The X-Linked Analog of the Hardy–Weinberg Law

Alan E. Stark

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

The system of mating that maintains a general genotypic distribution among females with respect to an X-linked locus is defined. In particular, it is shown that Hardy–Weinberg proportions can be maintained with non-random mating.

Dronamraju (2017, p. 82) asserts that 'population genetics' began as an attempt to 'marry Darwin's theory of evolution with the science of genetics that was founded by … Mendel'. He states that the 'next important step' (in population genetics theory) was the introduction of the Hardy–Weinberg law in 1908 (Hardy, 1908; Weinberg, 1908). He says that the law is strictly valid only if the following conditions are valid:

(1) the population must be large enough so that sampling errors can be ignored;
(2) there must be no mutation;
(3) there must be no selective mating;
(4) there must be no selection.

Conditions similar to the above four can be found in many introductory texts on population genetics, generally in relation to autosomal loci, as were the original formulations. It appears that Dronamraju is excluding assortative mating in the third of his conditions and may intend that the selection of mates should be random. As can be seen from references cited below, random mating is not a necessary requirement for the maintenance of Hardy–Weinberg proportions. The purpose of this paper is to show that the same is true of an X-linked locus.

We take locus Xg as envisaged by Mann et al. (1962) as archetypical. These authors give the set of estimated genotype frequencies reproduced in Table 1. They used the gene frequencies among males to represent the population frequencies and applied Hardy–Weinberg proportions to calculate the female genotype frequencies. In Table 1, the frequency of Xg^a in males is 0.6169 and the combined frequency of $Xg^a Xg^a$ and $Xg Xg^a$ in females is 0.8532. This is in approximate agreement with the frequencies of $Xg(a_+)$ (60% in males and 90% in females) given by Mueller and Young (1995). In our notation, introduced below, Xg is

given the label U, and Xg^a label T. Johnson (2011) and Tippett and Ellis (1998) review the XG system.

Mann et al. (1962) give 'certain rules which may be laid down for an X-borne dominant antigen' (p. 9); for example, that from the mating positive × positive, there can be no negative daughters. In this article, the alleles are treated as co-dominant. In their analysis, these authors calculate the proportions of expected matings by assuming, for example, that the frequency of Xg(a-) father by Xg(a-) mother is the product of the respective genotype frequencies, that is, equivalent to random mating frequencies. They compute the expected proportions of Xg(a+) and Xg(a-) male and female children from the observed numbers of female genotype frequencies and use them to calculate the expected numbers of male and female children of each type. These are then compared with the observed numbers from 50 sibships, finding satisfactory agreement with the hypothesis of X-linked inheritance. In their analysis, because of dominance, there are four mating types. In our analysis, there are six mating pair combinations.

Johnson (2011) states 'the function of the Xg^a protein is unknown' (p. 68). Tippett and Ellis (1998) state '... anti-Xg^a does not appear to be clinically significant' (p. 234). They give a table of gene frequencies that are reproduced in Table 2. In the light of this, it seems not unreasonable to treat the Xg locus as an example of a stable polymorphism with equally viable genotypes.

The monograph of Thomas Nagylaki (1977) gives a sound account of basic population genetics as it existed

TABLE 1

Estimated Genotype Frequencies at the Xg Locus

Males	Females
Xg^a Y 0.6169	Xg^a Xg^a 0.3806
Xg Y 0.3831	Xg^a Xg 0.4726
	Xg Xg 0.1468

TABLE 2

Xg^a Gene Frequencies in Different Populations

Population	Total tested	Xg^a	Xg
N. Europeans	11,716	0.66	0.34
Sardinians	322	0.76	0.24
Greeks	638	0.55	0.45
Barcelona, Spaniards	636	0.59	0.41
New York, Jamaica, Blacks	219	0.55	0.45
Singapore, Chinese	101	0.45	0.55
Japanese	529	0.68	0.32
Australian Aborigines	352	0.79	0.21
New Guineans	263	0.85	0.15

at the time of publication. In most respects, the theory as expounded by Nagylaki is still current. In the chapter entitled 'Panmictic Populations', Nagylaki starts with 'the genetic structure of a randomly mating population in the absence of selection, mutation, and random drift' (p. 33). He says of this theory: 'This part of population genetics was the first to be understood, and a thorough grasp of its principles is required for the formulation and interpretation of most evolutionary models' (p. 33).

