Chapter 2

Measuring Genetic Variability in Natural Populations by Allozyme Electrophoresis

James M. Bader

Department of Biology
Case Western Reserve University
Cleveland, Ohio 44106-7080
jxb14@po.cwru.edu

Jim received his B.S. from the University of Notre Dame and M.S. from Case Western Reserve University where he is the Manager of the Graduate and Undergraduate Biology Teaching Laboratories and the Assistant Director of the Center for Biology Education. He teaches courses in Introductory Biology, Microbiology, and Aquatic Biology.

Reprinted From: Bader, J. M. 1998. Measuring genetic variability in natural populations by allozyme electrophoresis. Pages 25-42, *in* Tested studies for laboratory teaching, Volume 19 (S. J. Karcher, Editor). Proceedings of the 19th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 365 pages.

Contents

Introduction

This laboratory is modeled after recent studies on zooplankton and zebra mussel genetic variability and concentrates on three local populations of bark beetles (*Ips pini)*. Investigations using allozyme analysis continue to provide valuable population-level information on the relatedness of populations and the amount of genetic variability within populations and between populations. This experiment uses the same approach that a typical research study might use. A preliminary screening of approximately 20 individuals of bark beetles at 20 different enzyme loci in each of the three populations is designed to identify polymorphic loci (enzyme loci that show variability in their electrophoretic pattern). Most of the loci will be monomorphic (showing no variability within or between populations) and are eliminated from further study. A more intense screening of the polymorphic loci (as many as 100 individuals from each population) will generate enough data to calculate measures of genetic identity and genetic distance.

In our course, we have 4 sections with 25 students in each section. The sections early in the week do the preliminary screens to identify polymorphic loci and the later sections screen as many individuals at these polymorphic loci as possible. All data are entered into a spreadsheet which grows as the week progresses; at the end of the week everyone gets a copy of the completed spreadsheet.

Materials

Organisms to be analyzed - locally collected populations of insects, zooplankton, plants, etc. They must be live or fresh frozen. Alcohol preserved specimens will not work. Any organism or any piece of an organism that contains living tissue that is small enough to grind is suitable.
3 MM filter paper
Tris glycine buffer (10X): 30 g Trizma base, 144 g Glycine. Make up to 1 liter. Dilute 1:9 with water.
Horizontal electrophoresis chambers - virtually any horizontal mini-gel unit can be adapted to run the cellulose acetate gels. Depending on the style, it may be possible to run two gels on the same unit at the same time.
Power supplies
60°C water bath for holding melted agar
Agar: 4 g Agar, 250 ml water. Heat to boiling to dissolve agar.
Light box
Documentation system : We use a Polaroid camera and 667 film to take pictures of each gel
Computer(s) with a spreadsheet program (such as Excel) is essential.
Sample well plates : Helena Laboratories Cat #4096, 2 for $125
Applicators : Helena Laboratories Cat #4090, $360
Cellulose acetate gels : Helena Laboratories Cat #3033, 100 gels for $350
Stain mixtures for each of the enzymes (recipes given in Appendix). Prepare ahead of time, freeze, and thaw just before use. Light and heat sensitive reagents are added just before use.

Notes for the Instructor

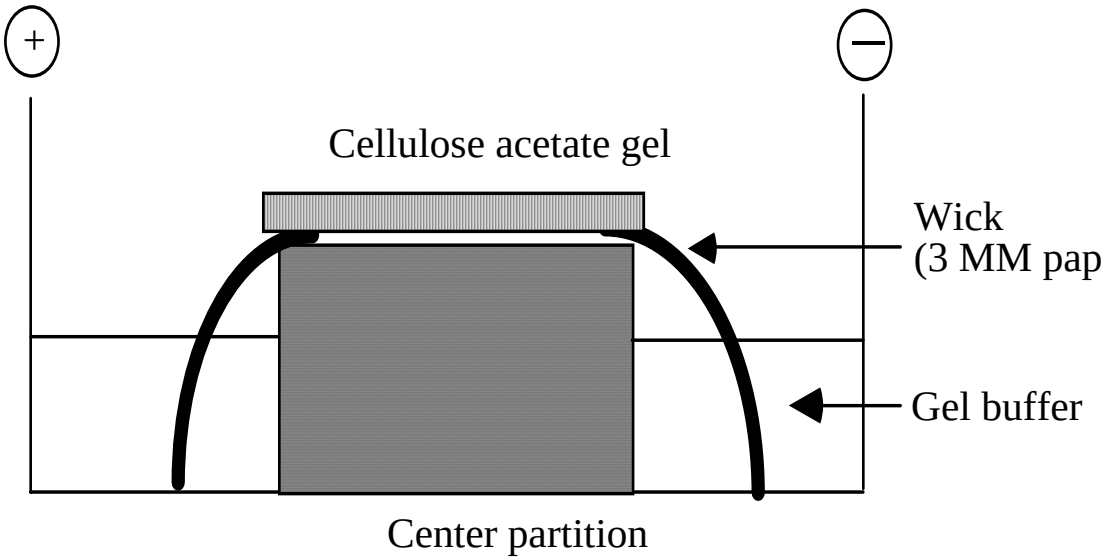
Before attempting this laboratory, I strongly recommend obtaining a copy of Hebert and Beaton (1989). The general methodology, stain recipes, and valuable background information for cellulose acetate electrophoresis are included in this book. A large number of more general sources dealing with allozyme analysis and other molecular methods are also available, such as Avise (1994) and Hillis *et al*. (1996).

