

INFERRING PURGING FROM PEDIGREE DATA

Davorka Gulisija^{1,2} and James F. Crow^{1,3}

¹*Department of Dairy Science and Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706*

²*E-mail: gulisija@calshp.cals.wisc.edu*

³*E-mail: jfcrow@wisc.edu*

Received April 20, 2006

Accepted December 27, 2006

The harmful effects of inbreeding can be reduced if deleterious recessive alleles were removed (purged) by selection against homozygotes in earlier generations. If only a few generations are involved, purging is due almost entirely to recessive alleles that reduce fitness to near zero. In this case the amount of purging and allele frequency change can be inferred approximately from pedigree data alone and are independent of the allele frequency. We examined pedigrees of 59,778 U.S. Jersey cows. Most of the pedigrees were for six generations, but a few went back slightly farther. Assuming recessive homozygotes have fitness 0, the reduction of total genetic load due to purging is estimated at 17%, but most of this is not expressed, being concealed by dominant alleles. Considering those alleles that are currently expressed due to inbreeding, the estimated amount of purging is such as to reduce the expressed load (inbreeding depression) in the current generation by 12.6%. That the reduction is not greater is due mainly to (1) generally low inbreeding levels because breeders in the past have tended to avoid consanguineous matings, and (2) there is essentially no information more than six generations back. The methods used here should be applicable to other populations for which there is pedigree information.

KEY WORDS: Inbred load, inbreeding depression, Jersey cattle, pedigree analysis, purging.

Genetic load and inbreeding depression can be reduced if deleterious recessive alleles are removed (purged) from the population by selection against homozygotes in previous generations. The purpose of this article is to show how to infer the amount of purging from pedigrees. Of course, an actual pedigree does not show the individuals that have been purged. We show that, with a few reasonable assumptions and approximations, we can assess the amount of purging that has occurred despite not having observed the purged individuals. The methods are applied to a pedigreed population of Jersey cattle.

A Brief Literature Survey

There are a number of experimental studies of purging, with varying results. The best data come from plants, as expected, because they are easy to control, can be studied in large numbers, and frequently permit self-fertilization. The large plant literature has been

summarized by Byers and Waller (1999), who found that of 34 studies, 14 reported significant evidence of purging, as determined by comparison of inbred individuals with different inbreeding histories.

Crnokrak and Barrett (2002) reviewed studies on 28 animal and plant species. They found strong evidence of purging, although the extent and patterns were quite inconsistent. Among other animal studies, the results are similarly variable. Ballou (1997) analyzed data from 25 captive populations and compared inbreeding depression with varying levels of ancestral inbreeding. The combined data support purging, but the effects are small and individually not significant. Although not studying purging per se, Wang and Hill (1999) showed that selection within inbred lines retarded the increase of homozygosity, as would be expected with purging. Lacy et al. (1996) did a thorough study of inbreeding effects in *Peromyscus polionotus*. Although inbreeding depression was usually found, there was great variability in different

populations and, of interest for this study, for different founding ancestors. The great variability among the contributions of different ancestors argues for relatively few alleles with major effects rather than many with minor effects.

Throughout the various studies, with all their inconsistencies, one clear conclusion emerges. If purging is detected at all, it is attributable mainly, if not entirely, to genes with large effects on viability or fertility. Studies of alleles with minor effects have been uniformly negative, at least in those studies that extend over only a few generations. Genetic load studies of *Drosophila* show that about $3/5$ of the inbreeding depression for viability is due to genes with major effects, the rest having very small effects, averaging less than 5% and grading imperceptibly into normal. There is very little contribution from alleles between these extremes (Greenberg and Crow 1960; Lewontin 1974, p. 50). Furthermore, genes with minor effects have proportionately greater heterozygous expression than those with major effects, approaching semidominance as the effect gets small (Greenberg and Crow 1960; Gillespie 2004, pp. 76–85). Fertility-reducing genes show a similar pattern (Temin 1966). This means that most of the selective removal of mildly deleterious effects is through their much more common occurrence in heterozygotes, so purging in autozygotes is not an important factor. This, too, would argue that whenever purging occurs it is mainly due to alleles with large effects. A second, more tentative, conclusion is that purging effects are not consistent nor large enough to support deliberate inbreeding as a mechanism for dealing with such problems as conserving small populations (Ballou 1997). For a general review of inbreeding and purging see Lynch and Walsh (1998), pp. 251–291.

Here we develop a first approximation to a quantitative theory by which the magnitude of expected purging can be assessed from pedigree information alone. Later we infer, by making assumptions about the frequency and fitness effects of recessive alleles, the purging actually realized. The methods are applied to a pedigreed population of Jersey cattle.

Here is the general idea. We can measure the effect of purging in reducing the load of an individual in the current generation by counting all its inbred ancestors, and weighting each by the degree of inbreeding, the prepurging frequency and fitness effect of recessive alleles, and the relative contribution to the current individual. We now develop this idea more formally.

Assumptions and Definitions

Assumptions: Throughout the study, we make four simplifying, but appropriate assumptions.

