1

An Overview of Quantitative Genetics

Version 11 July 2022

Any trait or feature that can be measured and assigned a value can be considered a **quantitative trait**, and the analysis of such traits forms the field of **quantitative genetics** (**QG**). Historically, this included traits such as human height, agricultural production characters (such as grain and milk yield), and morphological features in natural populations of plants and animals. Binary (presence/absence) traits also fall within the quantitative genetics framework, most notably human disease risk. It was also realized that behavioral and physiological traits fall within this framework as well, in part as an extension from the analysis of psychiatric diseases, but also independently driven by the work of behavioral and evolutionary ecologists. More recently, it was appreciated that this same machinery naturally extends to any measurable *genomic feature*, such as mRNA levels and splicing patterns, methylation status, or the configuration of binding factors on chromatin.

All of the above mentioned features typically show variation, either within, or between, populations. For a **simple** trait, this variation is very largely due to the action of one or two Mendelian genes, whose genotypic signals overpowers most of the environmental (and genetic background) noise. The analysis of such simple traits is a special subset of quantitative genetics. However, most traits are **complex** in the sense that their variability is due to variation in expected trait values over a very large number of different genotypes, overlaid with considerable noise from environmental effects. Historically, quantitative genetics formed the foundation for plant and animal breeding, evolutionary and ecological genetics, and human genetics. In the last decade, QG has also become an integral part of the genomics and post-genomics revolutions. Further, the field is strongly tied to advances in statistical methodology, many of which were motivated by questions arising from the analysis of complex traits (Appendices 1–9).

For much of its early history, quantitative genetics machinery (while build upon a Mendelian foundation; Chapters 4–6) was rather agnostic as to the genetics underlying a focal trait, relying instead on observations of phenotypes in sets of relatives (Chapters 7, 11–13, 22–32). This aspect of QG was regarded as a strength, as no genetic details were needed in order to proceed with a wide range of very useful analyses and predictions. This avoidance of any fine genetic details was not a deliberate oversight, but rather an operational decision given the information (or lack thereof) available at the time. However, starting in the early 1980s, the field quickly embraced the additional information offered by the emerging availability of a modest number of molecular markers, then moved on to exploiting SNP information, and finally to using whole genome sequencing (Chapters 8, 17–21, 31, 32). More recently, as the field realized that much of trait variation was not the result of variation in protein sequences, but rather due to subtle variation in regulation, QG fully embraced emerging concepts from functional genomics, to the point where many QG analyses are as focused on upstream molecular features as they are on the downstream final trait value (Chapter 21).

As a result of this fusion of classical quantitative genetics, modern statistics, and genomic/post-genomic data, quantitative genetics today is more relevant than at any time in its history. QG is the ultimate Phoenix, having risen from (according to its critics) the ashes of irrelevancy numerous times, only to become stronger, more integrated, and more relevant. In this volume, we attempt to not just present the important machinery in the analysis of complex traits, but also guide the reader on QG's incredible journey, from its founding in 1918 by Fisher into the current field that is fully intermeshed with genetics, genomics, evolution, statistics, and bioinformatics.

We start the first chapter in this volume with a brief overview of the history, and interplay, between classical, quantitative, population, and molecular genetics. We then briefly overview the rest of the book by discussing the major problems of interest to practitioners of quantitative genetics. The focus of this volume is on the genetics and analysis of quantitative traits, while their evolution and selection (for example, various breeding schemes) is examined in great detail in our second volume (Walsh and Lynch 2018; henceforth WL). A final (third) volume will complete our trilogy, and largely deals with multivariate issues (such as a deep analysis of G x E and multitrait selection).

A SHORT HISTORY OF QUANTITATIVE GENETICS

From Darwin and Mendel to Fisher 1918

Modern biology rests on the twin pillars of evolution offered by Darwin in his *Origin of Species* (1859) and genetics postulated by Mendel in his *Experiments in Plant Hybridization* (1865). While these two ideas appeared almost contemptuously, their impact was felt over very different timescales. Darwin's work had an immediate impact in reshaping biology, but Mendel's work was ignored until independently rediscovered by multiple groups over 35 years later (DeVries, Correns, and von Tschermak, all in 1900). As a result, an unnatural schism (in part centered around different assumed models of inheritance) developed between these pillars that lasted for almost two decades, with quantitative genetics emerging, in part, as a resolution and union of these grand concepts.

In order to explain the approximate intermediacy of progeny phenotypes with respect to those of parents, Darwin (1859) assumed that continuous characters exhibited **blending inheritance**. This was the common model of inheritance at the time, and postulates that parents mix some kind of biological vital fluid to generate the phenotype of their offspring. Fleeming Jenkin (1867), in what was the first paper in population genetics, pointed out a very serious problem with this model of inheritance. Jenkin, Regius Professor of Engineering at the University of Edinburgh, was asked to write a book review of Darwin's *Origins*. In which he noted that, under blending inheritance, ***half of any variation is removed each generation***, requiring some force to generate sufficient new variation to offset this loss.

Phrased in terms of modern statistics, and recalling (Chapter 3) that the variance of a product of a constant (a) times a random variable (x) is $\sigma^2(ax) = a^2\sigma^2(x)$, his argument was that

$$\sigma^2(z_o) = \sigma^2\left(\frac{z_f + z_m}{2}\right) = \frac{\sigma^2(z_f)}{2^2} + \frac{\sigma^2(z_m)}{2^2} = \frac{\sigma^2(z)}{2}$$

where z_f and z_m are the (assumed to be uncorrelated) paternal and maternal trait values, z_o is the offspring value, and we assume that the trait variance is the same in both sexes, so that $\sigma^2(z_f) = \sigma^2(z_m) = \sigma^2(z)$. Thus, any change achieved via natural selection would quickly be diluted away by the assumed nature of inheritance (also see Bulmer 2004). Mendelian genetics provides the solution to this quandary of reduced variation following reproduction. The particulate nature of genes implies that the variation generated by the transmission of alternative alleles in the gametes of heterozygous parents (the **segregation variance**) completely restores the genetic variance in each generation (Chapter 8).

Francis Galton, a cousin of Darwin, pointed out another problem. When he plotted the mean heights of offspring (measured at adult age) against the average height of their parents (corrected for differences between the sexes), he obtained a linear relationship (Figure 1.1). However, the slope of the line (being less than one) indicated that offspring were, on average, *less exceptional* than their parents. Parents whose average height was below the mean tended to have offspring taller than themselves but still below the mean. Parents above the mean tended to have offspring shorter than themselves. Galton (1886) called this trend **regression toward mediocrity** (and hence the origin of the term **regression** for such an analysis) and argued that this trend would erode away any selective progress. He concluded

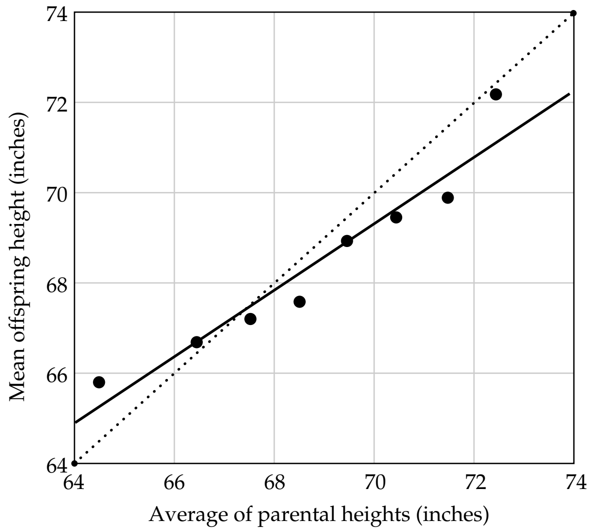

Figure 1.1 The relationship between height of adult children and the average height of their parents (in inches). Circles denote average offspring heights for different 1-inch classes of midparent heights. The best linear fit is given by the solid line, while the dotted line is the expected pattern if mean offspring height were equal to midparent height. (Data from Galton 1886.)

