# Admixture Models and the Breeding Systems of H. S. Jennings: A *GENETICS* Connection

**Noah A. Rosenberg**
Department of Biology, Stanford University, Stanford, California 94305-5020

*The lions of the field have published seminal work in our journal, from Bridges, Muller, and McClintock to Brenner, Horvitz, and Hartwell. Authors can publish in the same journal as did Luria and Delbruck, and Sewall Wright, and Ronald Fisher, and Crow and Kimura (and many other luminaries). And while GENETICS provides a professional and scientific thread that extends back to the founders of our field, it also points to our future.*

JOHNSTON 2014, p. 10

LONG-TIME *GENETICS* readers are bound to learn (see the perspective by Ganetzky and Hawley 2016, this issue) that the journal's first article, Vol. 1, p. 1, reported a foundational study whose results continue to occupy a central place in the field: the celebrated work of Calvin Bridges on "Nondisjunction as proof of the chromosome theory of heredity" (Bridges 1916).

What other treasures lie hidden in the 100-year record of *GENETICS*, and what links do they have to ongoing events in the field? *GENETICS* authors can delight in the fact that every so often, the search for the origin of an idea uncovers an unexpected connection in an early volume. This essay gives an account of one such quest that led to the *second* article in the first issue of the journal, Vol. 1, p. 53, by H. S. Jennings on "The numerical results of diverse systems of breeding" (Jennings 1916).

## Genetic Admixture

Amy Goldberg and I had been studying genetic admixture, a topic of current interest frequently featured in recent issues of *GENETICS* (*e.g.*, Lohmueller *et al.* 2011; Verdu and Rosenberg 2011; Gravel 2012; Patterson *et al.* 2012; Duchen *et al.* 2013; Loh *et al.* 2013; Liang and Nielsen 2014; Lohse and Frantz 2014; Sanderson *et al.* 2015; Sedghifar *et al.* 2015). In geneti-

cally admixed populations—populations formed from genetic mixing of two or more groups that have long been separated—just as in a classic $F_2$ cross, individual genomes consist of a mosaic of regions with different ancestries. The proliferation of population-genetic data, statistical advances in inferring the populations of origin of genomic segments, interest in ancient admixture, and the prospects for mapping trait loci in admixed populations formed by natural experiments have all helped give rise to substantial activity in the study of genetic admixture.

Our focus was sex-biased admixture, in which one or more of the ancestral sources for an admixed population has a difference in its levels of female and male contribution. In humans, sex-biased admixture has been common, often with a bias toward males in invading populations and toward females in native groups. Because the X chromosome and autosomes have different modes of inheritance, X chromosomes record signatures of the founding females and males differently from autosomes. Having already built a general mathematical model of the effect of sex-biased admixture on autosomal markers (Goldberg *et al.* 2014), our goal was to do the same for the X chromosome.

## An Admixture Model

Following our earlier model, we considered two source populations ("1" and "2"), one or both of which has a sex bias in its contributions to an admixed population. In a formulation with discrete generations, we examined the expected X-chromosomal admixture from a specific source population for a random female and a random male chosen in the admixed population at a specified generation—in other words, the admixture level predicted by the model for a

randomly chosen admixed female or male. A particular interest was "hybrid isolation," the case in which after the initial generation of admixture, no further contributions from the source groups enter the admixed population (Long 1991).

Discrete-time models are often studied by obtaining recursions that describe the state of a system in terms of that in the previous time step. Amy used a recursion approach and found that as a function of time, the expected X-chromosomal admixture could be computed using an iteratively defined sequence $J_n$, which after initial conditions $J_0 = 0$ and $J_1 = 1$ follows the formula $J_n = J_{n-1} + 2J_{n-2}$. If the initial generation of offspring $n = 1$ in an admixed population begins with a proportion $s_f$ of its mothers and $s_m$ of its fathers from source population 1, then under hybrid isolation, the random variables $H_{n,f}$, the proportion of X-chromosomal ancestry from source 1 for a random female in the admixed population in generation $n \geq 1$, and $H_{n,m}$, the corresponding admixture for a random admixed male, have expectations

