BREEDING for QUANTITATIVE TRAITS in PLANTS

Third Edition

Rex Bernardo
University of Minnesota–Twin Cities

Stemma Press
Woodbury, Minnesota

Printed in the United States of America

Stemma Press
1938 Bowsens Lane
Woodbury, MN 55125
http://stemmapress.com

ISBN 978-0-9720724-3-4

Library of Congress Control Number: 2019909100

Cover photo courtesy of Scott Bauer, USDA Agricultural Research Service
Back cover image courtesy of pngtree.com

The author used *Scientific WorkPlace version 5.5* ® to write and typeset this book

Contents

Preface

Breeding for Quantitative Traits in Plants was written as a textbook for a graduate course in the application of quantitative genetics to plant breeding. I hope it will also be useful as a reference for practicing plant breeders. My goal was to write a book that would help a student of plant breeding achieve the following objectives:

1. Understand fundamental concepts in plant breeding and population genetics;
2. Explore how quantitative genetics principles and modern tools can help a plant breeder design and implement a breeding program; and
3. Appreciate the theory, experimental approaches, and evidence that comprise the basis for these concepts and breeding strategies.

The first edition of this book, published in 2002, grew out of lecture notes for a class I twice taught at Purdue University. I wrote it from the perspective of one who has worked as a scientist in a seed company (1988 to 1997) and as a professor at a public university (since 1997). The first edition presented molecular marker applications mainly in the last two chapters. In the second edition (2010), molecular marker applications were integrated throughout the text. This third edition includes a major update on predicting performance via genomewide markers (Chapter 11). Outdated topics such as microarrays, generation means analysis, and mating designs with random parents have been deleted.

The coverage of the subject assumes that the reader has taken a graduate-credit course in plant breeding and a graduate-credit course in statistics.

Readers who need to review basic plant breeding are referred to my prequel textbook, **Essentials of Plant Breeding** (2014; stemmapress.com). Knowledge of basic matrix algebra is required; a brief introduction to matrix algebra is included for those without this background. I have attempted to keep the level of mathematics and statistics manageable. My goal was to stress concepts and principles, and derivations of equations are presented if they help reveal the underlying concepts or principles. Derivations that are primarily statistical rather than genetic in flavor are glossed over but in these instances readers are often referred to more specialized references.

The Scriptures tell us that "*Of making many books there is no end, and much study wearies the body*" (Ecclesiastes 12:12). I am indebted to many people who helped me in the sometimes wearisome task of writing and revising this book. Bill McFee, my former department head at Purdue University, was the first one to suggest (through a casual question during an after-dinner reception) that I write the first edition. Burle Gengenbach and Nancy Ehlke, my department heads in Agronomy and Plant Genetics at the University of Minnesota, provided an environment conducive to writing. Many of the revisions for the second edition were done while I was on a sabbatical leave at Limagrain Europe in France.

The late Wyman Nyquist reviewed the entire book manuscript for the first edition, helped me tighten-up loose passages, and saved me from embarrassing mistakes. Jim Anderson, Bill Beavis, Sofia Brandariz, José Crossa, Yang Da, John Dudley, Marcelo Pacheco, Sushan Ru, Bob Stupar, Dindo Tabanao, Dale Van Vleck, Jianming Yu, and Shengqiang Zhong each offered constructive comments and suggestions on some or on all of the chapters of the first, second, or third editions. My parents, Fernando and Emiliana Bernardo, edited a near-final version of the first edition. (I've wondered how many parents out there are comfortable in editing their son's technical writing.) Students in my University of Minnesota graduate course have provided comments that were helpful in preparing each edition. All of the remaining errors in the book are mine.

I dedicate this book to my wife, Chona, who has always been my biggest fan. She and our six adult F_1 progeny—Alyssa, Jessica, Michael, Matthew, Emily, and David—are a constant source of love and encouragement.

Rex Bernardo
Minneapolis, Minnesota
August 2019

2
Genetics of Breeding Populations

2.1 Genotype and Allele Frequencies in a Population

Quantitative traits are, by necessity, studied within the context of a population. Knowledge of how genes behave in populations is therefore fundamental regardless of whether the genes are known or whether they represent unknown alleles at a quantitative trait locus.