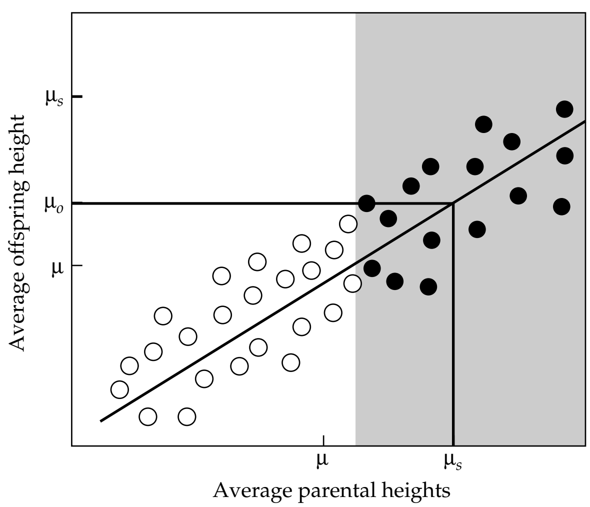

Figure 1.2 Response to selection. The solid line is Galton's regression. The closed circles represent selected parents. In the absence of selection, the mean height for both the parents and offspring is μ. However, if the average height of selected parents is μ_s, the expected height of offspring (μ_o), obtained by reading off the regression, is greater than μ but less than μ_s.

that evolution must be based upon **sports** (mutations of large effects) rather than on selection acting upon continuous variation. Galton's quantitative approach to the analysis of inheritance marked the founding of the **Biometrical school**, from which most of modern statistics can be traced (Stigler 1986, Crow 1993a).

Karl Pearson (1903) pointed out the fault in Galton's logic. Imagine that selection acts in such a way that only the largest parents reproduce, so that the mean height of reproductive adults after selection (μ_s) is greater than the mean height before selection (μ) (Figure 1.2). If μ_o is the average adult height of the offspring of the selected parents, then the **response to selection** across generations is $R = \mu_o - \mu$. Since the slope of the parent-offspring relationship is less than one, the adult offspring are not expected to be as tall (on average) as their selected parents. However, the mean offspring height will be greater than the mean height in the parental base population. Furthermore, the new mean height is stable—if selection on height is stopped, the new height is not expected to decay back to the original value. Thus, the regression toward mediocrity does not pose a serious problem for the theory of evolution by natural selection. Galton apparently failed to realize that selection starts from a new mean each generation and that it is to this new mean that the next generation of selection regresses. Although in disagreement with some of Galton's interpretations, Pearson was inspired greatly by Galton's quantitative approach to analysis and went on to develop many methods for the analysis of continuously distributed traits (such as regression and correlations).

Thus, at the time of Mendel's rediscovery in 1900, most of the focus on trait inheritance had been on *continuous* (those not clearly separable into discrete classes) traits (e.g., Galton 1886, 1889). In contrast, Mendelian genetics centered attention on the inheritance of *discrete* characters such as purple vs. white flower color, smooth vs. wrinkled seeds, and so on. Critically, Mendelian inheritance postulated that genes are *discrete particles*, and hence their heritable components can be recovered in future generations. Imagine crossing red and white parents to create a pink offspring. Under blending inheritance, pure red or pure white individuals could not be recovered as the blended fluid generating a pink offspring would not naturally resolve into its two separate fluid components. In contrast, if a pink offspring is the result of it carrying both a single red and a single white particle (the two alleles), these segregate cleanly in future generation, allowing for the potential of pure red, or pure white (as well as pink) offspring.

Rather than realizing the power of merging the quantitative analysis of inheritance with the particulate nature of transmission, a series of contentious debates instead ensued between the **Mendelians** (led by William Bateson) and the **Biometricians** (led by Pearson and W. F. R. Weldon). The major issues were whether discrete characters have the same hereditary and evolutionary properties as continuously varying characters. The Mendelians held that variation in discrete characters drove evolution through the appearance of new **macromutations** (mutations with large effects, akin to the appearance of sports suggested by Galton). In contrast, the biometricians thought that evolution occurred in very small steps by exploiting this continuously distributed variation. In essence, the Mendelians felt that small variation in continuous traits was not heritable, while the biometricians felt that it was. This schism between inheritance (Mendel) and evolution (Darwin) was largely driven by antagonistic personalities, rather than by facts, in the different camps (Provine 1971, 2000; Tabery 2004; but see Hogben 1974 for a different perspective), and significantly delayed the modern synthesis—the fusion of evolution by natural selection with Mendelian inheritance.

That continuous trait variation could have a discrete (Mendelian) underlying basis was suggested by the British mathematician G. Udny Yule, who gave formal proof for this idea in 1902. Unfortunately for Yule, the only thing that the Biometricians and the Mendelians could publicly agree on was the incompatibility of Mendelian genetics and the inheritance of continuous characters. The death of Weldon in 1906, followed by the publication of several key plant breeding experiments from 1908 to 1916, resulted in the rapid emergence of the **multiple-factor hypothesis**—continuous variation in a trait is generated, at least in part, by the impact of multiple underling genes.

George H. Shull, a major figure in American corn breeding, noted that self-fertilized corn strains were remarkably uniform in many continuous traits when compared to the outbred populations from which they are derived. His (Shull 1908) explanation (with modern terminology inserted and italicized) was that

> The obvious conclusion to be reached is that an ordinary cornfield is a series of very complex hybrids (*genotypes*) produced by the combination of numerous elementary species (*alleles*). Self-fertilization soon eliminates the hybrid elements (*removes the heterozygosity*) and reduces the strain to its elementary components (*each locus becomes homozygous, resulting in each inbred strain being composed of a single genotype*).

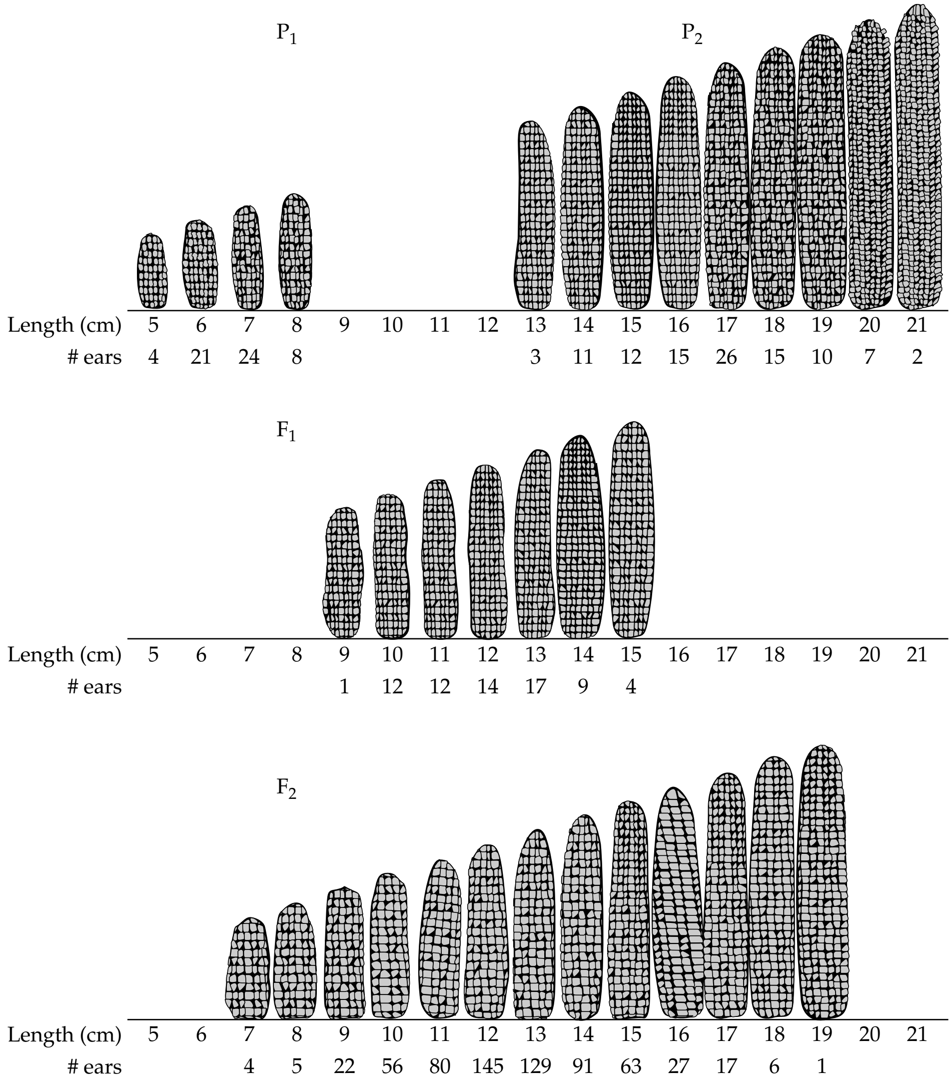

Figure 1.3 The distribution of ear size in the F_1 and F_2 generations formed by crossing two inbred lines of corn differing in ear length. The observed number of ears is given below each size class. The variation seen in the P_1, P_2 and F_1 populations is due entirely to environmental factors, as all individuals in each population have the same genotype. These three populations show roughly similar amounts of variation. In contrast, the F_2 generation shows considerably more variation, reflecting the diversity of genotypes in this population generated by segregation of genes in the F_1 parents. (Data from East 1911.)

