Teste de Hipótese sobre o coeficiente de coancestria de populações haplóides[1]

Joel Augusto Muniz[2], Silvio César Sartori Ito[2], Daniel Furtado Ferreira[2] e Ruben Delly Veiga[3]

Resumo – Neste trabalho foram estudadas as distribuições dos quadrados médios na análise de variância com as freqüências alélicas em amostras de indivíduos extraídas em populações haplóides, procurando avaliar o uso do teste F para testar a hipótese de nulidade do coeficiente de coancestria. Foi demonstrado que as expressões da soma de quadrados entre populações e da soma de quadrados entre indivíduos dentro de populações divididas pelas respectivas esperanças de quadrados médios possuem distribuições de χ^2 aproximadas, indicando que o quociente entre os quadrados médios na análise de variância tem distribuição F aproximada. Um estudo de simulação foi realizado comparando-se os valores experimentais simulados com os da distribuição teórica de F, procurando-se avaliar a validade das aproximações sugeridas. Os resultados mostraram que o teste F pode ser usado para testar a hipótese de nulidade do coeficiente de coancestria quando se trabalha com pelo menos cinco populações apresentando freqüência alélica média entre 0,3 e 0,7 utilizando-se no mínimo 50 indivíduos.

Termos para indexação: ascendência, haploidia, genética de população, análise estatística.

TEST OF HYPOTHESIS ON THE COEFFICIENT OF COANCESTRY IN HAPLOID POPULATIONS

ABSTRACT – The distribution of the allelic frequency mean squares in the analysis of variance of samples of individuals from haploid populations were studied to evaluate the Snedecor F test, for testing the null hypothesis relative to the coancestry coefficient. It was found that the distribution of the sum of the squares among populations and the sum of squares among individuals within the populations divided by the respective expected mean squares had distribution following approximately the χ^2, indicating that the quotient among the mean squares of the analysis of variance approximate the F distribution. A simulation study to validate the suggested approximations, comparing the simulated values to those of the theoretic F distribution, was carried out. The results showed that the F test may be used to test the coancestry coefficient null hypothesis, when the allelic frequency is between 0.3 and 0.7, working with 50 individuals.

Index terms: ancestry, haploidy, population genetics, statistical analysis.

INTRODUÇÃO

O conhecimento da estrutura genética fornece importantes subsídios para o entendimento da dinâmica evolutiva de populações, permitindo a estimação de parâmetros utilizados na área de Genética e Melhoramento de Plantas em estudos com as diversas espécies. A estimação de parâmetros genéticos de uma população com dados de freqüências alélicas, de acordo com Vencovsky (1992) e Weir (1996), pode ser feita no caso de populações diplóides utilizando a análise de variância em relação a uma variável binária y, que assume o valor 1, quando um determinado alelo, por exemplo A_1 de um loco, está no indivíduo e assume o valor zero, quando este alelo está ausente e presentes os alelos A_2, A_3, ..., A_u.

A técnica de análise é a mesma da Estatística Experimental, em que é feita a associação de um modelo aleatório que descreva a estrutura apresentada pelos dados. Esta técnica foi proposta inicialmente por Cockerham (1969), associando-se os conceitos de correlação como medida da probabilidade de identidade por descendência. No processo de estimação

[1]Aceito para publicação em 11 de fevereiro de 2000.
Financiado pelo CNPq

[2]Dep. de Ciências Exatas (DEX), Universidade Federal de Lavras (UFLA), Caixa Postal 37, CEP 37200-000 Lavras, MG. Bolsista do CNPq. E-mail: joamuniz @ufla.br, scsartori@hotmail.com, danielff@ufla.br

[3]DEX, UFLA, E-mail: delly@ufla.br

dos parâmetros, ocorrem algumas dificuldades, pois a variável indicadora **y** não apresenta distribuição normal. Além disto, em muitos casos os estimadores obtidos pelo método dos momentos correspondem ao quociente entre variáveis aleatórias que não têm distribuição estatística definida.

Cockerham & Weir (1983) apontaram os coeficientes de endogamia e de coancestria, bem como outras medidas de identidade por descendência dos genes, como parâmetros importantes em genética quantitativa e de populações. Esses parâmetros são úteis para informar sobre homozigosidade, deriva, endogamia e variação quantitativa. De acordo com Cockerham (1969), os conceitos e a maior parte da teoria envolvendo coancestria, endogamia, variância das freqüências alélicas e seus correspondentes tamanhos efetivos de população se deve aos trabalhos clássicos de Fisher & Wright.

