

METHODOLOGY

VARIANCE OF AN ESTIMATE OF OUTCROSSING RATE

R. Vencovsky

ABSTRACT

For clarification of an earlier work (Paiva *et al.*, *Brazil. J. Genetics* 16: 1003-1011, 1993) we derive an expression for the variance of an estimate of the outcrossing rate (t) in plant species. This expression is adapted to the special case in which we test the hypothesis of panmixia ($t=1$). An example is given for illustration.

INTRODUCTION

In dealing with molecular or biochemical genetic data we often have estimates of parameters for which no clear statement can be made about their variance. This is usually the case when the maximum likelihood procedure can not be used for estimation. Very often these estimates are ratios of genotypic frequencies or counts or involve even more complicated functions. In such circumstances, Fisher's formula, which is based on Taylor's series expansion, can be used for deriving an approximate expression of the variance of an estimate. We show an application of this method for an estimate of the outcrossing rate (t) in plants for the case in which t is obtained by the method of moments.

METHODS

We consider a population in equilibrium with inbreeding due to a certain amount of natural self-fertilization, such that genotypic frequencies remain constant over time for neutral loci. For a locus with an arbitrary number of alleles the equilibrium frequencies can be expressed as: $P_{uu} = p_u^2 + f p_u(1-p_u)$ for $A_u A_u$ homozygotes and $P_{uv} = 2p_u p_v(1-f)$ for $A_u A_v$ heterozygotes, if we assume the same

fixation index f ($f=F_{IS}$) for all heterozygotes. If this population has a constant rate (t) of outcrossing over time the following relation holds:

$$t = (1-f)/(1+f)$$

and an estimate of t is provided by

$$\hat{t} = (1 - \hat{f}) / (1 + \hat{f}).$$

A commonly used estimate of f , based on the method of moments, is given by

$$\hat{f} = \frac{\tilde{h} - \tilde{H}}{\tilde{h}} \quad (\text{Li, 1966; Weir, 1991})$$

with $\tilde{h} = 1 - \sum_u \tilde{p}_u^2$ being the expected frequency of heterozygotes, assuming panmixia, and $\tilde{H} = 1 - \sum_u \tilde{P}_{uu}$ being the observed frequency of heterozygotes in the sample. A tilde indicates that the corresponding quantities are sample values. The summation is over alleles ($u = 1, 2, \dots, M$) and \tilde{p}_u denotes the sample frequency of allele A_u .

For the variance of t we consider the expression

$$\text{var}(\hat{t}) = \frac{4}{(1+f)^4} \text{var}(\hat{f}) \quad (\text{Weir, 1991}). \quad (1)$$

Since $\hat{f} = \frac{\tilde{h} - \tilde{H}}{\tilde{h}} = 1 - T$, then $\text{var}(\hat{f}) = \text{var}(T)$ with $T = \tilde{H}/\tilde{h}$. An approximate expression for the variance of

\hat{f} can be obtained through Fisher's formula (Weir, 1991), which is expressed as follows, when counts of genotypes follow a multinomial distribution:

$$\text{var}(\hat{f}) = \text{var}(T) \approx n \sum_{u \leq v} \left(\frac{\partial T}{\partial n_{uv}} \right)^2 P_{uv} - n \left(\frac{\partial T}{\partial n} \right)^2 \quad (2).$$

In expression (2) n is the sample size; n_{uv} is the sample number of individuals with genotype $A_u A_v$ being n_{uu} for $A_u A_u$ homozygotes and n_{uv} ($u < v$) for $A_u A_v$ heterozygotes. Clearly $\sum_{u \leq v} n_{uv} = n$. P_{uv} is the population frequency of the $A_u A_v$ genotype ($u \leq v$), such that $\sum P_{uv} = 1$. For obtaining $\text{var}(T)$ the derivatives are evaluated at the expected values of n_{uv} , namely, $E(n_{uu}) = nP_{uu}$ and $E(n_{uv}) = nP_{uv}$.

RESULTS AND DISCUSSION

We express T as a function of the genotypic counts

$$\tilde{H} = 1 - (1/n)(n_{11} + n_{22} + \dots + n_{MM})$$

$$\tilde{h} = 1 - \sum_u \tilde{p}_u^2, \text{ with } \tilde{p}_u = (1/2n) (2n_{uu} + \sum_{u < v} n_{uv})$$

Deriving T with respect to these counts and proceeding as given by Fisher's Formula we get

$$\text{var}(\hat{f}) \approx \frac{1}{n h^4} \{ 2 H^2 [\sum_u p_u^2 (p_u + P_{uu}) + \sum_{u < v} p_u p_v P_{uv}] - 4 H h \sum_u p_u P_{uu} - (2H - Hh - h)^2 + h^2 (1 - H) \}$$

For estimation of $\text{var}(\hat{f})$ the allelic frequencies p_u , p_v , the genotypic frequencies P_{uu} , P_{uv} and the measures of heterozygosity, H and diversity h are replaced by the corresponding sample values.