The major implication of Shull's observation was that any variation for continuous characters that is lost upon inbreeding must have a Mendelian basis.

Additional support for the multiple-factor hypothesis came from H. Nilsson-Ehle (1909), a Swedish geneticist working with various cereal crops. Many of the characters that he examined yielded 3:1 ratios in the F_2 generation following the cross of two parental strains, consistent with expectations for a single segregating locus with one allele completely dominant over the other. However, there were some striking exceptions. For example, when red-seeded and white-seeded (allohexaploid) wheat strains were crossed, the F_1 progeny were identical in color and intermediate between the parents, but a diversity of colors, ranging from white to red, was seen in the F_2. Further, in some of the F_2 crosses, a ratio of 63 red:1

white seeds was observed. Nilsson- Ehle interpreted this to be the result of the segregation of three independent factors, the initial parents (both allohexapolids) being *AABBCC* and *aabbcc*, all members of the F_1 being *AaBbCc* and hence uniform in color, and the F_2 consisting of all possible genotypes, only one of which (*aabbcc*) gives rise to white seed. The probability of obtaining an *aabbcc* offspring from an *AaBbCc* $\times$ *AaBbCc* cross is $(1/2)^6 = 1/64$.

From these results, Nilsson-Ehle arrived at two general conclusions. First, ***sexual reproduction can produce a huge diversity of genotypes.*** For example, because a locus with two alleles A and *a* can produce three genotypes (*AA, Aa,* and *aa*), ten diallelic loci can produce $3^{10} \simeq 60,000$ genotypes. Second, given this huge potential diversity of genotypes, ***apparently new types appearing within a population may be the result of rare segregants rather than new mutations.*** Nilsson-Ehle, and independently East (1910), offered this as an explanation of **atavisms**—rare individuals that appear to be throwbacks to some previous population. Consider a hybrid population resulting from the cross of two pure lines differing at ten loci. The probability of obtaining a specific parental genotype in the F_2 or later generation is $(1/4)^{10}$. Hence, the population size has to be greater than $4^{10} \simeq 10^6$ for there to be much of a chance of observing a parental type. The common observation of **transgressive segregation** (Chapter 18), wherein some F_2 offspring can be more extreme than their parents, is also a natural outgrown of segregation over multiple loci. If parents are fixed for a mix of plus and minus alleles (increasing, and decreasing, trait values, respectively), then segregation can generate offspring with more more plus (or minus) alleles—and hence more extreme phenotypes—than found in either parent.

Subsequent studies quickly confirmed the ideas of Shull and Nilsson-Ehle. East (1911, 1916) and Emerson (1910; Emerson and East 1913) examined quantitative variation in a large number of plants. Typically, strains differing widely in some character were crossed and the variance of the resulting F_1 and F_2 generations recorded. In most of these crosses, especially when the parental populations were formed by repeated self-fertilizations, an **outbreak of variation** was seen in the F_2 (Figure 1.3). Such outbreaks of variation, resulting from the segregation of multiple genotypes from the F_1 heterozygotes, are consistent with the Mendelian model, but completely inconsistent with any blending hypothesis. An extensive historical review of the experimental verification of the multiple-factor hypothesis is given in Chapter 15 of Wright (1968).

The Danish botanist Wilhelm Johannsen (1903, 1909) was among the first to demonstrate that some of the variation in continuous characters was due to environmental rather than genetic causes. Starting from a common stock of beans, he produced several inbred lines. For each line, parental seeds of different weights were planted and the mean seed weight in the offspring measured. Johannsen observed that the variation *within a pure line* was not heritable—the mean seed weights of the offspring were essentially independent of the weights of the parental seed (Figure 1.4). In contrast to Galton's data (Figure 1.1), parent-offspring regressions *within lines* were flat (had slopes of zero). The lack of a within-line association in Johannsen's data arose because all parents within a line were genetically identical, in contrast to humans where parents were genetically different. To clarify the distinction between genetic and environmental effects, he coined the terms **genotype** (to denote genetically identical members of a pure line) and **phenotype** (the actual observed value for an individual—a compounding of genetic and environmental effects). From his observations, Johannsen concluded that natural selection could never move a character value beyond the level of variation seen in the original population. Like Galton, he felt that macromutations were essential for evolution.

However, Payne (1918) soon demonstrated that selection on *Drosophila* bristle number could result in flies with more extreme phenotypes than those seen in the base population (essentially transgressive segregation). Such a result has been observed in many selection programs involving economically important species of plants and animals: almost always, the range of observed variation underestimates the range of potential variation, often dramatically so (WL Chapters 24–26). By increasing the frequency of favored alleles, selection increases the probability of observing extreme genotypes. Recall from above that if the

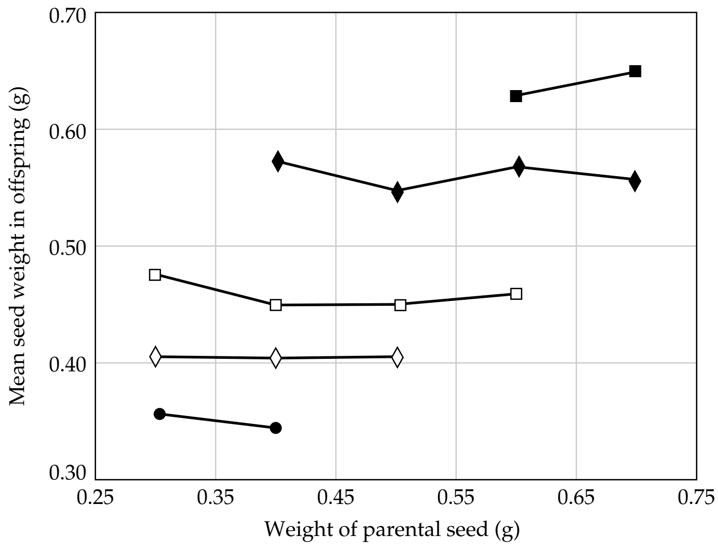

Figure 1.4 Mean offspring seed size as a function of parental seed size for some of Johannsen's pure lines. The data for the different lines are denoted by different symbols. If there is a heritable component to seed weight within a pure line, a line with positive slope is expected— larger parents should yield larger offspring. However, within each line, mean offspring size is essentially independent of the parental phenotype. (Data from Johannsen 1903.)

frequency of the favored allele at each of ten loci is 0.5, then the frequency of the most extreme genotype (a homozygote at all ten loci) is approximately one in a million. However, if selection advances the frequency of the favored allele at each locus to 0.9, the frequency of the most extreme genotype becomes $(0.9^2)^{10} \simeq 0.12$. In other words, about one in eight individuals would exhibit the most extreme genotype.

