Twin Research and Human Genetics (2021), **24**, 200–203
doi:10.1017/thg.2021.35

CAMBRIDGE UNIVERSITY PRESS

Article

Pseudo-Random Mating with Multiple Alleles

Alan E. Stark

School of Mathematics and Statistics, The University of Sydney, Sydney, New South Wales, Australia

Abstract

The conditions on the mating matrix associated with a stable equilibrium are specified for an autosomal locus with four alleles. An example illustrates how Hardy–Weinberg proportions are maintained with nonrandom mating. The ABO blood group provides an illustration.

Keywords: Autosomal locus; four alleles; Hardy–Weinberg law; nonrandom mating

(Received 23 July 2021; accepted 12 August 2021; First Published online 16 September 2021)

Li (1988) coined the term 'pseudo-random mating' to apply to his model which demonstrated that Hardy–Weinberg proportions can be maintained with nonrandom mating for an autosomal locus with two alleles.

Stark (1980) gave the following mating system, which was used to classify some systems of partial inbreeding, given here in the original notation to avoid confusion with the notation in the main part of the article:

$$f_{ij} = f_i f_j(1 + \mu d_i d_j/S + \nu e_i e_j/T), i = 0,1,2, j = 0,1,2, \qquad (1)$$

where $d_0 = -2p$, $d_1 = q - p$, $d_2 = 2q$, $S = 2pq(1 + \lambda)$, $e_0 = -p(1 - \lambda)/(q + \lambda p)$ $e_1 = 1$, $e_2 = -q(1 - \lambda)/(p + \lambda q)$, $T = pq(1 - \lambda)(1 + \lambda)/((q + \lambda p)\ (p + \lambda q))$, $f_0 = q^2 + \lambda pq$, $f_1 = 2pq(1 - \lambda)$, $f_2 = p^2 + \lambda pq$, $\mu = 2\lambda/(1 + \lambda)$. Terms f_0 etc are the genotype frequencies in equilibrium, and so are the Hardy–Weinberg frequencies when $\lambda = 0$. When $\lambda = 0$, the component involving λ in Eqn. (1) drops out but leaves the term involving ν so that the mating frequencies $\{f_{ij}\}$ are not random frequencies unless $\nu = 0$. This demonstrates the fact that Hardy–Weinberg frequencies can be maintained with nonrandom mating. Also, it clearly identifies the separation between Hardy–Weinberg frequencies and frequencies maintained by systems of mating with inbreeding.

We have shown for an autosomal locus how, with either two or three alleles, the parental distribution can be reproduced among offspring (Stark, 2021; Stark & Seneta, 2012, 2013). A corollary of this is the fact that Hardy–Weinberg proportions can be maintained by nonrandom mating. One of the requirements of paternity experts is expressed as follows: 'knowledge of genotype frequencies in defined populations in which the polymorphism is in Hardy-Weinberg equilibrium and random mating occurs' (Geserick & Wirth, 2012, p. 164). In his definition, Buckleton (2005b, p. 68) comes very close to one of the main points of this article when he writes 'the Hardy-Weinberg law is a statement of independence between alleles at one locus' but then includes random mating as one of the conditions that make the law true.

In the scenario sketched by Clayton and Buckleton (2005, pp. 224–226), fingernail clippings have been taken from a woman who has been assaulted and claims to have scratched her attacker. Suppose that evidence (E) consists of DNA from two individuals and can be fully explained by the presence of DNA from both the woman and her suspected attacker. The authors explain the forensic approach by considering the case when the woman's DNA at a particular locus is A_1A_2 and that of the other person is A_3A_4. They calculate the value of the likelihood ratio, a ratio of probabilities:

$$LR = \frac{\text{pr}(E|G_s,G_v,H_p)}{\text{pr}(E|G_v,H_d)},$$

where H_p is the hypothesis that the nail clippings contain the DNA of the complainant and the suspect, H_d is the hypothesis that the nail clippings contain the DNA of the complainant and an unrelated person, G_s is the genotype of the suspect and G_v is the genotype of the complainant. Clayton and Buckleton find the likelihood ratio to be $1/(2p_3p_4)$ by invoking the Hardy–Weinberg proportion for the suspect. The importance of the LR is seen in the relation:

$$\text{posterior odds} = \text{likelihood ratio} \times \text{priorodds}.$$

The prior odds are the odds on the hypothesis H_p before DNA evidence, and the posterior odds are the odds after DNA evidence. If the [subjective] probability of an event is p, the odds in favor of the event are $p/(1 - p)$. Good (1950, p. 62) has a note on terminology applied to odds (o): 'If $o = m/n$ it is often said that the odds are "m to n on" or "n to m against"'. Probability can be calculated from odds by $p = o/(1 + o)$. As Buckleton (2005a) points out, the weighing of evidence is subjective, so, while the calculation of the likelihood ratio may have a scientific basis, the subjective element might be disputable.