First, we define purging as the reduction in frequency of deleterious alleles because of selection against recessive alleles in individuals in earlier generations. This definition ignores partial dominance; all selection is in homozygotes. This is reasonable be-

cause heterozygous selection occurs with and without inbreeding and because in a short pedigree the effect of heterozygous selection is very small. Second, we make no distinction between homozygosity and autozygosity. Deleterious alleles, especially those with major effects, are so rare that identity in state arising from chance associations can be considered negligible. Instead, we assume that identity by descent (IBD) is the sole mechanism for expression of deleterious recessives and possible purging. Third, we assume that in the same path in a pedigree no two ancestors are autozygous for the same allele. Any purging in still earlier generations is ignored. This will not necessarily be the case in complex pedigrees, for example, when an inbred ancestor is a descendant of another. To keep the analysis simple, we ignore these complications. With low levels of inbreeding the error involved in this assumption, which is of the order of products of small quantities, will be small. But in Appendix 1 we show how in more complicated pedigrees this can be taken into account. Finally, for reasons given in the introduction, we assume that in pedigrees of only a few generations, substantial reduction in recessive allele frequency occurs only for those alleles with large fitness effects.

We note in passing that, although only alleles with major effects contribute to purging in few generations, because of their high frequency minor alleles make a substantial contribution to the genetic load.

Definitions: The total genetic load in a gamete is the reduction in fitness in a zygote that would result from the doubling of the chromosomes in that gamete (Morton et al. 1956). For a particular locus

$$L = sq_c \quad (1)$$

in which s is the homozygous selective disadvantage and q_c is the allele frequency in the current generation. Most of the total load is concealed by dominant alleles.

The expressed load (Morton et al. 1956) is the reduction in fitness that actually occurs in the population. This is

$$L_E = sq_c F \quad (2)$$

in which F is Wright's inbreeding coefficient.

Because of purging q_c , the frequency in the current generation, will be less than q_a , the frequency in the ancestral population before purging. We shall determine from the pedigree structure the amount by which q_a is reduced between the ancestral and current generation, even though purged individuals are not observed.

Purging Theory

THE OPPORTUNITY FOR PURGING, O

A pedigree contains only those individuals that have reproduced. An autozygote observed in such a genealogy is not a random

autozygote from an ancestral population as it would be without selection, but from a population that has undergone purging. An inbred ancestor in a pedigree has survived purging and therefore has a higher fitness than a random inbred individual. This is reflected by a change in allele frequency transmitted to the current generation.

The opportunity for purging is the potential for reduction in the total genetic load in the present generation as a consequence of having one or more inbred ancestors. (The actual reduction depends on the frequency and fitness of the deleterious alleles.) Most of this reduction is not expressed because most recessive alleles are heterozygous. We assess the opportunity for purging by the expected contribution of alleles from inbred ancestors to the present generation. An allele that was expressed due to IBD in ancestor j is weighted by its contribution to an individual i in the present generation. Such an allele has a probability $(1/2)^{n-1}$ of being transmitted along a particular path to descendant i , where n is the number of individuals in the path counting j and i . For example, if j is a grandparent of i , $n = 3$; n in the exponent is reduced by one because a homozygote transmits a copy of an allele with probability 1 rather than $1/2$. The amount of purging depends on the probability that the ancestor is autozygous, measured by Wright's inbreeding coefficient, F_j (Wright 1922, 1951; Malécot 1948). Thus the opportunity for purging of deleterious alleles in i due to a path from ancestor j is $(1/2)^{n-1} F_j$, where F_j is the inbreeding coefficient of ancestor j .

Considering all ancestors and paths to individual i , the opportunity for purging in individual i is

$$O_i = \sum_j \sum_k (1/2)^{n-1} F_j, \tag{3}$$

where the summation k is over all paths to i from ancestor j and summation j is over all inbred ancestors of i . In other words, O_i is the probability that an allele in i is a copy of an allele that was autozygous in some ancestor.

Averaged over all individuals in the current generation, the average opportunity for purging is

$$O = \sum_i O_i / N, \tag{4}$$

where N is the number of individuals in the current generation.

EXPRESSED OPPORTUNITY FOR PURGING, O_E

The effect of purging will be manifest only with autozygosity. The probability of a recessive allele that is transmitted from an ancestor to a current individual being expressed in that individual depends on its being autozygous. Following Morton et al. (1956) we denote this as the expressed opportunity. Therefore the expressed opportunity for purging is the potential for reduction in expressed load in the present generation as a consequence of having inbred ancestors. The expressed opportunity for purging in individual i due to inbreeding in ancestors j is

$$O_{Ei} = \sum_j 2F_{i(j)}F_j \tag{5}$$

in which $F_{i(j)}$ is the probability of an allele in i being derived from an allele in j and being autozygous in i . The product is doubled because a homozygous ancestor transmits an allele with probability 1 rather than $1/2$. $F_{i(j)}$ is known as the partial inbreeding coefficient (Lacy et al. 1996; Lacy 1997) and can be computed using path analysis or the tabular method (both are explained in Appendix 2).

The expressed opportunity for purging in individual i , O_{Ei} , is the probability that an allele, autozygous in i , is a copy of an allele that was autozygous in an ancestor. The sum of $2F_{i(j)}F_j$ across ancestors adds up to this probability, because we assume that no ancestors are autozygous for the same allele in the same path.

Averaged over all individuals in the current generation the expressed opportunity for purging is

$$O_E = \sum_i O_{Ei} / N. \tag{6}$$

Because our interest is in the reduction of the deleterious effects of inbreeding by ancestral purging, O_E is normalized by the current inbreeding, measured by Wright's (1922, 1951) inbreeding coefficient, F . So the normalized expressed opportunity for purging is

$$OP = \frac{O_E}{F} \quad O_E \leq F. \tag{7}$$

We use the ratio of averages rather than the average ratios to avoid the dubious practice of averaging of heterogeneous ratios.