Falconer (1964), e Hartl & Clark (1989) definiram endogamia como o acasalamento entre indivíduos relacionados por ascendência, tendo como primeiro efeito uma mudança nas freqüências genotípicas de Hardy-Weinberg, devido a um aumento na freqüência de genótipos homozigóticos à custa da freqüência de genótipos heterozigóticos.

Reynolds et al. (1983) utilizaram o coeficiente de coancestria θ como base para uma medida da distância genética da evolução a curto prazo, quando a divergência entre populações com um ancestral comum pode ser considerada como sendo exclusivamente devido à deriva.

Weir & Cockerham (1984) consideraram, para o caso de um dos alelos de um loco, as seguintes definições e notações: F, a correlação entre genes dentro de indivíduos ou endogamia; θ, a correlação entre genes de diferentes indivíduos da mesma população ou coancestria, e f, a correlação entre genes dentro de populações. Os três parâmetros correspondem às estatísticas F de Wright da seguinte forma:

$F = F_{IT}$, $\theta = F_{ST}$ e $f = F_{IS}$.

Os três parâmetros se relacionam através da expressão:

$f = (F - \theta)/(1 - \theta)$.

Para testar a nulidade do coeficiente de endogamia de uma população diplóide com dois alelos, utilizando-se a análise de variância com as freqüências alélicas de um grupo de n indivíduos, Cockerham (1969) sugeriu admitir que o quociente envolvendo os quadrados médios entre indivíduos e dentro de indivíduos tem distribuição de F aproximada, podendo-se aplicar o teste F de Snedecor.

Muniz et al. (1999) avaliaram o teste F proposto por Cockerham (1969) para testar o coeficiente de endogamia de uma população diplóide. Os autores verificaram que o teste F pode ser utilizado quando a freqüência alélica da população estiver entre 0,3 e 0,7 trabalhando-se com pelo menos 30 indivíduos, entre 0,25 e 0,75 com pelo menos 50 indivíduos, e entre 0,20 e 0,80 com pelo menos 100 indivíduos.

O objetivo do presente trabalho foi estudar as expressões das distribuições dos quadrados médios na análise de variância das freqüências alélicas de amostras de indivíduos extraídas de populações haplóides, para avaliar a validade do uso do teste F para testar a nulidade do coeficiente de coancestria.

MATERIAL E MÉTODOS

No caso de populações de indivíduos haplóides com dois alelos, a descrição de amostras de indivíduos pode ser feita por:

$$y_{ij} = p + a_i + e_{(i)j},$$

sendo:

y_{ij} a freqüência alélica do indivíduo j dentro da população i, correspondente aos valores de uma variável binária que assume o valor 1 se o alelo for A, e o valor zero, em caso contrário;

p a freqüência alélica média de todas as populações;

a_i o efeito da população i, com i = 1, 2, ..., r;

$e_{(i)j}$ o efeito do indivíduo j dentro da população i, com j = 1, 2, ..., n.

Este modelo é considerado aleatório, e apresenta todos os parâmetros independentes, e portanto:

$$E[p] = p,\ E[p^2] = p^2;$$

$$E[a_i] = 0,\ E[a_i^2] = Var[a_i] = \sigma_p^2 = p(1-p)\theta;$$

$$E[e_{(i)j}] = 0, E[e_{(i)j}^2] = Var[e_{(i)j}] = \sigma_I^2 = p(1-p)(1-\theta);$$

$$E[y_{ij}] = p,\ E[y_{ij}^2] = p \text{ e } Var[y_{ij}] = p(1-p),$$

sendo θ o coeficiente de coancestria entre os indivíduos das populações. As fontes de variação e as esperanças dos quadrados médios da análise de variância relativa ao modelo de acordo com Weir (1996) estão na Tabela 1.

Na Tabela 1, as expressões de E[QM] mostram que para

testar a hipótese de nulidade do coeficiente de coancestria H_0: $\theta = 0$,
deve-se assumir que o quociente
C = QMP/QMI
tem distribuição F de Snedecor com $v_1 = r\text{-}1$ e $v_2 = r(n\text{-}1)$ graus de liberdade.

A discussão sobre a validade deste critério é feita a partir do estudo das distribuições das somas de quadrados envolvidos, as quais, conforme o modelo são definidas por:

$$\text{SQPopulações} = \frac{1}{n}\sum_i y_{i.}^2 - \frac{y_{..}^2}{nr} = n\sum_{i=1}^{r}(\hat{p}_i - \hat{p})^2 \text{ e}$$

SQIndivíduos / populações =

$$\sum_{ij} y_{ij}^2 - \frac{1}{n}\sum_i y_{i.}^2 = n\sum_{i=1}^{r}\hat{p}_i(1-\hat{p}_i) = \sum_{ij}(y_{ij}-\hat{p}_i)^2,$$

sendo:
$\hat{p}_i$ a freqüência associada ao alelo A na amostra de indivíduos da população i;
$\hat{p}$ a freqüência associada ao alelo A nos indivíduos amostrados em todas as populações.

