Advanced Series in Agricultural Sciences 16

Advanced Series in Agricultural Sciences

Volumes already published in the series

Volume 1: *A. P. A. Vink*
Land Use in Advancing Agriculture

Volume 2: *H. Wheeler*
Plant Pathogenesis

Volume 3: *R. A. Robinson*
Plant Pathosystems

Volume 4: *H. C. Coppel, J. W. Mertins*
Biological Insect Pest Suppression

Volume 5: *J. J. Hanan, W. D. Holley, K. L. Goldsberry*
Greenhouse Management

Volume 6: *J. E. Vanderplank*
Genetic and Molecular Basis of Plant Pathogenesis

Volume 7: *J. K. Matsushima*
Feeding Beef Cattle

Volume 8: *R. J. Hanks, G. L. Ashcroft*
Applied Soil Physics

Volume 9: *J. Palti*
Cultural Practices and Infectious Crop Diseases

Volume 10: *E. Bresler, B. L. McNeal, D. L. Carter*
Saline and Sodic Soils

Volume 11: *J. R. Parks*
A Theory of Feeding and Growth of Animals

Volume 12: *J. Hagin, B. Tucker*
Fertilization of Dryland and Irrigated Soils

Volume 13: *A. J. Koolen, H. Kuipers*
Agricultural Soil Mechanics

Volume 14: *G. Stanhill*
Energy and Agriculture

Volume 15: *E. A. Curl, B. Truelove*
The Rhizosphere

Volume 16: *D. P. Doolittle*
Population Genetics: Basic Principles

Donald P. Doolittle

Population Genetics: Basic Principles

With 20 Figures

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo

Professor Dr. DONALD P. DOOLITTLE
Purdue University
Department of Animal Sciences
Lilly Hall
West Lafayette, IN 47907
USA

ISBN-13:978-3-540-17326-7 e-ISBN-13:978-3-642-71734-5
DOI: 10.1007/978-3-642-71734-5

Library of Congress Cataloging in Publication Data. Doolittle, Donald P. Population genetics. (Advanced series in agricultural sciences; 16) Includes index. 1. Population genetics. I. Title. II. Series. QH455.D66 1986 575.1'5 86-33877

2131/3130-543210

To Maria

Preface

I have for a number of years taught a course in population genetics for students interested in plant and animal breeding. The objective of the course has been to lay a foundation in population genetics for the concepts of quantitative genetics which are introduced in the last third of the course. I have not been able to find an appropriate text for this purpose.

For a quarter of a century, Falconer's *Introduction to Quantitative Genetics* has been the standard, and excellent, text in that subject. For my purposes, however, this text is not sufficiently detailed in the population genetics basis for quantitative theory.

A number of good texts in population genetics are available, of which Li's *First Course in Population Genetics* is didactically the best. But these texts are directed toward the genetics of natural populations, rather than domestic populations, breeding under human control. They also tend to treat quantitative genetics gingerly, if at all.

I have therefore developed the present text from my teaching notes.

The chapters of this book are labeled "Lectures". Each is intended to correspond approximately to the amount of material which can be covered in a 50-minute lecture. Divisions are, of course, dictated by the natural divisions of the subject matter, and the lectures are therefore not of uniform length. Nevertheless, in so far as possible, an attempt has been made to make the average length a lecture's worth.

Prerequisites for the course from which the book sprang are one course in genetics and one in statistical methods. Review lectures on basic genetic and statistical concepts have been included in the book, for review purposes.

Mathematical formulae engender in many people a near catatonic state of fright, making them impervious to the spoken or written word. There is no way to teach population and quantitative genetics without introducing algebraic expressions. As Earl Green points out in the first chapter of his *Genetics and Probability in Animal Breeding Experiments,* mathematical formulae are merely statements about mathematical relations among entities, written in a shorthand form. To emphasize this, formulae are incorporated into the flow of the discussion in the present text, though usually written on a separate line to enhance clarity. An attempt has been made to introduce formulae as gently as possible, and to explain fully what each means. For a biologist, after all, the importance of a formula is not its algebra, but the biological meaning of that algebra.

At the end of most lectures, one or two exercises have been introduced. Complete and detailed solutions for these exercises are given in an Appendix. The exercises are intended primarily to provide examples of the types of data to which the subject matter of the lecture applies, and to give the student the opportunity to manipulate such data. In a few cases, supplementary points about the theory are introduced via the exercises. The student should be encouraged to work out his own answers to the exercises before turning to the printed solution; this will enable him to gauge his progress in understanding the material.