Author for correspondence: Dr Alan E. Stark, Email: alans@exemail.com.au
Cite this article: Stark AE. (2021) Pseudo-Random Mating with Multiple Alleles. *Twin Research and Human Genetics* **24**: 200–203, https://doi.org/10.1017/thg.2021.35

In this article, we give the conditions for maintaining equilibrium for a system with four alleles. The principle could be extended to any number of alleles.

Boyd (1962, p. 335) begins his survey of blood groups, starting with ABO, as follows:

> The study of human genetics has consistently lagged behind that of lower forms. The classical Landsteiner blood groups were the first example of Mendelizing characteristics demonstrated in man, and even today few other normal hereditary characteristics have been as well studied as have the various blood-group systems.

In his tribute to Felix Bernstein, Crow (1993, p. 7) refers to 'that genetically refractory species *Homo sapiens*', citing two of Bernstein's major papers relating to the ABO system (Bernstein, 1924, 1925). Westhoff (2019) summarizes the evolving methodology and applications of genotyping.

The ABO system with four alleles as set out by Penrose (1973, pp. 25–27; p. 132) and Boyd (1962, pp. 335–337) is a relevant example. Ostrowski et al. (2020) outline the important role played by Ludwik Hirszfeld (1884–1954) in the introduction of ABO into medical practice. Apart from its importance in, for example, organ transplantation, it is studied in other specialities. Kahr et al. (2018) compared postpartum blood loss in type O and non-O women. They found a statistically significant but clinically minor increase in blood loss following delivery in women of type O. They suggest that O carriers may suffer from aggravated bleeding in the presence of additional obstetric bleeding pathologies.

Adapting the notation of Crew (1947, p. 65) the alleles would be written as H^{A1}, H^{A2}, H^{B} and H^{O}, but to conform with the general notation used with three alleles (in an earlier paper) are written as *A, B, C* and *D*. Genotypes are taken in order *AA, BB, CC, DD, AB, AC, AD, BC, BD, CD*, which are numbered from 1 to 10 and their frequencies denoted by G_i,i, = 1,2,...,10.

This article gives the conditions on the matrix of mating proportions such that the distribution of genotypes in the parents is reproduced in the offspring. The included numerical example illustrates how a parental distribution that follows the Hardy–Weinberg form can be maintained with nonrandom mating.

A Stable Population and Hardy–Weinberg Frequencies

Phenotypic identities are ignored so that the focus is on the 4 genes and 10 genotypes. There are 100 possible mating combinations, and the proportions are set out in a symmetric matrix with elements $c_{i,j}$, $i, j = 1, 2, \ldots, 10$.

The parental distribution is reproduced if the mating matrix obeys the following constraints:

$$c_{55} = 4c_{12}; c_{66} = 4c_{13}; c_{77} = 4c_{14}; c_{88} = 4c_{23}; c_{99} = 4c_{24}; c_{10,10} = 4c_{34};$$

$$c_{56} = 2c_{18}; c_{57} = 2c_{19}; c_{58} = 2c_{26}; c_{59} = 2c_{27}; c_{67} = 2c_{1,10}; c_{68} = 2c_{35};$$

$$c_{6,10} = 2c_{37}; c_{79} = 2c_{45}; c_{7,10} = 2c_{46}; c_{89} = 2c_{2,10}; c_{8,10} = 2c_{39}; c_{9,10} = 2c_{48};$$

$$c_{5,10} = c_{69} = c_{78}.$$

These restraints can be verified by calculating the progeny of each genotype and showing that it is the same as the proportion in the parents. For example, for genotype *AB*, the proportion in progeny is calculated from

Table 1 Nonrandom mating proportions which produce the same Hardy–Weinberg frequencies in offspring as in parents (multiplied by 8192)

	AA	*BB*	*CC*	*DD*	*AB*	*AC*	*AD*	*BC*	*BD*	*CD*	Total
AA	16	64	60	112	0	0	4	4	8	20	288
BB	64	1	99	136	24	12	16	0	0	40	392
CC	60	99	9	216	88	104	28	0	44	0	648
DD	112	136	216	68	32	36	80	48	60	12	800
AB	0	24	88	32	256	8	16	24	32	192	672
AC	0	12	104	36	8	240	40	176	192	56	864
AD	4	16	28	80	16	40	448	192	64	72	960
BC	4	0	0	48	24	176	192	396	80	88	1008
BD	8	0	44	60	32	192	64	80	544	96	1120
CD	20	40	0	12	192	56	72	88	96	864	1440

