

HISTOIRES D'IDENTITÉ (GÉNÉTIQUE)

(HI)STORIES OF (GENETIC) IDENTITY

Anne-Louise LEUTENEGGER ¹, Pierre DARLU ²

RÉSUMÉ

Cette note présente un point de vue historique sur les conditions dans lesquelles est apparu le concept d'identité génétique entre quatre allèles de deux individus. Plusieurs définitions possibles des états d'identité ont été proposées entre les années 1940 et 1970, avec des regroupements de ces états qui diffèrent d'un auteur à l'autre. L'analyse historique de la littérature montre que la définition et la description des états d'identité se sont mises en place quasi-simultanément, et très probablement indépendamment, en France et aux États-Unis.

Mots-clés : état d'identité génétique, parenté, consanguinité, identité par descendance, IBD.

ABSTRACT

This note takes a historical perspective on the concept of genetic identity states between four alleles of two individuals. Several definitions of these states were proposed between 1940 and 1970, and states were diversely merged according to the authors. The historical analysis of the publications shows that the definition and the description of the identity states occurred quasi-simultaneously and most probably independently both in France and in the United States of America.

Keywords : Genetic identity state, kinship, consanguinity, identity by descent, IBD.

-
1. U946, INSERM, Variabilité génétique et maladies humaines, 27 rue Juliette Dodu, 75010 Paris, e-mail : anne-louise.leutenegger@inserm.fr
 2. U535, INSERM, Épidémiologie génétique et structure des populations humaines, BP 1000 94817 Villejuif, e-mail : pierre.darlu@inserm.fr

Le milieu du XX^e s. fut l'occasion d'un changement conceptuel radical en ce qui concerne la façon d'appréhender les relations génétiques entre apparentés. Il n'était plus seulement question de rendre compte de ces relations en termes de corrélation, comme l'ont proposés Fisher (1918) et Wright (1921, 1922), mais de les exprimer en termes de probabilité. Ce changement de paradigme doit beaucoup à celui qui ne fut, un temps, qu'un obscur chercheur français, Gustave Malécot. Les travaux de Malécot sont maintenant largement reconnus et l'histoire de l'émergence de ses idées et de ses relations avec ses grands prédécesseurs, en particulier Fisher, Haldane et Wright a été amplement décrite par plusieurs auteurs (Nagylaki 1989 ; Bocquet-Appel 1996 ; Malécot, Bocquet-Appel 1996 ; Epperson 1999 ; Slatkin, Veuille 2002). La prééminence de la pensée de Malécot et sa conception probabiliste de la notion d'identité génétique ne doit cependant pas occulter l'existence de quelques traces d'antériorité, même si, dans son cas, il s'avère qu'il les ignorait. C'est le cas de la thèse de Cotterman, soutenue en 1940 mais publiée seulement en 1974 et sur laquelle nous reviendrons (Cotterman 1940)³.

Notre propos est, plus modestement, de relater comment les diverses relations d'identité entre les allèles de deux individus ont été évoquées et ont fini par s'imposer une fois exprimées en termes de probabilité.

Nous prendrons comme point de départ, la position actuelle, quitte à remonter dans le temps par la suite. On reconnaît 15 situations qui décrivent exhaustivement les relations d'identité entre les quatre allèles de deux individus diploïdes en un locus donné (voir *fig. 1a*). Considérons deux individus I et J, chacun ayant deux allèles à chaque locus (respectivement I1 et I2, J1 et J2). Leurs quatre allèles peuvent être différents (\neq) ou identiques (\equiv) par ascendance de 15 façons différentes. L'identité peut être établie entre les deux allèles d'un même individu, quand l'un ou l'autre de ces individus est consanguin (I1 \equiv I2 ou J1 \equiv J2), ou l'identité peut être entre allèles des deux individus (par ex. I1 \equiv J1 ou I1 \equiv J2), quand ils sont apparentés et que chacun d'eux a reçu ces allèles d'un même ancêtre commun. Enfin, l'identité peut également s'observer à la fois entre les allèles d'un même individu

(consanguinité) et entre les allèles des deux individus (par ex. I1 \neq I2 \equiv J1 \equiv J2) (parenté).

Ces 15 situations sont clairement exposées dans le livre de Jacquard (1970), publié en français et dans sa traduction anglaise qui sera publiée en 1974. La description de ces 15 situations se trouvait d'abord dans un article de Jacquard (1966) publié dans la revue *Population*. Cette description est illustrée d'une figure qui est reprise dans la figure 1a et permet de bien envisager toutes les situations possibles.

