Genetics and Molecular Biology, 30, 3, 664-666 (2007)

www.sbg.org.br

Research Article

On extending the Hardy-Weinberg law

Alan E. Stark

Balgowlah NSW, Australia

Abstract

This paper gives a general mating system for an autosomal locus with two alleles. The population reproduces in discrete and non-overlapping generations. The parental population, the same in both sexes, is arbitrary as is that of the offspring and the gene frequencies of the parents are maintained in the offspring. The system encompasses a number of special cases including the random mating model of Weinberg and Hardy. Thus it demonstrates, in the most general way possible, how genetic variation can be conserved in an indefinitely large population without invoking random mating or balancing selection. An important feature is that it provides a mating system which identifies when mating does and does not produce Hardy-Weinberg proportions among offspring.

Key words: Hardy-Weinberg law, non-random mating, general offspring distribution.

Received: December 15, 2006; Accepted: April 23, 2007.

Introduction

This paper gives a general mating system for an autosomal locus with two alleles. The population reproduces in discrete and non-overlapping generations. The system encompasses a number of special cases including the random mating model of Weinberg (1908) and Hardy (1908). It covers also the formulation of Li (1988) and Stark (2005) who showed that Hardy-Weinberg (H-W) frequencies can be maintained in large populations with non-random mating. Furthermore it subsumes the system of Stark (2006a) which demonstrates that Hardy-Weinberg proportions (HWP) can be attained in one round of non-random mating. It is more general than the last of these in that it produces an arbitrary distribution of genotypes in the offspring from an arbitrary distribution in the parents while maintaining the gene frequencies of the parents.

The next section defines the mating system. The following section demonstrates how it encompasses a number of special cases. The last section discusses the canonical representation of the model and includes some numerical examples.

The general mating system

Consider a population with respect to a single locus having alleles A and B with respective frequencies q and p, the same in males and females. Denote frequencies of genotypes AA, AB and BB among parents by f_0 , f_1 and f_2 and among offspring by g_0 , g_1 and g_2. Table 1 gives a mating system in which reciprocal crosses have the same frequency so that the roles of males and females can be reversed without changing the model. The 3×3 matrix of cell frequencies will be denoted by $[f_{ij}]$, i = 0, 1, 2; j = 0, 1, 2. Without loss of generality q is taken in the interval $0 < q \leq 1/2$. Since the elements of $[f_{ij}]$ are non-negative there are constraints on the values of F, G, s and t.

Summing the elements of Table 1 by rows and columns shows that the parental genotypic frequencies are: $f_0 = q^2 + Fpq$, $f_1 = 2pq - 2Fpq$, $f_2 = p^2 + Fpq$, F being Sewall Wright's *fixation index* Thus the parental frequencies are in the most general form, defined by values of q and F. Making the usual assumptions it can be seen that the distribution of genotypes among offspring is $g_0 = q^2 + Gpq$, $g_1 = 2pq - 2Gpq$, and $g_2 = p^2 + Gpq$. Because matrix $[f_{ij}]$ is symmetric the distribution of genotypes among offspring can be calculated from:

$$g_0 = f_{00} + f_{01} + 1/4\, f_{11}$$

$$g_1 = f_{01} + 2f_{02} + 1/2\, f_{11} + f_{21}$$

$$g_2 = f_{22} + f_{21} + 1/4\, f_{11}$$

Table 1 - General mating system.

M x F	*AA*	*AB*	*BB*
AA	$q^2 - s - t$	$Gpq + t$	$Fpq - Gpq + s$
AB	$Gpq + t$	$4s$	$2pq - 2Fpq - Gpq - 4s - t$
BB	$Fpq - Gpq + s$	$2pq - 2Fpq - Gpq - 4s - t$	$p^2 - 2pq + 2Fpq + 2Gpq + 3s + t$

M: male. F: female.

Send correspondence to Alan E. Stark. 3/20 Seaview Street, Balgowlah NSW, Australia 2093. E-mail: alans@exemail.com.au.