Genotypes are taken in order, *AA, BB, CC, DD, AB, AC, AD, BC, BD, CD*.

$$2c_{12} + c_{15} + c_{18} + c_{19} + c_{25} + c_{26} + c_{27} + \\ + (c_{55} + c_{56} + c_{57} + c_{58} + c_{59} + c_{68} + c_{69} + c_{78} + c_{79})/2.$$

Referring to the constraints and taking account of the symmetry of the mating matrix, this expression is equal to

$$\tfrac{1}{2}c_{55} + c_{51} + \tfrac{1}{2}c_{56} + \tfrac{1}{2}c_{57} + c_{52} + \tfrac{1}{2}c_{58} + \tfrac{1}{2}c_{59} + \\ + \tfrac{1}{2}c_{55} + \tfrac{1}{2}c_{56} + \tfrac{1}{2}c_{57} + \tfrac{1}{2}c_{58} + \tfrac{1}{2}c_{59} + c_{53} + \tfrac{1}{2}c_{5,10} + \tfrac{1}{2}c_{5,10} + c_{54}.$$

Collating the terms gives the sum of the elements in the 5th row of the mating matrix and so the frequency of genotype *AB* in the parents.

Table 1 is an example that illustrates how Hardy–Weinberg frequencies can be maintained with nonrandom mating. The gene frequencies are 6/32, 7/32, 9/32 and 10/32, and the genotype frequencies 36/1024, 49/1024, 81/1024, 100/1024, 84/1024, 108/1024, 120/1024, 126/1024, 140/1024 and 180/1024. Each element in the table is to be divided by 8192 to convert to a fraction.

Estimating Gene Frequencies

The frequencies of genes H^{A1}, H^{A2}, H^{B} and H^{O} are $\{p_1,p_2,p_3,p_4\}$ and are defined in terms of the parental frequencies as

$$p_1 = (2G_1 + G_5 + G_6 + G_7)/2,$$

$$p_2 = (2G_2 + G_5 + G_8 + G_9)/2,$$

$$p_3 = (2G_3 + G_6 + G_8 + G_{10})/2,$$

$$p_4 = (2G_4 + G_7 + G_9 + G_{10})/2.$$

In Crew's (1947, p. 65) notation, the correspondence between genotypes and phenotypes is as follows: A_1 ~ ($H^{A1}H^{A1}$, $H^{A1}H^{A2}$ & $H^{A1}H^{O}$); A_2 ~ ($H^{A2}H^{A2}$ & $H^{A2}H^{O}$); B ~ ($H^{B}H^{B}$ & $H^{B}H^{O}$); O ~ ($H^{O}H^{O}$); A_1B ~ ($H^{A1}H^{B}$); A_2B ~ ($H^{A2}H^{B}$). The procedure for estimating gene frequencies, from a sample, used by Hartl & Clark (1989, pp. 40–42) can be adapted for four alleles. Wherever necessary, the Hardy–Weinberg frequencies can be used to split the phenotype counts according to the following correspondences:

https://doi.org/10.1017/thg.2021.35 Published online by Cambridge University Press

$$A_1 \sim p_1^2 : 2p_1p_2 : 2p_1p_4;\ A_2 \sim p_2^2 : 2p_2p_4;\ B \sim p_3^2 : 2p_3p_4.$$

Penrose (1973, p. 132) gives the following phenotypic counts per thousand:

$$A_1 \sim 349;\ A_2 \sim 97;\ B \sim 85;\ O \sim 436;\ A_1B \sim 25;\ A_2B \sim 8.$$

The following gene frequencies are compatible with the phenotypic counts:

$$H^{A1} \sim 0.2088;\ H^{A2} \sim 0.0694;\ H^{B} \sim 0.0609;\ H^{O} \sim 0.6609$$

The method of Hartl and Clark (1989), like other methods, assumes that zygotes are formed by the union of independently drawn gametes.

Discussion

In the following quotation, Kingman (1980, p. 3) outlines an approach to population genetics which is similar to that given here in separating mating from zygote production and in contrast to the approach of Penrose (1934), which is described afterwards.

> It is convenient to have a definite model for the reproductive process in a monoecious randomly mating population. Suppose we have a population whose size N is held constant (for example, by constraints on living space or food supply). Direct attention to a particular locus. One can then imagine that each individual produces a very large number of cells called *gametes*, each of which contains only one gene at the locus. Half the gametes inherit copies of one of the individual's genes, the other half copies of the other. All the gametes produced by all the individuals are thrown into a pool, and an individual of the next generation is produced by drawing two gametes at random from the pool and combining them. The N individuals of the next generation are obtained by $2N$ independent drawings from the pool.

