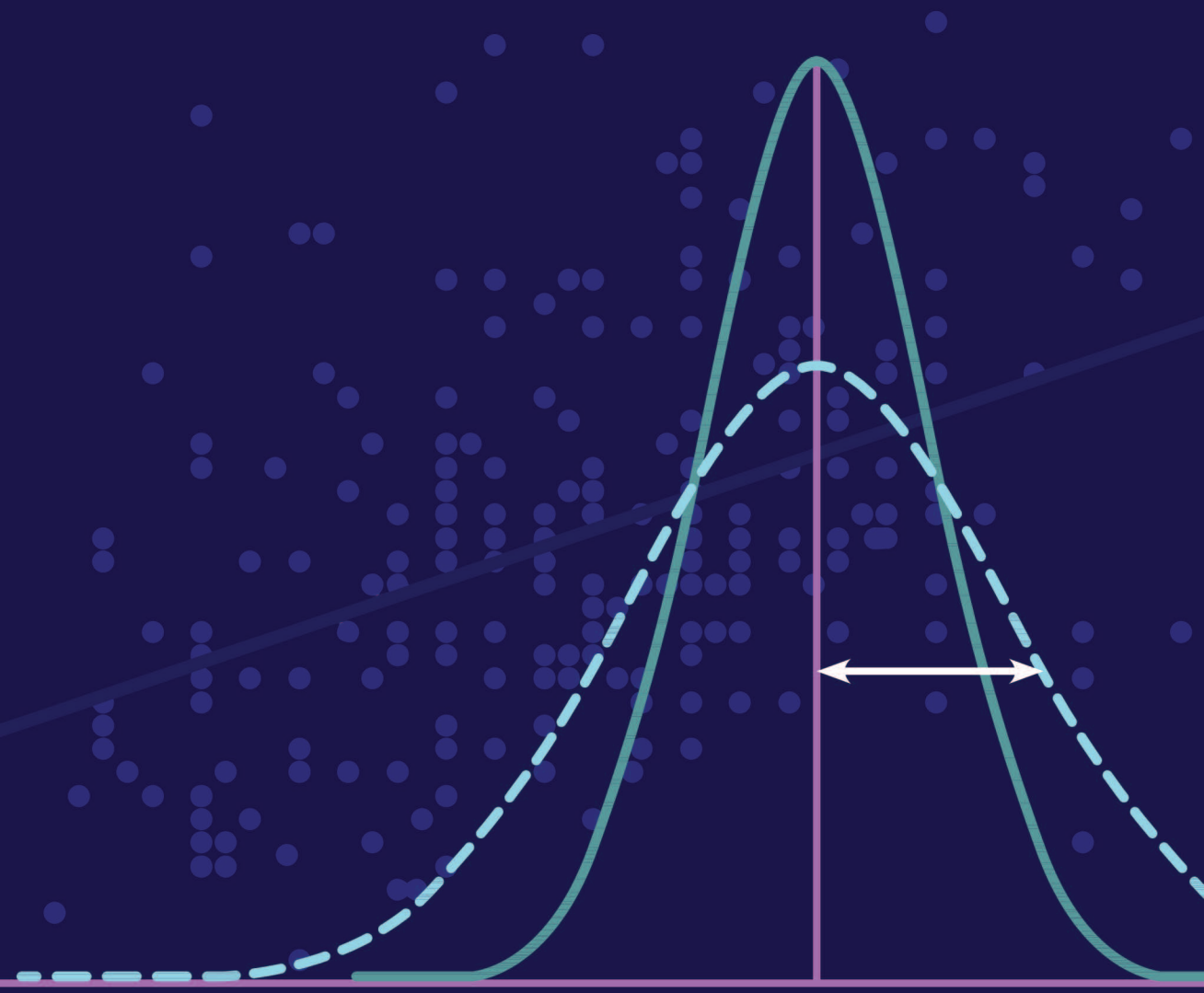
ARMANDO CABALLERO

Quantitative Genetics



Quantitative Genetics

Quantitative genetics is the study of continuously varying traits, which make up the majority of biological attributes of evolutionary and commercial interest. This book provides a much-needed up-to-date, in-depth, yet accessible text for this field. In lucid language, the author guides readers through the main concepts of population and quantitative genetics and their applications. Written to be approachable even to those without a strong mathematical background, applied examples, a glossary of key terms, and problems and solutions support students in grasping important theoretical developments and their relevance to real-world biology. This is an engaging, must-have textbook for advanced undergraduate and postgraduate students. Given its applied focus, it also equips researchers in genetics, genomics, evolutionary biology, animal and plant breeding and conservation genetics with the understanding and tools for genetic improvement, comprehension of the genetic basis of human diseases and conservation of biological resources.

Armando Caballero is Professor of Genetics at the University of Vigo, Spain, with research interests in quantitative and population genetics, conservation genetics and evolution. He has served as Associate Editor for the journals *Evolution*, *American Naturalist*, *Journal of Evolutionary Biology*, *Genetics Selection Evolution* and *Heredity.*

Quantitative Genetics

ARMANDO CABALLERO
University of Vigo, Galicia, Spain

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108481410
DOI: 10.1017/9781108630542

Originally published in Spanish as *Genética Cuantitativa* (2017) by Editorial Síntesis, Madrid, Spain.

First published in English by Cambridge University Press in 2020, translated and updated from the original Spanish edition by Armando Caballero

First published 2020

Printed in the United Kingdom by TJ International Ltd, Padstow Cornwall 2020

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Caballero, Armando, 1960– author.
Title: Quantitative genetics / Armando Caballero, University of Vigo, Spain.
Other titles: Genética cuantitativa. English
Description: Cambridge, United Kingdom ; New York : Cambridge University Press, 2020. | Originally published in Spanish as Genética Cuantitativa (2017) by Editorial Síntesis, Madrid, Spain. | Includes bibliographical references and index.
Identifiers: LCCN 2019039768 (print) | LCCN 2019039769 (ebook) | ISBN 9781108630542 (epub) | ISBN 9781108481410 (hardback) | ISBN 9781108722353 (paperback)
Subjects: LCSH: Quantitative genetics.
Classification: LCC QH452.7 (ebook) | LCC QH452.7 .C3313 2020 (print) | DDC 572.8/6–dc23
LC record available at https://lccn.loc.gov/2019039768

ISBN 978-1-108-48141-0 Hardback
ISBN 978-1-108-72235-3 Paperback

To Bill Hill

Contents

Preface *page* xiii
Preface to the Spanish Version xv

1 Continuous Variation 1
1.1 Quantitative Traits 1
1.2 Basic Concepts and Definitions 3
1.3 Historical Perspective 4
1.3.1 Beginnings of Quantitative Genetics: Heritable Variation and Evolution 4
1.3.2 The Development of the Central Body of Quantitative and Population Genetics 7
1.4 The Infinitesimal Model 10
Problems 13
Self-Assessment Questions 14

2 Forces of Change in the Allele Frequencies 15
2.1 Allele, Gamete and Genotype Frequencies 15
2.2 Hardy–Weinberg Equilibrium 16
2.3 Gametic or Linkage Disequilibrium 19
2.4 Forces of Change in the Allele Frequencies 23
2.5 Genetic Drift: The Ideal Population of Wright–Fisher 23
2.6 Change in Allele Frequencies by Mutation 26
2.7 Change in Allele Frequencies Due to Migration 27
2.8 Change in Allele Frequencies by Natural Selection 28
2.8.1 General Model of Fitness for a Locus 28
2.8.2 Change of Frequency of a Lethal Recessive Allele 30
2.8.3 Change in the Allele Frequency of a Favourable Allele 32
2.8.4 Overdominance and Underdominance 33
2.8.5 Natural Selection and Hardy–Weinberg Disequilibrium 37
2.8.6 Recreation of the Evolutionary Change in the Laboratory 38
Problems 39
Self-Assessment Questions 40

3 Components of Phenotypic Values and Variances 42
3.1 Decomposition of the Genotypic Value for a Locus 42
3.2 Decomposition of the Genotypic Variance for a Locus 46
3.3 Decomposition of Genotypic Values and Variances for More than One Locus 48
3.3.1 Gametic Disequilibrium Variance 48

