NOVEMBER 4 SEMINAR FOR POPULATION BIOLOGY RETREAT

DNA Fingerprinting in Tropical Trees

I. My background -- show 3 slides

II. My purpose is to describe our prospective studies of parentage in tropical forest trees. This is a natural extension of Hubbell's ecological and population data from a plot in Panama, using molecular techniques to establish parentage, and get quantitative measures of gene flow, dispersal, and individual RS in natural populations.

III. Previous work on parentage

A. Based on genetic variation; for example, if mother is known and genes can be identified in her and an offspring, then any different gene must have come from the father. Any adult with that gene could have been the father -- if more than one adult have it, the father can not be specified; obviously, the more alleles and the more loci to check, the less likely that an unrelated adult will have the correct combination.

B. Most common technique, established in last 10 years, based on isozyme variation, or individual variation in proteins; greatest drawback is limited variation: 5-10 loci with 2-4 alleles.

C. Consider 1 locus with 4 alleles, where mother is known (bottom line in lower table of slide) -- chance of eliminating all fathers but one is 20% with 5 loci,and 4 equally frequent alleles, falls to 1 in 1000 if one allele of 4 is common

D. In a study of lilies by Tom Meagher, only 50 out of 2200 offspring had a father specified, using 8 loci with 2-4 alleles

E. Meagher got more information using likelihood estimates -- when father's allele can be specified, but several adults carry it, ones that are homozygous for it are more likely to have been the father than ones heterozygous; using several loci, the likelihood that any non-excluded adult is the father is dependent on the number of loci for which the adult is homozygous for the correct allele; fathers can be ranked on likelihood, and the one with the highest chance assigned paternity, or fractional paternity can be assigned...

F. There are serious problems with this, only one of which I will mention; when likelihood of paternity is used in a large number of cases (Meagher identified 500 father's this way), many are

bound to be wrong -- is it better to use a sample of 50 you know are write, or to enlarge the sample but add some wrong?

IV. DNA fingerprinting has the potential to greatly improve these problems. Possible because highly variable sequences have been found, eg, 77 alleles at one locus in humans; applied in birds, seals, humans, others are working on mammals right now

A. This diagram illustrates the "hypervariable" regions -- not coding regions but tandem repeats of short sequences

B. Variability is in number of units of repeat sequence, not in the sequence itself. It is generated when these sequences pair illegitimately (but quite stably), where a crossover creates one chromosome with 18 and the other with 14 units; these crossovers occur with frequency of 10^{-3} or 10^{-4} per generation, which is 2-3 orders of magnitude higher than standard base substitution rates and explains high individual variation.ing rapid generation of variation

C. This variation can be readily revealed using a restriction enzyme cutting outside the field of repeats, which creates a DNA fragment whose length is dependent on the number of units in the array; different sized fragments are separable on a gel

D. The same core sequence appears many places in humans, and seals, and birds (50-60 in genome)

E. Here's what a gel looks like, with a probe of DNA of the repeat sequence; it hybridizes with all fragments, and different sized fragments are at different locations on the gel; there are about 20 alleles visible here per individual, and father and offspring must share one-half

F. Bands in the offspring not found in father must have come from mother; since there should be 10 of these, only a woman who matches at all 10 could be the mother; the chance that an unrelated individual matches at all 10 is remote; in the table I showed earlier, I give the probability that all but one parent be excluded, which for a modest fingerprinting probe showing 14 alleles is still 97%; to be contrasted with the same figures for an electrophoretic analysis.

G. Almost certain that these kinds of sequences occur in all eukaryotes; Since repeats are universal

V. Ecological application

A. We will assign parentage to randomly selected seedlings from Hubbell plot for a species with 50-100 adults; it is quite feasible to talk about 100 adults and 100 seedlings (map)

B. Part 1 -- pollen and fruit dispersal; andomly selected seedlings from the center of the plot so all parents are likely to be in quadrat; identifying two parents gives both pollen and fruit dispersal; we can readily test theories about gene flow, outcrossing, and species abundance could be tested, that is, does rarity reduce gene flow

C. Part 2 -- Measure of individual RS; randomly collect seedlings throughout the plot, but far enough from boundary that parents are in plot (this can be determined from previous study); exclude adults at edge of plot as well and tally number of seedlings produced by each

D. Both analyses could be repeated on saplings; differences between results on seedlings and saplings would suggest selection based on parent's location; for example, seedlings might be the product of many long distance matings, but offspring of neighbors are more frequent as saplings, suggesting selective advantage to close-by mates