Countless people have contributed to my understanding of population genetics and to the writing of this book. However, I owe a special debt of gratitude to:

Tom Doolittle and Jan Blackburn for the art work; the staffs of the Joan Staats Library at the Jackson Laboratory, Bar Harbor, Maine, and the Albert Mann Library at Cornell University, Ithaca, New York, for their invaluable assistance in finding references; Drs. Duwayne Englert of Southern Illinois University, Carbondale, and Wyman Nyquist, Purdue University, West Lafayette, Indiana, for reading and commenting upon the text; Dr. James F. Crow of the University of Wisconsin, Lansing, for his very detailed commentary; Dr. C. C. Li of the University of Pittsburgh, Pittsburgh, Pennsylvania, for his comments and encouragement; Dr. Dale Van Vleck of Cornell University for multiple acts of assistance: he introduced me to the publisher, sponsored my sabbatical leave at Cornell, during which most of the writing took place, and improved the text with his comments and suggestions; and to my wife, Maria, for her patience and her proof-reading skills.

Any virtues which this book may possess arise from their help; its faults are my responsibility.

DONALD P. DOOLITTLE

Contents

Part I The Hardy-Weinberg Law 1

1. The Hardy-Weinberg Law 2
2. Conditions for Hardy-Weinberg Equilibrium 8
3. Applications of the Hardy-Weinberg Law 12
4. Snyder's Ratios 18

Part II Constant Allele Frequencies 23

5. Multiple Alleles 24
6. Separate Sexes 27
7. Sex Linkage 31
8. Other Phenomena Associated with Sex 35
9. Linkage 41
10. Recombination Equilibrium 48
11. Polyploidy 53

Part III Systematic Forces 57

12. Selection 58
13. The General Selection Equation 64
14. Equilibrium Under Selection 69
15. Mutation 74
16. Interaction of Mutation with Selection 79
17. Migration 83

Part IV Dispersive Forces 89

18. The Finite Population Model 90
19. Inbreeding 92
20. The Development of Inbreeding 96
21. Genic Drift 100
22. Genotypic Drift 104
23. Effective Breeding Size 107
24. Interaction of Inbreeding with Systematic Forces 113

25. Coancestry 116
26. Applications of Coancestry 120
27. Tabular Calculation of Coancestry 123
28. Path Calculation of F 128
29. Systematic Inbreeding 131
30. Assortative Mating 137

Part V Quantitative Inheritance 141

31. Metric Traits 142
32. Evidence for the Theory of Metric Inheritance 148
33. Phenotypic Value 151
34. The Single Locus Model 154
35. Breeding Values 160
36. Dominance Deviations 164
37. Multiple Loci 169
38. Causal Components of Variance 172
39. Analytical Components 176
40. Heritability 183
41. Realized Heritability 187
42. Other Estimates of h^2 192
43. Environmental Variance 197
44. Repeatability 199

Appendix I Exercise Solutions 203

Appendix II Review Lectures 235

R1. Basic Genetics 236
R2. Probability 241
R3. Statistical Methods 248

References 256

Subject Index 261

Part I The Hardy-Weinberg Law

Lecture 1 The Hardy-Weinberg Law

The basic law of Mendelian genetics is the law of segregation. It allows us, if we know the genotypes of a pair of parents, to predict the genotypes, in their proportions, that will be found among the offspring of those parents. For example, if both parents are heterozygous *Aa*, we can predict that the three genotypes *AA*, *Aa* and *aa*, will occur among their offspring in an approximate 1: 2:1 ratio.

In a similar manner, the Hardy-Weinberg law is the basic law of population genetics. It allows us to predict the genotypic frequencies in the offspring generation of a population from those of the parent generation. The population is assumed to reproduce by random pairing of genotypes in the parents; Mendelian segregation then controls the frequencies of offspring genotypes produced by any given pair of parental genotypes.

The Hardy-Weinberg law derives its name from the two investigators who independently published it in 1908: George Hardy, a British mathematician, and Wilhelm Weinberg, a German physician. The central principle of the law had actually been stated by the American geneticist, W. E. Castle, as early as 1903 (Li 1967a), but Castle failed to generalize it.