REALIZED PURGING

We are interested in the actual amount of purging that has occurred. This depends on the frequency and fitness of the deleterious recessive alleles. Each opportunity for purging due to a path from an ancestor (j) is weighted by the frequency, q_a , and fitness reduction, s , of the deleterious allele. In accordance with standard selection terminology (Falconer and Mackay 1996), we call this the *realized* purging. This measures the effect of deleterious alleles that would otherwise be transmitted to the current generation, but are not represented because of being purged. Note that we employ Δ for absolute and δ for relative change in genetic load.

In a few generations, the fitness effect of an allele with large fitness effect hardly changes. The deleterious allele frequency is a property of the ancestral population and is independent of the pedigree. Thus, q_a is independent of the inbreeding level of j , of the path from j to i , and of inbreeding level of i . Therefore s and q_a can be treated as constants.

Thus, the realized expressed purging in individual i is

$$\Delta L_{Ei} = \sum_j s 2F_{i(j)}F_j q_a s, \tag{8}$$

where the sum is taken over all ancestors common to the parents of *i*. If the allele is lethal or sterilizing, $s = 1$.

We want to express the reduction of expressed load because of purging as a fraction of expressed load when purging is absent, L_{Ei} . Hence we normalize by sFq_a . The ancestral allele frequency, q_a , is used because in the absence of purging this frequency does not change. Thus the normalizing denominator is

$$L_{Ei} = sF_iq_a. \tag{9}$$

The purging effect, appropriately normalized, is the normalized realized expressed purging, which is the proportional reduction of the expressed deleterious recessives:

$$\delta L_E = \sum_i \Delta L_{Ei} / \sum_i L_{Ei}. \tag{10}$$

Substituting from (5) the reduction of the expressed genetic load and the total expressed load without purging are

$$\sum_i \Delta L_{Ei} = s^2q_a O_E N \quad \text{and} \quad \sum_i L_{Ei} = sq_a F N. \tag{11}$$

Dividing the first equation by the second and employing (7) the proportional decrease in the expressed genetic (inbred) load is

$$\delta L_E = sO_E / F = sOP. \tag{12}$$

So we reach our main conclusion: If $s = 1$, the proportional reduction in expressed genetic load due to purging is the normalized expressed opportunity for purging, OP . If $s < 1$, OP is multiplied by s . The value of s must be large, otherwise the amount of purging in a few generations is negligible. From *Drosophila* data, $s \approx 0.9$. Thus, we have measured the effect of deleterious alleles that do not appear in the pedigree because they have been purged. Remarkably, this effect of missing individuals can be inferred from the pedigree structure alone.

The corresponding formula for the proportional reduction in total load is

$$\delta L = sO. \tag{13}$$

If the individual could be rendered completely homozygous, the total load would be reduced by sO .

ALLELE FREQUENCY CHANGE

Opportunity for purging, O , and normalized expressed opportunity, OP , are properties of the pedigree only and are independent of the deleterious allele frequency, q_a . However, purging changes the frequency of the deleterious recessive alleles, so the purging effect reflects this change in allele frequency. The proportional change in allele frequency is thus a measure of the reduction in expressed load brought about by purging:

$$\delta L_E = \Delta L_E / L_E = (sq_a F - sq_c F) / sq_a F = (q_a - q_c) / q_a. \tag{14}$$

Thus,

$$q_c / q_a = 1 - sOP. \tag{15}$$

If we include alleles hidden in heterozygotes (which we designate by upper case), the proportional reduction in total load is

$$Q_c / q_a = 1 - sO. \tag{16}$$

Note that, although we do not know the absolute values of the q 's, we know their ratio.

It may seem strange that we distinguish between q_c and Q_c , because the probability of transmission of an allele from an ancestor to a descendant is the same whether the descendant is autozygous or allozygous. In a multigeneration pedigree with ancestors and descendants interspersed throughout, O and OP would be expected to be equal. But when opportunities for purging come in a recent generation they can be out of phase. This is because after purging occurs, the reduced load is passed to a subsequent generation, but this reduction is not realized until a still later generation. Hence $O > OP$.

An example on how the normalized expressed opportunity for purging is computed from a simple pedigree is given in Appendix 2.

Results: Data from Jersey Cattle

We examined pedigree records of 59,778 U.S. Jersey cows that had a first calving between 1995 and 2000, and at least six generations of known pedigrees (extracted from the American Jersey Cattle Association database). In this sample of cows, the inbreeding coefficients (F) range from 0.006 to 0.34, with the distribution skewed toward the high values (Fig. 1). The mean and median values of F are 0.068 and 0.063, with the 25th and 75th percentiles of the distribution at 0.045 and 0.083.

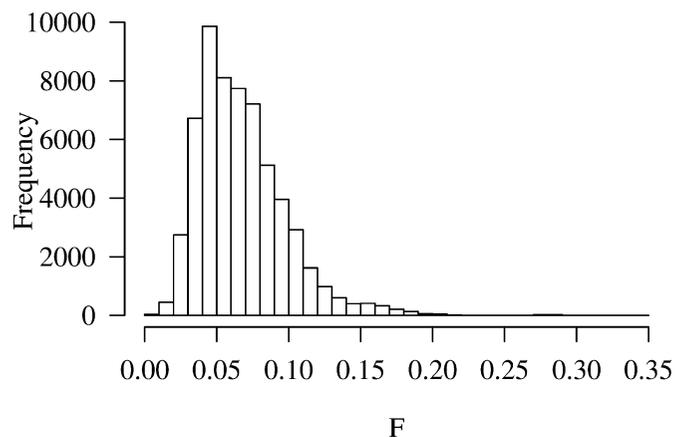


Figure 1. Frequency distribution of inbreeding coefficients for Jersey cows with at least six generations of known pedigree data.

