species with some post-mating should evolve to avoid mating with each other.

If we plot pre-mating isolation versus genetic distance, a surrogate for time since divergence, we see that sympatric species evolve pre-mating isolation much faster than post-mating isolation. These two pieces of evidence are hard to explain in any other way: pre-mating isolation seems to have evolved to prevent hybridization between species that are incompatible.

Potential problems with reinforcement

For a long time, reinforcement was invoked as a likely factor in speciation. In the late 1980s, however, doubt was thrown on the whole idea. Roger Butlin has been a particularly influential

critic. Butlin distinguished two different situations in which selection against cross-mating might arise:

- ***Reinforcement***, in which hybrids are formed, and are selected against incompletely, so that there can be some gene flow between the forms, and ...
- ***Reproductive character displacement***, in which hybrids are either not formed, or, if formed, are so inviable or infertile that they never survive.

In the second case, it is easy to imagine selection against courtship and gametic wastage of the wrong species. Because no gene flow ever takes place between the parental species, there is also no leakage of preferential mate choice genes to the other species, which could cause further hybridization.

However, where hybridization is successful, and gene flow occurs, it is much harder to imagine how reinforcement will work. Hybridization occurs freely, and hybrids quickly form a swarm in the centre of the hybrid zone. Because there are no "pure species", individuals don't know how they should mate even if genes for reinforcement were available. And even if they were able to choose optimally, genes for mate choice would themselves flow across the boundary between the forms, and as much inappropriate as appropriate mate choices could result.

Coyne and Orr's data seem to provide better evidence of reinforcement than the *Litoria* example because the process seems continuous, and because some hybrids do survive in crosses (isolation indices < 1.0). However, it is clear that the isolation indices used by Coyne and Orr are crude (although the best data we have). Possibly few of the sympatric species hybrids ever survive in the wild, so that gene flow is zero, even though hybridization is present; if so, many cases of apparent reinforcement could actually be examples of reproductive character displacement.

Recently, Mohamed Noor studied two sibling species of *Drosophila* (*D. pseudoobscura* and *D. persimilis*) which do hybridize in the wild, and for which some hybrids are known to be fertile. He clearly showed that, in areas of overlap, mating was more assortative than in areas where one species was absent. We don't yet know how common reinforcement is, but it does again seem likely that it is important in speciation.

5. What is the importance of linkage disequilibrium and its significance in modern methods of finding genes that affect phenotypic traits?

What is linkage disequilibrium, and how do we measure it?

Expected gametic frequencies if two genes are independently inherited can be obtained from allelic frequencies in population:

Allele		A	a	
	allele freq.	p_A	$1-p_A$	
B	p_B	p_Ap_B	$(1-p_A)p_B$	
b	$1-p_B$	$p_A(1-p_B)$	$(1-p_A)(1-p_B)$	Sum = 1

Non-randomness of the gametic frequencies by means of a deviation from two locus equilibrium

D is the gametic disequilibrium coefficient, or measure of deviation from 2 locus equilibrium, as follows:

Gametic frequencies = random + deviation

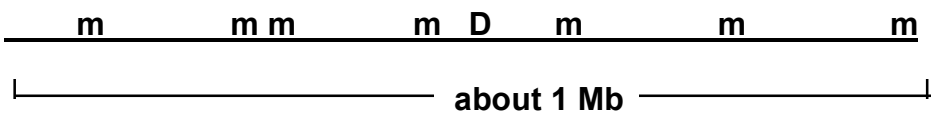
p_{AB}	=	p_Ap_B	**+ *D***
p_{Ab}	=	$p_A(1-p_B)$	**- *D***
p_{aB}	=	$(1-p_A)p_B$	**- *D***
p_{ab}	=	$(1-p_A)(1-p_B)$	**+ *D***

Obviously, the sum $p_{AB} + p_{Ab} + p_{aB} + p_{ab} = 1$

Gametic disequilibrium is usually referred to as linkage disequilibrium. This is somewhat confusing, because genes need not be linked to be in gametic disequilibrium (i.e. a significant value of ***D***). However, because recombination breaks down disequilibrium, tight linkage is a strong predictor of linkage disequilibrium, and can therefore be used to find genes.

2) **Linkage mapping** of Human loci when ***c*** = 0.01 or less

For example, disease locus **D** and marker loci **m**



In humans 1 million b.p. equivalent to ***c*** =1%
= 1 map unit (centimorgan)

Disequilibria significant between marker loci (like microsatellites)
and between markers and genetic disease loci ~ 1Mbp apart

Linkage disequilibria, are probably mainly originally due to genetic drift, but are preserved for long periods due to low recombination in the region of the genome where the changes have occurred. Linkage disequilibria useful for fine-scale gene mapping when loci are < 5 map units (c=0.05) apart. At these low rates of recombination, it becomes difficult to get sufficient families to do detailed recombination mapping. In contrast, disequilibrium mapping uses past recombination in the population to detect genes with pronounced phenotypic (e.g. a human disease) effects. The candidate region is the region which shows the highest linkage disequilibrium with the disease or phenotype being studied. This region can then be targeted for cloning, sequencing, or studying the regulatory properties.