Weinberg's 1908 paper presenting the law was only one of a number of valuable contributions he made to genetics, based on his observations as a practicing obstetrician (Stern 1962). Weinberg's other contributions included a method to estimate the frequency of monozygotic twins in man, and the method of "casting out probands" to correct for errors of ascertainment, as well as further papers on the subject of the Hardy-Weinberg law. But his original contribution to this law was little known, at least to English speaking geneticists, until Stern (1943) published a translation of relevant portions of the 1908 paper.

Hardy's effort was occasioned by his acquaintance with R. C. Punnett, a pioneer British Mendelian geneticist. In those early years of Mendelian genetics, opponents claimed that the law of segregation would require that 1:2:1 genotypic ratios, and 3:1 phenotypic ratios, be characteristic of most traits in any population. Because few traits showed anything like such ratios, the generality of Mendelian inheritance was called into doubt. Punnett challenged Hardy to show that Mendel's laws would not necessarily result in such ratios; and Hardy proceeded to do so.

Let us derive the law, using a somewhat more general argument than Hardy used. Imagine a population of diploid individuals in which a pair of alleles, *A* and *a*, are segregating. Three genotypes, *AA*, *Aa*, and *aa*, are then possible. We assign to these genotypes arbitrary relative frequencies D_t, $2H_t$, and R_t, respectively, in generation t of the population. The only restrictions on these frequencies are that they cannot be less than 0 or greater than 1, and that, since only the three geno-

Table 1.1. Derivation of the Hardy-Weinberg law

Matings	Frequency	Offspring genotypes		
		AA	*Aa*	*aa*
$AA \times AA$	D_t^2	D_t^2		
$AA \times Aa$	$2D_tH_t$	D_tH_t	D_tH_t	
$AA \times aa$	D_tR_t		D_tR_t	
$Aa \times AA$	$2D_tH_t$	D_tH_t	D_tH_t	
$Aa \times Aa$	$4H_t^2$	H_t^2	$2H_t^2$	H_t^2
$Aa \times aa$	$2H_tR_t$		H_tR_t	H_tR_t
$aa \times AA$	D_tR_t		D_tR_t	
$aa \times Aa$	$2H_tR_t$		H_tR_t	H_tR_t
$aa \times aa$	R_t^2			R_t^2
	1	$(D_t+H_t)^2$	$2(D_t+H_t)(H_t+R_t)$	$(H_t+R_t)^2$

types can occur,

$$D_t+2H_t+R_t=1.$$

Now we allow the individuals of generation t to mate at random among themselves. The frequencies of parental genotypic pairs, and of the consequent offspring genotypes produced by each, are summarized in Table 1.1.

The frequency of each parental genotypic combination is simply the product of the frequencies of the genotypes of the two parents. For example, the frequency of *Aa* x *Aa* is $(2H_t)(2H_t)=4H_t^2$. Mendelian segregation ratios, applied to these frequencies, yield the frequencies of offspring genotypes produced. For example, we expect a 1: 2:1 segregation of *AA*: *Aa*: *aa*, from the *Aa* x *Aa* mating, therefore, such matings should contribute H_t^2 *AA*, $2H_t^2$ *Aa* and H_t^2 *aa* to the offspring generation.

The column sums in Table 1.1 represent the overall frequencies of each genotype in generation t+1. A modicum of simple algebra is necessary to simplify these sums. Let us first deal with the sum of the mating frequencies column. This is

$$\begin{aligned}\Sigma\,(\text{freq.}) &= D_t^2+2D_tH_t+D_tR_t+2D_tH_t+4H_t^2+2H_tR_t+D_tR_t+2H_tR_t+R_t^2\\ &= D_t^2+4D_tH_t+2D_tR_t+4H_t^2+4H_tR_t+R_t^2,\end{aligned}$$

which takes the form

$$a^2+2ab+2ac+b^2+2bc+c^2=(a+b+c)^2,$$

with $a=D_t$, $b=2H_t$ and $c=R_t$. Hence,

$$\Sigma\,(\text{freq.})=(D_t+2H_t+R_t)^2=1, \tag{1.1}$$

because $D_t+2H_t+R_t=1$. The total frequency of matings must be 1, regardless of individual values of the genotypic frequencies. Thus, this sum reassures us that we have included all the possible parental genotypic combinations at the correct frequencies.

