THE CORRELATION BETWEEN RELATIVES IN A SIMPLE AUTOTETRAPLOID POPULATION[1]*

OSCAR KEMPTHORNE

Statistical Laboratory, Iowa State College, Ames, Iowa

Received June 3, 1954

IT is well known that the simple autotetraploid population

$$p^4A^4 + 4p^3qA^3a + 6p^2q^2A^2a^2 + 4pq^3Aa^3 + q^4a^4$$

is in equilibrium under random mating. The problem to be considered herein is that of the correlation between relatives in such a population. The assumption is that segregation is by chromosomes rather than chromatids; that is, if an individual is *abcd*, then there are six equally likely contributions to the offspring consisting of *ab*, *ac*, *ad*, *bc*, *bd*, and *cd*. We suppose that the phenotype (P) of an autotetraploid is made up additively of two parts which are uncorrelated, a genotypic value (G) and an environmental error (E). It will be supposed that relatives are correlated only because their genotypic values are correlated and that there is no correlation between their environmental errors. We shall consider the case when only one locus is involved, and there are m alleles involved, say A_i with frequencies p_i, $i = 1, 2, \cdots,$ m The equilibrium population may be specified briefly by its genotypic array:

$$\{\sum p_iA_i\}^4$$

The results given here are an extension of those of SEWALL WRIGHT (1938), who considered the case when the genes are additive in their effects.

THE BREAKDOWN OF GENOTYPIC VALUE

Consider an individual with genotype $A_iA_jA_kA_l$. Then we have the identity

$$\begin{aligned} A_iA_jA_kA_l = {} & \rho^4 + (A_i - \rho)\rho^3 + \rho(A_j - \rho)\rho^2 + (A_i - \rho)(A_j - \rho)\rho^2 + \rho^2(A_k - \rho)\rho \\ & + (A_i - \rho)\rho(A_k - \rho)\rho + \rho(A_j - \rho)(A_k - \rho)\rho + (A_i - \rho)(A_j - \rho)(A_k - \rho)\rho \\ & + \rho^3(A_l - \rho) + (A_i - \rho)\rho^2(A_l - \rho) + \rho(A_j - \rho)\rho(A_l - \rho) \\ & + (A_i - \rho)(A_j - \rho)\rho(A_l - \rho) + \rho^2(A_k - \rho)(A_l - \rho) + (A_i - \rho)\rho(A_k - \rho)(A_l - \rho) \\ & + \rho(A_j - \rho)(A_k - \rho)(A_l - \rho) + (A_i - \rho)(A_j - \rho)(A_k - \rho)(A_l - \rho). \qquad (1) \end{aligned}$$

where ρ is equal to $\sum_{i=1}^{m} p_iA_i$. Now let us expand each of the terms on the right hand side, and replace each genotype symbol by the genotypic value of that genotype. Using $A_iA_jA_kA_l$ also to denote the genotypic value of the genotype $A_iA_jA_kA_l$ on the left hand side we may write the result as

$$\begin{aligned} A_iA_jA_kA_l = {} & \mu + \alpha_i + \alpha_j + \beta_{ij} + \alpha_k + \beta_{ik} + \beta_{jk} + \gamma_{ijk} + \alpha_l \\ & + \beta_{il} + \beta_{jl} + \gamma_{ijl} + \beta_{kl} + \gamma_{ikl} + \gamma_{jkl} + \delta_{ijkl}, \qquad (2) \end{aligned}$$

---

* Part of the cost of formulae has been paid by the GALTON AND MENDEL MEMORIALFUND.

[1] Journal paper No. J-2610 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project 890.

where the terms are made to correspond one by one, and for instance μ is the population mean, α_i is equal to the expansion of $(A_i - \rho)\rho^3$ expressed in genotypic values, that is, it is equal to the following expression in genotypic values:

$$\sum_a p_a^3 A_i A_a^3 + 3\sum_{a \not\gtrless b} p_a^2 p_b A_i A_a^2 A_b + \sum_{a \not\gtrless b \not\gtrless c} p_a p_b p_c A_i A_a A_b A_c \text{ minus } \mu$$

and so on.