The initial preparation time for this experiment is quite extensive and the cost of materials and chemicals may be prohibitive for some programs, but after the experiment has been completed once, future preparation time and expense should be much less. Mixing stock solutions and then compiling the stain recipes for 20 different enzymes can be quite time consuming, but once the stocks are made, many can be frozen for future use, and thawed when needed.

One of the advantages of this procedure is that virtually any organism or part thereof can be used as the enzyme source. We have used *Drosophila*, copepods, and cladocerans in the past. We have not tried plant material, but as long as the tissue can be homogenized to burst the cells and release the enzymes, it should work. We have always used living material; fresh frozen tissue works, but preserved material is unreliable.

Another advantage of this experiment is the large number of enzyme stain recipes that have been extensively field tested and are readily available. Recipes for over 100 enzymes have been published in various sources including Harris and Hopkinson (1976), Richardson *et al.* (1986), Hebert and Beaton (1989) and Hillis *et al*. (1996). To keep things simple, we try to use only enzymes with the same buffer system (Tris-glycine). We like to use monomeric and dimeric enzymes because gel interpretation is more straightforward, although we have been known to be adventuresome and try tetrameric enzymes as well.

Just about any horizontal electrophoresis apparatus can be modified to run cellulose acetate gels. We use 3 different styles of gel boxes in our laboratory with no detectable effects on the resolution of enzymes. We use 4-5 layers of filter paper to act as wicks and to support the cellulose acetate gel so that it does not touch the center partition. The critical factor is to be sure that the gel is in contact with the buffer only at either end of the gel so that the current must pass through the gel during electrophoresis. A crude diagram of a typical setup is given below.



The biggest problem with this laboratory as we currently run it is that the sample application phase is the rate limiting step. It is critical that a small yet concentrated sample be applied in a

perfectly straight line. If this does not occur, it is impossible to determine whether you are dealing with two distinct allozymes or just two samples that were not applied at the same level on the gel. Ideally we would like to have one applicator per pair of students, but they are prohibitively expensive as purchased through Helena Laboratories (Beaumont, TX). An alternative method of application (which is less expensive and more rapid) but just as reliable would be an excellent addition to this exercise. We have tried some work with templates with limited success, but we plan to keep trying to solve this problem. In fact, all of the material supplied by Helena Laboratories is expensive. This includes the cellulose acetate gels, the sample applicators, and the sample wells. I have made a minimal effort trying to identify other sources, but I plan to increase my efforts and expand our supplies.

Student Outline

Introduction

Evolution is a necessary condition for the survival of a species. All organisms face stresses, whether it be from natural sources such as competition or predation, or from human impact such as pollution, habitat destruction, or exploitation of commercially valuable species.

As a result of these constantly changing stresses, the environment that most organisms live in is highly variable. If all the individuals within a population were genetically similar, it is likely that they would all be equally susceptible to biological forces, human intervention, and natural phenomena such as drought, floods, or drastic fluctuations in temperature. Individuals within populations must be able to adapt to these changing environments; this adaptive potential is the basis for natural selection and is largely genetically determined. In order for a species to survive, some populations of that species must exhibit sufficient genetic variability so that individuals within that population are able to adapt to the changing environment.

Genetic variability threatened : The Great Lakes fishery

A prime example of widespread loss of genetic diversity can be found in the Great Lakes fishery populations. Overexploitation combined with invasion by exotic species has placed severe stress on native fish species. Many commercially valuable species have been eliminated or reduced to a few isolated natural populations.

In a general sense, all organisms in the Great Lakes face the same forces which determine adaptive potential and the rate of evolution of a species. Selection, gene flow, and genetic drift have a significant impact on the genetic variability in fish populations in the Great Lakes.

- Although selection is a natural process, many of the forces that are sources of stress for the Great Lakes fish communities are of human origin. Size selective fishing gear, habitat elimination, alteration of prey species and pollution stresses are forms of selection that can favor particular individuals and reduce the overall genetic diversity of the population.
- Gene flow is the exchange of alleles between neighboring populations. Geography plays an important role in this process since the fewer barriers and the closer two populations are geographically, the more likely they are to exchange individuals and their alleles. The greater the rate of gene flow between populations, the more homogeneous those populations will become. Maintaining some degree of isolation between populations would reduce the amount of gene flow and ultimately would result in more genetic diversity in the species as a whole. Therefore any human activities which would affect the geographical distribution and movement of populations would have a direct impact on the amount of gene flow. In some situations in the Great Lakes, human activities have restricted the movement of individuals

through the construction of artificial barriers. In other situations, human interventions have removed natural barriers and provided the opportunity for movement of individuals. In addition, any activity which increases or decreases the population densities also effects the distance between populations.

• Genetic drift is most pronounced in small populations. If some individuals are removed from this small population, it is quite likely that unique alleles will be lost as well. In addition, small population size results in inbreeding which also reduces the amount of genetic variability in a population.

The decline of the Great Lakes fishery forced managers to develop practices which were designed to stock the lakes with fish that would eventually begin to reproduce and replenish the species. These stocked individuals were cultivated in hatcheries by obtaining eggs and milt from a small number of native fish, fertilizing them in the laboratory and raising the fry until they were of sufficient length to release. A large number of individuals can be reared in this manner; however, the genetic diversity in the stocked fish is not as great as that of the natural population since usually only a few males and females were used for fertilization.