En fait, ces 15 situations avait été préalablement décrites (*fig. 1b*) dans la thèse de Gillois (1964, 1965) thèse à laquelle Jacquard fait régulièrement crédit dans ses articles et livres successifs. L'originalité de Jacquard par rapport à la thèse de Gillois, outre un mode de représentation visuel très éclairant, est de regrouper ces 15 situations d'identité en neuf situations dites contractées (« *condensed* ») qu'il présente dans son livre (Jacquard 1970) puis dans la revue *Biometrics* en 1972. Dans sa thèse, Gillois propose également de regrouper les 15 situations mais en cinq situations (qu'il nomme « *restreintes* »).

La même année et le même mois (décembre, 1964), Gillois présentait sa thèse sur la relation d'identité en génétique et Harris (1964) publiait un article sur un sujet identique dans la revue *Genetics*. Dans cet article, comme Cockerham (1971) se plaira à le souligner en 1971, les 15 situations d'identité étaient aussi évoquées mais seulement pour permettre le calcul des probabilités des neuf situations d'identité, jugeant que elles seules sont « *relevant for the consideration of [his] paper* ». Ces neuf situations sont d'ailleurs les mêmes que celles qui seront retenues par Jacquard en 1970 (cependant sans référence à Harris 1964).

Cotterman (1940, 1974) fut, incontestablement, le premier à tenter d'explorer les diverses situations d'identités et à les exprimer en termes de probabilité, comme le soulignera Thompson (1974) dans son article étendant les situations d'identité entre deux individus à celles entre plusieurs individus. Cotterman définit aussi le concept de « *derivativeness* », concept qui connaîtra plus de succès sous le terme d'« *identity by descent* » (*ibd*). Cotterman proposait également une description des situations d'identité, mais n'en retenait paradoxalement que 12. Il lui en manque donc trois (*fig. 1c*) ! Ce fait ne semble pas avoir été souligné par les différents auteurs qui ont su rendre hommage à son travail. La thèse de Cotterman était d'ailleurs accompagnée d'une représentation en diagramme (*fig. 2*) qui ressemble et annonce celle

3. "The solution of many statistico-genetic problems is greatly simplified when the analysis proceeds from consideration of the probabilities of the various genic resemblances limited to the types of relatives involved. To the knowledge of the author, no systematic use of such procedure has heretofore been undertaken [...]"

