

Research Article

Open Access

# Toward a Unified Theoretical and Methodological Framework for Statistical and Quantitative Genetics in Molecular Breeding

Xuanjun Fang ✉

Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025, Hainan, China

✉ Corresponding email: [xuanjunfang@hitar.org](mailto:xuanjunfang@hitar.org)

Molecular Plant Breeding, 2026, Vol.17, No.1 doi: [10.5376/mpb.2026.17.0004](https://doi.org/10.5376/mpb.2026.17.0004)

Received: 15 Apr., 2026

Accepted: 25 Apr., 2026

Published: 30 Apr., 2026

**Copyright © 2026** Fang, This article was first published in Fenzi Zhiwu Yuzhong (Molecular Plant Breeding) in Chinese (24(9): 2817-2829), and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**

Fang X.J., 2026, Toward a unified theoretical and methodological framework for statistical and quantitative genetics in molecular breeding, Molecular Plant Breeding, 17(1): 50-76 (doi: [10.5376/mpb.2026.17.0004](https://doi.org/10.5376/mpb.2026.17.0004))

**Abstract** Advances in high-throughput sequencing and dense molecular markers have transformed the study of complex traits from phenotype- and pedigree-based inference to genome-scale, data-driven analysis. In this context, the relationship between statistical genetics and quantitative genetics has become increasingly important, yet conceptual ambiguity persists regarding their disciplinary roles. This study provides a systematic synthesis of their theoretical foundations, historical development, and conceptual distinctions.

Quantitative genetics is characterized as a problem- and theory-driven discipline focusing on the genetic architecture of complex traits and breeding strategies, whereas statistical genetics is defined as a methodology-driven field centered on model construction, inference, and analysis of high-dimensional genomic data. Through the examination of key paradigms such as QTL mapping, genome-wide association studies (GWAS), and genomic selection, we demonstrate that these two fields are not competing but highly complementary: quantitative genetics formulates biological questions and conceptual frameworks, while statistical genetics provides the inferential tools required to address them.

Building on this perspective, we propose an integrated framework based on the dimensions of “problem–method–data” and “theory–algorithm–application,” and further incorporate bioinformatics as a data-processing layer. This unified structure clarifies the roles of different disciplines within modern genetic research and highlights their coordinated interaction in the genomics era.

Finally, we discuss future directions in the context of molecular breeding, emphasizing the roles of multi-omics integration, artificial intelligence, and large-scale computation. We argue that deeper integration of theory and methodology is essential for improving the resolution and predictive power of complex trait analysis. This work provides a coherent conceptual framework for understanding the relationship between statistical and quantitative genetics and offers guidance for research design and interdisciplinary integration in modern genetics and breeding.

**Keywords** Statistical genetics; Quantitative genetics; QTL mapping; Genome-wide association studies (GWAS); Genomic selection; Molecular breeding; Multi-omics; Unified framework

## 1 Introduction

### 1.1 Background

The emergence of dense molecular markers and high-throughput sequencing has fundamentally reshaped how quantitative traits are studied in plants and other organisms. Classical quantitative genetics, rooted in Fisher’s reconciliation of Mendelian inheritance with continuous variation (Fisher, 1918) and codified in the work of Falconer and Mackay (1996) and Lynch and Walsh (1998), was largely developed in the absence of direct genotype information. Genetic architecture was inferred from phenotypic variation and pedigree-based relatedness, with concepts such as heritability, genetic correlations and the infinitesimal model providing a powerful, if abstract, description of complex traits (Hill, 2012; Walsh and Lynch, 2018).

The proliferation of molecular markers in the late twentieth century and, more recently, next-generation sequencing, enabled genome-wide characterization of variation at thousands to millions of loci. This transition has allowed quantitative-genetic questions to be addressed using QTL mapping, genome-wide association studies

(GWAS) and genomic prediction, turning previously latent genetic effects into observable marker–trait relationships (Lynch and Walsh, 1998; Hill, 2012; Jaganathan et al., 2020). Genomic selection, in particular, exploits dense marker maps to predict total genetic value using whole-genome regression models (Meuwissen et al., 2001; Spindel et al., 2015; Bhat et al., 2016). As a result, quantitative genetics has been extended from a phenotype- and pedigree-based discipline to a genomics-enabled science operating at genome-wide scale (Hill, 2012; Gienapp et al., 2017; Visscher and Goddard, 2019).

### 1.2 Problem statement

Alongside these methodological advances, there is growing conceptual confusion about the relationship between quantitative genetics and statistical genetics. Quantitative genetics is often described as the genetics of complex traits, emphasizing biological questions about how many loci, of what effect sizes and interactions, shape phenotypic variation and response to selection (Lynch and Walsh, 1998; Walsh and Lynch, 2018). Statistical genetics, by contrast, is typically framed as the development and application of statistical models and algorithms to infer genetic effects and structure from data, including GWAS, linkage analysis and genomic prediction (Posthuma et al., 2003; Laird and Lange, 2011; Schraiber et al., 2024).

In practice, however, the two labels are frequently used interchangeably, especially in the context of genome-wide analyses and big data. GWAS, genomic BLUP and Bayesian whole-genome regression are sometimes described as “modern quantitative genetics”, sometimes as “statistical genetics”, and sometimes as generic “genomic analysis” (Hill, 2012; Nelson et al., 2013; Serpico et al., 2023). This conflation obscures their overlapping yet distinct roles: quantitative genetics provides the conceptual and biological framework for understanding complex traits, whereas statistical genetics provides the inferential machinery that makes this framework operational with contemporary data (Posthuma et al., 2003; Hill, 2012; Schraiber et al., 2024).

### 1.3 Current controversies

This ambiguity manifests in several concrete controversies. First, disciplinary boundaries are blurred. Quantitative genetics has historically been tied to variance components, the animal model and prediction of response to selection, while statistical genetics evolved around linkage and association mapping and, later, GWAS and polygenic modeling (Lynch and Walsh, 1998; Posthuma et al., 2003; Laird and Lange, 2011). With the advent of genomic relationships and marker-based mixed models, these traditions have converged, making it unclear whether, for example, genomic heritability estimation or genomic prediction should be viewed as quantitative or statistical genetics—or both (Yang et al., 2011; Hill, 2012; Gienapp et al., 2017).

Second, there is persistent confusion between theoretical frameworks and statistical methodologies. Theoretical quantitative genetics, from Fisher’s variance decomposition through the infinitesimal model, provides a conceptual description of how many loci of small effect and environment combine to generate phenotypic distributions (Fisher, 1918; Lynch and Walsh, 1998; Walsh and Lynch, 2018). Statistical methods such as GWAS mixed models, REML estimation and Bayesian regression are tools for estimating parameters within that framework or for discovering loci consistent with it (Meuwissen et al., 2001; Yang et al., 2011; Visscher et al., 2017; Schraiber et al., 2024). Yet, as Serpico et al. (2023) emphasize, new genomic results are often interpreted using assumptions inherited from older quantitative-genetic models without making this distinction explicit, leading to conceptual slippage about causation, prediction and explanation.

Third, widely used methods such as QTL mapping and GWAS resist neat classification. QTL mapping was originally presented as a way to localize the factors underlying quantitative variation, clearly motivated by quantitative-genetic theory but implemented through likelihood-based and regression-based statistical models (Lynch and Walsh, 1998; Hill, 2012). GWAS similarly sits at the interface: it is a statistical-genetic technique whose outputs—effect sizes, variance explained, polygenicity—are interpreted in quantitative-genetic terms (Visscher et al., 2012; 2017; Hill, 2012). Disagreement over whether these approaches “belong” to quantitative or statistical genetics reflects deeper uncertainty about how to draw disciplinary lines in the genomics era.

## 1.4 Objective of this paper

This article responds to these tensions by taking a synthetic and critical perspective on the relationship between statistical and quantitative genetics, with particular attention to plant breeding in the era of genomic selection and big data. First, it aims to clarify conceptual definitions and boundaries by articulating quantitative genetics primarily as a theory-driven framework for complex traits and statistical genetics as a methodology-driven field focused on inference from genetic data (Lynch and Walsh, 1998; Posthuma et al., 2003; Hill, 2012; Schraiber et al., 2024). The goal is not to erect rigid disciplinary barriers, but to make explicit their different epistemological roles in modern genetics.

Second, the paper traces the historical co-evolution of the two fields from Fisher's 1918 foundation through the development of variance-component models and the animal model, the rise of QTL mapping and GWAS, and the advent of genomic prediction (Fisher, 1918; Lynch and Walsh, 1998; Meuwissen et al., 2001; Hill, 2012; Visscher and Goddard, 2019). This historical lens highlights the continuity of quantitative-genetic theory across technological eras and the increasing sophistication and partial independence of statistical-genetic methodologies.

Third, building on recent efforts to unify statistical traditions across genetics and evolutionary biology (Hill, 2012; Gienapp et al., 2017; Schraiber et al., 2024), the paper proposes an integrated conceptual framework for positioning key methods—such as QTL mapping, GWAS and genomic selection—at the interface between theory, data and inference. By making this structure explicit, the review aims to provide plant breeders and geneticists with clearer language and conceptual tools for navigating genomic-era methods, designing studies that are both statistically robust and biologically coherent, and situating new analytical developments within the broader landscape of quantitative and statistical genetics.

## 2 Concept and Methodological Framework of Statistical Genetics

### 2.1 Definition and disciplinary nature of statistical genetics

Statistical genetics can be understood as a methodology-driven discipline concerned with developing and applying statistical and computational methods to infer genetic mechanisms from high-dimensional genomic and phenotypic data. Its roots lie in early biometrical genetics and regression-based partitioning of variance (Fisher, 1918), but its contemporary identity is defined less by specific biological questions than by the inferential challenges posed by genome-scale data: multiple testing, population structure, cryptic relatedness, and complex dependency structures among markers (Hayes, 2013; Wang et al., 2019).

As such, statistical genetics is intrinsically interdisciplinary, drawing on probability theory, multivariate statistics, Bayesian inference, and machine learning, while being constrained by genetic principles such as linkage, segregation and quantitative-genetic models of inheritance (Moser et al., 2015; Evans et al., 2018). The field is also computationally intensive, relying on algorithm design and high-performance computing to scale methods to biobank- and breeding-program-level data (Weissbrod et al., 2019; Wang and Zhang, 2021). In plant breeding, statistical genetics provides the methodological backbone for genomic selection, GWAS, and multi-environment prediction, mediating between molecular data streams and breeding decisions (Xu et al., 2022; Montesinos-López et al., 2025).

### 2.2 Core research areas

Genome-wide association studies (GWAS) remain a central area, aiming to detect marker–trait associations across the genome while controlling for multiple testing and confounding. Standard GWAS pipelines encompass stringent quality control, single-marker or multi-marker tests, population structure correction, and post-GWAS fine-mapping and functional prioritization (Hayes, 2013; Wang et al., 2019; Weissbrod et al., 2019). Methodological advances include multi-locus models (e.g. MLM, FarmCPU, BLINK) and functionally informed fine-mapping frameworks such as PolyFun, which leverage annotations to improve power and localization (Weissbrod et al., 2019; Wang and Zhang, 2021).

Reliable GWAS and genomic prediction require accurate modeling of population structure and relatedness. Principal components, kinship matrices and model-based clustering are routinely used to capture ancestry and

familial relationships, reducing false positives and improving prediction (Hayes, 2013; Malle, 2022). Robust kinship estimators that remain valid under strong stratification exemplify the statistical-genetic focus on inference in realistic, heterogeneous populations (Manichaikul et al., 2010).

A second pillar concerns the genetic architecture of complex traits, particularly SNP-heritability and polygenic models. Whole-genome sequence and dense SNP arrays have motivated methods that estimate heritability from genome-wide markers, revealing pervasive polygenicity and sensitivity to assumptions about allele frequency and LD (Evans et al., 2018; Hou et al., 2019; Speed and Balding, 2019). Bayesian mixture models such as BayesR or BSLMM jointly address variant discovery, variance partitioning and prediction, offering a unified treatment of sparse large-effect and widespread small-effect loci (Moser et al., 2015).

In plant breeding, gene–environment interaction ( $G \times E$ ) has become a major research frontier for statistical genetics. Multi-environment trials, enviromics, and integrated genomic-enviromic prediction highlight that environmental covariates and  $G \times E$  terms can substantially increase prediction accuracy and reshape our understanding of adaptation (Technow et al., 2015; Cooper and Messina, 2021; Xu et al., 2022; Verbrigghe et al., 2025). These developments require models that accommodate heterogeneity of marker effects across environments and explicitly decompose main genetic and interaction components (Verbrigghe et al., 2025).

Finally, multi-omics integration extends statistical genetics beyond DNA polymorphisms to transcriptomic, metabolomic and phenomic layers. Multi-omics genomic prediction and association frameworks seek to exploit regulatory and metabolic intermediates to refine genotype–phenotype mapping, using Bayesian, kernel and machine-learning models tailored to heterogeneous data types (Amin et al., 2025; Montesinos-López et al., 2025). Emerging evidence indicates that model-based fusion strategies, rather than naive concatenation, are needed to consistently gain accuracy over genomic-only approaches (Montesinos-López et al., 2025).

### 2.3 Major methodological frameworks

Historically, statistical genetics built on classical regression and ANOVA, which underlie QTL mapping, diallel analysis, and early quantitative-genetic designs (Fisher, 1918). These linear models provide the basic language for fixed- and random-effect decomposition, hypothesis testing, and estimation of genetic parameters. Their limitations in capturing complex covariance structures and relatedness led to the central role of linear mixed models (LMMs), originally formalized for genetic evaluation in animal breeding (Henderson, 1975) and later adapted to GWAS and genomic prediction. LMMs model multiple random effects, absorb genome-wide markers into realized relationship matrices, and naturally accommodate unbalanced and multi-environment designs (Hayes, 2013; Yang et al., 2010). Their extensions, including compressed MLM, ECMLM and SUPER, have been crucial for scaling GWAS to large plant panels (Wang and Zhang, 2021).