We now state that the terms on the right-hand side of (2) have zero means and are uncorrelated in the population, that is: if we suppose that a random assignment of the numbers 1, 2, 3, 4 is made to the genes of each individual in the population, so as to give the genes an order, then the covariance of each term on the right hand side between two random individuals is zero. The fact that the means are zero is obvious. A couple of examples are sufficient to exhibit orthogonality of terms. Consider the terms α_i and $\alpha_{i'}$ for 2 random individuals: then for each, i and i' take the values a and $a' = 1, 2, \cdots, m$ with probability p_a and $p_{a'}$ independently. Hence

$$\begin{aligned} E(\alpha_i \alpha_{i'}) &= \sum_a p_a[(A_a - \rho)\rho^3] \sum_{a'} p_{a'}[(A_{a'} - \rho)\rho^3] \\ &= (\rho^4 - \rho^4)(\rho^4 - \rho^4) \\ &= 0. \end{aligned}$$

Again take β_{ik} for one individual and γ_{ikl} for another. We have

$$\begin{aligned} \underset{i,k,l}{E}\ (\beta_{ik}\gamma_{ikl}) &= \underset{ik}{E}\ \beta_{ik}\{\underset{l}{E}\ \gamma_{ikl}\} \\ &= \underset{ik}{E}\ \beta_{ik}\{0\} \\ &= 0. \end{aligned}$$

We see therefore that the genotypic value of an individual tetraploid is made up additively of the following parts:

(1) μ the population mean;

(2) four gene effects α_i, α_j, α_k, α_l which by analogy with diploid population may be called the additive effects of genes A_i, A_j, A_k and A_l;

(3) six terms arising from the interaction of two genes, β_{ij}, β_{ik}, β_{jk}, β_{il}, β_{jl}, β_{kl}, which are analogous to dominance deviations in the diploid case;

(4) four terms arising from the interaction of three genes, γ_{ijk}, γ_{ijl}, γ_{jkl} and γ_{ikl}, to which again there is no analogy in the diploid case;

(5) One term arising from the interaction of four genes, δ_{ijkl}, to which again there is no analogy in the diploid case.

The fact that the α_j's are the additively genetic effects of the genes A_i follows from the definition of additive genetic effects and the orthogonality or lack of correlation between the terms on the right-hand side of (2).

There appears to be no terminology established for terms like the γ's or δ's defined above. One might perhaps suggest that terms like β_{ij} should be called digene effects, terms like γ_{ijk} trigene effects, and terms like δ_{ijkl} quadrigene effects.

The values of the different contributions are exhibited in explicit form below.

Definitions:

$$y_{ijkl} = \text{genotypic value as deviation from population mean.}$$

$$\overline{A_iA_jA_kA_l} = y_{ijkl}$$

$$\overline{A_iA_jA_k} = \sum_a p_l y_{ijka}$$

$$\overline{A_iA_j} = \sum_a \sum_b p_a p_b y_{ijab}$$

$$\overline{A_i} = \sum_a \sum_b \sum_c p_a p_b p_c y_{iabc}$$

Contributions:

$$\alpha_i = \overline{A_i}\,,\ \alpha_j = \overline{A_j}\,,\ \alpha_k = \overline{A_k}\,,\ \alpha_l = \overline{A_l}$$

$$\beta_{ij} = \overline{A_iA_j} - \alpha_i - \alpha_j$$

$$\beta_{ik} = \overline{A_iA_k} - \alpha_i - \alpha_k$$

$$\beta_{il} = \overline{A_iA_l} - \alpha_i - \alpha_l$$

$$\beta_{jk} = \overline{A_jA_k} - \alpha_j - \alpha_k$$

$$\beta_{jl} = \overline{A_jA_l} - \alpha_j - \alpha_l$$

$$\beta_{kl} = \overline{A_kA_l} - \alpha_k - \alpha_l$$

$$\gamma_{ijk} = \overline{A_iA_jA_k} - \beta_{ij} - \beta_{ik} - \beta_{jk} - \alpha_i - \alpha_j - \alpha_k$$

$$\gamma_{ijl} = \overline{A_iA_jA_l} - \beta_{ij} - \beta_{il} - \beta_{jl} - \alpha_i - \alpha_j - \alpha_l$$

$$\gamma_{ikl} = \overline{A_iA_kA_l} - \beta_{ik} - \beta_{il} - \beta_{kl} - \alpha_i - \alpha_k - \alpha_l$$

$$\gamma_{jkl} = \overline{A_jA_kA_l} - \beta_{jk} - \beta_{jl} - \beta_{kl} - \alpha_j - \alpha_k - \alpha_l$$

$$\begin{aligned}\delta_{ijkl} = \overline{A_iA_jA_kA_l} - \gamma_{ijk} - \gamma_{ijl} - \gamma_{ikl} - \gamma_{jkl} - \beta_{ij} - \beta_{ik} - \beta_{il} \\ - \beta_{jk} - \beta_{jl} - \beta_{kl} - \alpha_i - \alpha_j - \alpha_k - \alpha_l\end{aligned}$$