Contributions from an ancestor to a descendant and partial inbreeding coefficients were calculated using a modification of a tabular method for computing additive relationships between animals (Henderson 1976; Ballou 1983; Lacy et al. 1996; Lacy 1997; Appendix 2).

The results are:

Total opportunity for purging:	$O = 0.17$	(eq. 4)
Expressed opportunity:	$O_E = 0.0086$	(eq. 6)
Mean inbreeding coefficient:	$F = 0.068$	(Fig. 1)
Normalized expressed opportunity:	$OP = 0.126$	(eq. 7)

Thus, the opportunity for purging is such as to reduce the autozygosity frequency of alleles with strong effect on fitness by about 12.6%. We can say that, with the level of inbreeding in Jersey cattle, the expressed genetic load in the current generation is reduced by about 12.6% because of ancestral inbreeding, provided the fitness of the homozygous allele is near zero. The total opportunity for purging, most of which is concealed by dominant alleles, is 0.17. If the population could be made completely autozygous for a locus, and $s = 1$, the reduction of total load (inbreeding depression) would be 17%.

We are aware that consideration of a multilocus case would require the additional assumption that fitness is a linear function of F . Nonlinearity in fitness- F relationship could hinder analyses of effects of purging. However, there is little evidence for nonlinearity in inbreeding decline for small values of F . In *Drosophila* there is sometimes slight synergistic fitness epistasis at inbreeding levels near $F = 1$, but not for inbreeding levels found in livestock (Temin et al. 1969). The number of lethal equivalents in Jersey cattle is estimated at about 1.1 per gamete (Pisano and Kerr 1961), which would predict a viability decrease of about 7% in our population.

Discussion

Using pedigree information, we have estimated the amount by which ancestral inbreeding has decreased the current total genetic load and level of inbreeding depression. In our inference, we have assumed that the alleles being purged are completely recessive. Yet it is known that almost all "recessive" alleles have deleterious heterozygous effects. The effect is small; in *Drosophila* the heterozygous effect of a recessive lethal from a wild population is 1–3%. So we would not expect in a pedigree of a half-dozen generations for this to be a significant factor.

Another assumption that we made is that the fitness of autozygotes is near zero. As mentioned earlier, data from *Drosophila melanogaster* indicate that most homozygous chromosomes are either lethal or sterile or have very small viability and fertility effects, with very few in between. This is the case for several

Drosophila species (Greenberg and Crow 1960; Dobzhansky et al. 1963; Temin 1966; reviewed in Lewontin 1974, p. 38). Nevertheless, there are a few alleles with viability or fertility of intermediate value, and neglecting them biases the results. If all mutations that cause a large viability decrease are included, the weighted average reduction in fitness, s , is about 0.9. Values for fertility reduction are almost the same. This suggests that our estimated purging values for major factors, based on $s = 1$, are about 10/9 of the true value, assuming that *Drosophila* data are relevant for cattle.

Our discussion of purging has dealt with the effects of consanguineous mating in large populations. A second kind of inbreeding effect can involve separation into discrete strains; and there are various intermediate situations. Examples of the second kind are sibling-mated lines of mice and selfed lines of maize. In the mouse inbreeding experiments summarized by Lynch and Walsh (1998, pp. 274–276), sibling-mated lines showed a sharp initial decline in litter size. Yet after some six generations the size increased to essentially normal. Incidentally, this adds to the already strong evidence that overdominance is not a major cause of inbreeding decline. At the same time, many lines became extinct, so that only about 1/10 of the original lines survived for 12 generations. The bulk of selection was between lines rather than within, the surviving lines being those that had the fewest deleterious alleles.

Similar results were found after self-fertilization of normally outcrossing plants. After two generations there was no further decline. Nevertheless, crosses between different lines showed substantially increased fitness, as if the different lines were fixed for different alleles. Again these observations argue that purging is caused by deleterious alleles with large effect.

The clear conclusion from our study and others is that the effect of purging from alleles of very small effect is negligible in the limited number of generations considered in domestic livestock, zoo animals, and most experimental studies. Nevertheless, purging of alleles with minor effects can become important in long continued inbreeding. This is apparent in the high performance of long-time inbred lines in mice in which there have been many generations of (unconscious?) selection between and within inbred lines. Similarly, long-time selection in inbred strains of maize has produced lines that, although not as good as hybrids, are as good as hybrids of a few generations ago. Because lethals and steriles are automatically eliminated, the improvement must be due to alleles of lesser effect. A particularly insightful study was done by Latter et al. (1995). They found that *Drosophila* strains that had experienced very slow inbreeding over some 200 generations showed considerably less inbreeding depression than when homozygosity was produced quickly. Nevertheless, as already emphasized, purging in pedigrees with a short number of generations is almost entirely due to alleles with major effects.