Bayesian methods have become another mainstay, particularly for whole-genome regression and joint modeling of many markers. The Bayesian alphabet (BayesA, BayesB, BayesC $\pi$ , Bayesian LASSO) and subsequent developments such as BayesR or BSLMM allow flexible prior structures on marker effects, enabling simultaneous shrinkage and variable selection under diverse genetic architectures (Meuwissen et al., 2001; Habier et al., 2011; Moser et al., 2015). Bayesian frameworks also underpin recent GWAS variable-selection methods tailored to LMMs, such as BGWAS, designed to control false discoveries in high-LD, high-dimensional settings (Williams et al., 2023).

The rise of machine learning (ML) has further diversified methodological options. In genomic prediction, support vector regression, random forests, gradient boosting and a variety of neural network architectures have been compared with LMMs and Bayesian regressions (Montesinos-López et al., 2021; John et al., 2022; Jones et al., 2023). Comparative studies generally report that classical linear and Bayesian models remain competitive or superior for many traits, especially in moderate-sized datasets, although ML methods may offer advantages under complex non-linear architectures or when integrating multiple data modalities (John et al., 2022; Jones et al., 2023; Montesinos-López et al., 2025). This reinforces the view of statistical genetics as methodologically pluralistic, with model choice driven by the interplay between biological assumptions and statistical properties.

## 2.4 Methodological trends and future directions

The current trajectory of statistical genetics is shaped by three intertwined trends. First, high-dimensional data analysis has become unavoidable: the number of markers, omics features and environmental covariates routinely exceeds sample size by orders of magnitude. This has motivated work on robust heritability estimation and architecture inference that remain valid under complex LD and allele-frequency spectra (Evans et al., 2018; Hou et al., 2019; Speed and Balding, 2019), as well as on multivariate, multi-trait models that share information across correlated phenotypes (Moser et al., 2015; Amin et al., 2025). Challenges of multiple testing, overfitting and interpretability are magnified in multi-omics and enviromic settings, where feature selection, regularization and careful cross-validation are indispensable (Guyon and Elisseeff, 2003; Marx, 2013; Montesinos-López et al., 2025).

Second, large-scale computation and algorithm development are now integral to methodological innovation. Software ecosystems such as GCTA, GEMMA, GAPIT and KING exemplify how algorithmic advances—sparse matrix methods, iterative solvers, parallelization—enable routine analysis of millions of markers in large cohorts (Manichaikul et al., 2010; Yang et al., 2010; Hayes, 2013; Wang and Zhang, 2021). Biobank-scale analyses have spurred new estimators and fine-mapping tools that remain computationally tractable while accounting for polygenicity and functional annotations (Hou et al., 2019; Weissbrod et al., 2019). For plant breeding, efficient pipelines that jointly handle GWAS, genomic prediction and G×E modeling are increasingly essential for integrating statistical genetics into operational decision-making (Xu et al., 2022; Verbrugghe et al., 2025).

Third, the integration of AI and deep learning is redefining the scope of statistical genetics without replacing its core principles. Deep learning methods, including multilayer perceptrons and convolutional and recurrent networks, have been explored for genomic selection and trait prediction, often with performance similar to traditional models but with greater capacity to capture non-linear interactions when very large training sets are available (Montesinos-López et al., 2021; John et al., 2022). Their particular promise lies in multi-omics and enviromic contexts, where hierarchical feature extraction and representation learning can model complex genotype–environment–phenotype relationships (Xu et al., 2022; Amin et al., 2025; Montesinos-López et al., 2025). At the same time, issues of data requirements, model tuning, and interpretability highlight the continued need for rigorous statistical thinking and for hybrid frameworks that blend mechanistic quantitative-genetic structures with flexible AI components (Technow et al., 2015; Jones et al., 2023; Amin et al., 2025).

Overall, statistical genetics has evolved from a supporting role in quantitative genetics to a central methodological hub in modern genetics and breeding. Its future development will depend on reconciling statistical rigour with algorithmic scalability, and on embedding increasingly sophisticated models within biologically coherent frameworks capable of exploiting the full depth of genomic, phenomic and environmental information.

## 3 Theoretical Framework and Development of Quantitative Genetics

### 3.1 Classical foundations of quantitative genetics

Quantitative genetics was originally conceived to resolve the apparent conflict between Mendelian inheritance of discrete factors and the continuous variation observed for most traits in natural and breeding populations. Fisher's 1918 paper established the polygenic theory of inheritance by showing that the joint segregation of many loci with small effects, combined with environmental variation, could generate approximately normal phenotypic distributions (Fisher, 1918). This “infinitesimal model” became the cornerstone of complex trait analysis, providing a statistical–genetic bridge between genotype and phenotype.

Within this framework, phenotypic variance is partitioned into components attributable to genetic and non-genetic causes. The genetic variance itself is further decomposed into additive, dominance and epistatic components, reflecting the linear contribution of alleles, interactions between alleles at a locus, and interactions among loci, respectively (Falconer and Mackay, 1996; Lynch and Walsh, 1998). Subsequent theoretical work clarified how these components depend jointly on gene action and allele frequency, and how epistasis at the level of genotypic values can be expressed through additive and non-additive variance components at the population level (Cheverud and Routman, 1995; Hill et al., 2008; Viana and Garcia, 2021).

Heritability and breeding value emerge as key derived concepts. Narrow-sense heritability quantifies the proportion of phenotypic variance attributable to additive genetic variance, the component most directly exploitable by selection (Falconer and Mackay, 1996). Henderson's formulation of best linear unbiased prediction (BLUP) embedded these ideas in a mixed-model framework, enabling prediction of individual breeding values by combining phenotypes and covariance structures derived from pedigrees (Henderson, 1975). The resulting "animal model" formalized the use of relationship matrices to propagate information across relatives and remains the template for modern genomic prediction (Calus, 2010; Crossa et al., 2010).

Collectively, polygenic inheritance, variance component partitioning, and BLUP-based breeding value prediction established a coherent theory for complex traits: populations are characterized by their genetic (co)variances, and selection response can be forecast from additive variance and selection intensity without explicit knowledge of causal loci.

### **3.2 Core research questions in quantitative genetics**

Against this theoretical background, quantitative genetics has been driven by three interrelated questions that connect abstract theory to breeding practice.

First, what is the genetic architecture of complex traits? This encompasses the number and effect size distribution of loci, the prevalence of pleiotropy and epistasis, and the contribution of dominance and gene–environment interaction. Classical theory allowed inference about the relative magnitudes of additive, dominance and epistatic variances from resemblance among relatives, but not direct identification of underlying loci. Contemporary syntheses show that, in many species and traits, most genetic variance is effectively additive even when biological interactions are pervasive (Hill et al., 2008; Mackay, 2014; Viana and Garcia, 2021). Recent large-scale analyses in humans similarly find that additive variance explained by common variants dominates, with little evidence for substantial dominance variance and currently imprecise estimates of epistasis (Hivert et al., 2021). These results reinforce the classical insight that epistatic gene action can manifest largely as additive variance at the population level, explaining why simple additive models often perform well for prediction even when molecular networks are highly non-linear (Mackay, 2014; Mackay and Anholt, 2024).

Second, how can genetic effects be estimated from data? Quantitative genetics approaches this problem through inference on variance components and prediction of breeding values rather than direct estimation of individual locus effects. Mixed models using pedigree-derived relationship matrices became standard tools for estimating additive variances and predicting breeding values in animal and plant breeding (Henderson, 1975; Crossa et al., 2010). Conceptually, this shifts focus from individual genes to aggregate genetic contributions captured through covariance among relatives, emphasizing prediction and response to selection over mechanistic dissection.

Third, how should breeding strategies be optimized given the inferred architecture and estimated effects? Classical selection theory addresses the design of mating schemes, management of inbreeding, and allocation of selection intensity across traits and generations, all expressed in terms of expected genetic gain per unit time and risk (Falconer and Mackay, 1996; Lynch and Walsh, 1998). The assumption that most usable variation is additive underpins the success of simple index and BLUP-based selection strategies in many breeding programs. At the same time, recognition that epistasis and dominance can influence long-term response, heterosis and inbreeding depression motivates continued interest in multi-trait, multi-environment and non-additive models (Mackay, 2014; Viana and Garcia, 2021).

These core questions—architecture, estimation, optimization—anchor quantitative genetics as a theory that is simultaneously predictive and strategic, directly informing how breeders exploit complex variation.

### **3.3 Development in the molecular/genomic era**

The arrival of molecular markers and high-throughput genotyping transformed how these longstanding questions are addressed, but did not displace the underlying theoretical core. Instead, quantitative genetics evolved from phenotype- and pedigree-based inference to genome-based analysis, with the same variance-partitioning and prediction logic now expressed in terms of marker data.

### 3.3.1 QTL theory and mapping

The pivotal contribution of Lander and Botstein (1989) was to show how classical linkage analysis could be extended to systematically map quantitative trait loci (QTL) using dense restriction fragment length polymorphism (RFLP) maps. Their interval mapping framework adapted human LOD-score methods to experimental crosses, enabling simultaneous estimation of QTL position and effect and providing power calculations for experimental design (Lander and Botstein, 1989). Shortly thereafter, Paterson et al. (1988) demonstrated, in tomato, that fruit quality traits could indeed be resolved into multiple Mendelian factors using a complete RFLP map, empirically validating Fisher's polygenic hypothesis in a crop species.

QTL mapping thus instantiated quantitative-genetic concepts—additive and dominance effects, variance explained by loci—at the level of specific genomic regions. Subsequent methodological extensions introduced composite and multiple-QTL models, mixed-model frameworks and multi-environment designs, but retained the same basic aim: to dissect complex traits into contributing loci while leveraging the variance component perspective developed in classical theory (Doerge et al., 1997; Mathews et al., 2008). Association mapping and, later, genome-wide association studies (GWAS) generalized these ideas to diverse populations, using historical recombination and linkage disequilibrium to localize QTL at higher resolution (Lander and Botstein, 1989; Mackay and Anholt, 2024). In both linkage and association designs, quantitative genetics supplies the conceptual language—effect sizes, variance explained, epistasis—while statistical genetics provides the inferential machinery.

### 3.3.2 Genomic selection

A more radical step came with the proposal of genomic selection (GS), which explicitly shifted emphasis from detecting individual loci to predicting breeding value from genome-wide markers (Meuwissen et al., 2001). The key idea is deeply quantitative-genetic: if markers are dense enough that every QTL is in linkage disequilibrium with at least one marker, then a regression of phenotypes on all markers can, in aggregate, capture most additive genetic variance, even when individual marker effects are small. Meuwissen et al. (2001) evaluated whole-genome regression models under various architectures and showed that correlations between true and genomic breeding values could reach  $\sim 0.85$ , high enough to support selection based purely on marker information.

This conceptual shift replaces the search for statistically significant QTL with prediction of genomic estimated breeding values (GEBVs), directly serving the core breeding objective of optimizing selection response. Subsequent work in livestock and crop species has confirmed that GS can substantially increase genetic gain per unit time and reconfigure the role of phenotyping, which becomes primarily a means to update prediction models (Heffner et al., 2009; Calus, 2010; Crossa et al., 2010; Hayes and Goddard, 2010; Spindel et al., 2015; Yanez et al., 2022). The mixed-model and Bayesian GS methods used in practice are natural extensions of Henderson's BLUP, with the pedigree relationship matrix replaced or augmented by genomic relationship matrices constructed from markers (Legarra et al., 2009; Calus, 2010).

### 3.3.3 Expansion of polygenic models

The genomic era has also driven an expansion and refinement of polygenic models. Genome-wide SNP data allow direct construction of genomic relationship matrices and high-dimensional regressions with tens or hundreds of thousands of predictors. Ridge-regression BLUP (RR-BLUP), Bayesian whole-genome regressions, penalized regression methods such as LASSO and elastic net, and extensions like MultiBLUP all aim to operationalize the infinitesimal idea under high dimensionality, while accommodating heterogeneous effect distributions across the genome (Li and Sillanpää, 2012; Speed and Balding, 2014; Meuwissen et al., 2021).

At the same time, theory and empirical work have revisited the role of non-additive effects in this new setting. Studies combining marker data with variance-component models suggest that, even in the genomic era, additive variance remains the predominant contributor to complex trait variation, although epistatic interactions can be biologically widespread and important for long-term evolution and network properties (Mackay, 2014; Hivert et al., 2021; Viana and Garcia, 2021; Mackay and Anholt, 2024). Genomic models that incorporate dominance and epistasis—through additional relationship matrices, interaction kernels or non-linear machine-learning

approaches—seek to capture these components when they are relevant for prediction or for understanding heterosis and genotype-by-environment interaction (Spindel et al., 2015; van Eeuwijk et al., 2019).

Thus, quantitative genetics has transitioned from phenotype- and pedigree-based inference to genome-based analysis without abandoning its theoretical core. Polygenic inheritance, variance component partitioning and breeding value prediction continue to provide the conceptual scaffold on which modern QTL mapping, GWAS and genomic selection are built, even as statistical implementations have grown more complex and computationally intensive.

## **4 Relationship Between Statistical Genetics and Quantitative Genetics**

### **4.1 Conceptual and paradigm differences**

Quantitative genetics and statistical genetics are best viewed as overlapping but epistemologically distinct enterprises. Quantitative genetics is fundamentally theory-driven: it formulates models of how many loci, each of small effect, together with environment, shape the distribution of complex traits, and then uses parameters such as heritability, genetic correlations and the G-matrix to describe evolutionary and breeding responses (Fisher, 1918; Lynch and Walsh, 1998; Hill, 2010; Walsh and Lynch, 2018). In this view, the primary questions are biological: what is the evolutionary potential of a population, how do selection and drift act on quantitative traits, and how do genetic covariances constrain trajectories (Mitchell-Olds and Rutledge, 1986; Stepan et al., 2002; Gienapp et al., 2017).

