

Comparison of GWAS results using different phenotypic statistic methods for a small grape population



Huimin Zhang^{1,2}, Lei He^{1,2}, Qihong Pan^{1,2*}

¹College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, 100083, China

²Key Laboratory of Viticulture and Enology, Ministry of Agriculture and Rural Affairs, Beijing 100083, China

*Corresponding author. Email: panqh@cau.edu.cn

INTRODUCTION

GWAS has been widely used in the study of dissecting complex quantitative traits in plants, which generally requires a association population including a large number of individuals. In this study, a genome-wide association study was conducted using an F1 population consisting of only 81 and 97 hybrids' Linalool contents in two years respectively, which was too small for GWAS analysis. We performed several pretreating methods of phenotype data using linear model and compared result with the original phenotypic data, the objective of which is to reveal the important SNPs masked by noise in small population.

RESULTS

Phenotypic data distribution of different statistic methods

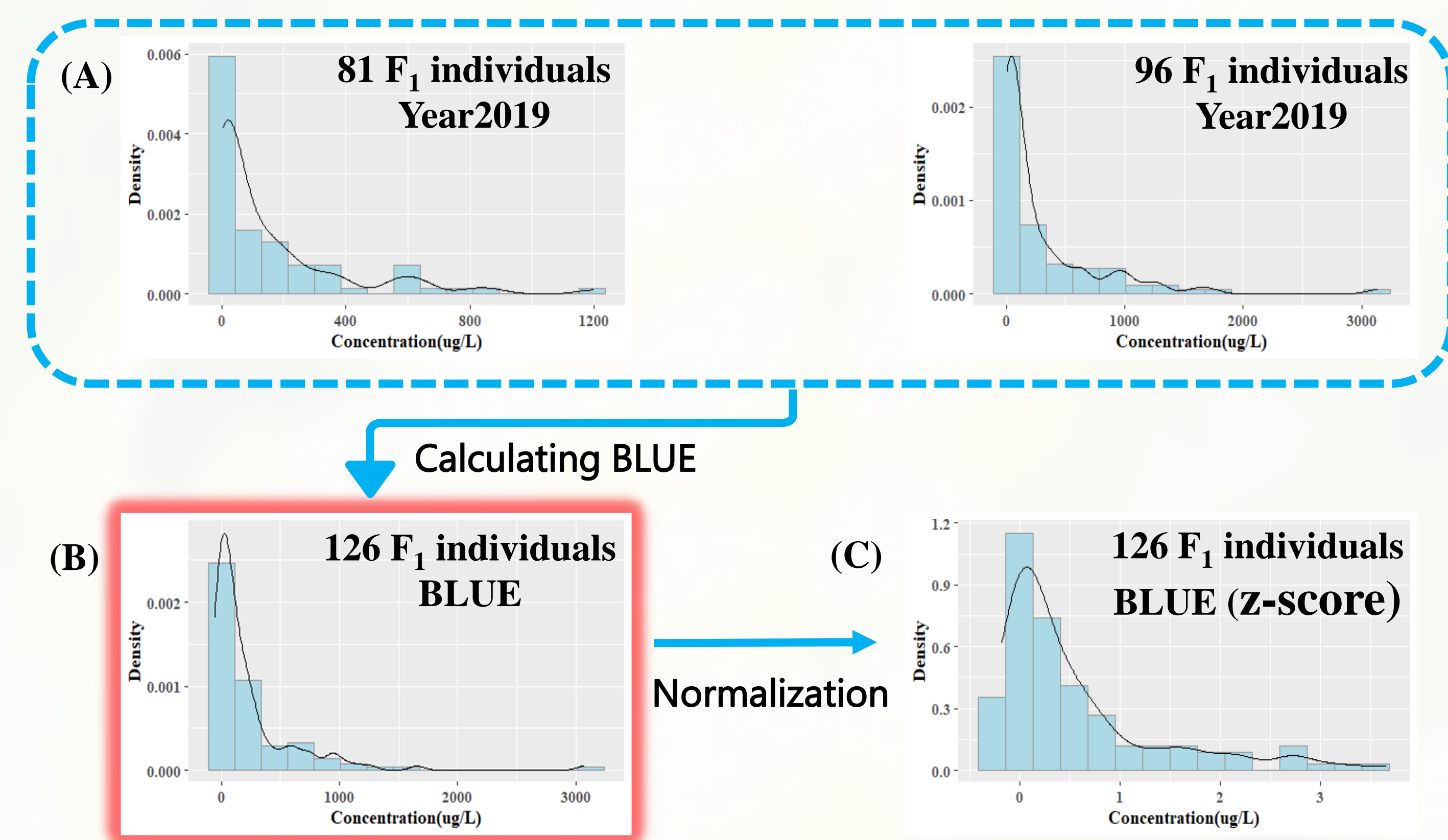


Fig. 1. Distribution of phenotypes used in comparison (A) Original phenotypic records of year 2019 and 2020 separately, (B) BLUE, (C) BLUE after Normalized using z-score.