ACKNOWLEDGMENTS

DG thanks her academic advisors, D. Gianola and K. A. Weigel, for support, valuable discussions, and suggestions. Associate Editor S. Otto provided comments that were unusually extensive, perceptive, and constructive. The article is much improved thereby. Research was supported by the American Jersey Cattle Association, the Wisconsin Agriculture Experiment Station, and by grant (National Science Foundation) NSF DEB-0089742. The USDA Animal Improvement Programs Laboratory and the American Jersey Cattle Association are acknowledged for providing the data.

LITERATURE CITED

- Ballou, J. D. 1983. Calculating inbreeding coefficients from pedigrees. Pp. 509–520 in C. M. Schonewald-Cox, S. M. Chambers, B. MacBryde, and L. Thomas, ed. *Genetics and conservation*. Benjamin/Cummings, Menlo Park, CA.
- . 1997. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.* 88:169–178.
- Byers, D. L., and D. M. Waller. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Ann. Rev. Syst. Ecol.* 30:479–513.
- Crnokrak, P., and C. H. Barrett. 2002. Purging the genetic load: a review of the experimental evidence. *Evolution* 56:2347–2358.
- Dobzhansky, Th., A. S. Hunter, O. Pavlovsky, B. Spassky, and B. Wallace. 1963. Genetics of natural populations. XXXI. Genetics of an isolated marginal population of *Drosophila pseudoobscura*. *Genetics* 48:91–103.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. 4th ed. Longman, Essex, U.K.
- Gillespie, J. H. 2004. *Population genetics. A concise guide*. 2nd ed. Johns Hopkins Univ. Press, Baltimore, MD.
- Greenberg, R., and J. F. Crow. 1960. A comparison of the effect of lethal and detrimental chromosomes from *Drosophila* population. *Genetics* 45:1153–1168.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* 32:69–83.
- Lacy, R. C. 1997. Errata. *Evolution* 51:1025.
- Lacy, R. C., G. Alaks, and A. Walsh. 1996. Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*. *Evolution* 50:2187–2200.
- Latter, B. D., J. C. Mulley, D. Reid, and L. Pascoe. 1995. Reduced genetic load revealed by slow inbreeding in *Drosophila melanogaster*. *Genetics* 139:287–279.
- Lewontin, R. C. 1974. *The genetic basis of evolutionary change*. Columbia Univ. Press, New York.
- Lynch, M., and B. Walsh. 1998. *Genetic analysis of quantitative traits*. Sinauer Associates, Sunderland, MA.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Masson, Paris.
- Morton, N. E., J. F. Crow, and H. J. Muller. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Natl. Acad. Sci. U.S.A.* 42:855–863.
- Pisano, J. F., and W. E. Kerr. 1961. Lethal equivalents in domestic animals. *Genetics* 46:773–786.
- Temin, R. G. 1966. Homozygous viability and fertility loads in *Drosophila melanogaster*. *Genetics* 53:27–46.
- Temin, R. G., H. U. Meyer, P. S. Dawson, and J. F. Crow. 1969. The influence of epistasis on homozygous viability depression in *Drosophila melanogaster*. *Genetics* 61:497–519.
- Wang, J. and W. G. Hill. 1999. Effect of selection against deleterious mutations on the decline in heterozygosity at neutral loci in closely inbreeding populations. *Genetics* 153:1475–1489.

Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Natur.* 56:330–338.

———. 1951. The genetic structure of populations. *Ann. Eugen.* 15:323–354.

Associate Editor: S. Otto

Appendix 1: Inferring Purging From Complex Pedigrees (DG)

The complete pedigrees that we used in analysis of Jersey cattle rarely extend past six generations with most of the autozygosity in recent generations. This justifies the assumption that no two ancestors in the same path in a pedigree were autozygous for the same allele. Applying this assumption for complex pedigrees can yield the (normalized) opportunity for purging, which is an approximation and not an exact probability. Here we show how exact probabilities are obtained when dealing with inbred ancestors descending from other autozygotes.

Autozygous ancestors whose ancestors had different levels of inbreeding may provide different opportunities for purging. This is because inbred individual descending from other autozygotes will be less likely to carry deleterious alleles than when their ancestors have not been inbred. Frequency of deleterious alleles among such individuals, q'_a , is

$$q'_a = q_a(1 - sOP), \quad (17)$$

where q_a is the allele frequency in the absence of purging.

It follows that (the unweighted) realized expressed purging (i.e., expressed load) in an ancestor descending from other autozygotes is

$$sFq'_a = sFq_a(1 - sOP) = sFq_a - sFsOPq_a = sFq_a - s^2O_Eq_a. \quad (18)$$

Because expressed opportunities for purging, O_E , may differ among ancestors, it follows that heterogeneous allele frequencies need to be accounted for in measures of purging for the current generation. This could be done by taking into account instances such as both the descendant and the inbred ancestor having common inbred ancestors (i.e., a nonnull probability that an allele was expressed in all pairs, triplets, etc., of such ancestors).

The probability that a closer ancestor in a pedigree, j , is autozygous for the same allele as its predecessor k is given by twice the partial inbreeding coefficient, $2F_{j(k)}$, weighted by the more distant ancestor's inbreeding coefficient, F_k . To compute the *Opportunity of Purging* in i , the opportunity for purging in a closer ancestor is discounted for what was already accounted for in its predecessor. Thus, negative $2F_kF_{j(k)}$ is weighted by the contribution of ancestor closer to the genome of i :

$$O_i = \sum (1/2)^{n_k-1} F_k + \sum (1/2)^{n_j-1} (F_j - 2F_{j(k)}F_k), \quad (19)$$

where n_k and n_j are the number of individuals in a path from i to k and i to j including i and the ancestor, the summation being over all paths to the ancestor and k being farther back than j .