Also, the presence of a large hatchery reared population poses potential problems for the native species. One concern, especially with salmonid species, is that a variety of locally adapted populations will be replaced with a smaller number of genetically homogeneous populations which would limit the adaptive potential of the species as a whole. A second concern is that hatchery raised populations may actually reduce the size of the natural population by increasing competition or introducing diseases the native populations have never encountered. Reduction in population size of the reproducing natural population would lead to inbreeding which is recognized to reduce genetic variability.

For these reasons, genetic variability in both natural and hatchery reared populations has become a significant concern to state and federal wildlife officials responsible for managing the Great Lakes fisheries. There is currently a growing consensus that determining the genetic composition of the fish populations of the Great Lakes should be a high priority.

Measurement of genetic variability

Traditional Mendelian methods of making crosses and scoring the phenotypes of the offspring in one or more generations is insufficient for a detailed estimate of genetic variability. The process is restricted to phenotypic characters primarily, which are limited in number. The process is too time consuming to wait for future generations in many species. The process does not always yield precise information on genotype (homozygous dominant *vs* heterozygote). There are too many gene loci in most organisms for this process to yield reliable estimates of the genetic variability.

These limitations can be overcome by using the techniques of molecular genetics. We know that nucleotide sequences of the DNA of a gene are transcribed into messenger RNA which are translated into a sequence of amino acids to form a polypeptide. If we could obtain an unbiased representative sample of all the structural gene loci, we could use data from these samples to estimate the amount of genetic variability. Ideally, it would be best to sequence the DNA in this subsample of gene loci. However, DNA sequencing also is a time consuming process, and some of these genes could be quite large. We are impatient, so we need a rapid procedure which would allow us to screen a sufficient subsample of a population at a large number of gene loci.

Instead of looking directly at DNA sequences, we could look at the protein product of those sequences. Theoretically, the amino acid sequence of the polypeptide product should reflect the DNA sequence that coded for it. Unfortunately, sequencing polypeptides that may be hundreds of amino

acids in length is also time consuming. But, do not lose heart. A combination of electrophoresis on cellulose acetate gels and enzyme assays allow us to screen as many as 100 individuals at 2-3 loci in as little as 30 minutes.

Allozyme analysis

The procedure is straightforward. Tissue samples are homogenized to release enzymes from the cells. The homogenates are placed into wells of the gel. The proteins are separated by electrophoresis in a cellulose acetate matrix. Separation can be achieved in as little as 20 minutes in this medium, which is far superior to the 2 hours typically required for starch gels.

After electrophoresis the gel is treated with a substrate that is specific for the enzyme of interest. A positive reaction will yield a dark spot where the protein migrated. This procedure takes advantage of the fact that all organisms will produce multiple forms of some enzymes (called allozymes). Each of these allozymes, where present, is known to have a slightly different amino acid sequence and is thought to be the product of a unique allele. In diploid organisms, there will be a combination of two alleles at each locus. These are designated as F (fast) and S (slow) or as 1 and 2 to distinguish between them. Therefore, the genotype at a gene locus coding for an enzyme can be inferred for each individual in the sample from the number and position of the spots observed on the gels.

If the enzyme is a monomer (the complete enzyme consists of only one polypeptide) the electrophoretic pattern shown in Figure 2.1a would result for homozygotes and heterozygotes. An enzyme whose quaternary structure is a dimer (the complete enzyme consists of two polypeptides) would show the electrophoretic pattern demonstrated in Figure 2.1b. In an individual heterozygous at a dimeric enzyme locus for slow (1) and fast (2) alleles, three enzyme associations will be evident (11, 12, 22). The heterozygote will show up twice as dark because both forms of each polypeptide will randomly associate, meaning that there is only one combination that will yield 11 or 22, but there are two combinations which will yield 12. The electrophoretic pattern for a tetrameric enzyme is given in Figure 2.1c.

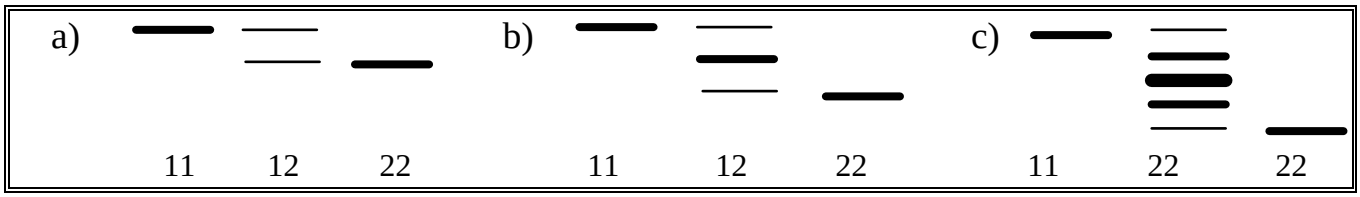


Figure 2.1. Electrophoretic pattern for a) a monomeric enzyme, b) a dimeric enzyme and c) a tetrameric enzyme. Homozygotes are represented by 11, 22 and heterozygotes are represented by 12.

By examining the electrophoretic pattern for a number of enzymes (loci), we can began to determine the genetic variability in the population and to compare that variability with other populations. Experience has shown that is necessary to screen at least 20 enzyme loci to get a representative sample of the variability in populations. Fewer loci would tend to underestimate the variability and more loci does not seem to change the estimate significantly.

