Conditional probabilities of identity of genes at a locus linked to a marker

C. CHEVALET, M. GILLOIS and Jacqueline VU TIEN KHANG*

I.N.R.A., Laboratoire de Génétique cellulaire,
* Station d'Amélioration génétique des Animaux,
Centre de Recherches de Toulouse, B.P. 12, F 31320 Castanet-Tolosan

Summary

A method is given for determining the probabilities that genes are identical by descent, at a locus linked to a marker where phenotypic data are available. For tight linkage, such conditional probabilities of identity may differ very much from unconditional ones ; they depend on the dominance relationships between alleles at the marker locus, and on the allelic frequencies. Applications discussed refer to the calculation of general two-locus descent measures, to the validation of pedigrees from polymorphism analysis, and to the statistical detection of an association between a quantitative trait and a marked region of the genome.

Key words : *Identity by descent of genes ; recombination ; genetic polymorphism.*

Résumé

Probabilités conditionnées d'identité entre gènes en un locus lié à un marqueur

On donne une méthode pour calculer les probabilités d'identité entre des gènes d'un locus, conditionnées par l'observation des phénotypes en un locus marqueur lié. La méthode proposée réunit 2 approches : les probabilités d'identité conditionnelles, au locus marqueur, sont calculées en construisant les événements possibles du processus de la ségrégation des gènes ; les probabilités conditionnelles d'identité en un locus voisin sont calculées ensuite par la méthode des arbres géniques, modifiée pour tenir compte des conditions réalisées au locus marqueur. Les résultats dépendent des relations de dominance entre allèles et des fréquences alléliques au locus marqueur ; pour une liaison étroite, ils peuvent être très différents des résultats non conditionnés. Les applications discutées concernent une méthode générale de calcul des coefficients d'identité à 2 locus, l'étude de la cohérence entre des données généalogiques et un polymorphisme génétique, et les méthodes de détection statistique d'une association entre un caractère quantitatif et une région marquée du génôme.

Mots clés : *Identité entre gènes ; recombinaison ; polymorphisme génétique.*

I. Introduction

Several methods of computing probabilities that genes at a single locus are identical by descent have been published (GILLOIS, 1964, 1966 ; CHEVALET, 1971 ; COCKERHAM, 1971 ; NADOT & VAYSSEX, 1973 ; DENNISTON, 1974 ; CHEVALET et al., 1977 ; VU TIEN KHANG et al., 1979). Except for the works by WEIR & COCKERHAM (1969a), COCKERHAM & WEIR (1973) and DENNISTON (1975), papers dealing with 2 linked loci have been mostly concerned with the change in time of mean descent measures in populations with fixed mating rules (COCKERHAM & WEIR, 1968, 1973 ; GALLAIS, 1970 ; WEIR & COCKERHAM, 1969b, 1973, 1974). In this paper, we consider the case where some phenotypic information is available at a marker locus linked to the locus at which probabilities are to be computed. We give a solution to this problem which has not been previously studied in a general way, and we specify some rules that allow : (1) computation of probabilities of identity of genes at a marker locus, conditional on the observation of phenotypes among relatives, and (2) derivation of probabilities of identity at a locus linked to such a marker. Methods used here combine the approaches developed by CHEVALET (1971), on the one hand, and by GILLOIS (1964) and VU TIEN KHANG et al. (1979), on the other hand.

Applications outlined in the discussion are : general calculation of two-locus descent measures from pedigree data, validation of pedigrees from polymorphism analysis, and detection of associations between a marker locus and linked genes contributing to a quantitative character.

II. Conditional probabilities of identity of genes at a marker locus

We consider a diploid population and a marker locus. We assume that the system is autosomal, regular and thoroughly described : any genotype gives rise to a unique phenotype, and dominance relationships between alleles and allelic frequencies are known. We first recall the expression for the a priori probability of the observed phenotypes, given the pedigree [formulae (1), (2)] ; then we show how to derive the conditional probabilities of identity at the marker locus given the observed phenotypes [formulae (3), (4)].

As an example, we use the pedigree made up of one mother and her 2 offspring born of 2 unrelated fathers (fig. 1), and the human ABO blood group system as a marker.

In a pedigree, let N be the number of parent-offspring links. Two mutually exclusive and equiprobable segregational events may correspond to each link, at any locus, which will be called *elementary* events. A gene transmitted to some offspring by 1 zygote originates from, and is therefore identical to, 1 of the 2 homologous genes carried by this zygote. So, a pedigree with N links gives rise to 2^N possible events (fig. 1), each of which is made up of N elementary events. Let ω denote any one of these mutually exclusive events. In any one of them, every gene in the population is given the name of the founder gene from which it derives ; genes bearing the same name in some event ω are identical by descent, and if genes in a fixed set of genes are isonymous in M events, out of the 2^N possible ones, their probability of being identical by descent is : $M.2^{-N}$. Monte Carlo simulations of the segregational process make this approach feasible for large pedigrees and small subsets of genes (CHEVALET, 1971).
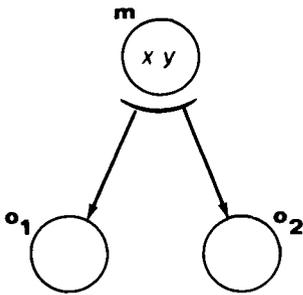
event	gene transmitted by the mother (m)	
	to (o_1)	to (o_2)
ω_1	x	x
ω_2	x	y
ω_3	y	x
ω_4	y	y

FIG. 1

The 2^2 exclusive events associated with the pedigree made up of 1 mother m
and her 2 offspring o_1 and o_2 born of 2 unrelated fathers.

