

A New Estimate of Genetic Load from Inbreeding Data

Ademar Freire-Maia

SUMMARY

Inbreeding effect revealed as mortality and morbidity in general may be ascribed to deleterious mutations generally present at different loci of the ancestors common to each consanguineous mating. A new coefficient (E) is suggested to measure this effect. On the basis of this theory, a new method to estimate the load of mutations disclosed by inbreeding has been presented elsewhere, and is now extended. In this method, the load is estimated without introducing gene frequency notions. Genetic loads estimated through our method are in rather close agreement to those estimated through the classical theory of Morton, Crow and Muller (MCM).

In populations characterized by sexual reproduction and cross-fertilization, deleterious mutations are maintained by exceedingly low frequencies due to the action of negative selective forces. However, once the number of loci giving rise to such mutations is rather large, each individual in the population probably carries at least one mutation which, if made homozygous, would cause death or impair the zygote in some manner. This load of mutations, which in a randomly mating population is hardly detected, may be disclosed, at least in part, through inbreeding studies.

The estimate of the load of mutations disclosed by inbreeding has been made possible, firstly, due to the theory of inbreeding developed in a series of papers by Sewall Wright (cf., e.g., Wright, 1921, 1922), and, secondly, due to the theory of genetic equivalents developed by Morton, Crow and Muller (MCM) in 1956, and Morton (1960).

An Estimate of Inbreeding Effect

The coefficient of correlation between uniting gametes, which measures the frequency of loci made homozygous as a result of inbreeding, has been defined as the coefficient F of inbreeding (Wright, 1922). This coefficient is especially important in studies of genetic conditions due to deleterious recessive mutations at a single locus.

Let us assume, for an instant, the situation where there is one lethal equivalent per gamete (i.e., one *lethon*, according to Freire-Maia's 1964 terminology). Let us hypothesize also that (a) lethal equivalents are constituted by full lethals, and

¹ The research projects of our Department are performed under the sponsorship of the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), US National Institutes of Health, the Rockefeller Foundation, and the World Health Organization.

(b) lethals are alleles at the same locus of the *ancestors common to the consanguineous couple* (for convenience, *common ancestors*). In this hypothetical situation, the probability of homozygosis (death) in the inbred offspring would be given directly by coefficient F . However, it is clear that lethal equivalents are not constituted by alleles at the same locus of the common ancestors.

We would like to propose here an opposite assumption regarding the loci occupied by lethal equivalents, namely, that, in general, deleterious mutations carried by common ancestors are located at different loci.² This would hold true, not only when one thinks over mutations carried by each common ancestor, but also when one considers mutations carried by different common ancestors. In a population in equilibrium characterized by an average of 1 lethon, the probability of homozygosis by descent for any one lethal allele would therefore be given by $F/2n$, where F is Wright's coefficient of inbreeding and n is the number of common ancestors. Since there are $2n$ loci involved, the probability of escaping death due to inbreeding is equal to $(1 - F/2n)^{2n}$. The inbreeding effect may then be measured by

$$E = 1 - (1 - F/2n)^{2n}. \quad [1]$$

In control samples, where the coefficient of inbreeding assumes very low values, being practically equal to zero, the ratio $F/2n$ also approaches to zero; in this case, $E = 0$.

Coefficient E , which especially suits to studies of genetic load, may be indirectly calculated on the basis of coefficient F (equation [1]) provided the number of common ancestors is known and no inbred ancestor occurs. In cases where inbred ancestors are involved, a more accurate estimate of coefficient E should be calculated directly from the pedigrees.

It is to be noted that equation [1] involves an approximation, namely, that in general only one common ancestor carries a particular detrimental gene. However, this approximation certainly is not of importance. As stated earlier, it is generally accepted that the frequency of individual lethal genes in the populations is rather low (according to Morton, 1964, " $a_i + b_i F$ is small for each locus and the number of loci is large"). Therefore, the contribution to the inbreeding effect of particular lethal genes present in two (or more) unrelated common ancestors is negligible, especially if compared to the total inbreeding effect, which is to be measured. Let us assume, for instance, that a lethal gene a has a population frequency $q = 0.01$, where $1 - q = p$. The probability that only one ancestor, common to a first cousin marriage, carries this lethal allele is $4p^3q = 0.0388$, the probability of death in the inbred offspring being $4p^3qF/2n = 0.0006$ (where n is the number of common ancestors, as before). Let us see now the situation where both the common ancestors carry