Statistical genetics, by contrast, is methodology-driven. Its central focus is the development of inferential and computational tools to learn about genetic effects and structure from data, rather than the formulation of new biological principles (Doerge, 2002; Schraiber et al., 2024). GWAS, QTL mapping, fine-mapping, and genomic prediction pipelines are framed as solutions to problems of variable selection, multiple testing, confounding by relatedness, and high-dimensional regression (Doerge, 2002; Hill, 2012; Qi et al., 2024). The field is therefore anchored in statistical and algorithmic criteria—bias, variance, false discovery rate, computational scalability—while drawing constraints from quantitative-genetic theory.

The two traditions also differ in their primary data dependencies. Classical quantitative genetics operated largely with phenotypes and pedigrees, estimating relatedness from recorded ancestry and inferring genetic parameters without direct observation of loci (Falconer and Mackay, 1996; Lynch and Walsh, 1998; Hill, 2010). Statistical genetics, in its contemporary form, is inseparable from dense molecular markers and genomic “big data”, extending similar variance-component and regression frameworks to settings with millions of SNPs, multi-omics layers and large cohorts (Doerge, 2002; Hill, 2012; Bazakos et al., 2017; Qi et al., 2024). Yet this contrast is not absolute: modern “genomic quantitative genetics” now uses realized genomic relationships to estimate classical parameters, blurring the earlier phenotype/pedigree versus genomic divide (Hill, 2012; Gienapp et al., 2017).

The persistence of confusion in the literature reflects this blending. Quantitative genetics has always relied on statistical models, leading some to equate it with its methods and to downplay its role as a conceptual bridge between genotype and phenotype (Hill, 2010; Serpico et al., 2023). Conversely, statistical genetics inheriting many of its core models directly from quantitative genetics—most obviously the animal model and mixed-model frameworks—can appear to be merely the “applied arm” of quantitative genetics rather than a distinct, methodologically oriented discipline (Hadfield and Nakagawa, 2010; Schraiber et al., 2024).

### **4.2 Historical evolution of the two fields**

#### **4.2.1 Classical era**

The classical era of quantitative genetics began with Fisher’s reconciliation of Mendelian inheritance and biometrical variation, establishing the variance-component framework that still underpins modern analyses (Fisher, 1918). Subsequent formalization by Falconer and later Falconer and Mackay (1996) and the comprehensive synthesis by Lynch and Walsh (1998) codified quantitative genetics as the study of complex traits using phenotypic records and pedigree structures. Key objects such as heritability, additive and non-additive variance,

and the G-matrix emerged as central descriptors of genetic architecture and constraints on evolution (Mitchell-Olds and Rutledge, 1986; Stepan et al., 2002; Walsh and Lynch, 2018).

In this period, what would later be called statistical genetics was largely embedded within quantitative genetics as “biometrical methods”: regression, ANOVA, and covariance analysis. The conceptual asymmetry was clear: statistical tools served the quantitative-genetic theory of inheritance and response to selection.

#### 4.2.2 Molecular marker era

The molecular marker era introduced RFLPs, microsatellites and later SNPs, and with them the possibility of mapping QTLs rather than inferring architecture solely from variances. Lander and Botstein (1989) provided the seminal interval-mapping framework, explicitly formulating QTL detection as a likelihood-based statistical problem. Subsequent developments integrated multiple markers, maximum likelihood, and permutation procedures, marking a distinct methodological turn (Doerge, 2002).

Lynch and Walsh (1998) placed QTL mapping squarely within quantitative genetics, treating it as an extension of variance decomposition to marker-segregating populations. Yet as methods became more sophisticated—modeling epistasis, multiple QTL, and complex designs—the domain increasingly overlapped with the emerging identity of statistical genetics, which responded to computational and inferential challenges of these analyses (Leal, 2001; Doerge, 2002).

#### 4.2.3 Genomic era

The genomic era, with dense SNP arrays and affordable sequencing, transformed both disciplines. Genome-wide association studies generalized QTL mapping to diverse populations, while genomic selection used all markers to predict genetic values without necessarily identifying causal loci (Meuwissen et al., 2001; Hill, 2012; Bazakos et al., 2017). Mixed linear models with marker-based kinship matrices, REML estimation and Bayesian whole-genome regressions became standard (Yang et al., 2011; Hill, 2012; Moser et al., 2015).

Here, the continuity of quantitative-genetic theory is striking: GWAS mixed models and genomic BLUP are direct implementations of the animal model using genomic rather than pedigree relationships (Yang et al., 2011; Hill, 2012; Gienapp et al., 2017; Schraiber et al., 2024). At the same time, the methodological complexity exploded, motivating a distinct statistical-genetic literature addressing population structure, LD, high-dimensionality and functional fine-mapping (Doerge, 2002; Qi et al., 2024). Recent work unifying GWAS, animal models and phylogenetic regression under a single covariance-based quantitative-genetic framework illustrates both the deep conceptual continuity and the diversification of statistical traditions around it (Hadfield and Nakagawa, 2010; Schraiber et al., 2024).

### 4.3 Interaction and integration mechanisms

The co-evolution of the two fields can be understood as a cycle in which quantitative genetics formulates biological problems and conceptual quantities, while statistical genetics provides tools to estimate and test them. Quantitative genetics posits that additive variance and the G-matrix govern short-term evolvability and correlated responses (Mitchell-Olds and Rutledge, 1986; Stepan et al., 2002; Walsh and Lynch, 2018). Statistical genetics then develops estimators of these quantities using pedigree, marker, or multi-omic data in increasingly complex designs (Hadfield and Nakagawa, 2010; Gienapp et al., 2017).

Conversely, methodological advances feed back into theory. Genomic-relatedness-based estimates of heritability and partitioning of variance by functional annotation have revised understandings of polygenicity and the contribution of rare versus common variants, prompting re-evaluation of classical assumptions (Yang et al., 2011; Hill, 2012; Visscher et al., 2017; Galas et al., 2020). High-dimensional GWAS and multi-trait models have clarified that the “infinitesimal” description remains surprisingly robust while also exposing situations where variance-based summaries mask complex interactions (Hill, 2012; Galas et al., 2020).

In plant breeding, this interaction is particularly visible. Quantitative genetics sets the breeding objectives and theoretical expectations about selection response, G×E and long-term genetic variance (Hill, 2010; Bernardo,

2020). Statistical genetics operationalizes these aims via GWAS to identify loci, genomic prediction models to rank candidates, and simulation frameworks to optimize breeding schemes (Doerge, 2002; Vieira et al., 2025). Bernardo's (2020) reflection that quantitative genetics in plant breeding has become "increasingly empirical and computational and less grounded in theory" underscores how statistical-genetic methods have come to dominate practice even as they rely on quantitative-genetic constructs.

#### 4.4 Toward a unified conceptual framework

Given this intertwined history, a rigid disciplinary boundary is neither realistic nor desirable. A more productive view is to adopt an explicit "theory–algorithm–application" framework. In this scheme, quantitative genetics provides the theory layer: models of how genetic and environmental factors generate phenotypic distributions, definitions of parameters such as heritability and the G-matrix, and conceptual tools such as QLE and adaptive landscapes (Fisher, 1918; Hill, 2010; Neher and Shraiman, 2011; Walsh and Lynch, 2018). Statistical genetics inhabits the algorithm layer, designing estimation and testing procedures—mixed-model solvers, Bayesian regressions, variable-selection GWAS, information-theoretic measures—that target these theoretical quantities under realistic data constraints (Henderson, 1975; Doerge, 2002; Moser et al., 2015; Galas et al., 2020; Qi et al., 2024). Breeding and evolutionary studies constitute the application layer, where theory-aligned quantities estimated by statistical-genetic algorithms inform decisions on selection, conservation and intervention (Hill, 2012; Bazakos et al., 2017; Gienapp et al., 2017; Bernardo, 2020).

Alternatively, a "problem–method–data" perspective may help resolve current confusion. Quantitative genetics formulates the problems—predicting response to selection, quantifying constraints, understanding  $G \times E$ —largely independent of specific data modalities (Mitchell-Olds and Rutledge, 1986; Steppan et al., 2002). Statistical genetics develops methods tuned to particular data regimes: from phenotypes and pedigrees to whole-genome sequences and multi-omics matrices (Doerge, 2002; Hill, 2012; Bazakos et al., 2017; Qi et al., 2024). On this view, the same quantitative-genetic problem (e.g. estimating heritability) may be addressed by different statistical-genetic methods depending on whether only pedigree, SNP array or multi-omic information is available (Yang et al., 2011; Gienapp et al., 2017).

Such frameworks clarify that the distinction between quantitative and statistical genetics is not ontological but epistemological. Quantitative genetics asks what needs to be known about complex traits to understand and predict their evolution; statistical genetics asks how, given particular data and constraints, those quantities can be estimated reliably. Confusion persists because in practice the same individuals often do both kinds of work, using the same linear mixed models and Bayesian machinery for both conceptual exploration and routine data analysis (Hill, 2012; Serpico et al., 2023; Schraiber et al., 2024). Moreover, influential texts such as Lynch and Walsh (1998) and Walsh and Lynch (2018) treat statistical methodology and quantitative-genetic theory in a unified way, reflecting the historical entanglement of the two.

Recognizing their different epistemological roles does not require re-drawing hard disciplinary borders. Instead, it encourages explicit articulation, in any given study, of which elements are grounded in quantitative-genetic theory (e.g. assumptions about additivity, equilibrium, or  $G \times E$ ) and which belong to statistical-genetic implementation (e.g. choice of estimator, regularization, treatment of LD). In the genomics era—characterized by high-dimensional data, complex dependence structures and multi-layered phenotyping—this clarity becomes essential for interpreting results, comparing methods, and designing breeding strategies that are both statistically sound and biologically coherent.

As illustrated in Figure 1, the relationships among quantitative genetics, statistical genetics, and bioinformatics can be understood within a unified analytical framework that links biological questions, methodological approaches, and data resources. In this framework (Figure 1A), quantitative genetics occupies the problem-oriented layer, focusing on the genetic architecture and inheritance of complex traits; statistical genetics represents the methodological layer, providing model-based tools for inference and prediction; and bioinformatics forms the data-processing layer, transforming raw genomic and omics data into structured inputs for downstream

analysis. This hierarchical yet interconnected structure emphasizes that advances in complex trait research are not driven by any single discipline, but by the coordinated interaction among theory, methods, and data.

The evolutionary perspective (Figure 1B) further highlights the continuity and transformation of this integrated system. While classical quantitative genetics was primarily based on phenotypic observations and pedigree information, the introduction of molecular markers enabled QTL mapping and bridged theoretical models with genomic localization. In the genomic era, high-throughput sequencing and large-scale datasets have accelerated the development of statistical genetics, leading to genome-wide association studies and genomic prediction approaches. Despite these methodological shifts, the underlying theoretical foundation of quantitative genetics remains largely intact, underscoring the persistence of core concepts across technological transitions.

Finally, the QTL analysis workflow (Figure 1C) exemplifies how this integration operates in practice. From data acquisition (genotypic and phenotypic information), through statistical modeling (linkage analysis, GWAS, and mixed models), to biological interpretation (QTL detection and candidate gene identification), each step reflects the interplay between data processing, statistical inference, and genetic theory. Taken together, Figure 1 provides a conceptual synthesis of the disciplinary relationships and their functional roles in modern genetics research, reinforcing the view that quantitative genetics and statistical genetics are distinct in orientation yet fundamentally interdependent in application.

## **5 QTL Mapping as a Paradigm of Integration Between Quantitative Genetics and Statistical Genetics**

### **5.1 Quantitative genetic basis of QTL**

The very notion of a quantitative trait locus (QTL) is rooted in classical polygenic inheritance theory. Fisher's 1918 synthesis showed that continuous variation can be explained by the joint segregation of many Mendelian loci of small effect, in combination with environmental noise (Fisher, 1918). Subsequent quantitative genetic theory formalized this insight in terms of additive, dominance and epistatic variance components and their contribution to heritability and response to selection (Falconer and Mackay, 1996; Lynch and Walsh, 1998). In this framework, causal loci remained abstract contributors to variance; quantitative genetics operated at the level of variance components and covariances among relatives, not named genes.

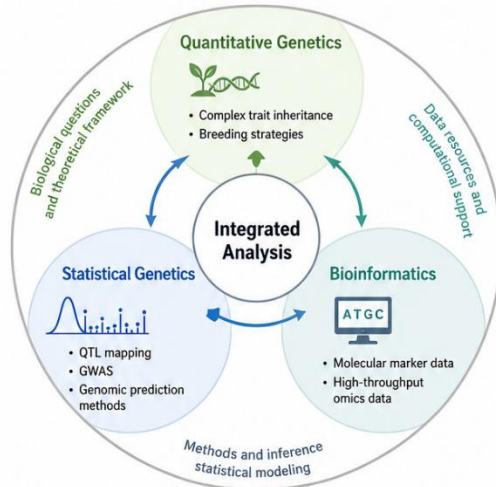
QTL theory makes this abstract polygenic model genomically explicit. A QTL is simply a genomic region where allelic substitution changes the expectation of a quantitative phenotype, i.e. a localized realization of the additive and non-additive effects that, in aggregate, constitute genetic variance (Lynch and Walsh, 1998; Mackay, 2001). The decomposition of phenotypic variance into components attributable to specific chromosomal segments rather than only to "additive variance" at the whole-genome level is conceptually continuous with Fisher's variance partitioning; it is a refinement of the same theory rather than a departure (Fisher, 1918; Falconer and Mackay, 1996; Lynch and Walsh, 1998).