Les 2^2 événements exclusifs asssociés au pedigree constitué d'une mère m et de ses 2 enfants o_1 et o_2
nés de 2 pères non apparentés.

Genes carried by the mother are named x and y.
Les gènes portés par la mère sont désignés par x et y.

For any event ω, genes belonging to zygotes of known phenotypes are found in some 1 identity situation. Conditional on ω, the probability of any list G of genotypes, for these zygotes, reads :

$$\Pr (G/\omega) = \delta (G,\omega). \prod_u p_u^{n_u(G,\omega)}, \qquad (1)$$

where :

p_u is the frequency of the u-th allele A_u at the marker locus ;

$n_u (G,\omega)$ is the number of distinct founder genes that should be associated with allele A_u, so as to realize the list G under the condition ω ;

$\delta (G,\omega)$ is either 1, if G is allowed by ω, or 0 if it is not (for instance, a zygote cannot be heterozygous in a situation where its 2 genes are identical).

It should be stressed here that in the specification of events ω, homologous genes of paternal and maternal origin, within a zygote, do not play symmetrical roles. It follows that any heterozygotic phenotype $[A_uA_v]$ must be split into the 2 possible *ordered* genotypes (A_uA_v) and (A_vA_u). With the convention that the gene of paternal origin is cited first, and following the notation in figure 1, we get, for example :

$$\Pr (m = (OB), o_1 = (OB), o_2 = (OO)/\omega_1) = 0$$
$$\Pr (m = (OB), o_1 = (OB), o_2 = (OO)/\omega_1) = 1 . p_O^2 . p_B^2$$

With dominance interactions between alleles, any list P of phenotypes may be realized by a series of genotypic lists G_α. Extension of formula (1) gives :

$$\Pr (P/\omega) = \sum_\alpha \delta (G_\alpha,\omega) . \prod_u p_u^{n_u(G_\alpha,\omega)} \qquad (2)$$

If this probability is not zero for at least some one ω event, application of Bayes'rule yields :

$$\Pr(\omega/P) = \frac{\sum_\alpha \delta(G_\alpha,\omega) \cdot \prod_u p_u^{n_u(G_\alpha,\omega)}}{\sum_{\omega'} \sum_\alpha \delta(G_\alpha,\omega') \prod_v p_v^{n_v(G_\alpha,\omega')}} \qquad (3)$$

Now, we are generally interested in the identity situations occurring among genes of a small subset. Let Ω be such a situation, and $\Delta(\Omega,\omega)$ the function equal to 1 if Ω is implied by ω, and to 0 if not. We get :

$$\Pr(\Omega/P) = \sum_\omega \Delta(\Omega,\omega) \cdot \Pr(\omega/P), \qquad (4)$$

where terms $\Pr(\omega/P)$ are given by formula (3).

As an example, we take as Ω the situation where genes of offspring originating from the mother are identical (fig. 1). The unconditional probability of Ω is 1/2 ; applications of formula (4) give :

$$\Pr(\Omega/m = [O], o_1 = [O], o_2 = [O]) = \frac{1}{2},$$

$$\Pr(\Omega/o_1 = [AB], o_2 = [O]) = 0,$$

$$\Pr(\Omega/m = [AB], o_1 = [B], o_2 = [B]) = 1,$$

$$\Pr(\Omega/m = [B], o_1 = [B], o_2 = [B])$$
$$= ((p_B + p_O)^3 + p_B^2\, p_O)/((p_B + p_O)^3 + p_B^2\, p_O + p_B\, (p_B + p_O)\, (p_B + 3p_O)),$$

$$\Pr(\Omega/m = [A], o_1 = [A], o_2 = [B], \textit{and the father of } o_1 \textit{ is } [O]) = 0.$$

The last example shows that some phenotypic data may be important although they refer to zygotes which do not carry any gene that can give rise to some gene found in the Ω situation. Hence no simple rule could be stated, that permitted eliminating useless information.

As in the case of unconditional probabilities, an approximate calculus may be proposed, based on Monte Carlo simulations. Many independent events ω_ℓ ($\ell = 1, 2, ..., L$) are generated. For each one, the conditional probability, $\Pr(P/\omega_\ell)$, is derived by formula (2). Then exact formula (4) is replaced by an estimated value :

$$\text{Est}.\Pr(\Omega/P) = \frac{\sum_{\ell=1}^{L} \Delta(\Omega, \omega_\ell) \cdot \Pr(P/\omega_\ell)}{\sum_{\ell=1}^{L} \Pr(P/\omega_\ell)}, \qquad (5)$$

provided that the denominator is not zero.