$$E[H_{n,f}] = \frac{J_{n+1}s_f + J_n s_m}{2^n} \tag{1}$$

$$E[H_{n,m}] = \frac{J_n s_f + J_{n-1}s_m}{2^{n-1}}. \tag{2}$$

A quick check of the *On-Line Encyclopedia of Integer Sequences* (Sloane 2015) revealed that our sequence $J_n$, which begins 0, 1, 1, 3, 5, 11, 21, 43, was known as the Jacobsthal numbers. The approximate doubling in the successive terms of the sequence produced an approximate 2-to-1 weighting of the female and male contributions $s_f$ and $s_m$ in Equations 1 and 2—a sensible result given that in a population with equally many males and females, the number of X chromosomes residing in females is twice as large as the corresponding number for males.

At the same time, however, the appearance in our problem of the sequence $J_n$ was peculiar. Its generative formula, $J_n = J_{n-1} + 2J_{n-2}$, was reminiscent of the famous recursion for the Fibonacci numbers, $F_n = F_{n-1} + F_{n-2}$. We knew of the result that the number of genealogical ancestors of an individual according to haplodiploid inheritance—such as on the X chromosome—follows the Fibonacci sequence (*e.g.*, Basin 1963). That a related sequence appeared in another X-chromosomal problem could not have been a coincidence. But where did such number sequences, formed by simple linear recursions, enter population genetics?

## The Fibonacci Numbers

A search for *Fibonacci numbers* and *population genetics* quickly led us to a much-chronicled episode (Wright 1969, p. 171; Provine 1971, p. 136; Crow 1987) involving the computation of the probability of heterozygosity of an autosomal diploid locus after $n$ generations of inbreeding by sib mating, beginning from an initial cross of *AA* and *aa* parents. Such calculations of probabilities of homozygosity and heterozygosity were important in understanding the consequences of animal and plant

breeding schemes for the properties of the lines produced. What level of inbreeding was generated by different breeding systems with their own characteristic sequences of crosses?

The sib-mating problem was difficult, and the first report of a solution, by Pearl (1913), was mistaken. In one of a series of corrections to the erroneous claim, a brief note by Jennings (1914) used a discrete-time recursion to show that the probability of heterozygosity in the $n$th generation, including as generation $n = 1$ the sibs born from the initial cross, is $F_n/2^{n-1}$. He commented, "The formula turns out to be a combination of the successive powers of 2, with the successive terms of the Fibonacci series, which appears in so curious a way in various natural phenomena" (Jennings 1914, p. 693).

We had traced the source of the Fibonacci numbers. They arose, however, in an autosomal calculation. Surely someone had looked at the X.

## The Breeding Systems of Jennings (1916)

The 1914 report by Jennings was short, focused on correcting the earlier error. Jennings's (1916) *GENETICS* article was next. In this longer study, a 37-page treatise dense with calculations, Jennings considered not only the case of sib mating and autosomal loci from the exchange with Pearl, but also several other breeding systems: for example, random mating, self-fertilization, parent-by-offspring mating with the same parent bred over multiple generations of offspring, and parent-by-offspring mating with parents bred with their own offspring and no further. The various cases had characteristic discrete-time recursions describing the proportions of homozygotes or heterozygotes in consecutive generations.

Jennings's article was impressive. Its problems were challenging, and the number of cases far exceeded what a modern author would put into a single manuscript. At the same time, it was hard to follow. Figures were limited to a few schematic diagrams, and the article was almost all math, with little context. One wants to know more about the interest at the time of such schemes as parent-by-offspring mating in which each individual is used for two, and exactly two, successive generations.