² Slatis (1960) described a different coefficient, based on a similar principle. Specifically speaking on the European Bison, Slatis defined his coefficient as "the likelihood of genetic death if each ancestor in the foundation herd possessed a single recessive lethal gene, l , and if each of these lethals was at a different locus".

the same lethal gene. The probability of this event is equal to $4p^2q^2 = 0.0004$; by taking into consideration that the common ancestors' offspring may not be homozygous for the allele in question, the probability of death in the inbred offspring will be $4p^2q^2 \cdot 0.0278 = 0.0001$. Therefore, the probability of death, due to the presence of the lethal allele in only one common ancestor is 60 times higher than the same probability when both common ancestors carry the lethal. This last probability is therefore more than 60 times lower than the total inbreeding effect. If one assumes $q = 0.001$, the probability of death due to only one common ancestor carrying the lethal allele would be 600 times greater than the corresponding value from both the common ancestors carrying the lethal; this last probability is therefore more than 600 times lower than the total inbreeding effect.

An Estimate of Genetic Load

Based on the reasonings leading to equation [1], a new method to estimate the load of mutations disclosed by inbreeding has been presented elsewhere (Freire-Maia and Freire-Maia, 1964, 1965a). As in MCM's theory, this method assumes that different causes of death (genetic and environmental) are independent in action, i. e., "nonsynergistic" in the sense given by Muller (1950). This assumption seems sound (Freire-Maia, 1964). The basic equation in this method may be expressed by

$$S_f = (1-E)^B, \quad [2]$$

where S_f is the probability of escaping death due to inbreeding, and B (as in MCM's theory) is a slight underestimate of the mean number of lethons (i. e., Σqs). Equation [2] may also be expressed as $S_f = (1-F/2n)^{2nB}$ (Freire-Maia and Freire-Maia, 1964, 1965a).

For those who have no opportunity to use an electronic computer, the estimate of B through equation [2] affords a great advantage, since it can be easily obtained with the aid of only a calculating machine (or even without it). The results are equivalent to those obtained through MCM's method, as may be shown. If one introduces into our formula a simplification similar to the one made by MCM in the development of their method, the formula would turn out to be $S_f = e^{-BF}$, since $\log_e S_f = B \log_e (1-E)$. Compare that equation to the fundamental one in MCM'S theory, i. e., $S = e^{-A-BF}$. The only difference is that our simplified formula measures the probability S_f of escaping death due to inbreeding, whereas MCM's measures the probability S of escaping death due to all the causes (genetic and environmental). Since both methods intend to detect the genetic load disclosed by inbreeding, an estimate of A is not essential for the estimate of B ; both formulas lead therefore to equivalent estimates of the number of lethons. It should be kept in mind, however, that the simplified equation involves an approximation which is not present in equation [2]. When there is no interest in obtaining a very reliable estimate of B , MCM's and our theory provide a simple approach which is exactly the same in both meth-

ods, namely $B = [\log_e (S_o/S_i)] / F$, where S_o and S_i are the rates of survivors in the control and inbred samples, respectively (see Morton, 1964, and Freire-Maia and Freire-Maia, 1964).

In our theory, we tried to solve the problem of simultaneous homozygosis without making allowance for the frequency of individual lethal genes in the population. Instead, we tried to estimate the frequencies, in the ancestors common to *each* consanguineous couple, of lethals which are carried by them. This has seemed justified since there is no matter if the ancestors common to *different* consanguineous matings carry the same or different lethals, provided a lethal occurs no more than once in the common ancestors. In other words, the inbreeding effect may be the same even for genes with different frequencies, provided the number of lethons is maintained constant, and each gene occurs no more than once in the common ancestors. This is explained by the fact that, by definition, a given lethal gene is identical to any other one regarding the behaviour and effect, and that it does not matter whether a genetic death is caused by a specific lethal gene or by any other lethal gene.

Although formula [2] provides an estimate of the damage brought to expression through inbreeding, it does not provide an estimate of the damage in the inbred sample which is *not* due to inbreeding. Instead of it, MCM used an estimate of the expressed damage in a randomly mating population (A).