Calculation of genetic variation

One measure of genetic variation is the proportion of polymorphic loci or polymorphism. This is simply the proportion of loci examined that show evidence of more than one allele. In Table 2.1, the results of an allozyme analysis of a marine worm are summarized. As can be seen from the table, 11 of the 39 loci examined (0.282) were determined to be polymorphic. You should also note that some loci, even though they show more than one allele, are designated invariant (as opposed to polymorphic) because the most common allele has a frequency greater than 0.95. This cutoff point of 0.95 has been arbitrarily chosen to offset the problems of small sample size, but it has no real biological basis. A value of 0.98 could just as easily be chosen, and for some data sets, this would significantly alter the polymorphism calculation. For this reason, polymorphism is an inadequate descriptor of genetic variability.

A more suitable measure of genetic variation in a population is the average frequency of heterozygous individuals per locus (H) of the population. This is calculated by first determining the frequency of heterozygotes at each locus and then averaging these frequencies over all loci. In the data set listed in Table 1, the observed frequency of heterozygotes is listed for each locus. These are then averaged over all loci to give a value of H = 0.072 or 7.2%. This value may be determined from the raw data as just described, or it may be estimated from the frequency data where :

$$h_j = 1 - \Sigma\, q_k^2$$

and h_j is the heterozygosity estimated at locus *j* and q_k is the frequency of the *kth* allele at that locus. To determine the estimated heterogeneity for the whole population, average the h_j values for all loci.

What does this value mean? How does it compare to other populations and other species? Table 2.2 summarizes a partial list of the results of electrophoretic surveys obtained for 69 plant species and 125 animal species in which a sufficient number of loci have been examined. Since the time this list was compiled, many more species have been screened. In general, it appears that invertebrates have more genetic variability than vertebrates, although there are certainly numerous exceptions. Plants that are self-fertilizing show less genetic variability that those that are not self fertilizing.

This variability as measured by the average heterozygosity (H) represents the adaptive potential of the species and is an extremely powerful evolutionary force. For instance, humans appear to have an average heterozygosity of 6.7%. If we assume that there are 30,000 structural genes in the human genome (a very conservative estimate), a person would be heterozygous at (30,000 x 0.067) = 2010 loci. An individual heterozygous at *n* loci theoretically could produce 2^n different gametes. In our example, this gives the possibility of producing 10^{605} different gametes, an impossible number when one considers that the total number of neutrons and protons in the universe is estimated at 10^{76}. Although not all of these different combinations are actually produced in the gametes, the number of possibilities is certainly staggering. Somewhere in that genetic recombination, individuals should be produced that are well suited to their ever changing environment.

Table 2.1. Allelic frequencies at 27 variable loci in 120 individuals of the marine worm *Phoronopsis viridis*. The numbers used to represent alleles (*1, 2, 3,* etc.) indicate increasing mobility, in an electric field, of the proteins encoded by the alleles.

	Frequency of Allele						Heterozygosity		Poly-morphic? (<0.95)
Gene locus	1	2	3	4	5	6	Observed	Expected	
Acph-1	0.995	0.005					0.010	0.010	No
Acph-2	0.009	0.066	0.882	0.014	0.005	0.024	0.160	0.217	Yes
Adk-1	0.472	0.528					0.224	0.496	Yes
Est-2	0.008	0.992					0.017	0.017	No
Est-3	0.076	0.924					0.151	0.140	Yes
Est-5	0.483	0.396	0.122				0.443	0.596	Yes
Est-6	0.010	0.979	0.012				0.025	0.041	No
Est-7	0.010	0.990					0.021	0.021	No
Fum	0.986	0.014					0.028	0.028	No
Gpd	0.005	0.995					0.010	0.010	No
G3pd	0.040	0.915	0.017	0.011	0.011	0.006	0.159	0.161	Yes
G6pd	0.043	0.900	0.057				0.130	0.185	Yes
Hk-1	0.996	0.004					0.008	0.008	No
Hk-2	0.005	0.978	0.016				0.043	0.043	No
Idh	0.992	0.008					0.017	0.017	No
Lap-3	0.038	0.962					0.077	0.017	No
Lap-4	0.014	0.986					0.028	0.027	No
Lap-5	0.004	0.551	0.326	0.119			0.542	0.576	Yes
Mdh	0.008	0.987	0.004				0.025	0.025	No
Me-2	0.979	0.021					0.042	0.041	No
Me-3	0.017	0.824	0.159				0.125	0.296	Yes
Odh-1	0.992	0.008					0.017	0.017	No
Pgi	0.995	0.005					0.010	0.010	No
Pgm-1	0.159	0.827	0.013				0.221	0.290	Yes
Pgm-3	0.038	0.874	0.071	0.017			0.185	0.229	Yes
Tpi-1	0.929	0.071					0.141	0.133	Yes
Tpi-2	0.008	0.004	0.962	0.013	0.013		0.076	0.074	No
Averages (including 12 invariant loci)									
Heterozygosity							0.072	0.094	
Polymorphism						11/39 = 0.282			

After F.J. Ayala et al. 1974. *Biochemical Genetics 18*:413.

Table 2.2. Genetic variation in natural populations of some major groups of animals and plants.