However, the estimation is generally biased. The bias decreases as L increases, but increases with the unknown frequency of forbidden ω events.

Computer programs, written in the Fortran 77 language and following algorithms derived from formulae (4) and (5), are available by contacting the authors.

III. Conditional probabilities of identity of genes at a locus linked to the marker

At a locus, linked to the marker with a non-zero probability of recombination, λ, and where no phenotypic information is available, every segregational event allowed by the pedigree is possible, but the probabilities attached to the 2^N events are not equal. Consider, at the marker locus, one of the possible events, ω, with its conditional probability Pr (ω/P) (formula (3)). Let ω' be 1 event at the 2nd locus. For every one of the N parent-offspring links, ω and ω' indicate from which parental chromosomes both offspring genes derive : ω and ω' state either that both genes originate from the same chromosome in the parent (fig. 2, case 1), or that they derive from the 2 homologous chromosomes in the parent (case 2). In the 1st case, there is no recombination, and the elementary event attached to that link in ω' has probability $(1 - \lambda)$, conditional on the elementary event attached to that same link in ω ; in the 2nd case a recombination is involved for that link, and the conditional probability is λ. Denoting by p (ω, ω') the number of links of the 1st kind, we have :

$$\text{Pr } (\omega'/\omega) = (1 - \lambda)\, p^{(\omega,\ \omega')}.\lambda^{N\ -p(\omega,\ \omega')},$$

and further :

$$\text{Pr } (\omega'/P) = \sum_{\omega} (1 - \lambda)^{p\,(\omega,\ \omega')} . \lambda^{N-p\,(\omega,\ \omega')} . \text{Pr } (\omega/P). \tag{6}$$

This formula (6) is the counterpart of formula (3), for a locus linked to the marker.

parent

ω ω' ω ω'

offspring

case 1

no recombination

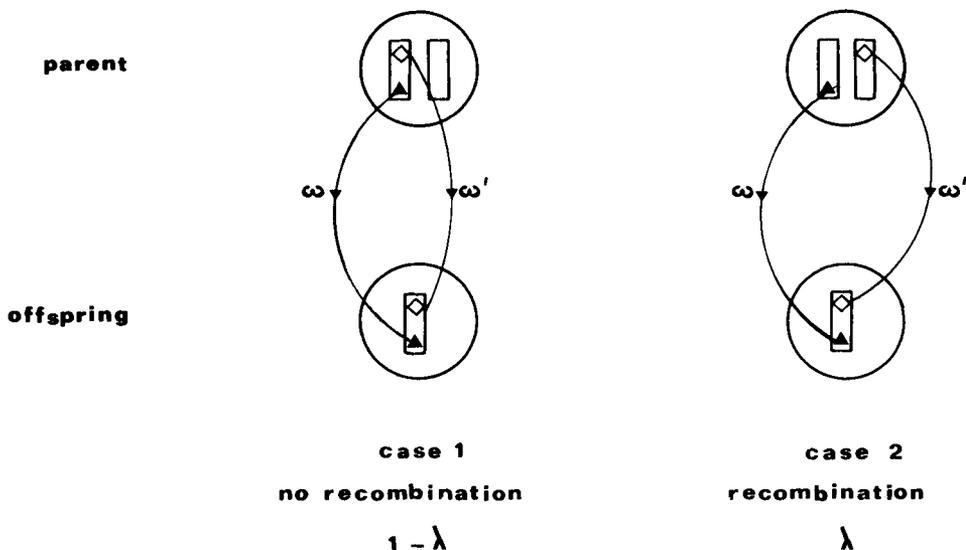$1 - \lambda$

case 2

recombination

λ

Fig. 2

The 2 possible joint segregations of 2 loci, along a parent-offspring link.

Les 2 ségrégations conjointes possibles de 2 locus liés,
le long d'un chaînon parent-enfant.

Calculations may be made simpler, using the method of genic trees (Vu Tien Khang *et al.*, 1979). In this algebraic approach, the set of genes at some one locus is identified with the set made up of the N parent-offspring links (a, b) in the pedigree. So, the gene carried by some zygote (b) and originating in its parent (a) is denoted by the pair (a, b) (fig. 3a, 3b). A relation of ancestry between genes is derived from that described in the zygotic network, in such a way that both genes (a', a) and (a", a) carried by zygote (a) are said to be the « 1-ancestors » of the (a, b) gene transmitted by (a) to any offspring (b). So, the set of genes is given a structure of partially ordered set, in which any segregation event ω can be represented in the following way. The 2 mutually exclusive elementary events attached to 1 link between parent (a) and offspring (b) are described in the genic set by the 2 *gametic* links ((a', a), (a, b)) and ((a", a), (a, b)). According to any event ω, every gene (a, b) is then linked upwards to only 1 gene among its 2 « 1-ancestors », and downwards to 0, one or several genes (b, c) carried by the offspring (c) of (b).
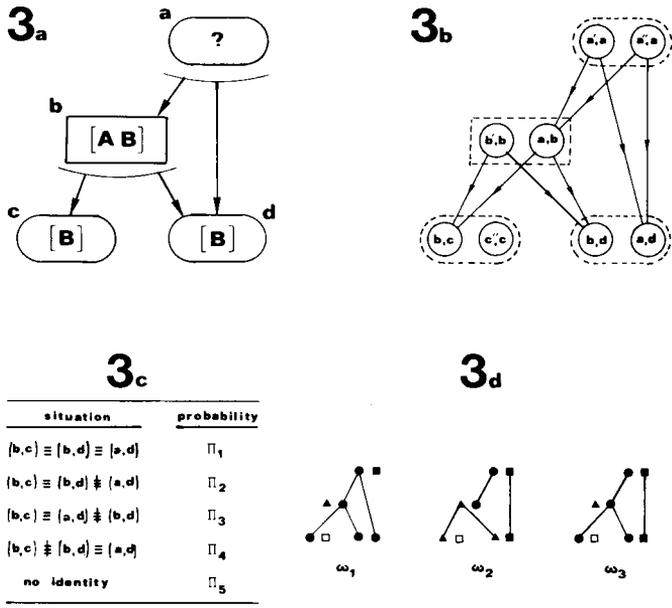