Early empirical demonstrations, such as the dissection of tomato fruit traits into multiple Mendelian factors using a complete RFLP map (Paterson et al., 1988), illustrated that classical polygenic models could be resolved into discrete genomic segments, vindicating the quantitative-genetic view of complex traits. Later syntheses emphasized that QTL mapping is fundamentally an exercise in estimating the positions and effect sizes of loci whose joint contribution corresponds to the variance components of classical quantitative genetics (Mackay, 2001; Du et al., 2016). QTL therefore represent the genomic instantiation of polygenic models: they are where additive, dominance and epistatic effects, long treated as abstract parameters, become linked to specific regions of the genome.

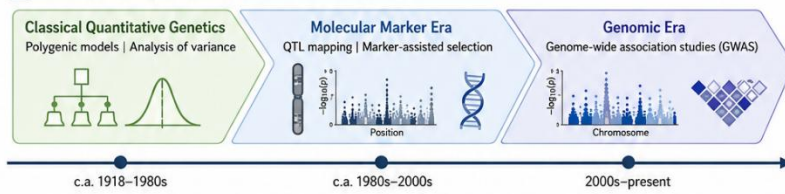
### **5.2 Statistical genetic implementation of QTL mapping**

If QTL are conceptually quantitative-genetic objects, their detection and characterization are irreducibly statistical. QTL mapping is a paradigmatic statistical-genetic enterprise: it turns the qualitative question "does a locus affect a trait?" into a problem of model-based inference under uncertainty.

**A Conceptual Framework of Disciplinary Relationships**



**B Evolutionary Timeline of Genetic Disciplines**



**C QTL Analysis Workflow**

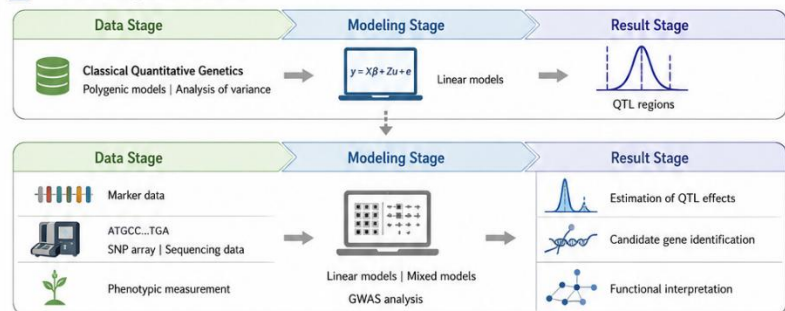


Figure 1 Integrated framework illustrating the relationships among quantitative genetics, statistical genetics, and bioinformatics, along with their historical evolution and the analytical workflow of QTL mapping.

Figure Caption: (A) Conceptual framework showing the hierarchical relationship among quantitative genetics (problem-oriented layer), statistical genetics (methodological layer), and bioinformatics (data-processing layer), highlighting their roles in integrated analysis. (B) Evolutionary timeline of genetic disciplines, illustrating the transition from classical quantitative genetics based on phenotypic and pedigree data, through the molecular marker era characterized by QTL mapping, to the genomic era dominated by genome-wide association studies (GWAS). (C) Workflow of QTL analysis, demonstrating the progression from data acquisition (phenotypic and molecular marker data), through statistical modeling (linear and mixed models), to result interpretation, including QTL detection, effect estimation, and candidate gene identification

5.2.1 Linkage mapping approaches

The seminal contribution of Lander and Botstein (1989) was to adapt human LOD-score linkage analysis to the mapping of quantitative traits using dense RFLP maps. Their interval mapping method treats the putative QTL position as a parameter along the chromosome and uses likelihood-ratio tests to infer both location and effect size, fully exploiting recombination information between flanking markers (Lander and Botstein, 1989). The earlier theoretical proposal that RFLP maps could enable systematic mapping of complex traits (Lander and Botstein 1986; Botstein et al. 1980) thus crystallized into a formal statistical framework.

Subsequent developments—composite interval mapping, multiple-QTL models and mixed-model-based approaches—extended this framework to account for background genetic effects, multiple linked loci, epistasis and QTL  $\times$  environment interactions (Jiang and Zeng, 1995; Yang et al., 2007; Du et al., 2016). These methods remain deeply statistical: they rely on likelihood or F-statistics, permutation-based thresholds, and Bayesian estimation to control genome-wide error rates and to navigate high-dimensional parameter spaces (Doerge, 2002; Yang et al., 2007).

#### 5.2.2 Association mapping (GWAS)

Association mapping generalizes QTL analysis from controlled crosses to diverse natural or breeding populations. Here, linkage disequilibrium (LD) replaces controlled recombination as the source of mapping information: historical recombination events and drift create marker–QTL correlations that can be exploited at much finer scale than in biparental linkage populations (Mackay, 2001; Du et al., 2016). Genome-wide association studies (GWAS) formalize this as a multiple-testing problem over genome-wide markers, typically using linear or logistic regression, mixed models, and stringent control of type I error (Visscher et al., 2012; 2017).

In plants, LD-based association mapping provides high-resolution dissection of QTL intervals identified in linkage studies, enabling the nomination of candidate genes and alleles (Du et al., 2016). Yet, as Visscher et al. (2012; 2017) emphasize, the GWAS design is purely statistical: it aims to detect significant marker–trait associations, not to estimate variance components per se. The quantitative-genetic concepts of effect size and variance explained are then derived from these statistically identified associations.

#### 5.2.3 Mixed model applications

A critical development in association-based QTL mapping was the widespread adoption of linear mixed models (LMMs) to control confounding by population structure and relatedness. Yu et al. (2006) introduced a unified mixed-model approach that incorporates both population structure (Q matrix) and kinship (K matrix), showing markedly improved control of false positives and false negatives in human and maize association studies. Zhang et al. (2010) further improved computational efficiency via compressed mixed models and “population parameters previously determined” (P3D), enabling routine GWAS in large plant panels.

These models are conceptually continuous with Henderson’s animal model and with genomic REML frameworks used to estimate SNP heritability (Henderson, 1975; Yang et al., 2010; 2011). However, in the QTL context, they are deployed as statistical tools: random effects absorb background polygenic variation and relatedness so that fixed marker effects can be tested for association with minimal bias (Yu et al., 2006; Zhang et al., 2010; Du et al., 2016). QTL detection thus becomes a problem of parameter estimation and hypothesis testing within carefully specified linear mixed models.

### 5.3 Disciplinary interpretation of QTL mapping

Viewed through the lens of disciplinary roles, QTL mapping is quintessentially quantitative-genetic in its objectives and statistical-genetic in its implementation. The core questions—how many loci underlie a trait, what are their effect sizes and modes of action, how do they interact, and how much variance do they explain—are classic quantitative-genetic questions about genetic architecture and its consequences for selection (Lynch and Walsh, 1998; Mackay, 2001; Visscher et al., 2017).

At the same time, the procedures used to answer these questions—interval mapping, composite interval mapping, multitrait models, GWAS mixed models, Bayesian multi-QTL frameworks—are designed, evaluated and improved using the criteria of statistical genetics: power, bias, control of false discoveries, robustness to heterogeneity, and computational tractability (Lander and Botstein, 1989; Jiang and Zeng, 1995; Doerge, 2002; Yu et al., 2006; Yang et al., 2007; Zhang et al., 2010).

This duality is explicit in methodological papers that present QTL mapping as a route to infer full genetic architecture, including epistasis and QTL  $\times$  environment interaction, through increasingly sophisticated mixed and Bayesian models (Jiang and Zeng, 1995; Yang et al., 2007; Du et al., 2016). Quantitative genetics specifies what

“genetic architecture” means and why it matters for breeding and evolution; statistical genetics specifies how, given finite and noisy data, one can infer it with quantified uncertainty.

#### **5.4 Impact of QTL studies on both fields**

QTL mapping has reshaped quantitative genetics by providing empirical validation and refinement of its core concepts. Empirically, QTL studies across plants, animals and model organisms have confirmed that complex traits are typically influenced by many loci with a spectrum of effect sizes, often with a few moderate-effect QTL and a long tail of small effects, broadly consistent with polygenic theory (Mackay, 2001; Visscher et al., 2012; 2017). Integrated linkage–LD studies, such as those in *Populus*, have revealed intricate architectures involving additive, dominance and epistatic effects and trait-specific gene–gene networks, enriching quantitative-genetic views of epistasis and pleiotropy (Du et al., 2016).

These results have also informed debates about “missing heritability” by quantifying how much phenotypic variance is captured by detected QTL versus undetected polygenic background (Visscher et al., 2012; 2017). Large-scale GWAS and variance-components analyses using genomic relationships (Yang et al., 2010; 2011) have shown that a substantial fraction of heritability can be recovered by common variants, supporting the infinitesimal model while also highlighting the limits of locus-wise mapping.

For statistical genetics, QTL mapping has been an engine of methodological innovation. It stimulated the development of interval and composite interval mapping, multi-trait and multi-environment models, full-QTL mixed models, Bayesian multi-locus frameworks, and permutation-based multiple-testing procedures (Lander and Botstein, 1989; Jiang and Zeng, 1995; Doerge, 2002; Yang et al., 2007; Du et al., 2016). Later, the challenges of GWAS—population structure, cryptic relatedness, LD, high dimensionality—drove the widespread adoption and refinement of mixed models, kinship matrices, and computational tricks (Yu et al., 2006; Zhang et al., 2010; Yang et al., 2010; 2011).

Conceptually, QTL mapping forced statistical genetics to grapple with inferential problems where the number of potential predictors (markers) vastly exceeds sample size, where signals are correlated through LD, and where effect sizes are heterogeneous and context dependent. This has contributed to broader advances in high-dimensional inference, Bayesian variable selection, and multiple-testing correction that now permeate genomic prediction and multi-omics analysis (Doerge, 2002; Yang et al., 2010, 2011; Visscher et al. 2012).

In this sense, QTL mapping stands as a paradigmatic bridge: it operationalizes quantitative-genetic theory in a statistically coherent way, and it pushes statistical methodology forward in direct response to biologically meaningful questions about complex traits. That it cannot be cleanly assigned to one discipline is precisely what makes it such a powerful example of interdisciplinary integration in modern genetics and plant breeding.

## **6 Revisiting Classical and Modern Quantitative Genetics**

### **6.1 Background of the distinction**

The distinction between “classical” and “modern” quantitative genetics has emerged largely in response to technological change rather than a rupture in theoretical foundations. Fisher’s 1918 synthesis, which reconciled Mendelian inheritance with continuous variation via the infinitesimal model and analysis of variance, provided the conceptual template for complex trait genetics that remains central today (Fisher, 1918; Visscher and Goddard, 2019; Visscher and Walsh, 2019). For much of the twentieth century, this framework was implemented with phenotypes and pedigrees, using expected relationships among relatives to estimate genetic variances and predict response to selection (Falconer and Mackay, 1996; Lynch and Walsh, 1998).

The advent of molecular markers, and later high-throughput sequencing, expanded what quantitative geneticists could observe and manipulate. RFLPs, microsatellites, and SNPs enabled QTL mapping and later genome-wide association studies (GWAS), linking Fisherian variance components to specific genomic regions (Lander and Botstein, 1989; Walsh, 2014; Zargar et al., 2015). High-density marker panels and genotyping-by-sequencing then

made it feasible to saturate genomes with markers and to implement genomic selection (Meuwissen et al., 2001; He et al., 2014; Zargar et al., 2015).

Within this context, the terminology of “classical” versus “modern” quantitative genetics crystallized. “Classical” refers to the phenotype- and pedigree-based theory and methods codified by Falconer and Mackay (1996) and Lynch and Walsh (1998). “Modern” typically denotes the incorporation of dense molecular data, genome-wide prediction, and big-data computation into essentially the same conceptual framework (Walsh, 2014; Visscher and Goddard, 2019; Yin et al., 2023).

## **6.2 Comparison of the two conceptual systems**

Classical quantitative genetics was developed for situations where causal loci were unobserved. Inference rested on patterns of resemblance among relatives, from which additive, dominance and epistatic variance components could be estimated, and on BLUP-based prediction of breeding values using pedigree relationships (Henderson, 1975; Falconer and Mackay, 1996; Lynch and Walsh, 1998). Its core objects—heritability, the G-matrix, correlated response to selection—are population-level summaries that do not depend on knowledge of particular genes (Walsh, 2014; Barton, 2022).

In contrast, “modern” quantitative genetics is strongly genomic data-driven. QTL mapping framed the identification of genomic regions underlying trait variation as a statistical problem in linkage and linkage-disequilibrium mapping (Lander and Botstein, 1989; Meuwissen and Goddard, 2007; Zargar et al., 2015). GWAS extended this to genome-wide scans across diverse populations, while genomic selection used thousands of markers simultaneously to predict total genetic value without requiring locus-wise significance (Meuwissen et al., 2001; He et al., 2014; You et al., 2020). Whole-genome resequencing and sequence-based prediction further intensified this trend, leveraging very dense variant data to improve genomic prediction accuracy (Meuwissen and Goddard, 2010).

Yet the theoretical core is continuous across eras. Genomic BLUP and Bayesian whole-genome regressions are direct descendants of Henderson’s mixed-model framework, with the pedigree relationship matrix replaced by a genomic relationship matrix derived from markers (Henderson, 1975; Meuwissen et al., 2001; Goddard et al., 2011). The key parameters—additive variance, heritability, breeding values—retain the same definitions, even when estimated from SNPs rather than pedigrees (Yang et al., 2011; Visscher et al., 2017). “Modern” quantitative genetics thus represents an expansion of the data and method space, not a replacement of classical theory (Walsh, 2014; Visscher and Walsh, 2019).

## **6.3 Role of statistical genetics in modern quantitative genetics**

The rise of “modern” quantitative genetics has been inseparable from the growth of statistical genetics as a methodology-driven field. Linear mixed models, originally developed for genetic evaluation using pedigree (Henderson, 1975), were generalized to marker-based kinship and implemented in scalable software such as GAPIT and HIBLUP, enabling GWAS and genomic prediction in large plant and animal datasets (Wang and Zhang, 2021; Yin et al., 2023). Bayesian whole-genome regression methods (e.g. BayesA/B/C) and high-dimensional shrinkage models operationalize the infinitesimal idea under genomic big data, handling  $p \gg n$  settings typical of SNP and sequence data (Meuwissen et al., 2001; Meuwissen and Goddard, 2010; Bermingham et al., 2015).