Our theory also provides an estimate of A , as follows. In any inbred sample, let $(1-pM_i)$ be the probability of escaping death due to inbreeding, and $(1-pA_i)$ be the probability of escaping death due to other factors (genetic and environmental). Therefore, the fraction of survivors to all causes of death will be given by

$$S = 1 - pA_i - pM_i + (pA_i)(pM_i). \quad [3]$$

Since so far there is no method to know the value of (pA_i) , an estimate of it, namely, the probability of death in a randomly mating sample (pA) , will be used. As can be seen from equation [3], by assuming $F = 0$ (therefore $pM_i = 0$), this probability is equal to A , the expressed damage in the randomly mating sample. By substituting (pA_i) by its estimate A , equation [3] turns out to be

$$1 - pM_i = S/(1-A). \quad [4]$$

Since the probability of escaping the inbreeding effect is assumed to be given both by equations [2] and [4], by equalling their terms results that

$$S = (1-A)(1-E)^B \quad [5]$$

and

$$\log_e S = \log_e (1-A) + B \log_e (1-E). \quad [6]$$

From equation [6], estimates of A and B may then be obtained by the weighted regression on $\log_e (1-E)$ of the natural logarithm of the number of survivors. According to maximum-likelihood theory, the appropriate weights are $NS/(1-S)$, where S is the expected frequency of survivors and N is the total number of obser-

vations (cf. MCM). As in MCM's theory, the observed value of S must be used as a trial value, in order the weights may be obtained by iteration. When there is a small number of deaths in the noninbred groups, and the inbreeding levels are low, virtually the same estimates of A and B may be obtained from the simple approximations

$$-\log_e S = A + BE \quad [7]$$

or

$$1 - S = A + BE. \quad [8]$$

Note that equation [7] is rather similar to MCM's fundamental equation, namely, $-\log_e S = A + BF$.

Equation [6] has been applied to data from the literature, as well as to some theoretical situations, and the results have been compared to those obtained through MCM's theory (Table). One may conclude that the results regarding to the estimates of B are equivalent. However, MCM's method systematically led to overestimates of the expressed damage in the control sample ($F = O$), thus confirming the previous observation by Freire-Maia (1964). These overestimates may be slight, if the damage is low; however, with increasing damage there is an increasing overestimate, which may even lead to nonsense values, namely higher than 1. For these situations, Freire-Maia (1964) suggested that, in randomly mating populations ($F = O$), the expressed damage can be more accurately estimated by $(1 - e^{-A})$. This leads, however, to a curious situation, since in MCM's theory the amount of expressed damage in a randomly mating population ($F = O$) is defined by A , whereas its best estimate would be given by $(1 - e^{-A})$. In our theory (equation [6]), this damage is given by A , and its best estimate is also given by A .

The values of A obtained through MCM's method are practically the same from $\log_e (1 - A)$ in equation [6] [let us make $\log_e (1 - A) = SEPT$ (see Table)]. Whereas in MCM's theory A is the intercept of the regression in a semi-logarithmic scale of the number of survivors on F , in our theory $SEPT$ is the intercept of the regression in a logarithm scale of the frequency of survivors on the reciprocal of E . Therefore, A in MCM's theory and $SEPT$ in ours, have *per se* no biological meaning, unless they are good estimates of the expressed damage in the control ($F = O$) sample. This will occur only when the damage is low (Table). In any circumstance, however, the damage may be properly estimated by A from equation [6].

Naturally the comparative evaluation of differences between MCM's and our theory does not reflect on the excellence and refinement of the MCM's theory. As a matter of fact, the results obtained from both methods generally present a rather good agreement, especially regarding the estimates of B . It seems very important to emphasize this point, since the theories are based on different, but not opposite, assumptions.