Nagylaki (1977) gives theory for several alleles at an autosomal locus whereas we take only two. Also, he uses ordered genotype frequencies, so that in his notation, when P_{ij} ($= P_{ji}$) designates the frequency of ordered A_iA_j genotypes, $2P_{ij}, i \neq j$, is the frequency of unordered A_iA_j heterozygotes. For convenience, we use unordered genotypes so that a single subscript serves to distinguish genotypes. Using his notation, Nagylaki calculates allele frequencies as

$$p_i = \sum_j P_{ij}.$$

It is relevant to quote from Nagylaki's (1977, p. 34) monograph:

By Mendel's Law of Segregation, p_i is the frequency of A_i in the gametic output of the population.

If mating occurs without regard to the genotype at the A-locus, random union of gametes yields the genotypic proportions

$$P'_{ij} = p_i p_j$$

in the next generation. Therefore, the gene frequencies do not change,

$$p'_i = \sum_j P'_{ij} = p_i,$$

and Hardy–Weinberg proportions,

$$P'_{ij} = p'_i p'_j,$$

are attained in a single generation.

Nagylaki (1977, p. 34) states further that 'The most important aspect of the Hardy-Weinberg law is the constancy of the allelic frequencies'.

Nagylaki also considers matings explicitly but soon resorts to random mating. He introduces different frequencies for the two sexes and reaches the identity

$$P'_{ij} = Q'_{ij} = \tfrac{1}{2}(p_i q_j + p_j q_i),$$

where P and p apply to male and Q and q apply to female entities. He uses the same approach for autosomal and X-linked loci to reach what he calls 'generalized Hardy–Weinberg proportions' (Nagylaki, 1977, p. 36).

The point which we emphasize is that, in respect of Hardy–Weinberg equilibrium, using frequencies of mating pairs, it is not necessary to invoke random mating. It is implicit in a formula of Stark (1980) that, for an autosomal locus, Hardy–Weinberg frequencies are consistent with non-random mating. Stark (2006) showed that Hardy–Weinberg frequencies can be attained in a single round of non-random mating. Stark and Seneta (2013, 2014) show how general genotypic proportions can be maintained.

Nagylaki (1977) shows that for an X-linked locus with random mating, not only are gene frequencies equalized in the two sexes, but Hardy–Weinberg proportions are approached rapidly. In this paper, we assume that gene frequencies in males are the same as those in females.

The object of this paper is to show how a general equilibrium at an X-linked locus can be sustained in females. Just as in autosomal loci, Hardy–Weinberg frequencies can be maintained with non-random mating at an X-linked locus. The condition required to maintain equilibrium is given in the next section. The boundaries of the region of admissible points of equilibrium are given in the following section.

The Mating System

This is a model for a single X-linked locus with two alleles U and T with frequencies in the population q and p ($q + p = 1$). We have in mind the human population in which females have two X chromosomes and males one. We assume that the population is in equilibrium, the genotypes equally viable and the gene frequencies the same in both sexes. The frequencies of genotypes UU, UT, TT in the females are denoted, respectively, f_0, f_1, f_2 ($f_0 + f_1 + f_2 = 1$) and the frequencies of male hemizygotes U and T are denoted, respectively, m_0 and m_1 ($m_0 + m_1 = 1$). The frequency of U in females is $f_0 + \tfrac{1}{2} f_1 = q$ and in males is $m_0 = q$. Without loss of generality, q is taken to be less than or equal to ½.