The use of disequilibrium, D, can thus quickly narrow search for the "candidate loci". Successfully in a number recent studies of human genes, and also in stickleback body armour genes.

7. Discuss the evidence and theory behind the assertion that: "Selection is unimportant for most polymorphisms".

Classical vs. Balance Schools

Before the 1960s (in the days before there was any data about protein or DNA variation) there were two schools of population geneticists: the classical and balance schools. The classical school believed that polymorphisms, the existence of more than one allele in a population of genes, were rare. They argued that natural selection was a mainly a purifying force that removed any deleterious alleles that may arise or would drive any advantageous alleles to fixation. Therefore, they believed that individuals were homozygous for most loci. In contrast, the balance school believed that polymorphisms were common. Polymorphisms at the various loci were thought to be maintained by different forms of balancing selection that favoured heterozygostes over homozygotes. Both schools of thought agreed that natural selection was the force driving molecular evolution.

In the mid-1960s the technique of protein electrophoresis was discovered allowing investigation into the levels of enzyme polymorphism. The results showed that large amounts of genetic variation was present in natural populations appearing to vindicate the balance school's beliefs. The balance school held that these high levels of polymorphism were maintained by balancing selection. Others argued that maintaining these high levels of polymorphism at thousands of loci by balancing selection would be very costly. Summed over multiple loci this high genetic load would be large enough to drive populations to extinction!

The neutral theory of molecular evolution

However, the high levels of polymorphism can explained without encountering excessive genetic load simply by dropping the assumption that natural selection is the driving force of molecular evolution and instead allowing the majority of mutations fixed to be neutral and therefore have no effect on fitness. Two papers, by Kimura in 1968 and by King and Jukes in 1969 first proposed this neutral theory of evolution. Since then it has become one of the most important and controversial theories in evolutionary biology.

In his paper, Kimura made some simple calculations.

If μ = mutation rate per gene per generation, and N = effective population size

Number alleles in a diploid population = $2N$

Number of new mutations per generation = $2N\mu$

Most of the time a new neutral allele will be quickly lost from the population by genetic drift. But sometimes it will drift into the population and get fixed, that is, it will replace (or substitute) the original allele in the population.

The probability that the new allele will drift to fixation = 1/2N (this is equivalent to the probability of reaching into a bag containing 2N marbles and pulling out the only red marble).

The rate of substitution of an allele by a new allele = $2N\mu \times 1/2N = \mu$

Thus the rate of neutral molecular evolution is independent of population size and is simply equal to the neutral mutation rate.

The average time for a neutral mutation to drift to fixation is 4N generations. Therefore, while the rate of origin and fixation of new mutations (μ) is independent of population size, the rate of progress of the mutation through the population is proportional to the population size. Therefore, under the neutral theory, polymorphisms in a large population are simply a result of lots of neutral mutations arising and passing through the population at a slow rate such that at any one time there are several different alleles at a particular locus drifting through the population.

According to the neutralists, most mutations are either deleterious and are selectively removed, or are "effectively neutral", in which case there is a small probability that they are fixed. Natural selection is incorporated, but as a purifying force, removing deleterious mutations and with only a small role in fixing new mutations. As we have seen above, the probability of fixation of a neutral allele by drift is 1/2N. If this probability is bigger than the selection pressure, the influence of drift is greater than that of selection and the mutation is effectively neutral. So, the neutral theory does not argue that most mutations are completely neutral, but that any selection pressures are outweighed by the effects of drift.

On the other hand, according to the selectionists, mutations are fixed because they confer a selective advantage and that neutral mutations are rare.

Some predictions from the neutral theory

1. There is a constant rate, or molecular clock, of sequence evolution
2. There is an inverse relationship between the rate of substitution and the degree of functional constraint acting on a gene, such that functionally constrained genes or gene regions evolve at the lower rate and vice versa.

Functional constraints and the rate of substitution

According to the neutral theory most mutations are deleterious and the rest are neutral (advantageous mutations are very rare). However, genes will differ in the proportion of mutations that are deleterious. The higher the functional constraint on the gene, the greater is the strength of negative selection removing mutations. In a gene with high functional constraints the vast proportion of mutants will be deleterious and be removed by selection, leaving only a small fraction of neutral mutations which will result in a low rate of substitution. In a less constrained gene a larger fraction of the mutations will be neutral leading to a higher substitution rate.

Variation in rates between and within genes

Substitution rates in non-coding regions – pseudogenes, introns.

