Genet. Res., Camb. (2004), **83**, *pp.* 113–120. *With 6 figures.* © 2004 Cambridge University Press
DOI: 10.1017/S0016672303006633 Printed in the United Kingdom

On the prediction of simultaneous inbreeding coefficients at multiple loci

JULES HERNÁNDEZ-SÁNCHEZ*, CHRIS S. HALEY AND JOHN A. WOOLLIAMS
Roslin Institute, Midlothian, EH25 9PS, Scotland, UK

(*Received 7 August 2003 and in revised form 31 October and 24 November 2003*)

Summary

A new deterministic method for predicting simultaneous inbreeding coefficients at three and four loci is presented. The method involves calculating the conditional probability of IBD (identical by descent) at one locus given IBD at other loci, and multiplying this probability by the prior probability of the latter loci being simultaneously IBD. The conditional probability is obtained applying a novel regression model, and the prior probability from the theory of digenic measures of Weir and Cockerham. The model was validated for a finite monoecious population mating at random, with a constant effective population size, and with or without selfing, and also for an infinite population with a constant intermediate proportion of selfing. We assumed discrete generations. Deterministic predictions were very accurate when compared with simulation results, and robust to alternative forms of implementation. These simultaneous inbreeding coefficients were more sensitive to changes in effective population size than in marker spacing. Extensions to predict simultaneous inbreeding coefficients at more than four loci are now possible.

1. Introduction

The inbreeding coefficient F is one of the most widely used parameters in population and quantitative genetics. Statistically, F is the probability that the two alleles at a locus in an individual are identical by descent (IBD), and two alleles are IBD if they descend from, and are exact copies of, the same ancestral allele (Malécot, 1948).

An increase in F lowers genetic variance within populations, and can lead to inbreeding depression and to a higher frequency of genetic diseases (Wright, 1977). These adverse consequences of high F are of prime interest in a wide range of biological sciences. For example, genetic disorders, such as cardiovascular disease, are more common in small and isolated populations than in large and open ones (Wright *et al.*, 1999), and breeding programmes aim at maximizing genetic gain whilst restricting F (Bijma *et al.*, 2000).

Wright (1922) calculated F for each individual by tracing pedigree loops. Malécot (1948) overcame the computational difficulties of Wright's method in large and complex pedigrees by expressing F as a population average. Currently, marker data can be combined with pedigree records to calculate F at each chromosomal position for each individual (Thompson, 1994).

In the above, F denoted the expected inbreeding at a single locus, a single individual or a single population. Weir & Cockerham (1969, 1974) developed a recurrence equation to predict the average inbreeding coefficient at two loci simultaneously, in a population. No extensions were available for three or more loci. Such multilocus inbreeding coefficients were key elements in a method for predicting co-ancestry at a locus, given population history and marker data, but without pedigree information (Hernández-Sánchez, Haley & Woolliams, unpublished). Moreover, they can also be useful in mapping genes, designing breeding programmes, detecting population structure, and studying linkage disequilibrium.

We have developed an efficient and simple approximation, based on multiple regression models, and using the theory of digenic measures of Weir and Cockerham, to predict simultaneous F at three and four loci. We describe this method, test its accuracy,

* Corresponding author. Tel: +44 (0)131 527 4449. Fax: +44 (0)131 440 0434. e-mail: Jules.Hernandez@bbsrc.ac.uk

Table 1. *Parameters and variables*

F, F_i	Inbreeding coefficient, or IBD probability, at any locus or at locus i, respectively
F_{ij}	Simultaneous inbreeding coefficient at loci i and j
$F_{i\|jk}$	Conditional IBD at locus i given IBD at loci j and k
η_{ij}, $\eta_{i,jk}$, $\eta_{ij,jk}$	$=F_{ij}-F_iF_j$, $F_{ijk}-F_iF_{jk}$ and $F_{ijk}-F_{ij}F_{jk}$, respectively
$y_i(\bar{y}_i)$	Indicator variable: 1 if locus i is IBD, 0 otherwise (Expected y_i)
$y_{ij}(\bar{y}_{ij})$	Indicator variable: 1 if loci i and j are IBD, 0 otherwise (Expected y_{ij})
σ_i^2	Variance of IBD at locus i, equal to $F_i(1-F_i)$
$\sigma_{i,j}$, $\sigma_{i,ij}$, $\sigma_{i,jk}$, $\sigma_{ij,ik}$	Covariances among loci i, j and k
N	Effective and actual sample size
T	Total number of generations
d	Vector of interloci distances in cM

and study the sensitivity of simultaneous inbreeding to changes in population parameters.