The same principle applies to computation of expressed opportunity for purging. Thus, with two autozygous ancestors in a path, the realized expressed purging, ΔL_{Ei} becomes

$$\begin{aligned} \Delta L_{Ei} &= \sum 2F_{i(k)}F_kq_a s^2 + \sum 2F_{i(j)}(F_jq_a s^2 - 2F_{j(k)}F_k s q_a s^2) \\ &= \sum 2F_{i(k)}F_kq_a s^2 + \sum 2F_{i(j)}(F_j - 2F_{j(k)}F_k s)q_a s^2 \\ &= \sum 2F_{i(k)}F_kq_a s^2 + \sum 2F_{i(j)}(F_j - O_{Ej(k)s})q_a s^2 \\ &= \sum 2F_{i(k)}F_kq_a s^2 + \sum 2F_{i(j)}(F_j - F_j O P_{j(k)s})q_a s^2 \\ &= \sum 2F_{i(k)}F_kq_a s^2 + \sum 2F_{i(j)}F_j(1 - O P_{j(k)s})q_a s^2. \end{aligned} \quad (20)$$

By induction, in the case of multiple inbred ancestors in a path, the opportunity for purging becomes

$$\begin{aligned} O_i &= \sum_j (1/2)^{n_j-1} F_j - \sum_{j1 < j2} (1/2)^{n_{j1}-1} 2F_{j1(j2)}F_{j2} \\ &+ \sum_{j1 < j2 < j3} (1/2)^{n_{j1}-1} 2^2 F_{j1(j2)}F_{j2(j3)}F_{j3} \\ &+ \dots + (-1)^{m+1} \sum_{j1 < j2 < \dots < jm} (1/2)^{n_{j1}-1} 2^{m-1} \left(\prod_{jk < jl} F_{jk(jl)} \right) F_{jm} \\ &+ \dots + (-1)^{n+1} (1/2)^{n_{j1}-1} 2^{j_1-1} F_{j1(j2)} \dots F_{jn-1(jn)} F_{jn}, \end{aligned} \quad (21)$$

where $j1 < j2$ indicates that $j2$ is farther back in the pedigree than $j1$.

The expressed opportunity for purging becomes

$$\begin{aligned} O_{Ei} &= \sum_j 2F_{i(j)}F_j - \sum_{j1 < j2} 2^2 F_{i(j)}F_{j1(j2)}F_{j2} \\ &+ \sum_{j1 < j2 < j3} 2^3 F_{i(j)}F_{j1(j2)}F_{j2(j3)}F_{j3} \\ &+ \dots + (-1)^{m+1} \sum_{j1 < j2 < \dots < jm} 2^m F_{i(j)} \left(\prod_{jk < jl} F_{jk(jl)} \right) F_{jm} \\ &+ \dots + (-1)^{n+1} 2^{jn} F_{i(j)}F_{j1(j2)} \dots F_{jn-1(jn)}F_{jn}. \end{aligned} \quad (22)$$

Likewise, the realized expressed purging can be written as

$$\begin{aligned} \Delta L_{Ei} &= \sum_{j=1} 2F_{i(j)}F_jq_a s^2 - \sum_{j1 < j2} 2^2 F_{i(j1)}F_{j1(j2)}F_{j2}q_a s^3 \\ &+ \sum_{j1 < j2 < j3} 2^3 F_{i(j1)}F_{j1(j2)}F_{j2(j3)}F_{j3}q_a s^4 \\ &+ \dots + (-1)^{m+1} \sum_{j1 < j2 < \dots < jm} 2^m F_{i(j1)} \left(\prod_{jk < jl} F_{jk(jl)} \right) F_{jm}q_a s^{m+1} \\ &+ \dots + (-1)^{n+1} 2^{jn} F_{i(j1)}F_{j1(j2)} \dots F_{jn-1(jn)}F_{jn}q_a s^{n+1} \end{aligned} \quad (23)$$

Note that for alleles with $s = 1$, $\Delta L_{Ei} = s q_a O_{Ei}$ (O_{Ei} is given in eq. 22).

The frequency of an allele in autozygotes is reduced through purging by about $s q_a OP$. If purging is from rare alleles of major effect, the product of q_a and OP is very small. So we conclude that the effect of purging is to cause minimal changes in the allele frequencies across ancestral generations in U.S. Jerseys and we are justified in computing O and O_E using equations (3) and (5) in the text.

Appendix 2: Computing Procedures (DG)

For Jersey cattle, contributions from an ancestor to a descendant (eq. 3) were calculated using a tabular method for computing additive relationships between animals (Henderson 1976) with the distinction that for each ancestor in question its previous genealogy was disregarded. This was done because we were only interested in alleles descending from a given ancestor to an individual in the current generation and not any allele that could be a copy of some ancestral sequence common to both.

Partial inbreeding coefficients were computed using a modification of a tabular method for computing additive relationships between animals (Lacy et al., 1996; Lacy 1997); however, for clearer understanding of the concept, we describe it using both a path analysis and the tabular method.

PATH ANALYSIS AND COMPUTATION OF PARTIAL INBREEDING COEFFICIENTS

The general principle of computing partial inbreeding coefficients, $F_{i(j)}$ via path analysis is to add all inbreeding paths from the ancestor j to individual i . The contribution of the ancestor through intermediate ancestor x is weighted by the fraction of j 's genome that is transmitted through x , $f_{x(j)}$. If x is inbred, this weight is augmented by the proportion of the total inbreeding coefficient of x that is due to j , $F_{x(j)}$. Note that in these analyses the earlier history of an ancestor is ignored.