Synonymous vs. non-synonymous mutations rates

Testing the neutrality of mutations using d_N/d_S:

1) Sequence copies of the gene of interest from a variety of species.
2) Construct a phylogeny of the species using the sequence or other data.
3) Identify synonymous and non-synonymous mutations.
4) Calculate the average synonymous rate of subsititution, d_S, the average non-synonymous rate of substitution, d_S, and the ratio, $\omega = d_N/d_S$.

We assume that synonymous mutations are neutral. As we have seen, due to functional constraints, in most genes $d_N < d_S$, and $\omega < 1$.
If $d_N > d_S$, $\omega > 1$. The coding changes are occurring more rapidly than silent changes. This is indicative of positive selection to change the amino acid sequence.

Positive selection – evidence against the neutral model?

Examples – mutation rates within the major histocompatibility complex and HIV envelop proteins.

However, the procedure described above for detecting positive selection is insensitive. Another problem is that the procedure calculates a single value of ω for one gene. It is possible that only a few parts of a protein are under strong positive selection. If this is the case, averaging over the whole gene will mean that the $\omega > 1$ signal from the bits under positive selection will be swamped by the $\omega < 1$ signal coming from the majority of the gene.

Some improvements have been made to detecting positive selection, many of them coming from Ziheng Yang here at UCL.

So who is correct, the neutralists or the selectionists?

It seems that both genetic drift and natural selection determine the evolution of mutations. Neutralists are probably correct in that most mutations are neutral, especially in non-coding DNA and synonymous sites. However, evidence of natural selection is sometimes evident at non-synonymous sites when molecular evolution over a short time period is examined.

BIOL B242 Evolutionary Genetics 2006 Marks

Question 9

		BB	*BC*	*BD*	*CC*	*CD*	*DD*	*Sum*	
	Obs	150	104	424	18	129	198	**1023**	
		freq							
a	B	**p=0.404692**							1
	C	**q=0.131476**							1
	D	**r=0.463832**							1
	Sum	1							
		Null hypothesis: Hardy-Weinberg equilibrium							
		p^2	2pq	2pr	q^2	2qr	r^2		
b	Exp freq	0.163776	0.106415	**0.375418**	0.017286	0.121966	0.21514	**1**	2
	E	167.5425	108.8622	384.0528	17.68353	124.7708	220.0882	**1023**	
c	(O-E)^2/E	1.836788	0.217162	4.155106	0.005664	0.143354	2.216791	**8.574863**	

Chi sq. val is between 7.81 (P=0.05) and 11.34 (P=0.01) on table, therefore **0.01<P<0.05. So reject null!** 4

d There should be three degrees of freedom, because there are 6 classes of genotypes, but one degree of freedom is taken off for each independent parameter estimated from the data, including 1 df for getting the total numbers, and 2 df for the two independent gene frequencies (the third is obtained by subtraction from the total). So 6-1-2 = **3** df remaining 2

Relative fitnesses:

e	O/E	0.895295	0.955336	**1.104015**	1.017896	1.033896	0.899639	3
	Fitnesses relative							
f	to BD:	**0.810945**	**0.865329**	**1.00000**	**0.921995**	**0.936487**	**0.81488**	3

g i) Chromosomal inversions may be heterotic, either positively or negatively. In this case, the suggestion is heterozygote advantage, because the BD heterozygote is more common than expected under Hardy-Weinberg.
ii) Another possibility is that B-containing genotypes like to mate disassortively with D-containing genotypes, and therefore heterozygotes become more common than expected.

To test (i), rear crosses between BD and BD (or other genotypes that can produce BD offspring) and test for the frequency of BD relative to other genotypes in the progeny against the Mendelian expectation. If survival is enhanced, we would expect to see BD genotypes surviving better than other genotypes. However, laboratory conditions may be too benign, if the selection occurs in the wild. In the case of no significant difference in this first test, transplant progeny out into the wild, and see whether recapture rates differ between genotypes as an estimate of relative fitness in natural environments

To test (ii), test the mating behaviour of the genotypes to see if BB x DD crosses occur experimentally with greater frequency than BB x BB or DD x DD. This could also be done on many other genotypes, to compare the mating preferences of both males and females for particular genotypes 4

h *Drosophila pseudoobscura* is another fly that has many inversion polymorphisms. These are known to be related to the environment, such that particular inversions do better at high or low elevations or in different seasons. There is also strong heterozygote advantage in Drosophila pseudoobscura inversion polymorphisms.

Inversions can cause problems for heterozygotes because some crossing over forms will produce duplications, deletions and chromosome breaks. However, in Drosophila the male undergoes no crossing over, and females selectively shunt duplication and deletion products into polar bodies, leading to mainly fertile eggs. Thus the potential deleterious effects are avoided, and are outweighed by the strong heterozygote advantage.

(Any other suitable examples also allowed) 4

TOTAL 25