Genetic stability of a synthetic variety

Estabilidad genética de una variedad sintética

Rodríguez-Pérez JE, J Sahagún-Castellanos, JL Escalante-González, JJ López-Reynoso, C Villanueva-Verduzco

Abstract. Synthetic varieties (SVs) have been assumed to be genetically stable populations through generations. However, it has been recognized that the randomness of the genetic mechanism, the presence of parents with heterozygous genotypes, and the finite sample sizes of the individuals that represent each parent of a SV may cause gene loss, which may make it difficult to obtain the expected genotypic array of a given SV. To study this issue, the number of non-identical by descent (NIBD) genes in the sample of m plants of each parental line was considered as a random variable (Ym). The objectives were: (1) to determine the mean [E(Ym)] and variance [Var(Ym)] of Ym, and (2) to calculate the average loss of NIBD genes of each parent (Δ_{NIBD}). Parents were assumed unrelated and their assumed inbreeding coefficient was F. It was found that E(Ym) and Var(Ym) were $2-F^m$ and $F^m(1-F^m)$, respectively, and that Δ_{NIBD} was F^m. Evidently this gene loss was larger as m was smaller and F increased when F < 1. Furthermore, if F < 1 the mean tended to 2 as m was larger, whereas the variance and gene loss tended to 0. Finally, if parents are pure lines (F=1) or m is large, Var(Ym) and Δ_{NIBD} reduce to 0.

Keywords: *Allium cepa* L.; *Zea mays* L.; Population stability; Genotypic array; Random mating.

Resumen. Las variedades sintéticas (VSs) se han asumido como poblaciones genéticamente estables generación tras generación. Sin embargo, se ha reconocido que el azar del mecanismo genético, la presencia de progenitores con genotipos heterocigóticos y el número finito de representantes de los progenitores de una VS pueden hacer que se pierdan genes, lo que puede dificultar la obtención del arreglo genotípico esperado de tal VS. Para estudiar este tópico, el número de genes no idénticos por descendencia (NIPD) en los m representantes de cada línea progenitora se consideró como una variable aleatoria (Ym). Los objetivos fueron: (1) determinar la media [E(Ym)] y la varianza de Ym [Var(Ym)], y (2) determinar la pérdida de genes NIPD de cada progenitor (Δ_{NIPD}). Se supuso que el coeficiente de endogamia de los progenitores (no emparentados) fue F. Se encontró que E(Ym) y Var(Ym) fueron $2-F^m$ y $F^m(1-F^m)$, respectivamente, y que para F < 1, Δ_{NIPD} fue F^m. Evidentemente, esta pérdida de genes se incrementó cada vez que F fue mayor, sin llegar a 1, y cada vez que m fue más pequeña. Además, si F < 1, cuando m crece E(Ym) tendió a 2, en tanto que Var(Ym) y Δ_{NIPD} tendieron a 0. Finalmente, basta que se usen líneas puras (F=1) o que m sea grande para que Var(Ym) y Δ_{NIPD} se reduzcan prácticamente a 0.

Palabras clave: *Allium cepa* L., *Zea mays* L.; Estabilidad poblacional; Arreglo genotípico; Apareamiento aleatorio.

INTRODUCTION

A synthetic variety (SV) of a crop such as corn (*Zea mays* L.) or onion (*Allium cepa* L.), among others, is the population generated by random mating of several parents, each represented by *m* plants (Márquez-Sánchez, 1992; Sahagún, 1994). Due to its origin, a SV has been considered as a genetically heterogeneous population that keeps gene and genotypic frequencies constant through generations, as stated in the Hardy-Weinberg principle (Falconer & Mackay, 2001). The stability of the gene and genotypic frequencies of a SV should grant producers reliable seed production of the variety used on their plot. However, the randomness of the genetic mechanism, the finite sample sizes of plants representing each of the parents, usually lines, and the presence of parental genotypes formed by two non-identical by descent (NIBD) genes alter the frequencies of these genes and genotypes, even during the formation of SVs (Sahagún-Castellanos, 2015). This, in turn, means that the SV-generating process does not ensure the assumed genetic stability because there may be a loss of NIBD genes. Since it is believed that neither the population formed by parents nor that resulting from random mating (the SV) can incorporate genes from another source, change in gene frequencies can only occur by loss. This problem can occur not only with the usual type of parents (lines) but also with another type of parents, for example, single- (Sahagún & Villanueva, 1997), trilinear- (Márquez-Sánchez, 2010) and double-crosses (Sahagún & Villanueva, 2003; Sahagún et al., 2005).