3.3.2 Epistatic Variance 51
3.4 Concepts of Heritability and Genetic Correlation 54
3.4.1 Degree of Genetic Determination 54
3.4.2 Heritability 55
3.4.3 Genetic Correlation between Traits 57
3.5 The Environmental Deviation and Its Contribution to Phenotypic Variance 59
3.5.1 Phenotypic Plasticity 59
3.5.2 Genotype–Environment Interaction 61
3.5.3 Genotype–Environment Correlation 63
3.5.4 Decomposition of the Environmental Variance: Repeatability 63
Problems 65
Self-Assessment Questions 66

4 Inbreeding and Coancestry 67
4.1 The Coefficients of Inbreeding and Coancestry 67
4.1.1 Calculation of F and f from Genealogies 69
4.1.2 Molecular Coancestry and Inbreeding Coefficients 71
4.1.3 The Expected Inbreeding in the Ideal Population 76
4.2 Populations with Regular Inbreeding 78
4.2.1 Highly Inbred Lines 78
4.2.2 Inbreeding in Large Populations 80
4.3 Modulation of Inbreeding by the Effect of Mutation and Selection 81
4.3.1 Mutation–Drift Equilibrium 82
4.3.2 The Purged Inbreeding Coefficient 82
4.4 Inbreeding in Subdivided Populations 83
4.4.1 Wright's F Statistics 83
4.4.2 Change of Base Population in the Estimation of Inbreeding 87
4.4.3 Migration–Drift Equilibrium 88
4.4.4 Estimation of F_{ST} with Genetic Markers 90
Problems 93
Self-Assessment Questions 93

5 Effective Population Size 95
5.1 Definition and Methods for the Prediction of the Effective Population Size 95
5.2 Prediction of N_e in Unselected Populations 97
5.2.1 Absence of Self-Fertilization 97
5.2.2 Different Number of Males and Females 98
5.2.3 Variable Population Size across Generations 98
5.2.4 Non-Random Contribution of Parents to Offspring 99
5.2.5 Partial Mating among Relatives 101
5.2.6 Overlapping Generations 102
5.2.7 Different Models of Inheritance and Reproduction 103

5.3 Prediction of N_e in Selected Populations 104
5.3.1 Cumulative Effect of Selection 104
5.3.2 The Impact of Linkage 106
5.4 Prediction of N_e in Subdivided Populations 107
5.4.1 Prediction with Different Models of Population Structure 107
5.4.2 General Model 107
5.5 Applications of the Theory of Effective Population Size to Conservation 109
5.5.1 Contributions with Minimal Variance 109
5.5.2 Mating Systems 111
5.6 Estimation of Effective Population Size Using Demographic Methods 113
5.7 Estimation of Effective Population Size Using the Allelism of Lethals Method 114
5.8 Estimation of Effective Population Size with Molecular Markers 116
5.8.1 Heterozygosity Excess Method 116
5.8.2 Linkage Disequilibrium Method 116
5.8.3 Temporal Method 117
5.8.4 Coancestry and Sib Frequency Methods 118
5.8.5 Methods with Multiple Information Sources 118
5.9 Estimates of N_e/N in Natural Populations 119
Problems 119
Self-Assessment Questions 120

6 Estimation of Genetic Values, Variances and Covariances 122
6.1 Estimation of Heritability with Simple Experimental Designs 122
6.1.1 Estimation Based on the Degree of Resemblance between Parents and Offspring 123
6.1.2 Estimation Based on the Degree of Resemblance between Sibs 125
6.1.3 Effect of Assortative Mating 129
6.1.4 Estimation Based on the Degree of Similarity between Twins 129
6.1.5 Coefficients of Additive and Dominance Relationships 131
6.2 Estimation of Genetic Correlation 132
6.3 Estimation of Variance Components and Prediction of Additive Values with Complex Structure of the Data 133
6.3.1 Estimation of Mean and Variance by Maximum Likelihood 134
6.3.2 REML Estimation with the Animal Model 136
6.3.3 Prediction of Additive Values by BLUP 138
6.3.4 Example of BLUP Prediction and BLUE Estimation 140
6.3.5 Use of Molecular Coancestries with Genetic Markers 142
6.3.6 Comparison between Estimates of Heritability with Genealogical and Molecular Data 144
Problems 145
Self-Assessment Questions 146

7 Mutation 148
7.1 Estimation and Analysis of Mutation in Quantitative Traits 148
7.1.1 Probability of Fixation of a Mutation 150
7.1.2 Estimation of the Rate of Mutation and Mutational Effects 152
7.1.3 Estimation of the Dominance Coefficient 157
7.1.4 Beneficial Mutations and Summary of Mutational Parameters for Fitness 160
7.1.5 Combined Effect of Mutations and Environmental Factors 161
7.1.6 Estimation of Mutational Parameters for Neutral or Quasi-Neutral Traits 163
7.2 Implications of Deleterious Mutations in Populations of Large Size 165
7.2.1 Equilibrium between Deleterious Mutation and Selection 165
7.2.2 Mutation Load 166
7.2.3 Estimation of the Average Dominance Coefficient in the Mutation–Selection Equilibrium 168
7.2.4 Estimation of Mutational Parameters in the Mutation–Selection Equilibrium from Data of Panmictic and Inbred Populations 169
7.3 Mutation and Recombination 170
7.3.1 Evolutionary Advantage of Recombination 170
7.3.2 The Hill–Robertson Effect 172
Problems 173
Self-Assessment Questions 174

8 Consequences of Inbreeding 175
8.1 Effects of Inbreeding on the Mean and Variance of Quantitative Traits 176
8.1.1 Decomposition of the Genotypic Value and Variance in a Non-Panmictic Population 176
8.1.2 Estimation of Inbreeding Depression and Inbreeding Load 179
8.2 Inbreeding in Panmictic Populations of Reduced Census Size 185
8.2.1 Redistribution of Within- and Between-Line Genetic Variance 185
8.2.2 Genetic Differentiation in Quantitative Traits 188
8.2.3 Mutational Meltdown and Purging of the Inbreeding Load 191
8.3 Evolution of Inbreeding in Natural Populations 193
8.4 Crossbreeding and Heterosis 196
8.4.1 General and Specific Combining Abilities 197
8.5 Applications in Conservation 201
8.5.1 The Minimum Effective Size of a Viable Population 202
8.5.2 Consequences of Conservation Methods 203
8.5.3 Molecular Variation as a Complement to Quantitative Genetic Variation 204
Problems 205
Self-Assessment Questions 206

9 Artificial Selection 207
9.1 Principles of Artificial Selection and Its Applications 207
9.1.1 Response to Selection and Its Prediction 208
9.1.2 Correlated Response 211
9.1.3 Measure of the Response 212
9.1.4 Asymmetry of the Response 214
9.1.5 Change of Allele Frequency Due to Selection 216
9.2 Effect of Selection on Genetic Variance 218
9.2.1 The Bulmer Effect 218
9.2.2 The Effective Population Size with Artificial Selection 219
9.2.3 Combined Effect of Genetic Drift and Selection on Genetic Variance 220
9.3 Long-Term Response 221
9.3.1 Prediction of the Long-Term Selection Response 221
9.3.2 Comparison with Empirical Data 224
9.4 Family, Within-Family and BLUP Selection 228
9.5 Use of Molecular Markers 231
Problems 232
Self-Assessment Questions 233

10 Natural Selection 234
10.1 Quantitative Traits and Fitness 234
10.2 The Response to Natural Selection 235
10.3 Directional, Stabilizing and Diversifying Selection 238
10.3.1 The Intensity of Real and Apparent Stabilizing Selection 240
10.3.2 Selection in Heterogeneous Environments 242
10.3.3 Genetic Variance and Natural Selection 245
10.4 Genomic Footprint of Natural Selection 247
Problems 247
Self-Assessment Questions 248

11 Genomic Analysis of Quantitative Traits 250
11.1 Mapping of Quantitative Trait Loci 250
11.2 Genome-Wide Association Studies 253
11.3 Detection of the Footprint of Selection 258
11.3.1 Principles of Methods Based on Polymorphism and Divergence 258
11.3.2 Linkage Disequilibrium Methods 259
11.3.3 Detection of Diversifying and Convergent Selection 261
11.4 Genomic Selection 263
Problems 268
Self-Assessment Questions 268

Solutions to Problems and Self-Assessment Questions 269

Glossary 286
References 299
Index 318

Preface

I want to thank Cambridge University for allowing me to produce the English version of my Spanish book and for the continuous help and support. For this version, I have updated some content and references, correcting a few minor errors present in the Spanish version. I want to thank Professor Carlos López-Fanjul for having read the complete draft, making additional comments to improve the book, and to Professor Peter Keightley for reading Chapters 7 and 8 and making useful suggestions and corrections. I also want to thank Raquel Sampedro for help in managing the text and Miguel Toro, Beatriz Villanueva, Antonio Carvajal and Humberto Quesada for helpful discussions on specific parts of the book. Any errors in the content are, however, my sole responsibility, and I would appreciate any comments in this regard. Finally, I want to express my gratitude to my family for their understanding and support during the preparation of the manuscript.

