Computational Molecular Biology 2026, Vol.16, No.3, 159-180
http://bioscipublisher.com/index.php/cmb

Research Article Open Access

Genome-wide Relationship Matrix-Based Heritability Estimation: Statistical Interpretation, Comparability, and Practical Diagnostics in the GCTA-GREML Framework

Running title: Interpreting SNP Heritability with GCTA-GREML

Xuanjun Fang
Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025
Corresponding author: xuanjunfang@hitar.org

Computational Molecular Biology, 2026, Vol.16, No.3 doi: 10.5376/cmb.2026.16.0012
Received: 05 Apr., 2026
Accepted: 08 May, 2026
Published: 20 May, 2026

Preferred citation for this article:
Fang X.J., 2026, Genome-wide relationship matrix-based heritability estimation: statistical interpretation, comparability, and practical diagnostics in the GCTA-GREML framework, Computational Molecular Biology, 16(3): 159-180 (doi: 10.5376/cmb.2026.16.0012)

Abstract Heritability, as a core concept, plays a critical role in explaining trait variation and predicting selection response. Traditional heritability estimation relies on pedigree information but is limited by pedigree completeness and environmental confounding. With the development of high-throughput genotyping and genome-wide association studies, the restricted maximum likelihood method based on genomic relationship matrices (GCTA/GREML) has provided a new pathway for estimating the heritability of complex traits. This study reviews the theoretical framework and statistical assumptions of the GCTA and GREML families, elucidates their logic in variance decomposition and differences from pedigree-based models, and focuses on analyzing the sources and interpretive boundaries of the "missing heritability" problem. Further, the study explores methodological extensions such as the LOCO strategy, functional annotation partitioning, and bivariate analysis, and discusses their application value in complex trait dissection and crop breeding, supported by both simulation and empirical studies. The results indicate that GCTA/GREML not only promotes a paradigm shift in heritability research but also provides theoretical support for genomic selection and molecular breeding design. In the future, with the accumulation of sequencing data and multi-environment big data, this method is expected to more comprehensively uncover the genetic basis of complex traits.

Accordingly, this review focuses on clarifying the statistical interpretation of SNP-based heritability estimation rather than providing a general tutorial. Specifically, we (i) outline the statistical conditions required for meaningful comparisons between SNP-based and pedigree-based heritability estimates; (ii) formally define the estimand targeted by GREML and clarify its relationship to the concept of missing heritability; (iii) organize commonly used GREML extensions into a unified framework based on their inferential goals, assumptions, and diagnostic boundaries; and (iv) propose a workflow-oriented checklist to guide the interpretation of SNP heritability estimates in practice.

Keywords SNP heritability; Genome-wide relationship matrix (GRM); GCTA-GREML; Missing heritability; Statistical interpretation; Diagnostic workflow

1 Introduction

Heritability, as a central concept in quantitative genetics, has played a fundamental role in explaining the sources of trait variation and guiding genetic improvement practices since Fisher established the framework of analysis of variance. It is defined as the proportion of phenotypic variance that can be attributed to genetic differences, and serves as a cornerstone concept in both quantitative genetics and breeding science (Yang et al., 2017; Srivastava et al., 2023). In breeding, heritability not only provides a quantitative scale for evaluating the potential for trait selection, but also constitutes a key parameter for predicting selection response and optimizing population improvement strategies (Zhu and Zhou, 2020). High heritability indicates that genetic variation accounts for a large proportion of phenotypic variance, thereby leading to higher efficiency of artificial selection; conversely, low heritability suggests a dominant role of environmental variation and consequently limited selection effectiveness. Therefore, whether in the design of crop and livestock breeding strategies or in the genetic epidemiology of complex human diseases, accurate estimation of heritability remains an unavoidable core issue at both theoretical and practical levels (Yang et al., 2017).