A **population** is a group of interbreeding individuals that exist together in time and space. **Breeding populations** are created by breeders to serve as a source of cultivars that meet specific breeding objectives. A population can be characterized in terms of its **genotype frequencies** and **allele frequencies** at a locus. Suppose a diploid breeding population is segregating at a locus with two alleles, A_1 and A_2 (Table 2.1). The genotype frequencies refer to the proportion of individuals that have a particular genotype. In our example in Table 2.1, the genotype frequencies are $P_{11} = 240/600 = 0.40$ for the A_1A_1 homozygotes, $P_{12} = 240/600 =$

TABLE 2.1. Frequencies of A_1A_1, A_1A_2, and A_2A_2 individuals.

Genotype	Number	Frequency	Frequency after random mating
A_1A_1	240	$P_{11} = 0.40$	$p^2 = 0.36$
A_1A_2	240	$P_{12} = 0.40$	$2pq = 0.48$
A_2A_2	120	$P_{22} = 0.20$	$q^2 = 0.16$
Total	600		

0.40 for the A_1A_2 heterozygotes, and $P_{22} = 120/600 = 0.20$ for the A_2A_2 homozygotes.

A population of 600 diploid plants is equivalent to a population of 1200 alleles. The allele frequencies refer to the proportion of each allele in the population, with no distinction being made between an allele present in a heterozygote or in one of the two homozygotes. The frequency of the dominant allele (i.e., A_1) is denoted by p and is equal to

$$p = P_{11} + \frac{1}{2}P_{12}$$

The frequency of the recessive allele (i.e., A_2) is denoted by q and is equal to

$$q = P_{22} + \frac{1}{2}P_{12}$$

In other words the frequency of an allele is equal to the frequency of individuals that are homozygous for that allele plus one-half the frequency of the heterozygote. In our example, p is equal to $0.40 + 0.40/2 = 0.60$, whereas q is equal to $0.20 + 0.40/2 = 0.40$. The sum of allele frequencies at the locus is $p + q = 1$. The frequency of one allele is easily determined from the frequency of the other allele, e.g., $q = 1 - p$.

Allele frequencies are generally unknown in breeding populations created from non-inbred parents or from three or more inbred parents. But breeding populations in both self-pollinated and cross-pollinated crops are often created by crossing two inbreds (Tables 1.1 and 1.2). As such the allele frequencies at segregating loci are known even if the alleles themselves are unknown. In F_1 and F_2 populations the allele frequencies are $p = q = 0.50$ at all loci that differ between the two parental inbreds. Likewise, the allele frequencies in a **backcross** generation to either parent can be predicted. Suppose one parent has the A_1A_1 genotype whereas the other parent has the A_2A_2 genotype. The first backcross generation (i.e., BC_1) to the A_1A_1 parent will have allele frequencies of $p = 0.75$ and $q = 0.25$. With continued backcrossing to the A_1A_1 parent, the frequencies of the A_1 allele are 0.875 in the BC_2; 0.9375 in the BC_3; 0.96875 in the BC_4; and 0.984375 in the BC_5. The value of q is therefore reduced by 50% with each subsequent backcross, from 0.50 in the F_2 to 0.25 in the BC_1; 0.125 in the BC_2; 0.0625 in the BC_3; 0.03125 in the BC_4; and 0.015625 in the BC_5.

2.2 Hardy-Weinberg Equilibrium

Suppose the population described in Table 2.1 is mated at random. Mating is dictated purely by chance in a **random-mating** population. In other words an individual in a random-mating population is equally likely to mate with any other individual in the population. The gametes produced

by the 240 A_1A_1 plants will all have the A_1 allele. Among the gametes produced by the 240 heterozygotes, 50% will have the A_1 allele whereas 50% will have the A_2 allele. The frequency of gametes with the A_1 allele is therefore $[P_{11} + \frac{1}{2}P_{12}] = [240/600 + \frac{1}{2}(240/600)] = 0.60 = p$. The frequency of gametes with the A_2 allele is $q = 0.40$.

