

EL COEFICIENTE DE ENDOGAMIA DE UNA POBLACIÓN FINITA AUTOESTERIL

J. Sahagún-Castellanos

Instituto de Horticultura. Departamento de Fitotecnia. Universidad Autónoma Chapingo. Km. 38.5 Carretera México-Texcoco, C. P. 56230. Chapingo, Estado de México. MÉXICO. Correo-e: jsahagun@correo.chapingo.mx

RESUMEN

El coeficiente de endogamia tiene importancia en el mejoramiento genético de los cultivos agrícolas ya que se relaciona con el número efectivo de una población (tamaño de la población ideal cuya tasa de endogamia es igual a la de la población bajo estudio) que a su vez se relaciona con la probabilidad de deriva genética, con la probabilidad de la extinción de un gene y con la respuesta a la selección. Sin embargo, para el caso de las especies autoestériles [como el tomate de cáscara (*Physalis ixocarpa* Brot.) y algunas especies de ornato del género *Primula*], la fórmula existente para el coeficiente de endogamia de la generación t que relaciona a éste con el de las dos generaciones anteriores no es aplicable si no se conocen los coeficientes de endogamia de las dos primeras generaciones. Para estas especies en este estudio se hizo una derivación teórica de los coeficientes de endogamia de las generaciones 1 y 2 (F_1 y F_2) con base en el modelo de una población que es igual al de la población ideal, excepto que se excluyó la autofecundación y se consideró un tamaño de muestra para avanzar de una generación a la siguiente igual a N. Estos coeficientes fueron $F_1 = 0$ y $F_2 = (4N-5)/\{2[4N(N-1)-1]\}$. Con éstos y con la fórmula del coeficiente de endogamia de la generación t arriba mencionada ya es posible determinar numéricamente el coeficiente de endogamia de cualquier generación de la población que es objeto de estudio.

PALABRAS CLAVE ADICIONALES: respuesta a la selección, identidad por descendencia, mejoramiento genético, número efectivo, tomate de cáscara.

THE ENDOGAMIC COEFICIENTE OF A FINITE SELF STERILE POPULATION

SUMMARY

The endogamic coefficient plays an important role in the genetic improvement of agricultural crops since it is related to the effective number of a population (the ideal population number whose endogamic rate is equal to that of the population being studied) which is related to the probability of genetic derivation, to the probability of the extinction of a gene, and to the selection response. However, in the case of self sterile species (such as the husk tomato (*Physalis ixocarpa* Brot. and some ornamental species in the genera *Primula*), the existing formula for the endogamic coefficient of generation t which is related to the case of the two previous generations is not applicable if the endogamic coefficients of the first two generations is unknown. The theoretical derivation of the endogamic coefficients of the first and second generations (F_1 and F_2) of these species was made based on the model of a population that is equal to an ideal population, except that self-fertilization is excluded and the sample size was considered in order to advance from one generation to the next to be equal to N. These endogamic coefficients were $F_1 = 0$ and $F_2 = (4N-5)/\{2[4N(N-1)-1]\}$. With these and with the endogamic coefficient of generation t previously mentioned, it is now possible to numerically determine the endogamic coefficient of whatever population is being studied.

ADDITIONAL KEY WORDS: selection response, identity by descent, genetic improvement, effective number, husk tomato

INTRODUCCIÓN

Algunos cultivos de ornato del género *Primula* y el tomate de cáscara (*Physalis ixocarpa* Brot.) son ejemplos de especies que tienen en común la autoesterilidad. Esta característica hace que las poblaciones de estas especies no incluyan individuos generados por autofecundación, lo que de una u otra manera se refleja en su coeficiente de endogamia.

En lo general, el coeficiente de endogamia de una población es importante por varias razones. Es bien sabido que este coeficiente se relaciona en forma inversa con el tamaño de población, en tanto que su relación con la depresión endogámica es directa. También, aunque quizá en un contexto más teórico, el coeficiente de endogamia tiene una relación inversa con el número o tamaño efectivo de una población (v.g., Falconer, 1989). Desde un punto

de vista aplicado, esto es muy importante ya que este número a su vez se relaciona con la variabilidad genética y con la respuesta a la selección.

El tamaño efectivo de una población es un concepto que fue creado por Wright (1931) con el fin de que el análisis de una población cualquiera pueda ser hecho en términos de una población estándar, llamada población ideal (Falconer, 1989). En términos del coeficiente de endogamia, el número efectivo o tamaño efectivo de una población es el tamaño de la población ideal que genera la misma tasa de endogamia que la de la población objeto de estudio.

