Genetic analysis of phenotypic variation measured in the Oregon Wolfe Barley collection

Halide Aydin, Aliz Fodor, Ingo Giebel, and Alessia Russo

QBio305: Population and Quantitative Genetics

Institute of Plant Genetics, Heinrich-Heine-Universität Düsseldorf

Prof. Dr. Maria von Korff Schmising, Dr. Timo Hellwig, Dr. Agatha Walla

December 22, 2023

Contents

List of tables

List of figures

1 Introduction

Task definition: Paper about the phenotypic variation measured in the Oregon Wolfe Barley collection. This paper has to follow the structure of a scientific journal paper and should include the following analyses:

- ANOVA of plant height You find the data in

`dta---ANOVA_QTLanalysis.csv`

, which includes the measurements of all groups. The group is actually not a replicate, because you just measured the same plant several times. However, for the purpose of this exercise pretend that the group is a valid replicate, and do not include it in your model.

- Genetic linkage mapping of the qualitative trait your group has scored (which marker is in closest linkage?)

- QTL analysis of plant height You find the data in

`dta---ANOVA_QTLanalysis.csv`

.

Submit your paper as a pdf to agatha.walla@hhu.de and timo.hellwig@hhu.de. Deadline: December 22, 2023

Introduction: The Oregon Wolfe Barley collection. Text of 500 words describing the OWB population. See https://barleyworld.org/owb and (Costa et al., 2001).

Linkage maps and mapping qualitative traits: Text of 200 words.

Quantitative traits and QTL mapping Text of 200 words.

2 Materials and methods

Plant population and phenotyping: How large was your population? Which traits did we score, and how did we score them in detail?

Side note: the plants were 2 month old.

Statistics: Which statistical methods did we use and how did we implement them? Which are potential problems with these approaches? Which software/packages and versions did we use?

3 Results

3.1 Analysis of variance for the genotype effect on the plant height

To investigate the association between the height of the plants and their genotype, we conducted a single-factor analysis of variance (ANOVA). The analysis revealed a highly significant impact of the genotype on the plant height (p-value < 2×10^{-16} for both the raw and the cleansed data in which three outliers were removed). As the corresponding F-value for the cleansed data (834.4) is substantially greater than the corresponding F-value for the raw data (198.1), the actual p-value for the cleansed data is actually even much smaller than the p-value for the raw data. Figure 1 shows density plots of the raw height data for the different genotypes.