The objectives of this research were: (1) to derive the variance and mean of the number of NIBD genes from samples of the *m* representatives of each parent of the SV, and (2) to determine the loss of NIPD genes that occurs in the intragenerational journey between the virtual population constituting each parental line, and the sample of *m* individuals representing it and which are used to form the synthetic variety.

MATERIALS AND METHODS

This study was based on the model of a locus of a diploid species reproduced by random mating. Each of the L parents of the SV was considered as a population whose inbreeding coefficient is F, and that it has no relationship with any of the remaining L-1 parents. It was also considered that the *m* plants of the sample of each parent originated from an endogamous process, as selfing supplemented by selection leading to the formation of a pure line. The parent i (i = 1, 2, 3,…,L) was visualized as a virtual population containing two genotypes: (1) those formed by two identical by descent genes (A_iA_i), and (2) those formed by two non-identical by descent genes (A_iB_i) with frequencies F and 1–F, respectively. Summarizing, the genotypic array of parent i ($GEAP_i$) was:

$$GEAP_i = FA_iA_i + (1 - F) A_iB_i$$
$$i = 1,2,...,L \tag{1}$$

The representatives of each parent of a synthetic variety were considered as the plants that result in a random sample with *m* size with replacement, taken from the population that is the corresponding parent (Equation 1). To study the effects of random sampling with replacement, it was defined the random variable Ym representing the number of non-identical by descent (NIBD) genes reaching each sample, and its mean [E(Ym)] and variance [Var(Ym)] were derived. The number of NIBD genes lost in the formation of the sample of each parent (Δ_{NIBD}) was also determined as an indicator of the intragenerational, and consequently intergenerational, genetic stability of the synthetic variety. This loss was expressed as the difference between the NIBD genes contained by the parent and those carried by the sample that represents it. For this, it was considered that when $0 \leq F < 1$, the parent i (i = 1, 2, …, L) contained either 2 NIBD genes (A_i and B_i) or only 1 (A_i) when F = 1 (Equation 1).

RESULTS

Mean and variance of the number of NIBD genes. If the inbreeding coefficient of the parents is F, $0 \leq F \leq 1$ and *m* = 1. For example, the expected number of NIBD genes in the sample of a parent [E(Y_1)], according to Equation 1, is:

$$E(Y_1) = (1)F + (2)(1 - F)$$
$$= 2 - F$$

while the variance of the same variable [Var (Y_1)] is:

$$Var (Y_1) = (1)^2F + (2)^2 (1 - F) - (2 - F)^2$$
$$= F(1 - F)$$

Similarly, if *m* = 2, according to Equation 1, the genotype pairs that can generate random sampling with replacement that is done to represent line i are: 1) A_iA_i and A_iA_i; 2) A_iB_i and A_iB_i, and 3) A_iA_i and A_iB_i; moreover, the probabilities of occurrence of these events are, in the same order: 1) F^2, 2) $(1-F)^2$, and 3) 2F(1–F). With this information, the mean [E(Y_2)] and the variance [Var(Y_2)] of the random variable (Y_2) representing the number of NIBD genes are derived as shown below:

$$E(Y_2) = (1)F^2 + (2)(1 - F)^2 + (2)[2F(1 - F)]$$
$$= 2 - F^2$$

and

$$Var(Y_2) = (1^2)F^2 + (2^2)(1 - F)^2 + (2)^2 [2F(1 - F)] - [2 - F^2]^2$$
$$= F^2 (1 - F^2)$$

In general, when there are m representatives of each line, the mean and variance of Ym can be derived based on the consideration of only two results of the random sampling of size m with replacement: 1) the resulting m genotypes are A_iA_i, which occurs with an F^m probability, and 2) any result different from the above, with an occurrence probability of $1 - F^m$. As the numbers of NIBD genes associated with these two results are 1 and 2, respectively, then:

$$E(Ym) = (1)F^m + (2)(1 - F^m)$$
$$= 2 - F^m, 0 \leq F \leq 1 \qquad (2)$$

and

$$Var(Ym) = (1)^2F^m + (2)^2 (1 - F^m) - (2 - F^m)^2$$
$$= F^m (1 - F^m), 0 \leq F \leq 1 \qquad (3)$$

Loss of NIBD genes. Equation 3 implies that when F = 0 or F = 1 the variance of the number of NIBD genes is zero. This is because, according to Equation 1, in each parent all individuals have the same genotype (A_iB_i if F = 0 or A_iA_i if F = 1). That is, the sample representing each parent cannot be genotypically variable. Evidently, in these two cases (F=0 and F=1) no NIBD gene loss occurs in the step from the virtual populations (the parents) to the corresponding sets of m individuals that represent them.

