

Breeding for
QUANTITATIVE
TRAITS in PLANTS

Second Edition

Rex Bernardo

University of Minnesota—Twin Cities

Stemma Press
Woodbury, Minnesota

© 2010 by Rex Bernardo

All rights reserved. No part of this book may be reproduced, stored, or transmitted by any means without written permission from the author. The information in this book is deemed highly reliable, but the author and the publisher cannot be held responsible for any consequences of using the information it contains.

Printed in the United States of America

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

ISBN 978-0-9720724-1-0

Library of Congress Control Number: 2010906366

Cover photo courtesy of Scott Bauer, USDA Agricultural Research Service

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

Contents

I	Plant Breeding and Population Genetics	1
1	Introduction	3
1.1	Plant Breeding and Quantitative Traits	3
1.2	Genetic Effects and the One-Locus Model	5
1.3	Modes of Reproduction in Plants	6
1.4	Population Structures and Types of Cultivars	7
1.5	Breeding Self-Pollinated and Cross-Pollinated Crops	10
2	Genetics of Breeding Populations	15
2.1	Genotype and Allele Frequencies in a Population	15
2.2	Hardy-Weinberg Equilibrium	16
2.3	Linkage	20
2.4	Linkage Disequilibrium and Lack of Random Mating	22
2.5	Molecular Markers and Linkage Maps	25
2.6	Small Populations	27
2.7	Selection	31
2.8	Assortative Mating	34
2.9	Inbreeding and Relatedness	35
2.9.1	Concepts of Inbreeding and Relationship	35
2.9.2	Identity by Descent in Pedigrees	39
2.9.3	Estimating Relatedness with Molecular Markers	45

II	Means of Genotypes and Breeding Populations	51
3	Phenotypic and Genotypic Values	53
3.1	Phenotype as a Function of Genes and Environment	53
3.2	Population Mean for a One-Locus Model	56
3.3	Effects of Alleles	58
3.4	Breeding Values and Dominance Deviations	62
3.5	Means and Values in F_2 and BC_1 Populations	65
3.6	Two-Locus Model: Epistatic Effects	65
3.7	Testcross Effect of an Allele	67
3.8	General and Specific Combining Ability	70
4	Selecting Parents to Maximize Mean Performance	73
4.1	Parental Selection in Inbred Development	73
4.2	Mean of Recombinant Inbreds	75
4.3	Fixation of Favorable Alleles	77
4.4	F_2 versus Backcross Populations	79
4.5	Heterotic Groups and Testcross Means	81
5	Mapping Quantitative Trait Loci	85
5.1	Linkage Mapping of QTL	85
5.1.1	General Approach	85
5.1.2	Means of Marker Genotypes	86
5.1.3	Single-Factor Analysis and Interval Mapping	90
5.1.4	Multiple-Marker Analysis	92
5.2	Significance Tests and False Discovery Rate	95
5.3	Selective Genotyping and Phenotyping	99
5.4	Association Mapping	101
5.4.1	Marker-Trait Associations and Population Structure	101
5.4.2	Linkage Disequilibrium	103
5.5	Introgression Libraries and AB-QTL Analysis	106
5.6	Gene Expression Profiling	107
5.7	Candidate Genes and Comparative Mapping	110
III	Variation in Breeding Populations	113
6	Phenotypic and Genetic Variances	115
6.1	Variation due to Genes and Environment	115
6.2	Additive Variance and Dominance Variance	117
6.3	Epistatic Variance	121
6.4	Genetic Variances from a Factorial Model	123
6.5	Covariance between Relatives	126
6.6	Variance among Testcross, Half-Sib, and Full-Sib Families .	129
6.7	Covariance between Single Crosses	131