FIG. 3

Conditional probabilities of identity of genes at a marker locus
(ABO human blood group system).

Probabilités conditionnelles d'identité des gènes à un locus marqueur
(système sanguin humain ABO).

(a) The zygotic network and the phenotypic data ;
    Le réseau zygotique et les données phénotypiques ;
(b) The associated gametic network ;
    Le réseau gamétique associé ;
(c) List of identity situations between the 3 genes (b, c), (b, d) and (a, d) ;
    Liste des situations d'identité entre les 3 gènes (b, c), (b, d) et (a, d) ;
(d) Admissible segregation events at the marker locus.
    Evénéments possibles au locus marqueur.

Therefore in any ω, the subset made up of all the genes which are identical to one founder gene has the structure of a tree, the root of which is the founder gene, and the whole set of genes is split into disjoint subsets which are trees in the ordered set of genes (fig. 3d). The basic rule is then as follows. One considers situations Id ($\mathscr{S}$) according to which all genes in a subset $\mathscr{S}$ are identical. This situation arises for segregation events in which the genes involved pertain to a tree, in the ordered set of genes. To any situation Id ($\mathscr{S}$), corresponds a family ($\mathscr{A}_i$, m_i ; $i = 1, ..., I$) of trees $\mathscr{A}_i$, where m_i is the number of genes in $\mathscr{A}_i$. The unconditional probability of Id ($\mathscr{S}$) is then written :

$$\Pr\left(\mathrm{Id}\ (\mathscr{S})\right) = \sum_{i=1}^{I} \left(\frac{1}{2}\right)^{m_i - 1},$$

(GILLOIS, 1964 ; VU TIEN KHANG et al., 1979) (fig. 4a).

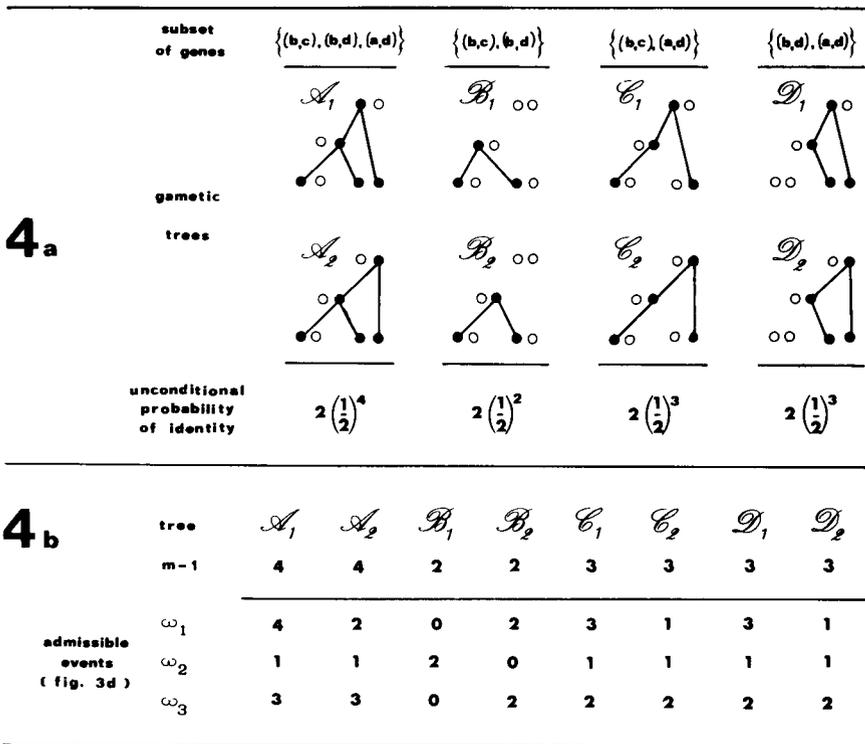|	tree	$\mathscr{A}_1$	$\mathscr{A}_2$	$\mathscr{B}_1$	$\mathscr{B}_2$	$\mathscr{C}_1$	$\mathscr{C}_2$	$\mathscr{D}_1$	$\mathscr{D}_2$
	m − 1	4	4	2	2	3	3	3	3
admissible events (fig. 3d)	ω_1	4	2	0	2	3	1	3	1
	ω_2	1	1	2	0	1	1	1	1
	ω_3	3	3	0	2	2	2	2	2

FIG. 4

Conditional probabilities of identity at a locus linked to a marker.
Probabilités conditionnelles d'identité à un locus lié à un marqueur.