Shortly after the rediscovery, a new branch of quantitative biology started to be developed based on the implications of Mendelian segregation of discrete particles (alleles) in a population. Initially, some geneticists thought that an allele showing dominance would eventually take over a population. Hardy (1908) and Weinberg (1908) showed this was not the case, marking the formal beginnings of **population genetics**, the science of the behavior of alleles (and genotypes) under various evolutionary forces. Population genetics is highly intertwined with, but still distinct from, quantitative genetics. The former is built around allele and genotype frequencies, while the latter is concerned with the behavior of traits, which depends on *the effects of alleles* in addition to their frequencies. WL Chapters 2–11 gives a complete review of modern population genetics, and details how the merger of these two fields is now largely complete.

Fisher 1918: The Formal Foundation of the Quantitative Genetics

The growing consensus by 1916 in favor of the multiple-factor hypothesis set the stage for R. A. Fisher's (1918) brilliant paper, *On the correlation between relatives on the supposition of Mendelian inheritance*. This paper formed the conceptual bridge between Mendelian inheritance with discrete genes and evolution acting on continuous traits (ideas hinted at earlier by Yule 1902, 1906). It was also extremely important in the history of statistics. The term *variance* was first introduced in this paper, along with the method of ANOVA (the analysis of variance), that would lead to Fisher's development of the field of **experimental design** (Fisher 1935). Despite its enormous importance, Fisher wrote this paper as a high school teacher in 1916, and it was initially rejected by the Royal Society of London. As Crow (1972) succinctly stated, "It was apparently too mathematical for the Mendelists and too Mendelian for the biometricians." Norton and Pearson (1976) published the original reviews of Fisher's paper by Karl Pearson and R. C. Punnett, noting that formally Fisher withdrew the paper, rather than having it rejected. It was eventually published by the Royal Society of Edinburgh with the help of Leonard Darwin, Charles's son.

Fisher's paper was build around three key ideas. First, (diploid) parents pass along only part of their genotypic value to their offspring, as they only pass along single alleles, not diploid genotypes (Fisher's decomposition of the genotypic value, Chapter 4). Second, he showed that appropriate summary statistics (the additive genetic and other **variance components**) largely describe the resemblance between relatives (Chapter 7). Lastly, he showed how to estimate these variance components given phenotypic resemblance between known sets of relatives (Chapters 7, 22–30).

Fisher used a bit of a simplification for his work, the so-called **infinitesimal model**, which assumed a large number of genes, each of very small individual effect, underlay most trait variation. His notion of such a genetic model has been often, incorrectly, seized upon as proof that QG is a bit out of touch with reality. As we will see, the machinery of modern QG can easily handle any desired genetic structure.

1920–1950: The Biometrical Age of Quantitative Genetics

Fisher's brilliant demonstration that many short-term prediction problems could be complete described using only variance components lead to the rapid expansion of the age of **biometrical genetics**, using only phenotypic information to extract QG information. At nearly the same time as Fisher, Sewall Wright (1921a–1921d), using his method of path analysis (Appendix 2), independently developed many of Fisher's results, and further examined the implications of inbreeding. During the 1902s and 1930s, Fisher and Wright were joined by J. B. S. Haldane to lay the foundations of modern population genetics (Chapter 1 of WL).

In 1937, Jay Lush published his famous *Animal Breeding Plants*, helped greatly by weekly trips from Iowa State to the University of Chicago to attend Wright's course on quantitative genetics. Lush appears to be the first to formally state the **breeders equation**, $R = h^2 S$, predicting selection response R as the product of heredity h^2 and the selection differential $S = (\mu_s - \mu)$. This was the formalization, in quantitative-genetic terms, of Pearson's idea outlined in Figure 1.2. Quantitative genetics showed that the expected slope of the midparent-offspring regression was h^2, the ratio of the additive genetic variance to the total phenotypic variance. Lush's book ushered the power of quantitative genetics into the breeding world. Breeders, in turn, went on to develop much of the early machinery of QG in response to needs arising from specific applications.

The biometrical age saw the full fruition of ANOVA designs to estimate a variety of important, and useful, variance components. For example, the various **North Carolina designs** (Chapters 23 and 25) exploited special features of various plant breeding systems to cleanly estimate dominance variance in an attempt to more fully understand the basis of heterosis in maize. Additional line-cross methods were developed for insight into other features, such as the effective number of segregating factors (Chapter 11), inbreeding and heterosis (Chapters 12 and 13), and estimating the impact of G x E (Chapter 27).

Biometrical methodology was concerned with genetics only in the sense of estimating the appropriate genetic variance components, very much treating any other genetic details as a bit of a black box. This was not really a problem for the geneticist as the time, as the gene itself was also a black box. This changed with the rise of molecular biology, starting with the formal demonstration that DNA was the genetic material (Hershey and Chase 1952), and a model for its structure (Watson and Crick 1953).

1950–1980: The Rise of Molecular Genetics and its Diverence from Quantiative Genetics

In large part, the interaction between so-called classical genetics and quantitative genetics that played out during the rise of molecular genetics parallels much of the debate between the early Mendelians and biometricians, but with less rancor. During the development of molecular genetics in the 1950s (nicely reviewed by Judson 1979 and Mukherjee 2016), the notion of a gene grew from a fairly nebulous concept into a rather concrete, well understood, object build around a DNA sequence. This resulted in any statistical description of inheritance from quantitative genetics (variances of traits instead of following genes) becoming far less appealing. Indeed, to many geneticists, quantitative genetics appeared as an anachronism, a crutch that biologists no longer needed. As massive advances were made in the dissection of specific, individual genes with discrete, and highly reproducible, phenotypes, the result was a diminished interest in continuous traits. Indeed, Fisher's notion of an infinitesimal model—a large number of loci, each with very small effects—was anathema to this molecular way of thinking.

Thus, for three decades, the once harmonious field of genetics bifurcated into molecular and quantitative arms. During this time, biometrical approaches continued to be developed, cumulating with Charles Henderson's introduction (1949, 1950, 1963) of linear **mixed models** (BLUP) that elegantly extended the rigid ANOVA-based designs for variance component estimation to the completely general, and very complex, pedigrees that were common in animal breeding. It is greatly ironic that these models would be repurposed many decades later to play key roles in the genomic analysis of complex traits (Chapters 20 and 21).

1980–2000: Molecular Markers, QTL mapping, and the Age of Semi-major Genes

By the early 1980s, one of the byproducts of the molecular revolution was access to an ever-growing number of molecular-based markers. Geneticists, almost from the time of rediscovery, were well versed in using markers to map genes by exploiting the excess of parental gametes seen under linkage (Chapter 5). While a few **quantitative trait loci** (**QTLs**) were mapped using classical (phenotypic) markers (such as Sax's work in 1923 showing linkage in beans between loci for spotting and seed weight), the lack of widespread markers with trivial phenotypic effects prevented most attempts at QTL mapping. The bottleneck was technical (lack of markers), not operational (linkage-based mapping was well developed). In the early 1960's, the first molecular markers (allozymes, variants in the protein sequences produced by alternative alleles at a single gene) were developed, resulting in a modest set of markers (a few dozen) for many species. Starting a decade later, the development of DNA-based markers, such as RFLPs or STRs (Chapter 8), initally offered up to a few hundred markers. Their availability sparked widespread QTL mapping, especially in crops and laboratory species. As detailed in Chapters 17–19, linkage-based QTL mapping involves the isolation of small chromosomal regions that influence a trait of interest by applying linkage analysis to line-cross derivatives. Such studies routinely found relatively small ($\sim$ tens of megabases) chromosomal segments that appeared to account for nontrivial fractions of the trait variation (5% or more; Chapter 18).

The detection of such **semi-major** genes (or, more correctly, *genomic segments*) of modest to large effect was at odds with infinitesimal-like model thinking, and reinforced the view of many molecular geneticists on the importance of single genes (a resurrection of the Mendelian-Biometrican debate). The detection of such QTLs started to slightly open the genetical black box of quantitative traits, suggesting a road to redemption (at least from the molecular geneticists) whereby one would fairly rapidly be able to isolate a modest number of alleles that generated the bulk of trait variation. Alas, the actual isolation of such candidate genes proved extremely elusive (Chapters 17–20). Eventually, this led to the realization that the large estimated effects of such QTLs were very often the result substantially inflated effects for markers declared to be statistically significant (so-called **Beavis effects**; Chapter 18) and the presence of tightly linked clusters of trait-influencing genes within the detected region (the QTL) that fractionate upon finer mapping (Chapter 18).

