Chapter 4 Diversity

Introduction

Heterozygosity

Gene diversity

Heterozygosity

$$\tilde{H}_l = \sum_u \sum_{v \neq u} \frac{n_{luv}}{n}$$

$$\tilde{H} = \frac{1}{m} \sum_{l=1}^{m} \tilde{H}_l$$

Within-Population Variance of Heterozygosity

$$E\left(\tilde{H}_l\right) = H_l$$

$$Var\left(\tilde{H}_l\right) = \frac{1}{n} H_l\left(1 - H_l\right)$$

$$E(\tilde{H}) = H = \frac{1}{m}\sum H_l$$

$$Cov\left(\tilde{H}_l, \tilde{H}_{l'}\right) = \frac{1}{n}\left(H_{ll'} - H_l H_{l'}\right)$$

where $H_{ll'} = E(X_{jl} X_{jl'})$

$$Var\left(\tilde{H}\right) = \frac{1}{nm^2}\sum_l H_l\left(1 - H_l\right) + \frac{1}{nm^2}\sum_l \sum_{l' \neq l}\left(H_{ll'} - H_l H_{l'}\right)$$

The variance of an average of single-locus statistic must involve two-locus parameters.

The variance among the sample heterozygosities at individual loci may be used to estimate the variance of average heterozygosity, but….

$$S_H^2 = \frac{1}{m-1}\sum_l \left(\tilde{H}_l - \tilde{H}\right)^2$$

$$E\left(\frac{S_H^2}{m}\right) - Var\left(\tilde{H}_l\right)$$

$$= \frac{1}{m(m-1)}\left[\sum_l \left(H_l - H\right)^2 - \frac{1}{n}\sum_l \sum_{l' \neq l} Cov\left(\tilde{H}_l - \tilde{H}_{l'}\right)\right]$$

Total Variance of Heterozygosity

When variation between replicate populations is taken into account, to accommodate the dependence between different sample members, caused by genetic sampling, write

$$E(x_{jl}x_{j'l}) = M_l\,,\ j \neq j'$$

where M_l is the probability that two individuals in a population are heterozygous at locus l

$$E\left(\tilde{H}_l\right) = H_l$$

$$Var\left(\tilde{H}_l\right) = \left(M_l - H_l^2\right) + \frac{1}{n}\left(H_l - M_l\right)$$

Again, $Var\left(\tilde{H}_l\right) \neq 0$ as $n \to \infty$

$$E\left(\tilde{H}\right) = H = \frac{1}{m}\sum_l H_l$$

$$Var\left(\tilde{H}\right) = \frac{1}{m^2}[\cdot\ \cdot\ \cdot] + \frac{1}{m^2 n}[\cdot\ \cdot\ \cdot] \qquad (4.6)$$

Split-plot design (Table 4.1)

$$X_{ijl} = \alpha_i + \beta_{ij} + r_l + (\alpha r)_{il} + (\beta r)_{ijl}$$

GENE DIVERSITY

$$D_l = 1 - \sum_u P_{lu}^{\ 2} \qquad (\text{why?})$$

$$D = 1 - \frac{1}{m} \sum_l \sum_u P_{lu}^{\ 2}$$

$$\text{MLE of D} : \hat{D} = 1 - \frac{1}{m} \sum_l \sum_u \tilde{P}_{lu}^{\ 2}$$

Whthin-population variance of gene diversity

$$E(\tilde{D}_l) = \left(1 - \frac{1+f}{2n}\right) D_l$$

The Presence of the f term indicates that the expected value of D depends on genotype as well as allele frequencies.

$$Var(\hat{D}_l) = \frac{2(1+f)}{n}\left[\sum_u P_{lu}^{\ 3} - \left(\sum_u P_{lu}^{\ 2}\right)^2\right]$$

$$Var(\hat{D}) = \frac{2}{nm^2} \sum_l (1+f_l)\left[\sum_u P_{lu}^{\ 3} - \left(\sum_u P_{lu}^{\ 2}\right)^2\right]$$

$$+ \frac{2}{nm^2} \sum_l \sum_{l' \neq l} \sum_u \sum_v P_{lu} P_{l'u} \Delta_{lu,l'u}$$

Total Variance of Gene Diversity

$$E\left(\tilde{D}_l\right) = D_l\left[\left(1 - \theta\right) - \frac{1}{2n}\left(1 + F - 2\theta\right)\right]$$

Cockerham (1967) defined a group coancestry coefficient

$$\theta_L = \theta + \frac{\left(F - \theta\right)}{n} + \frac{\left(1 - F\right)}{2n}$$

that refers to the identity of a random pair alleles among the 2n alleles of n individuals.

$$E\left(\tilde{D}_l\right) = \left(1 - \theta_L\right) D_l$$

$$Var\left(\tilde{D}_l\right) = \ldots\ldots\ \left(\text{pg } 154\right)$$

The total variance for gene diversity is a very cumbersome expression

Estimating the variance of Diversity

$$Var\left(\tilde{D}\right) = \ldots\ldots\ (\text{pg } 155)$$

It is often suggested that different loci play the role of replicate populations, however, so that the variance among diversities at different loci could serve as an estimate of the total variance

$$E\left(\frac{S_D{}^2}{m}\right) = Var\left(\tilde{D}_l\right) + \frac{1}{m}\left[\sum_l \left(\mathrm{D}_l - D\right)^2 - \sum_l \sum_{l' \neq l} Cov\left(\tilde{D}_l - \tilde{D}_{l'}\right)\right]$$

Example

	BB	Bb	bb	Total
AA	520	3	20	543
Aa	72	6	1	79
aa	502	38	72	612
Total	810	113	311	1234

	CC	Cc	cc	Total
AA	375	25	143	543
Aa	44	27	8	79
aa	391	61	160	612
Total	810	113	311	1234

	CC	Cc	cc	Total
BB	770	83	241	1094
Bb	4	48	25	47
bb	36	12	45	93
Total	810	113	311	1234

	BB	Bb	bb	Total
AA	520	3	20	543
Aa	72	6	1	79
aa	502	38	72	612
Total	1094	47	93	1234

l	$\tilde{D}_l$	l, l'	$\hat{\Delta}_{l,l'}$
A	0.4984	A, B	0.0510
B	0.1710	A, C	0.0117
C	0.4182	B, C	0.0540

l	u	$\tilde{p}_{lu}$	$\tilde{P}_{lu,lu}$	u, v	$\tilde{P}_{lu,lv}$
A	1	0.4720	0.4400	1,2	0.0640
	2	0.5280	0.4959		
B	1	0.9056	0.8784	1,3	0.0381
	2	0.0944	0.0754		
C	1	0.7022	0.6564	2,3	0.0916
	2	0.2978	0.2520		

Summary