6.8	Covariance between Selfed Relatives in an F_2 Population . . .	133
6.9	Heritability	135
6.10	Usefulness Criterion	137
6.11	Linkage and Genetic Variances	138
6.12	Genetic Variances in Autotetraploid Species	139
6.13	Maintenance of V_A by Epistasis	141
6.14	Molecular Markers and Trait Variation	143
7	Estimating Genetic Variances	147
7.1	Do Breeders Need to Estimate Genetic Variances?	147
7.2	Mating Designs	149
7.2.1	Approach	149
7.2.2	Assumptions	151
7.3	One-Factor Design	152
7.4	Factorial Design	156
7.5	Diallel	160
7.6	Nested Design	163
7.7	Parent-Offspring Covariance	165
7.8	Precision of Estimates of Genetic Variances	166
7.8.1	Confidence Intervals	166
7.8.2	Resampling Methods	169
7.9	Epistatic Variance	170
7.10	Predicting Genetic Variances	172
7.11	Generation Means Analysis	173
8	Genotype \times Environment Interaction	177
8.1	Genotypic Values in Different Environments	177
8.2	Coping with Genotype \times Environment Interaction	180
8.3	V_ϵ , V_{GE} , and Number of Replications and Environments	182
8.4	Partitioning Environments into Homogeneous Subgroups	186
8.4.1	Cluster Analysis	186
8.4.2	Principal Components Analysis	189
8.5	Stability Analysis	191
8.6	Multiplicative Models	195
8.6.1	AMMI Model	195
8.6.2	Sites Regression Analysis	199
8.7	QTL \times Environment Interaction	200
8.8	Analysis of External Data	202
IV	Selection in Breeding Populations	205
9	Inbred and Testcross Selection	207
9.1	Variance among and within Selfed Families	207
9.2	Variance at Different Selfing Generations	209

9.3	Selection among versus within Families	210
9.4	Selection in F_2 versus BC_1 Populations	213
9.5	Stage of Evaluation	214
9.5.1	Selection during Early Generations	214
9.5.2	Early versus Late Testing in Hybrid Crops	215
9.5.3	Effectiveness of Early Testing	217
9.6	Choosing a Suitable Tester	219
9.7	Selection with Major QTL	222
9.8	F_2 Enrichment	224
9.9	Difficulties in Selection with Many QTL	225
10	Recurrent Selection	231
10.1	Recurrent Selection in Cultivar Development	231
10.2	Response to Selection	233
10.3	Predicted Response to Recurrent Selection	236
10.3.1	Intrapopulation Selection	236
10.3.2	Interpopulation Selection	239
10.3.3	Comparing Recurrent Selection Methods	241
10.4	Increasing the Response to Selection	244
10.5	Theoretical Efficiency of Selection with Markers	245
10.6	Marker-Assisted Recurrent Selection	249
10.6.1	Approach in MARS	249
10.6.2	Gain per Unit Time and Cost	252
10.7	Long-Term Selection	253
10.7.1	Illinois Oil and Protein Selection Experiment	253
10.7.2	Selection Limits	255
11	Best Linear Unbiased Prediction	259
11.1	Usefulness of BLUP	259
11.2	Matrices	262
11.2.1	Types of Matrices	262
11.2.2	Matrix Operations	263
11.2.3	Usefulness of Matrices	267
11.3	BLUP for Self-Pollinated Crops: Inbreds	268
11.3.1	Linear Model	268
11.3.2	Interpretation of Random Effects	269
11.3.3	Mixed-Model Equations	271
11.4	Properties of $\hat{\beta}$ and $\hat{\mathbf{u}}$	275
11.5	BLUP for Cross-Pollinated Crops: Single Crosses	277
11.5.1	Genetic Effects	277
11.5.2	Mixed-Model Equations	280
11.6	BLUP within a Breeding Population	282
11.7	Estimation of Variances	283
11.8	Association Mapping by Mixed-Model Analysis	286
11.8.1	Single Population or Heterotic Pattern	286

