



XVI Simpósio Brasileiro de Melhoramento Animal

Piracicaba, SP – 04 e 05 de agosto de 2025

Ant colony optimization for epistasis detection in simulated complex traits

Mateus G. dos Santos^{*1}, Cassiane G dos Santos¹, Marcelo J Bock¹, Renata Veroneze¹, Ricardo S. Ferreira¹,
Marcos Vinicius G. B. da Silva², Simone E. F. Guimarães¹, Guilherme J. R. Rosa³

¹Universidade Federal de Viçosa, Viçosa/MG, Brasil; ²Embrapa Gado de Leite, Juiz de Fora/MG; ³University of Wisconsin-Madison, Madison/WI.

^{*}Autor correspondente – mateus.g.santos@ufv.br

Epistasis, or gene-gene interaction, is fundamental to the expression of complex traits in plants and animals, including livestock. However, its detection in genome-wide association studies remains difficult due to the vast number of possible SNP combinations and typically small effect sizes, especially in traits with low additive variance. This study evaluates the performance of an Ant Colony Optimization (ACO)-based approach for identifying epistatic interactions associated with a simulated binary trait. ACO is a metaheuristic that explores combinatorial spaces using probabilistic search guided by heuristic and pheromone-based feedback. Here, a composite fitness function combining mutual information and the Gini score evaluates SNP combinations in relation to phenotypes, penalizing suboptimal solutions to refine the search. Simulated genotypic and phenotypic data were generated in AlphaSimR for 2,000 animals. We simulated a genome with one chromosome, 1 Morgan, and 100 SNPs. Traits with low heritability ($h^2 = 0.05$), and three levels of additive-by-additive to additive variance ratios ($raa = 0.05, 0.10, 0.15$) were simulated. SNP panels were LD-pruned, removing SNPs with $r^2 \geq 0.30$ to evaluate performance with and without LD structure. Two genetic scenarios were tested: 10 quantitative trait loci (QTL) with 5 interactions; and 30 QTLs with 15 interactions, each with 500 replicates. Performance of ACO was compared to logistic regression (LR) using a 5% false discovery rate and evaluated using recall, precision, and F1-score. A result was considered a true positive if the method detected either the exact QTL or an SNP in LD ($r^2 \geq 0.30$). A binomial generalized linear model with logit link was used to analyze the factorial experiment at a 5% significance level. For exact detection, LD pruning based on the highest marginal effect consistently improved the performance of recall and precision, increasing by 562% and 865%, respectively, compared to analyses using raw SNP files. LR was particularly sensitive to LD structure, showing a 158% improvement in F1 score when LD was removed. In contrast, the ACO algorithm was largely unaffected by LD structure or the raa , likely due to its use of nonlinear functions that do not explicitly depend on these factors. In LD-based detection, LD pruning once again improved overall performance, increasing precision and F1 scores by reducing false positives, especially for LR, which was highly sensitive to LD and performed poorly when LD remained. Recall, however, declined as the number of interactions increased, likely due to dilution of the interaction signal, further affecting LR's performance. ACO, on the other hand, consistently outperformed LR in these more complex scenarios. Although ACO occasionally showed slightly lower metrics than LR, it offered significant advantages in runtime and memory efficiency. Unlike LR, ACO does not require exhaustive pairwise comparisons or prior LD pruning, and returned better results for key metrics such as F1-score. In many cases where LR failed to detect epistatic combinations, ACO still identified meaningful associations, highlighting its robustness. Due to its efficiency and low computational demands, ACO is a viable approach for traits with low heritability and complex genetic architecture in livestock.

Keywords: additive-by-additive epistasis, metaheuristics, low heritability.

Acknowledgements: High-performance Computing System of the Universidade Federal de Viçosa