Current methods for calculating the average number of lethons acting before birth depend on relatively unreliable data based on the frequency of prenatal deaths. Taking in consideration this fact, Frota-Pessoa (1966) suggests a variant for the meth-

Table. Amount of expressed damage in a randomly mating population (*A*) and load of mutations disclosed by inbreeding (*B*)

Estimates of <i>A</i>			Estimates of <i>B</i>	
<i>AFM</i>	<i>AMCM</i>	<i>SEPT</i>	<i>BFM</i>	<i>BMCM</i>
0.018	0.018	0.018	0.19 ± 0.18	0.20 ± 0.18
0.019	0.019	0.019	0.55 ± 0.37	0.53 ± 0.36
0.020	0.021	0.021	0.20 ± 0.16	0.20 ± 0.16
0.032	0.033	0.033	0.13 ± 0.10	0.13 ± 0.10
0.041	0.042	0.042	−0.33 ± 0.02	−0.32 ± 0.07
0.081	0.085	0.085	0.75 ± 0.70	0.72 ± 0.68
0.096	0.101	0.101	0.39 ± 0.49	0.39 ± 0.49
0.173	0.189	0.189	0.98 ± 0.41	1.00 ± 0.40
0.181	0.200	0.200	1.58 ± 0.32	1.59 ± 0.32
0.200	0.223	0.223	−0.49 ± 0.43	−0.50 ± 0.43
0.272	0.318	0.318	1.31 ± 0.60	1.32 ± 0.60
0.279	0.327	0.327	1.94 ± 0.57	1.95 ± 0.58
0.282	0.331	0.331	0.53 ± 0.47	0.53 ± 0.47
0.336	0.410	0.410	0.93 ± 0.54	0.94 ± 0.55
0.381	0.480	0.480	1.11 ± 0.83	1.12 ± 0.84
0.433	0.568	0.568	2.53 ± 0.78	2.56 ± 0.79

NOTES: The data have been obtained from a large sample from the State of Espírito Santo, Brazil (cf. Freire-Maia, 1969).

AMCM and *BMCM* = estimates according to MCM's method;

AFM and *BFM* = estimates according to our method.

For a definition of *SEPT*, see text.

od by MCM, which avoids using such data at all. Making L_f = average number of offspring born alive to the couples in the $F = f$ class, per couple per year of cohabitation; C = average number of conceptions per couple per year of cohabitation; and $L_f/C = S_f$, Frota-Pessoa added $\log_e C$ to both sides of MCM's fundamental equation and obtained $-\log_e L_f = A' + BF$, where $A' = A + \log_e C$.

Our method also provides estimates of genetic load from data on the frequencies of alive births (which are notoriously more reliable than those on prenatal deaths). From equation [6] one may directly conclude that

$$\log_e L_f = \log_e C(1-A) + B \log_e (1-E). \quad [9]$$

Although leading to more reliable estimates of B , Frota-Pessoa's variant and equation [9], do not provide better estimates of A than do MCM's method or equation [6], respectively. As in Frota-Pessoa's variant, equations [6] and [9] may also be employed to check if inbreeding acts with different intensities for the two sexes (for details, see Frota-Pessoa, 1966). As a matter of fact, the same holds true for the original MCM's method. Another variant to MCM's method has also been sug-

gested by Frota-Pessoa (1966) for evaluating the load of mutations due to sex-linked genes which manifest themselves in the inbred female offspring. As previously, this suggestion may also be easily used in our method.

The load of mutations expressed as undetected abortions may be disclosed through a method suggested by Freire-Maia and Freire-Maia (1965b).

Genetic loads may be easily calculated from equation [6] through a program, written in FORTRAN language by Dr. Ivan Jelinek Kantor, for the IBM 1620 electronic computer (see Appendix to the paper by Freire-Maia and Freire-Maia, 1965a). A copy of the program may be freely obtained on request to the Centro de Computação Eletrônica, Escola Politécnica, Universidade de São Paulo, São Paulo, Brasil. Investigators who do not have access to an electronic computer may submit their data to be processed.

ACKNOWLEDGMENT. The author is deeply grateful to his brother, Dr. Newton Freire-Maia, for permanent guidance and encouragement.

This paper has been mainly made at the Centro de Computação Eletrônica da Universidade de São Paulo, where the author remained during one year. Thanks are due to the whole team of the center and particularly to Drs. Ernesto Vita Jr. and J. Dion de M. Teles, and Mr. Ivan Jelinek Kantor. A part of this paper has also been made when the author was a Professor at the Faculdade de Filosofia, Ciências e Letras de Araraquara. For discussions in connexion with the preparation of this paper, true thanks are due to Dr. Newton Freire-Maia, Dr. Newton E. Morton, Dr. Henrique Krieger, Lic. Sci. Francisco A. Marçallo, and Dr. Warwick E. Kerr. For reading and criticizing the manuscript, the author is also indebted to Drs. H. M. Slatis, N. E. Morton, N. Freire-Maia, W. E. Kerr, and H. Krieger. Of course, none of these friends are responsible for the concepts here sustained.