Statistical genetics thus supplies the analytical, computational and modeling frameworks that allow quantitative-genetic theory to be applied in the genomic era. High-throughput genotyping and sequencing generate massive marker matrices; linear mixed models, genomic relationship matrices, and Bayesian regressions translate these into estimates of genomic heritability and genomic breeding values (Yang et al., 2011; Visscher et al., 2017; Yin et al., 2023). High-dimensional feature selection and sparse factor models further extend this toolkit to multi-trait and multi-omic settings (Bermingham et al., 2015; Qu et al., 2022; Amin et al., 2025).

It is therefore important to distinguish modern quantitative genetics from statistical genetics. The former denotes the application and extension of quantitative-genetic theory in the presence of genomic data; the latter denotes the statistical and algorithmic machinery—mixed models, Bayesian inference, identity-by-descent modeling, feature selection—used to analyze such data (Doerge, 2002; Walsh, 2014; Schraiber et al., 2024). Modern quantitative genetics depends heavily on statistical genetics, but the two are not coextensive: many statistical-genetic developments (e.g. coalescent-based LD mapping, effective population size estimation) answer questions not primarily framed in quantitative-genetic terms (Meuwissen and Goddard, 2007; Barton, 2022).

#### **6.4 Rationality and limitations of this classification**

It is rational to talk about “classical” and “modern” quantitative genetics insofar as these labels capture technological and methodological epochs. Classical quantitative genetics was constrained by phenotypic and pedigree information; modern practice routinely integrates dense markers, multi-environment trials, and sometimes multi-omics data into quantitative-genetic models (He et al., 2014; Walsh, 2014; You et al., 2020; Yin et al., 2023). Likewise, statistical genetics has gained a relative methodological independence, with dedicated literatures on GWAS, genomic prediction algorithms, and mixed-model optimization that sometimes proceed with minimal reference to the underlying quantitative-genetic theory (Doerge, 2002; Wang and Zhang, 2021; Yin et al., 2023).

However, this classification has important limitations. First, it can obscure the theoretical continuity of quantitative genetics across eras. Fisher’s variance decomposition and the infinitesimal model remain the conceptual anchor for both phenotype-based and genome-based analyses (Fisher, 1918; Visscher and Walsh, 2019; Barton, 2022). Visscher and Goddard (2019) explicitly argue that Fisher’s 1918 framework is still the foundation for GWAS and genomic prediction a century later, implying that “modern” developments refine rather than replace classical theory.

Second, the language of “modern quantitative genetics” risks conflating theory and method, inviting the mistaken impression that the adoption of genomic data automatically constitutes a new discipline. In reality, most genomic prediction and GWAS models are recognizable as quantitative-genetic mixed models augmented with richer covariance structures and priors (Henderson, 1975; Meuwissen et al., 2001; Yang et al., 2011; Visscher et al., 2017). Statistical genetics has become methodologically sophisticated and partially autonomous, but its central models remain anchored in quantitative-genetic assumptions about additivity, variance components, and the relationship between markers and QTL (Walsh, 2014; Yin et al., 2023; Schraiber et al., 2024).

Finally, strict labels may hinder integrative thinking. As Walsh (2014) emphasized, quantitative genetics already “plays a unique role in biology, serving as the conduit between purely statistical descriptions of trait inheritance and more genetically informed views.” From this perspective, “modern” quantitative genetics is best seen as an extension and enrichment of classical theory enabled by genomics and by the methodological contributions of statistical genetics, rather than as a separate field. Recognizing this helps maintain conceptual coherence while acknowledging the genuine epistemic shift brought by whole-genome data and big-data computation.

## **7 Relationship Between Statistical Genetics and Bioinformatics**

### **7.1 Core differences between the two fields**

Bioinformatics and statistical genetics emerged from the same genomic revolution but occupy distinct epistemological niches. Bioinformatics is primarily concerned with capturing, structuring, and annotating biological information: assembling genomes, aligning reads, calling variants, harmonizing formats, and adding functional and contextual annotation so that the resulting datasets are computable and interoperable (Nielsen et al., 2011; Tatusova et al., 2016; Ahmed et al., 2021; Clark and Lillard, 2024). In this sense, bioinformatics addresses “what is in the data”: which bases are present, which variants and genes can be reliably identified, how they map to reference coordinates, and what is known about their function.

Statistical genetics, by contrast, is centrally about model-based inference: given a curated dataset of genotypes, phenotypes and covariates, what can be inferred about genetic effects, heritability, genetic architecture and

causality? It focuses on estimation, hypothesis testing, prediction and uncertainty quantification, typically via regression, mixed models, Bayesian methods and high-dimensional machine learning (Yang et al., 2010; Richardson et al., 2016; Visscher et al., 2017; Johri et al., 2021). It addresses “what the data mean” in terms of genetic contribution to traits, disease risk, or breeding value.

These roles overlap—bioinformatics pipelines themselves often embed statistical models, and statistical genetics increasingly relies on complex computational infrastructures—but the primary emphasis differs: data engineering and representation versus statistical inference and interpretation (Morris and Baladandayuthapani, 2017; Johri et al., 2021). Confusion persists in practice because many tools and pipelines now integrate both layers, and because both fields draw on statistics and computation, albeit with different goals.

## 7.2 Division of roles in the research pipeline

### 7.2.1 Bioinformatics data preprocessing

In modern sequencing-based studies, the early stages of the pipeline are chiefly bioinformatic. Raw reads must be quality-controlled, aligned to a reference genome, and transformed into variant calls (Nielsen et al., 2011; Ahmed et al., 2021; Dotolo et al., 2022). Widely used next-generation sequencing (NGS) pipelines (e.g. GATK-based workflows) implement complex chains of alignment, duplicate marking, local realignment, base-quality recalibration and variant calling, each step with non-trivial algorithmic and statistical assumptions (Nielsen et al., 2011; Ahmed et al., 2021; Dotolo et al., 2022).

Subsequent annotation and structuring—conversion to standardized formats (VCF, BED), functional annotation of variants, integration with gene models and ontologies, and harmonization across cohorts—are hallmark bioinformatic tasks (Tatusova et al., 2016; Ahmed et al., 2021; Clark and Lillard, 2024). Studies of RAD-seq and WGS/WES pipelines highlight that seemingly technical choices at this stage (de novo vs reference mapping, filtering thresholds, handling of missing data) can dramatically affect downstream population genetic and association inferences (Nielsen et al., 2011; Shafer et al., 2017).

Conceptually, this is still about defining the dataset: deciding which reads, sites and samples are of sufficient quality, how they are encoded, and which external knowledge (functional annotation, databases) is attached. Errors or biases introduced here propagate into all later statistical analyses (Shafer et al., 2017; Johri et al., 2021).

### 7.2.2 Statistical genetics modeling and inference

Once a high-quality, well-annotated genotype (or multi-omics) matrix exists, the focus shifts to statistical genetics. GWAS typically rely on linear or mixed models that relate phenotypes to genotypes under assumptions about population structure, relatedness and effect distributions (Yang et al., 2010; Visscher et al., 2012; 2017). Genomic prediction and polygenic risk scoring extend these models to high-dimensional regression and shrinkage frameworks aimed at maximizing predictive accuracy rather than locus discovery (Yang et al., 2010; Richardson et al., 2016; Orliac et al., 2022). Marker-based REML and related methods use genomic relationships to estimate SNP-heritability and partition variance by genomic annotations (Yang et al., 2011; Visscher et al., 2017; Orliac et al., 2022).

All these tasks are downstream of bioinformatics preprocessing. They assume that genotypes are already called, filtered and harmonized, and that any remaining uncertainty can be modeled via statistical error structures. Critical methodological debates—e.g. about model misspecification, multiple testing, LD structure, and baseline model choice in population genomics—are firmly within statistical genetics (Richardson et al., 2016; Johri et al., 2021; Orliac et al., 2022).

Thus, the research workflow has a clear directional dependency: bioinformatics generates analyzable data; statistical genetics extracts biological meaning. When these roles are conflated, it becomes difficult to diagnose whether discrepancies across studies arise from preprocessing differences or from alternative modeling assumptions (Shafer et al., 2017; Johri et al., 2021; Kui et al., 2026).

### 7.3 Synergy and integration

Despite these conceptual distinctions, the two fields are increasingly intertwined. Multi-omics projects that combine genomics, transcriptomics, epigenomics and proteomics require both sophisticated bioinformatic integration of heterogeneous data types and advanced statistical models for joint analysis (Richardson et al., 2016; Morris and Baladandayuthapani, 2017; Clark and Lillard, 2024). Integrative genomics methods explicitly sit at this interface, using network models, Bayesian hierarchical structures and multiset statistics to connect different assay layers and extract coherent biological signals (Richardson et al., 2016; Morris and Baladandayuthapani, 2017).

The big data scale of contemporary biobanks and sequencing consortia further drives convergence. Efficient GWAS, fine-mapping and genomic prediction at biobank scale depend on highly optimized pipelines that merge high-performance computing, automated QC and imputation (Lam et al., 2019; Orliac et al., 2022). Projects such as RICOPILI exemplify this fusion: they bundle stringent QC, imputation and basic association and polygenic scoring in a single workflow, blurring the practical line between bioinformatics and statistical genetics while still conceptually separating preprocessing and inference components (Lam et al., 2019).

Machine learning and AI reinforce this trend. Deep learning models for variant calling, chromatin structure prediction or integrative biomarker discovery rely on bioinformatic pipelines for feature construction and on statistical principles for training, validation and uncertainty assessment (Morris and Baladandayuthapani, 2017; Dotolo et al., 2022; Clark and Lillard, 2024; Halder et al., 2024). High-dimensional modeling in integrative genomics similarly demands both careful handling of data structures (missingness, batch effects, cross-platform harmonization) and sophisticated inference frameworks that can accommodate complex dependencies and multiple hypothesis testing (Richardson et al., 2016; Johri et al., 2021).

Confusion between statistical genetics and bioinformatics persists partly because many research teams, tools and publications span the full pipeline, and because both fields rely heavily on computation and statistics. A clearer conceptual distinction—bioinformatics as primarily about representation and organization of biological data, statistical genetics as about inference and prediction from those data—can sharpen problem formulation, clarify responsibilities in interdisciplinary collaborations, and improve reproducibility. For plant breeding and genetics, where operational decisions depend on both accurate variant catalogs and robust genetic models, making these epistemological roles explicit is essential for designing pipelines that are simultaneously technically sound and biologically interpretable.

## 8 Future Perspectives and Challenges in Quantitative and Statistical Genetics

### 8.1 Methodological challenges

The methodological frontier of quantitative and statistical genetics is dominated by the high-dimensionality of modern data. In genomic prediction and SNP-heritability estimation, the number of markers typically far exceeds the number of individuals ( $p \gg n$ ), making overfitting, instability of effect estimates, and confounding via linkage disequilibrium central concerns (Yang et al., 2011; Hill, 2012; Visscher et al., 2017). Penalized regression, Bayesian whole-genome regression, and mixed-model approaches mitigate these issues, yet they often rely on strong assumptions (e.g. Gaussian effect priors) that may not hold across traits or species (Meuwissen et al., 2001; Hill, 2012; Crossa et al., 2025).

At the same time, there is pressure to model increasingly complex genetic architectures, including epistasis, pleiotropy, and genotype-by-environment interaction ( $G \times E$ ). Incorporating higher-order interactions and reaction norms into prediction models remains statistically and computationally demanding, especially when crossed with multi-environment and multi-trait designs (Jarquín et al., 2014; van Eeuwijk et al., 2016; Li and Gutierrez, 2023; Dwivedi et al., 2024). Kernel methods, random regressions and non-linear models can capture some of this complexity but raise questions about over-parameterization and identifiability in finite breeding data sets (Messina et al. 2018; Crossa et al., 2025).

These developments sharpen the trade-off between prediction accuracy and interpretability. Genomic BLUP and related linear models are relatively transparent and align closely with quantitative-genetic theory, but may underexploit non-linear signal. Machine- and deep-learning models promise higher accuracy with multi-omics and enviromic inputs but often function as black boxes, complicating biological interpretation and breeding decisions (Pérez-Enciso, 2021; Jeon et al., 2023; Crossa et al., 2025). Reconciling the need for explainable models with the empirical gains of complex architectures is a central methodological challenge.

## **8.2 Biological challenges**

Methodological sophistication has not eliminated key biological uncertainties. The “missing heritability” problem—where GWAS loci explain only a fraction of pedigree-based or twin-based heritability—remains an important conceptual issue (Manolio et al., 2009). Work using genomic-relationship matrices has shown that a substantial portion of heritability can be captured when all SNPs are fitted jointly, supporting a highly polygenic, near-infinitesimal model (Yang et al., 2011; Hill, 2012; Visscher et al., 2017). Nevertheless, limited power for rare variants, imperfect tagging, structural variants, and context-dependent effects all contribute to gaps between observed and expected variance explained.

A second unresolved challenge concerns the functional interpretation of non-coding variants, which constitute the bulk of GWAS hits. Large-scale functional genomics and xQTL studies are beginning to map links between non-coding variants, chromatin accessibility, gene expression and downstream phenotypes, but mechanistic understanding is still fragmentary (Manolio et al., 2009; Bykova et al. 2022; Yang et al., 2024). Deep-learning models that predict regulatory activity and variant effects from sequence or epigenomic context offer promising tools, yet their integration with quantitative-genetic parameters (e.g. variance explained,  $G \times E$ ) is in its infancy (Yang et al., 2024; Chen, 2025).