The general formula for computing partial inbreeding coefficients by path analysis is

$$F_{i(j)} = \sum_x \sum_p (1/2)^n (f_{x(j)} + F_{x(j)}), \quad (24)$$

where the summations are over all ancestors common to the parents of i including j itself and over all paths to each ancestor (x); n is number of individuals in the path from i to x and $f_{x(j)}$ is a contribution of an allele in j to x ($\sum_k (1/2)^{n-1}$ with k indicating the sum over all paths to x from ancestor j).

Note that if $x = j$ then $f_{x(j)} = 1$ and $F_{x(j)} = F_j$. When j is the only common ancestor to parents of i , the formula for partial inbreeding becomes the formula for Wright's inbreeding coefficient: $F_{i(j)} = F_i$.

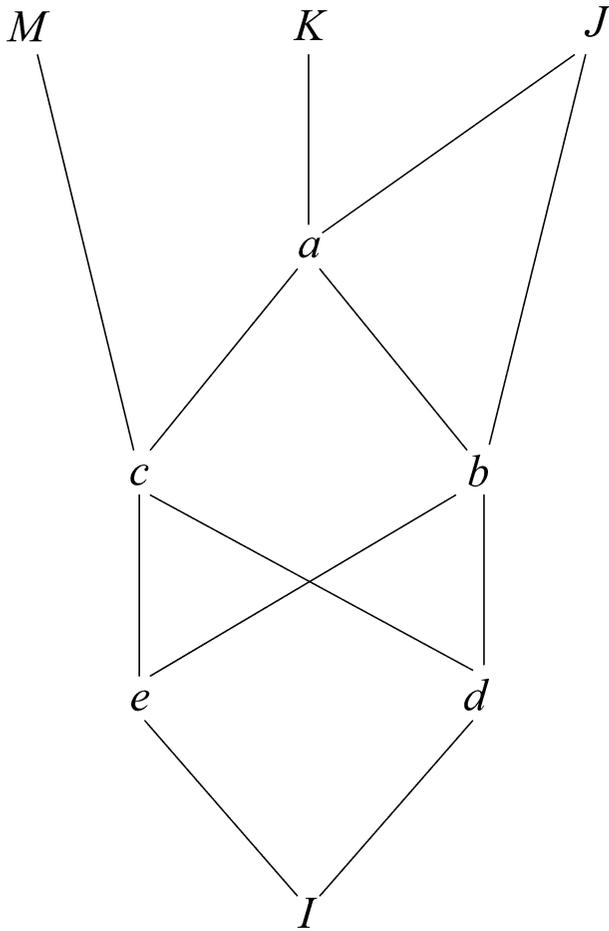


Figure 2. Path diagram of a sample pedigree.

Here, we describe the computation of partial inbreeding coefficient using a path diagram (Fig. 2). Note that in the simple pedigree given, we use capital letters to denote founding individuals (*M*, *K*, and *J*) and a descendant in the current generation (*I*). The following example also illustrates the principle for computing the normalized opportunity for purging.

Calculations are as follows:

Path Contribution to $F_{I(J)}$

For founder *J*

$$eca\underline{j}bd : \left(\frac{1}{2}\right)^6 = \frac{1}{64}$$

$$eb\underline{j}acd : \left(\frac{1}{2}\right)^6 = \frac{1}{64}$$

$$eca\underline{b}d : \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right) = \frac{1}{64}; \text{ } 1/2 \text{ is the proportion of } a \text{ received from } J$$

$$eb\underline{a}cd : \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right) = \frac{1}{64}; \text{ } 1/2 \text{ is the proportion of } a \text{ received from } J$$

$$eb\underline{d} : \left(\frac{1}{2}\right)^3 \times \left(\frac{3}{4} + \frac{1}{4}\right) = \frac{1}{8}; \text{ } 3/3 \text{ is the proportion of } b \text{ received from } J \text{ and } 1/4 \text{ is the partial inbreeding of } b \text{ due to } J$$

$$ec\underline{d} : \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{4}\right) = \frac{1}{32}; \text{ } 1/4 \text{ is the proportion of } c \text{ received from } J,$$

thus,

$$F_{I(J)} = \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{8} + \frac{1}{32} = \frac{7}{32}.$$

Likewise, for founder *K*

$$eca\underline{b}d : \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right) = \frac{1}{64}; \text{ } 1/2 \text{ is the proportion of } a \text{ received from } K$$

$$eb\underline{a}cd : \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right) = \frac{1}{64}; \text{ } 1/2 \text{ is the proportion of } a \text{ received from } K$$

$$eb\underline{d} : \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{4}\right) = \frac{1}{32}; \text{ } 1/4 \text{ is the proportion of } b \text{ received from } K$$

$$ec\underline{d} : \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{4}\right) = \frac{1}{32}; \text{ } 1/4 \text{ is the proportion of } c \text{ received from } K$$

$$F_{I(K)} = \frac{1}{64} + \frac{1}{64} + \frac{1}{32} + \frac{1}{32} = \frac{3}{32}$$

and, for founder *M*

$$ec\underline{d} : \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right) = \frac{1}{16} = F_{I(M)}; \text{ } 1/2 \text{ is the proportion of } c \text{ received from } M.$$