O teste F foi avaliado por um estudo de simulação, utilizando-se um programa desenvolvido no software SAS (Statistical Analysis System), comparando-se a distribuição do quociente entre quadrados médios com a distribuição F de Snedecor, em diferentes combinações de tamanho de amostra e freqüências alélicas. Foram construídos conjuntos de três, cinco e dez populações com mesma freqüência alélica, isto é, admitindo-se a hipótese $H_0 : \theta = 0$, com as seguintes freqüências alélicas médias (p): 0,10, 0,20, 0,30, 0,40, 0,50, 0,60, 0,70, 0,80 e 0,90 e simulados 1.000 experimentos para cada grupo de populações (r = 3, 5 e 10), considerando-se os seguintes números de indivíduos (n) amostrados com reposição: 10, 20, 30, 40, 50, 100 e 200, totalizando-se 279.000 experimentos em todos os grupos de populações.

Na simulação de cada experimento, utilizou-se a função RANUNI (SAS Institute, 1990) na geração dos efeitos aleatórios. Essa função produz valores de uma variável aleatória com distribuição uniforme no intervalo entre zero e 1.

Em cada tamanho de amostra, foram calculadas, nos 1.000 experimentos, as estatísticas: média ($\bar{\mathbf{x}}$), variância (s^2), percentil 95 (P_{95}) e percentil 99 (P_{99}) para a abscissa na distribuição F de Snedecor associada à variável aleatória, nível de significância (ns), definida em linguagem SAS por:
ns = 1 - Prob F (x; ngl; dgl), sendo:

$$\text{Prob F (x;ndf;ddf)} = \int_0^x f(F)\,df,$$

a probabilidade obtida diretamente no SAS, de que uma variável aleatória tendo distribuição F de Snedecor, com n graus de liberdade no numerador e d graus de liberdade no denominador, assuma valores menores que um determinado x, obtido em cada experimento pelo quociente

$$C = \frac{\text{QM Indivíduos}}{\text{QM Genes dentro de Indivíduos}}.$$

Os resultados das estatísticas nos 1.000 experimentos para cada conjunto de população, nos diversos tamanhos de amostra, foram comparados com os valores teóricos da distribuição F de Snedecor.

RESULTADOS E DISCUSSÃO

Subtraindo-se e somando-se a freqüência alélica média de todas as populações, p, na expressão que define a soma de quadrados de populações (SQP), na análise de variância da Tabela 1, obtém-se:

$$\text{SQP} = n\sum_{i=1}^{r}\left[\hat{p}_i - p - (\hat{p} - p)\right]^2$$

$$= n\sum_{i=1}^{r}\left[(\hat{p}_i - p)^2 + (\hat{p} - p)^2 - 2(\hat{p}_i - p)(\hat{p} - p)\right]$$

$$= n\sum_{i=1}^{r}(\hat{p}_i - p)^2 - rn(\hat{p} - p)^2.$$

Admitindo a hipótese de nulidade do coeficiente de coancestria, H_0: $\theta = 0$, a expressão que define a esperança do quadrado médio de populações na Tabela 1, corresponde a:

E[QMP] = p(1 - p),

Tabela 1. Análise de variância com as freqüências de um alelo em amostras de n indivíduos haplóides extraídos de r populações.

Causa de variação	GL	SQ	QM	E [QM]
Populações	r - 1	SQP	QMP	$p(1-p)[(1-\theta)+n\theta]$
Indivíduos/populações	r(n - 1)	SQI	QMI	$p(1-p)[(1-\theta)]$

podendo-se escrever a seguinte expressão:

$$\frac{(r-1)QMP}{E[QMP]} = \frac{SQP}{E[QMP]} = \frac{n\sum_{i=1}^{r}(\hat{p}_i - p)^2 - rn(\hat{p}_i - p)^2}{p(1-p)}$$

$$= \sum_{i=1}^{r}\left[\frac{\hat{p}_i - p}{\sqrt{p(1-p)/n}}\right]^2 - \left[\frac{\hat{p} - p}{\sqrt{p(1-p)/m}}\right]^2$$