Note that the gene frequencies among the offspring are identical to those of the parents. However the genotypic distribution among the offspring is arbitrary being determined by G which plays the same role in offspring as F does in parents. In particular, taking $G = 0$ gives Hardy-Weinberg proportions (HWP) among the offspring. A numerical illustration is given in Table 2 which is discussed in the final section. It is specified by $q = 0.4$, $F = 1/6$, $G = 0$, $s = 0.05$, $t = 0.02$ and $f_0 = 0.2$, $f_1 = 0.4$, $f_2 = 0.4$. The distribution among offspring is $g_0 = 0.16$, $g_1 = 0.48$, and $g_2 = 0.36$. Clearly mating is not random yet the offspring proportions are Hardy-Weinberg.

Special cases

Random mating is defined in Table 1 by putting $s = 1/4\, f_1^2$ and $t = f_0 f_1$. The offspring are distributed in HWP so that $G = 0$ completes the specification.

The mating system given by Li (1988) is reproduced in Table 3. Since both parents and offspring are distributed in HWP both $F = 0$ and $G = 0$. Li's parameters and those of Table 1 are related by $a = pq^2(1 + q) - (s + t)$ and $b = s - p^2q^2$. In Li's model random mating is defined by the pair of conditions $a = 0$ and $b = 0$ so that s and t in Table 1 are then $s = p^2q^2$ and $t = 2pq^3$.

The model given by Stark (2006a) is obtained by taking $G = 0$. A particular case is obtained by taking $F = 1/2\,(p - q)/p$ and forcing $f_{00} = 0$ and $f_{11} = 0$. This case is given by Table 4 and considered further in the next section.

Table 2 - Mating scheme with $q = 0.4$, $F = 1/6$, $G = 0$, $s = 0.05$ and $t = 0.02$.

M x F	*AA*	*AB*	*BB*
AA	0.09	0.02	0.09
AB	0.02	0.20	0.18
BB	0.09	0.18	0.13

M: male. F: female.

Table 3 - Li's symmetric non-random mating model.

M x F	*AA*	*AB*	*BB*
AA	$q^4 + a$	$2pq^3 - a - b$	$p^2q^2 + b$
AB	$2pq^3 - a - b$	$4p^2q^2 + 4b$	$2p^3q + a - 3b$
BB	$p^2q^2 + b$	$2p^3q + a - 3b$	$p^4 - a + 2b$

M: male. F: female.

Table 4 - Mating scheme with $F = 1/2\,(p - q)/p$, $G = 0$, $s = 0$ and $t = q^2$.

M x F	*AA*	*AB*	*BB*
AA	0	q^2	$1/2q(p - q)$
AB	q^2	0	pq
BB	$1/2q(p - q)$	pq	$p(p - q)$

M: male. F: female.

The canonical representation of Table 1

It is instructive to examine $[f_{ij}]$ through its canonical form

$$f_{ij} = f_i f_j(1 + \rho x_i x_j + \sigma y_i y_j),\ (i = 0, 1, 2;\ j = 0, 1, 2). \quad (1)$$

Formula (1) is a particular example of the representation of a discrete bivariate probability distribution which Lancaster (1969, p. 90) refers to as "Fisher's Identity". Denote the vector of values $\{x_0, x_1, x_2\}$ by $\boldsymbol{x}$ and $\{y_0, y_1, y_2\}$ by $\boldsymbol{y}$. Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ attribute two sets of values to the genotypes of the parents, the same for males and females. To simplify the exposition it helps to define some expressions involving the elements of $[f_{ij}]$ and the parental genotypic frequencies f_0, f_1 and f_2:

$$W = f_{00} f_{11} f_{22} + 2 f_{01} f_{02} f_{12} - (f_{00} f_{12}^2 + f_{11} f_{02}^2 + f_{22} f_{01}^2). \quad (2)$$