Preface to the Spanish Version

Quantitative genetics is the branch of genetics dedicated to the study of quantitative traits, which are all those biological attributes with continuous variation. The vast majority of traits of evolutionary interest and commercial value in animals and domestic plants are quantitative, so that quantitative genetics contributes to the understanding of adaptation and evolution of living beings and provides the necessary tools for the genetic improvement and conservation of biological resources. A large number of human diseases are also of a quantitative nature, so that the study of quantitative genetics is fundamental for their understanding.

This book aims to build the foundations on which the subject is based, trying to maintain a balance between explanations of the most basic concepts and the most modern methods of analysis. It is therefore intended to be not only an introduction for students, particularly postgraduates, but also a reference book for researchers. Understanding this subject requires an elementary knowledge of mathematics and statistics, and, whenever possible, I have attempted to explain, or at least indicate, the origin of the equations that reflect the different concepts. Either the original sources of the main contributions or useful reviews or books have been cited so that the reader can get more insight into each subject. Each of the 11 chapters includes some problems and questions, solved at the end of the book, and there is a glossary of the most important terms.

I want to express my gratitude to several colleagues who have contributed with data or comments or by amending errors. First of all, and especially, I have to thank Professor Carlos López-Fanjul for having read the complete draft and for making numerous corrections, comments and suggestions that have substantially improved the content and its presentation. I owe him, as well as Professor Bill Hill of the University of Edinburgh, a great deal of my knowledge in this field, as well as teaching me how to carry out high-quality research. I also want to thank Professors Aurora García-Dorado and Miguel Toro for their help with various chapters. Other colleagues have contributed comments and discussions for specific topics or have contributed data, photos or figures, among whom are my colleagues from the University of Vigo, Humberto Quesada, Paloma Morán, Emilio Rolán-Alvarez, Juan Galindo and Daniel Estévez-Barcia, and from other institutions, Beatriz Villanueva, Jesús Fernández, Almudena Fernández and Andrés Legarra. I also want to express my gratitude to the coordinator of the series of which this book is a part, César (Mario) Benito, for reading the complete draft, making corrections to the text and suggesting improvements. Any errors that remain in the book are, however, my responsibility. Finally, I have to thank my wife, Esther, and my children, Alberto and Laura, for their understanding and support during the long work sessions.

1 Continuous Variation

Concepts to Study

- Quantitative traits
- Meristic and threshold traits
- Genotypic and phenotypic values
- Additivity, dominance and epistasis
- Major gene
- Pleiotropy
- Fitness and its main components
- Infinitesimal model

Objectives for Learning

- To understand the definition of quantitative trait, the reason for the different names by which they are known and their types depending on whether they are expressed with continuous or discrete observable variation
- To distinguish the concepts of phenotypic and genotypic value and the types of intra-and interlocus gene action
- To understand the concept of fitness and its main components
- To know the definitions of major gene and pleiotropy
- To understand the infinitesimal model and the partition of phenotypic variation in genetic and environmental components

1.1 Quantitative Traits

Some heritable characteristics are qualitative, with an expression clearly identifiable in discrete classes. Such is the case of attributes like some differences in colour, shape or structure, by which individuals of a population or species can be classified. The analysis of this type of simple character was what allowed Mendel to describe the bases of inheritance and many other geneticists, later, to understand the relation between this and the chromosomal behaviour during reproduction, as well as the interactions between genes. However, most of the traits that we find in nature present a continuous variation. Even some of the seemingly discrete attributes, such as colour, may show gradual variation if analysed in detail. These types of characters with gradual variation are called quantitative traits and, sometimes, metric or continuous traits. Among them it is possible to find many with purely continuous variation, with a priori infinite possibilities of expression, whose analysis is based on measurement, such as body dimensions or weight, but also those with discrete variation whose characterization is carried out by counting, the so-called meristic traits, such as the morphological structures that vary in number, or any discrete characteristic that implies numerical variation in its

expression, such as the number of offspring of the same birth or the number of matings that an individual carries out throughout its life. Even some characters whose expression is displayed in only a few possible categories, the so-called threshold traits, such as susceptibility or resistance to certain substances, death or survival or the circumstance of suffering, or not, a disease, often imply an underlying continuous variation.

In fact, there are an indefinitely large number of heritable biological characteristics that can present a continuous variation, observable as such or underlying, and this becomes evident when verifying that practically all variable characters that have been studied in laboratory organisms or domestic species have a certain hereditary component, which can be exploited by artificial selection. The potential of genetic change is evident in the enormous morphological differences generated, for example, between dog breeds, where the weight of the largest breeds is of the order of 200 times greater than that of the smallest ones, or the tremendous improvement obtained in the productive traits of domestic animals and plants, for example, the duplication of milk production in cattle in the last 50 years, it being possible to attribute at least half of this increase to genetic changes (Hill, 2014).

There are two fundamental premises that characterize this variation. First is the genetic control by a large enough number (say greater than 5) of generally unknown genes – the reason why the quantitative traits are often called polygenic or multifactorial traits. Second is that its expression is influenced, to a greater or lesser extent, by environmental factors intrinsic or extrinsic to the individual. As we will see in the different chapters of this book, these are precisely the two complications that characterize the analysis of continuous variation: (1) the ignorance of all or a large part of its genetic bases and (2) the ignorance of the relative influence of inheritance and environment in its phenotypic expression.

The importance of quantitative traits is fundamental. First, from an evolutionary point of view, all the attributes of a reproductive nature that are the direct object of natural selection, such as viability, fertility or mating success, are quantitative. Second, in the practical aspects of human consumption, the large majority of productive characteristics of plants and domestic animals, such as milk production, egg laying, quality and quantity of meat, plant biomass and a long endless list, are also quantitative. A large part of the development of quantitative genetics has occurred thanks to the continuous interest to improve animal and plant production through artificial selection and crossing methods. Third, with respect to human longevity and well-being, many of the most common diseases in our species, such as cancer, psychiatric disorders, autoimmune disorders and excess cholesterol or blood pressure, among many others, are also quantitative traits, which in the medical jargon are also called complex traits. A large part of the current genetic research focuses precisely on finding out the genetic basis of the aforementioned characters and on delimiting the relative importance of the environment in their expression. This has allowed us to find thousands of genes involved in hundreds of traits, both in wild species and those of economic interest, as well as in our own (MacArthur et al., 2017).