The union of two gametes that have the A_1 allele leads to an individual with the A_1A_1 genotype. With random mating the probability of having an A_1A_1 individual is

$$P_{11(RM)} = p^2$$

where the (RM) subscript indicates random mating. Likewise, the probability of having an A_2A_2 individual is

$$P_{22(RM)} = q^2$$

Finally, an individual with the A_1A_2 genotype results from the union of a gamete that has the A_1 allele and a gamete that has the A_2 allele. This event occurs in either of two ways: the male gamete has the A_1 allele and the female gamete has the A_2 allele, or the male gamete has the A_2 allele and the female gamete has the A_1 allele. With random mating the probability of having an A_1A_2 individual is therefore

$$P_{12(RM)} = 2pq$$

In our example, the genotype frequencies after one generation of random mating are $p^2 = 0.36$ for A_1A_1, $2pq = 0.48$ for A_1A_2, and $q^2 = 0.16$ for A_2A_2 (Table 2.1). If the population size of 600 is kept constant, the expected numbers of plants with each genotype are 216 with A_1A_1, 288 with A_1A_2, and 96 with A_2A_2. The resulting allele frequencies after random mating are

$$\begin{aligned} p_{(RM)} &= P_{11(RM)} + \frac{1}{2}P_{12(RM)} \\ &= p^2 + pq \\ &= p \end{aligned}$$

and

$$\begin{aligned} q_{(RM)} &= P_{22(RM)} + \frac{1}{2}P_{12(RM)} \\ &= q^2 + pq \\ &= q \end{aligned}$$

Random mating therefore changes the genotype frequencies of the population in Table 2.1 but it does not change the allele frequencies.

What happens after a second generation of random mating? The frequency of gametes that have the A_1 allele is again equal to p. Likewise,

the frequency of gametes that have the A_2 allele is equal to q. The resulting genotype frequencies therefore remain $P_{11(RM)} = p^2$ for A_1A_1, $P_{12(RM)} = 2pq$ for A_1A_2 and $P_{22(RM)} = q^2$ for A_2A_2. Subsequent generations of random mating lead to the same allele and genotype frequencies.

G.H. Hardy was a British mathematician and W. Weinberg was a German physician who, in 1908, independently deduced the relationship between allele frequencies and genotype frequencies under random mating. There are three key features of **Hardy-Weinberg equilibrium**. First, the allele frequencies remain constant from generation to generation. Second, the square of the array of allele frequencies is equal to the array of genotype frequencies, i.e., $(p+q)^2 = p^2 + 2pq + q^2$. The observed and equilibrium genotype frequencies of the population in Table 2.1 are unequal. This population is therefore not in Hardy-Weinberg equilibrium but, as we have seen, one generation of random mating will lead to equilibrium genotype frequencies. Third, if allele frequencies change due to external factors, one generation of random mating will lead to a new set of equilibrium genotype frequencies. Suppose the frequency of A_1 changes from 0.60 to 0.80 due to selection. The equilibrium frequencies after one generation of random mating become $p^2 = 0.64$ for A_1A_1, $2pq = 0.32$ for A_1A_2, and $q^2 = 0.04$ for A_2A_2 (Fig. 2.1). In addition to random mating and the absence of selection, other conditions needed for Hardy-Weinberg equilibrium are a large population and the absence of mutation and migration.

Several characteristics of random-mated F_2 and random-mated backcross populations can be gleaned from Fig. 2.1. First, the frequency of heterozygotes in a population in Hardy-Weinberg equilibrium is maximum when p

