

Biometry and Population Genetics

Thesis topics
6. September 2023

Note that all thesis topics are described in English but you can of course write your thesis in German as well if nothing else is stated in the description. All our theses require that you have basic knowledge in R programming from our core modules. Please contact the potential supervisor for more information on the topics. You can start anytime if not stated otherwise.

Marker imputation with machine learning

With the reduced cost and increased prevalence of genotyping, companies are accumulating large amounts of marker data. The ongoing development of SNP arrays has resulted in different arrays with a limited number of common markers. Therefore, one challenge is to use datasets generated with different SNP arrays in a common genomic selection approach.

In your thesis, you will start by creating a small literature overview of marker imputation. The core of the thesis will be the application, comparison and evaluation of different imputation methods, both traditional and based on machine learning.

Target group: Master students of Nutzpflanzenwissenschaften or Agrobiotechnology

Supervisor: Philipp Heilmann (philipp.g.heilmann@agrar.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN or MP-236-EN completed successfully, strong interest in programming, ideally already some experience with machine learning.

Genome-wide prediction of resistance scores in wheat

Selection has been used in plant breeding for a very long time in order to identify the best individuals from a breeding population. The availability of marker data expanded the possibilities for selection: Markers can be used to estimate the genotypic value of individuals and select those with the highest estimated genotypic value. For example, from 300 selection candidates, 100 are evaluated for their yield in field trials. They are then genotyped with 5,000 SNP markers and an effect is assigned to each of the markers. The other 200 individuals then only have to be genotyped and their potential yield is estimated from the marker effects without the need to conduct any additional field trials. This is called genomic selection or genome-wide prediction.

Resistance scores are a challenge for genome-wide prediction because the standard methods are best suited for data on a metric scale. Resistance scores, however, are often on an ordinal scale. Transformation of resistance scores, for example the logit or arcsin transformation, can be used to improve predictions. The focus of your thesis will be on the comparison of different transformation methods on the accuracy of predictions.

Target group: Bachelor students of Agrarwissenschaften

Supervisor: Dr. Carola Zenke-Philippi (carola.zenke@agrار.uni-giessen.de)

Requirements: BK-005 completed successfully.

Simulation study: Haplotype-based cross-planning in wheat, barley, or oat

Selection has been used in plant breeding for a very long time in order to identify the best individuals from a breeding population and improve the performance of crops by producing progeny only from those. The availability of marker data expanded the possibilities for selection: Markers can be used to estimate the genotypic value of individuals and select those with the highest estimated genotypic value. For example, from 300 selection candidates, 100 are evaluated for their yield in field trials. They are then genotyped with 5,000 SNP markers and an effect is assigned to each of the markers. The other 200 individuals then only have to be genotyped and their potential yield is estimated from the marker effects without the need to conduct any additional field trials. This is called genomic selection or genomewide prediction.

Data from genomic selection can also be used in order to plan crosses: The marker effects are used to identify the combinations of genotype with the most promising offspring, i.e. those offspring which combines the most positive alleles for the trait of interest.

In your thesis, you will investigate the potential of haplotype blocks for this purpose. Haplotype blocks can be used instead of single SNPs to make the predictions. Here is how they work: The genome is partitioned in many segments, so-called haplotype blocks, based on different criteria. The marker effects are then estimated for the blocks, not for the SNPs. It remains unclear, however, how the haplotype blocks should be ideally built and if they really have advantages over the use of single SNPs.

You will conduct a simulation study with the R package SelectionTools with a data set with either wheat, barley, or oat lines.

Target group: Master students of Nutzpflanzenwissenschaften or Agrobiotechnology

Supervisor: Dr. Carola Zenke-Philippi (carola.zenke@agrار.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN or MP-236-EN completed successfully.

Literature to start with: Osthusenrich et al. 2017 PLoS ONE 12(12): e0188839, Osthusenrich et al. 2018 Frontiers in Plant Science 9:1899, Difabachew et al. 2023 Frontiers in Plant Science 14:1168547

Haplotype stacking in wheat

Selection has been used in plant breeding for a very long time in order to identify the best individuals from a breeding population and improve the performance of crops by producing progeny only from those. The availability of marker data expanded the possibilities for selection: Markers can be used to estimate the genotypic value of individuals and select those with the highest estimated genotypic value. This is called genomic selection or genomewide prediction. Another alternative is to identify chromosome stretches (haplotype blocks) that carry a positive net effect on the trait of interest and try to select genotypes that carry as many haplotype blocks with positive effects as possible (“haplotype stacking”).

The project “HaploSelekt” aims to apply haplotype stacking in wheat. In your thesis, you will investigate which haplotype blocks are promising candidates for selection. You will work with real data from the project. The calculations will be done in R.

Target group: Master students of Nutzpflanzenwissenschaften or Agrobiotechnology

Supervisor: Dr. Carola Zenke-Philippi (carola.zenke@agrار.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN or MP-236-EN completed successfully.

Literature to start with: Difabachew et al. 2023 *Frontiers in Plant Science* 14:1168547

Grain protein deviation in wheat

The price that a farmer gets for one dt of wheat strongly depends not only on the grain moisture but also on the protein content. Unfortunately, grain yield and grain protein concentration are negatively correlated: When one goes up, the other one goes down. “However, some wheat genotypes consistently deviate from this relationship, a phenomenon known as Grain Protein Deviation (GPD) (Monaghan et al., 2001). GPD is defined as the deviation from the negative regression between grain protein and yield as described by Monaghan et al. (2001). GPD is negative in wheats developed to have high contents of starch and low contents of protein for distilling or livestock feed, and positive in some high protein wheat genotypes bred for breadmaking” (Mosleth et al. 2020). This means that a high PPD is desirable in many elite wheat lines.

In your thesis, you will characterize the GPD of two wheat populations from ongoing breeding programs and examine its relationship to other yield parameters.

Target group: Bachelor students of Agrarwissenschaften

Supervisor: Dr. Carola Zenke-Philippi (carola.zenke@agrار.uni-giessen.de)

Requirements: BK-005 completed successfully, BP-041 is recommended but not mandatory.

Literature to start with: Monaghan et al. 2001 *Euphytica* 122:309–317, Mosleth et al. 2020 *Field Crops Research* 255:107896

Different measures for heritability in wheat and/or oat

There are different definitions for the heritability. They all aim to somehow quantify how much of the observed phenotypic variability in a field trial can be attributed to genetics. However, this can be difficult to measure, particularly if the field trials under consideration are unbalanced. Different measures for the heritability have been proposed as a solution to this problem.

In your thesis, you will investigate different measures for the heritability, calculate them for data from ongoing breeding programs and wheat and/or oat and compare their properties.

Target group: Master students of Nutzpflanzenwissenschaften or Agrobiotechnology

Supervisor: Dr. Carola Zenke-Philippi (carola.zenke@agrار.uni-giessen.de)

Requirements: MK-119-EN and either MK-002 or MK-002-EN completed successfully.

Literature to start with: Visscher et al. 2008 *Nature Reviews Genetics* 9: 255–266, Piepho and Möhring 2007 *Genetics* 177: 1881–1888