Quantitative genetics focuses mainly on the analysis of continuous variation and its applications to the study of evolution, animal and plant breeding and medicine. However, its imbrication with population genetics is total, since the analysis of quantitative traits is always carried out in the context of a population. This latter can be defined, in this context, as a set of individuals that constitute a reproductive unit, that is, those that are connected by a spatial and

temporal relationship and share the same gene pool, the genes of which they are carriers. Population genetics includes the study of the heritable variation in general, qualitative or quantitative, and the forces of change of the gene frequencies that act on it. Such study is essential to understand biological evolution because this is, fundamentally, the result of the spatial-temporal change of the genetic composition of the populations. Since, as we have already indicated, most traits of evolutionary importance are quantitative, a fundamental objective of population genetics and, therefore, of quantitative genetics, is the understanding of the evolutionary process by means of the study of the genetic (mutation, chromosome segregation, recombination, etc.), ecological and demographic (population census size, migration, geographic distribution of populations, etc.) and adaptive processes (natural selection) that act on populations. The basic questions refer to which are the forces that maintain the variability in the populations, what role they play with respect to each other and how the integration of the phenotypes in the environments is accomplished, that is, how does the adaptation of the organisms occur to the environment in which they live. The application during millennia of artificial selection on a great number of quantitative traits, and the analysis and development of mathematical methods to reveal the consequences of this type of selection, has allowed us to understand, in turn, a good part of the way in which natural selection operates. The instruments used in quantitative genetics encompass the statistical characterization of populations for different traits, the performance of controlled experiments and the formulation of theoretical models. The latter are the primary tool, allowing a simplified description of the observed phenomena, which facilitates the prediction of natural processes and the response to artificial selection of domestic plants and animals.

1.2 Basic Concepts and Definitions

The genome of an individual is constituted by a great variety of elements, with the traditional concept of the gene being a functional unit located in a unique position of the genome. For practical purposes we will use the term *locus* to refer to a genomic element located in a fixed position of the genome, which may have different variants (alleles) which may or may not have an impact on the expression of the trait in the individual. Although the term *gene* refers to a functional unit, in this book we will use locus and gene interchangeably in most cases. For a particular locus, the genotype of a diploid individual can be homozygous, if it carries two copies of the same allele, or heterozygous, if it carries different alleles. As we will see in the next chapter, the genetic description of the populations can be done in terms of the allele or gene frequencies when the genotypes are distinguishable. For most quantitative characteristics, however, this is not possible, and the analysis will be limited to the calculation of general trait parameters in the population, such as means and variances. Sometimes there are loci with such a large effect on a trait that it allows us to distinguish their genotypic classes. These loci are called major genes, unlike the rest, which are called minor genes.

For a quantitative trait, the effects of the alleles of the different genes of which an individual is a carrier can, in a simplified form, be added to constitute the genotypic value of the individual, in which case it is said that there is additive gene action or additivity, but the relationships between alleles intra- or interlocus can be more complex. The partial or total prevalence of the effect of one allele on another at a given locus constitutes the concept of

dominance, analogous to that applied for qualitative characteristics (the dominant allele is the one that prevails over the recessive allele). Likewise, the lack of additivity between the effects of different loci is called, generically, epistasis or epistatic interaction. The expression of the trait for the individual, its phenotypic value, is the result of the combination of its genotypic value and the effect of the environmental factors that surround it, as well as a possible interaction between both that we will discuss in later chapters. Finally, when a given locus has an effect on more than one quantitative trait, it is said that there is pleiotropy. All these concepts will be explained in greater depth in later chapters.

Any trait with a continuous genetic basis is an object of study by quantitative genetics. In the evolutionary context, however, the most relevant characters are those related to the reproduction of individuals. Darwin's (1859) Theory of Natural Selection is fundamentally based on the competition of individuals for resources, which determines their survival and/or differential reproduction, which constitutes their fitness. This trait can be defined as the contribution of the individual with offspring to the next or future generations, and it is the subject of direct natural selection. As we will see in later chapters, the indirect impact of natural selection on any other quantitative trait depends on the relationship between this latter trait and fitness. For natural selection to act, it is enough that there are inherited differences in fitness between the different individuals. Given the abstraction of its definition, fitness is difficult to evaluate, and in practice, it is specified in more empirically accessible traits, grouped under the name of the main components of fitness that are, among others of lesser importance, viability or survival, fecundity and mating success. A substantial part of quantitative genetics is dedicated to the study of the genetic variation for these traits and their implication in the context of evolution, animal and plant breeding and conservation.

1.3 Historical Perspective

1.3.1 Beginnings of Quantitative Genetics: Heritable Variation and Evolution

The beginnings of quantitative genetics, like those of genetics in general, are linked to the development of the theory of evolution and the elucidation of the bases of inheritance. The evolutionary theory consists essentially of two complementary parts. One of them proposes the mechanism of natural selection whose action can result in adaptive change. The Theory of Natural Selection by Charles Darwin (1859) offers an adaptation mechanism that represents the response of the individual to the constant challenge posed by the variation of the environment, in which individuals are passive subjects selected by it. The other part of the evolutionary theory specifies the hereditary principle, whose purpose is to guarantee the relative permanence of the acquired change. In this respect, Darwin held to the opinions of his time, accepting, as was general then, the mixed inheritance of the characters of both parents, which explained the similarity between the parents and their progeny. However, this type of inheritance implied a continuous loss of variation with the passing of generations, which provoked important criticisms. Gregor Mendel's experiments, based on the study of qualitative pea characters, and his idea of particle inheritance, published in 1865, although almost totally unknown until 1900, could explain the resemblance between parents and offspring without implying a constant loss of genetic variation. At the end of the nineteenth century, William Bateson distinguished

two types of variability: continuous and discontinuous. The first one referred to the small differences between the organisms that constituted, for Darwin, the raw material on which natural selection would act. The second was, however, the one that Bateson considered essential to explain the observable variability. This notion was later developed by Hugo de Vries, who introduced at the beginning of the twentieth century the term *mutation* to designate the hereditary changes he considered as a norm: those with large and discontinuous effects. The legacy of Bateson and de Vries was embraced by other researchers, and the new science, baptized by Bateson as genetics, was built, in a sense, as the strongest opponent of Darwin's theory, since it presented mutation as the engine of evolutionary change.

During the period between the publication of Mendel's experiments and his rediscovery, Francis Galton, Darwin's cousin, also tried to elucidate the principles of hereditary transmission. However, Galton's (1889) approach was different from Mendel's. On the one hand, he was interested in the transmission of those characters that presented a continuous variation instead of focusing on those that showed clearly differentiable alternative forms. On the other, his study focused on the measurement of the average similarity between individuals with a certain degree of kinship in order to achieve a predictive model. Galton pointed out the fact that by representing the average height of the offspring against the average height of the parents, a linear relationship was obtained (Figure 1.1). However, the slope of the line indicated that children's height deviated less from the average population than that of their parents, that is, parents who were shorter than average had children with a stature somewhat greater than their own, and those who had a height higher than the average had shorter children. Galton called this circumstance 'regression towards mediocrity' and interpreted that this would prevent any selective progress, concluding that evolution should be based on variants of great effect rather than on the result of the action of selection on continuous variation. Karl Pearson would later prove that the regression to mediocrity deduced by Galton did not necessarily imply a problem for evolution. The linear regression technique, widely used in statistical analysis in all fields, comes from these Galton studies and was formally developed by Pearson along with the correlation technique. Likewise, other statistical

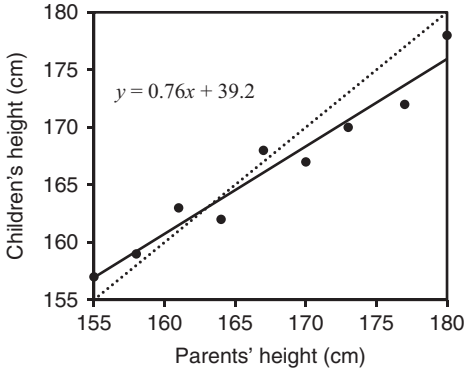


Figure 1.1 Type of representation similar to those made by Galton, comparing the average height of children with that of their parents. The regression line presents a slope of 0.76, lower than that which would be expected with a perfect resemblance relationship between parents' and children's heights (dotted line).

techniques, such as the analysis of variance and various hypothesis testing methods, introduced by Ronald A. Fisher (1918), also have their origin in the analysis of quantitative traits. Finally, the path analysis of Sewall Wright (1921), applied to the study of inbreeding, has had some applications in the social and ecological sciences. Therefore, it can be said that quantitative genetics and its applications, such as animal genetic improvement (Gianola and Rosa, 2015), have contributed in an essential way to the development of many of the statistical techniques of universal use.