2. Materials and methods

Table 1 shows the main variables and parameters used, where inbreeding coefficient and IBD probability are synonymous. The inbreeding coefficient at locus A is F_A. The simultaneous inbreeding coefficient at two loci, A and B, is $F_{AB}=F_AF_{B|A}$, where $F_{B|A}=F_B+\eta_{AB}/F_A$ is the conditional probability of locus B being IBD given that locus A is IBD, and where $\eta_{AB}=F_{AB}-F_AF_B$ is the identity disequilibrium coefficient (Weir & Cockerham, 1969). We used exact recurrence formulae to calculate F_{AB} directly for different population histories (Weir & Cockerham, 1969, 1973, 1974; Weir *et al.*, 1980). The Appendix contains the original formulae for F_{AB} when a monoecious population is mating and selfing at random.

(i) *Prediction of three-loci inbreeding coefficients*

The simultaneous inbreeding coefficient at three neutral loci A, B and C, where C is located between A and B, is

$$F_{ABC}=F_{AB}F_{C|AB}, \tag{1}$$

where $F_{C|AB}$ is the conditional probability of IBD at locus C given IBD at loci A and B simultaneously. The effect of linkage is already accounted for in the two-loci theory of Weir and Cockerham. The conditional probability in (1) is obtained assuming IBD at loci A and B, and then using this information to predict IBD at a linked locus C. This method obtains IBD probabilities as means of indicator random variables y_i taking a value of 1 when locus i is IBD or 0 otherwise, and also $y_{ij}=1$ when loci i and j are simultaneously IBD or 0 otherwise. Hence, one can calculate an approximation to $F_{C|AB}$ using the following regression equation:

$$F_{C|AB}\approx F_C+\mathbf{X}'\mathbf{R}, \tag{2}$$

where $\mathbf{R}$ is a vector of partial regression coefficients, and

$$\mathbf{X}'=[y_A-\bar{y}_A,\ y_B-\bar{y}_B]=[1-F_A,\ 1-F_B],$$

is the transposed vector of corrected indicator variables, since this probability is conditional on loci A and B being IBD, i.e. $y_A=y_B=1$. The vector $\mathbf{R}$ is obtained as $\mathbf{V}^{-1}\mathbf{G}$, where $\mathbf{V}$ is a covariance matrix of IBD among loci A and B, and $\mathbf{G}$ a vector of covariances between IBD at loci A or B and IBD at locus C (see upper left 2×2 corner of matrix $\mathbf{V}$, and the first two elements from vectors $\mathbf{G}'$ and $\mathbf{X}'$ in Table 2).

The diagonal elements of $\mathbf{V}$ are the variances $\sigma_i^2=\overline{y_i^2}-(\bar{y}_i)^2=F_i(1-F_i)$, since $\overline{y_i^2}=\bar{y}_i=F_i$, where $i=$A or B. The off-diagonal elements of $\mathbf{V}$ are the covariances $\sigma_{i,j}=\bar{y}_{ij}-\bar{y}_i\bar{y}_j=F_{ij}-F_iF_j=\eta_{ij}$ for $i\neq j=$A or B. The transpose of $\mathbf{G}$ is $\mathbf{G}'=[\eta_{AC},\ \eta_{CB}]$. Therefore, $F_{C|AB}$ can be approximated as

$$F_{C|AB}\approx F_C + \frac{[(1-F_A),\ (1-F_B)]\begin{bmatrix} F_B(1-F_B) & -\eta_{AB} \\ -\eta_{AB} & F_A(1-F_A)\end{bmatrix}\begin{bmatrix}\eta_{AC}\\ \eta_{BC}\end{bmatrix}}{F_A(1-F_A)F_B(1-F_B)-\eta_{AB}^2}, \tag{3a}$$

which simplifies to:

$$F_{C|AB}\approx F_C+(1-F)(\eta_{AC}+\eta_{CB})/(F(1-F)+\eta_{AB}), \tag{3b}$$

when $F_A=F_B=F$. This prediction model does not include a regression term corresponding to loci A and B being simultaneously IBD, i.e. $y_{AB}=1$, because the additional covariance in $\mathbf{G}$, i.e. $\sigma_{AB,C}=F_{ABC}-F_{AB}F_C$, requires F_{ABC}.

Two special cases can be studied here. First, if loci A, B and C are completely unlinked to each other, and assuming all identity disequilibrium parameters are zero on average (which is true for a very large population with a mating structure that excludes selfing), then (3a) and (3b) simplify to $F_{C|AB}=F_C$, reflecting the fact that IBD at locus C does not depend on IBD at

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

Table 2. *Matrix* **V** *of covariances of IBD among loci A, B and D, and vector* **G** *of covariances between IBD at these loci and IBD at locus C. This table provides the necessary elements to predict both* $F_{C|AB}$ *and* $F_{C|ABD}$. *See Table 1 for relevant notation*