The pedigree and the phenotypic data are those of figure 3a.
Le pedigree et les données phénotypiques sont ceux de la figure 3a.
(a) Gametic trees corresponding to identity within subsets of genes ;
   Arbres gamétiques correspondant à l'identité de sous-ensembles de gènes ;
(b) Values of the numbers p (ω, $\mathscr{A}$), cf. formula (7).
   Valeurs des nombres p (ω, $\mathscr{A}$), cf. formule (7).

Considering the situation Id_2 ($\mathscr{S}$) at the 2nd locus linked to the marker, any tree $\mathscr{A}_i$ may still be realized, and is the union of ω' events whose probabilities, conditional on some ω at the 1st locus, are given by (6). Except for the root, any one of the genes in $\mathscr{A}_i$ has its origin specified among its 2 « 1-ancestors » ; the specification is either identical to (fig. 2, case 1), or different from (case 2), that given in the ω event. Letting p (ω, $\mathscr{A}_i$) be the number of coincidences between ω and $\mathscr{A}_i$, we then have :

$$\text{Pr (Id}_2 (\mathscr{S})/\omega) = \sum_{i=1}^{I} (1 - \lambda)^{p\,(\omega,\mathscr{A}_i)} . \lambda^{m_i-1-p\,(\omega,\,\mathscr{A}_i)}$$

and, conditional on a phenotypic list P_1 at the 1st marker locus :

$$\text{Pr (Id}_2 (\mathscr{S})/P_1) = \sum_{\omega} \sum_{i=1}^{I} (1 - \lambda)^{p\,(\omega,\,\mathscr{A}_i)} . \lambda^{m_i-1-p\,(\omega,\,\mathscr{A}_i)} . \text{Pr } (\omega/P_1) \qquad (7)$$

Extensions to identity situations involving 2 or more independent groups of identical genes follow the corresponding analysis by genic trees, for one-locus situations (VU TIEN KHANG *et al.*, 1979). An example is presented in figures 3 and 4, with the results given in appendix.

IV. Discussion

In the 1st instance, it seems worth citing some technical by-products of the previous analysis.

The rules combining segregational events at 2 linked loci (formulae (6), (7)) have a direct application to the computation of two-locus descent measures in the general case, when no phenotypic information is available. Consider 2 subsets, $\mathscr{S}$ and $\mathscr{T}$, in the set of genes denoted by parent-offspring links, and the situation :

$$\Omega = (Id_1 (\mathscr{S}), Id_2 (\mathscr{T}))$$

stating that gene of $\mathscr{S}$ at locus 1 are identical, and that genes of $\mathscr{T}$ at locus 2 are identical. Identity of $\mathscr{S}$ genes is made possible by a set of I trees $\mathscr{A}_i$, with m_i genes in $\mathscr{A}_i$ ($i = 1, ..., I$) (resp. : J trees $\mathscr{B}_j$, n_j, $j = 1, 2..., J$ for $\mathscr{T}$). Any pair of trees $(\mathscr{A}_i, \mathscr{B}_j)$ makes the situation Ω possible. Assuming that Id_1 ($\mathscr{S}$) is realized by the tree $\mathscr{A}_i$, we have to compute the probability of Id_2 ($\mathscr{T}$) through the tree $\mathscr{B}_j$ and conditional on $\mathscr{A}_i$ at the 1st locus. Relative to $\mathscr{A}_i$, any *gametic* link ((a, b), (b, c)) in $\mathscr{B}_j$ may be of 3 kinds :

(1) *gametic* link ((a, b), (b, c)) is found in $\mathscr{A}_i$, then we are in case 1 of figure 2, the conditional probability of the elementary event is $(1 - \lambda)$.

(2) *gametic* link ((a, b), (b, c)) is not in $\mathscr{A}_i$, but the *gametic* link ((a', b), (b, c)) is in $\mathscr{A}_i$: according to the events $\mathscr{B}_j$ the gene (b, c) transmitted by zygote (b) to zygote (c) originates in the parent (a) of (b), while in the events $\mathscr{A}_i$, the gene (b, c) originates in the other parent (a') of (b) ; this situation implies a recombination event during meiosis in zygote (b), we are in case 2, the conditional probability is λ.

(3) neither ((a, b), (b, c)) nor ((a', b), (b, c)) is found in $\mathscr{A}_i$, then the segregation event described in this *gametic* link of $\mathscr{B}_j$ is independent of those specified by $\mathscr{A}_i$, the probability attached to the elementary event is $\dfrac{1}{2}$.