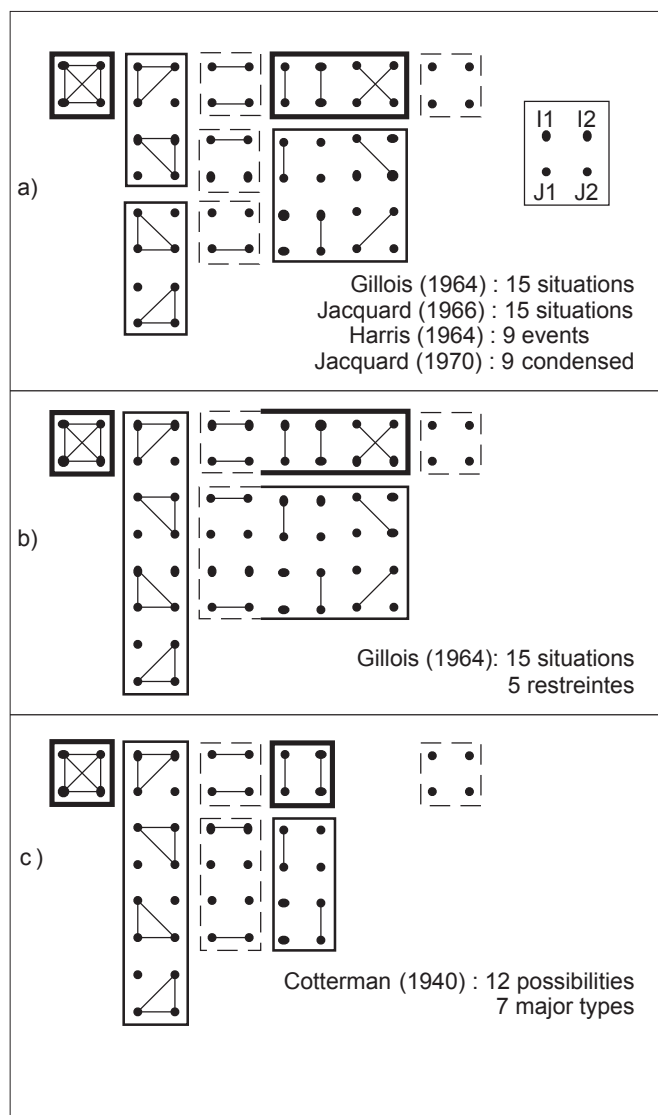


Fig. 1 - Représentation des situations d'identité entre les quatre allèles (I1, I2, J1, J2) des deux individus I et J.

Un trait joignant deux allèles signifie une relation d'identité entre eux. Les cadres rassemblent les différents regroupements proposés par les auteurs. Leur épaisseur (épais, fin ou tirets) rend compte du degré d'identité par descendance ($ibd = 2, 1$ et 0 respectivement).

Dans la figure 1b, deux cadres regroupent des situations d' ibd hétérogènes.

Fig. 1—Representation of the identity states between four alleles (I1, I2, J1, J2) of two individuals, I and J. Identity between two alleles is indicated by a line connecting them. Frames bring together different states according to different authors. The thickness (thick, thin or dotted) of the frames indicates the degree of identity by descent ($ibd = 2, 1$ et 0 respectively). In the figure 1b, two frames merge heterogeneous ibd states.

utilisée par Jacquard (1966) (fig. 1a) mais organise les situations d'identité différemment. La coïncidence entre la présentation de Jacquard et celle de Cotterman apparaît manifeste par la publication simultanée dans le même ouvrage (Ballonoff 1974) de l'ancienne thèse de Cotterman (datant alors de 25 ans) et de l'article en anglais correspondant à la partie sur les états d'identité du livre de Jacquard (1970, 1972). Cotterman propose, lui aussi, des situations condensées (« major types »), mais au nombre de sept, et qui ne correspondent ni aux regroupements proposés par Gillois ni à ceux de Jacquard (fig. 1c)... En 1983, Cotterman (1983) revient sur ses propres situations d'identité mais en reste toujours aux 12 situations de 1940.

Pour tenter d'expliquer les raisons pour lesquelles les situations d'identité sont différemment considérées d'un auteur à l'autre, revenons sur la figure 1c et 2. Cotterman ne reconnaît que 12 situations car il n'attribue pas le même rôle à I et J. En effet, il opère une distinction entre les deux allèles de l'individu I (par exemple leur origine paternelle ou maternelle), tandis qu'il ne se pose pas la question d'une telle différence d'origine pour les allèles de l'individu J. C'est pourquoi, les situations notées 2 et 3 chez Cotterman (fig. 2) sont considérées comme bien distinctes puisque c'est l'allèle I2 qui est en situation d'identité avec un allèle de J dans le premier cas, mais I1 dans le second cas. En revanche la présence de situations comme $[I1 \neq I2, J1 \neq J2, I1 \equiv J2]$, $[I1 \neq I2, J1 \neq J2, I2 \equiv J1]$ ou $[I1 \equiv I2, I1 \equiv J2, I2 \equiv J1]$, n'ont plus de justification dans la logique de Cotterman⁴, J1 et J2 étant interchangeables, non différenciés par leur origine. Ce sont les trois situations manquantes. Son regroupement en sept types majeurs (« major types »), selon les pointillés dans la figure 2 suit une évidente logique. Les pointillés verticaux définissent des colonnes à $ibd = 2, 1$ ou 0 , de gauche à droite respectivement, tandis que les pointillés horizontaux définissent $0, 1$ ou 2 individus consanguins, de haut en bas respectivement... En adoptant une telle représentation, la situation où tous les allèles sont en relation d'identité est représentée cinq fois, mais n'est comptée (et numérotée) qu'une fois (colonne de droite et ligne du bas).

4. "There are found to be 12 ways in which the four allelic genes can be connected by the equivalent relation, provided that when both relatives are allozygous [*i.e.* non identiques par ascendances], the ordering of the genes in one individual is regarded as arbitrarily fixed" (Cotterman 1983).