Under the assumption of panmixia or Hardy-Weinberg proportions we have $f=0$ and $t=1$, $P_{uu} = p_u^2$, $P_{uv} = 2p_u p_v$ and the variance of \hat{f} reduces to

$$\text{var}(\hat{f} | HWE) = \frac{1}{n h^2} [(2 - h) (1 - h) - 2 \sum p_u^3] \quad (4).$$

This variance is appropriate for testing the null hypothesis $H_0: f=0$. In this case, for \hat{t} we have $H_0: t=1$ and the corresponding variance $\text{var}(\hat{t} | HWE) = 4\text{var}(\hat{f} | HWE)$.

Example

For application of these expressions we may consider the genotypic counts of "dendé" oil palms (*Elaeis guineensis* Jacq.) obtained by Santos (1991) for the IDH (Isocitrate dehydrogenase) isozyme system (Table I).

Table I - Number of plants (n_{uv}) and corresponding frequencies (\tilde{P}_{uv}) for genotypes of the IDH isozyme system in "dendé" oil palms (Santos, 1991).

| Genotype | n_{uv} | \tilde{P}_{uv} |
|-------------------------------|----------|------------------|
| A ₁ A ₁ | 10 | 0.1031 |
| A ₁ A ₂ | 12 | 0.1237 |
| A ₁ A ₃ | 15 | 0.1546 |
| A ₂ A ₂ | 13 | 0.1340 |
| A ₂ A ₃ | 21 | 0.2165 |
| A ₃ A ₃ | 26 | 0.2680 |
| | 97 | 0.9999 |

From Table I we get $\tilde{p}_1 = 0.2423$; $\tilde{p}_2 = 0.3041$; $\tilde{p}_3 = 0.4536$.

We reject the hypothesis that these data fit the Hardy-Weinberg equilibrium (HWE) proportions ($\chi^2_{HWE} = 10.38^{**}$; 3 df.). We have:

$$\tilde{h} = 0.643055, \quad \tilde{H} = 0.494845, \text{ such that } \hat{f} = 0.2305.$$

A goodness-of-fit test assuming the inbreeding equilibrium model gives $\chi^2 = 0.27ns$; (2 df.). Hence we may proceed to estimate t . The result $t = 0.6254$ suggests mixed self and random mating.

The corresponding estimated variances are:

$$\hat{\text{var}}(\hat{f}) = 0.006002; \quad \sqrt{\hat{\text{var}}(\hat{f})} = 0.077;$$

$$\hat{\text{var}}(\hat{t}) = 0.010473; \quad \sqrt{\hat{\text{var}}(\hat{t})} = 0.102.$$

Under the null hypothesis, $H_0: f=0$ or $H_0: t=1$, and we have:

$$\hat{\text{var}}(\hat{f} | HWE) = 0.005310 \text{ and } \hat{\text{var}}(\hat{t} | HWE) = 0.021240.$$

With a sample size of $n=97$ we may test $H_0: t=1$ through

$$z = \frac{0.625 - 1}{\sqrt{0.021240}} = -2.57^{**}.$$

Being a one-sided test, this computed z is significant at the 1% level of probability. Using the Hardy-Weinberg goodness-of-fit chi-square ($\chi^2_{HWE} = 10.38^{**}$) we reject the hypothesis of panmixia. Alternatively we accept a model that assumes mixed self and random mating with $\hat{t} = 0.625 \pm 0.102$ and a condition of equilibrium under inbreeding with $\hat{f} = \hat{F}_{IS} = 0.230 \pm 0.077$.

ACKNOWLEDGMENTS

Publication supported by FAPESP.

RESUMO

Para esclarecer um trabalho anterior (Paiva *et al.*, *Brazil. J. Genetics* 16: 1003-1011, 1993) apresentamos a dedução da expressão da variância de uma estimativa da taxa de fertilização cruzada (t) em espécies vegetais. A expressão é adaptada para o caso especial em que

desejamos testar a hipótese de panmixia ($t=1$). Damos um exemplo numérico para ilustrar as deduções feitas.

REFERENCES

- Li, C.C. (1966). *Population Genetics*. The University of Chicago Press, Chicago, pp. 366.
- Paiva, J.R., Kageyama, P.Y. and Vencovsky, R. (1993). Outcrossing rates and inbreeding coefficients in rubber trees (*Hevea brasiliensis* (Willd. Ex ADR. de Juss.) Mueller. Arg.). *Brazil. J. Genetics* 16: 1003-1011.
- Santos, M. de Miranda (1991). Polimorfismo isoenzimático de população subespontânea de dendê (*Elaeis guineensis* Jacq.) do Estado da Bahia e sua relação genética com seis procedências africanas. Doctoral Thesis, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP.
- Weir, B.S. (1991). *Genetic Data Analysis*. Sinauer Associates, Inc., Sunderland, pp. 377.

(Received March 23, 1994)