Biometry began with Galton and also the conflict between biometric and Mendelian conceptions of inheritance. For the Mendelians, with Bateson in the lead, the object of the science of inheritance was to develop a model of the process of transmission from parents to offspring of those factors that determined observable characteristics of individuals. For the biometricians, led by Pearson and Weldon, the objective, on the contrary, was to measure the phenotypic resemblance between the individuals of a population, since they doubted that the Mendelian laws could be applied to the traits with continuous variation. These two visions of inheritance also reflected a different conception of evolution. For the Mendelians or mutationists, this would be produced by rapid changes of great magnitude, while for the biometricians, it would occur continuously and gradually, more in accordance with the Darwinian model. The Mendelians argued that the quantitative characteristics reflected, basically, environmental differences, creating the debate of inheritance versus environment, or nature–nurture. A methodological interpretation that led on this and other occasions to error is to think that the observable is the general norm. Given that the effects of the major genes are the only ones observed empirically, it is easy to assume that only these exist and to ignore the contribution of others of smaller effect. Something very similar occurred with the denial until the 1960s of the additive gene action. When studying major genes, generally recessive, it was assumed that this was the general form of gene action. That is, again, the experimental need was taken as a general rule. Udny Yule (1902) was the first to seek the connection between both visions of inheritance, studying the application of Mendelian laws to panmictic populations, although the general formulation was proposed in 1908 independently by Godfrey Hardy in England and Wilhelm Weinberg in Germany, arriving at the Hardy–Weinberg principle that we will study in the next chapter.

Wilhelm Johannsen (1903) managed to take a big step towards the reconciliation of the two positions, showing the intervention of the medium in the expression of quantitative traits in his theory of pure lines. With his experiments he showed that the continuous variation observed for the trait weight in beans was produced by the combined influence of the genes and the environment. The bean is, like the pea used by Mendel, an autogamous legume species. Starting from 19 seeds that differed in weight, Johannsen established 19 lines by continuous self-fertilization that also differed in the average weight of the seeds, between 34 and 64 cg. He also observed a certain variation in weight between the individuals of each of the lines around their average. He showed first that any individual of a line with heavy or light beans gave rise to descendants which maintained the average parental weight. Johannsen also found that if the same line was selected for several generations to increase or decrease seed weight, no response was obtained, which indicated that there was no genetic variation within that line. His conclusion was that, since it is a selfer and each line was produced by self-fertilization from a single initial seed, each line would be formed by individuals with the same

homozygous genetic constitution (pure lines) and that, therefore, the great variation between the average weights of the lines was due to the different genetic constitution of each of them, while the smaller variation in weight observed among the individuals of a given line came from environmental sources. Johannsen coined the terms genotype, to denote the genetically identical individuals of each line, and phenotype, for the observable value of the trait in each individual that would be the result of genetic and environmental effects.

Another important advance in the reconciliation between Mendelian and biometric hypotheses came from the hand of George Shull in 1908, studying characters of corn, and, above all, Herman Nilsson-Ehle, working with cereals, who attributed the hereditary determination of a trait to the segregation of several genes of similar and cumulative effects, giving support to the multifactorial hypothesis of quantitative traits or theory of the polymeric factors. Through crosses of varieties of wheat with flowers of different colours, Nilsson-Ehle (1909) found that several genes contributed to the variation of colour tones of the flower, from white to intense red, and that the individual effects of the genes were small and summable. Subsequent studies by other researchers (mainly R. A. Emerson and E. M. East with corn) confirmed the multifactorial hypothesis of quantitative traits.

1.3.2 The Development of the Central Body of Quantitative and Population Genetics

Both quantitative genetics and theoretical population genetics were developed gradually from the early twentieth century and reached maturity in the early 1930s with the publication of three essential works: *The Genetic Theory of Natural Selection*, by Ronald A. Fisher (1930), *Evolution in Mendelian Populations*, by Sewall Wright (1931), and *The Causes of Evolution*, by John B. S. Haldane (1932). Fisher mathematically demonstrated that natural selection, by acting on the genetic variability of populations, could perfectly explain the evolutionary change. Fisher's legacy to the theoretical body of quantitative genetics is essential, in particular the proposal of the Fundamental Theorem of Natural Selection, which sets the theoretical basis of the consequences of the action of natural selection on fitness, and the development of the so-called infinitesimal model, which proposes an interpretation of the nature of continuous variation in discrete Mendelian terms. Using the techniques of analysis of variance, also devised by him, the phenotypic variance of a certain character can be broken down into a series of components attributable to different genotypic and environmental causes. Thus, the genotypic variance can be ascribed to an additive component, due to the average effect of genes, and others due to the effects of dominance and epistasis, on which we will get insight in later chapters. The additive genetic component can be estimated with relative ease and is of great importance because it is the determinant of the immediate response to selection and family resemblance. The proportion of the phenotypic variance explained by this additive component constitutes the concept of heritability.

Among the different authors who carried out notable extensions to Fisher's theoretical central body is Jay Lush, who is considered the father of animal breeding. His book *Animal Breeding Plans* (Lush, 1945) gathers numerous applications of quantitative genetics to the genetic improvement of animals, among which stands out the well-known 'breeder's equation', by which selection response is predicted in terms of the heritability of the trait and the selection differential applied, which is the selection pressure exerted on the population.

The development of selection indices, through which individual and family information of a trait can be combined to obtain a greater response, is also due to him (Lush, 1947). Michael Lerner also made important contributions to quantitative theoretical and practical genetics collected in his book *Population Genetics and Animal Improvement* (Lerner, 1950), and more evolutionary aspects, such as the development of the theory of genetic homeostasis (Lerner, 1954), which explains the greater plasticity of heterozygotes in the face of environmental variation. In the decades of the 1960s and 1970s the contributions of Alan Robertson stood out, with the theory of limits to selection under the infinitesimal model and the Second Theorem of Natural Selection, which explains how natural selection can result in adaptation through the trait-fitness additive covariance. Together with William (Bill) Hill he also made important contributions to animal breeding, as well as to population and evolutionary genetics, highlighting his predictions about the behaviour under selection of genes physically linked in the chromosome. Hill also extended Robertson's selection limits theory by introducing the impact of mutation. Many of these applications have been collected in the different editions of Douglas Falconer's famous book, *An Introduction to Quantitative Genetics* (Falconer and Mackay, 1996), and a lucid historical summary of the most important contributions of quantitative genetics to animal breeding can be seen in Hill's (2014) review.

Haldane made important contributions to the theoretical body of population genetics, with immediate application in quantitative genetics, establishing expressions on the fate of advantageous mutations in populations and deriving the conditions of polymorphisms in some situations. His most interesting contribution is the deduction of the balance that is reached between the appearance of deleterious mutations, those that reduce fitness, and their elimination by selection. Natural selection acts as a purging factor of the deleterious variability that is constantly generated by mutation, and the populations carry a 'mutation load', a concept due to Hermann Muller, which is simply equal to the diploid mutation rate. Thus, a high mutation rate could lead to negative implications for the population, particularly in asexual species (Muller, 1932). In these latter, the accumulation of deleterious mutations is very fast due to the lack of genetic recombination, which would allow their better elimination by natural selection. Thus, each time a mutation is fixed in an asexual population, the genome with fewer mutations will carry one more, without the possibility of going back to its previous state, a phenomenon known as Muller's Ratchet. The classic studies of Terumi Mukai in the 1960s and 1970s indicated that the deleterious mutation rate for viability in *Drosophila* was very high and mutational effects were small but high enough to be harmful, corroborating Muller's vision. The high mutation load could only be overcome by a greater effectiveness of selection with a synergistic effect of mutations, that is, when the combined effect of two or more mutations is greater than the sum of the effects of each of them separately.

Wright's contribution to population genetics is probably the most extensive and is summarized in the four volumes of his book *Evolution and the Genetics of Populations* published between 1968 and 1978. Wright emphasized the importance of epistatic interactions to produce advantageous combinations of genes on which selection could act and, especially, developed most of the theoretical body of work on the possible impact of chance (genetic drift) on evolution. We owe him the concept of effective population size, which allows us to quantify the effects of inbreeding and genetic drift and which was developed later by James F. Crow and Motoo Kimura. Wright discussed intensively with Fisher and

Haldane about the relative role of evolutionary forces in populations. For Fisher and Haldane, the most advantageous scenario from the evolutionary point of view was that of a large population rich in variability, where selection could act on individual genes of predominantly additive gene action. For Wright, on the contrary, the most favourable situation, gathered in his Shifting Balance Theory, would be that of a population subdivided into small isolated sub-populations in which genetic drift could expose to selection novel combinations of interactive genes, where dominant and epistatic gene actions could play an important role. One of the best introductions to this and other debates on evolutionary issues is found in Crow's (1986) book.