$$= \sum_{i=1}^{r} Z_{1i}{}^2 - Z_2{}^2.$$

Pode-se demonstrar que a freqüência associada ao aleloA na amostra de indivíduos tomada na população i , ($\hat{p}_i$), tem distribuição binomial com média p e variância p(1 - p)/n, e que a freqüência associada ao alelo A nos indivíduos amostrados em todas as populações tem distribuição binomial com média p e variância p(1 - p)/rn e, em conseqüência Z_{1i} e Z_2 tem, respectivamente, distribuição aproximadamente normal com média zero e variância 1, se a hipótese de nulidade H_0: θ = 0 for verdadeira e se as freqüências $\hat{p}_i$ e $\hat{p}$ forem estimadas em amostras grandes extraídas de populações com a freqüência alélica média p que garanta a validade do teorema do limite central. Para estas condições, $\sum_{i=1}^{r} Z_{1i}{}^2$ tem distribuição de χ^2 aproximada, com r graus de liberdade, enquanto $Z_2{}^2$ tem distribuição de χ^2 aproximada com um grau de liberdade, e, conseqüentemente,

$$\frac{(r-1)QMP}{E[QMP]} \approx \chi^2{}_{(r-1)} \text{ e } \frac{QMP}{E[QMP]} \approx \frac{\chi^2{}_{(r-1)}}{r-1}.$$

Pelo mesmo raciocínio, subtraindo-se e somando-se a freqüência paramétrica, p, na expressão que define a soma de quadrados de indivíduos dentro de populações (SQI), na análise de variância da Tabela 1, obtém-se:

$$SQI = \sum_{ij}\left(y_{ij} - \hat{p}_i\right)^2$$

$$= \sum_{ij}\left[\left(y_{ij} - p\right) - \left(\hat{p}_i - p\right)\right]^2$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{n}\left(y_{ij} - p\right)^2 - n\sum_{i=1}^{r}\left(\hat{p}_i - p\right)^2.$$

Assumindo a hipótese de nulidade do coeficiente de coancestria, H_0: θ = 0, a expressão que define a esperança do quadrado médio de indivíduos dentro de populações na Tabela 1, pode ser escrita como

$$E[QMI] = p(1-p),$$

podendo-se escrever o seguinte quociente:

$$\frac{(r-1)QMI}{E[QMI]} = \frac{SQI}{E[QMI]} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n}(y_{ij} - p)^2 - n\sum_{i=1}^{r}(\hat{p}_i - p)^2}{p(1-p)}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{n}\left[\frac{y_{ij} - p}{\sqrt{p(1-p)}}\right]^2 - \sum_{i=1}^{r}\left[\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}\right]^2$$

$$= \sum_{i=1}^{r}\ \sum_{j=1}^{n} Z_{ij}{}^2 - \sum_{i=1}^{r} Z_{1i}{}^2 .$$

A freqüência alélica do indivíduo j dentro da população i, y_{ij}, corresponde aos valores de uma variável binária, que tem distribuição de Bernoulli com média p e variância p(1 - p). Mas a média da variável Z_{ij} é $E[Z_{ij}] = 0$ e a variância é $Var[Z_{ij}] = 1$, e portanto, Z_{ij} tem média e variância iguais à média e variância de uma distribuição normal padronizada, e, conseqüentemente,

$$\sum_{i=1}^{r}\sum_{j=1}^{n} Z_{ij}{}^2 ,$$

corresponde à soma de rn valores ao quadrado de uma variável com as características semelhantes a uma distribuição normal padronizada, podendo ser considerada como uma variável com características semelhantes a uma χ^2 com rn graus de liberdade.

Portanto,

$$\frac{(r-1)QMI}{E[QMI]} = \sum_{i=1}^{r}\sum_{j-1}^{n} Z_{1i}{}^2 - \sum_{i-1}^{r} Z_{1_i}{}^2 \approx \chi^2_{(rn)} - \chi^2_r \approx \chi^2_{[r(n-1)]},$$

indicando que

$$\frac{QMI}{E(QMI)} \approx \frac{\chi^2_{[r(n-1)]}}{r(n-1)} .$$

Considerando mais uma vez a hipótese de nulidade H_0: θ = 0, então

E[QMP] = E[QMI] = p(1-p), e, conseqüentemente,

$$\frac{QMP/E[QMP]}{QMI/E[QMI]} = \frac{QMP}{QMI} \approx \frac{\chi^2_{(r-1)}/r-1}{\chi^2_{[r(n-1)]}/r(n-1)} \approx F[(r-1), r(n-1)],$$

e portanto, os resultados obtidos no desenvolvimento teórico mostram que o quociente QMP/QMI na análise de variância apresentada na Tabela 1 tem distribuição aproximada de F, podendo ser usado como teste para a hipótese de nulidade H_0: q = 0. Este teste F tem utilidade no estudo de estrutura de populações, quando se utilizam dados de freqüências alélicas.