Finally we can envisage the partition of genotypic variance by the equation

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_T^2 + \sigma_F^2$$

where

$$\tfrac{1}{4}\sigma_A^2 = \sum_{1=l}^{m} p_i \alpha_i^2$$

$$\tfrac{1}{6}\sigma_D^2 = \sum_{i,j} p_i p_j \beta_{ij}^2$$

$$\tfrac{1}{4}\sigma_T^2 = \sum_{i,j,k} p_i p_j p_k \gamma_{ijk}^2$$

and

$$\sigma_F^2 = \sum_{ijkl} p_i p_j p_k p_l \delta_{ijkl}^2 .$$

These components of variance will be found useful in characterising correlations between relatives.

THE CORRELATIONS BETWEEN RELATIVES

Because of the assumption

$$P = G + E$$

where E is a random variable completely independent from genotype and from relative to relative, the correlation of two individuals say I and II reduces to

$$\frac{Cov(G_I , G_{II})}{\sigma_P^2} .$$

Hence we have to evaluate the covariances between relatives.

We shall make the assumption that the two individuals are related because the sire genes of both come from some set of genes in the population and the dam genes come from another set of genes in the population which is independent of the former set. There is assumed to be no inbreeding in the pedigree, so that the genes on the sire side are statistically independent of the genes on the dam side. Also it will be assumed that there is no inbreeding in the ancestries leading to either sire genes or dam genes, so that the frequency with which the sire genes are any particular ordered pair say A_i and A_j is equal to $p_i p_j$ the frequency in the original population. The same assumption is made about the dam genes.

Let the genes A_i , A_j , A_k , A_l of any individual be ordered so that genes A_i and A_j are genes contributed by the sire of the individual and genes A_k and A_l are contributed by the dam.

Let A_{s1} , A_{s2} , $\cdots$, A_{sk} be the set of genes which can be sire genes of the two relatives. These are a random set of the totality of possible genes. If one relative receives A_{s1} say and another does not receive A_{s1} then the covariance between the two relatives with respect to additive effects is zero. The covariance will be equal to $\frac{1}{4}\sigma_A^2$, if both relatives receive A_{s1} . Let p_{s1}^I be the probability that the first relative receives gene A_{s1} and p_{s1}^{II} the probability that the second relative receives gene A_{s1} . Then the covariance between the relatives because of their both inheriting gene A_{s1} is

$$P_{s1}^I P_{s1}^{II} \{\text{additive effect of } A_{s1}\}^2$$

which on the average will be equal to $P_{s1}^I P_{s1}^{II} \frac{1}{4}\sigma_A^2$.

Now let

$$\phi = \sum_{\substack{\text{sire}\\ \text{alleles}}} P_S^I P_S^{II} ,$$

then the covariance in additively genetic variance because of sire genes is equal to $\phi(\frac{1}{4}\sigma_A^2)$. Let ϕ' be defined similarly for the dam genes. Then the total covariance with respect to the additive gene effects is

$$(\phi + \phi')\tfrac{1}{4}\sigma_A^2$$

When we come to the digenic effects we have to calculate the sum of the probabilities that any pair of original genes are inherited by both relatives on the sire side. Let this be ψ and let the corresponding number of the dam side be ψ'. Also we need the sum of the probabilities that the same gene will be inherited on the sire side and the same gene on the dam side. This is equal to $\phi\phi'$. Then the covariance with respect to what we have called digenic effects is equal to

$$[\phi\phi' + \psi + \psi'] \frac{\sigma_D^2}{6}$$

In the case of trigenic effects we need the sum of the probabilities that (1) both relatives receive the same pair of genes from their sires and only one gene in common from their dams, or (2) vice versa. This is easily seen to be

$$(\phi\psi' + \phi'\psi),$$

and the covariance is equal to

$$(\phi\psi' + \phi'\psi) \frac{\sigma_T^2}{4}$$

Finally in the case of quadrigenic effects the covariance is equal to

$$\psi\psi'\sigma_F^2 .$$

Combining all the terms we have the result that the covariance between the relatives is equal to

$$(\phi + \phi')\tfrac{1}{4}\sigma_A^2 + [\phi\phi' + \psi + \psi']\tfrac{1}{6}\sigma_D^2 + (\phi\psi' + \phi'\psi)\tfrac{1}{4}\sigma_T^2 + \psi\psi'\sigma_F^2$$