Although with the works of Fisher, Haldane and Wright the mechanisms that act on the genetic variability present in populations and, especially, the mode of action of natural selection on genetic variability were established in the 1930s, its excessively mathematical view meant that a few more years were needed before the neo-Darwinian theory of evolution was popularized thanks to the biological foundation provided by other fields of biology. The so-called modern synthesis represented the integration of the different biological disciplines, previously very separated from each other, in a global context, that of evolutionary biology. Modern population and quantitative genetics represent the contribution of genetics to the synthesis. This contribution conferred to Darwinian evolutionism the capacity to elaborate mathematical models that allow us to treat microevolutionary processes in a general way, thus becoming the core of neo-Darwinism.

The first half of the century of neo-Darwinism was dominated by the pre-eminence given to the selective force as an agent of evolutionary change, with a panselectionist view proposed by the experimentalists and headed by T. Dobzhansky. At the end of the 1960s, an anti-selection reaction occurred, whose most representative aspect is the formulation of the so-called Neutralism, formally developed by Kimura (1983) in his book *The Neutral Theory of Molecular Evolution*. This is not an opposition to neo-Darwinism but an orthodox version of it, in which it is desired to objectively establish the evolutionary importance of the forces that modify the composition of gene pools. Its essential contribution lies in not considering natural selection as a proven fact, but in establishing the null hypothesis against which the selection alternative can be verified in each specific case. A large part of molecular variation behaves as practically neutral, and selection can act very subtly when moulding such variation. The neutral theory was refined with the qualification of *quasi-neutral*, defended by Tomoko Ohta, one of Kimura's collaborators, and the study of the impact of the hitchhiking of deleterious or advantageous mutations on neutral variation, particularly in genomic regions of low recombination, is one of the most active areas of current research (Charlesworth and Charlesworth, 2010).

Kimura's contribution to quantitative and theoretical population genetics is not restricted to the neutral theory, and can be compared in quality and magnitude with that of Wright, Fisher and Haldane. In addition to developing most of the theoretical framework of the study of molecular variation, he made important advances in aspects related to linkage, population structure and sexual reproduction. Most of the theoretical framework of population genetics is condensed in his book with James Crow, *An Introduction to Population Genetics Theory* (Crow and Kimura, 1970), which still has full validity in many aspects.

Many of the fields of study of quantitative genetics developed in the second half of the twentieth century focused on prediction models of genetic variability and evolutionary change, the

elucidation of the nature of quantitative traits and the study of the consequences of inbreeding, selection, population structure and crossing (Barton and Turelli, 1989; Lynch and Walsh, 1998; Walsh and Lynch, 2018). In the area of animal breeding, research focused on the development of more efficient selection methods, making use of information from relatives for the prediction of genetic values and the estimation of fixed effects (Henderson, 1984). In the last 50 years the development of molecular genetic techniques has been strengthened in the study of genetic variability at the level of DNA itself. The discovery of restriction enzymes in the late 1960s and, above all, the polymerase chain reaction technique, in the late 1980s, triggered the development of protocols to allow molecular analysis to be carried out from a few loci to the sequencing of complete genomes. This has exponentially increased the battery of genetic markers available for the elucidation of kinships, with applications in animal and plant breeding, such as genomic selection (Hayes et al., 2013), and in the mapping of genes, with repercussions on animal and plant improvement (Blasco and Toro, 2014), as well as in medicine (Visscher et al., 2017).

1.4 The Infinitesimal Model

As we have previously emphasized, the defining characteristics of quantitative traits are their polygenic nature and the modulation of gene effects by environmental influences, precisely the reasons that explain the continuous variation presented by many of them. To illustrate this, let us consider the possibility that a quantitative trait is controlled by a single locus with two alleles, A and a, so that the presence of allele a determines the addition of 1/2 unit of the trait, while allele A does not add anything to its value. From the offspring of a heterozygous individual Aa, reproduced by self-fertilization, we will find the distribution of genotypic values represented in Figure 1.2a. There will be, therefore, three genotypic classes of individuals, AA, Aa and aa, whose genotypic values are 0, 1/2 and 1 units, respectively.

Suppose now that there are two loci involved in the determination of the trait and, again, an individual heterozygous for the two loci self-fertilizes. If the effect of one of the alleles of each locus is now 1/4, to maintain the same total range of variation as in the case of one locus, and the other is null, and if the individual genotypic values are obtained as the sum of the effects of the alleles of the two loci, we have that there are five genotypic classes in the population (Figure 1.2b). In general, the frequencies of the classes are obtained with the development of the binomial $(0.5 + 0.5)^n$, where n is the number of loci. When more loci are considered, the number of genotypic classes increases at a rate of $2n + 1$ (Figures 1.2c and 1.2d). With 32 loci it can be observed that the genotypic classes begin to be so numerous that the unit of measurement of the trait can become incapable of distinguishing them. If, in addition, the loci had different effects, the number of genotypic classes would increase substantially. Note also that genotypic classes of more extreme values have a non-negligible probability when the number of loci is small, but when this increases the probability of finding individuals from the extreme classes is negligible and huge populations would be needed to detect their presence. For example, with 32 loci, the probability of appearance of an individual of the most extreme class (with 64 alleles that produce no effect, that is, with genotypic value 0, or with 64 alleles that add 1/64 effect, that is, with genotypic value 1) is only 5.42×10^{-20}.

If to the possibility that there are many genotypic classes we add the effect of the environment to configure the phenotypic classes, we can find that the phenotypic variation is

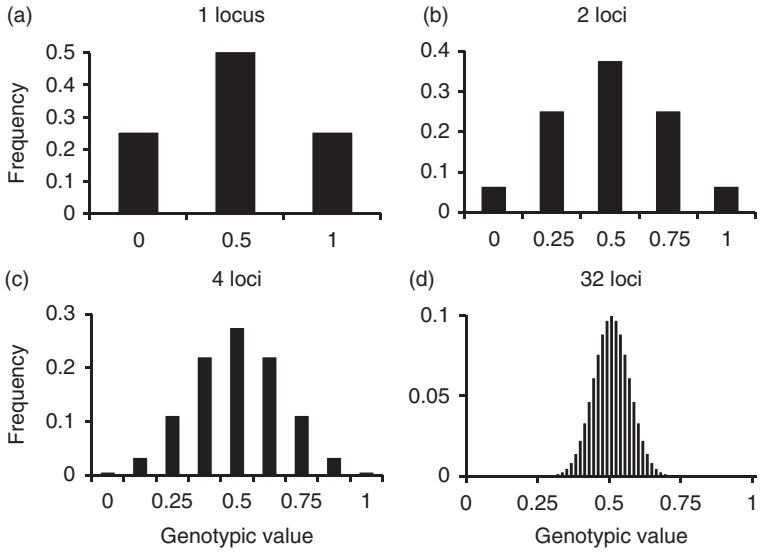


Figure 1.2 Distribution of frequencies of individuals from a population classified by their genotypic value for a quantitative trait controlled by 1, 2, 4 or 32 loci. In each case, the population comes from the self-fertilization of a single individual heterozygous for the loci involved. For each locus, one allele has no effect on the trait, and the other has effect $1/(2n)$, where n is the number of loci, and the effects of the different loci are added to obtain the genotypic value.

purely continuous, with an indefinitely large number of classes that could be approximated by a normal distribution, as can be deduced from Figure 1.2. The vast majority of quantitative traits, following the law of large numbers, have a phenotypic distribution in the form of a Gaussian bell, either in the original units or by making a change of scale, such as taking logarithm or arcsine transformations (see Falconer and Mackay, 1996, chapter 7).