11.8.2	<i>QK</i> Model for Multiple Subpopulations	290
11.9	Genomewide Selection by BLUP	293
11.9.1	Marker-Based Selection without QTL Mapping	293
11.9.2	Accuracy of Genomewide Predictions	296
11.9.3	Non-BLUP Methods for Genomewide Selection	298
12	Heterosis and Hybrid Prediction	301
12.1	History and Importance	301
12.2	Genetics of Heterosis	305
12.3	Design III and the Average Level of Dominance	309
12.4	Identifying Heterotic Groups and Patterns	311
12.5	Initial Approaches for Predicting Hybrid Performance	313
12.6	BLUP of Untested Single Crosses	315
12.6.1	Approach	315
12.6.2	Steps in Single-Cross Prediction	316
12.7	Three-Way Crosses, Double Crosses, and Synthetics	320
13	Selection for Multiple Traits	323
13.1	Genetic Correlation between Traits	323
13.2	Correlated Response to Selection	326
13.3	Tandem Selection and Independent Culling Levels	328
13.4	Index Selection	330
13.5	Multiple-Trait Improvement in Applied Breeding	334
14	Epilogue: Designing a Breeding Program	337
	References	341
	Index	380

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 at a seed company (1988 to 1997) and as a professor at a public university (since 1997). The first edition, which I used as a textbook in my graduate course at the University of Minnesota, presented molecular-marker applications mainly in the last two chapters. But rapid developments since the first edition was published have made molecular markers increasingly cheap and abundant, to the extent that molecular markers should no longer be considered as simply an add-on to a breeding program. Molecular-marker applications are therefore integrated throughout the text in this second edition.

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. 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 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 this second edition were done while I was on a sabbatical leave at Limagrain Europe in Riom, France and I thank Pascal Flament and his colleagues for being such gracious hosts.

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, José Crossa, Yang Da, John Dudley, Marcelo Pacheco, Bob Stupar, Dindo Tabanao, Dale Van Vleck, and Jianming Yu each offered constructive comments and suggestions on some or on all of the chapters of the first or second 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.) All the remaining errors in the book, however, are mine.

I dedicate this book to my wife, Chona, who is my biggest fan. She and our six children—Alyssa, Jessica, Michael, Matthew, Emily, and David—are a constant source of love and encouragement.

Rex Bernardo
Saint Paul, Minnesota
June 2010

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 generic 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 the example in Table 2.1, the genotype frequencies are $P_{11} = 240/600 = 0.40$ for the A_1A_1 homozygote, $P_{12} = 240/600 =$

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 heterozygote, and $P_{22} = 120/600 = 0.20$ for the A_2A_2 homozygote.

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 would 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

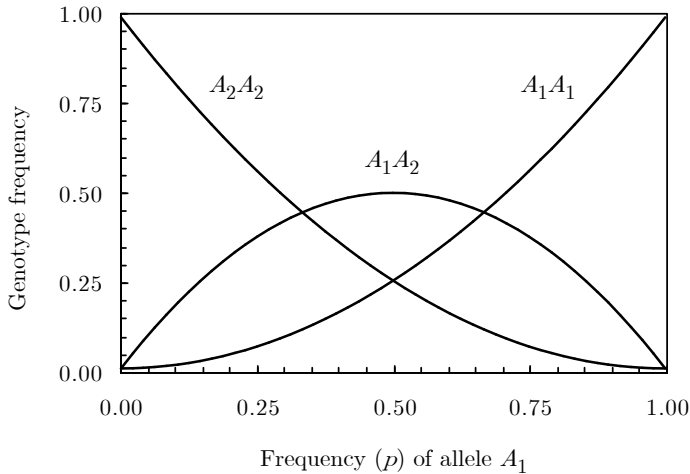


FIGURE 2.1. Genotype frequencies at Hardy-Weinberg equilibrium.

$= q = 0.50$ (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 more than 60 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 i th 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 populations, assortative mating, selection, and inbreeding during the development of progeny. Some of these procedures, such as inbreeding and the use

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