2000 - present: SNPs, GWAS, Whole Genome Sequencing, and the Rise of Quantitative Genomics

The use of linkage-based approaches greatly limited QTL mapping resolution, which was set by the number of recombinants between markers within the population sample of gametes. One or two crossovers per chromosome was expected to be the norm under most linkage designs, resulting in blocks of nonrecombining material typically longer than tens of megabases. With many genomes on the scale of around 3000 megabases (such as humans), only 300 well-chosen markers could detect most of the recombination events. Increasing marker density quickly become redundant, as they would tend to be inherited as a block in the gamete sample due to insufficient number of recombinations.

Thus, the massively increasing number of usable DNA-based markers (such as **single nucleotide polymorphisms**, **SNPs**), that were starting to be generated by the late 1990s for model species (such as humans), could not be fully exploited by the limited number of recombinants in a typical linkage mapping population. New approaches would be required, as the mapping roadblock was no longer markers, but rather methodology, leading to the rise of **association mapping** (Chapters 17, 20, and 21). Today such **GWAS** (**genome-wide association studies**) are the standard fine-mapping approach in most settings.

The foundation for GWAS rests on the use of **linkage disequilibrium** (**LD**; Chapter 5)—correlations between closely-linked alleles at different loci—in random samples of unrelated individuals from a large population to fine-map genes. The reach of LD in a population is typically on the order from less than a kilobase (kb) to a few hundred kb, depending on the population (Chapter 5). LD thus offers much greater resolution than linkage mapping (the latter on the megabase scale). The GWAS logic, as developed in Chapters 5, 8, and 20, is that a random population sample of individuals contains very small blocks of very deep shared ancestry, generating the LD. Such blocks have persisted through hundreds of potential rounds of recombination, randomizing all but the tightest of linked markers.

DNA scoring technology initially used hybridization on DNA chips of up to a few million SNPs, and then moved onto using the scored SNPs to **impute** (predict) the genotypes of unscored, but very closely linked, SNPs (Chapter 20). Finally, **whole genome sequencing** (**WGS**) became affordable, at least on small scales. As we detail in Chapter 21, WGS provides very little gain in power when causal sites are in LD with **common SNPs** (a SNP whose minor allele frequency is greater than one to five percent). The power of WGS only shows itself when causal alleles are very rare, which in turn requires massive sample sizes. Fortunately, GWAS, being collections of unrelated individuals, is well suited to scale to very large sizes, potentially in the millions of individuals scored, each with millions of SNPs either directly scored or imputed.

As detailed in Chapter 21, GWAS " hits" typically only account for a very tiny fraction of trait variation. While at first blush this appears to be consistent with Fisher's model, there is actually a subtle difference. There is an inverse relationship between effect size and allele frequency such that large-effect alleles also tend to be very rare. Thus, a "small" GWAS hit could either be a common allele of very small effect or a very rare allele of very large effect. Even more disconcerting, most hits are in noncoding regions, typically rather far from genes (coding regions). Indeed, assigning a GWAS hit to a particular gene is far from trivial (Chapter 21). It appears that most trait variation is generated by regulatory variation, often acting at great distance from the target site. As a result, quantitative geneticists have started to turn their sights to GWAS mapping of genomic features (**quantitative genomics**), such as mRNA expression level, splicing isoform ratios, or methylation/chromatin configurations (Chapter 21). The idea is to leverage this information on upstream effects to both obtain better biological intuition of the processes generating trait variation, and also to more accurately isolate downstream targets.

Beside its obvious implications for mapping, dense marker technology also allowed classical approaches based on the fraction of shared ancestry to be greatly refined. For example, relative-based methods assume that two random full sibs each share (on average) one allele at a locus via common ancestry. However, there is considerable variation in the expected value, and the actual **realized relationships** can be accurately estimated using marker information (Chapters 8, 31, and 32).

Thus, we have come full circle, with the genomic (i.e., association) data providing strong support for a version of infinitesimal-like model for the genetic architecture of many traits. A quantitative-genetic framework is thus required when using association data for genomic prediction (the framework for genomic selection and genomic-based individualized medicine; Chapters 31 and 32). A quick glance at any of the front-line journals in genomics highlights this growing merger with quantitative genetics, as many functional-genomic features are now routinely examined in a quantitative-genetics framework. Perhaps there is no better exemplar for the journey of QG than human height. Galton's analysis sparked the rise of modern statistics, motivated Fisher's 1918 paper, and provided the framework for modern breeding. At the genomic level, at the time of this writing, the largest GWAS is a 5.4 million individual meta-analysis (Appendix 6) of human height, which found that over 12,000 sites account for height variation, with a potentially far larger number of much rarer sites waiting to be discovered (Example 21.15).

Quantitative Genetics, the Rise of Statistics, and the Current Statistical Revolution

As menioned, the impact of early quantitative-genetic theory extends well beyond the biological sciences. It laid the foundations for modern theoretical and applied statistics. Out of a need for quantitative methods to describe the distributions of continuously distributed characters, Galton provided the empirical motivation for Pearson's formal development of the theory of regression and correlation (Provine 1971, Stigler 1986). Fisher's 1918 paper introduced the concept of variance-component partitioning, upon which the principles of analysis of variance (ANOVA) are based, and his subsequent contributions had a profound influence on the development of methods for experimental design (Appendix 9) and hypothesis testing (Fisher 1925, 1935, 1956). Wright's (1921a) method of path analysis (Appendix 2) beaome broadly embraced by the social and ecological sciences.

As the field of QG continued to mature, more advanced methods arose from interactions between statisticians and quantitative geneticists. Starting with Henderson's work on mixed linear models, the power of random-effects models for many quantitative genetics problems became apparent. As computational advances allowed Fisher's method of maximum likelihood (Appendix 4) to be feasible in more general settings, it quickly gained prominence in quantitative genetics. The ML-based variance estimation method of REML (Chapter 32) arose from the concern of adjusting the estimated variance for the very large number of fixed effects typically fitted by animal breeders. REML later went on to be one of the workhorse methods in the genomic analysis of complex traits (Chapter 32).

More recently, the complex, high-dimensional nature of modern QG data sets, coupled with further computational and algorithmic advances (Appendix 8), led to the widespread use of Bayesian methods (Appendix 7). As data sets, especially with marker information, became both more complex and more numerous, the field embraced evolving statistical concepts such as false discovery rate (FDR) control for multiple comparisons (Appendix 6) and meta-analysis approaches for combining information from multiple data sets (Appendix 6). After the early sojourn of path analysis through the social sciences, it returned with a vengeance in quantitative genetics in the guise of **mediation analysis** to examine the impact of molecular intermediates on potentially downstream traits values (Chapter 21). Path analysis also birthed the important field of **Mendelian randomization** for controlling for confounding factors in complex epidemiological settings using quantitative genetic features (Appendix 2). Advances in QG continue drive new methods. For example, with whole gene sequences becoming more routinely used in QG analysis, new statistics on sparse, high-dimensional data will need to be developed. Thus, QG has been both a driver, and a grateful recipient, of advances in statistical theory and methodology. One final point is worst mentioning. While the impact of the genomics revolution has been obvious, what is less appreciated is that a nearly comparable revolution has occurred in modern statistics. In part, much of this new machinery arose from dealing the size, scale, and complexity of quantitative-genetic data sets.

THE MAJOR GOALS OF QUANTITATIVE GENETICS

Despite the great diversity fields using quantitative genetics machinery, most of the questions being examined revolve around a few key issues. Hence, before proceeding, we give a very brief summary of the main issues to be addressed in the remainder of the book:

The Nature of Quantitative-trait Variation

As we have just reviewed, the expression of quantitative characters is typically influenced by both genetic and environmental factors, with patterns of variation that are qualitatively consistent with Mendelian expectations. However, many questions remain. An obvious one, of interest from many perspectives, is, *How much of the standing variation in populations is due to genetic causes and how much to environmental ones*? From the standpoint of evolution and applied breeding programs, genetic components of variance are of particular interest because they determine the rates at which characters respond to artificial and natural selection. Environmental variance reduces the efficiency of the response to selection by causing the phenotypes of selected individuals to deviate from their underlying genotypic values. From the perspective of human genetic disorders, the degree to which the expression of undesirable traits is determined by genetic vs. environmental causes has broad implications for the development of preventative procedures and genetic counseling strategies.