The mating scheme is represented by

$$\begin{bmatrix} U \times UU & U \times UT & U \times TT \\ T \times UU & T \times UT & T \times TT \end{bmatrix}$$

with commensurate mating frequencies given by the matrix

$$C = \begin{bmatrix} f_{00} & f_{01} & f_{02} \\ f_{10} & f_{11} & f_{12} \end{bmatrix}.$$

The row sums of C are q and p and the column sums are f_0, f_1, f_2, so these are the parental frequency distributions. We use C in the extended row vector form

$$u' = \{f_{00}, f_{01}, f_{02}, f_{10}, f_{11}, f_{12}\}.$$

To follow the progression of generations, which are assumed to be discrete and non-overlapping, we need Mendel's coefficients of heredity, given in matrix form, for female offspring, by

$$M = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

Then, the frequency distribution of juvenile females is calculated from

$$j' = (Mu)'$$

which in detail is

$$j = \{f_{00} + \tfrac{1}{2}f_{01}, \tfrac{1}{2}f_{01} + f_{02} + f_{10} + \tfrac{1}{2}f_{11}, \tfrac{1}{2}f_{11} + f_{12}\}'.$$

The distribution of male juveniles is q of type U and p of type T.

The main point of interest is to specify the properties of C which satisfy

$$j = \{f_0, f_1, f_2\}', \tag{1}$$

that is, that the distribution of juvenile females is the same as the distribution of adult females. The special distribution

$$a^* = \{q^2, 5pq, p^2\}' \tag{2}$$

will be referred to as the Hardy–Weinberg distribution even though both Hardy and Weinberg considered only the case of an autosomal locus. Other distributions can be put in the form

$$a = \{q^2 + Fpq, 2pq - 2Fpq, p^2 + Fpq\}'. \tag{3}$$

Parameter F, as well as q, serves to specify details of the system.

Equation (1) is satisfied if

$$f_{01} = 2f_{10}. \tag{4}$$

If $q = 1/3$, $F = 1/12$, $a = \{7/54, 22/54, 25/54\}'$,

$$C = \frac{1}{54}\begin{bmatrix} 3 & 8 & 7 \\ 4 & 14 & 18 \end{bmatrix} \tag{5}$$

has property (4), thereby satisfying condition (1). Matrix (5) is only one of an infinite number which could be found to satisfy (1). The force of (4) can be seen by exploiting the fact that the elements of the first row of C sum to q, as do the sum of the elements of the first column and half of each element of the second column, leading to the identity:

$$f_{00} + f_{01} + f_{02} = f_{00} + f_{10} + \tfrac{1}{2}(f_{01} + f_{11}). \tag{6}$$

Substituting from (4) into (6) leads to the implied property

$$f_{11} = 2f_{02}. \tag{7}$$

Identity (4) ensures that juvenile females of type UU have frequency f_0, (7) that those of type TT have frequency f_2, and the heterozygotes have frequency f_1 since the frequencies sum to unity. Thus, given the marginal sums of C, nominating elements f_{01} and f_{10}, which conform to (4) and are compatible with marginal quantities, enable the construction of C satisfying (1).

Taking $F = 0$ produces the Hardy–Weinberg distribution among adult females. Random mating, defined by

$$f_{sd} = m_s \times f_d, \ s = 0, 1; d = 0, 1, 2, \tag{8}$$

satisfies (1), but is only one of an infinite number of mating schemes with this property.