In crops, similar questions arise around regulatory variants, structural variation and pangenome diversity underlying adaptation and stress tolerance (Bayer et al., 2021; Varshney et al., 2021). Quantitative and statistical genetics must increasingly interact with functional genomics to move from statistical association to causal and mechanistic models of complex traits.

## **8.3 Emerging technological directions**

The most dynamic technological trajectory involves multi-omics integration. Combining genomics with transcriptomics, epigenomics, metabolomics and detailed phenomics promises richer models of the pathways from genotype to phenotype and of gene–environment interplay (Ritchie et al., 2015; Xu et al., 2022; Alemu et al., 2025). However, heterogeneous data structures, differing sample sizes across omics layers, and batch and cohort effects make integrative modeling technically challenging. Meta-dimensional and multi-stage strategies, as well as network-based and causal-inference approaches, are being developed to address these issues (Ritchie et al., 2015; Dugourd et al., 2021; Alemu et al., 2025).

Simultaneously, artificial intelligence (AI) and machine learning are transforming how large genomic and omic datasets are analyzed. Machine-learning and deep-learning methods are being used for genomic prediction, multi-trait and multi-environment modeling, variant prioritization, and functional annotation (Xu et al., 2022; Jeon et al., 2023; Athanasopoulou et al., 2025; Crossa et al., 2025). AI-driven models can learn high-order interactions and latent structure across heterogeneous data, but raise concerns about overfitting, generalizability across populations and environments, and interpretability (Pérez-Enciso, 2021; Chen, 2025). Future directions likely lie in hybrid approaches that blend the scalability and feature-learning capacity of AI with the inferential discipline of statistical genetics and the constraints of quantitative-genetic theory (Yang et al., 2024; Chen, 2025).

## **8.4 Applications in molecular breeding**

In molecular breeding, the convergence of these methodological and biological developments is reshaping precision breeding and genomics-enabled selection. Genomic selection, originally proposed as a whole-genome regression framework (Meuwissen et al., 2001), is now widely viewed as a central engine for accelerating genetic gain by shortening breeding cycles, increasing selection intensity and exploiting complex trait variation

(Grattapaglia et al., 2018; Voss-Fels et al., 2019; Varshney et al., 2021). Current work focuses on optimizing training population design, model updating, and cross-population and cross-environment prediction (Grattapaglia et al., 2018; Merrick and Carter, 2021; Escamilla et al., 2025).

Future gains depend on integrating multi-omics and environmental data into genomic-enabled prediction. Genomic-enviromic prediction frameworks, smart or “digital” breeding concepts, and integrated genomic-enviromic prediction (iGEP) schemes seek to use spatiotemporal environmental covariates and high-throughput phenotyping to model G×E explicitly and to forecast performance in untested environments and climates (Messina et al., 2018; Xu et al., 2022; Jeon et al., 2023; Li and Gutierrez, 2023). Multi-omics information may further refine predictions by capturing regulatory and metabolic intermediates, though robust integrative models that consistently outperform genomic-only approaches remain an active research target (Ritchie et al., 2015; Xu et al., 2022; Crossa et al., 2025).

Looking ahead, achieving sustainable improvements in complex traits under climate change and resource constraints will require tight co-evolution of methods, biology and data resources. Statistical genetics must continue to innovate in high-dimensional modeling, causal inference, and explainable AI; quantitative genetics must refine theories of polygenicity, G×E and long-term response in genomics-rich contexts; and breeding programs must invest in dense, high-quality phenotypic, genomic and environmental databases at operational scales (Hill, 2012; Grattapaglia et al. 2018; Bernardo, 2020; Xu et al., 2022). The most promising future lies not in any single technique, but in integrated frameworks where statistical methodology, biological interpretation and large-scale data are jointly designed to unlock the full potential of complex trait genetics for crop improvement.

## **9 Conclusion**

### **9.1 Main conclusions**

This study provides a systematic analysis of the conceptual foundations, developmental trajectories, and interrelationships between statistical genetics and quantitative genetics. Overall, the two disciplines differ fundamentally in their academic orientation. Quantitative genetics is primarily concerned with the genetic mechanisms underlying complex traits and their improvement strategies, and is inherently problem- and theory-driven. In contrast, statistical genetics focuses on model construction and data analysis, exhibiting a distinctly methodological character.

However, in the context of modern genetic research, these differences have not led to disciplinary separation. Instead, driven by the availability of genomic-scale data, the two fields have developed a highly complementary relationship. Quantitative genetics formulates scientific questions and provides theoretical frameworks, whereas statistical genetics enables these questions to be tested and implemented through continuously evolving analytical methods in high-dimensional data settings.

Research paradigms such as QTL mapping, genome-wide association studies (GWAS), and genomic selection clearly demonstrate that the genetic dissection of complex traits can no longer be achieved within a single disciplinary framework. Rather, these approaches underscore that progress in modern genetics relies on the deep integration of theory and methodology, rather than rigid disciplinary boundaries.

### **9.2 Integration of theory and methodology**

As the disciplines continue to evolve, the traditional dichotomy between “quantitative genetics” and “statistical genetics” has become increasingly insufficient to fully capture contemporary research practices. In this study, a unified conceptual framework is proposed from the perspectives of “problem–method–data” and “theory–algorithm–application,” aiming to structurally integrate the two fields.

Within this framework, quantitative genetics primarily defines research questions and provides theoretical interpretations, while statistical genetics offers the tools for model implementation and inference. Molecular and multi-omics data serve as the foundational resources supporting this analytical process.

This integrative perspective not only helps clarify conceptual ambiguities but also provides a more coherent understanding of how modern genetic research operates. Particularly in the genomic era, the boundary between theory and methodology is becoming increasingly blurred, with continuous interaction between the two driving complex trait research toward higher resolution and improved predictive accuracy. Consequently, viewing statistical genetics as merely a replacement or simple extension of quantitative genetics fails to accurately reflect its role within the broader disciplinary system.

### 9.3 Implications for future research

From a forward-looking perspective, the relationship between statistical genetics and quantitative genetics will increasingly be characterized by deeper interdisciplinary integration. On the one hand, the challenges posed by high-dimensional data and complex modeling in the study of complex traits require researchers to possess both a solid foundation in genetic theory and advanced statistical modeling skills. On the other hand, the incorporation of multi-omics data, artificial intelligence, and large-scale computational technologies continues to expand the scope of research, further intensifying the interplay among disciplines.

Future research will therefore need not only methodological innovation but also closer integration between theory and application. In applied fields such as molecular breeding, this integration will directly influence prediction accuracy and decision-making efficiency. Promoting the coordinated development of statistical genetics and quantitative genetics at the conceptual, methodological, and practical levels is essential for advancing the understanding of complex trait genetics and achieving precision improvement.

Overall, a clear understanding of the relationship between these two disciplines not only contributes to more precise academic communication but also provides forward-looking guidance for research design and talent development in related fields.

### Author Contributions

Xuanjun Fang conducted the study, including literature review, data analysis, and drafting and revising the manuscript. The author has read and approved the final version of the manuscript.

### Acknowledgments

This work was supported by the Major Program of the National Natural Science Foundation of China (Grant No. 30490254).

### References

- Ahmed Z., Renart E.G., and Zeeshan S., 2021, Genomics pipelines to investigate susceptibility in whole genome and exome sequenced data for variant discovery, annotation, prediction and genotyping, *PeerJ*, 9: e11724.  
<https://doi.org/10.7717/peerj.11724>
- Alemu R., Sharew N.T., Arsano Y.Y., Ahmed M., Tekola-Ayele F., Mersha T.B., and Amare A.T., 2025, Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: techniques, translation, and equity issues, *Human Genomics*, 19(1): 8.  
<https://doi.org/10.1186/s40246-025-00718-9>
- Amin A., Zaman W., and Park S., 2025, Harnessing multi-omics and predictive modeling for climate-resilient crop breeding: from genomes to fields, *Genes*, 16(7): 809.  
<https://doi.org/10.3390/genes16070809>
- Athanasopoulou K., Michalopoulou V.I., Scorilas A., and Adamopoulos P.G., 2025, Integrating artificial intelligence in next-generation sequencing: advances, challenges, and future directions, *Current Issues in Molecular Biology*, 47(6): 470.  
<https://doi.org/10.3390/cimb47060470>
- Barton N., 2022, The "New Synthesis", *Proceedings of the National Academy of Sciences of the United States of America*, 119(30): e2122147119.  
<https://doi.org/10.1073/pnas.2122147119>
- Bayer P.E., Golicz A.A., Scheben A., Batley J., and Edwards D., 2020, Plant pan-genomes are the new reference, *Nature Plants*, 6(8): 914-920.  
<https://doi.org/10.1038/s41477-020-0733-0>
- Bazakos C., Hanemian M., Trontin C., Jiménez-Gómez J.M., and Loudet O., 2017, New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype, *Annual Review of Plant Biology*, 68: 435-455.  
<https://doi.org/10.1146/annurev-arplant-042916-040820>
- Bermingham M., Pong-Wong R., Spiliopoulou A., Hayward C., Rudan I., Campbell H., Wright A., Wilson J., Agakov F., Navarro P., and Haley C., 2015, Application of high-dimensional feature selection: evaluation for genomic prediction in man, *Scientific Reports*, 5(1): 10312.  
<https://doi.org/10.1038/srep10312>

- Bernardo R., 2020, Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE, *Heredity*, 125(6): 375-385.  
<https://doi.org/10.1038/s41437-020-0312-1>
- Bhat J., Ali S., Salgotra R., Mir Z., Dutta S., Jadon V., Tyagi A., Mushtaq M., Jain N., Singh P., Singh G., and Prabhu K., 2016, Genomic selection in the era of next generation sequencing for complex traits in plant breeding, *Frontiers in Genetics*, 7: 221.  
<https://doi.org/10.3389/fgene.2016.00221>
- Botstein D., White R.L., Skolnick M., and Davis R.W., 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *American Journal of Human Genetics*, 32(3): 314.
- Bykova M., Hou Y., Eng C., and Cheng F., 2022, Quantitative trait locus (xQTL) approaches identify risk genes and drug targets from human non-coding genomes, *Human Molecular Genetics*, 31(R1): R105-R113.  
<https://doi.org/10.1093/hmg/ddac208>
- Calus M.P., 2010, Genomic breeding value prediction: methods and procedures, *Animal*, 4(2): 157-164.  
<https://doi.org/10.1017/S1751731109991352>
- Chen L., 2025, Can classical statistics and deep learning converge on explainable, causally driven target discovery?, *DNA Research*, 32(5): dsaf024.  
<https://doi.org/10.1093/dnares/dsaf024>
- Cheverud J.M. and Routman E.J., 1995, Epistasis and its contribution to genetic variance components, *Genetics*, 139(3): 1455-1461.  
<https://doi.org/10.1093/genetics/139.3.1455>
- Clark A.J. and Lillard Jr J.W., 2024, A comprehensive review of bioinformatics tools for genomic biomarker discovery driving precision oncology, *Genes*, 15(8): 1036.  
<https://doi.org/10.3390/genes15081036>
- Cooper M. and Messina C.D., 2021, Can we harness "enviromics" to accelerate crop improvement by integrating breeding and agronomy?, *Frontiers in Plant Science*, 12: 735143.  
<https://doi.org/10.3389/fpls.2021.735143>
- Crossa J., de los Campos G., Pérez P., Gianola D., Burgueno J., Araus J.L., Makumbi D., Singh R.P., Dreisigacker S., Yan J., Arief V., Banziger M., and Braun H.J., 2010, Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers, *Genetics*, 186(2): 713-724.  
<https://doi.org/10.1534/genetics.110.118521>
- Crossa J., Martini J., Vitale P., Pérez-Rodríguez P., Costa-Neto G., Fritsche-Neto R., Runcie D., Cuevas J., Toledo F., Li H., De Vita P., Gerard G., Dreisigacker S., Crespo-Herrera L., Pierre C., Bentley A., Lillemo M., Ortiz R., Montesinos-López O., and Montesinos-López A., 2025, Expanding genomic prediction in plant breeding: harnessing big data, machine learning, and advanced software, *Trends in Plant Science*, 30(7): 756-774.  
<https://doi.org/10.1016/j.tplants.2024.12.009>
- Doerge R.W., 2002, Mapping and analysis of quantitative trait loci in experimental populations, *Nature Reviews Genetics*, 3(1): 43-52.  
<https://doi.org/10.1038/nrg703>
- Doerge R.W., Zeng Z.B., and Weir B.S., 1997, Statistical issues in the search for genes affecting quantitative traits in experimental populations, *Statistical Science*, 12(3): 195-219.  
<https://doi.org/10.1214/ss/1030037909>
- Dotolo S., Abate E., Roma C., Guido D., Preziosi A., Tropea B., Palluzzi F., Giaco L., and Normanno N., 2022, Bioinformatics: from NGS data to biological complexity in variant detection and oncological clinical practice, *Biomedicines*, 10(9): 2074.  
<https://doi.org/10.3390/biomedicines10092074>
- Du Q., Gong C., Wang Q., Zhou D., Yang H., Pan W., Li B., and Zhang D., 2016, Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies, *New Phytologist*, 209(3): 1067-1082.  
<https://doi.org/10.1111/nph.13695>
- Dugourd A., Kuppe C., Sciacovelli M., Gjerga E., Gabor A., Emdal K.B., Vieira V., Bekker-Jensen D.B., Kranz J., Bindels E.M.J., Costa A.S.H., Sousa A., Beltrao P., Rocha M., Olsen J.V., Frezza C., Kramann R., and Saez-Rodriguez J., 2021, Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses, *Molecular Systems Biology*, 17(1): MSB20209730.  
<https://doi.org/10.15252/msb.20209730>
- Dwivedi S., Heslop-Harrison P., Amas J., Ortiz R., and Edwards D., 2024, Epistasis and pleiotropy-induced variation for plant breeding, *Plant Biotechnology Journal*, 22(10): 2788-2807.  
<https://doi.org/10.1111/pbi.14405>
- Escamilla D., Li D., Negus K., Kappelmann K., Kusmec A., Vanous A., Schnable P., Li X., and Yu J., 2025, Genomic selection: essence, applications, and prospects, *The Plant Genome*, 18(2): e70053.  
<https://doi.org/10.1002/tpg2.70053>
- Evans L., Tahmasbi R., Vrieze S., Abecasis G., Das S., Gazal S., Bjelland D., De Candia T., Goddard M., Neale B., Yang J., Visscher P., and Keller M., 2018, Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits, *Nature Genetics*, 50(5): 737-745.  
<https://doi.org/10.1038/s41588-018-0108-x>
- Falconer D.S., and Mackay T.F.C., 1996, *Introduction to Quantitative Genetics* (4th ed.), Longman, Harlow, UK.
- Fisher R.A., 1918, The correlation between relatives on the supposition of Mendelian inheritance, *Transactions of the Royal Society of Edinburgh*, 52(2): 399-433.  
<https://doi.org/10.1017/S0080456800012163>