Kingman's model can be adapted for the present purpose by supposing N to be a very large number to eliminate random changes to the composition of the population.

Penrose (1934) is the edited version of an essay written for a competition. At the time when he wrote it, the Hardy–Weinberg distribution was widely known but not everyone realized that random mating was an assumption, not an inference from Hardy–Weinberg proportions.

Penrose applies the phrase 'the principle of random mating', in respect of an autosomal locus with two alleles, to the Hardy–Weinberg distribution. He does not attribute it to any individual thus giving the impression that it was widely known at the time. He writes: 'the principle of random mating is one of the most valuable concepts in human genetics.' (p. 25). The details of his definition are important:

> There are three genotypes formed by a pair of allelomorphic genes, D and R. If these genes are distributed at random in the general population, the three types will have the following frequencies, where x is the frequency of the gene D and $(1 - x)$ is the frequency of the gene R.

He then gives the familiar distribution of genotypes $\{x^2,2x(1-x),(1-x)^2\}$. This is followed by the remark: 'If there is random mating in the population the frequencies of these types remain constant.' (p. 26).

The important point to note in the above is the notion that random mating and the Hardy–Weinberg distribution are somehow equivalent. This notion has appeared countless times in the literature and continues to appear (Cassidy, 2021, p. 72). This article demonstrates, yet again, the flaw in the notion. The assumption embodied in $\{x^2,2x(1-x),(1-x)^2\}$ is that the zygote is formed by the union of two gametes drawn independently from the gene pool. This is possible by one of the uncountable number of mating combinations, including random mating of parents, set out above.

Penrose (1934, p. 26) mentions the ABO system:

> Now, in a homogeneous population, we should expect to find the gene responsible for agglutinogen A distributed according to the principle of random mating: the same should apply to B. If the two dominant genes are distributed *independently* in the population, we can infer a certain theoretical relation between the sizes of the classes of people having different blood types. If the two dominant genes are allelomorphic, as suggested by Bernstein, we get another theoretical distribution. Snyder has shown that, in practically every instance where a large number of individuals has been examined, the proportions of the four groups are in agreement with the expectation calculated on Bernstein's hypothesis. This result not only supports very strongly the theory that the two agglutinogens are determined by allelomorphic genes, but also fortifies our belief in the truth of the principle of random mating as applied to man. The subvarieties of the agglutinogen A have also been shown to be due to allelomorphic factors.

Penrose (1934, pp. 45–47) used data on ABO that he had collected to see whether they supported Bernstein's theory that the genes were allelomorphic and concluded that they did. Penrose and Penrose (1933) recorded the ABO type of 1000 patients of the Royal Eastern Counties Institution for Mental Defectives. As noted above, Crow (1993) reviews Felix Bernstein's important contributions.

Buchanan and Higley (1921) show the uncertainty about the genetics of ABO that existed then. Some of it was about whether an existing pathology affected agglutination. This is a different question from the notion that antigens may sometimes play a biological role (Garratty, 1996). There have been many studies looking at the association between ABO type and fitness.

Geserick and Wirth (2012) sketch the advances that enabled more accurate forensic testing, from Landsteiner's (1901) discovery of the ABO system to serum proteins, the HLA system, erythrocyte enzymes and DNA markers, a progression from phenotype to genotype level.

Ostrowski et al. (2020) is a tribute to Ludwik Hirszfeld (1884–1954), one of the pioneers of ABO research. Ludwik and his wife Hanka published their study on ABO distribution under Ludwik's German name (Hirschfeld & Hirschfeld, 1919). It contains many fascinating details, including highlights of Ludwik's collaboration with von Dungern (von Dungern & Hirschfeld, 1910).

Ludwik and Hanka give the theory of 'Landsteiner's Law of Iso-agglutination' (Hirschfeld & Hirschfeld, 1919, p. 676). They write about 'race' problems, where they mean biochemical race, which would now be referred to as allele. They give a long table of ABO phenotypic proportions observed in various races, in which the term is used conventionally. Much of these data they collected themselves among troops involved in World War II. They attempt some segregation analysis that treats A and B as unlinked loci, a few years before Felix Bernstein proposed his single locus theory. Because of the trend in phenotypic frequencies from East to West, they speculate that India was a 'cradle of one part of humanity' (p. 679). Having accepted 'Mendel's Law', they suggest the possible forensic use of ABO (p. 676).