Organisms	Number of species	Ave number of loci per species	Average polymorphism	Average heterozygosity
Invertebrates				
Drosophila	28	24	0.529	0.150
Wasps	6	15	0.243	0.062
Other insects	4	18	0.531	0.151
Marine invertebrates	14	23	0.439	0.124
Land snails	5	18	0.437	0.150
Vertebrates				
Fishes	14	21	0.306	0.078
Amphibians	11	22	0.336	0.082
Reptiles	9	21	0.231	0.047
Birds	4	19	0.145	0.042
Mammals	30	28	0.206	0.051
Plants				
Self-pollinating	33	14	0.179	0.058
Outcrossing	36	11	0.511	0.185
Overall averages				
Invertebrates	57	22	0.469	0.134
Vertebrates	68	24	0.247	0.060
Plants	69	13	0.345	0.121

After Ayala, 1984. Modern Genetics

A third method of estimating the genetic variability in a population involves the Hardy-Weinberg law. The Hardy-Weinberg law says that under certain conditions, allele frequencies do not change from generation to generation. Even though the alleles may not be expressed (as determined by Mendel's laws of inheritance) they are still present in the population in the same frequency.

In order for the Hardy-Weinberg law to be valid, specific assumptions must be satisfied. However, these conditions are rarely met in nature. However, this does not invalidate the usefulness of the Hardy-Weinberg calculation. If allelic frequencies deviate significantly from Hardy-Weinberg predictions, then we can assume that some evolutionary forces are at work.

Testing the hypothesis of Hardy-Weinberg equilibrium

Deviation from Hardy-Weinberg equilibrium would suggest that one or more of the necessary assumptions is not valid. In some cases, it may be possible to rule out potential sources of deviation, leaving a logical mechanism that may account for the observed deviation.

Assume that we obtain the following results after screening 126 individuals at the *bio* locus :

bio^1/bio^1	84
bio^1/bio^2	28
$\underline{bio^2/bio^2}$	14
Total	126

To determine whether the observed numbers occur with the frequencies predicted by the Hardy-Weinberg law, first calculate *p*, the frequency of the most common allele (1).

$$p = \frac{(84 \times 2) + 28}{252} = \frac{196}{252} = 0.78$$

This means that the frequency of the (2) allele is assumed to be 1 - p = 1 - 0.78 = 0.22. Using these allele frequencies, we calculate the expected number of the three genotypes given a total of 126 observations.

Table 2.3. Expected frequencies if in Hardy-Weinberg equilibrium.

Genotype	Frequency	Number
bio^1/bio^1	$p^2 = 0.60$	76
bio^1/bio^2	$2pq = 0.35$	44
bio^2/bio^2	$q^2 = 0.05$	6

Table 2.4. Chi-square calculation to test for Hardy -Weinberg equilibrium at the *bio* locus.

	bio^1/bio^1	bio^1/bio^2	bio^2/bio^2
Observed	84	28	14
Expected	76	44	6
(Observed-expected)	8	16	8
$(\text{Observed-expected})^2$	64	256	64
$(\text{Observed-expected})^2/\text{exp}$	0.84	5.82	10.67
$\chi^2 = \Sigma \frac{(\text{obs - exp})^2}{\text{exp}}$	17.33		

To determine whether these expected values are significantly different from observed ratios, we perform a chi-square goodness of fit test as described in Table 2.4.

The number of degrees of freedom in this case is one, not two as might have been expected. Remember that the number of degrees of freedom is the number of classes minus the number of independent values obtained from the data that are used for calculating the expected numbers. In this example, if we know the number of homozygous dominant individuals, we can use that information to determine p. This value is then used to calculate q as well as the expected number of individuals in each class. Therefore the total number of classes is in essence two (p and q) and the number of degrees of freedom is one. With one degree of freedom and a probability level of .05, the critical chi-square value is 3.84. Our value of 17.33 is greater than the critical value so we reject the hypothesis that the population is in Hardy-Weinberg equilibrium at the *bio* locus.

Interpopulation analysis

The preceding calculations (percent polymorphism, heterozygosity, and Hardy-Weinberg equilibrium) are suitable for determining the genetic variability within a single population. More often, we wish to compare local populations within a species to estimate the amount of gene flow and genetic drift between populations. These forces along with natural selection lead to the creation of new species. It is sometimes necessary to determine the amount of genetic differentiation taking place during speciation.

Estimates of genetic differentiation between two populations can be obtained by using allozyme data. Once again it is assumed that the genes encoding these enzymes are a random subsample of all the structural genes in the population. The results obtained from the study of a suitable sample of gene loci can be extrapolated to the whole genome.

A number of statistical methods have been formulated in an attempt to quantify the degree of genetic differentiation between populations. Two such statistical calculations formulated by Nei

(1972) are (1) genetic identity *I*, which estimates the proportion of genes that are identical in the two populations being compared, and (2) genetic distance *D*, which estimates the accumulated number of gene differences per locus that have occurred over evolutionary time. If the rate of gene substitutions per year is constant, the genetic distance is linearly related to evolutionary time.

One advantage to these measures of genetic differentiation is that they apply to any population, whether they be haploid, diploid, tetraploid, or selfing. This is because the definitions of *I* and *D* depend solely on gene frequencies rather than on genotype frequencies as in the Hardy-Weinberg equation.