Il est intéressant de remarquer, comme l'ont fait Crow et Kimura (1970) et Thompson (1975), que, dans le cas où aucun individu n'est consanguin (première ligne de la fig. 2), Cotterman (1940) a été le premier à introduire les « k-coefficients » ($ibd = 2, 1$ ou 0) qui suffisent à décrire les relations entre les quatre allèles de deux individus. Dans ce cas, Cotterman donne correctement les probabilités d' $ibd = 2, 1$ ou 0 (nommées $c_2, 2*c_1, c_0$) pour des relations simples (parent-enfant, germains) bien que n'ayant que quatre « major types » (au lieu de six).

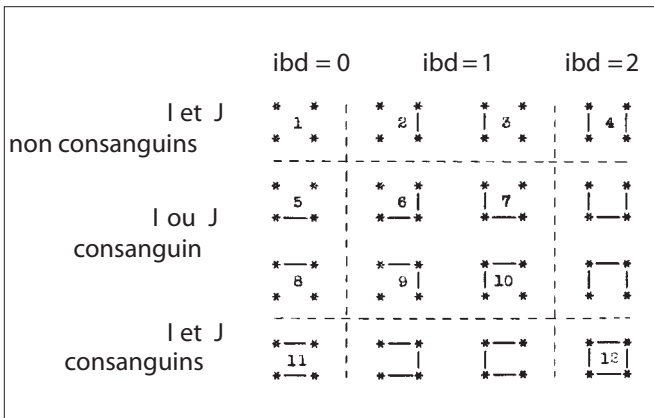


Fig. 2 - Représentation des identités entre les allèles de deux individus (Cotterman 1940).

Les situations numérotées sont de Cotterman (les situations non numérotées sont identiques à la situation 12). Les mentions d' ibd (en colonne) et de consanguinité (en ligne) ont été rajoutées.

Fig. 2—Diagram showing the identity between the alleles of two individuals I and J (Cotterman 1940). States are numbered according to Cotterman (non numbered ones are identical to the state 12). Mentions of the ibd (column) and of the consanguinity (line) have been added to the original.

Le regroupement proposé par Gillois à partir des 15 situations qu'il a identifiées pour la première fois dans sa thèse est différent. Ses cinq situations restreintes reviennent à ne plus distinguer entre les individus I et J, mais à considérer seulement les quatre allèles débarrassés de leur « personnalisation » en I ou J. Dans cette logique d'une « urne » contenant quatre allèles, on a bien les cinq situations restreintes suivantes : celles où les quatre allèles sont identiques, puis trois allèles identiques, deux, un et, finalement, aucun. Il n'est pas évident qu'il s'agisse de regroupements correspondant à une situation biologique réaliste. En tout cas, la situation restreinte de Gillois, ne

permet pas de différencier l'identité entre deux allèles en situation de consanguinité et en situation de parenté. La figure 1b montre bien que les regroupements de Gillois ne coïncident pas avec les regroupements par ibd opérés dans la figure 1a et 1c.

Enfin les neuf situations condensées de Jacquard (1970) et de Harris (1964) ne considèrent plus l'ordre des allèles de chaque individu, donc ne différencient plus leur origine maternelle et paternelle. En revanche, elles font toujours la distinction selon que I et J sont consanguins ou pas.

Ces quelques rappels historiques n'ont que l'ambition de préciser ce qui revient à chacun dans la description des situations d'identité. Ils montrent comment l'originalité d'un concept, celui de la probabilité d'identité, a surgi indépendamment dans des contextes différents (les États-Unis et la France). Mais si le travail de Malécot a eu du mal à dépasser un cercle étroit de scientifiques français, il faut également reconnaître que la thèse de Cotterman n'avait pas non plus, à l'époque, bénéficié d'une grande diffusion et n'avait probablement pas franchi l'Atlantique en son temps. Par ailleurs, la concomitance parfaite dans la parution des travaux de Gillois, sans doute ignorant de la thèse de Cotterman, et de ceux de Harris (qui ne cite pas davantage Cotterman) reste une mince consolation pour le premier, car la diffusion d'une thèse, quelle qu'en soit sa qualité, ne peut rivaliser avec celle d'un article paru dans une des meilleures revues du domaine. Cependant, contrairement à ce que laisse entendre Lange (1997), il est peu probable, ou il resterait à prouver, que Gillois et Jacquard, ou même Harris, aient « tiré profit » du travail pionnier de Cotterman. Enfin, l'intrigant reste la façon dont les uns et les autres ont envisagé les regroupements des différentes situations d'identité. Cotterman prend bien en compte l'origine des allèles de l'un des individus (maternelle ou paternelle, I1 et I2 par exemple), mais ne prend pas en considération cette origine dans la relation d'identité avec les allèles (J1 et J2) de l'autre individu. Harris, malgré sa clairvoyance dans la description de toutes les situations, n'en exploite qu'une partie ; Gillois, en probabiliste, focalise sur l'urne des allèles, mais pas sur les individus. Il faut reconnaître à Jacquard l'avantage d'avoir su tirer partie des avancées précédentes et d'avoir proposé une version complète qui clarifie définitivement les différentes relations d'identité.