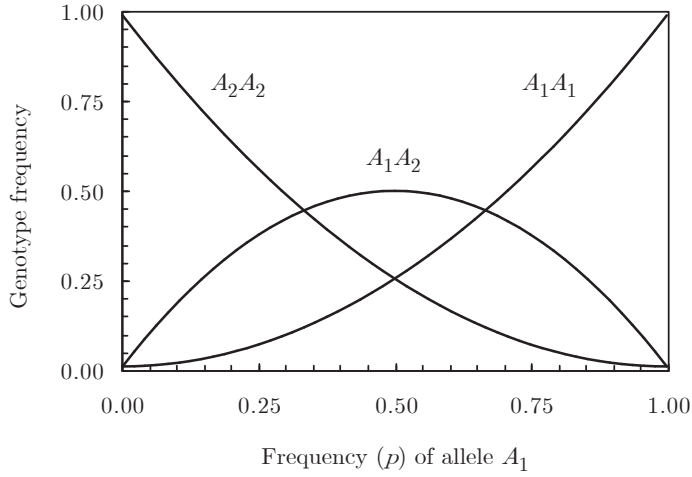

FIGURE 2.1. Genotype frequencies at Hardy-Weinberg equilibrium.

$= q = 0.50$ (Fig. 2.1). This result indicates that the proportion of heterozygotes is maximized in an F_2 population. Second, if p is greater than $\frac{2}{3}$, then the proportion of heterozygotes is intermediate between the proportions of the two homozygotes, i.e., $p^2 > 2pq > q^2$. This result indicates that if a BC_1 or any other backcross generation is random mated, the resulting proportion of A_1A_2 genotypes will always be intermediate to the proportions of the A_1A_1 and A_2A_2 genotypes. Third, if an allele is rare, then that allele will be present mostly in heterozygotes rather than in homozygotes. For example, the expected value of p in the BC_4 generation is 0.96875. The proportion of heterozygotes, which carry the rare A_2 allele (i.e., $2pq = 0.06055$), is 62 times greater than the proportion of A_2A_2 individuals (i.e., $q^2 = 0.00098$).

We gain a better understanding of Hardy-Weinberg equilibrium by considering a locus with more than two alleles. Suppose a locus has three alleles, A_1, A_2, and A_3. The observed frequencies of the six genotypes are in Table 2.2.

Let P_{ij} denote the observed frequency of the A_iA_j genotype. The frequency of the ith allele (denoted by p_i) is equal to the frequency of the homozygotes for the allele, plus half the sum of the frequencies of all heterozygotes that carry the allele, i.e., $p_i = P_{ii} + \frac{1}{2}\sum_{i<j} P_{ij}$. The frequency of A_1 is $p_1 = [P_{11} + \frac{1}{2}(P_{12} + P_{13})] = [0.15 + \frac{1}{2}(0.25 + 0.35)] = 0.45$. The frequencies of the other alleles are $p_2 = 0.15$ for A_2 and $p_3 = 0.40$ for A_3.

At Hardy-Weinberg equilibrium the array of genotype frequencies is equal to the square of array of allele frequencies, i.e., $(p_1 + p_2 + p_3)^2 = p_1^2 + 2p_1p_2 + 2p_1p_3 + p_2^2 + 2p_2p_3 + p_3^2$. The expected frequency of A_1A_1 homozygotes is therefore $P_{11(RM)} = p_1^2 = 0.2025$. The expected frequency of A_1A_2 heterozygotes is $P_{12(RM)} = 2p_1p_2 = 0.135$. The equilibrium genotype frequencies are equal to the square of the allele frequencies regardless of the number of alleles at the locus. With five alleles, for example, the array of genotype frequencies is equal to $(p_1 + p_2 + p_3 + p_4 + p_5)^2$.

Populations in Hardy-Weinberg equilibrium represent idealized populations. Much of the theory and methodology in quantitative genetics has been formulated on the assumption that the reference population is in Hardy-Weinberg equilibrium. An F_2 population from two inbreds is in Hardy-Weinberg equilibrium at a single locus. Breeders, however, routinely use procedures that cause deviations from Hardy-Weinberg equilibrium. These procedures include the lack of random mating, the use of small population sizes, assortative mating, selection, and inbreeding during the development of progeny. Some of these procedures, such as inbreeding and

TABLE 2.2. Genotype frequencies at a locus with three alleles.

Frequency	A_1A_1	A_2A_2	A_3A_3	A_1A_2	A_1A_3	A_2A_3
Observed	0.15	0	0.20	0.25	0.35	0.05
Equilibrium	0.2025	0.0225	0.16	0.135	0.36	0.12