A Tabela 2 apresenta os valores da média e da variância da distribuição F de Snedecor obtidos na análise de variância com três, cinco e dez populações haplóides, variando-se o número de indivíduos. Os valores foram obtidos usando-se as expressões teóricas:

$$E(F) = \frac{n}{n-2},$$

$$Var(F) = \frac{2n^2(m+n-2)}{m(m-2)^2(n-4)}, \text{ sendo:}$$

n o número de graus de liberdade associados ao resíduo;

m o número de graus de liberdade associados a populações.

As Tabelas 3, 4 e 5 ilustram a distribuição dos valores da abcissa associada à variável aleatória nível de significância (ns), obtida em estudo de simulação para validação dos resultados teóricos, envolvendo amostras de diversos tamanhos, extraídas, respectivamente, em três, cinco e dez populações haplóides com dois alelos variando-se freqüências alélicas.

Comparando-se os valores das estatísticas: média, variância, percentil 0,95 e percentil 0,99 obtidos para a variável nível de significância nos 1.000 experimentos simulados nos diversos tamanhos de amostra e números de populações (Tabelas 3, 4 e 5), com os valores teóricos da distribuição F de Snedecor (Tabela 2), percebe-se que utilizando-se 50 indivíduos, o critério sugerido por Cockerham (1969) apresentou estatísticas similares à distribuição F quando as populações apresentaram freqüência alélica média entre 0,30 e 0,70, desde que o número de indivíduos seja de no mínimo 50. Neste caso, o teste F pode ser usado para testar a hipótese de nulidade H_0: q = 0 associada ao coeficiente de coancestria de populações haplóides.

Para tamanho de amostra inferior a 50, ou nos casos em que as populações apresentaram frequência alélica média fora do intervalo (0,30; 0,70), as estatísticas obtidas no estudo de simulação não concordaram com os valores teóricos da distribuição de F. Neste caso, a utilização do teste F para testar a nulidade do coeficiente de coancestria não seria válida.

Tabela 2. Média (E) e variância (Var) da distribuição teórica de F na análise de variância de três, cinco e dez populações (r) haplóides variando-se o número de indivíduos (n) amostrados.

n	r = 3		r = 5		r = 10	
	E(F)	Var(F)	E(F)	Var(F)	E(F)	Var(F)
10	1,0800	1,3796	1,0465	0,6277	1,0224	0,2616
20	1,0363	1,1551	1,0215	0,5561	1,0105	0,2403
30	1,0235	1,0980	1,0139	0,5359	1,0069	0,2339
40	1,0173	1,0717	1,0103	0,5264	1,0051	0,2308
50	1,0137	1,0565	1,0082	0,5209	1,0040	0,2291
100	1,0067	1,0274	1,0040	0,5102	1,0020	0,2256
200	1,0033	1,0135	1,0020	0,5050	1,0010	0,2239

Tabela 3. Distribuição da abcissa associada à variável aleatória nível de significância no teste da hipótese H_0: $\theta = 0$ em N = 1.000 experimentos com amostras de tamanhos diferentes (n), extraídas em três populações de indivíduos haplóides com mesma freqüência alélica ($p_1 = p_2 = p_3$).

n	Estatísticas[1]	Valores das freqüências alélicas (p_1, p_2 e p_3)								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
10	$\overline{F}$	1,0137	1,0937	1,0887	1,0766	1,1038	1,1066	1,1239	1,0901	1,0516
	S_F^2	0,4985	1,3799	1,3123	1,5286	1,5778	1,5510	1,6926	1,1389	0,8412
	P_{95}	0,7970	0,8782	0,8845	0,8859	0,8859	0,8859	0,8852	0,8782	1,000
	P_{99}	0,8845	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
20	$\overline{F}$	1,0553	1,0425	1,0607	1,0424	1,0373	1,0150	1,0832	1,0917	1,0544
	S_F^2	1,0670	1,1391	1,2407	1,2287	1,1948	1,1117	1,2517	1,3082	1,0576
	P_{95}	1,000	0,9248	0,9311	0,9374	0,9382	0,9374	0,9311	0,9151	0,8920
	P_{99}	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Continua

Tabela 3. Continuação.