GENE IDENTITY NEAR A MARKER LOCUS

Let p_{ij} and q_{ij} be the numbers of $\mathscr{B}_j$ *gametic* links of the 1st and the 2nd kinds, respectively. We get :

$$\Pr\left(\mathrm{Id}_1\left(\mathscr{G}\right),\, \mathrm{Id}_2\left(\mathscr{G}\right)\right) = \sum_{i=1}^{I}\left(\frac{1}{2}\right)^{m_i-1}\sum_{j=1}^{J}(1-\lambda)^{p_{ij}}\,\lambda^{q_{ij}}\left(\frac{1}{2}\right)^{n_j - p_{ij} - q_{ij} - 1} \tag{8}$$

Extensions to situations involving several groups of identical genes at both loci follow the same rationale. This analysis provides a generalization of that given by DENNISTON (1975), since it is not restricted to situations found between genes taken from « regular » (non-inbred) relatives. We may stress that working within a set of genes (the set of parent-offspring links considered as genes, chromosomes or gametes) makes the analysis general and simpler : it does not exclude inbreeding, and only 2 basic meiotic events (fig. 2) need be considered, instead of 5 when working directly with « regular » relatives (DENNISTON, 1975, fig. 4).

The calculations done to get the conditional probabilities of identity at a marker locus involve verifying whether genealogical and phenotypic data are compatible. As such, the method may provide an extension of those used to validate genealogies in domestic animals, or to test paternity in humans from analysis in trios or in sibships (CHASTANG, 1976). For that, we have only to decide whether lists G_α of genotypes are admissible or not ; the complete list of alleles and their dominance relations should be sufficient, and the calculations of allelic frequencies products useless (formulae (1) and (2)). The method might improve the efficiency of classical ones in complex situations with dominance between alleles, or with incomplete phenotypic information (fig. 5).
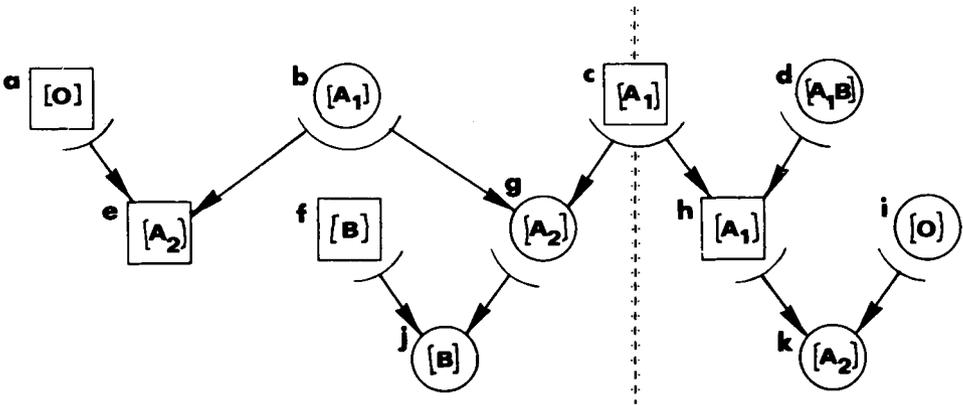

FIG. 5

An example of joint exclusion in a 3 generation pedigree.

Exemple d'exclusion dans un pedigree.

The genetic system is the A_1A_2BO human blood group system (A_1 dominant over A_2 and O, A_2 dominant over O, and B dominant over O). Every trio (mother-father-offspring) is admissible ; the left and right parts of the pedigree are admissible subsets ; the whole set becomes admissible if the phenotype of any of a, b, c, d, e, i, j or k is unknown.

Le marqueur est le système sanguin humain A_1A_2BO (A_1 est dominant sur A_2 et sur O ; A_2 et B sont dominants sur O). Chacun des trios (mère-père-enfant) est admissible ; les parties gauche et droite du pedigree constituent des sous-ensembles admissibles ; l'ensemble de la généalogie devient admissible si le phénotype de a, b, c, d, e, i, j ou k est inconnu.

The main results of this work, as expressed in formulae (3), (4) and (7), provide a solution to a problem which had no solution, except in the special circumstance ot a locus linked to a locus forced to remain heterozygous (HANSET, 1974, and references therein). Such studies are based on the generation matrix method, which is tractable only when dealing with very small populations and fixed mating rules, such as sib mating. On the contrary, our method can handle any pedigree and a large class of genetic markers, although its use should be restricted to rather sparse genealogical networks or to few generations, so as to avoid some kind of « combinatorial explosion » ; this practical restriction is mainly due to the fact that, for example, the conditional inbreeding coefficient of some zygote may depend on the phenotypes of unrelated individuals, a situation that forbids most simplifications available in the calculation of unconditioned probabilities. An interesting feature of the conditional probabilities is their dependence upon allelic frequencies. This dependence may turn out to be very important : the probability of identity of genes in zygotes (c) and (d) (fig. 3), is highly increased if alleles B and O are rare, even for loose linkage (appendix(e)). This shows how the description of polymorphisms in a population may change the genetical interpretation of correlations between traits measured in related individuals, if the genes contributing to these traits are linked to the markers. This evidence motivated the present work, and should lead to its main field of application.