Admissible Points of Equilibria

The mating matrix C must be consistent with various mathematical, as well as biological constraints. These are conveniently depicted by points within and on the sides of a figure defined by the pair of coordinates F and f_{01}, using a unique planar figure for each value of q. Figure 1 displays the admissible region, whose vertices are QVDE, for $q = 2/5$. Given a value of q, for a given F, admissible definitions of C are represented by points (values of f_{01}) along the vertical line above F within or on the boundary of the appropriate polygon QVDE. The base of the defining triangle extends from $-q/p$ to 1, the admissible range of F. The maximum height of the triangle is $f_{01} = q$, when $F = (2p - 1)/(2p)$, the mid-point of the base. The equation of the side of the triangle from $-q/p$ to the vertex is $f_{01} = 2(q^2 + Fpq)$ and of the side from 1 to the vertex is $f_{01} = 2pq - 2Fpq$. When $\tfrac{1}{3} < q \leq \tfrac{1}{2}$, the line with equation $f_{01} = 2p(q - p - 2Fq)$ is another boundary. These three equations, together with $f_{01} = 0$, define the admissible region of the system for a specified value of q. Point Z in Figure 1 has $\{F, f_{01}\}$ coordinates $\{0, 0\}$, Q has coordinates $\{(2q - 1)/(2q), 0\}$ and point E coordinates $\{(3pq - 1)/(3pq), 2q - \tfrac{2}{3}\}$. Figure 1 shows the point of random mating (R) when $F = 0$ and
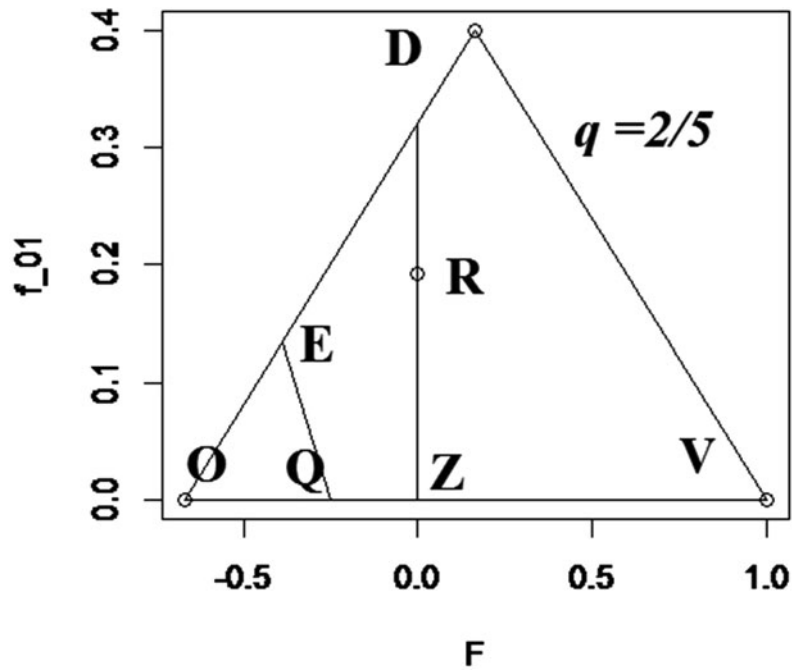
FIGURE 1
Admissible region, defined by QVDE, when $q = 2/5$; R indicates point of random mating.

$f_{01} = {}^{24}/125$ highlighting the fact that random mating is only one point of an infinite number, on the vertical line through R, which are consistent with Hardy–Weinberg frequencies. When $q \leq \frac{1}{3}$, points O and Q coalesce, so that the boundary of the admissible region is OVD.

References

Dronamraju, K. (2017). *Popularizing science: The life and work of JBS Haldane*. New York, NY: Oxford University Press.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28*, 49–50.

Johnson, N. C. (2011). XG: The forgotten blood group system. *Immunohematology, 27*, 68–71.

Mann, J. D., Cahan, A., Gelb, A. G., Fisher, N., Hamper, J., Tippett, P.,... Race, R. R. (1962). A sex-linked blood group. *Lancet, 1*, 8–10.

Mueller, R. F., & Young, I. D. (1995/6). *Emery's elements of medical genetics*. Edinburgh, Scotland: Churchill Livingstone.

Nagylaki, T. (1977). *Selection in one- and two-locus systems*. Berlin, Germany: Springer-Verlag.

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

Stark, A. E. (2006). A clarification of the Hardy–Weinberg law. *Genetics, 174*, 1695–1697.

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

Stark, A. E., & Seneta, E. (2014). Hardy–Weinberg equilibrium as foundational. *International Journal of Statistics in Medical Research, 3*, 198–202.

Tippett, P., & Ellis, N. A. (1998). The Xg blood group system: A review. *Transfusion Medicine Reviews, 12*, 233–257.

Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahresh Verein f Vaterl Naturk Württem, 64*, 368–382.