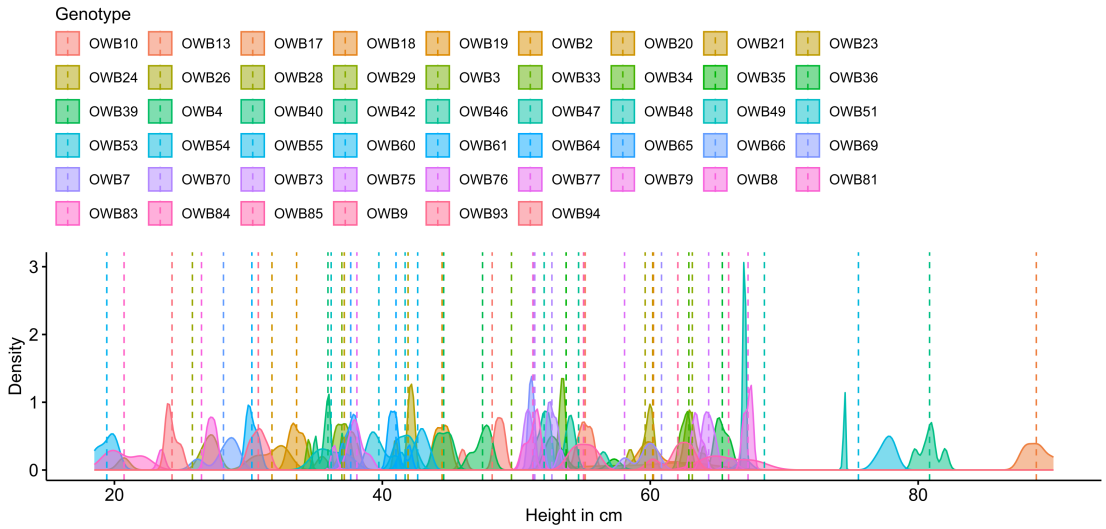


Figure 1: Density plots of the raw height data for the different genotypes

3.2 ANOVA homogeneity of variance assumption

ANOVA assumes that the data are normally distributed and that the variance across the groups is approximately equal. These assumptions were assessed with some diagnostic plots and tests. Figure 2 shows the ANOVA residuals (raw data) versus the fitted values. This plot clearly shows three outliers (in the genotypes OWB33, OWB76, and OWB54) which were later removed in the cleansed data because they severely affected the normality of the variance.

Furthermore, a Levene's test for homogeneity of variance was done. The p-value for that test was 0.7938, which means that the variances across the groups were not significantly different from each other, i.e., the homogeneity assumption of the variance was met.

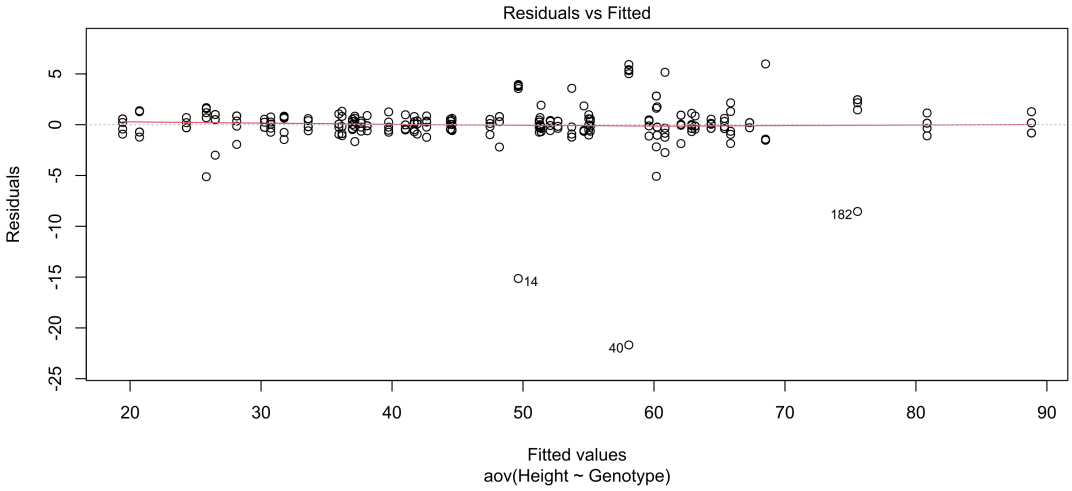
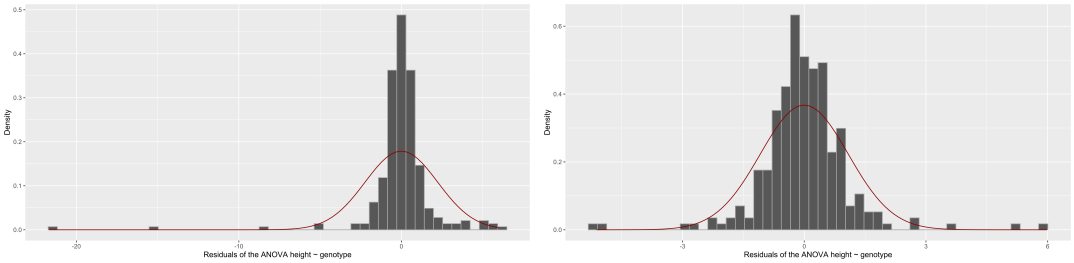


Figure 2: ANOVA residuals vs fitted values. Used for visual assessment of the homogeneity of variance assumption and to identify any outliers.