Ces différentes façons de concevoir les relations d'identité entre quatre allèles de deux individus, dont l'histoire montre qu'elles ont surgi indépendamment dans des contextes scientifiques divers, ont conduit, par la

suite, à des développements, généralisations et utilisations multiples. On peut citer en particulier les analyses de grandes généalogies et les études de maladies complexes en populations isolées (Genin, Clerget-Darpoux 1996 ; Thompson 2000 ; Abney *et al.* 2000) ainsi que l'estimation

génomique des relations de parenté, initiée en 1975 par Thompson et poursuivie plus récemment grâce aux cartes denses du génome (Thompson 1975, 2008 ; Leutenegger 2003 ; Leutenegger *et al.* 2003 ; Browning 2008).

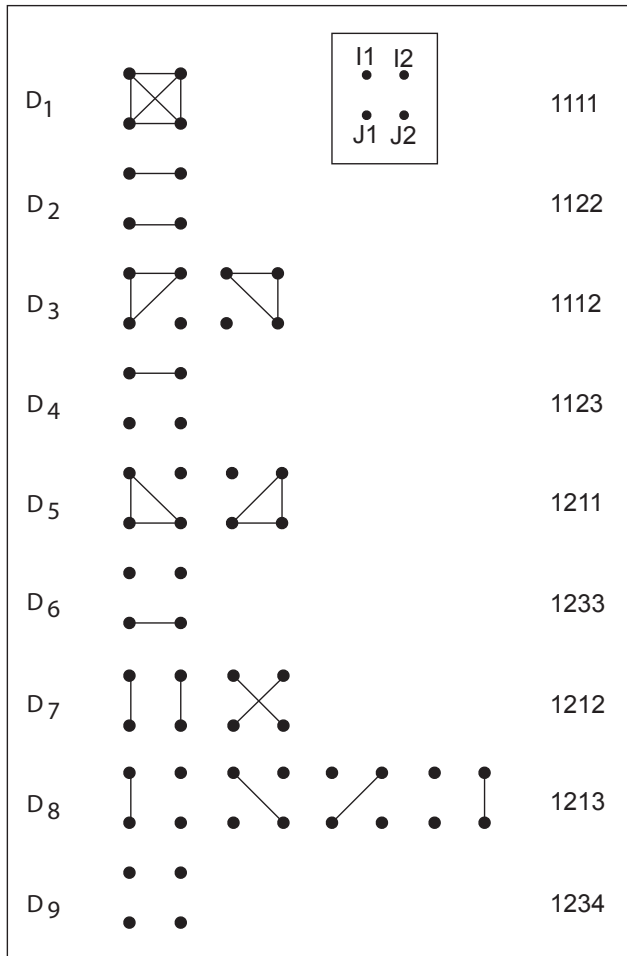


Fig. 3 - Présentation actuellement en vigueur des neuf états d'identité condensés, avec leur dénomination de un à neuf (Harris 1964 ; numérotation selon Jacquard 1970), leur représentation graphique (Jacquard 1970) et leur représentation numérique (Thompson 1974, 2000).

Fig. 3—Current presentation of the 9 condensed identity states, numbered from 1 to 9 (Harris 1964; numbering according to Jacquard 1970), with their graphic (Jacquard 1970) and numeric (Thompson 1974, 2000) presentations.