METHODS

Population



P1
'Muscat Bianco'
Muscat cultivar

Year 2019
Year 2020



P2
'Petit Manseng'
Non-muscat cultivar

81 authentic F1 individuals
96 authentic F1 individuals

phenotypic statistic methods

BLUE (Best Linear Unbiased Estimator)

— genotypic values calculated as fixed effects

Z-score

— transform the phenotypic data to normal distribution

GWAS model-MLM

- Perform Mixed Linear Model in GEMMA
- Use 3 PCs as covariates
- Calculate Relatedness Matrix as random effect

$$Y = \text{SNP} + Q \text{ (or PCs)} + \text{Kinship} + e$$

(fixed effect) (fixed effect) (random effect)

General Linear Model (GLM)
Mixed Linear Model (MLM)

Manhattan Plots and Q-Q plots of GWAS

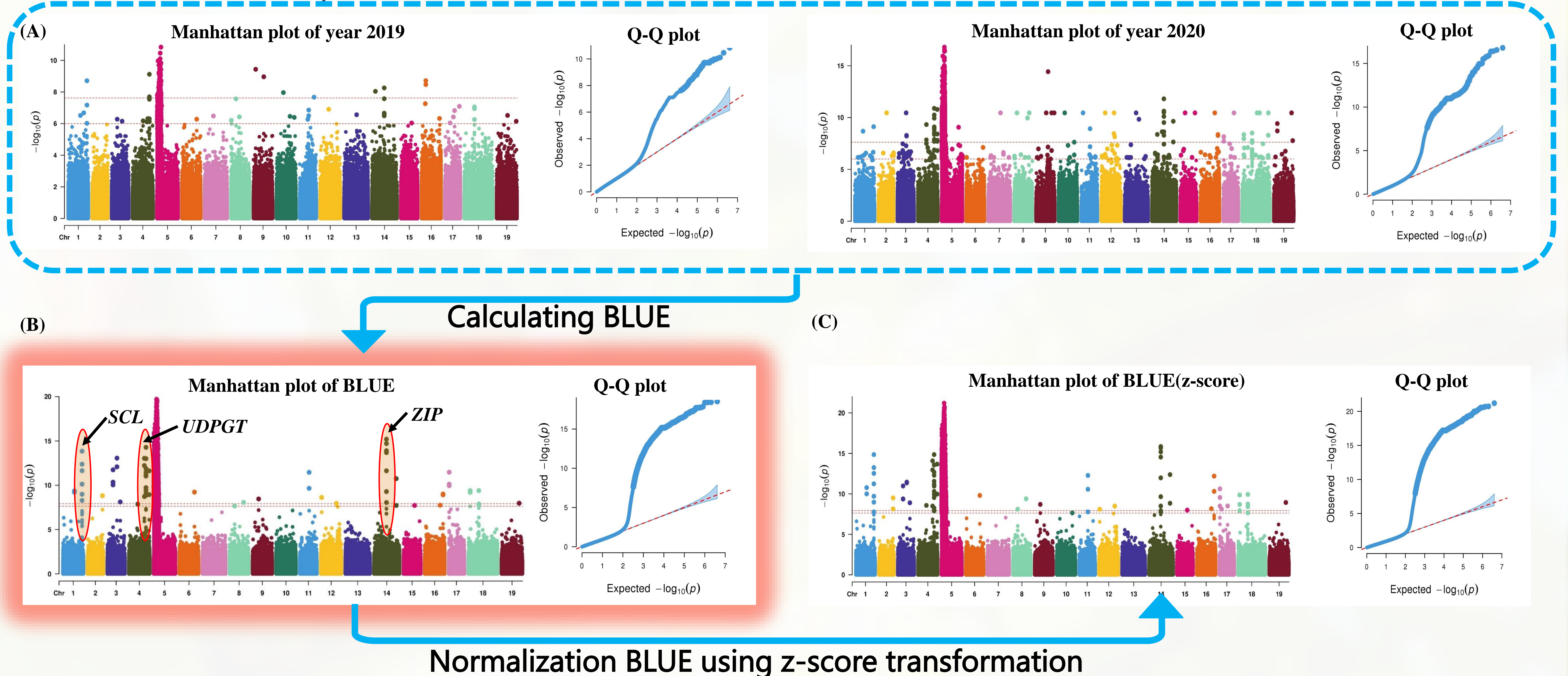


Fig. 2. Manhattan plots and Q-Q plots of different phenotypic statistic methods (A) results of year 2019 and 2020 separately, (B) result of BLUE using a larger dataset of 126 individuals, (C) result of BLUE after Normalized using z-score.

CONCLUSION

- Using BLUE instead of simple phenotypic records in a small population could get more information about trait-related SNPs and candidate genes.
- Skewed distribution of phenotype might not influence GWAS outcome.

ACKNOWLEDGEMENT

This research was supported by China Agriculture Research System of MOF and MARA (CARS-29).