3.3 ANOVA normality of variance assumption

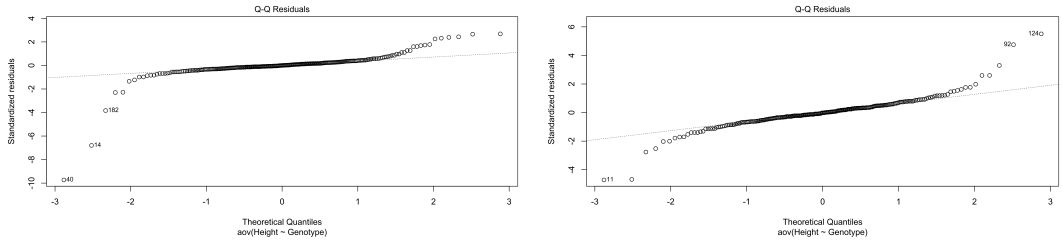
The normality of variance was assessed for both the raw data and the cleansed data in which the three most extreme outliers were removed. Figures 3, 4, and 5 show a visual comparison of the raw and the cleansed data: a histogram of the ANOVA residuals, a Q-Q plot of the ANOVA residuals, and a Cook's distance plot of the ANOVA residuals, respectively.

Furthermore, a Shapiro-Wilk test for normality was done for both the raw and the cleansed data. According to the p-values of the Shapiro-Wilk test (raw data < 2.2×10^{-16}, cleansed data 7.823×10^{-14}), both the raw and the cleansed data was not normally distributed in a narrow sense. Nevertheless, the plots reveal that the data conform at least roughly to a normal distribution, in particular the cleansed data without the three most extreme outliers.



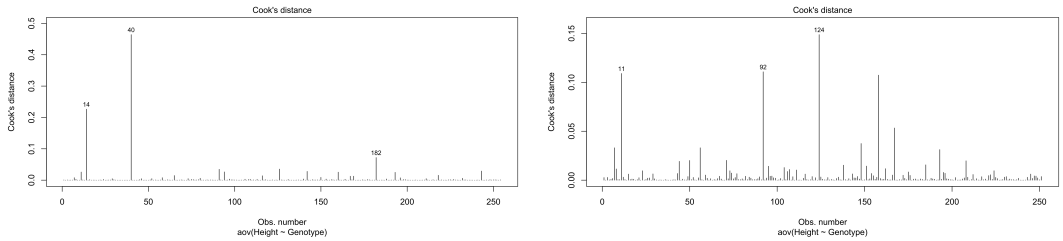
i) Raw data **ii)** Cleansed data

Figure 3: Histogram of the ANOVA residuals



i) Raw data **ii)** Cleansed data

Figure 4: Q-Q plot of the ANOVA residuals



i) Raw data **ii)** Cleansed data

Figure 5: Cook's distance plot of the ANOVA residuals

3.4 QTL analysis of plant height

In order to assess the impact of the genetic markers on the plant height, an ANOVA was done for all markers. Our chosen significance threshold of $5\,\%$ was Bonferroni corrected by dividing the threshold by the number of tests.

The ANOVAs showed that the plant height is mainly effected by a genomic region within gene 2H. But there are also markers within the genes 1H, 3H, 5H, and 6H which have a significant impact on the plant height. Figure 6 illustrates the impacts of the genetic markers on the plant height.

3.5 Genetic linkage mapping of the leaf variegation trait

In order to localize any genes determining the leaf variegation, a genetic linkage mapping of the leaf variegation was done. We found that the locus at position 192.8 of gene 2H near the marker scssr08447 highly influences the leaf variegation trait, having a logarithm of odds (LOD) score of 3.83. Figure 7 shows the genetic linkage map where the marker for the leaf variegation trait was moved to the position with the highest LOD score.