BIBLIOGRAPHIE

- ABNEY (M.), McPEEK (M.S.), OBER (C.) 2000, Estimation of variance components of quantitative traits in inbred populations, *American Journal of Human Genetics* 66, 2: 629-650.
- BALLONOFF (P.) 1974, Genetics and social structure: mathematical structuralism in population genetics and social theory, *Benchmark Papers in Genetics*, vol. 1, Ballonoff Ed., Dowden, Hutchinson & Ross, Inc. Stroudsburg, Pennsylvania, 504 p.
- BOCQUET-APPEL (J.-P.) 1996, Interview de Gustave Malécot (mai 1993-avril 1994), *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, n.s., 8 : 105-114.
- BROWNING (S.R.) 2008, Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes, *Genetics* 178, 4: 2123-2132.
- COCKERHAM (C.C.) 1971, Higher order probability functions of identity of alleles by descent, *Genetics* 69: 235-246.
- COTTERMAN (C.W.) 1940, *A calculus for Statistico-genetics*, PhD., Ohio State University, Columbus.
- COTTERMAN (C.W.) 1974, A calculus for statistico-genetics, in *Genetics and social structure: Mathematical Structuralism in Population Genetics and social theory*, *Benchmark Papers in Genetics*, vol. 1, Ballonoff Ed., Dowden, Hutchinson & Ross, Inc. Stroudsburg, Pennsylvania, p. 157-272.
- COTTERMAN (C.W.) 1983, Relationship and probability in mendelian populations, *American Journal of Medical Genetics* 16: 393-440.
- CROW (J.F.), KIMURA (M.) 1970, *An introduction to population genetics theory*, A Harper International Edition, Harper & Row, Publishers, New York. 591 p.
- EPPERSON (B.K.) 1999, Gustave Malécot, 1911-1988, population genetics founding father, *Genetics* 152: 477-484.
- FISHER (R.A.) 1918, The correlation between relatives on the supposition of Mendelian inheritance, *Transaction of the Royal Society, Edinburgh* 52: 399-433.
- GENIN (E.), CLERGET-DARPOUX (F.) 1996, Consanguinity and the sib-pair method: an approach using identity by descent between and within individuals, *American Journal of Human Genetics* 59, 5: 1149-1162.
- GILLOIS (M.) 1964, *La relation d'identité en génétique*, Université de Paris, p. 299.
- GILLOIS (M.) 1964, *La relation d'identité en génétique*, Thèse de doctorat d'État, Faculté des Sciences de Paris, série A, 299 p.
- HARRIS (D.L.) 1964, Genotypic covariances between inbred relatives, *Genetics* 50: 1319-1348.
- JACQUARD (A.) 1966, Logique du calcul des coefficients d'identité entre deux individus, *Population* 21 : 751-776.
- JACQUARD (A.) 1970, *Structures génétiques des populations*, Masson & C^{ie}, Paris.
- LANGE (K.) 1997, *Mathematical and Statistical Methods for Genetic Analysis*, Springer-Verlag, New York Inc.
- LEUTENEGGER (A.L.) 2003, *Estimation of random genome sharing; consequences for linkage*, Thèse de doctorat, Université Paris-Sud, Paris and University of Washington, Seattle.
- LEUTENEGGER (A.L.), PRUM (B.), GENIN (E.), VERNY (C.), LEMAINQUE (A.), CLERGET-DARPOUX (F.), THOMPSON (E.A.) 2003, Estimation of the inbreeding coefficient through use of genomic data, *American Journal of Human Genetics* 73, 3: 516-523.
- MALÉCOT (G.), BOCQUET-APPEL (J.-P.) 1996, Interview de Gustave Malécot (mai 1993-avril 1994), *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, n.s., 8 : 105-114.
- NAGYLAKI (T.) 1989, Gustave Malécot and transition from classical to modern population genetics, *Genetics* 122: 253-258.
- SLATKIN (M.), VEUILLE (M.) 2002, *Modern developments in theoretical population genetics, the legacy of gustave*

Malécot, Slatkin, Veuille (eds), Oxford University Press Inc, New York, 264 p.

THOMPSON (E.A.) 1974, Gene identities and multiple relationships, *Biometrics* 30: 667-680.

THOMPSON (E.A.) 1975, The estimation of pairwise relationships, *Annals of Human Genetics* 39: 173-188.

THOMPSON (E.A.) 2000, *Statistical inferences from genetic data on pedigrees*, Institute of Mathematical Statistic, Beachwood.

THOMPSON (E.A.) 2008, The IBD process along four chromosomes, *Theoretical Population Biology* 73: 369-373.

WRIGHT (S.) 1921, Systems of mating I-V, *Genetics* 6: 111-178.

WRIGHT (S.) 1922, Coefficients of inbreeding and relationship, *American Naturalist* 56: 330-338.