The formulation for genetic identity is derived as follows : Assume that you have two populations, X and Y. For a given locus, let x_i and y_i be the frequencies of the *i*th alleles in populations X and Y respectively. Then :

$$I = \Sigma x_i y_i \ / (\Sigma x_i^2 \Sigma y_i^2)^{0.5}$$

For multiple loci, the overall similarity or identity is :

$$I = J_{XY} / (J_X J_Y)^{0.5}$$

where J_X, J_Y and J_{XY} are the arithmetic means across all loci of Σx_i^2, Σy_i^2, and $\Sigma x_i y_i$ respectively. The formulation for genetic distance, *D*, is given by :

$$D = -\ln I$$

The values for *I* may range from 0 (no alleles in common) to 1 (the same alleles, and in the same frequencies are found in both populations). Genetic distance, *D*, may range from a value of 0 (no allelic substitutions) to infinity. Generally speaking, closely related populations tend to show *I* values > 0.9 and *D* values < 0.1. Divergent populations or separate species tend to have genetic identities (*I*) < 0.8 and genetic distances (*D*) > 0.2.

Proper interpretation of these values may even be used to determine whether populations are undergoing genetic differentiation during speciation. Local populations which are genetically similar have *I* values in excess of 0.97. As speciation takes place, the *I* values decrease and the *D* values increase.

Another measure of genetic distance has been proposed by Rogers (1972). For a given locus with *m* alleles, let x_i and y_i be the frequencies of the *i*th allele in populations X and Y just as we assumed above. Rogers' D is defined as :

$$D = [0.5\Sigma(x_i - y_i)^2]^{0.5}$$

where the summation is over all alleles. When data from more than one gene is considered, the arithmetic mean of such values across loci provides the overall genetic distance estimate. Rogers' D values are between zero and one. Rogers' similarity value is given by :

$$S = 1 - D$$

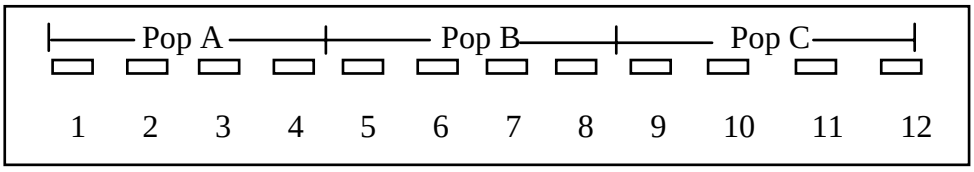
Procedure

This procedure will work for any organism that can be homogenized in whole or in part. If the organism is small (0.5 - 5 mm) the entire animal or plant may be used. For larger organisms, samples of tissues are homogenized and analyzed. Liver, muscle, and eye are commonly used in vertebrates.

We will be assessing the genetic variability in three populations of the pine engraver beetle (*Ips pini).* As the name implies, pine engravers breed in the inner bark of pine trees. This species of bark beetle is widespread throughout North America with populations breeding in dying or freshly dead trees. Most population comparisons have focused on chemical differences in the male pheromone ipsdienol which attracts females and other males resulting in groups of beetles occupying the same log. Eastern and western populations have been shown to produce chemically distinct forms of ipsdienol (Mary Reid, personal communication), and populations in British Columbia, separated by only 400 km, also produce distinct forms (Miller *et al.* 1989).

Two of the three populations we will be using are separated by approximately 25 km and are located in separate watersheds in the Kananaskis area about 70 km west of Calgary. The third population comes from near Whitecourt, northwest of Edmonton and about 400 km from Kananaskis.

1. Each pair of students will process 12 individuals from each of the populations. You will determine the genotypes of each of those 12 individuals at a single enzyme locus. Other groups will be assigned different individuals and different loci. We will try to screen at least 50 individuals at 12 different loci in all populations. We will then pool the class data for analysis.
2. If the organism to be used is small enough, place an individual directly into the sample well and add 10 µl of water. Use a spatula to grind the animals once they are in the wells. Pine engraver beetles are just a little too large to be ground directly in the wells, so they were ground in 50 µl of water in a microfuge tube and the homogenates frozen until used. In order to compare the three populations more easily, load 10 µl of your homogenate from each of four individuals from each population as shown.



3. After all the animals have been homogenized, remove a cellulose acetate gel from the beaker where it has been soaking in Tris-glycine buffer for at least 20 minutes. Blot the gel dry between two pieces of 3MM filter paper. It is essential that all moisture is removed from the gel surface.
4. Notice that the plate has a dull upper side (which is the acetate coating) whereas the back is shiny mylar. Place the plate mylar side down on the aligning base. The applicator is designed to fit into the sample wells. Each lane has what looks like two small staples mounted in plastic. They are spaced closely enough together that a thin film of sample is picked up with each application. Insert the applicator into the slots on the sample well plate and depress the lever 3-4 times to pick up homogenate on the teeth of the applicator. Transfer the applicator to the aligning base and apply the sample to the acetate gel by gently depressing and holding the lever for 5 seconds. Repeat so that two applications of homogenate are applied to the gel. Two rows of twelve samples (load zones) will fit on a 76 x 76 cm plate, but it is suggested that only one set be applied initially approximately 2 cm from the end until you know how quickly the proteins are going to migrate. If you are going to load two zones, the first load zone should be applied near one end of the plate and the second should be positioned near the center of the gel. This will allow you to screen 24 individuals on the same gel. A total of 3-4 gels can be generated from the same 10 µl of homogenate.
5. Once loaded, the plates are rested on the wicks set up in the electrophoresis chamber. The gels are placed acetate side down on the wicks, being careful that the areas of application nearest the end do not come into contact with the wicks. Place a standard glass microscope slide at both ends of the

gel where the gel makes contact with the wick to be sure there is a uniform electric field during electrophoresis.