Acknowledgment. I thank the reviewer for many constructive comments.

https://doi.org/10.1017/thg.2021.35 Published online by Cambridge University Press

References

Bernstein, F. (1924). Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen. *Klinische Wochenschrift*, *3,* 1495–1497.

Bernstein, F. (1925). Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, *37*, 237–370.

Boyd, W. C. (1962). Blood groups and soluble antigens. In W. J. Burdette (Ed.), *Methodology in human genetics* (pp. 335–365). Holden-Day.

Buchanan, J. A., & Higley, E. T. (1921). The relationship of blood-groups to disease. *The British Journal of Experimental Pathology*, *2,* 247–255.

Buckleton, J. (2005a). A framework for interpreting evidence. In J. Buckleton, C. M. Triggs & S. J. Walsh (Eds.), *Forensic DNA evidence interpretation* (pp. 27–63). CRC Press.

Buckleton, J. (2005b). Population genetic models. In J. Buckleton, C. M. Triggs & S. J. Walsh (Eds.). *Forensic DNA evidence interpretation* (pp. 65–122). CRC Press.

Cassidy, M. (2021). *Biological evolution: An introduction.* Cambridge University Press.

Clayton, T., & Buckleton, J. (2005). Mixtures. In J. Buckleton, C. M. Triggs & S. J. Walsh (Eds.), *Forensic DNA evidence interpretation* (pp. 217–274). CRC Press.

Crew, F. A. E. (1947). *Genetics in relation to clinical medicine.* Oliver and Boyd.

Crow, J. F. (1993). Felix Bernstein and the first human marker locus. *Genetics*, *133,* 4–7.

Garratty, G. (1996). Association of blood groups and disease: do blood group antigens and antibodies have a biological role? *History and Philosophy of the Life Sciences*, *18,* 321–344.

Geserick, G., & Wirth, I. (2012). Genetic kinship investigation from blood groups to DNA markers. *Transfusion Medicine and Hemotherapy*, *39,* 163–175.

Good, I. J. (1950). *Probability and the weighing of evidence.* Charles Griffin & Company.

Hartl, D. L., & Clark, A. G. (1989). *Principles of population genetics* (2nd ed.). Sinauer Associates.

Hirschfeld, L., & Hirschfeld, H. (1919). Serological differences between the blood of different races. The result of researches on the Macedonian front (Paper read before the Salonika Medical Society, June 5th, 1918). *Lancet, October 18,* 675–679.

Kahr, M. K., Franke, D., Brun, R., Wisser, J., Zimmermann, R., & Haslinger, C. (2018). Blood group O: A novel risk factor for increased postpartum blood loss? *Haemophilia*, *24*, e207–e212.

Kingman, J. F. C. (1980). *Mathematics of genetic diversity.* Society for Industrial and Applied Mathematics.

Landsteiner, K. (1901). Über Agglutinationserscheinungen normalen menschlichen Blutes. *Wiener klinische Wochenschrift*, *14,* 1132–1134.

Li, C. C. (1988). Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics*, *119*, 731–737.

Ostrowski, J., Rutkowski, B., & Rutkowski, P. (2020). Ludwik Hirszfeld (1884–1954) – Pioneer of blood type testing Significance for organ transplants. *Archives of Hellenic Medicine*, *37,* 63–67.

Penrose, L. S. (1934). *The influence of heredity on disease.* H. K. Lewis & Co.

Penrose, L. S. (1973). *Outline of human genetics* (3rd ed.). Heinemann Educational Books.

Penrose, M., & Penrose, L. S. (1933). The blood-group distribution in the Eastern Counties of England. *The British Journal of Experimental Pathology*, *14,* 160–161.

Stark, A. E. (1980). Inbreeding systems: classification by a canonical form. *Journal of Mathematical Biology*, *10,* 305.

Stark, A. E. (2021). A misconception about the Hardy-Weinberg law. *Twin Research and Human Genetics*, *24*, 160–162.

Stark, A. E., & Seneta, E. (2013). A reality check on Hardy-Weinberg. *Twin Research and Human Genetics*, *16*, 782–789.

Stark, A., & Seneta, E. (2012). On S.N. Bernstein's derivation of Mendel's law and 'rediscovery' of the Hardy-Weinberg distribution. *Genetics and Molecular Biology*, *35*, 388–394.

von Dungern, E., & Hirschfeld, L. (1910). Über Vererbung gruppenspezifischer Strukturen des Blutes. *Zeitschrift für Immunitätsforschung*, *6,* 284–292.

Westhoff, C. M. (2019). Blood group genotyping. *Blood*, *133,* 84–1820.

https://doi.org/10.1017/thg.2021.35 Published online by Cambridge University Press