Conversely, when 0 < F < 1, the parent i (i = 1, 2, …, L) contains genotypes A_iA_i and A_iB_i. This implies that the number of non-identical by descent (NIBD) genes contained by the sample of each parent is itself a random variable that, according to Equation 2, has a mean equal to $2-F^m$. Based on these arguments, the change, or loss, of NIBD (Δ_{NIBD}) genes that occurs between the number of such genes of each parent $[(NIBD)_p]$ and that which is expected in the random sample that represents it $[(NIBD)_M]$ is:

$$\Delta_{NIBD} = (NIBD)_p - (NIBD)$$
$$= 2 - (2 - F^m) \qquad (4)$$
$$= F^m, 0 < F < 1$$

DISCUSSION

When F = 0 and F = 1 there is no gene loss (Δ_{NIPD} = 0), which can be explained as follows: if F = 0 or F = 1, the genotypic array of parent i (i = 1, 2, …, L) and that of the sample that represents it should be equal to (Equation 1): A_iA_i (if F = 1) or A_iB_i (if F = 0). And if the genotypic arrays do not change, the gene frequencies also do not change and Var(Ym) = 0; and, finally, if there is no variability in the number of NIBD genes from the parent to the sample, there should be no loss of these genes.

In addition, when 0 < F < 1, the expected number of NIBD genes that carry the m plants representing a parent, $2-F^m$ (Equation 2), is larger as F is smaller and m grows. This suggests that if the parents have an inbreeding level F, and 0 < F < 1, increasing m should reduce the loss of NIBD genes. In fact, when m tends to infinity this loss tends to zero. The increase in m, however, has less effect as F is increased (Table 1); for example, when F = 0.25, F = 0.50 and F = 0.75, gene loss is virtually avoided with $m = 4$, $m = 8$ and $m = 20$, respectively. In more detail, in the case where F = 0.75, for example, the lowest gene frequency of the progenitor i is that of B_i and is 0.125 (Equation 1), whereas

Table 1. Expected number [E(Ym)], variance [Var(Ym)] and loss of non-identical by descent genes (Δ_{NIBD}) produced by randomly sampling with m size with replacement a parental line of a synthetic variety. The inbreeding coefficient of the lines is F.
Tabla 1. Número esperado [E(Ym)], varianza [Var(Ym)] y pérdida de genes ancestrales (Δ_{NIBD}) producido por un muestreo al azar de tamaño m con reemplazo de una línea parental de una variedad sintética. El coeficiente de endogamia de las líneas es F.

m	F = 0.25			F = 0.50			F = 0.75			F = 1		
	E(Ym)	Var(Ym)	Δ_{NIBD}	E(Ym)	Var(Ym)	Δ_{NIBD}	E(Ym)	Var(Ym)	Δ_{NIBD}	E(Ym)	Var(Ym)	Δ_{NIBD}
1	1.75	0.188	0.250	1.50	0.250	0.500	1.25	0.188	0.750	1.00	0.00	0.0
2	1.94	0.059	0.063	1.75	0.188	0.250	1.44	0.246	0.563	1.00	0.00	0.0
3	1.98	0.015	0.016	1.88	0.109	0.125	1.58	0.243	0.422	1.00	0.00	0.0
4	1.99	0.004	0.004	1.94	0.059	0.063	1.68	0.216	0.316	1.00	0.00	0.0
5	2.00	0.001	0.001	1.97	0.030	0.031	1.76	0.181	0.237	1.00	0.00	0.0
6	2.00	0.000	0.000	1.98	0.015	0.016	1.82	0.146	0.178	1.00	0.00	0.0
8	2.00	0.000	0.000	2.00	0.004	0.004	1.90	0.090	0.100	1.00	0.00	0.0
10	2.00	0.000	0.000	2.00	0.001	0.001	1.94	0.053	0.056	1.00	0.00	0.0
15	2.00	0.000	0.000	2.00	0.000	0.000	1.99	0.013	0.013	1.00	0.00	0.0
20	2.00	0.000	0.000	2.00	0.000	0.000	2.00	0.003	0.003	1.00	0.00	0.0

with F = 0.25 this frequency is 0.375 (and that of A_i should be 0.625) and, according to the results shown in Table 1, with a sample size of 4 the loss of NIBD genes achieved is virtually zero, while with F = 0.75 obtaining the same result requires a sample size of 20. In addition to the coincidence in gene loss, the variances of these two cases: ($m = 20$, F = 0.75) and ($m = 4$, F = 0.25) are also equal (Table 1, Equation 3).