- Galas D., Kunert-Graf J., Uechi L., and Sakhanenko N., 2020, Toward an Information Theory of Quantitative Genetics, *Journal of Computational Biology*, 28(6): 527-559.  
<https://doi.org/10.1089/cmb.2020.0032>
- Gienapp P., Fior S., Guillaume F., Lasky J.R., Sork V.L., and Csilléry K., 2017, Genomic quantitative genetics to study evolution in the wild, *Trends in Ecology & Evolution*, 32(12): 897-908.  
<https://doi.org/10.1016/j.tree.2017.09.004>
- Goddard M.E., Hayes B.J., and Meuwissen T.H., 2011, Using the genomic relationship matrix to predict the accuracy of genomic selection, *Journal of Animal Breeding and Genetics*, 128(6): 409-421.  
<https://doi.org/10.1111/j.1439-0388.2011.00964.x>
- Grattapaglia D., Silva-Junior O., Resende R., Cappa E., Müller B., Tan B., Isik F., Ratcliffe B., and El-Kassaby Y., 2018, Quantitative Genetics and Genomics Converge to Accelerate Forest Tree Breeding, *Frontiers in Plant Science*, 9: 1693.  
<https://doi.org/10.3389/fpls.2018.01693>
- Guyon I., and Elisseeff A., 2003, An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3(Mar): 1157-1182.
- Habier D., Fernando R.L., Kizilkaya K., and Garrick D.J., 2011, Extension of the Bayesian alphabet for genomic selection, *BMC Bioinformatics*, 12(1): 186.  
<https://doi.org/10.1186/1471-2105-12-186>
- Hadfield J.D. and Nakagawa S., 2010, General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters, *Journal of Evolutionary Biology*, 23(3): 494-508.  
<https://doi.org/10.1111/j.1420-9101.2009.01915.x>
- Halder A., Agarwal A., Jodkowska K., and Plewczynski D., 2024, A systematic analyses of different bioinformatics pipelines for genomic data and its impact on deep learning models for chromatin loop prediction, *Briefings in Functional Genomics*, 23(5): 538-548.  
<https://doi.org/10.1093/bfpg/elae009>
- Hayes B. and Goddard M., 2010, Genome-wide association and genomic selection in animal breeding, *Genome*, 53(11): 876-883.  
<https://doi.org/10.1139/G10-076>
- Hayes B., 2013, Overview of Statistical Methods for Genome-Wide Association Studies (GWAS), In: Gondro C., van der Werf J., and Hayes B. (eds), *Genome-Wide Association Studies and Genomic Prediction*, *Methods in Molecular Biology*, 1019: 149-169.  
[https://doi.org/10.1007/978-1-62703-447-0\\_6](https://doi.org/10.1007/978-1-62703-447-0_6)
- He J., Zhao X., Laroche A., Lu Z.X., Liu H., and Li Z., 2014, Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, *Frontiers in Plant Science*, 5: 484.  
<https://doi.org/10.3389/fpls.2014.00484>
- Heffner E.L., Sorrells M.E., and Jannink J.L., 2009, Genomic selection for crop improvement, *Crop Science*, 49(1): 1-12.  
<https://doi.org/10.2135/cropsci2008.08.0512>
- Henderson C.R., 1975, Best linear unbiased estimation and prediction under a selection model, *Biometrics*, 31(2): 423-447.  
<https://doi.org/10.2307/2529430>
- Hill W.G., 2010, Understanding and using quantitative genetic variation, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537): 73-85.  
<https://doi.org/10.1098/rstb.2009.0203>
- Hill W.G., 2012, Quantitative genetics in the genomics era, *Current Genomics*, 13(3): 196-206.  
<https://doi.org/10.2174/138920212800543110>
- Hill W.G., Goddard M.E., and Visscher P.M., 2008, Data and theory point to mainly additive genetic variance for complex traits, *PLoS Genetics*, 4(2): e1000008.  
<https://doi.org/10.1371/journal.pgen.1000008>
- Hivert V., Sidorenko J., Rohart F., Goddard M., Yang J., Wray N., Yengo L., and Visscher P., 2021, Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals, *The American Journal of Human Genetics*, 108(5): 786-798.  
<https://doi.org/10.1016/j.ajhg.2021.02.014>
- Hou K., Burch K., Majumdar A., Shi H., Mancuso N., Wu Y., Sankararaman S., and Pasiunic B., 2019, Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture, *Nature Genetics*, 51(8): 1244-1251.  
<https://doi.org/10.1038/s41588-019-0465-0>
- Jaganathan D., Bohra A., Thudi M., and Varshney R.K., 2020, Fine mapping and gene cloning in the post-NGS era: advances and prospects, *Theoretical and Applied Genetics*, 133(5): 1791-1810.  
<https://doi.org/10.1007/s00122-020-03560-w>
- Jarquín D., Crossa J., Lacaze X., Du Cheyron P., Daucourt J., Lorgeou J., Piraux F., Guerreiro L., Pérez P., Calus M., Burgueño J., and de los Campos G., 2014, A reaction norm model for genomic selection using high-dimensional genomic and environmental data, *Theoretical and Applied Genetics*, 127(3): 595-607.  
<https://doi.org/10.1007/s00122-013-2243-1>
- Jeon D., Kang Y., Lee S., Choi S., Sung Y., Lee T.H., and Kim C., 2023, Digitalizing breeding in plants: A new trend of next-generation breeding based on genomic prediction, *Frontiers in Plant Science*, 14: 1092584.  
<https://doi.org/10.3389/fpls.2023.1092584>

- Jiang C. and Zeng Z.B., 1995, Multiple trait analysis of genetic mapping for quantitative trait loci, *Genetics*, 140(3): 1111-1127.  
<https://doi.org/10.1093/genetics/140.3.1111>
- John M., Haselbeck F., Dass R., Malisi C., Ricca P., Dreischer C., Schultheiss S., and Grimm D., 2022, A comparison of classical and machine learning-based phenotype prediction methods on simulated data and three plant species, *Frontiers in Plant Science*, 13: 932512.  
<https://doi.org/10.3389/fpls.2022.932512>
- Johri P., Aquadro C., Beaumont M., Charlesworth B., Excoffier L., Eyre-Walker A., Keightley P., Lynch M., McVean G., Payseur B., Pfeifer S., Stephan W., and Jensen J., 2021, Recommendations for improving statistical inference in population genomics, *PLoS Biology*, 20(5): e3001669.  
<https://doi.org/10.1371/journal.pbio.3001669>
- Jones D., Fornarelli R., Derbyshire M., Gibberd M., Barker K., and Hane J., 2023, The pursuit of genetic gain in agricultural crops through the application of machine-learning to genomic prediction, *Frontiers in Genetics*, 14: 1186782.  
<https://doi.org/10.3389/fgene.2023.1186782>
- Kui N., Yu Y., Choi J., McCaw Z.R., Li X., Huff C., and Sun R., 2026, Large Impact of Genetic Data Processing Steps on Stability and Reproducibility of Set-Based Analyses in Genome-Wide Association Studies, *Genetics*, iyag079.  
<https://doi.org/10.1093/genetics/iyag079>
- Laird N.M. and Lange C., 2011, *The fundamentals of modern statistical genetics*, Springer.  
<https://doi.org/10.1007/978-1-4419-7338-2>
- Lam M., Awasthi S., Watson H., Goldstein J., Panagiotaropoulou G., Trubetsky V., Karlsson R., Frei O., Fan C., De Witte W., Mota N., Mullins N., Skarabis N., Huang H., Neale B., Daly M., Mattheissen M., Walters R., and Ripke S., 2019, RICOPIIL: rapid imputation for COnsortias PIpeLIne, *Bioinformatics*, 36(3): 930-933.  
<https://doi.org/10.1093/bioinformatics/btz633>
- Lander E.S. and Botstein D., 1989, Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics*, 121(1): 185-199.  
<https://doi.org/10.1093/genetics/121.1.185>
- Leal S.M., 2001, Genetics and analysis of quantitative traits, *American Journal of Human Genetics*, 68(2): 548-549.  
<https://doi.org/10.1086/318209>
- Legarra A., Aguilar I., and Misztal I., 2009, A relationship matrix including full pedigree and genomic information, *Journal of Dairy Science*, 92(9): 4656-4663.  
<https://doi.org/10.3168/jds.2009-2061>
- Li Z. and Gutierrez L., 2023, Statistical methods for analyzing multiple environmental quantitative genomic data, *Frontiers in Genetics*, 14: 1212804.  
<https://doi.org/10.3389/fgene.2023.1212804>
- Li Z. and Sillanpää M.J., 2012, Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection, *Theoretical and Applied Genetics*, 125(3): 419-435.  
<https://doi.org/10.1007/s00122-012-1892-9>
- Lynch M. and Walsh B., 1998, *Genetics and analysis of quantitative traits*, Sinauer, 1: 535-557.
- Mackay T.F. and Anholt R.R., 2024, Pleiotropy, epistasis and the genetic architecture of quantitative traits, *Nature Reviews Genetics*, 25(9): 639-657.  
<https://doi.org/10.1038/s41576-024-00711-3>
- Mackay T.F., 2001, The genetic architecture of quantitative traits, *Annual Review of Genetics*, 35(1): 303-339.  
<https://doi.org/10.1146/annurev.genet.35.102401.090633>
- Mackay T.F.C., 2014, Epistasis and quantitative traits: using model organisms to study gene-gene interactions, *Nature Reviews Genetics*, 15(1): 22-33.  
<https://doi.org/10.1038/nrg3627>
- Malle S., 2022, Population structure and relatedness for genome-wide association studies, In *Genome-Wide Association Studies*, Springer US: 185-196.  
[https://doi.org/10.1007/978-1-0716-2237-7\\_12](https://doi.org/10.1007/978-1-0716-2237-7_12)
- Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K., Sale M., and Chen W.M., 2010, Robust relationship inference in genome-wide association studies, *Bioinformatics*, 26(22): 2867-2873.  
<https://doi.org/10.1093/bioinformatics/btq559>
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F.C., McCarroll S.A., and Visscher P.M., 2009, Finding the missing heritability of complex diseases, *Nature*, 461(7265): 747-753.  
<https://doi.org/10.1038/nature08494>
- Marx V., 2013, The big challenges of big data, *Nature*, 498(7453): 255-260.  
<https://doi.org/10.1038/498255a>
- Mathews K.L., Malosetti M., Chapman S., McIntyre L., Reynolds M., Shorter R., and Van Eeuwijk F., 2008, Multi-environment QTL mixed models for drought stress adaptation in wheat, *Theoretical and Applied Genetics*, 117(7): 1077-1091.  
<https://doi.org/10.1007/s00122-008-0846-8>
- Messina C.D., Technow F., Tang T., Totir R., Gho C., and Cooper M., 2018, Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP), *European Journal of Agronomy*, 100: 151-162.  
<https://doi.org/10.1016/j.eja.2018.01.007>
- Meuwissen T. and Goddard M., 2010, Accurate prediction of genetic values for complex traits by whole-genome resequencing, *Genetics*, 185(2): 623-631.  
<https://doi.org/10.1534/genetics.110.116590>