The frequencies of the three genotypes in generation t+ are the totals of the *AA*, *Aa* and *aa* columns in Table 1.1.

$$\begin{aligned} D_{t+1} &= D_t^2 + 2D_tH_t + H_t^2 = (D_t + H_t)^2; \\ 2H_{t+1} &= 2D_tH_t + 2D_tR_t + 2H_t^2 + 2H_tR_t \\ &= 2D_t(H_t + R_t) + 2H_t(H_t + R_t) \\ &= 2(D_t + H_t)(H_t + R_t); \text{ and} \\ R_{t+1} &= H_t^2 + 2H_tR_t + R_t^2 = (H_t + R_t)^2. \end{aligned} \tag{1.2}$$

Now let us consider the frequencies of the alleles *A* and *a* in generation t; we will use p_t and q_t to represent these frequencies, respectively. Because all the genes in *AA* homozygotes, half of those in *Aa* heterozygotes, and none of those in *aa* homozygotes are *A* genes,

$$p_t = 1(D_t) + 0.5(2H_t) + 0(R_t) = D_t + H_t; \tag{1.3a}$$

by similar reasoning,

$$q_t = 0(D_t) + 0.5(2H_t) + 1(R_t) = H_t + R_t. \tag{1.3b}$$

Note that

$$p_t + q_t = D_t + H_t + H_t + R_t = D_t + 2H_t + R_t = 1.$$

Again, p+q must equal 1 if *A* and *a* are the only alleles in the population, providing a check on our calculations.

We now see that the offspring genotypic frequencies are simple functions of the frequencies of the two alleles in the parent generation. The frequencies of offspring genotypes are, from Eqs. (1.2) and (1.3):

$$\begin{aligned} D_{t+1} &= (D_t + H_t)^2 = p_t^2; \\ 2H_{t+1} &= 2(D_t + H_t)(H_t + R_t) = 2p_tq_t; \\ R_{t+1} &= (H_t + R_t)^2 = q_t^2. \end{aligned} \tag{1.4}$$

This implies, of course, that parental genotypic frequencies are irrelevant in determining genotypic frequencies among the offspring. The latter are determined by parental allelic frequencies; any two populations which have different allelic frequencies among the parents will produce different offspring genotypic distributions.

Here, then, was Hardy's answer to the critics of Mendelian inheritance. Mendel's law of segregation does not require all populations to come to a 1: 2:1 genotypic ratio. A distribution is attained which will differ from population to population, and from trait to trait within the same population, depending on allelic frequencies. A footnote to Hardy's paper states that even Udney Yule, one of the most vehement critics of Mendelian genetics, admitted the cogency of this argument.

It also follows, of course, that if two populations have the same allelic frequencies, they will produce offspring in the same genotypic ratio, even if the genotypic distributions of the parents differ. Consider four populations whose genotypic ra-

tios D_t: $2H_t$: R_t are

0.60: 0 : 0.40

0.20: 0.80: 0

0.40: 0.40: 0.20

0.36: 0.48: 0.16.

In all four populations, despite wide disparities in genotypic frequencies in generation t, allele frequencies are $p_t = 0.6$, $q_t = 0.4$. Therefore, all four populations will yield offspring in the Hardy-Weinberg ratio

$$p_t^2 : 2p_tq_t : q_t^2 = 0.36 : 0.48 : 0.16.$$

In any one of these populations,

$$D_{t+1} = 0.36 = p_t^2,\ 2H_{t+1} = 0.48 = 2p_tq_t,\ R_{t+1} = 0.16 = q_t^2;$$

The same ratio will recur in generation $t+2$.

From Eqs. (1.3),

$$p_{t+1} = p_t^2 + p_tq_t = p_t(p_t + q_t) = p_t;$$
$$q_{t+1} = p_tq_t + q_t^2 = (p_t + q_t)q_t = q_t.$$

In Table 1.2, we have substituted p_t^2 for D_t, $2p_tq_t$ for $2H_t$, and q_t^2 for R_t in Table 1.1. Thus, Table 1.2 represents the parental genotypic frequencies in generation $t+1$, and the offspring frequencies in $t+2$, expressed as functions of generation t allele frequencies. From this table, we see that the offspring ratio is, indeed, the Hardy-Weinberg ratio, even when the parents are already in that ratio. For example,

$$D_{t+2} = p_t^4 + 2p_t^3q_t + p_t^2q_t^2 = p_t^2(p_t^2 + 2p_tq_t + q_t^2) = p_t^2,$$