Matrix **V** of co-variances among loci *A*, *B* and *D*

V	*A*	*B*	*D*	*AB*	*AD*	*BD*
A	$F_A(1-F_A)$	η_{AB}	η_{AD}	$F_{AB}(1-F_A)$	$F_{AD}(1-F_A)$	$\eta_{A,BD}$
B		$F_B(1-F_B)$	η_{BD}	$F_{AB}(1-F_B)$	$\eta_{B,AD}$	$F_{BD}(1-F_B)$
D			$F_D(1-F_D)$	$\eta_{D,AB}$	$F_{AD}(1-F_D)$	$F_{BD}(1-F_D)$
AB				$F_{AB}(1-F_{AB})$	$\eta_{AB,AD}$	$\eta_{AB,BD}$
AD					$F_{AD}(1-F_{AD})$	$\eta_{AD,BD}$
BD						$F_{BD}(1-F_{BD})$

Vector **G**′ of covariances between loci *A*, *B* or *D* and locus *C*

G′	*A*	*B*	*D*	*AB*	*AD*	*BD*
C	η_{AC}	η_{BC}	η_{DC}	$\eta_{C,AB}$	$\eta_{C,AD}$	$\eta_{C,BD}$

Vector **X**′ of corrected indicator variables of IBD

A	*B*	*D*	*AB*	*AD*	*BD*
$1-F_A$	$1-F_B$	$1-F_D$	$1-F_{AB}$	$1-F_{AD}$	$1-F_{BD}$

loci A and B in the absence of linkage. Nevertheless, (3a) and (3b) can account for random associations between unlinked loci (e.g. those caused by drift) because the terms η_{ij} can be greater than zero even without linkage, and hence allowing for $F_{C|AB} > F_C$. Second, if loci A and B are completely linked so that $F_A = F_B = F$, $\eta_{AC} = \eta_{BC}$ and $\eta_{AB} = F(1-F)$, then (3a) reduces to $F_{C|AB} = F_{C|A} = F_C + \eta_{AC}/F$, which appeared above in the two-loci case. If locus C lies in between A and B, then $\eta_{AC} = F(1-F)$ and $F_C = F$, leading to $F_{C|AB} = 1$, reflecting the fact that conditional on A and B being IBD, C must be also IBD with probability 1 if linkage is complete.

(ii) *Prediction of four-loci inbreeding coefficients*

A similar procedure can be used to predict the simultaneous inbreeding coefficient at four loci, F_{ABCD}. Assuming the locus order A-B-C-D, then

$$F_{ABCD} = F_{ABD} F_{C|ABD}. \quad (4)$$

The parameter F_{ABD} is equivalent to the parameter F_{ABC} of the previous section. The conditional probability $F_{C|ABD}$ can be calculated using an extended version of model (2) that includes the interactions due to loci pairs AB, AD and BD, as well as the main effects of each locus (A, B and D). The new vector **R** of partial regression coefficients was obtained with the full matrix **V** and the full vector **G** given in Table 2. This table also shows the full vector **X** of corrected indicator variables for IBD. The additional elements in matrix **V** are variances and covariances of the form σ^2_{ij}, and $\sigma_{i,ik}$, $\sigma_{i,jk}$, $\sigma_{ij,ik}$, respectively, and the additional elements in vector **G** are covariances of the form $\sigma_{ij,C}$, where $i \neq j \neq k =$ A, B, D. In the special case of completely unlinked loci in a large population, the vector **G** is zero, and therefore $F_{C|ABD} = F_C$, because the conditioning is over uninformative markers. On the other hand, if C is completely linked to any of the other three loci, then $F_{C|ABD} = 1$.

(iii) *Simulation results versus deterministic predictions*

Empirical observations of F_{ABC} and F_{ABCD} were obtained with computer simulations. In each replicate, a population was created with N unrelated and non-inbred founders, and its evolution by drift was monitored over 100 discrete generations (T). The size N remained constant across generations, with random mating and selfing. Under these circumstances, N denoted both the actual and the effective population size. The IBD status of three (ABC) and four loci (ABCD) was scored as 1/0 in each individual within each generation, and averages were taken over 1000 (or 10 000) replicates. The distances between adjacent loci, in centimorgans (cM), were given in vector **d**. These distances were transformed to recombination rates using Haldane's mapping function (Haldane, 1919). Empirical results were compared against deterministic predictions obtained with equations (1) and (4). An exact knowledge of population parameters (**d**, N and T) was assumed.

Other population histories were considered to demonstrate the generality of (1) and (4). For example,