- Meuwissen T., van den Berg I., and Goddard M., 2021, On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL, *Genetics Selection Evolution*, 53(1): 19.  
<https://doi.org/10.1186/s12711-021-00607-4>
- Meuwissen T.H. and Goddard M.E., 2007, Multipoint identity-by-descent prediction using dense markers to map quantitative trait loci and estimate effective population size, *Genetics*, 176(4): 2551-2560.  
<https://doi.org/10.1534/genetics.107.070953>
- Meuwissen T.H., Hayes B.J., and Goddard M., 2001, Prediction of total genetic value using genome-wide dense marker maps, *Genetics*, 157(4): 1819-1829.  
<https://doi.org/10.1093/genetics/157.4.1819>
- Mitchell-Olds T. and Rutledge J.J., 1986, Quantitative genetics in natural plant populations: a review of the theory, *The American Naturalist*, 127(3): 379-402.  
<https://doi.org/10.1086/284490>
- Montesinos-López O., Montesinos-López A., Mosqueda-González B., Delgado-Enciso I., Chavira-Flores M., Crossa J., Dreisigacker S., Sun J., and Ortiz R., 2025, Genomic prediction powered by multi-omics data, *Frontiers in Genetics*, 16: 1636438.  
<https://doi.org/10.3389/fgene.2025.1636438>
- Montesinos-López O., Montesinos-López A., Pérez-Rodríguez P., Barrón-López J., Martini J., Fajardo-Flores S., Gaytán-Lugo L., Santana-Mancilla P., and Crossa J., 2021, A review of deep learning applications for genomic selection, *BMC Genomics*, 22(1): 19.  
<https://doi.org/10.1186/s12864-020-07319-x>
- Morris J.S. and Baladandayuthapani V., 2017, Statistical contributions to bioinformatics: design, modelling, structure learning and integration, *Statistical Modelling*, 17(4-5): 245-289.  
<https://doi.org/10.1177/1471082X17698255>
- Moser G., Lee S.H., Hayes B.J., Goddard M.E., Wray N.R., and Visscher P.M., 2015, Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model, *PLoS Genetics*, 11(4): e1004969.  
<https://doi.org/10.1371/journal.pgen.1004969>
- Neher R.A. and Shraiman B.I., 2011, Statistical genetics and evolution of quantitative traits, *Reviews of Modern Physics*, 83(4): 1283-1300.  
<https://doi.org/10.1103/RevModPhys.83.1283>
- Nelson R.M., Pettersson M.E., and Carlborg Ö., 2013, A century after Fisher: time for a new paradigm in quantitative genetics, *Trends in Genetics*, 29(12): 669-676.  
<https://doi.org/10.1016/j.tig.2013.09.006>
- Nielsen R., Paul J.S., Albrechtsen A., and Song Y.S., 2011, Genotype and SNP calling from next-generation sequencing data, *Nature Reviews Genetics*, 12(6): 443-451.  
<https://doi.org/10.1038/nrg2986>
- Orliac E.J., Trejo Banos D., Ojavee S.E., Läll K., Mägi R., Visscher P.M., and Robinson M.R., 2022, Improving GWAS discovery and genomic prediction accuracy in biobank data, *Proceedings of the National Academy of Sciences*, 119(31): e2121279119.  
<https://doi.org/10.1073/pnas.2121279119>
- Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E., and Tanksley S.D., 1988, Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms, *Nature*, 335(6192): 721-726.  
<https://doi.org/10.1038/335721a0>
- Pérez-Enciso M., 2021, Breeding beyond genomics, *Journal of Animal Breeding and Genetics*, 138(3): 275-276.  
<https://doi.org/10.1111/jbg.12547>
- Posthuma D., Beem A.L., De Geus E.J., Van Baal G.C.M., Von Hjelmborg J.B., Iachine I., and Boomsma D.I., 2003, Theory and practice in quantitative genetics, *Twin Research and Human Genetics*, 6(5): 361-376.  
<https://doi.org/10.1375/136905203770326367>
- Qi T., Song L., Guo Y., Chen C., and Yang J., 2024, From genetic associations to genes: methods, applications, and challenges, *Trends in Genetics*, 40(8): 642-667.  
<https://doi.org/10.1016/j.tig.2024.04.008>
- Qu J., Runcie D.E., and Cheng H., 2022, Mega-scale mixed models for genome-wide prediction with thousands of high-throughput phenotyping traits, In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production*, Wageningen Academic Publishers: 1294-1297.  
[https://doi.org/10.3920/978-90-8686-940-4\\_308](https://doi.org/10.3920/978-90-8686-940-4_308)
- Richardson S., Tseng G.C., and Sun W., 2016, Statistical methods in integrative genomics, *Annual Review of Statistics and Its Application*, 3(1): 181-209.  
<https://doi.org/10.1146/annurev-statistics-041715-033506>
- Ritchie M.D., Holzinger E.R., Li R., Pendergrass S.A., and Kim D., 2015, Methods of integrating data to uncover genotype-phenotype interactions, *Nature Reviews Genetics*, 16(2): 85-97.  
<https://doi.org/10.1038/nrg3868>
- Schraiber J.G., Edge M.D., and Pennell M., 2024, Unifying approaches from statistical genetics and phylogenetics for mapping phenotypes in structured populations, *PLoS Biology*, 22(10): e3002847.  
<https://doi.org/10.1371/journal.pbio.3002847>
- Serpico D., Lynch K.E., and Porter T.M., 2023, New historical and philosophical perspectives on quantitative genetics, *Studies in History and Philosophy of Science*, 97: 29-33.  
<https://doi.org/10.1016/j.shpsa.2022.11.009>

- Shafer A.B., Peart C.R., Tusso S., Maayan I., Brelsford A., Wheat C.W., and Wolf J.B., 2017, Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference, *Methods in Ecology and Evolution*, 8(8): 907-917.  
<https://doi.org/10.1111/2041-210X.12700>
- Speed D. and Balding D.J., 2014, MultiBLUP: improved SNP-based prediction for complex traits, *Genome Research*, 24(9): 1550-1557.  
<https://doi.org/10.1101/gr.169375.113>
- Speed D. and Balding D.J., 2019, SumHer better estimates the SNP heritability of complex traits from summary statistics, *Nature Genetics*, 51(2): 277-284.  
<https://doi.org/10.1038/s41588-018-0279-5>
- Spindel J., Begum H., Akdemir D., Virk P., Collard B., Redoña E., Atlin G., Jannink J., and McCouch S., 2015, Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines, *PLoS Genetics*, 11(2): e1004982.  
<https://doi.org/10.1371/journal.pgen.1004982>
- Steppan S.J., Phillips P.C., and Houle D., 2002, Comparative quantitative genetics: evolution of the G matrix, *Trends in Ecology & Evolution*, 17(7): 320-327.  
[https://doi.org/10.1016/S0169-5347\(02\)02505-3](https://doi.org/10.1016/S0169-5347(02)02505-3)
- Tatusova T., DiCuccio M., Badretdin A., Chetvermin V., Nawrocki E., Zaslavsky L., Lomsadze A., Pruitt K., Borodovsky M., and Ostell J., 2016, NCBI prokaryotic genome annotation pipeline, *Nucleic Acids Research*, 44(14): 6614-6624.  
<https://doi.org/10.1093/nar/gkw569>
- Technow F., Messina C.D., Totir L.R., and Cooper M., 2015, Integrating crop growth models with whole genome prediction through approximate Bayesian computation, *PLoS ONE*, 10(6): e0130855.  
<https://doi.org/10.1371/journal.pone.0130855>
- Van Eeuwijk F., Bustos-Korts D., Millet E., Boer M., Kruijer W., Thompson A., Malosetti M., Iwata H., Quiroz R., Kuppe C., Muller O., Blazakis K., Yu K., Tardieu F., and Chapman S., 2019, Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding, *Plant Science*, 282: 23-39.  
<https://doi.org/10.1016/j.plantsci.2018.06.018>
- Van Eeuwijk F.A., Bustos-Korts D.V., and Malosetti M., 2016, What should students in plant breeding know about the statistical aspects of genotype × environment interactions?, *Crop Science*, 56(5): 2119-2140.  
<https://doi.org/10.2135/cropsci2015.06.0375>
- Varshney R.K., Roorkiwal M., Sun S., Bajaj P., Chitikineni A., Thudi M., Singh N.P., Du X., Upadhyaya H.D., Khan A.W., Wang Y., Garg V., Fan G., Edwards D., and others, 2021, A chickpea genetic variation map based on the sequencing of 3,366 genomes, *Nature*, 599: 622-627.  
<https://doi.org/10.1038/s41586-021-04066-1>
- Verbrigghe N., Muylle H., Pegard M., Rietman H., Đorđević V., Čeran M., and Roldán-Ruiz I., 2025, Disentangling soybean G×E effects in an integrated genomic prediction and machine learning-GWAS workflow, *Plant Methods*, 21(1): 119.  
<https://doi.org/10.1186/s13007-025-01434-0>
- Viana J.M.S. and Garcia A.A.F., 2022, Significance of linkage disequilibrium and epistasis on genetic variances in noninbred and inbred populations, *BMC Genomics*, 23(1): 286.  
<https://doi.org/10.1186/s12864-022-08335-9>
- Vieira R.A., Nogueira A.P.O., and Fritsche-Neto R., 2025, Optimizing the selection of quantitative traits in plant breeding using simulation, *Frontiers in Plant Science*, 16: 1495662.  
<https://doi.org/10.3389/fpls.2025.1495662>
- Visscher P.M. and Bruce Walsh J., 2019, Commentary: Fisher 1918: the foundation of the genetics and analysis of complex traits, *International Journal of Epidemiology*, 48(1): 10-12.  
<https://doi.org/10.1093/ije/dyx129>
- Visscher P.M. and Goddard M.E., 2019, From RA Fisher's 1918 paper to GWAS a century later, *Genetics*, 211(4): 1125-1130.  
<https://doi.org/10.1534/genetics.118.301594>
- Visscher P.M., Brown M.A., McCarthy M.I., and Yang J., 2012, Five years of GWAS discovery, *The American Journal of Human Genetics*, 90(1): 7-24.  
<https://doi.org/10.1016/j.ajhg.2011.11.029>
- Visscher P.M., Wray N.R., Zhang Q., Sklar P., McCarthy M.I., Brown M.A., and Yang J., 2017, 10 years of GWAS discovery: biology, function, and translation, *The American Journal of Human Genetics*, 101(1): 5-22.  
<https://doi.org/10.1016/j.ajhg.2017.06.005>
- Voss-Fels K.P., Cooper M., and Hayes B.J., 2019, Accelerating crop genetic gains with genomic selection, *Theoretical and Applied Genetics*, 132(3): 669-686.  
<https://doi.org/10.1007/s00122-018-3270-8>
- Walsh B. and Lynch M., 2018, *Evolution and selection of quantitative traits*, Oxford University Press.  
<https://doi.org/10.1093/oso/9780198830870.001.0001>
- Walsh B., 2014, Special issues on advances in quantitative genetics: introduction, *Heredity*, 112(1): 1-3.  
<https://doi.org/10.1038/hdy.2013.115>
- Wang J. and Zhang Z., 2021, GAPIT version 3: boosting power and accuracy for genomic association and prediction, *Genomics, Proteomics & Bioinformatics*, 19(4): 629-640.  
<https://doi.org/10.1016/j.gpb.2021.08.005>

- Wang M.H., Cordell H.J., and Van Steen K., 2019, Statistical methods for genome-wide association studies, In *Seminars in Cancer Biology*, 55: 53-60.  
<https://doi.org/10.1016/j.semcancer.2018.04.008>
- Weissbrod O., Hormozdiari F., Benner C., Cui R., Ulirsch J., Gazal S., Schoech A., Van De Geijn B., Reshef Y., Márquez-Luna C., O'Connor L., Pirinen M., Finucane H., and Price A., 2019, Functionally informed fine-mapping and polygenic localization of complex trait heritability, *Nature Genetics*, 52(12): 1355-1363.  
<https://doi.org/10.1038/s41588-020-00735-5>
- Williams J., Xu S., and Ferreira M.A., 2023, BGWAS: Bayesian variable selection in linear mixed models with nonlocal priors for genome-wide association studies, *BMC Bioinformatics*, 24(1): 194.  
<https://doi.org/10.1186/s12859-023-05316-x>
- Xu Y., Zhang X., Li H., Zheng H., Zhang J., Olsen M., Varshney R., Prasanna B., and Qian Q., 2022, Smart breeding driven by big data, artificial intelligence, and integrated genomic-environmental prediction, *Molecular Plant*, 15(11): 1664-1695.  
<https://doi.org/10.1016/j.molp.2022.09.001>
- Yanez J.M., Barria A., Lopez M.E., Moen T., Garcia B.F., Yoshida G.M., and Xu P., 2023, Genome-wide association and genomic selection in aquaculture, *Reviews in Aquaculture*, 15(2): 645-675.  
<https://doi.org/10.1111/raq.12750>
- Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., and Visscher P.M., 2010, Common SNPs explain a large proportion of the heritability for human height, *Nature Genetics*, 42(7): 565-569.  
<https://doi.org/10.1038/ng.608>
- Yang J., Das Adhikari S., Wang H., Huang B., Qi W., Cui Y., and Wang J., 2024, De novo prediction of functional effects of genetic variants from DNA sequences based on context-specific molecular information, *Frontiers in Systems Biology*, 4: 1402664.  
<https://doi.org/10.3389/fsysb.2024.1402664>
- Yang J., Lee S.H., Goddard M.E., and Visscher P.M., 2011, GCTA: a tool for genome-wide complex trait analysis, *The American Journal of Human Genetics*, 88(1): 76-82.  
<https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yang J., Zhu J., and Williams R.W., 2007, Mapping the genetic architecture of complex traits in experimental populations, *Bioinformatics*, 23(12): 1527-1536.  
<https://doi.org/10.1093/bioinformatics/btm143>
- Yin L., Zhang H., Tang Z., Yin D., Fu Y., Yuan X., Li X., Liu X., and Zhao S., 2023, HIBLUP: an integration of statistical models on the BLUP framework for efficient genetic evaluation using big genomic data, *Nucleic Acids Research*, 51(8): 3501-3512.  
<https://doi.org/10.1093/nar/gkad074>
- You X., Shan X., and Shi Q., 2020, Research advances in the genomics and applications for molecular breeding of aquaculture animals, *Aquaculture*, 526: 735357.  
<https://doi.org/10.1016/j.aquaculture.2020.735357>
- Yu J., Pressoir G., Briggs W., Bi I., Yamasaki M., Doebley J., McMullen M., Gaut B., Nielsen D., Holland J., Kresovich S., and Buckler E., 2006, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, *Nature Genetics*, 38(2): 203-208.  
<https://doi.org/10.1038/ng1702>
- Zargar S., Raatz B., Sonah H., Bhat J., Dar Z., Agrawal G., and Rakwal R., 2015, Recent advances in molecular marker techniques: insight into QTL mapping, GWAS and genomic selection in plants, *Journal of Crop Science and Biotechnology*, 18(5): 293-308.  
<https://doi.org/10.1007/s12892-015-0037-5>
- Zhang Z., Ersoz E., Lai C.-Q., Todhunter R.J., Tiwari H.K., Gore M.A., Bradbury P.J., Yu J., Arnett D.K., Ordovas J.M., and Buckler E.S., 2010, Mixed linear model approach adapted for genome-wide association studies, *Nature Genetics*, 42(4): 355-360.  
<https://doi.org/10.1038/ng.546>

**Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.