n	Estatísticas[1]	Valores das freqüências alélicas (p_1, p_2 e p_3)								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
30	$\overline{F}$	1,0186	1,0272	1,0206	1,0424	1,0461	1,0196	1,0735	1,0518	0,9574
	S_F^2	1,0340	1,1558	1,2029	1,2287	1,2696	1,2139	1,4068	1,3354	0,9826
	P_{95}	0,9166	0,9397	0,9515	0,9374	0,9578	0,9565	0,9490	0,9454	0,9166
	P_{99}	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
40	$\overline{F}$	1,0021	0,9835	1,0194	1,0262	1,0027	0,9769	1,0128	1,0364	1,0195
	S_F^2	0,9745	0,9239	1,1670	1,2807	1,1164	1,0068	1,1039	1,1176	1,0970
	P_{95}	0,9320	0,9488	0,9620	0,9656	0,9677	0,9652	0,9610	0,9519	0,9320
	P_{99}	1,000	0,9620	1,000	0,9675	1,000	0,9677	0,9663	1,000	1,000
50	$\overline{F}$	0,9977	0,9868	1,0102	0,9877	1,0351	1,0146	1,0599	1,0306	1,0673
	S_F^2	1,0156	0,8845	0,9283	0,9494	1,1689	1,1236	1,0895	1,0346	1,0780
	P_{95}	0,9370	0,9547	0,9660	0,9707	0,9244	0,9226	0,9665	0,9589	0,9337
	P_{99}	1,000	0,9665	1,000	0,9739	0,9741	0,9739	0,9718	1,000	1,000
100	$\overline{F}$	1,0107	0,9950	0,9982	1,0240	1,0425	1,0116	1,0194	0,9917	1,0165
	S_F^2	1,0084	1,0338	1,0050	1,1605	1,1142	0,9968	1,0292	0,9600	1,0788
	P_{95}	0,9604	0,9410	0,9432	0,9590	0,9609	0,9478	0,9515	0,9388	0,9525
	P_{99}	1,000	0,9813	0,9849	0,9866	0,9868	0,9862	0,9847	0,9813	0,9712
200	$\overline{F}$	0,9932	0,9815	0,9805	0,9993	1,0220	1,0090	0,9856	0,9732	0,9836
	S_F^2	0,9564	0,9765	1,0410	1,0110	1,0265	1,0598	1,0231	1,0290	1,0316
	P_{95}	0,9449	0,9590	0,9479	0,9524	0,9545	0,9530	0,9467	0,9583	0,9437
	P_{99}	0,9822	0,9896	0,9920	0,9931	0,9868	0,9930	0,9920	0,9900	0,9836

[1] $\overline{F}$: valor obtido para a média da abscissa associada à variável nível de significância; S_F^2 : valor obtido para a variância da abscissa associada à variável nível de significância; P_{95}: valor obtido para o percentil 0,95 da abscissa associada à variável nível de significância; P_{99}: valor obtido para percentil 0,99 da abscissa associada à variável nível de significância.

Tabela 4. Distribuição da abcissa associada à variável aleatória nível de significância no teste da hipótese H_0: $\theta = 0$ em N = 1.000 experimentos com amostras de tamanhos diferentes (n), extraídas em cinco populações de indivíduos haplóides com mesma freqüência alélica ($p_1 = p_2 = p_3 = p_4 = p_5$).

n	Estatísticas[1]	Valores das freqüências alélicas (p_1, p_2, p_3, p_4 e p_5)								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
10	$\overline{F}$	1,0137	1,0456	1,0355	1,0479	1,0370	1,0119	1,0287	1,0486	1,0185
	S_F^2	0,4985	0,6174	0,6163	0,7541	0,6745	0,5989	0,5955	0,5754	0,4935
	P_{95}	0,9208	0,9345	0,9638	0,9638	0,9460	0,9420	0,9620	0,9510	0,9345
	P_{99}	0,9933	0,9851	0,9871	0,9896	0,9896	0,9896	0,9871	0,9798	0,9733
20	$\overline{F}$	1,0311	1,0176	1,0090	1,0215	1,0373	0,9902	1,0213	1,0347	1,0186
	S_F^2	0,4982	0,5962	0,5821	0,5801	1,1948	0,5538	0,5821	0,6123	0,5975
	P_{95}	0,9403	0,9511	0,9542	0,9629	0,9581	0,9948	0,9511	0,9539	0,9403
	P_{99}	0,9763	0,9906	0,9919	0,9936	0,9936	0,9568	0,9913	0,9917	0,9797

Continua

Tabela 4. Continuação.