The basic model of variation, proposed by Fisher, is the infinitesimal model, where the genotypic value of an individual (G) is determined by the joint effect of many loci (theoretically an indefinitely large number), with independent segregation, whose effects are small (in theory infinitesimal) and additively cumulative. An environmental deviation (E) is added to the genotypic value of the individual to determine the phenotype (P) of the individual, that is,

$$P = G + E. \tag{1.1}$$

The characteristics of the infinitesimal model are described in Figure 1.3. The distribution of phenotypic values is determined by a normal distribution, whose phenotypic mean $\overline{P}$ is equal to the genotypic mean $\overline{G}$. The reason for this equality is that it is assumed that environmental deviations can increase or decrease the phenotype of the individual with equal probability and similar magnitude, that is, $\overline{E} = 0$. This assumption is fundamental when it comes to carrying out the partition of the genetic variance in its components, as we will see in Chapter 3. The genotypic values are then distributed as a normal distribution with mean $\overline{G}$ and genotypic variance V_G. Environmental deviations are also distributed as a normal distribution

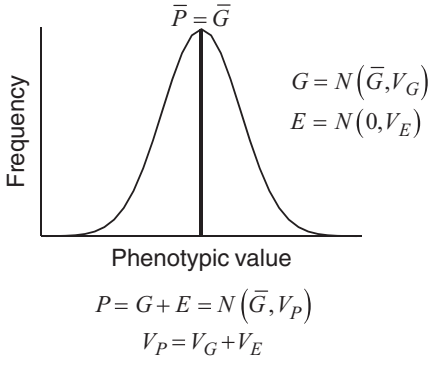


Figure 1.3 Characteristics of Fisher's infinitesimal model. The phenotypic value (P) is the result of the sum of the genotypic value (G) and the environmental deviation (E), which follows the normal distribution, so that the phenotypic variance V_P is the sum of the genotypic V_G and environmental V_E variances.

with mean 0 and environmental variance V_E. The phenotypic values are distributed, likewise, with normal distribution and phenotypic variance

$$V_P = V_G + V_E. \tag{1.2}$$

One of the fundamental objectives of quantitative genetics is to quantify the proportion of the phenotypic variation due to its genetic and environmental components, since the resemblance between relatives, the relative influence of environmental factors and the response to natural and artificial selection, will depend on this partition, as we will see in other chapters. On some occasions, the variance of the trait is scaled to the square of the mean (squared coefficient of variation, $CV_P^2 = V_P/\bar{P}^2$) to avoid scale effects and to be able to compare the magnitude of the variation between different quantitative traits.

The characteristics of the infinitesimal model are always violated in practice, since the number of genes is not infinite and their effects are not equal and infinitesimal. In fact, the experimental results indicate that most gene effects are of small magnitude with a lower proportion of loci with effects of great magnitude. In addition, quantitative trait loci are frequently subject to interactions, and their frequencies show dependence. However, the model serves as a predictive approach in a large number of scenarios. Later versions of the model include, in addition, dominance, the possibility of frequency dependence between loci (the so-called linkage disequilibrium) and interaction between effects (epistasis), circumstances that we will study in later chapters.

We have previously commented that quantitative traits can also be expressed discretely instead of continuously. Figure 1.4 illustrates two examples of such traits in *Drosophila melanogaster*. Panel (a) presents the number of sternopleural bristles in a large sample of individuals from a laboratory population. Note that the distribution is approximately normal with a slight asymmetry. Panel (b) shows the distribution of the number of pupae produced per female, a measure of fecundity, in the same population. This trait, like other main components of fitness, usually shows asymmetry towards low values due to the presence of deleterious genes of substantial effect that, although rare, can be carried by some individuals, positioning them very far from the average.

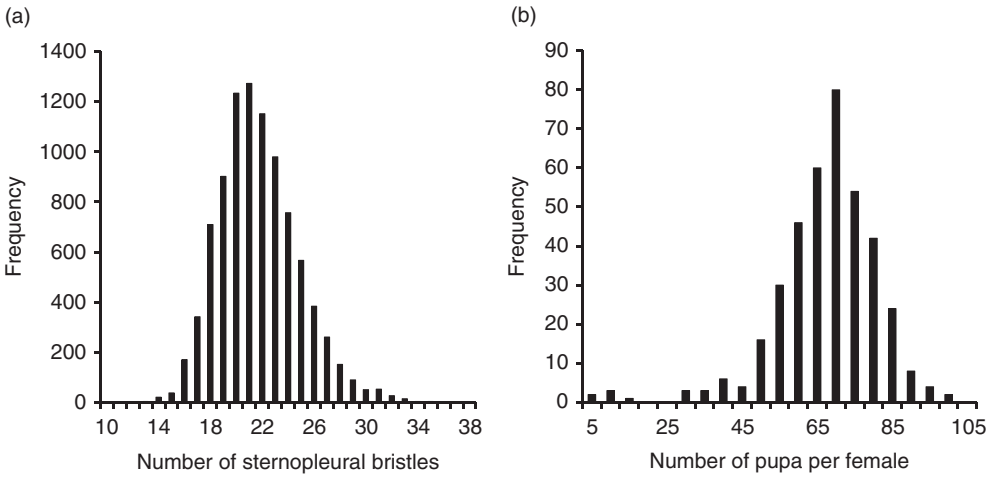


Figure 1.4 Distribution of two quantitative traits with discrete expression in a population of *Drosophila melanogaster* maintained in the laboratory with a large census size. (a) Number of sternopleural bristles (sum of both side plates). Vilas' (2014) data, corresponding to 9202 individuals (design of 1200 full-sib families with 10 individuals per family). The mean bristle number was $\overline{P} = 21.74$, and the variance $V_P = 9.80$. (b) Productivity per female measured as the number of pupae produced 11 days after mating with a male. Vilas' (2014) data corresponding to 388 females of the population. The mean pupae number was $\overline{P} = 64.35$, and the variance $V_P = 188.42$.

The most extreme case of quantitative traits with discrete expression are threshold traits, where only two or three phenotypic classes occur. A typical example is the susceptibility to a disease: individuals suffer it or not. The idea is that an underlying trait, which is called propensity, or liability in the context of human diseases, is a continuous trait determined perhaps by several or many loci and environmental factors, and that if a certain 'threshold' value of the liability is exceeded, the trait changes of phenotypic expression (Figure 1.5). This model can be applied to a large number of human diseases whose polygenic nature has been clearly demonstrated. The proportion or percentage of affected individuals is called, in general, incidence and, in the context of human diseases, prevalence, where incidence is the new number of cases in a given period of time.

Problems

1.1 In the cross between two pure lines, a heterozygous hybrid was obtained for 20 biallelic loci that affect a quantitative trait. For these loci, one allele has no effect on the trait and the other increases it by one unit. (a) How many genotypic classes would be found in the offspring by self-fertilization of the hybrid? (b) How often would descendants be found with a heterozygous genotype at all 20 loci? (c) With what probability would descendants with a phenotypic value equal to 10 be found?

1.2 The following table shows the number of sternopleural bristles in 50 individuals of *Drosophila melanogaster*. Knowing that the genetic variance of the trait is $V_G = 4$, deduce

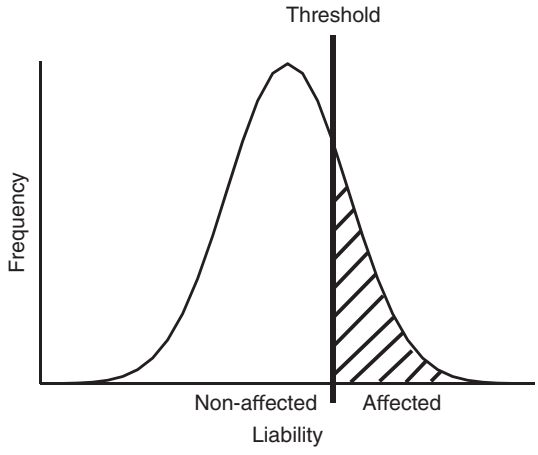


Figure 1.5 Model of threshold trait applicable, for example, to suffering a certain disease or not. There is an underlying continuous trait (liability in the context of human diseases) determined by the combined effect of a group of genes and environmental effects. Once a certain threshold of the underlying trait has been overcome, the disease is manifested.

the value of the environmental variance (V_E) and the phenotypic (CV_P) and genotypic (CV_G) coefficients of variation.