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

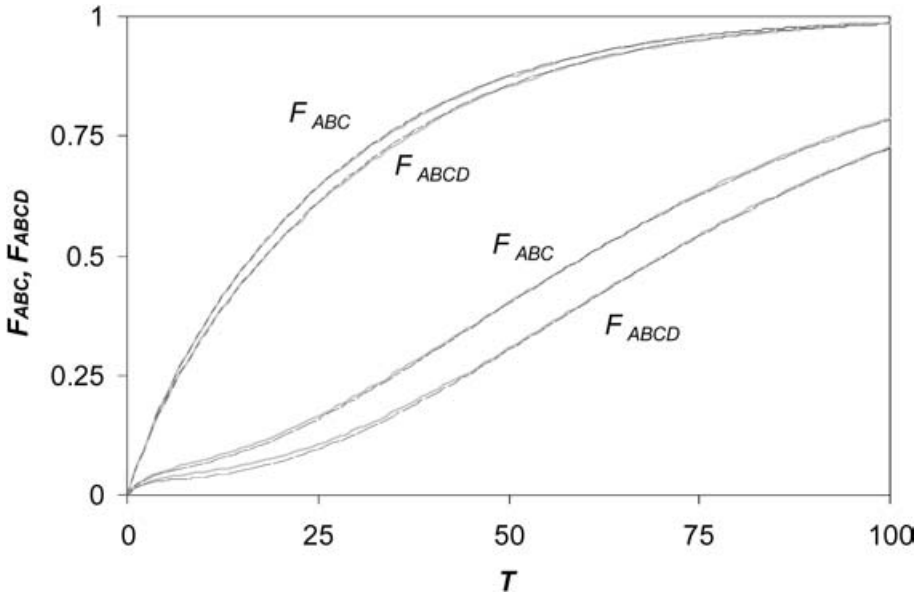

Fig. 1. Empirical averages over 10 000 replicates (continuous lines) versus deterministic predictions (dashed lines) of F_{ABC} and F_{ABCD}, plotted against generations (T). A monoecious population mated at random including selfing. Two different scenarios were considered: $N=10$ and $\mathbf{d}=[1, 1, 1]$ cM in the two upper graphs, and $N=20$ and $\mathbf{d}=[10, 10, 10]$ cM in the two lower graphs.

we also studied the behaviour of simultaneous inbreeding coefficients using the previous population model excluding selfing, and when $N=\infty$ and there were equal proportions of random mating and random selfing. In the latter scenario, computer simulations were generated for a population with 1000 individuals, in which 500 random individuals selfed each generation whereas the other 500 mated at random conditional on being non-inbred and unrelated.

The issues studied regarding (1) and (4), were: (a) accuracy of predictions, and robustness of method of implementation for a finite monoecious population with or without selfing, and also for an infinite population with a fixed proportion of selfing, and (b) sensitivity of F_{ABC} and F_{ABCD} to changes in N or $\mathbf{d}$ in a finite monoecious population with random mating and selfing.

3. Results

(i) *Accuracy of predictions, and robustness of implementation*

Deterministic predictions of F_{ABC} and F_{ABCD} obtained with (1) and (4), respectively, are compared against simulation results (averages of 10 000 replicates) in a monoecious population mating at random with (Fig. 1), or without selfing (Fig. 2). The predictions were more accurate with tight linkage (1 cM spacing), than with loose linkage (10 cM spacing), although differences were small.

This high accuracy was achieved without fitting an interaction between loci A and B in the model for predicting $F_{C|AB}$, because F_{ABC} is required in such an interaction. Although an iterative scheme could be developed, the accuracy of the approximation makes such a scheme unnecessary, at least when linkage is tight. Similarly, two but not three loci interactions were fitted to the model that predicted $F_{C|ABD}$.

In general, we observed slightly larger prediction errors for F_{ABCD} than for F_{ABC}. One potential reason for this is that F_{ABC} is included in the model for predicting F_{ABCD}, and therefore the prediction error of F_{ABC} is also included in the prediction error of F_{ABCD}.

Fig. 3 compares deterministic and empirical predictions of inbreeding at one, two, three and four loci in an infinite population undergoing 50% random mating and 50% selfing. Empirical predictions were obtained averaging over 1000 replicated histories of a large population ($N=1000$). It can be seen that all inbreeding coefficients reach an equilibrium between 0 and 1 in only 5 generations. Only when migration and mutation are negligible compared with the effect of drift in finite populations, do these inbreeding coefficients reach 1 eventually, i.e. the population becomes fixed at all loci.

Fig. 4 demonstrates that the prediction method was robust to the precise form used, given that (1) and (4) can be written differently with respect to locus order. For example, the lines showing F_{ABCD} using either $F_{ABC}F_{D|ABC}$ or $F_{ABD}F_{C|ABD}$ were indistinguishable, within each set of distances $\mathbf{d}$. Likewise, F_{ABC} can be expressed as in (1), or alternatively, as $F_{AC}F_{B|AC}$ or $F_{BC}F_{A|BC}$ (results not shown). Hence, for a given set of distances $\mathbf{d}$, the choice of loci to obtain conditional probabilities is of negligible significance.