El estudio de la endogamia de poblaciones finitas de individuos diploides monoicos, panmícticos, en donde no ocurre migración, mutación ni selección, como el maíz (*Zea mays* L.), puede ser hecho con base en el modelo de población ideal. Similarmente, el estudio de la endogamia de las poblaciones de especies que, como el tomate de cáscara, difieren de las especies como el maíz en la incapacidad que aquéllas tienen para autofecundarse, también puede basarse en el modelo de población ideal, excepto que no ocurre la autofecundación. Estudios de este tipo se han hecho para poblaciones que difieren de la ideal en un solo aspecto, como el anterior; por ejemplo se ha derivado el número efectivo para poblaciones en que los sexos no están igualmente representados (Falconer, 1989).

Aunque existe una fórmula del coeficiente de endogamia de las poblaciones de individuos autoincompatibles (Crow y Kimura, 1970), ésta no permite la determinación numérica de coeficiente de endogamia de cada una de las generaciones, sólo expresa la relación entre coeficiente de endogamia de una generación con el de las dos generaciones anteriores. Por esta razón y porque tiene aplicaciones en el manejo genotécnico de poblaciones de individuos autoestériles, se visualizó esta investigación con el objetivo de derivar probabilísticamente las partes que hagan posible la determinación numérica del coeficiente de endogamia de cualquier generación de una población que no es ideal debido a que no ocurre la autofecundación.

MÉTODOS Y MARCO TEÓRICO

El estudio se basó en un modelo de población que es el de la población ideal (Falconer, 1989); excepto que no ocurre la autofecundación. El punto de partida fue una población de tamaño infinito de individuos monoicos, diploides y no endogámicos ni emparentados que, exceptuando la autofecundación, se reproduce por apareamiento aleatorio. De esta población se obtiene una muestra aleatoria de tamaño N , para que el apareamiento al azar, sin autofecundación, de los individuos que la integran formen la generación 1. De esta generación también se obtiene una muestra aleatoria para que, como en el caso anterior, se produzca una nueva generación, la

generación 2, y así sucesivamente. A continuación se determinará el coeficiente de endogamia de las generaciones 0 y 1 a la vez que irán siendo definidas la notación y metodologías utilizadas en estas derivaciones. Posteriormente se hará una derivación detallada del coeficiente de endogamia de la generación 2 que es la parte central de los resultados obtenidos en este trabajo. Esta derivación se hará en términos teóricos, basados en los principios probabilísticos básicos. En este sentido es de considerarse que en especies autoincompatibles la disponibilidad de datos experimentales relativos al tema es prácticamente nula.

Sea $A_{p1} A_{p2}$ el genotipo del p -ésimo individuo de la muestra de la población base ($p=1,2,\dots,N$). En virtud de que esta población está formada por individuos no endogámicos ni emparentados, el coeficiente de endogamia de los individuos de la muestra debe ser cero. El apareamiento aleatorio sin autofecundación de los individuos de la muestra produce una población (la generación 1) cuyo arreglo genotípico (Kempthorne, 1957) es

$$(AG)_1 = \sum_{p \neq q}^n \sum_{k=1}^2 \sum_{l=1}^2 \frac{1}{4N(N-1)} A_{pk} A_{ql} \quad (1)$$

Como los genes de cada genotipo de la generación 1 provienen de individuos no emparentados la probabilidad de que tales genes sean idénticos por descendencia [$P(A_{pk} \equiv A_{ql})$] siempre será cero. Por esta razón el coeficiente de endogamia de la generación 1 (F_1) debe ser cero; es decir,

$$F_1 = \sum_{p \neq q}^N \sum_{k=1}^2 \sum_{l=1}^2 \frac{1}{4N(N-1)} P[A_{pk} \equiv A_{ql}] = 0. \quad (2)$$

La generación 2 se formará por el apareamiento aleatorio (sin autofecundación) de los N individuos que integrarán la muestra aleatoria tomada de la generación 1. El coeficiente de endogamia de la generación 2 debe ser diferente de cero toda vez que el apareamiento entre individuos que portan un gene en común es un evento cuya probabilidad de ocurrencia ya es diferente de cero.