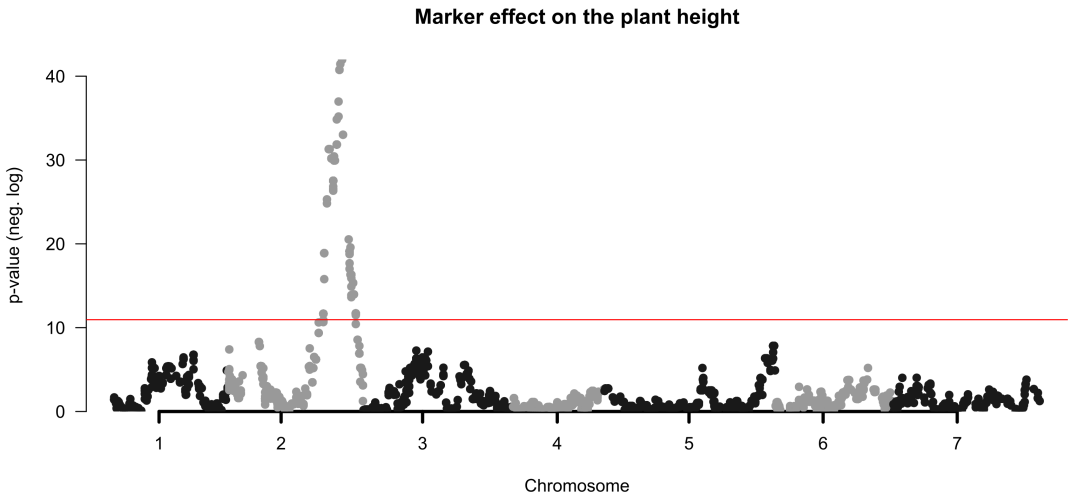


Figure 6: Manhattan plot of the genome-wide p-values for the genetic markers

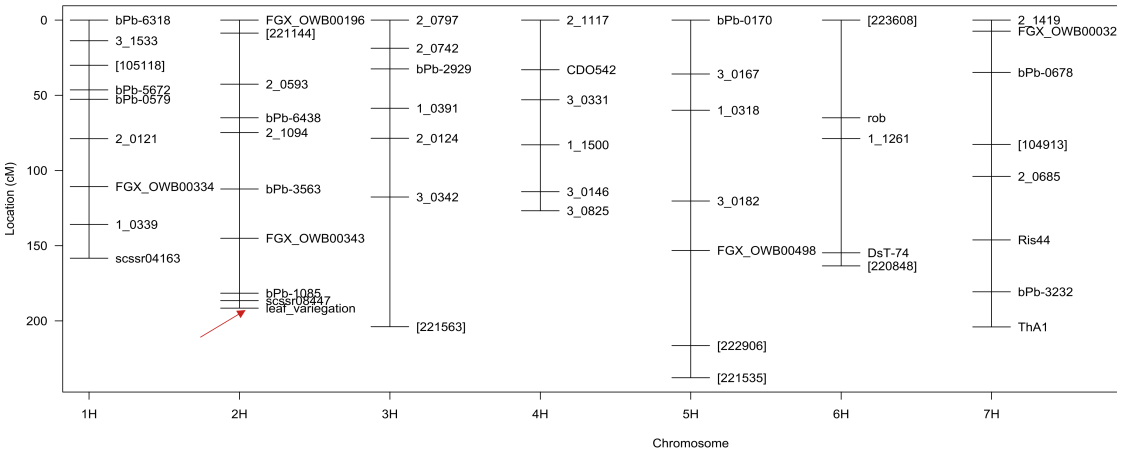


Figure 7: Genetic linkage mapping of the leaf variegation trait leaf_var (moved to the position with the highest LOD score of 3.83)

4 Discussion

TODO...

References

Costa, J. M., Corey, A., Hayes, P. M., Jobet, C., Kleinhofs, A., Kopisch-Obusch, A., Kramer, S. F., Kudrna, D., Li, M., Riera-Lizarazu, O., Sato, K., Szucs, P., Toojinda, T., Vales, M. I., & Wolfe, R. I. (2001). Molecular mapping of the oregon wolfe barleys: A phenotypically polymorphic doubled-haploid population. *Theoretical and Applied Genetics*, *103*(2-3), 415–424. https://doi.org/10.1007/s001220100622

A Appendix A

The first appendix...

B Appendix B

The second appendix...

C Appendix C

The third appendix...

Index