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

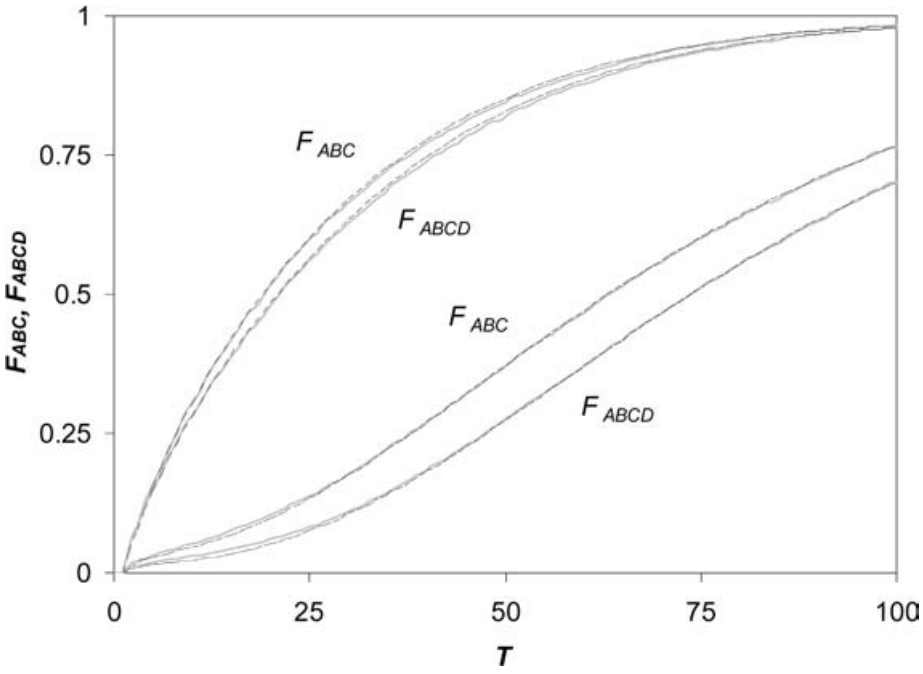

Fig. 2. Empirical averages over 10 000 replicates (continuous lines) versus deterministic predictions (dashed lines) of F_{ABC} and F_{ABCD}, plotted against generations (T). A monoecious population mated at random excluding selfing. Two different scenarios were considered: $N=10$ and $\mathbf{d}=[1, 1, 1]$ cM in the two upper graphs, and $N=20$ and $\mathbf{d}=[10, 10, 10]$ cM in the two lower graphs.

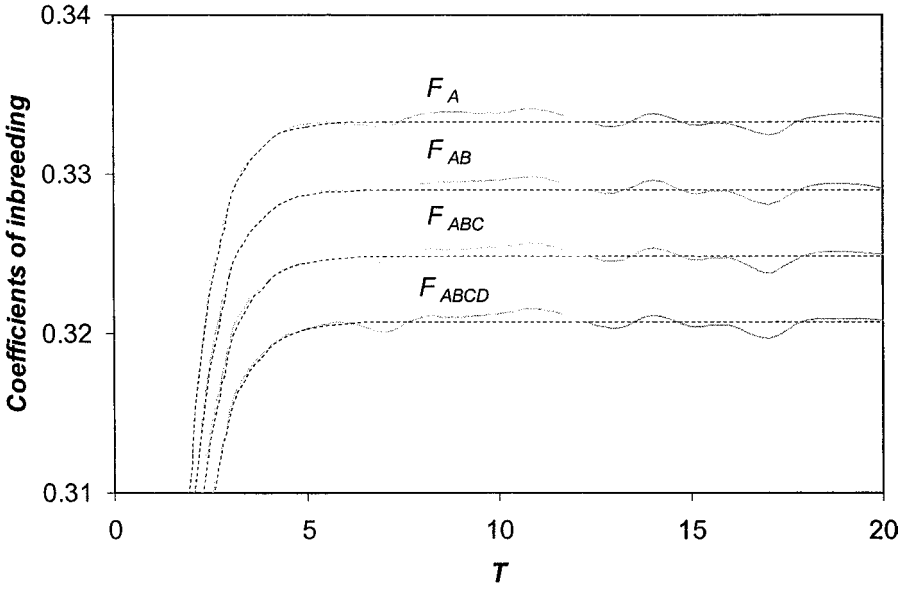

Fig. 3. Empirical averages over 1000 replicates (continuous lines) versus deterministic predictions (dashed lines) of F_A, F_{AB}, F_{ABC} and F_{ABCD}, plotted against generation time (T). A large monoecious population mated at random with one half of all matings being selfing. Deterministic predictions assumed an infinite population, whereas empirical predictions were averages for a population of size $N=1000$. The distance between adjacent loci was 1 cM.

(ii) *Sensitivity of* $\mathrm{F_{ABC}}$ *and* $\mathrm{F_{ABCD}}$ *to changes in* N *or* ***d***

Fig. 5 shows deterministic predictions of F_{ABC} for equidistant markers, spaced 1 cM from each other, and N equal to 5, 10, 20, 50, 100 or 200. Fig. 6 shows deterministic predictions of F_{ABC} for $N=10$, and equidistant markers spaced $\frac{1}{2}$, 1, 2, 5, 10, 20 or 30 cM from each other (F_{ABC} was very similar for distances $\geqslant 30$ cM between adjacent markers). It is easy to appreciate that F_{ABC} responds with a wider range of values to changes in N (Fig. 5) than in **d** (Fig. 6). Hence, wrong effective population sizes may lead to more strongly biased predictions of F_{ABC} than wrong genetic distances.