$$X = f_0 f_{12}^2 + f_1 f_{02}^2 + f_2 f_{01}^2 - (f_0 f_{11} f_{22} + f_1 f_{00} f_{22} + f_2 f_{00} f_{11}). \quad (3)$$

$$Y = f_0 f_1 f_{22} + f_0 f_2 f_{11} + f_1 f_2 f_{00}. \quad (4)$$

Next form the following quadratic in ν:

$$(W + X + Y)\nu^2 + (W + X)\nu + W = 0. \quad (5)$$

Solve the quadratic and designate the two solutions of ν as ρ and σ. Then ρ is the correlation of $\boldsymbol{x}$ in female parents with $\boldsymbol{x}$ in male parents and σ is the correlation of $\boldsymbol{y}$ in females with $\boldsymbol{y}$ in males.

Finally the vector $\boldsymbol{x}$ can be calculated by solving the set of equations

$$\Sigma_j f_i x_i = 0,\ \Sigma_j f_i x_i^2 = 1,\ \Sigma_i \Sigma_j f_{ij} x_i x_j = \rho, \quad (6)$$

and the vector $\boldsymbol{y}$ from

$$\Sigma_j f_i y_i = 0,\ \Sigma_j f_i y_i^2 = 1,\ \Sigma_i \Sigma_j f_{ij} y_i y_j = \sigma. \quad (7)$$

Some modification of the solution to Eqs. (5) - (7) is necessary for special cases. For example the formulation given by Table 1 can include the cases $f_0 = 0$ and $f_0 = f_2 = 0$.

The solution of Eqs. (5) - (7) involves rather unwieldy algebraic expressions although solutions can be obtained for particular numerical examples. One root of (5) is zero if $W = 0$. Suppose this is ρ, then (1) reduces to

$$f_{ij} = f_i f_j(1 + \sigma y_i y_j),\ (i = 0, 1, 2;\ j = 0, 1, 2). \quad (8)$$

However, even this may not yield simple expressions. One case is that given by Stark (2006a) where the entries in (8) are defined by

$$y_0 = T^{-1/2} p(F - 1)/(q + Fp),$$

$$y_1 = T^{-1/2},$$

$$y_2 = T^{-1/2} q(F - 1)/(p + Fq),$$

and

$T = pq(1 - F^2)/((q + Fp)(p + Fq))$.

Then, in Table 1, $G = 0$, $s = 1/4\, f_1^2(1 + \sigma T^{-1})$ and $t = f_0 f_1(1 + \sigma T^{-1} p(F - 1)/(q + Fp))$.

A special case of the preceding example is given in Table 4. The canonical form is expressed by $\rho = 0$, $\sigma = -q(3-4q)/(2-3q)$, $y_0 = -\tau$, $y_1 = \tau$, $y_2 = -\tau q/(2-3q)$, where $\tau = 1/\sqrt{(-\sigma)}$. These terms satisfy Eqs. (5) and (7).

Simplifying Li's model (Table 3) by putting $b = a$ yields another example: then $F = 0$, $G = 0$; also $W = 0$, $X = -2apq$, $Y = 2pq(a + p^2q^2)$, $s = p^2q^2 + a$, $t = 2pq^3 - 2a$ and

$$\rho = 0,\ \sigma = a/(p^2q^2),$$

$$x_0 = -2p/\sqrt{(2pq)},$$

$$x_1 = (q - p)/\sqrt{(2pq)},$$

$$x_2 = 2q/\sqrt{(2pq)},$$

$$y_0 = -p/q,\ y_1 = 1,\ y_2 = -q/p.$$

Note that in this case the vector $\boldsymbol{x}$ is a set of *additive* values, that is with the property $x_2 - x_1 = x_1 - x_0$, as pointed out by Stark (2006b), and the set $\boldsymbol{y}$ is that given by Stark (2005). Since $\rho = 0$ the elements in $[f_{ij}]$ are obtained from Eq. (8).