Table 1.2. Hardy-Weinberg equilibrium

Matings	Frequency	Offspring genotypes		
		AA	*Aa*	*aa*
$AA \times AA$	p_t^4	p_t^4		
$AA \times Aa$	$2p_t^3q_t$	$p_t^3q_t$	$p_t^3q_t$	
$AA \times aa$	$p_t^2q_t^2$		$p_t^2q_t^2$	
$Aa \times AA$	$2p_t^3q_t$	$p_t^3q_t$	$p_t^3q_t$	
$Aa \times Aa$	$4p_t^2q_t^2$	$p_t^2q_t^2$	$2p_t^2q_t^2$	$p_t^2q_t^2$
$Aa \times aa$	$2p_tq_t^3$		$p_tq_t^3$	$p_tq_t^3$
$aa \times AA$	$p_t^2q_t^2$		$p_t^2q_t^2$	
$aa \times Aa$	$2p_tq_t^3$		$p_tq_t^3$	$p_tq_t^3$
$aa \times aa$	q_t^4			q_t^4
	1	p_t^2	$2p_tq_t$	q_t^2

and so forth. A population in the Hardy-Weinberg genotypic ratio in one generation will produce offspring in the same ratio in the next.

Any population will reach the Hardy-Weinberg ratio in generation $t+1$, and remain in that ratio in generations $t+2$, $t+3$, etc. All of the populations above will produce the ratio 0.36: 0.48: 0.16 from generation $t+1$ on, as long as they continue to reproduce under Hardy-Weinberg conditions. This has been termed Hardy-Weinberg equilibrium (Wentworth and Remick, 1916).

Allele frequencies among the zygotes of generation $t+1$ will be the same as among their generation t parents:

$$p_{t+1(z)} = p_t^2 + 0.5(2p_tq_t) = p_t^2 + p_tq_t = p_t(p_t+q_t) = p_t;$$
$$q_{t+1(z)} = 0.5(2p_tq_t) + q_t^2 = q_t(p_t+q_t) = q_t; \qquad (1.5)$$

where $p_{t+1(z)}$ is the *A* allele frequency among the zygotes of generation $t+1$. If allele frequencies are constant throughout the life of generation $t+1$, when that generation mates at random, it will again produce zygotes in the same Hardy-Weinberg ratio, $p_t^2 : 2p_tq_t : q_t^2$. The population will be in Hardy-Weinberg equilibrium.

Suppose, however, that allele frequencies change during generation $t+1$, so that adults of that generation have *A* alleles in the frequency

$$p_{t+1(m)} \neq p_{t+1(z)} = p_t.$$

Random mating among these adults will produce generation $t+2$ zygotes in the ratio

$$p_{t+1(m)}^2 : 2p_{t+1(m)}q_{t+1(m)} : q_{t+1(m)}^2.$$

This is a Hardy-Weinberg ratio, but one based on $p_{t+1(m)}$ rather than on p_t.

Because genotypic frequencies change, there is no equilibrium. Genotypic frequencies can remain the same from generation to generation only if allele frequencies remain the same. Constant allele frequency is an essential condition for Hardy-Weinberg equilibrium.

However, each generation produces zygotes in a Hardy-Weinberg ratio based on the allele frequencies of their parents. This ratio is solely a consequence of random mating. If parents mate at random, offspring zygotes will occur in the Hardy-Weinberg ratio, even if the genotypic ratio may change at later stages of the offspring generation. Random mating is an essential condition for the Hardy-Weinberg ratio; constant allele frequency is not.

In deriving the Hardy-Weinberg law, we tacitly assumed constant allele frequencies, simply by assuming that none of the forces which cause allele frequency to change were operative. This is only one of a set of rather narrowly limited assumptions implicit in the derivation. Hardy himself admitted the limitations, as did Weinberg; and the latter, in his 1909 and 1910 papers, made some attempts to broaden the limits. In the next lecture, we will make these assumptions explicit.

Lecture 1 Exercises

1. The four populations

a) 0.60: 0 : 0.40

b) 0.20: 0.80: 0

c) 0.40: 0.40: 0.20

d) 0.36: 0.48: 0.16

are stated in the text to produce the same genotypic distribution in generation $t+1$. Prepare tables in the form of Table 1.1 to show that this is true.