RESULTADOS

De acuerdo con el arreglo genotípico de la generación 1 (Ecuación 1), para una muestra de más de un individuo ($N \geq 2$) entre los apareamientos de los N individuos que formarán la generación 2 se incluirá los que ocurren entre dos individuos que portan genes en común e idénticos por descendencia. Más específicamente, la fuente de genotipos

formados por dos genes idénticos por descendencia serán las cruza $A_{pk}A_{ql} \times A_{pk}A_{rs}$, en donde q y r son diferentes de p (p, q, r = 1, 2, ..., N; k, l, s = 1, 2).

El número de cruza (NC) que pueden producir genotipos $A_{pk}A_{pk}$ ($A_{pk}A_{ql} \times A_{pk}A_{rs}$) es

$$NC = [(2)(2)(N-1)][(2)(2)(N-1)] - (2)(2)(N-1)$$

Esto es así porque si q y r no pueden ser iguales a p, para un valor particular de p tanto q como r podrán diferir de N-1 formas de este valor; y en cada una de éstas, l y s pueden tomar 2 valores (el 1 y el 2); además, los progenitores $A_{pk}A_{ql}$ y $A_{pk}A_{rs}$ también ocurren en 2 formas en el arreglo genotípico de la generación 1 (Ecuación 1), la anterior y como $A_{ql}A_{pk}$ y $A_{rs}A_{pk}$, respectivamente; y, finalmente, del número $[(2)(2)(N-1)][(2)(2)(N-1)]$ se tiene que exceptuar los $(2)(2)(N-1)$ casos en que ocurre que $ql = rs$, puesto que si esto ocurriera, se formaría una cruz inexistente en esta fase (una cruz entre dos individuos homocigóticos). Resumiendo,

$$NC = [(2)(2)(N-1)][(2)(2)(N-1)] - (2)(2)(N-1) \\ = 4(N-1)[4N-5] \quad (3)$$

En el arreglo genotípico de la generación 1 (Ecuación 1) se consignan $4N(N-1)$ genotipos, cada uno con una misma frecuencia. Como la muestra aleatoria de tamaño N, que debe producir la generación 2, puede incluir o no los dos progenitores presentes en cada cruz de la forma $A_{pk}A_{ql} \times A_{pk}A_{rs}$, capaz de producir genotipos $A_{pk}A_{pk}$, habrá que calcular la probabilidad de que estos 2 progenitores ingresen a la muestra. Para calcular ésta se requiere determinar el número de formas en que en la muestra se incluyan los dos genotipos involucrados en la cruz y el número de muestras diferentes de tamaño N que pueden ser formadas. El primer número es igual al número de combinaciones o subconjuntos diferentes de N-2 genotipos que se pueden formar con los $4N(N-1)-2$ genotipos no involucrados en la cruz; y el segundo es el número de muestras diferentes de tamaño N que pueden ser obtenidas; así, la probabilidad en cuestión (PIM) es

$$PIM = \frac{\binom{2}{2} \binom{4N(N-1)-2}{N-2}}{\binom{4N(N-1)}{N}} \\ = \frac{1}{4[4N(N-1)-1]} \quad (4)$$

Dado que dos individuos que portan un gen A_{pk} (por ejemplo, los ya referidos $A_{pk}A_{ql}$ y $A_{pk}A_{rs}$) ya son parte de la muestra, la probabilidad de que éstos se crucen es $1/[N(N-1)]$, y la probabilidad de que cada individuo que surja de esta cruz sea de genotipo $A_{pk}A_{pk}$ es $1/4$. De acuerdo con estos dos resultados y con los de las Ecuaciones 3 y 4, es posible determinar la frecuencia del genotipo $A_{pk}A_{pk}$ en la segunda generación $[P(A_{pk}A_{pk})_2]$. Esta frecuencia debe expresarse como

$$P(A_{pk}A_{pk})_2 = 4(N-1)[4N-5] \frac{1}{4[4N(N-1)-1]} \frac{1}{N(N-1)} \frac{1}{4} \\ = \frac{4N-5}{4N[4N(N-1)-1]} \quad (5)$$