Many methods exist for partitioning phenotypic variance into its various components (Chapters 22–32), all of which are based on the principle that the phenotypic resemblance between relatives provides information on the degree of genetic differentiation among individuals (Chapter 7). With a sufficiently dense set of molecular markers, we no longer need a *known* collections of relatives, as we can estimate relatedness within a population sample (Chapter 8). Provided this sample has sufficient variance in relatedness, we can apply standard QG methods to estimate variance components (Chapters 8, 31, and 32).

Despite the flexibility of relative-based estimation procedures, virtually all of the methods have some undesirable theoretical and methodological features. For example, the assumed additivity of G and E in the base model—which generates a simple dichotomy between genetic and environmental sources of variation—is often overly simplistic, as genotype $\times$ environment interaction may exist (Chapter 27). Examples include drug interactions with specific genotypes and local adaptation. In the latter, the performance of a line (or population) is disproportionately better (or worse) relative to other lines in a particular environment (i.e., G and E are no longer additive).

A further complication is that there are several components of both environmental and genetic variation (Chapters 4–6). The different components influence the resemblance between relatives to different degrees, and as a consequence, have substantially different influences on the evolutionary process. They also impact expressions for the joint disease risk occurrence over different sets of relatives. The additive component of the genetic variance (also known as the variance of breeding values) is of particular interest because it is the primary determinant of the degree to which offspring resemble their parents, which governs the rate of response of a character to selection. When significant nonadditive variance is present, breeders of species that can self (or be cloned) have a variety of options to exploit this other variance (WL Chapters 21–23). Similarly, when dominance is present, disease risk is greater than for a pair of full sibs than between parent and offspring.

Ultimately, the pool of genetic variation in a population must be due to a quasi-balance between the forces of selection and random genetic drift, both of which tend to eliminate variation, and the replenishing force of mutation (Chapters 15 and 21; WL Chapters 7 and 28). Recent work has shown that the mutational rate of production of new variation for quantitative traits is remarkably high (Chapter 15), in part due to a very large number of potential mutational target sites (Chapter 21). Further, different traits may have different **genetic architectures** (the underlying number of loci and the joint distribution of allele effects and frequencies at such loci). These architectures are functions of both the underlying molecular mechanisms generating trait variation *and* the past evolution history of their underlying genetic variants. For example, early-onset diseases likely face different evolutionary pressures than do diseases that appear after reproduction, potentially resulting in rather different genetic architectures. Likewise, a trait itself may be under no selection, but its underlying variants could still be due to pleiotropic effects on in other selected traits (Chapters 21 and 26; WL Chapter 28). The machinery from Chapters 18–21 is starting to provide insight into the size-frequency distribution, which, in turn, informs us somewhat about past evolutionary history.

The Consequences of Inbreeding and Outcrossing

The mean phenotypes of progeny from consanguineous matings commonly differ from those of progeny from random-bred parents. Such inbreeding effects are almost always deleterious (**inbreeding depression**), generally increasing linearly with the degree of relatedness between parents (Chapter 12). These observations are consistent with the presence of deleterious recessive alleles segregating at the loci underlying quantitative variation. The mechanisms and consequences of inbreeding have a number of practical implications, e.g., in the design of breeding programs for captive populations of endangered species, the maintenance of inbred lines for biomedical and agricultural research, and the nature of increased disease risk in small, closed human populations. Patterns of fitness decline with inbreeding can also provide insight into the rate and average effects of deleterious mutation (Chapter 12).

While crosses between relatives *within* a population can be deleterious, crosses *between* individuals from different populations often exhibit "hybrid vigor" in the F_1 generation, only to be followed by substantial fitness decline in the next (F_2) generation. The pattern of change in mean phenotypes in line crosses can yield insight into the mode of gene action, particularly with regard to interaction between genes at different loci (Chapters 11 and 13). In particular, dominance is a necessary, but not sufficient, condition for both inbreeding depression and hybrid vigor. Conversely, as the individuals being crossed come from increasingly divergent populations, **outbreeding depression** is often seen. Here, F_1 hybrids have reduced fitness, in the extreme, being sterile or inviable. Unlike hybrid vigor, epistasis is required for outbreeding depression, and understanding the genetics of this processes is vital to attempts to understand the mechanisms of speciation (Chapter 17).

The Constraints on the Evolutionary Process

Fundamental questions arise in evolutionary biology and in selective breeding programs as to the factors that limit the rate of phenotypic evolution. As noted above, when selection operates on a single trait, the response to selection is roughly proportional to the additive genetic variance for the trait (Chapter 3). However, if the same genes influence the expression of different traits, then an evolutionary change in one trait will necessarily lead to changes in the correlated traits. This can impede the breeding/evolutionary process when there is a conflict in the fitness consequences of selection operating directly on a trait and that operating on correlated traits. Questions of evolutionary trade-offs have long been the focus of evolutionary ecology, but many of the ideas in that field (e.g., life-history theory) have developed out of simple energetic arguments or comparative surveys on the *phenotypes* of different species. An unambiguous understanding of the constraints on the evolution of systems of complex characters requires information on the magnitude and direction of *genetic* correlations between characters. Quantitative-genetic methodology provides a powerful, but data demanding, means of elucidating these issues (Chapter 26). Ultimately, we would like to know the extent to which patterns of genetic correlations within populations are reflected in multivariate patterns of differentiation among species.

A related issue is the unresolved question about the nature of **pleiotropy** (one variant influencing multiple traits), the component (along with linkage disequilibrium) underlying genetic correlations. How common is this, and when it occurs, are its effects limited to a set of traits in within some developmental module, or are they more widespread? Does the effect size on one trait influence the extent of pleiotropy? For example, is a major mutation for one trait more likely to also influence additional traits? Resolving the pleiotropic structure of new mutations is critical for a better understanding the evolutionary forces that maintain quantitative genetic variation (WL Chapter 28).

The Estimation of Breeding Values and Predicting Phenotypes

Plant and animal breeding programs are based on choosing the best individuals to serve as parents for the next generation. In the case where offspring are clones of the selected parents (such as choosing among inbred individuals, and then selfing each to produce offspring), the goal is to choose the parents with the best genotypic (or line) values, as these are exactly what their offspring receive. In the case where offspring are produced by sexually crossing parents, offspring are a mixture of their parental genotypes. In this setting, the goal of breeders is to chose those parents with the best **breeding values** (Chapter 4), as the expected offspring mean from two parents is the average of their breeding values. Phenotypes are an imperfect predictor of either of these measures (line/genotypic and breeding values). The precision of these estimates can be improved by using information from relatives in a mixed-model framework (BLUP; Chapter 31). More recently, molecular marker information can be used to improve these estimates (**genomic selection**; Chapters 31 and 32). A related issue is precision medicine, the use of an individual's marker information to predict their future phenotype (the notion of a **polygenic risk score**).

Statistical and Mathematical Tools for Quantitative Genetics

A number of reviews of helpful quantitative tools are scattered throughout this book. While we have tried to concentrate these either where they needed, or summarize them in specific sections (such as the Appendixes), we also encourage the reader to use the very extensive subject index when perusing for information about particular tool.

Chapter 2 introduces basic properties of univariate distributions, such as moments (mean, variance, skew, kurtosis), as well as some less common distributions that arise in certain quantitative-genetics applications (such as the hypergeometric distribution for sampling without replacement and the negative binomial for the waiting time for exactly k events to occur). Bivariate distributions and their metrics (such as correlations, covariances, and regressions) are introduced in Chapter 3, which also examines conditional probability, and, in particular, the powerful theorem of Bayes. Appendix 1 develops the Taylor-series based **delta method** for obtaining approximations for means, variance, and covariances of more complex statistics (such as ratios or products). Appendix 2 introduces the method of path analysis for exploring causality, with an extended discussion on Mendelian Randomization approaches.