The ratios D_t: $2H_t$: R_t are given below for five populations. Determine p_t and q_t and the equilibrium genotypic proportions for each population.

a) 0.50: 0 : 0.50

b) 0.50: 0.50: 0

c) 0.09: 0.10: 0.81

d) 0.36: 0.16: 0.48

e) 0.45: 0.45: 0.10

(N.B.: In each case, $p_t+q_t=p_t^2+2p_tq_t+q_t^2=1$. Use these relationships to check your results.)

3. Much of the algebra involved in population genetics consists of manipulations based upon the fact that $p+q=1$ (and that therefore, $p=1-q$, $q=1-p$). Use this fact to verify that

a) $p+2q=1+q=2-p$

b) $p^2-q^2=p-q$

c) $p^2+q^2=1-2pq$

d) $(p-q)^2=1-4pq$

e) $p^2+4pq+4q^2=(1+q)^2$

Lecture 2 Conditions for Hardy-Weinberg Equilibrium

We have derived the Hardy-Weinberg law under a set of assumed conditions, only a few of which were stated explicitly. We must now make all of these assumptions explicit, defining terms clearly.

We introduced, without defining, the concept of a population of organisms. The population geneticist defines a population as a set of freely interbreeding organisms. This definition implies that all members of the population are of the same species; otherwise, free interbreeding would not be possible. For a population in nature, it also implies geographic propinquity of members of the population; individuals living too far apart cannot freely interbreed. "Too far apart", of course, will be conditioned by the nature of the species; it may be a much shorter distance for a sessile plant than for a mobile animal, for example. Populations living under domestication can be limited in breeding range, in the most general meaning of that term. A plant or animal breeder may restrict the breeding range of a given individual to other members of the same breed or variety, even though the individual could breed fertilely with any other member of the same species. At the same time, domesticated plants and animals can, with man's help, transcend normal geographic limitations on breeding range. Individuals, or their gametes, under artificial insemination, can be transported across or even between continents to breed.

We assumed, in deriving the law, that breeding occurred only within the population, that is, that all of the parents of one generation of the population were members of an earlier generation of the same population. In general, we will continue to assume breeding within the population. When it becomes necessary to transgress this limitation, as in discussing crossing or migration, we will clearly indicate that we are doing so.

We also assumed that generations were discrete, that is, that members of generation t mated only with other members of generation t. The model is that of an annual plant, with this year's generation dead and gone before next year's can come to flower. Many populations are not subject to this kind of restriction. An individual from generation t may still be alive and fertile to mate with members of generation $t+1$, $t+2$, etc. Overlapping generations are probably more common than discrete generations in nature. In domestic populations, a superior individual will often be retained to breed with descendant generations.

Nonetheless, the assumption of discrete generations will greatly simplify the discussion of many of the factors whose effects we will be investigating, without any great sacrifice in generality (Crow and Kimura 1970, Hill 1972c, 1979). We will therefore make this assumption in most of our discussions.

A given generation number applies to the zygotes of that generation, the adults which develop from those zygotes, and the gametes those adults produce.

That is to say, generation t zygotes develop into generation t adults, which produce generation t gametes. Generation $t+1$ starts with the zygotes formed by the union of generation t gametes.

We assumed Mendelian segregation in deriving the Hardy-Weinberg law. The population was regarded as a congeries of matings, the proportions of genotypic combinations being determined by the parental genotypic ratio. The genotype of the offspring produced by each mating was then determined by Mendelian segregation from the parental genotypes. We will continue to make this assumption in the discussions which follow. Segregation is modified in the case of sex-linked, partially sex-linked and holandric genes, but the basic mechanism is still Mendelian. Extrachromosomal genes, if such exist, may not segregate according to Mendel's law; some types of aneuploidy may also cause non-Mendelian segregation. Such conditions will not fall within the purview of our discussions.

In deriving the law, we allowed "the individuals of generation t to mate at random," without defining what "to mate at random" meant. The model of random mating which we assumed can be described as follows.