Methods to correlate genetic polymorphism and variation in quantitative traits have been developed, using the segregation of a marker gene among full sibs or half sibs (NEIMAN-SÖRENSEN & ROBERTSON, 1961 ; JAYAKAR, 1970 ; HASEMAN & ELSTON, 1972 ; SOLLER & GENIZI, 1978). Developments in gene mapping (OLLIVIER & SELLIER, 1982 ; PEARSON *et al.*, 1982) and evidence that a great body of polymorphic loci should be identified by means of molecular biology techniques (SOUTHERN, 1975 ; WYMAN & WHITE, 1980 ; SKOLNICK & WHITE, 1982) are promising for the genetic improvement of domestic animals.

Statistical methods of detection are based on an analysis of variance aimed at pointing out the significance of effects attached to the identity states of genes linked to the marker locus. In the case of a population of half sibs from an heterozygous father, the statistical setting is evident (SOLLER & GENEZI, 1978), but it becomes intricate in the case of a hierarchical scheme (s sires, d dams per sire, n offspring per dam). Results of this paper provide a way to the automatic statement of the needed statistical model. For n zygotes, any identity situation may be represented by an (n × r) matrix X, of which element X_{ij} is the number of genes from j-th origin carried by the i-th zygote : X_{ij} is equal to 0, 1, or 2, row sums are equal to 2, and XX' is 4 times the matrix of conditional coefficients of kinship. Given the set of identity situations represented by X_k, with their conditional probabilities q_k, the matrix of conditional coefficients of kinship for the marked region of the genome is :

$$\phi_c = \frac{1}{4} \sum_k q_k X_k X_k'. \tag{9}$$

The problem of detecting whether the marked region contributes to the variability of a trait Y is amenable to a mixed model analysis of variance, with 1st and 2nd moments given by :

$$E(Y) = X_o \gamma_o$$
$$Cov(Y, Y') = 2 \sigma_1^2 \phi_c + \sigma^2 \Sigma$$

where :

γ_o is a vector of unknown fixed effects, and X_o a known incidence matrix ;

$\sigma^2 \Sigma$ is the variance-covariance matrix of observations, assuming no contribution of the marked region, Σ being taken as known ;

σ_1^2 is the contribution of the marked region to the genetic additive variance, and ϕ_c is given by (9).

The hypothesis to be tested is ($\sigma_1^2 = 0$), against ($\sigma_1^2 \neq 0$). The general way to compute the matrix ϕ_c should make it possible to compare various mating schemes, in order to optimize the power of the test within fixed experimental limits.

Conversely, observation of the segregations at marked loci may be used as a tool to improve selection, provided that a quantitative linkage has been proven between the markers and the traits selected for. SOLLER & BECKMAN (1982) have stressed that it should be quite efficient for within family selection and preliminary results indicate that a typical situation in which marker-assisted selection might prove most useful is combined selection for a trait with low heritability.

Acknowledgments

References

CHASTANG C., 1976. « PATER » : a program for parenthood diagnosis. *Computer Programs in Biomedicine*, **5**, 251-258.

CHEVALET C., 1971. Calcul automatique des coefficients d'identité. *Ann. Génét. Sél. Anim.*, **3**, 449-462.

CHEVALET C., GILLOIS M., NASSAR R., 1977. Identity coefficients in finite populations. I. Evolution of identity coefficients in a random mating dioecious population. *Genetics*, **86**, 697-713.

COCKERHAM C.C., 1971. Higher order probability functions of identity of alleles by descent. *Genetics*, **69**, 235-246.

COCKERHAM C.C., WEIR B.S., 1968. Sib mating with two linked loci. *Genetics*, **60**, 629-640.

COCKERHAM C.C., WEIR B.S., 1973. Descent measures for two loci with some applications. *Theor. Popul. Biol.*, **4**, 300-330.

DENNISTON C., 1974. An extension of the probability approach to genetic relationships : one locus. *Theor. Popul. Biol.*, **6**, 58-75.

DENNISTON C., 1975. Probability and genetic relationships : two loci. *Ann. Hum. Genet.*, **39**, 89-104.

GALLAIS A., 1970. Covariances entre apparentés quelconques avec linkage et épistasie. II. Evolution en régime d'autofécondation. *Ann. Génét. Sél. Anim.*, **2**, 417-427.

GILLOIS M., 1964. *La relation d'identité en génétique*. Thèse, Faculté des Sciences, Paris, 294 p.

GILLOIS M., 1966. Le concept d'identité et son importance en génétique. *Ann. Génét.*, **9**, 58-65.

HANSET R., 1974. Consanguinité par croisement frère × sœur, avec hétérozygotie forcée et viabilités différentes. *Ann. Génét. Sél. Anim.*, **6**, 345-368.

HASEMAN J.K., ELSTON R.C., 1972. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.*, **2**, 3-19.

JAYAKAR S.D., 1970. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics*, **26**, 451-466.

NADOT R., VAYSSEIX G., 1973. Apparentement et identité. Algorithme du calcul des coefficients d'identité. *Biometrics*, **29**, 347-359.

NEIMANN-SÖRENSEN A., ROBERTSON A., 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agric. Scand.*, **11**, 163-196.

OLLIVIER L., SELLIER P., 1982. Pig genetics : a review. *Ann. Génét. Sél. Anim.*, **14**, 481-544.