The most useful mathematical machinery in QG is linear (matrix) algebra, which is introduced in Chapter 9, with additional topics (such as generalized inverses) discussed in Appendix 3. Chapters 10, 31, and 32 build heavily on matrix machinery to present both the general linear model (GLM) and the mixed model (an extension of the GLM to include random effects). The latter is one of the main workhorse tools in QG, arising in GWAS (Chapter 20), G x E (Chapter 27), estimation of genetic and breeding values (Chapters 31 and 32), genomic selection and prediction (Chapter 31), meta-analysis (Appendix 6), and experimental design (Appendix 9), to name just a few. The scale on which a trait is measured impacts its analysis, and Chapter 14 discusses various topics in this area. Scales of analysis also serve as an introduction to generalized linear models—such as logistic regressions and log-linear models—which more easily handle discrete (i.e., zero/one or count) data than standard linear models. These are widely used in both GWAS (Chapter 20) and in the analysis of threshold traits (Chapter 30).

Chapter 16 introduces *mixture distributions* (not to be confused with *mixed models*), which are weighted sums of simpler distributions (such as a weighted sum of normals). These are commonly used in the search for major genes (Chapter 16) and in QTL mapping (Chapters 16 and 17). Fisher's method of Maximum likelihood (ML)—which provides a unified approach for estimation, assigning standard errors, and hypotheses testing—is widely used in QG, and is reviewed in Appendix 4. Bayesian methods (Appendix 7) are the extension of ML methods to more general settings. Their analysis uses straightforward, but computationally demanding, iterative approaches, such as MCMC, which a are reviewed in Appendix 8.

Table 1.1 Suggested topic-specifc pathways through this book.

Topic	Suggested Chapters
Basic quantitative genetics	2–7, 11–13, 22–24, 26, 27
Analytic tools	2, 3, 9, 10, 14, 30–32, Appendicies 1–9
Agricultrue and breeding	11–14, 17–20, 25–27, 31–32, Appendix 9
Evolutionary quantitative genetics	12–15, 21, 26–29, Appendicies 4–8
Human genetics	8–10, 12, 16-17, 19–21, 23, 24, 30–32, Appendices 2, 4–8
Genomic applications	8–10, 16–21, 31–32, Appendices 2, 4–8

Finally, calculation of statistical power is critical before any experiment. Appendix 5 examines power for simple t and z tests, as well as for both fixed and random effects ANOVA, general linear models, and for certain ML settings. Modern quantitative genetic datasets typically have a large number of tested parameters, raising issues with controlling for multiple testing. Various approaches for control (such as FDR and sequential Bonferroni) are reviewed in Appendix 6. This appendix also examines the related field of meta-analysis: how best to combine signals over a number of studies. Lastly, Appendix 9 reviews basic features in Fisher's theory of experimental design, focusing on analysis under heterogeneous environments.

NAVIGATING THE SCIENCE OF QUANTITATIVE GENETICS

The science of quantitative genetics has been around for over a hundred years, its main principles having been outlined independently by Ronald Fisher (1918) and Sewall Wright (1921a–d). It has served as the theoretical basis for most plant and animal breeding programs for almost a century (Lush 1937, 1945; Hanson and Robinson 1963; Turner and Young 1969; Namkoong 1979; Mayo 1980, 1987; Hallauer and Miranda 1981; Mather and Jinks 1982; Pirchner 1983; Henderson 1984a; Wricke and Weber 1986; Ollivier 1988; Hill and Mackay 1989; Gianola and Hammond 1990; Weller 1994; Bos and Caligari 1995; Falconer and Mackay 1996; Kearsey and Pooni 1996; Cameron 1997; Simm 1998; Kinghorn et al. 2000; Gallais 2003; Hallauer et al. 2010; Bernardo 2010, 2020). It has also played an important role in our understanding of the inheritance of complex human genetic disorders and underpins much of disease genomics (Lander and Schork 1994; Risch and Merikangas 1996; Jansen and Nap 2001; de Koning and Haley 2005; Gibson and Weir 2005; Rockman and Kruglyak 2006; Wellcome Trust Case Control Consortium 2007; Altschuler et al. 2008; Visscher et al. 2012, 2017; Boyle et al. 2017a). Finally, it has become also well entrenched in the field of evolutionary biology (Bulmer 1980; Lande 1988; Barton and Turelli 1989; Boake 1994; Walsh and Lynch 2018).

Hence, this book attempts to reach a very broad audience. As such, the reader might the detailed level of discussion on some topics to be of less interests that others. How, then, should one navigate this volume? Chapters 2–7 gives the basics that all QG practitioners should know. From here, paths can diverge dramatically given the interests and questions of the reader. Table 1.1 provides some guidance on how to proceed. We have tried to write chapters in such a way that their dependencies on specific previous Chapters are clear, but otherwise make them as independent as possible. Hence, we encourage the reader to feel free to skip around topics. Further, we have endeavored to develop a very detailed index to help the reader find hidden gems of interest on a particular topic that might appear in a somewhat out of the way location. Enjoy!

Literature Cited

Altschuler, D., M. Daly, and E. S. Lander. 2008. Genetic mapping in human disease. *Science* 322: 881–888. [1]

Barton, N. H., and M. Turelli. 1989. Evolutionary quantitative genetics: how little do we know? *Ann. Rev. Genetics* 23: 337–370. [1]

Bernardo, R. 2010. *Breeding for quantitative traits in plants,* 2nd Ed. Stemma Press, Woodbury, MN. [1]

Bernardo, R. 2020. *Breeding for quantitative traits in plants,* 3rd Ed. Stemma Press, Woodbury, MN. [1]

Boake, C. R. B. (ed.) 1994. *Quantitative genetic studies of behavioral evolution.* Univ. Chicago Press, Chicago. [1]

Bos, I., and P. Caligari. 1995. *Selection methods in plant breeding.* Chapman and Hall, London, UK. [1]

Boyle, E. A., Y. I. Li, and J. K. Pritchard. 2017a. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177–1186. [1]

Bulmer, M. G. 1980. *The mathematical theory of quantitative genetics.* Oxford Univ. Press, NY. [1]

Cameron, N. D. 1997. *Selection indices and prediction of genetic merit in animal breeding*. CAB International, Wallingford, UK. [1]

Crow, J. F. 1972. The origins of theoretical population genetics. *Amer. J. Hum. Genet.* 24: 608–609. [1]

Crow, J. F. 1993a. Francis Galton: count and measure, measure and count. *Genetics* 135: 1–4. [1]

Darwin, C. 1859. *The origin of species by means of natural selection.* Murray, London. [1]

de Koning, D.-J., and C. S. Haley. 2005. Genetical genomics in humans and model organisms. *Trends Genet.* 21: 377–381. [1]

East, E. M. 1910. A Mendelian interpretation of variation that is apparently continuous. *Am. Nat.* 44: 65–82. [1]

East, E. M. 1911. The genotype hypothesis and hybridization. *Am. Nat.* 45: 160–174. [1]

East, E. M. 1916. Studies on size inheritance in *Nicotiana. Genetics* 1: 164–176. [1]

Emerson, R. A. 1910. The inheritance of sizes and shapes in plants. *Am. Nat.* 44: 739–746. [1]

Emerson, R. A., and E. M. East. 1913. The inheritance of quantitative characters in maize. *Bull. Agric. Exp. Sta. Neb.* 2. [1]

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics.* 4th Ed. Longman Sci. and Tech., Harlow, UK. [1]

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Royal Soc. Edinburgh* 52: 399–433. [1]

Fisher, R. A. 1925. *Statistical methods for research workers.* Hafner, NY. [1]

Fisher, R. A. 1935. *The design of experiments.* 8th Ed. Hafner, NY. [1]

Fisher, R. A. 1956. *Statistical methods and scientific inference.* 13th Ed. Hafner, NY. [1]

Gallais, A. 2003. *Quantitative genetics and breeding methods in autopolyploid plants.* INRA, Paris, FR. [1]