In cases when the sire and dam genes are not distinct sets, the multipliers of σ_A^2, σ_D^2, σ_T^2, σ_F^2 are analogous quantities relating to the probabilities that genes of the two relatives are identical by descent.

EXAMPLES OF THE USE OF THE FORMULA

(1) *Parent-offspring*

We have

$$\begin{aligned}\phi &= (\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2}) = 2,\\ \phi' &= 0\\ \psi &= (\tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6}) = 1,\\ \psi' &= 0.\end{aligned}$$

Hence the covariance is equal to

$$\tfrac{1}{2}\sigma_A^2 + \tfrac{1}{6}\sigma_D^2$$

(2) *Full-sib*
We have

$$\begin{aligned}\phi &= 1/4 \times 4 = 1,\\ \phi' &= 1/4 \times 4 = 1,\\ \psi &= 1/36 \times 6 = 1/6,\\ \psi' &= 1/36 \times 6 = 1/6,\end{aligned}$$

so that the covariance is equal to

$$1/2\sigma_A^2 + 2/9\sigma_D^2 + 1/12\sigma_T^2 + 1/36\sigma_F^2 .$$

(3) *Double first cousin*
We find

$$\begin{aligned}\phi &= 1/2 = \phi'\\ \psi &= \frac{8}{6^3} = \psi',\end{aligned}$$

so that the covariance is equal to

$$1/4\sigma_A^2 + \frac{70}{6^4}\sigma_D^2 + \frac{2}{6^3}\sigma_T^2 + \frac{64}{6^6}\sigma_F^2 .$$

(4) *Parent-grandoffspring*
Here we are

$$\begin{aligned}\phi &= 1, \phi' = 0,\\ \psi &= 1/6, \psi' = 0,\end{aligned}$$

so that the covariance is equal to

$$1/4\sigma_A^2 + 1/36\sigma_D^2$$

NOTES ON THE RESULTS

The results given for the four types of relatives enable the components σ_A^2, σ_D^2, σ_T^2 and σ_F^2 to be estimated by equating observed covariances to expected covariances, care having been taken to eliminate environmental correlations. Since in the original population under the assumptions made

$$\begin{aligned}\sigma^2 &= \sigma_G^2 + \sigma_E^2\\ &= \sigma_A^2 + \sigma_D^2 + \sigma_T^2 + \sigma_F^2 + \sigma_E^2 ,\end{aligned}$$

we may also estimate σ_E^2. If one wishes one may then estimate ratios of the components, in the same way as ratios in the diploid case are calculated to estimate what are termed heritabilities.

It is of interest to note that if one had a character which is determined by one locus apart from an additive effect of environment one could use the formulae given here and the corresponding well known ones for the diploid case to determine whether the inheritance of that character was diploid or autotetraploid.

The formulae given may be extended with no change in symbolism to deal with the case when we have an arbitrary number of loci among which there is no linkage and also no epistacy.

The present treatment may be extended to the case of an arbitrary number of independently segregating loci. This will be reserved for a later note, although we remark that the epistatic components now consist of a large number of types: in the case of interactions between two loci there are 10 distinct types of epistatic variance which enter differently into covariances between relatives.

The extension to higher degrees of polyploidy will be worked out by the same methodology.

SUMMARY

The correlations between relatives in a single-locus autotetraploid population in which there is no inbreeding are expressed by a simple formula involving numbers which can be calculated when one knows the chain of relationship of the relatives. Some notes on the results are given.

ACKNOWLEDGMENT

I am indebted to Dr. Sewall Wright for suggestions about presentation.

LITERATURE CITED

WRIGHT, SEWALL, 1938 The distribution of gene frequencies in populations of polyploids. Proc. Nat. Acad. Sci. **24:** 372–377.