PEARSON P.L., RODERICK T.M., DAVISSON M.T., LALLEY P.A., O'BRIEN S.J., 1982. Report of the committee on comparative mapping. *Cytogenet. Cell. Genet.*, **32**, 208-220.

SKOLNICK M.H., WHITE R., 1982. Strategies for detecting and characterizing restriction fragment length polymorphims (RFLP's). *Cytogenet. Cell Genet.*, **32**, 58-67.

SOLLER M., BECKMANN J.S., 1982. Restriction fragment length polymorphisms and genetic improvement. *2nd World Congress on Genetics applied to Livestock Production, Madrid, October 4-8, 1982,* Editorial Garsi, Madrid, **7**, 396-404.

SOLLER M., GENIZI A., 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics*, **34**, 47-55.

SOUTHERN E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503-517.

VU TIEN KHANG J., de ROCHAMBEAU H., CHEVALET C., GILLOIS M., 1979. Analyse des pedigrees et calcul des coefficients d'identité par les arbres géniques. *Biom. J.*, **21**, 367-387.

WEIR B.S., COCKERHAM C.C., 1969a. Pedigree mating with two linked loci. *Genetics*, **61**, 923-940.

WEIR B.S., COCKERHAM C.C., 1969b. 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. *Genet. Res. Camb.*, **21**, 247-262.

WEIR B.S., COCKERHAM C.C., 1974. Behavior of pairs of loci in finite monoecious populations. *Theor. Popul. Biol.*, **6**, 323-354.

WYMAN A.R., WHITE R., 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 6754-6758.

Appendix

Results of the calculations associated with the example of figures 3 and 4.

Résultats des calculs associés à l'exemple des figures 3 et 4.

(a) Unconditional probabilities of identity situations (cf. figure 3c).

Probabilités non conditionnelles des situations d'identité.

$$\pi_1 = \frac{1}{8}, \ \pi_2 = \frac{3}{8}, \ \pi_3 = \frac{1}{8}, \ \pi_4 = \frac{1}{8}, \ \pi_5 = \frac{1}{4}$$

(b) Conditional probabilities of admissible events at the marker locus (cf. formula (3) and fig. 3d) :

Probabilités conditionnelles des événements admissibles au locus marqueur.

$$\text{Pr}(\omega_1/P_1) = \frac{1}{1 + 2(p_o + p_B)}, \ \text{Pr}(\omega_2/P_1) = \text{Pr}(\omega_3/P_1) = \frac{p_o + p_B}{1 + 2(p_o + p_B)}$$

(c) Conditional probabilities of identity situations at the marker locus.

Probabilités conditionnelles des situations d'identité au locus marqueur.

$$\pi_1 = \frac{1}{1 + 2\,(p_o + p_B)}, \quad \pi_2 = \frac{2\,(p_o + p_B)}{1 + 2\,(p_o + p_B)}, \quad \pi_3 = \pi_4 = \pi_5 = 0$$

(d) Conditional probabilities of identity situations at a locus linked to the marker with a recombination fraction λ, ($\mu = 1 - \lambda$) :

Probabilités conditionnelles des situations d'identité à un locus lié au marqueur avec une fréquence de recombinaison λ, ($\mu = 1 - \lambda$) :

$$\pi_1\,(\lambda) = (\lambda^2 + \mu^2)\,(\mu^2 + 2\lambda\mu\,(p_o + p_B))/(1 + 2\,(p_o + p_B)),$$

$$\pi_2\,(\lambda) = (\lambda^2 + \mu^2)\,(1 - \mu^2 + 2\,(1 - \lambda\mu)\,(p_o + p_B))/(1 + 2\,(p_o + p_B)),$$

$$\pi_3\,(\lambda) = \pi_4\,(\lambda) = \lambda\mu\,(\lambda^2 + \mu^2 + 4\,\lambda\mu\,(p_o + p_B))/(1 + 2\,(p_o + p_B)),$$

$$\pi_5\,(\lambda) = 4\,\lambda\mu\,(\lambda\mu + (\lambda^2 + \mu^2)\,(p_o + p_B))/(1 + 2\,(p_o + p_B))$$

(e) Conditional values of the inbreeding coefficient f_d of (d), and of the kinship coefficient ϕ_{cd} between (c) and (d) as functions of λ, for frequencies p_A of allele A equal to .30 and .90 ($p_A = 1 - p_O - p_B$).

Valeurs conditionnées du coefficient de consanguinité f_d de (d) et du coefficient de parenté ϕ_{cd} entre (c) et (d), en fonction de λ, pour des fréquences p_A de l'allèle A égales à 0,30 et 0,90 ($p_A = 1 - p_o - p_B$).

λ	0	.01	.02	.05	.10	.20	.30	.40	.50	p_A
f_d	.417	.410	.404	.386	.360	.320	.292	.270	.250	.30
	.833	.810	.788	.724	.630	.480	.373	.300	.250	.90
ϕ_{cd}	.354	.348	.341	.323	.295	.250	.218	.198	.188	.30
	.458	.448	.437	.407	.363	.290	.238	.205	.188	.90