Another system is defined by $f_{ij} = f_i f_j(1 + \rho x_i x_j)$, where $x_0 = -2pV^{-1/2}$, $x_1 = (q - p)V^{-1/2}$, $x_2 = 2qV^{-1/2}$ and $V = 2pq(1 + F)$. This model was given by Stark (1976a, 1976b). It has the property that if ρ is fixed at value $2F/(1 + F)$ then the parental distribution characterized by q and F is reproduced in the offspring, that is $G = F$. Again $\boldsymbol{x}$ is additive and the correlation between mates based on $\boldsymbol{x}$ is $\rho = 2F/(1 + F)$. In the notation of Table 1, $s = 1/4 f_{11}$ and $t = f_{01} - Gpq = f_{01} - Fpq$.

Table 2 was introduced earlier. Its canonical form is:

$$\rho = 2/5,\ \sigma = -1/8$$

$$x_0 = -2\sqrt{(5/6)} = -1.826,\ x_1 = \sqrt{(5/6)} = 0.913,\ x_2 = 0,$$

$$y_0 = y_1 = \sqrt{2/3} = 0.816,\ y_2 = -\sqrt{(3/2)} = -1.225$$

Table 5 contains the numerical example defined by $q = 1/3$, $F = 1/4$, $G = -1/4$, $s = 1/18$ and $t = 1/18$. The distribution of parental types is $f_0 = 3/18$, $f_1 = 6/18$, $f_2 = 9/18$ and the distribution among offspring is $g_0 = 1/18$, $g_1 = 10/18$, and $g_2 = 7/18$. The terms to be substituted in formula (1) are as follows:

$$\rho = (1 + \sqrt{73})/18 = 0.530,\ \sigma = (1 - \sqrt{73})/18 = -0.419$$

$$x_0 = 2u - v = 1.074,$$

$$x_1 = -(u + v) = -1.392,$$

$$x_2 = v = 0.570,$$

$$y_0 = u + 2v = 1.961,$$

$$y_1 = u - v = 0.252,$$

$$y_2 = -u = -0.822,$$

where $u = \sqrt{(73 + 3\sqrt{73})}/\sqrt{146}$ and $v = \sqrt{(73-3\sqrt{73})}/\sqrt{146}$.

Table 5 - Mating scheme with $q = 1/3$, $F = 1/4$, $G = -1/4$, $s = 1/18$ and $t = 1/18$.

M x F	*AA*	*AB*	*BB*
AA	0	0	3/18
AB	0	4/18	2/18
BB	3/18	2/18	4/18

M: male. F: female.

The preceding examples show that the mating system given in Table 1 is a general model which conserves genetic variation but allows genotypic distributions which are not exclusively in Hardy-Weinberg form. In fact it provides a mating system which identifies when mating does and does not produce Hardy-Weinberg proportions among offspring.

References

Hardy GH (1908) Mendelian proportions in a mixed population. Science 28:49-50.

Lancaster HO (1969) The Chi-Squared Distribution. John Wiley & Sons, Inc., New York, 356 pp.

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

Stark AE (1976a) Generalisation of the Hardy-Weinberg law. Nature 259:44-44.

Stark AE (1976b) Hardy-Weinberg law: Asymptotic approach to a generalized form. Science 193:1141-1142.

Stark AE (2005) The Hardy-Weinberg principle. Genet Mol Biol 28:485-485.

Stark AE (2006a) A clarification of the Hardy-Weinberg law. Genetics 174:1695-1697.

Stark AE (2006b) Stages in the evolution of the Hardy-Weinberg law. Genet Mol Biol 29:589-594.

Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. Jahresh Verein f vaterl Naturk Württem 64:368-382. English version: On the demonstration of heredity in Man. In: Boyer SH (ed) Papers on Human Genetics. Prentice-Hall, Englewood Cliffs, 1963, pp 4-15.

Associate Editor: Paulo A. Otto