24	24	21	21	21	20	19	19	18	24
23	25	22	24	22	22	18	20	22	28
23	27	23	20	23	21	26	27	27	28
26	26	24	26	27	26	20	22	21	18
18	21	17	21	18	20	23	22	34	22

Self-Assessment Questions

1 Threshold traits are those determined by two or three loci.
2 Quantitative traits are affected by many loci and, therefore, are also called polygenic traits.
3 Additivity, or additive gene action, implies that the phenotypic value of the heterozygote is intermediate between that of the homozygotes.
4 Major genes are those that affect qualitative traits, while minor genes affect quantitative ones.
5 Threshold traits are a type of meristic trait.
6 When a locus has an effect on two or more traits, we speak of pleiotropy.
7 In the infinitesimal model, the phenotypic mean is not expected to be equal to the genotypic mean.
8 In the infinitesimal model, the loci have generally small effects, but there may be large-effect genes.
9 The main components of fitness generally present asymmetric phenotypic distributions.
10 The incidence in a threshold trait is the proportion of individuals affected by that trait.

2 Forces of Change in the Allele Frequencies

Concepts to Study

- Allele, gamete and genotype frequencies
- Hardy–Weinberg equilibrium
- Expected heterozygosity or gene diversity and allelic diversity
- Gametic or linkage disequilibrium
- Genetic drift
- The ideal population of Wright–Fisher
- Equilibrium between mutation and back-mutation
- Migration models
- Selection and dominance coefficients
- Types of within-locus gene action
- Stable and unstable equilibria
- Antagonistic pleiotropy and marginal overdominance

Objectives for Learning

- To learn how to calculate allele frequencies from genotype frequencies
- To know the conditions for Hardy–Weinberg equilibrium, its implications and the definition of expected heterozygosity and allelic diversity
- To know how to calculate linkage disequilibrium
- To understand the process of genetic drift
- To learn the basic characteristics of the Wright–Fisher idealized population
- To know how to calculate the changes in allele frequency by mutation and how an equilibrium between mutation and back-mutation is reached
- To understand the homogenizing effect of migration and the different population models used for its description
- To know the general model of fitness and the concepts of selection coefficient and dominance coefficient
- To understand how the changes in allele frequency for deleterious or beneficial alleles take place with different types of within-locus gene action
- To learn how to distinguish stable and unstable models of allele frequencies
- To comprehend the concepts of antagonistic pleiotropy and marginal overdominance
- To understand the impact of selection on the test for Hardy–Weinberg equilibrium

2.1 Allele, Gamete and Genotype Frequencies

The genetic description of a population can be done at three different levels, the locus, the gamete or the individual genotype, by specifying the different variants in each case

Table 2.1 *Illustration of the calculation of genotype frequencies and allele or gene frequencies*

Genotype	*AA*	*Aa*	*aa*	Total
No. individuals	40	50	10	100
Genotype frequency	0.4	0.5	0.1	1
Allele	*A*	*a*		
Number	130	70		200
Allele frequency	$p = 0.65$	$q = 0.35$		1

(allele, gamete or genotype) and their respective frequencies. A population of a diploid species is composed of individuals (genotypes) that reproduce by the union of their gametes to form zygotes that will give rise to the individuals of the next generation, hence the interest of a genotypic and gametic description. But genotypes and gametes are sets of alleles, two for each locus in the first case and one in the second, hence the interest of the allelic description.

Consider the simplest case: a biallelic locus A with alleles *A* and *a*, therefore, genotypes *AA*, *Aa* and *aa*, and suppose that in a population formed by 100 individuals, the number of those corresponding to each genotype is 40, 50 and 10, respectively (Table 2.1). Genotype frequencies are normally given in relative values, so that their sum is unity. From the genotype frequencies we obtain the allele frequencies, p that of *A* and $q = 1 - p$ that of *a*, which are also the gamete frequencies because the gametes carry a single copy of each of the chromosomes of the individual that produce them. Since homozygotes for one locus carry two copies of the relevant allele, and heterozygotes only one, allele frequencies can be obtained quickly as the sum of the frequency of the homozygotes carrying the allele in question plus half the frequency of the heterozygotes, that is, $p = 0.4 + 0.5/2 = 0.65$ and $q = 0.1 + 0.5/2 = 0.35$.

Suppose now that we consider two loci A and B, with alleles *A* and *a*, and *B* and *b*, respectively, there being nine possible genotypes formed by the combination of the three genotypes for each locus: *AABB*, *AaBB*, . . . *aabb*. The constitution of the four gametes that can exist will be *AB*, *Ab*, *aB* and *ab*, whose gametic frequencies we will denominate P_{AB}, P_{Ab}, P_{aB}, P_{ab}, respectively, and whose values will depend on the allele frequencies for each locus and the possible physical and/or genetic association between the two loci in question.

2.2 Hardy–Weinberg Equilibrium

In order to establish the genetic constitution of a filial population based on the genetic description of the parental population, it is necessary to specify the way in which the gametes produced by the parents are united in pairs to form the offspring. The simplest case is one in which gametic paring is random, called panmixia. If there were no differentiation in allele frequencies between sexes and we consider male and female gametes that carry the allele *A* with frequency p or the allele *a* with frequency q, random pairing generates the descendants shown in Figure 2.1, whose expected genotype frequencies are obtained developing the binomial expression $(p + q)^2$, that is,

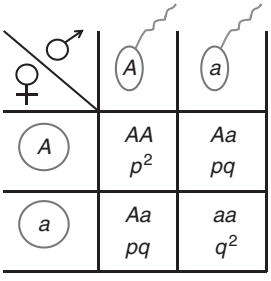


Figure 2.1 Calculation of the expected genotype frequencies from the allele or gametic frequencies in a scenario under panmixia. The frequency of the gamete carrying allele *A* is *p* and that of allele *a* is *q* in both sexes.

$$\begin{array}{lccc} \text{Genotypes:} & AA & Aa & aa \\ \text{Genotype frequencies:} & p^2 & 2pq & q^2. \end{array} \tag{2.1}$$

This is known as the Hardy–Weinberg principle or equilibrium, since it was independently deduced in 1908 by the English mathematician G. H. Hardy and the German physician W. Weinberg. Although very elementary, this principle is basic in population and quantitative genetics, since it greatly simplifies most of the theoretical developments formulated for the processes that affect populations. Note that random pairing is only required for the locus (or loci) under study. Although gametes do not unite in a strictly random manner, for most of the genome the assumption is usually valid. The possible discrepancy between the expected and observed genotype frequencies can provide information about the forces of change acting on the locus, as we will see later. In the example described in the previous section, the observed genotype frequencies were 0.4, 0.5 and 0.1 for the genotypes *AA*, *Aa* and *aa*, respectively, and the corresponding allele frequencies were $p = 0.65$ and $q = 0.35$. Therefore, the expected frequencies of the three genotypes in Hardy–Weinberg equilibrium would be $p^2 = 0.4225$, $2pq = 0.455$ and $q^2 = 0.1225$, which are similar to those observed, although to determine this it would be necessary to carry out a statistical test that will be illustrated in a problem solved at the end of the chapter.

The expected genotype frequencies in Hardy–Weinberg equilibrium for a locus are illustrated in Figure 2.2 as a function of allele frequencies. Note that the maximum frequency of heterozygotes is 0.5, and that when one allele is rare, most of its copies are found in heterozygous individuals. For example, if allele *A* were at a frequency of $p = 0.01$, the expected frequency of *AA* homozygotes could be only 0.0001, one in 10,000 individuals, while heterozygotes *Aa* would constitute approximately 0.2% of the population, or 2 out of 1000 individuals. This observation will be essential for some of the arguments that we will develop in later sections.

The Hardy–Weinberg principle can refer not only to a locus but to other elements, such as a chromosomal organization (for example, an inversion or a translocation). The extension to multiallelic loci is immediate. Suppose that locus A has *n* alleles, A_1, A_2, ... A_n, with frequencies p_1, p_2, ... p_n, respectively. The frequency of the homozygous