According to the formula of the variance of Ym, $F^m(1 - F^m)$, the maximum variability of this variable occurs when $F^m = 0.5$, as in the case F = $(0.5)^{0.5}$ and $m = 2$, or when $m = 1$ and F = 0.5, etc. With F = 0.5 and $m=1$ the parent i (i = 1,2,…, L) has the genotypic array: $0.5A_iA_i + 0.5A_iB_i$, and as the sample size is 1, the probability of losing B_i is 0.5 (which is the probability of the randomly-extracted genotype being A_iA_i). In addition, the probability that both are preserved or that A_i is fixed is also 0.5. On the other hand, for any value of F different from 0 and 1($0 < F < 1$), the greatest variance of Ym as a function of the magnitude of m always corresponds to the greatest loss of NIBD genes (Equations 3 and 4; Table 1).

Summarizing, if $0 < F < 1$ the variance and number of lost NIBD genes tend to zero as m increases, and they approach this value faster as F becomes smaller.

In a broader context, as the parents are unrelated, if $0 < F < 1$ the expected number and variance of NIBD genes in the L parents of the synthetic variety should be, according to Equations 2 and 3, $L(2-F^m)$ and $LF^m(1-F^m)$, respectively. In addition, the expected number of NIBD genes should be LF^m. By contrast, when F = 0 or F = 1 both Var(Ym) and NIBD gene loss in each of the L parents are equal to zero, regardless of the magnitude of m.

CONCLUSIONS

The mean and variance of the number of non-identical by descent (NIBD) genes of the m representatives of a parent whose inbreeding coefficient is F are $2-F^m$ and $F^m(1-F^m)$, respectively. According to these results, the average reduces when F is large and/or m is small, and the variance [Var(Ym)] is maximized when $F^m = 0.5$. On the other hand, with $0 < F < 1$ the loss of NIBD genes that occurs between a parent and the sample that it represents (Δ_{NIBD}) is F^m, and both Δ_{NIBD} and Var(Ym) tend to zero, with greater speed in Δ_{NIBD} and with speed increases in both when F is smaller. Finally, when the lines are pure (F = 1) and F = 0, the loss of NIBD genes is always nil.

REFERENCES

Falconer, D.S. & T.F.C. Mackay (2001). Introducción a la Genética Cuantitativa. Acribia SA. España. 469 p.

Márquez-Sánchez, F. (1992). Inbreeding and yield prediction in synthetic maize cultivars made with parental lines: I. Basic methods. *Crop Science* 32: 271-274. doi:10.2135/cropsci1992.0011183X003200020013x

Márquez-Sánchez, F. (2010). Inbreeding coefficient and mean prediction of maize synthetics of three-way lines hybrids. *Maydica* 55: 227-229. http://www.maydica.org/articles/56_341.pdf

Sahagún-Castellanos, J. (1994). Sobre el cálculo del coeficiente de endogamia de variedades sintéticas. *Agrociencia Serie Fitociencia* 5: 67-78. http://www.redalyc.org/pdf/302/30240406.pdf

Sahagún-Castellanos, J. (2015). Genetic stability of synthetics derived from double-cross or three-way line hybrids. *Revista Chapingo Serie Horticultura* 21: 147-155. doi: 10.5154/r.rchsh.2014.04.017

Sahagún-Castellanos, J., J.E. Rodríguez-Pérez & A. Peña-Lomelí (2005). Predicting yield of synthetics derived from double crosses. *Maydica* 50: 129-136. http://www.maydica.org/articles/50_129.pdf

Sahagún-Castellanos, J. & C. Villanueva-Verduzco (1997). Teoría de las variedades sintéticas formadas con híbridos de cruza simple. *Revista Fitotecnia Mexicana* 20: 69-79.

Sahagún-Castellanos, J. & C. Villanueva-Verduzco (2003). Coeficiente de endogamia de las variedades sintéticas con cruzas dobles. *Agrociencia* 37: 641-655. http://www.colpos.mx/agrocien/Bimestral/2003/nov-dic/art-9.pdf