References

- FREIRE-MAIA A. (1969). Human genetic studies in areas of high natural radiation. I. Methodology. *Acta Genet. Med. Gemellol.*, **18**: 175-212.
- FREIRE-MAIA N. (1965a). Estimativa do efeito genético dos casamentos consanguíneos. *Cienc. Cult.*, **17**: 539-547.
- FREIRE-MAIA N. (1965b). Genetic load and undetected abortions. *Amer. J. Hum. Genet.*, **17**: 93-94.
- FREIRE-MAIA N. (1964). On the methods available for estimating the load of mutations disclosed by inbreeding. *Cold Spring Harb. Symp. Quant. Biol.*, **29**: 31-40.
- FREIRE-MAIA A. (1964). Estimate of the genetic load disclosed by inbreeding. *Genetics*, **50**: 527-529.
- FROTA-PESSOA O. (1966). Lethal equivalents acting before birth. *Amer. J. Hum. Genet.*, **18**: 57-61.
- MORTON N. E. (1960). The mutational load due to detrimental genes in man. *Amer. J. Hum. Genet.*, **12**: 348-364.
- CROW J. F., MULLER H. J. (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proc. Nat. Acad. Sci. USA*, **42**: 855-863.
- MULLER H. J. (1950). Our load of mutations. *Amer. J. Hum. Genet.*, **2**: 111-176.
- SLATIS H. M. (1960). An analysis of inbreeding in the European bison. *Genetics*, **45**: 275-287.
- WRIGHT S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, **6**: 111-123.
- (1922). Coefficients of inbreeding and relationship. *Amer. Nat.*, **56**: 330-338.

RIASSUNTO

Gli effetti della consanguineità su mortalità e morbilità possono, in genere, essere imputati alle mutazioni dannose solitamente presenti su diversi loci degli antenati comuni di ciascuna coppia consanguinea. Per misurare tali effetti, viene suggerito un nuovo coefficiente (E). Sulla base di questa teoria, è stato presentato altrove, e viene qui ampliato, un nuovo metodo per stimare il peso delle mutazioni rivelate dalla consanguineità, senza ricorrere alle frequenze geniche. Il peso genetico stimato con questo metodo concorda con quello stimato con la teoria classica di Morton, Crow e Muller (MCM).

RÉSUMÉ

Les effets de la consanguinité sur la mortalité et la morbidité peuvent, en général, être attribués aux mutations nuisibles de différents loci des ancêtres communs du couple consanguin. Un nouveau coefficient (E) est proposé pour la mesure de ces effets. Sur la base de cette théorie une nouvelle méthode, déjà présentée, est ici développée, pour estimer le poids des mutations révélées par la consanguinité, sans faire recours aux fréquences géniques. Le poids génétique ainsi estimé concorde avec l'estime par la méthode classique de Morton, Crow et Muller (MCM).

ZUSAMMENFASSUNG

Der Einfluss der Blutsverwandtschaft auf Mortalität lässt sich im allgemeinen auf schädliche Mutationen zurückführen, die gewöhnlich an verschiedenen "loci" der gemeinsamen Vorfahren jedes blutsverwandten Paares vorkommen. Um diese Folgen zu messen, wird ein neuer Koeffizient (E) vorgeschlagen. Auf Grund dieser Theorie wurde bereits anderswo eine neue Methode angegeben, die hier erweitert wird, um ohne Zuhilfenahme der Genfrequenz den Einfluss der durch die Blutsverwandtschaft aufgedeckten Mutationen zu bestimmen. Der auf diese Weise geschätzte Erbeinfluss stimmt mit den Werten der klassischen Theorie nach Morton, Crow und Muller (MCM) überein.

Dr. A. FREIRE-MAIA, Departamento de Genética, Faculdade de Ciências Médicas e Biológicas, Botucatu, Est. São Paulo, Brazil.