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

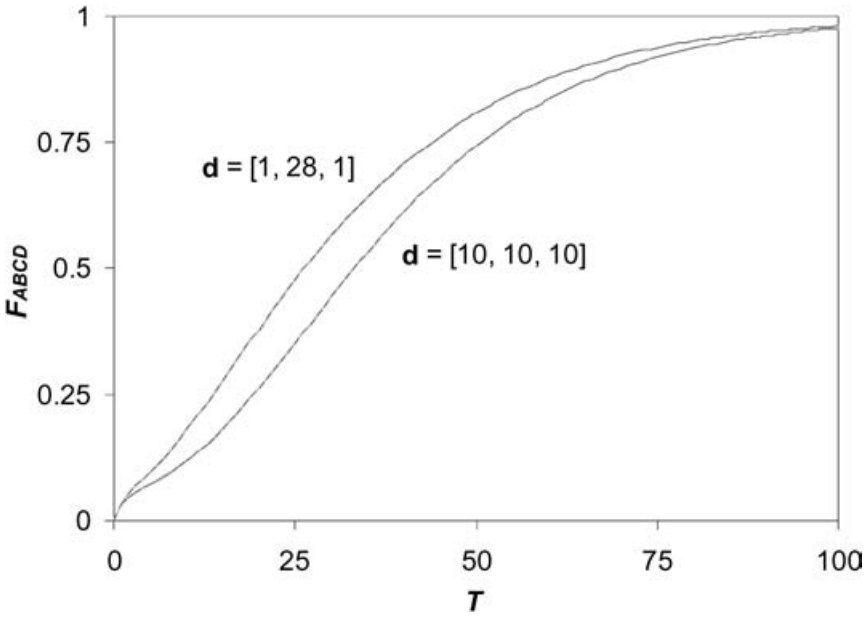

Fig. 4. F_{ABCD} predicted with $F_{ABC}F_{D|ABC}$ and $F_{ABD}F_{C|ABD}$, for two different sets of distances (**d**) in cM, plotted against generation time (T). The plots overlap completely within each set **d**, proving that $F_{ABC}F_{D|ABC}=F_{ABD}F_{C|ABD}$. There were differences in F_{ABCD} between scenarios representing equidistant and non-equidistant loci. In both cases, the total length was 30 cM, and $N=10$.

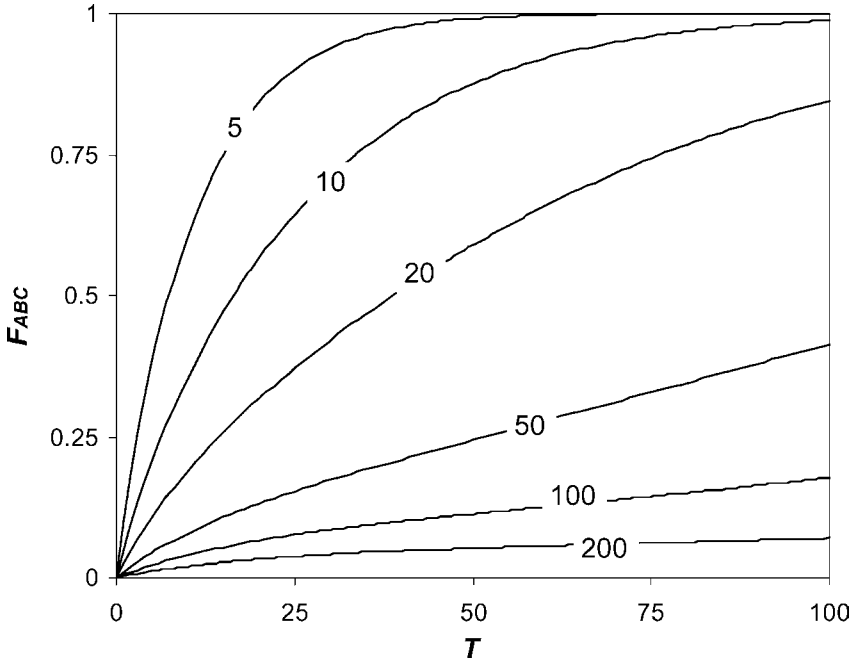

Fig. 5. Deterministic predictions of F_{ABC} at generation T for $N=5$, 10, 20, 50, 100, or 200, given equidistant markers, spaced 1 cM from each other.

The high sensitivity of F_{ABC} and F_{ABCD} to changes in N was also observed (results not shown) when F_{ABC} was predicted with $N=50$, and F_{ABCD} with $N=10$ (**d** was the same for both). Under this situation, and because of a stronger drift effect in the latter case, four loci became inbred faster than three.