Thus, Wright's inbreeding coefficient is

$$F_I = F_{I(J)} + F_{I(K)} + F_{I(M)} = \frac{3}{8}, \text{ as expected because } J, K, \text{ and } M \text{ are unrelated.}$$

Individuals *b*, *d*, and *e* are inbred with

$$F_b = \frac{3}{16}, F_d = \frac{3}{16} \text{ and } F_e = \frac{3}{16},$$

thus $O_I =$

$$\left(\frac{1}{2}\right)^2 \times F_b + \left(\frac{1}{2}\right) \times F_d + \left(\frac{1}{2}\right) \times F_e = \frac{1}{16} + \frac{3}{32} + \frac{3}{32} = \frac{1}{4}.$$

Because *b* is the only inbred ancestor common to the parents of *I*, the expressed opportunity for purging in *I* is

$$O_{EI} = O_{EI(b)} = 2 \times F_{I(b)} \times F_b = \frac{1}{16},$$

and normalized expressed opportunity for purging,

$$OP_I = \frac{\frac{1}{16}}{\frac{3}{8}} = \frac{1}{6}.$$

From this pedigree, the inbred load in individual *I* is reduced by 1/6 because of purging, provided $s = 1$.

TABULAR METHOD FOR COMPUTING PARTIAL INBREEDING COEFFICIENTS

Partial inbreeding coefficients can be calculated using a modification of the tabular method (Lacy 1997) for calculating additive relationship or kinship matrices, via partial kinship matrices. A partial kinship matrix traces only alleles descending from an ancestor of interest. To compute partial inbreeding coefficients due to all ancestors in the population, a distinct partial kinship matrix is built for each ancestor. The sum of the partial kinship matrices corresponding to all possible ancestors unrelated to all other individuals in a pedigree other than their descendants (founders) yields a matrix of kinship coefficients between individuals. Here we illustrate the principle using the same simple pedigree as in the previous section. Recall that in the simple pedigree given, capitol letters are used to denote founding individuals (*M*, *K*, and *J*) and a descendant in the current generation (*I*). Partial kinships due to founder *J* for all individuals in the pedigree are given in Table 1.

The symmetric partial kinship matrix can be built in the following manner: Let an entry, $f_{xy} = f_{yx}$, in the *x*th row and *y*th column (or vice versa) of the partial kinship matrix, be the coefficient of kinship between *x* and *y* due to a particular founder; $x, y = 1, 2, \dots, n$ are individuals ordered from the oldest to the youngest. Note that $2f_{xy} = f_{x(y)}; f_{x(y)}$ is defined in the previous section. The matrix is filled in as follows:

Step 1: Founders (*J*, *K*, or *M*)

All f_{jz} and f_{zj} are set to 0, except the diagonal entry corresponding to the founder of interest that is set, $f_{jj} = 1/2$ (this represents the kinship of a founder with itself). Par-

Table 1. Partial kinship matrix for founder *J*. The entries in the table are the partial kinships of the individuals in the corresponding rows and columns.

Individual	<i>J</i>	<i>K</i>	<i>M</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>I</i>
<i>J</i>	1/2	0	0	1/4	3/8	1/8	1/4	1/4	1/4
<i>K</i>	0	0	0	0	0	0	0	0	0
<i>M</i>	0	0	0	0	0	0	0	0	0
<i>a</i>	1/4	0	0	1/4	1/4	1/8	3/16	3/16	3/16
<i>b</i>	3/8	0	0	1/4	1/2	1/8	5/16	5/16	5/16
<i>c</i>	1/8	0	0	1/8	1/8	1/8	1/8	1/8	1/8
<i>d</i>	1/4	0	0	3/16	5/16	1/8	5/16	7/32	17/64
<i>e</i>	1/4	0	0	3/16	5/16	1/8	7/32	5/16	17/64
<i>I</i>	1/4	0	0	3/16	5/16	1/8	17/64	17/64	23/64

tial kinships of all other possible founders ($j \neq z$) with all individuals (*x*) in the pedigree are $f_{zx} = 0$.

Step 2: Intermediate ancestors (*a*, *b*, *c*, *d*, and *e*)

Off-diagonal elements $f_{xw} = 1/2(f_{mw} + f_{pw})$, where *w* is older than *x* and *m* and *p* are the parents *x*. Diagonal elements $f_{xx} = f_{xj} + (1/2) \times f_{mp}$ (here, f_{xj} is the entry corresponding to the founder of interest (*j*) and *x*). This value is used instead of the 1/2 representing the coefficient of kinship with itself because we are only interested in alleles derived from founder *j*. f_{mp} is used instead of the inbreeding coefficient of individual *x*.

For example, $f_{Ja} = 1/2 (1/2 + 0) = 1/4$ because this is kinship between parent and offspring, but $f_{ac} = 1/2 (1/4 + 0) = 1/8$ because we are considering only alleles in *a* that descended from *J*; $f_{aa} = 1/4$ and $f_{cc} = 1/8$, but $f_{bb} = 3/8 + 1/8 = 1/2$, this is because *b* received 3/4 of its genome from *J* and *b* is also inbred due to *J* ($f_{Ja} = F_{b(J)} = F_b = 1/4$).

Step 3: Partial inbreeding coefficients

The partial inbreeding coefficient with respect to founder *j* of a descendant *i* (with parents *m* and *p*) is given by $F_{i(j)} = f_{mp}$.

For example, $f_{de} = F_{I(J)} = 7/32$.