Perhaps the X chromosome appeared in this article? The article had a table of contents with *Sex-Linked Factors* as one of its sections. At first, however, the section was unenlightening. But in obtaining the recursion for each breeding system, Jennings computed the first several terms, conveniently collecting the numerical values into a comprehensive table. And there in the table, right under the Fibonacci sequence, in a row annotated simply by "G," was the Jacobsthal sequence— the very numbers we had just derived!

## The Connection Between Admixture and Breeding Systems

We returned to the section on *Sex-Linked Factors*, finding that it contained the mysterious G. Knowing now that G (Figure 1)

TABLE I. *Fundamental series in Mendelian breeding.*

| n | x + 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | Formula |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | 2x | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 | 2048 | 4096 | 8192 | 16384 | 32768 | 65536 | 131072 | 262144 | 524288 | 1048576 | $2^n$ |
| C | 2x + 1 | 0 | 1 | 3 | 7 | 15 | 31 | 63 | 127 | 255 | 511 | 1023 | 2047 | 4095 | 8191 | 16383 | 32767 | 65535 | 131071 | 262143 | 524287 | 1048575 | $2^n-1$ |
| D | 2x − 1 |  | 2 | 3 | 5 | 9 | 17 | 33 | 65 | 129 | 257 | 513 | 1025 | 2049 | 4097 | 8193 | 16385 | 32769 | 65537 | 131073 | 262145 | 524289 | $2^{n-1}+1$ |
| E |  |  |  | 2 | 5 | 11 | 23 | 47 | 95 | 191 | 383 | 767 | 1535 | 3071 | 6143 | 12287 | 24575 | 49151 | 98303 | 196607 | 393215 | 786431 | $2^{n-1}+2^{n-2}-1$ |
| F | Sum of two preceding | 0 | 1 | 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | 55 | 89 | 144 | 233 | 377 | 610 | 987 | 1597 | 2584 | 4181 | 6765 |  |
| G | 2x + 1, then 2x − 1 | 0 | 1 | 1 | 3 | 5 | 11 | 21 | 43 | 85 | 171 | 341 | 683 | 1365 | 2731 | 5461 | 10923 | 21845 | 43691 | 87381 | 174763 | 349525 | $(=B_{n-1}-G_{n-1})$ |
| H | G − F | 0 | 0 | 0 | 1 | 2 | 6 | 13 | 30 | 64 | 137 | 286 | 594 | 1221 | 2498 | 5084 | 10313 | 20858 | 42094 | 84797 | 170582 | 342760 |  |
| I | B − G − F | 1 | 0 | 2 | 3 | 8 | 16 | 35 | 72 | 150 | 307 | 628 | 1276 | 2587 | 5228 | 10546 | 21235 | 42704 | 85784 | 172179 | 345344 | 692286 | $(=G_{n+1}-F_n)$ |
| J | $B_n - F_{n+1}$ | 0 | 1 | 2 | 5 | 11 | 24 | 51 | 107 | 222 | 457 | 935 | 1904 | 3863 | 7815 | 15774 | 31781 | 63939 | 128488 | 257963 | 517523 | 1037630 |  |
| K | $B_n - F_{n+2}$ | 0 | 0 | 1 | 3 | 8 | 19 | 43 | 94 | 201 | 423 | 880 | 1815 | 3719 | 7582 | 15397 | 31171 | 62952 | 126891 | 255379 | 513342 | 1030865 | $(=J_n-F_n)$ |
| L | $B_n - F_{n-1} - G_{n-1}$ |  | 2 | 2 | 6 | 11 | 24 | 48 | 99 | 200 | 406 | 819 | 1652 | 3324 | 6683 | 13420 | 26930 | 54003 | 108240 | 216856 | 434323 | 869632 |  |
| M | 3B_n − F_{n+2} | 2 | 4 | 9 | 19 | 40 | 83 | 171 | 350 | 713 | 1447 | 2928 | 5911 | 11911 | 23966 | 48165 | 96707 | 194024 | 389035 | 779667 | 1561918 | 3128017 | $(=B_{n+1}+K_n)$ |

GENETICS 1: Ja 1916

**Figure 1** The table of number sequences from Jennings (1916). Row F contains the Fibonacci numbers, and row G, the Jacobsthal numbers.

was the Jacobsthal sequence unlocked the section. Jennings (1916) had examined a case involving the X chromosome, crossing a female $AA$ to a male $aY$ with sole X chromosome $a$. Random mating occurs among the offspring, continuing for $n$ generations. What are the frequencies for allele $A$ in females and males in the $n$th generation of the newly formed line?