Differences in distribution of loci over the same genetic distance caused significant differences in the expected multilocus inbreeding coefficient. For example, Fig. 4 shows differences in F_{ABCD} between equidistant loci (**d**=[10, 10, 10]) and non-equidistant loci (**d**=[1, 28, 1]), over a 30 cM chromosomal region. This result indicates that it is more likely to sample four loci simultaneously inbred when they are unevenly distributed over a given distance than when they are evenly distributed.

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

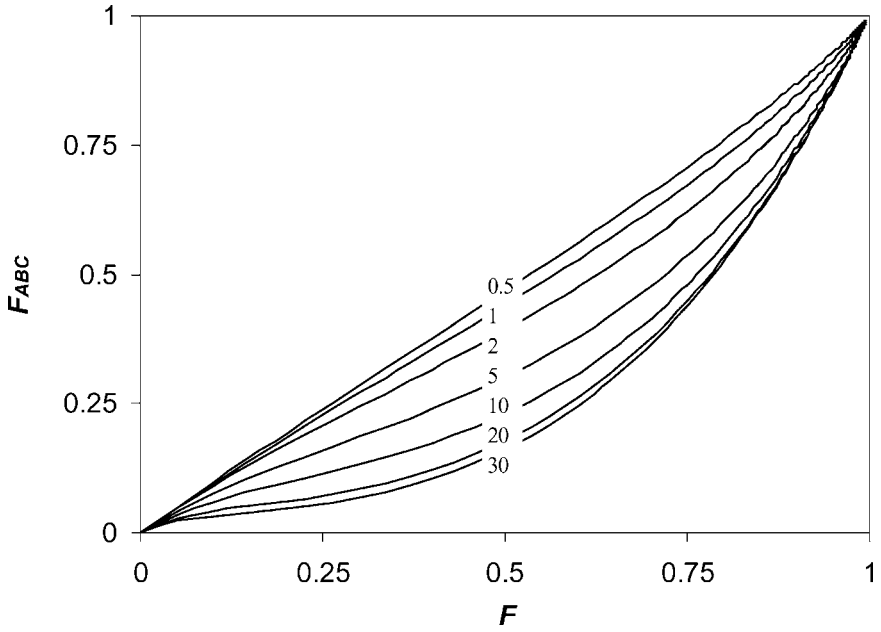

Fig. 6. Deterministic predictions of F_{ABC} plotted against the single locus F for $N=10$. The plots of F_{ABC} for $d \geqslant 30$ cM were very similar, so they are not shown. The genetic distances (in cM) between adjacent and equidistant loci are given in the graph.

4. Discussion

Population parameters such as inbreeding coefficients are the basic building blocks of quantitative genetic theory. Hence, the usefulness of this theory depends on our ability to predict these parameters in accurate and efficient ways. We have contributed to the theory of inbreeding by developing a new method, which is both accurate and instantaneous, to predict simultaneous inbreeding coefficients at three (F_{ABC}) and four (F_{ABCD}) loci in any generation.

This method uses a multiple linear regression model to predict the conditional probability of IBD at a locus given IBD at other linked loci, and then multiplies this probability by the prior probability of the latter loci being IBD. Although most of our results have been obtained assuming a finite population with constant size, and randomly mating and selfing over T discrete generations, the same level of accuracy has been observed without selfing and for a fixed proportion of selfing in an infinite population. Deterministic predictions are very accurate when the parameters N, (effective) population size, and **d**, genetic distances, are known without error. A violation of any of the previous assumptions reduces the accuracy of the method.

This theory can accommodate other population histories as long as simultaneous inbreeding coefficients at two loci can be predicted accurately. Weir *et al.* (1980) predicted exact two-loci coefficients for a monoecious population with random mating without selfing, for a dioecious population randomly mating, and for a dioecious population with a hierarchical mating structure. Moreover, Weir & Cockerham (1969, 1974) also considered mating systems where gametes are sampled without replacement (i.e. monogamy) in either monoecious or dioecious populations. In the case of an infinite population, inbreeding can only arise if there is a non-zero probability of random selfing, and for this case, exact equations for the simultaneous inbreeding coefficient at two loci can be found in Weir & Cockerham (1973).

The sensitivity of F_{ABC}, i.e. the range of possible values, is greater to N than to **d**. This fact makes multilocus inbreeding coefficients more robust to misspecifications of **d** than of N. In practice, a researcher could reduce the standard error of **d** by increasing the sample size, because **d** is usually obtained from observed recombination rates, which are more reliably estimated with large samples. On the contrary, N is less controllable experimentally because it depends on the particular evolution of each population and our knowledge of its history. Accurate methods for estimating N given molecular information may complement this methodology. Nevertheless, the need for accurate N to predict multilocus F is no different from the case of predicting single-locus F, and likewise, the need for accurate **d** is common to all tasks involving multiple loci.

The order in which loci were considered did not affect the accuracy of predictions. For example, (1) and (4) were equivalent to $F_{AC}\,F_{B|AC}$ and $F_{ABC}\,F_{D|ABC}$, respectively. However, the distribution of loci over a given distance, although not affecting the accuracy,

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press

had a significant effect on the predicted value. For example, four unevenly distributed loci were more likely to be simultaneously inbred than four evenly distributed ones.