n	Estatísticas[1]	Valores das freqüências alélicas (p_1, p_2, p_3, p_4 e p_5)								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
30	$\overline{F}$	1,0114	1,0092	0,9986	0,9914	1,0137	1,0000	1,0326	1,0128	0,9857
	S^2_F	0,5256	0,5345	0,5247	0,5408	0,5569	0,5744	0,5933	0,5671	0,5028
	P_{95}	0,9441	0,9430	0,9454	0,9572	0,9398	0,9547	0,9414	0,9608	0,9441
	P_{99}	0,9836	0,9894	0,9848	0,9909	0,9921	0,9852	0,9855	0,9920	0,9844
40	$\overline{F}$	1,0065	0,9834	0,9925	0,9973	1,0083	0,9917	1,0075	1,0089	0,9907
	S^2_F	0,5310	0,5026	0,5237	0,5490	0,5275	0,4865	0,5461	0,5199	0,4936
	P_{95}	0,9503	0,9449	0,9557	0,9544	0,9500	0,9539	0,9428	0,9484	0,9559
	P_{99}	0,9918	0,9871	0,9891	0,9907	0,9886	0,9899	0,9855	0,9836	0,9918
50	$\overline{F}$	0,9977	1,0134	1,0088	1,0085	1,0130	0,9909	1,0012	1,0057	1,0353
	S^2_F	0,5048	0,5283	0,4567	0,5191	0,5194	0,4990	0,4827	0,5188	0,5655
	P_{95}	0,9442	0,9389	0,9338	0,9464	0,9519	0,9382	0,9497	0,9526	0,9454
	P_{99}	0,9901	0,9934	0,9889	0,9850	0,9926	0,9822	0,9865	0,9871	0,9932
100	$\overline{F}$	0,9875	0,9957	0,9835	0,9915	1,0170	1,0207	1,0236	0,9993	1,0139
	S^2_F	0,4781	0,5045	0,4667	0,5103	0,5412	0,5048	0,5254	0,5165	0,5315
	P_{95}	0,9578	0,9506	0,9468	0,9599	0,9626	0,9595	0,9519	0,9563	0,9548
	P_{99}	0,9899	0,9920	0,9883	0,9924	0,9916	0,9945	0,9927	0,9889	0,9861
200	$\overline{F}$	0,9829	0,9901	0,9912	0,9915	1,0243	1,0376	0,9906	1,0111	1,0064
	S^2_F	0,5140	0,5333	0,5045	0,5103	0,5149	0,5545	0,5477	0,5386	0,5276
	P_{95}	0,9911	0,9905	0,9515	0,9599	0,9398	0,9538	0,9618	0,9455	0,9440
	P_{99}	0,9492	0,9546	0,9904	0,9924	0,9815	0,9917	0,9915	0,9908	0,9866

[1] $\overline{F}$: valor obtido para a média da abscissa associada à variável nível de significância; S^2_F : valor obtido para a variância da abscissa associada à variável nível de significância; P_{95}: valor obtido para o percentil 0,95 da abscissa associada à variável nível de significância; P_{99}: valor obtido para percentil 0,99 da abscissa associada à variável nível de significância.

Tabela 5. Distribuição da abcissa associada à variável aleatória nível de significância no teste da hipótese H_0:$\theta = 0$ em N = 1.000 experimentos com amostras de tamanhos diferentes (n), extraídas em dez populações de indivíduos haplóides com mesma freqüência alélica ($p_1 = p_2 = ... = p_{10}$).

n	Estatísticas[1]	Freqüências alélicas (p_1, p_2,..., e p_{10})								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
10	$\overline{F}$	1,0029	1,0182	1,0132	1,0198	1,0270	1,0096	1,0148	1,0297	1,0068
	S^2_F	0,2164	0,2536	0,2338	0,2892	0,2883	0,2552	0,2714	0,2535	0,2384
	P_{95}	0,9437	0,9493	0,9379	0,9572	0,9545	0,9489	0,9417	0,9488	0,9437
	P_{99}	0,9713	0,9946	0,9869	0,9923	0,9931	0,9930	0,9869	0,9893	0,9897
20	$\overline{F}$	1,0235	0,9986	0,9932	0,9911	1,0035	0,9946	1,0081	1,0115	1,0063
	S^2_F	0,2322	0,2404	0,2466	0,2322	0,2437	0,2257	0,2492	0,2561	0,2611
	P_{95}	0,9365	0,9469	0,9597	0,9612	0,9513	0,9458	0,9568	0,9469	0,9516
	P_{99}	0,9798	0,9940	0,9948	0,9946	0,9919	0,9890	0,9905	0,9929	0,9886

Continua

Tabela 5. Continuação.