We simplified Jennings's equations, shifting the index in his paragraph 58 so that, like in our admixture model—where generation 1 is the first generation of the existence of the admixed population—the $n$th generation includes the result of the initial cross followed by $n-1$ generations of random mating. Using our notation, Jennings obtained for the frequency $p_{n,\mathrm{f}}$ of allele $A$ in females and the corresponding frequency $p_{n,\mathrm{m}}$ in males

$$p_{n,\mathrm{f}} = \frac{J_{n+1}}{2^n} \tag{3}$$

$$p_{n,\mathrm{m}} = \frac{J_n}{2^{n-1}}. \tag{4}$$

Jennings's allele frequencies accord exactly with the expected admixture proportions from Equations 1 and 2 in the special case that all females arise from source 1 ($s_\mathrm{f} = 1$) and all males arise from source 2 ($s_\mathrm{m} = 0$). In other words, at generation $n$, the frequency of X-chromosomal allele $A$ in females and males in a breeding system where a female $AA$ is crossed to a male $aY$, and random mating continues for $n-1$ generations, exactly equals the expected X-chromosomal admixture level in females and males for source 1 in an admixture model in which females enter from source 1 and males from source 2 (Table 1).

Mathematically, the computations answer the same question: whereas we compute the admixture from a source population for a random female or male at an X-chromosomal locus, Jennings computes for an X-chromosomal locus the "admixture" from the "source population" consisting of the single $AA$ female. With this view of Jennings's scheme, the frequency of allele $A$ records the admixture proportion from the source population that provided the females. In a sense, breeding-system computations of levels of homozygosity and heterozygosity that were at the forefront of the field a century ago are a version of modern admixture computations popular now.

## Perspective

The appearance in two problems of the same number sequence $J_n$ was the linchpin in identifying a connection between similar families of models separated by 99 years. The question remains, however: What other links between models of admixture and models of breeding systems can be unearthed deep in the *GENETICS* record, and what insights can they bring to the modern study of admixture? The analogy of admixture as a form of cross between populations is familiar. But in some cases, the parallel is more than an analogy, as admixture and breeding-system models can make use of exactly the same mathematics.

Beyond its relationship to admixture models, the article of Jennings (1916) is noteworthy for a number of other reasons, including an early assortative mating model and an early single-locus analysis of selection against deleterious recessives. It is perhaps most often remembered for its identification of an oscillation in the approach to Hardy–Weinberg equilibrium on the X chromosome. Because each male X chromosome was in a female one generation in the past, the allele frequencies in males "chase" the frequencies in females, lagging by one generation. Unlike for autosomes, equilibrium is approached over time, rather than achieved immediately. A correspondence with this pattern can also be seen in Goldberg and Rosenberg (2015): the expected admixture for a specific source population follows the same oscillation (Figure 2). Indeed, Equations 3 and 4 give exactly the female and male frequencies of allele $A$ in generation $n \geq 1$ during the approach to Hardy–Weinberg equilibrium for a population that begins in generation 0 exclusively with $AA$ females and $aY$ males (Crow and Kimura 1970, p. 46). In parallel to the Hardy–Weinberg setting, unlike for autosomal loci (Goldberg et al. 2014), the expected admixture on the X chromosome for a random female or male under hybrid isolation is not constant over time.