Galton, F.. 1886. Regression towards mediocrity in hereditary stature. *J. Antropol. Inst. Great Britain and Ireland* 15: 246–263. [1]

Galton, F. 1889. *Natural inheritance.* Macmillan, London. [1]

Gianola, D., and K. Hammond (eds.) 1990. *Advances in statistical methods for genetic improvement of livestock.* Springer-Verlag, Berlin. [1]

Gibson, G., and B. Weir. 2005. The quantitative genetics of transcription. *Trend Genet.* 21: 616–623. [1]

Hallauer, A. R., and J. B. Miranda. 1981. *Quantitative genetics in maize breeding.* Iowa State Univ. Press, Ames. [1]

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho. 2010. *Quantitative genetics in maize breeding*, 3rd Ed. Springer-Verlag, New York, NY. [1]

Hanson, W. D., and H. F. Robinson (eds.) 1963. *Statistical genetics and plant breeding.* Natl. Acad. Sci., Natl. Res. Council Publ. 982. Washington, D. C. [1]

Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28: 49–50. [1]

Henderson, C. R. 1949. Estimates of changes in herd environment. *J. Dairy Sci.* 32: 706. [1]

Henderson, C. R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* 21: 309–310. [1]

Henderson, C. R. 1963. Selection index and the expected genetic advance *In* W. D. Hanson and H. F. Robinson (eds.), *Statistical genetics and plant breeding*, pp. 141–163. Natl. Acad. Sci., Natl. Res. Council Publ. No. 982, Washington, D. C. [1].

Henderson, C. R. 1984a. *Applications of linear models in animal breeding*. Univ. Guelph, Guelph, Ontario. [1]

Hershey, A., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36: 39–56. [1]

Hill, W. G., and T. F. C. Mackay (eds.) 1989. *Evolution and animal breeding.* CAB International, Wallingford, UK. [1]

Hogben, L. 1974. The origins of theoretical population genetics. *Brit. J. Hist. Sci.* 7: 176–179. [1]

Jansen, R. C. , and J.-P. Nap. 2001. Genetical genomics: the added value from segregation. *Trend Genet.* 17: 388–391. [1]

Jenkin, F. 1867. Origins of species. *North British Review* 46: 277–318. [1]

Johannsen, W. 1903. *Über Erblichkeit in Populationen und in Reinen Linien.* Gustav Fischer, Jena, Germany. [1]

Johannsen, W. 1909. *Elemente der exakten Erblichkeitslehre.* Gustav Fischer, Jena, Germany. [1]

Judson, H. F. 1979. *The eighth day of creation: Makers of the revolution in biology.* Simon and Schuster, New York, NY. [1]

Kearsey, M. J., and H. S. Pooni. 1996. *The genetical analysis of quantitative traits.* Chapman and Hall, London. [1]

Kinghorn, B. P., J. H .J. Van der Werf, and M. Ryan (Eds.) 2000. *Animal breeding: Use of new technologies.* The Post Graduate Foundation in Veterinarian Science of the University of Sydney, Australia. [1]

Lande, R. 1988. Quantitative genetics and evolutionary theory. *In* B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong (eds.), *Proceedings of the second international conference on quantitative genetics*, pp. 71–84. Sinauer Assoc., Sunderland, MA. [1]

Lander, E. S., and N. Schork. 1994. Genetic dissection of complex traits. *Science* 265: 2037–2048. [1]

Lush, J. L. 1937. *Animal breeding plans.* Iowa State Univ. Press, Ames. [1]

Lush, J. L. 1945. *Animal breeding plans, 3rd Edition.* Iowa State University Press, Ames, IA. [1]

Mather, K., and J. L. Jinks. 1982. *Biometrical genetics.* 3rd Ed. Chapman and Hall, NY. [1]

Mayo, O. 1980. *The theory of plant breeding.* Clarendon Press, Oxford, UK. [1]

Mayo, O. 1987. *The theory of plant breeding, 2nd Edition.* Clarendon, Oxford, UK. [1]

Mendel, G. 1865. Versuche uber pflanzen-hybriden. *Vorgelegt in de Sitzungen*. [1]

Mukherjee, S. 2016. *The gene: An intimate history*. Simon and Schuster, New York, NY. [1]

Namkoong, G. 1979. *Introduction to quantitative genetics in forestry.* Dept. of Agriculture Technical Bulletin No 1588. U. S. Governemnt Printing Office, Washington, DC. [1]

Nilsson-Ehle, H. 1909. Kreuzungsuntersuchungen an Hafer und Weizen. *Lunds Univ. Årsskrift,* n. s., series 2, vol. 5, no. 2: 1–122. [1]

Norton, B., and E. S. Pearson. 1976. A note on the background to, and refereeing of, R. A. Fisher;s 1928 paper 'On the correlation between relatives on the supposition of Mendelian inheritance' . *Notes Records Roy. Soc. Lond* 31: 151–162. [1]

Ollivier, L. 1988. Current principles and future prospects in selection of farm animals. *In* B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong (eds.), *Proceedings of the second international conference on quantitative genetics,* pp. 438–450. Sinauer Assocs., Sunderland, MA. [1]

Payne, F. 1918. The effect of artificial selection on bristle number in *Drosophila ampelophila* and its interpretation. *Proc. Natl. Acad. Sci. USA* 4: 55–58. [1]

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans. Royal Soc. Lond.* A 200: 1–66. [1]

Pirchner, F. 1983. *Population genetics in animal breeding.* 2nd Ed. Plenum, NY. [1]

Provine, W. B. 1971. *The origins of theoretical population genetics.* Univ. Chicago Press, Chicago. [1]

Provine, W. B. 2001. *The origins of theoretical population genetics: With a new afterword*. University of Chicago Press, Chicago, IL. [1]

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. *Science* 273: 1516–1517. [10]

Rockman, M. V., and L. Kruglyak. 2006. Genetics of global gene expression. *Nature* 7: 862–872. [1]

Simm, G. 1998. *Genetic improvement of cattle and sheep*. Farming press, Suffolk, UK. [1]

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8: 552–560. [1]

Shull, G. H. 1908. The composition of a field of maize. *Rpt. Am. Breed. Assoc.* 4: 296–301. [1]

Stigler, S. M. 1986. *The history of statistics.* Harvard Univ. Press, Cambridge, MA. [1]

Tabery, J. G. 2004. The "Evolutionary Synthesis" of George Udny Yule. *J. History Biol.* 37: 73–101. [1]

Turner, H. N., and S. S. Y. Young. 1969. *Quantitative genetics in sheep breeding*. Cornell University Press, Ithaca, NY. [1]

Visscher, P, M. A. Brown, M. I. McCarthy, and J. Yang. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90: 7–24. [1]

Visscher, P, and 6 others. 2017. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 9101: 5–22. [1]

Walsh, B., and M. Lynch. 2018. *Evolution and selection of quantitative traits*. Oxford University Press, Oxford, UK. [1]

Watson, J. D., and F. H. Crick. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171: 737–738. [1]

Weinberg, W. 1908. Ueber den Nachweis der Vererbung beim Menschen. *Jh. Ver. Vaterl. Naturk. Wurttemb.* 64: 368–382. [1]

Wellcome Trust Case Control Consortium (WTCCC). 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–676. [1]

Weller, J. I. 1994. *Economic aspects of animal breeding*. Chapman and Hall, London, UK. [1]

Wright, S. 1921a. Correlation and causation. *J. Agric. Res.* 20: 557–585. [1]

Wright, S. 1921b. Systems of mating. I. The biometric relations between parents and offspring. *Genetics* 6: 111–123. [1]

Wright, S. 1921c. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* 6: 111–123. [1]

Wright, S. 1921d. Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics* 6: 144–161. [1]

Wright, S. 1968. *Evolution and the genetics of populations. I. Genetic and biometric foundations.* Univ. Chicago Press, Chicago. [1]

Yule, G. U. 1902. Mendel's laws and their probable relations to intra-racial heredity. *New Phytologist* 1: 193-–207, 222—238. [1]

Yule, G. U. 1906. On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws—a preliminary note. *Report of the Conference on Genetics* 140–142. [1]