We choose an individual from generation t of the population "at random", which is to say that our choice is dictated purely by chance; every member of generation t has equal probability of being chosen. The probability that we will choose an individual of any given genotype, say *AA*, is therefore the frequency of that genotype in generation t, D_t. We then choose a second individual, again at random, to mate with this first. The first is assumed to be still available for the second choice, so that selfing is possible. What is more important, the choice of the first individual does not change the genotypic frequencies; the probability of choosing *AA* on the second choice is still D_t. The choices are independent, and the probability of choosing any pair of genotypes is the product of the frequencies of the two genotypes. From this pair of mates we produce a single offspring, whose genotypic probabilities are dictated by Mendelian segregation from the parental genotypes. Innumerable repetitions of this procedure produce generation $t+1$ of the population.

Random mating as described above implies randomization in two phases. Mates are chosen and paired at random; the random choice of a gamete from each mate is implicit in Mendelian segregation. We can combine these in a single randomization phase. Suppose we obtain equal numbers of gametes from each member of generation t, and pool them. Assuming equal fertility, each individual will contribute an equal number of gametes to the pool. The pool will then contain *A* and *a* gametes in the ratio p_t: q_t. Gamete frequencies are equal to the frequency of the corresponding allele, because each gamete carries one gene.

If we now draw a random gamete from the pool, the probability of choosing a gamete carrying a particular allele will be the frequency of that gamete. If we unite pairs of randomly drawn gametes, the probability of producing a given genotype will be the product of the frequencies of the gametes in that genotype. Innumerable repetitions of this procedure will give the results summarized in the modified Punnett square of Table 2.1. The genotypic ratio will be the Hardy-Weinberg ratio, p_t^2: $2p_tq_t$: q_t^2, the same ratio we derived from random mating. Random union of gametes is equivalent to random mating of individuals.

Table 2.1. Random union of gametes

First gamete	Second gamete	
	A	a
	p_t	q_t
A p_t	$p_t^2 AA$	$p_t q_t Aa$
a q_t	$p_t q_t aA$	$q_t^2 aa$

As we consider the effects of changing the assumptions under which we derived the Hardy-Weinberg law, the equivalence of random mating and random union of gametes will provide us with a convenient shortcut. In many cases, we will not have to consider the effect of changes on genotypic ratios at all, because, given random mating, there is a 1:1 relationship between the gamete array and the ratio of genotypes.

Under certain circumstances, e.g., self sterility, natural populations will not mate at random. A breeder may also impose nonrandom mating on a domesticated population. We will investigate the consequences of such nonrandom mating.

Random union of gametes implies that any two gametes from any two individuals can successfully unite to form an offspring. We know, however, that in most species, two types (sexes) of gamete occur, and that only the union of gametes of opposite sex will form a viable zygote. In most animal species, the two sexes exist in separate individuals. Even in plants, where both often come from the same individual, male and female gametes are formed. We must, then, explore the effects of separation of the sexes upon equilibrium.

When sexes are separate, there may be an hereditary mechanism for determining sex, such as the sex chromosomes of many higher animal species. Do genes carried on the sex chromosomes reach equilibrium in the same manner as autosomal genes?

We derived the Hardy-Weinberg law for two alleles in a diploid. What are the effects of multiple alleles, of polyploidy? We also considered only one locus; how do two or more loci approach equilibrium simultaneously?

We have seen that it was necessary to assume constant allele frequencies to produce Hardy-Weinberg equilibrium. This meant assuming equal viability, fertility and fecundity of all genotypes, and ruling out mutation and migration. What effect do inequalities of viability and reproductive capability have? Do populations subject to mutation and migration reach equilibrium?

Finally, we implicitly assumed, by equating probabilities with frequencies, that our population was infinitely large. This assumption was also necessary to avoid changes in allele frequency due to sampling error. Because no real population can be infinitely large, we must examine the effects of finite size.

The above assumptions indicate that the Hardy-Weinberg equilibrium can apply to a population only under rather narrowly defined conditions. Before examining the effects of changing these conditions, we must first ask whether any real population can exist to which the Hardy-Weinberg assumptions do apply.

Lecture 2 Exercises

1. The student can simulate random mating, using any of a number of randomization devices (dice, poker chips, tables of random numbers, computer programs for generating random results). The device should be arranged to provide three classes ("genotypes") of results in an arbitrarily predetermined ratio D_t: $2H_t$: R_t. A series of trials are made, the genotype from each being recorded. Pairs of successive genotypes represent matings. Each student should make at least a hundred matings to obtain a fair sample. If all students in the class use the same ratio of genotypes, pooled results from the whole class should give an even better representation of a large population.

2. Random union of gametes can be simulated using two result classes in the ratio p_t: q_t.