Some earlier textbooks, including Crow and Kimura (1970) and Li (1976), discuss the X-chromosomal Hardy–Weinberg

**Table 1 Parallels between the X-chromosomal hybrid-isolation admixture model of Goldberg and Rosenberg (2015) and a breeding-system model of Jennings (1916)**

| Model feature | Admixture model | Breeding-system model |
|---|---|---|
| The initial "cross" | An admixed population is formed by females from source population 1 and males from source population 2 | A cross occurs between a female $AA$ and a male $aY$ |
| Type of mating after the initial cross | Random mating occurs in the admixed population | Random mating occurs among the progeny |
| No. generations of mating after the initial cross | $n - 1$ | $n - 1$ |
| Description of the population of interest | Admixed population under hybrid isolation | Inbred line |
| Quantities computed for the population | Expected X-chromosomal admixture proportion from source population 1 in the admixed population in generation $n$, in females ($p_{n,f}$) and males ($p_{n,m}$) | Frequency of X-chromosomal allele $A$ in the inbred line in generation $n$, in females ($p_{n,f}$) and males ($p_{n,m}$) |
| Mathematical results[a] | $p_{n,f} = J_{n+1}/2^n$ | $p_{n,f} = J_{n+1}/2^n$ |
| | $p_{n,m} = J_n/2^{n-1}$ | $p_{n,m} = J_n/2^{n-1}$ |
| Source | Goldberg and Rosenberg (2015, equations 12 and 13) | Jennings (1916, paragraph 58) |

[a] $J_n$ represents entry $n$ in the Jacobsthal number sequence.

oscillation. In later books, however, the oscillating pattern and its connection to recursive sequences are sometimes lost, as, in some cases, is the reference to Jennings (1916).
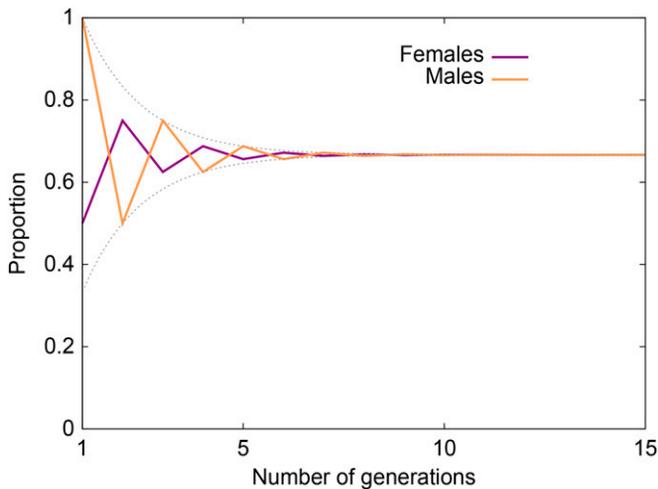


**Figure 2** Admixture proportions and allele frequencies for females and males in an X-chromosomal model. The values plotted can be interpreted equivalently as (1) the X-chromosomal proportions of admixture from source population 1 in a random female and a random male in the $n$th generation, in an admixture model in which the mothers of the offspring individuals present in generation 1 all enter from source population 1 and the fathers all enter from source population 2 (Goldberg and Rosenberg 2015); (2) in generation $n$, the frequencies in females and males of an X-chromosomal allele $A$ after a cross of an $AA$ female and an $aY$ male, followed by $n - 1$ generations of random mating (Jennings 1916); and (3) in generation $n$, the frequencies in females and males of an X-chromosomal allele $A$ after the $n$th generation of random mating, starting in generation 0 from a population of $AA$ females and $aY$ males (Jennings 1916). The value for females is given by Equation 3 and the value for males by Equation 4. Noting that the closed-form expression for the Jacobsthal numbers is $J_n = [2^n - (-1)^n]/3$, it follows that at generation $n$, both quantities plotted are constrained by a lower bound $(2^n - 1)/(3 \times 2^{n-1})$ and an upper bound $(2^n + 1)/(3 \times 2^{n-1})$.