6. Subject the gel to electrophoresis for 15 minutes at 200 volts. During this time, one student can consult the handout which gives the ingredients in the stain recipes. Find the enzyme you have been assigned. Note that some of the ingredients are labeled with a single asterisk (*) indicating they are optional ingredients in the recipe. Some ingredients have no asterisk which means they are required. Other ingredients are identified with a double asterisk (**) indicating that they are heat or photosensitive. All of the optional and required ingredients for your enzyme recipe have been mixed and frozen in a scintillation vial. While your gel is running, thaw your mixture by rubbing it between your palms. Add the ** ingredients just before staining. The other student can rinse the sample well plate and adding homogenates from another four individuals from each of the populations. Again, load the wells as indicated in Step 2.
7. After electrophoresis, remove the gel from the chamber. Place the gel on a level surface with its mylar side down. Add the necessary ** ingredients to your stain mixture. Add 2 ml of the agar solution which is in the 60°C water bath. Swirl the mixture and immediately pour the mixture over the entire surface of the gel. This must be done within 10-15 seconds of adding the agar to avoid clumps in the stain mixture. The plate should remain undisturbed for about one minute while the agar solidifies.
8. Place the plates in the dark while the staining reaction takes place.
9. Once the plate has stained sufficiently to resolve the enzymes, place the gel on the white light box to photograph it. After photographing the gel, remove the agar/stain overlay by holding the plate under cold running water or by peeling the overlay off and rinsing the plate with water.

Gel interpretation

Gel interpretation should be straight forward for enzymes that are well resolved. Consult the handout with the stain recipes to determine whether your enzyme is a monomer, a dimer, or perhaps a tetramer. Refer back to Figure 2.1 to distinguish between homozygotes and heterozygotes for these three groups of enzymes. Score each individuals as 11, 12, 22, 13, 23, 33, etc. depending on the banding pattern you see. Enter your data into the class spreadsheet.

Data analysis

1. First calculate descriptive statistics for each of the three populations.
   - % polymorphic loci
   - average number of alleles/locus
   - average observed heterozygosity (H)
   - agreement with Hardy-Weinberg equilibrium at each locus
2. How do the populations compare? Are they similar or very different?
3. Do any of the populations exhibit unusually high or low levels of genetic variability?
4. Conduct pairwise comparisons of the populations to investigate interpopulation genetic variability.
   - genetic identity (*I*)
   - genetic distance (*D*)
5. Do these populations, theoretically of the same species, show similar genetic composition?
6. Do two populations appear to be more closely related to each other than to the third population?

Literature Cited

Avise, J.C. 1994. Molecular Markers, Natural History, and Evolution. Chapman & Hall, Inc. New York, NY.

Harris, H. and D.A. Hopkinson. 1976. Handbook of Enzyme Electrophoresis in Human Genetics. American Elsevier, New York, NY.

Hebert, P.D.N., and M.J. Beaton. 1989. Methodologies for Allozyme Analysis Using Cellulose Acetate Electrophoresis. Helena Laboratories, Beaumont, TX.

Hillis, D.M., C. Moritz, and B.K. Mable. (eds). 1996. Molecular Systematics, 2nd ed. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Miller, D.R., J.H. Borden, and K.N. Slessor. 1989. Inter- and intrapopulation variation of the pheromone ipsdienol produced by male pine engravers, *Ips pini* (Say) (Coleoptera : *Scolytidae*). Journal of Chemical Ecology. 15:233-247.

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

Richardson, B.J., P.R. Baverstock, and M. Adams. 1986. Allozyme Electrophoresis. A handbook for animal systematics and population studies. Academic Press, New York, NY.

Rogers, J.S. 1972. Measures of genetic similarity and genetic distance. Studies in Genetics VII, University of Texas Publication 7213:145-153.

Appendix A
Staining recipes for selected enzymes

The recipes given here are sufficient for one agar overlay. For large groups, it is often convenient to mix these in bulk if you do not want to have to concoct individual stain reactions in scintillation vials; we generally make sufficient volume for 10 overlays (assuming 20 drops per ml). Components marked with an * are optional ingredients in the recipe. Components marked with an ** are sensitive to heat or light and should be added to the stains just before use. Freeze aliquots at -20°C for up to 2 years. We have found that samples that have turned completely purple are no longer good, but samples that are mostly yellow are still usable.

Aldehyde dehydrogenase (AD)
- 0.6 ml Tris-HCl, pH 8.0*
- 1 drop benzaldehyde
- 5 drops MTT
- 1.5 ml NAD
- 5 drops PMS**
- 2 ml agar

Aldehyde oxidase (AO)
- 0.6 ml Tris-HCl, pH 8.0*
- 1 drop benzaldehyde
- 5 drops MTT
- 5 drops PMS**
- 4 ml agar

Adenylate kinase (AK)
- 0.6 ml Tris-HCl, pH 7.0*
- 1.5 ml NAD

1.5 ml ADP solution (0.1 g ADP, 3.15 g D-glucose, 10 ml water)
6 drops $MgCl_2$
5 drops MTT
5 drops PMS**
15 µl Hexokinase **
5 µl G6PDH**
2 ml agar