Theoretically, this method can be extended to predict any n-loci ($n>4$) simultaneous inbreeding coefficient. However, as these coefficients are approximations and they depend on all lower simultaneous inbreeding coefficients (from 1 locus to $n-1$ loci), the prediction error may increase with n.

These coefficients are useful in many areas of research. For example, they have been incorporated in a new method for predicting co-ancestry among individuals without pedigree. These co-ancestry coefficients will be used in gene mapping. Moreover, these coefficients may be used in designing marker assisted selection/introgression programs in animal breeding, in studying linkage disequilibrium in populations, or detecting population structure.

Appendix

Parameters used in calculating $\Theta_{11(t+1)}$, which is equivalent to F_{AB}

$\Theta_{11(t)}$	Digametic digenic descent measure, equivalent to F_{AB}
$\Gamma_{11(t)}$	Trigametic digenic descent measure
$\Delta_{11(t)}$	Quadrigametic digenic descent measure
$\Theta_{1(t)}$	Digametic monogenic descent measure, equivalent to F
λ	$(1-2c)$, where c is the recombination rate
N	Effective population size
$\mathbf{\Omega}$	Matrix of sampling probabilities assuming random mating
t	Time variable (generations)

Formulae to calculate $\Theta_{11(t+1)}$, from Appendix B in Weir & Cockerham (1974):

$$\Theta_{11(t+1)} = \Omega_{11}\Theta_{11(t)} + \Omega_{12}\Gamma_{11(t)} + \Omega_{13}\Delta_{11(t)} + [(1-\lambda^2)/2N]\Theta_{1(t)} + [(1+\lambda^2)/4N].$$

$$\Gamma_{11(t+1)} = \Omega_{21}\Theta_{11(t)} + \Omega_{22}\Gamma_{11(t)} + \Omega_{23}\Delta_{11(t)} + [(2N-1)/2N^2]\Theta_{1(t)} + (1/4N^2).$$

$$\Delta_{11(t+1)} = \Omega_{31}\Theta_{11(t)} + \Omega_{32}\Gamma_{11(t)} + \Omega_{33}\Delta_{11(t)} + [(2N-1)/2N^2)\Theta_{1(t)} + (1/4N^2).$$

$$\Theta_{1(t)} = 1-((2N-1)/2N)^t.$$

$$\Omega_{ij} = \begin{bmatrix} \frac{(1+\lambda)^2}{4} - \frac{\lambda}{2N} & \frac{(N-1)(1-\lambda^2)}{2N} & \frac{(N-1)(1-\lambda)^2}{4N} \\ \frac{1+\lambda}{4N} - \frac{\lambda}{4N^2} & \frac{(N-1)[N+1+\lambda(N-2)]}{2N^2} & \frac{(N-1)(2N-3)(1-\lambda)}{4N^2} \\ \frac{2N-1}{4N^3} & \frac{(N-1)(2N-1)}{N^3} & \frac{(N-1)(2N-1)(2N-3)}{4N^3} \end{bmatrix}.$$

We acknowledge the UK Biotechnology and Biological Sciences Research Council (BBSRC) for funding this work. We also thank two anonymous referees and the editor for helpful comments that improved this manuscript.

References

Bijma, P., Woolliams, J. A. & Van Arendonk, J. A. M. (2000). Genetic gain of pure line selection and combined crossbred purebred selection with constrained inbreeding. *Animal Science* **72**, 225–232.

Haldane, J. B. S. (1919). The combination of linkage values, and the calculation of distance between the loci of linked factors. *Journal of Genetics* **8**, 299–309.

Malécot, G. (1948). *Les mathématiques de l'hérédité*. Paris: Masson.

Wright, S. (1977). *Evolution and the Genetics of Populations*, vol. 3. Chicago: University of Chicago Press.

Thompson, E. A. (1994). Monte Carlo likelihood in genetic mapping. *Statistical Science* **9**, 355–366.

Weir, B. S. & Cockerham, C. C. (1969). Group inbreeding with two linked loci. *Genetics* **63**, 711–742.

Weir, B. S. & Cockerham, C. C. (1973). Mixed self and random mating at two loci. *Genetical Research* **21**, 247–262.

Weir, B. S. & Cockerham, C. C. (1974). Behavior of pairs of loci in finite monoecious populations. *Theoretical Population Biology* **6**, 323–354.

Weir, B. S., Avery, P. J. & Hill, W. G. (1980). Effect of mating structure on variation in inbreeding. *Theoretical Population Biology* **18**, 396–429.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* **56**, 330–339.

Wright, A. F., Carothers, A. D. & Pirastu, M. (1999). Population choice in mapping genes for complex diseases. *Nature Genetics* **23**, 397–404.

https://doi.org/10.1017/S0016672303006633 Published online by Cambridge University Press