In a perspective on H. S. Jennings, Crow (1987) commented that the computational methods of Jennings (1916) in systems of breeding were soon superseded by Sewall Wright's general approach to the inbreeding coefficient produced by an arbitrary pedigree. Nevertheless, this one article uncovered the X-chromosomal behavior of the Hardy–Weinberg model, anticipated questions of interest in modern admixture investigations, and moreover, contributed to expanding the long list of biological examples of the Fibonacci numbers.

A quest in the *GENETICS* archive is full of surprises. *GENETICS* reports the date in 1915 on which Bridges's manuscript was received as October 21 and Jennings's as August 26. It is tempting to speculate how the history of *GENETICS* might have unfolded differently had the editors not inverted the timing of arrival in arranging the publication order for the journal's first issue.

## Acknowledgments

## Literature cited

Basin, S., 1963 The Fibonacci sequence as it appears in nature. Fibonacci Quarterly 1: 53–56.

Bridges, C. B., 1916 Non-disjunction as proof of the chromosome theory of heredity. Genetics 1: 1–52.

Crow, J. F., 1987 Seventy years ago in *Genetics*: H. S. Jennings and inbreeding theory. Genetics 115: 389–391.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Burgess, Minneapolis.

Duchen, P., D. Živković, S. Hutter, W. Stephan, and S. Laurent, 2013 Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics 193: 291–301.

Ganetzky, B. S., and R. S. Hawley, 2016 The centenary of *GENETICS*: Bridges to the future. Genetics 202. DOI: 10.1534/genetics.115.180182.

Goldberg, A., and N. A. Rosenberg, 2015 Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. Genetics 201: 263–279.

Goldberg, A., P. Verdu, and N. A. Rosenberg, 2014 Autosomal admixture levels are informative about sex bias in admixed populations. Genetics 198: 1209–1229.

Gravel, S., 2012 Population genetics models of local ancestry. Genetics 191: 607–619.

Jennings, H. S., 1914 Formulae for the results of inbreeding. Am. Nat. 48: 693–696.

Jennings, H. S., 1916 The numerical results of diverse systems of breeding. Genetics 1: 53–89.

Johnston, M., 2014 On the GSA journals. *GSA Reporter*, winter 2014, pp. 10–11.

Li, C. C., 1976 *First Course in Population Genetics*. Boxwood Press, Pacific Grove, CA.

Liang, M., and R. Nielsen, 2014 The lengths of admixture tracts. Genetics 197: 953–967.

Loh, P.-R., M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell *et al.*, 2013 Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193: 1233–1254.

Lohmueller, K. E., C. D. Bustamante, and A. G. Clark, 2011 Detecting directional selection in the presence of recent admixture in African-Americans. Genetics 187: 823–835.

Lohse, K., and L. A. F. Frantz, 2014 Neanderthal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. Genetics 196: 1241–1251.

Long, J. C., 1991 The genetic structure of admixed populations. Genetics 127: 417–428.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland *et al.*, 2012 Ancient admixture in human history. Genetics 192: 1065–1093.

Pearl, R., 1913 A contribution towards an analysis of the problem of inbreeding. Am. Nat. 47: 577–614.

Provine, W. B., 1971 *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago.

Sanderson, J., H. Sudoyo, T. M. Karafet, M. F. Hammer, and M. P. Cox, 2015 Reconstructing past admixture processes from local genomic ancestry using wavelet transformation. Genetics 200: 469–481.

Sedghifar, A., Y. Brandvain, P. Ralph, and G. Coop, 2015 The spatial mixing of genomes in secondary contact zones. Genetics 201: 243–261.

Sloane, N. J. A., 2015 *The On-Line Encyclopedia of Integer Sequences*. Available at: https://oeis.org.

Verdu, P., and N. A. Rosenberg, 2011 A general mechanistic model for admixture histories of hybrid populations. Genetics 189: 1413–1426.

Wright, S., 1969 *Evolution and the Genetics of Populations Vol. 2: The Theory of Gene Frequencies* University of Chicago Press, Chicago.

*Communicating editor: M. Johnston*