En virtud de que p puede tomar N valores y k sólo 2, la frecuencia de individuos homocigóticos en la generación dos o, en otras palabras, el coeficiente de endogamia de la segunda generación (F_2), de acuerdo con la Ecuación 5 es

$$F_2 = 2N \frac{4N-5}{4N[4N(N-1)-1]} \\ = \frac{4N-5}{2[4N(N-1)-1]} \quad (6)$$

DISCUSIÓN

Se encontró que los coeficientes de endogamia de las generaciones 0 y 1 son iguales a 0. El primero por el origen de la muestra (un subconjunto de la población base, formada por individuos no endogámicos ni emparentados), y el de la generación 1 por las consideraciones que condujeron a la Ecuación 2. En virtud de que también se derivó una fórmula para el coeficiente de endogamia de la generación 2 (Ecuación 6) y de que existe una fórmula para F_t en términos de F_{t-1} y F_{t-2} (Crow y Kimura, 1970), ya es posible calcular el coeficiente de endogamia de cualquier generación de la población ideal en que no se produce autofecundación. Esto no sería posible sin el conocimiento del coeficiente de la generación 2 (F_2), porque para calcular el coeficiente de endogamia de la generación 3 se requeriría conocer el coeficiente de endogamia de las generaciones 2 y 1; a su vez, como sin F_2 no se puede determinar F_3 , tampoco se podría determinar T_4 , etc. La fórmula de Crow y Kimura (1970) es

$$F_t = \frac{1+F_{t-2}}{2N} + (1-1/N)F_{t-1} \quad (7)$$

La determinación del coeficiente de endogamia de una población como la que es objeto de estudio debe ser importante para el fitomejorador porque se relaciona con el número efectivo, y éste a su vez: 1) guarda una relación directa con la magnitud de la respuesta a la selección (Robertson, 1960), y 2) guarda una relación inversa con la probabilidad de deriva genética y con la probabilidad de la extinción de un gene y, en consecuencia, con la disminución de variabilidad genética (Kimura y Crow, 1963).

Debido a que tiene una fuente de endogamia que no tiene la población que es objeto de estudio (la que carece de autofecundación), la población ideal debe tener un mayor coeficiente de endogamia en todas las generaciones. Sin embargo, la diferencia entre los coeficientes de ambas poblaciones para los tamaños de muestra comunes en el mejoramiento genético no es muy grande ya que la tasa de endogamia de la población ideal es, aunque mayor, muy similar al de la otra población. Estas tasas son $1/(2N)$ y $1/[2N+1]$, respectivamente (Falconer, 1989).

Evidentemente, los modelos de población como el utilizado en este estudio no se ajustan perfectamente a las poblaciones reales, y en ocasiones las diferencias pueden ser importantes y se debe ser cauteloso. Por ejemplo, Barret *et al.* (2004) en poblaciones naturales de *Narcissus longispathus*, una especie autocompatible de polinización libre, estimaron con base en técnicas de electroforesis aplicadas a semillas y en un programa en FORTRAN (Ritland, 1990) que la tasa de autofecundación y el coeficiente de endogamia promedios fueron 0.37 y 0.324, respectivamente. En los progenitores, sin embargo, el coeficiente de endogamia estimado no difirió estadísticamente de cero debido, según sugieren los

autores (Barret *et al.*, 2004), a depresión endogámica.

CONCLUSIONES

Los coeficientes de endogamia de las generaciones 1 y 2 de una población ideal, excepto que no se produce autofecundación y el tamaño de muestra es N , son 0 y $(4N-5)/\{2[4N(N-1)-1]\}$, respectivamente. Con estos dos coeficientes y la fórmula ya conocida (que expresa el coeficiente de endogamia de una generación en términos de los coeficientes de las dos generaciones anteriores (Ecuación 7), ya es posible determinar numéricamente el coeficiente de endogamia de cualquier generación de esta particular población ideal.

LITERATURA CITADA

- BARRET, S. C. H.; COLE, W. W.; HERRERA, C. M. 2004. Mating patterns and genetic diversity in the wild Daffodil *Narcissus longispathus* (Amaryllidaceae). *Heredity* 92: 459-45.
- CROW, J. F.; KIMURA, M. 1970. An Introduction to Population Genetics Theory. Harper and Row Publishers. New York. 591 p.
- FALCONER, D. S. 1989. Introduction to Quantitative Genetics. Third edition. Longman. 438 p.
- KEMPTHORNE, O. 1957. An Introduction to Genetic Statistics. John Wiley and Sons Inc. New York. 545 p.
- KIMURA, M.; CROW, J. F. 1963. The measurement of effective population number. *Evolution* 17: 279-288.
- RITLAND, K. 1990. A series of FORTRAN computer programs for estimating plant mating systems. *J. Heredity* 81: 235-237
- ROBERTSON, A. 1960. A theory of limits in artificial selection. *Proc. of the Royal Soc. of London* 153: 234-239.
- WRIGHT, S. 1931. Evolution in mendelian populations. *Genetics* 16: 97-159.