Alcohol dehydrogenase (ADH)
0.6 ml Tris-HCl, pH 7.0•
1.5 ml NAD
5 drops MTT
3 drops ethanol
5 drops PMS**
2 ml agar

Xanthine dehydrogenase (XDH)
1.0 ml Tris-HCl, pH 8.0*
1.5 ml NAD
20 drops hypoxanthine
5 drops MTT
5 drops PMS**
2 ml agar

Arginine kinase (ARK)
0.5 ml Tris-HCl, pH 8.0*
1.5 ml NAD
5 drops $MgCl_2$
5 drops Phospho-L-arginine
5 drops ADP solution (0.1 g ADP, 3.15 g D-glucose, 10 ml water)
5 drops MTT
5 drops PMS**
10 µl Hexokinase**
10 µl G6PDH**
2 ml agar

Isocitrate dehydrogenase (IDH)
1.0 ml Tris-HCl, pH 7.0*
1.5 ml NADP
15 drops DL-Isocitric acid
8 drops $MgCl_2$
5 drops MTT
5 drops PMS**
2 ml agar

Malate dehydrogenase (MDH)
- 1.0 ml Tris-HCl, pH 8.0*
- 1.5 ml NAD
- 13 drops Malic substrate (180 ml water, 20 ml Tris-HCl, pH 9, 3.68 g L-Malic acid)
- 5 drops MTT
- 5 drops PMS**
- 2 ml agar

Phosphoglucomutase (PGM)
- 1.0 ml Tris-HCl, pH 8.0*
- 1.5 ml NAD
- 5 drops ml $MgCl_2$
- 5 drops Glucose-1-phosphate solution (250 mg G-1-P Grade III, 250 mg G-1-P Grade VI, 5.0 ml water)
- 5 drops MTT
- 5 drops PMS**
- 20 µl G6PDH**
- 2 ml agar

6-Phosphogluconate dehydrogenase (6PGDH)
- 0.6 ml Tris-HCl, pH 8.0*
- 1.5 ml NADP
- 6 drops 6-phosphogluconic acid
- 6 drops $MgCl_2$
- 5 drops MTT
- 5 drops PMS**
- 2 ml agar

Malate dehydrogenase NADP+ (ME)
- 0.6 ml Tris-HCl, pH 8.0*
- 1.5 ml NADP
- 13 drops Malic substrate (180 ml water, 20 ml Tris-HCl, pH 9, 3.68 g L-Malic acid)
- 2 drops $MgCl_2$
- 5 drops MTT
- 5 drops PMS**
- 2 ml agar

Glucose-6-phosphate isomerase (GPI)
- 1.0 ml Tris-HCl, pH 8.0*
- 1.5 ml NAD
- 5 drops Fructose-6-phosphate
- 5 drops MTT
- 5 drops PMS**
- 10 µl G6PDH**
- 2 ml agar

Appendix B

Stock solution recipes for enzyme stain mixtures.

The volumes given in the following recipes are sufficient for assembling 10 stain mixtures for the enzymes listed in Appendix A. Sigma catalog numbers are given.

0.09 M Tris-HCl (pH 7.0)	1.11 g Tris 8.75 ml 1 M HCl V_T = 100 ml	Sigma T-1503
0.09 M Tris-HCl (pH 8.0)	2.22 g Tris 12.4 ml 1 M HCl V_T = 100 ml	
0.20 M Tris-HCl (pH 9.0)	2.47 g Tris 3 ml 1 M HCl V_T = 100 ml	
TG buffer	30 g Tris 144 g glycine V_T = 1 liter Dilute 1:9 before use	
MTT (10 mg/ml) Thizolyl Blue	300 mg/30 ml dH_2O	Sigma M-2128
NAD (2 mg/ml)	300 mg/150 ml dH_2O	Sigma N-7381
NADP (2 mg/ml)	90 mg/45 ml dH_2O	Sigma N-0505
$MgCl_2$ (20 mg/ml)	320 mg/16 ml dH_2O	
ADP solution	200 mg ADP 6.3 g glucose 20 ml dH_2O	Sigma A-2754 Sigma G-5000
Hypoxanthine (10 mg/ml)	100 mg/10 ml dH_2O	Sigma H-9377
Phospho-L-arginine (20 mg/ml)	50 mg/2.5 ml dH_2O	Sigma P-5139
DL-Isocitric acid (100 mg/ml)	0.75 g/7.5 ml dH_2O	Sigma I-1252
Malic substrate	45 ml dH_2O 5 ml Tris-HCl, pH 9.0 920 mg Malic acid Adjust to pH 8.0	Sigma M-6876
Glucose-1-phosphate solution	250 mg Glucose-1-phosphate Grade III 250 mg Glucose-1-phosphate Grade VI 5 ml dH_2O	Sigma G-7000 Sigma G-1259

6-Phosphogluconic acid (20 mg/ml)	60 mg/3 ml dH_2O	Sigma P-7871
Fructose-6-phosphate (20 mg/ml)	50 mg/ 2.5 ml dH_2O	Sigma F-3627
PMS (2 mg/ml)	60 mg/ 30 ml dH_2O	Sigma P-9625
Hexokinase (250 U/ml)	125 U/0.5 ml dH_2O	Sigma H-5500
G6PDH (300 U/ml)	180 U/0.675 ml dH_2O	Sigma G-5885
Agar (1.6%) Bacteriological Grade	4 g/250 ml dH_2O	