n	Estatísticas[1]	Freqüências alélicas (p_1, p_2,..., e p_{10})								
		0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
30	$\bar{F}$	1,0071	1,0015	0,9802	0,9783	1,0030	0,9966	1,0090	0,9981	0,9897
	S^2_F	0,2488	0,2307	0,2081	0,2517	0,2512	0,2345	0,2202	0,2300	0,2327
	P_{95}	0,9481	0,9591	0,9595	0,9538	0,9481	0,9543	0,9393	0,9563	0,9481
	P_{99}	0,9894	0,9890	0,9886	0,9957	0,9896	0,9886	0,9890	0,9911	0,9824
40	F	0,9955	0,9805	0,9897	1,0006	1,0007	0,9939	0,9969	1,0078	1,0059
	S^2_F	0,2281	0,2050	0,2172	0,2275	0,2263	0,1998	0,2194	0,2232	0,2353
	P_{95}	0,9503	0,9550	0,9362	0,9514	0,9533	0,9428	0,9472	0,9495	0,9396
	P_{99}	0,9925	0,9893	0,9874	0,9909	0,9950	0,9930	0,9878	0,9888	0,9880
50	F	0,9830	0,9912	0,9972	1,0083	1,0165	1,0139	1,0220	1,0152	1,0245
	S^2_F	0,2213	0,2190	0,2177	0,2304	0,2273	0,2052	0,2397	0,2416	0,2441
	P_{95}	0,9486	0,9574	0,9358	0,9444	0,9468	0,9477	0,9404	0,9507	0,9493
	P_{99}	0,9934	0,9902	0,9818	0,9847	0,9927	0,9817	0,9876	0,9912	0,9916
100	$\bar{F}$	0,9790	0,9872	0,1000	0,9922	1,0032	1,0173	0,9959	1,0066	1,0229
	S^2_F	0,2236	0,2289	0,2127	0,5055	0,2186	0,2290	0,2337	0,2436	0,2484
	P_{95}	0,9613	0,9568	0,9526	0,9613	0,9524	0,9391	0,9558	0,9539	0,9477
	P_{99}	0,9938	0,9979	0,9887	0,9905	0,9889	0,9899	0,9938	0,9912	0,9879
200	F	0,9852	0,9719	0,9910	0,9874	1,0179	1,0259	1,0101	1,0148	1,0117
	S^2_F	0,2271	0,2146	0,2209	0,2105	0,2269	0,2306	0,2351	0,2267	0,2209
	P_{95}	0,9458	0,9499	0,9590	0,9460	0,9402	0,9370	0,9469	0,9449	0,9435
	P_{99}	0,9916	0,9909	0,9925	0,9848	0,9940	0,9905	0,9871	0,9904	0,9852

(1) $\bar{F}$: valor obtido para a média da abscissa associada à variável nível de significância; S^2_F : valor obtido para a variância da abscissa associada à variável nível de significância; P_{95}: valor obtido para o percentil 0,95 da abscissa associada à variável nível de significância; P_{99}: valor obtido para percentil 0,99 da abscissa associada à variável nível de significância.

CONCLUSÃO

O teste F de Snedecor pode ser usado para testar a nulidade do coeficiente de coancestria de populações haplóides, desde que se tenham pelo menos cinco populações com freqüência alélica média entre 0,3 e 0,7, trabalhando-se no mínimo com 50 indivíduos.

REFERÊNCIAS

COCKERHAM, C. C. Variance of gene frequency. **Evolution**, Lawrence, v. 23, n. 1, p. 72-74, 1969.

COCKERHAM, C. C.; WEIR, B. S. Variance of actual inbreeding. **Theoretical Population Biology**, San Diego, v. 23, n. 1, p. 85-109, 1983.

FALCONER, D. S. **Introduction of quantitative genetics**. New York : Ronald, 1964. 365p.

HARTL, D. L.; CLARK, A. G. **Principles of population genetics**. Sunderland : Sinauer Associates, 1989. 681p.

MUNIZ, J. A.; BARBIN, D.; VENCOVSKY, R.; VEIGA, R. D. Teste de hipótese sobre o coeficiente de endogamia de uma população diplóide. **Ciência e Agrotecnologia**, Lavras, v. 23, n. 2, p. 410-420, 1999.

REYNOLDS, J.; WEIR, B. S.; COCKERHAM, C. C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, Bethesda, v. 105, p. 767-779, 1983.

SAS INSTITUTE (Cary, Estados Unidos). **Statistical analysis system/graph software**: reference: version 6. Cary, 1990. v. 1, p. 794.

VENCOVSKY, R. Análise de variância de freqüências alélicas. **Brazilian Journal of Genetics**, Ribeirão Preto, v. 15, n. 1, p. 56-60, 1992. Suplemento.

WEIR, B. S. **Genetic data analysis II**: methods for discrete population genetic data. Sunderland : Sinauer Associates, 1996. 445p.

WEIR, B. S.; COCKERHAM, C. C. Estimating F-statistics for the analysis of population structure. **Evolution**, Lawrence, v. 38, n. 6, p. 1358-1370, 1984.