

Biometry and Population Genetics

Thesis topics
12. Januar 2026

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.

Genome-wide association study of stem length in Carnation

Stem length is a determinant factor in cut flower quality control. Important traits in cut flower production, like harvesting time and vase life are directly influenced by stem length (Varu *et al.*, 2010). Therefore, longer, stronger and faster growing stems are part of the breeding goals for cut flower breeders.

Carnation is part of the top economically important ornamental crops in the cut flower trade. Despite this, and similar to other important cut flower cultures like Rose and Chrysanthemum, research based on genomic resources is still underutilized in ornamental plant breeding (Schultz *et al.*, 2016). However, this is changing with the increasing availability of genetic marker and genomic data.

A well know method that exploits genetic marker data is genome-wide association studies (GWAS). Here, the idea is to pair phenotypic data of a specific trait (e.g. stem length) with genetic marker data from a given set of individuals. Differences in the phenotype and genotype between individuals are used to detect which, if any, of the evaluated markers shows a strong relation (association) to the trait under observation. This approach is useful for understanding which and how many regions of the subject species' genome can be influencing a given trait.

In your thesis, you will conduct a genome-wide association study with a cut Carnation data set to investigate underlying genetic factors of stem length as a trait.

Target group: Master students of Nutzpflanzenwissenschaften (Crop Science), Informationstechnologie, or Agrobiotechnology.

Supervisor: Hugo Tavera (hugo.tavera@agrار.uni-giessen.de).

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN (would be ideal), or MP-240-EN completed successfully.

Focus: R programming

Literature to start with: For GWAS in ornamentals, Schultz et al., 2016. *Frontiers in Plant Science* 7:1798 and Tavera et al., 2025 *Frontiers in Plant Science* 16:1673111. For the trait "stem length", Varu, D. K., & Barad, A. V., 2010. *Journal of horticultural sciences*, 5(1), 42-47.

Master thesis: Genomic prediction of the segregation variance - approaches, improvements, applications and perspectives

In recent years, several analytical derivations of the segregation variance σ_g of a cross based on genotypic data have been published. The concept has since then been employed for selection of crossing partners and planning crosses with high performance and diversity. However, critical studies have highlighted current problems with prediction accuracy and suggested improvements for existing approaches.

In your thesis, you will write a comprehensive literature review on the topic of genomic prediction of the segregation variance and its applications in breeding programs. You should also discuss suggested modifications and future perspectives of the concept.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

Supervisor: Dr. Eva Herzog (eva.herzog@agrar.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN or MP-240-EN completed successfully, basics of plant breeding, good skills in maths

Work focus: Literature research, mathematical formulas for quantitative genetics

Literature to start with: Osthushenrich et al. 2017 PLoS One 12.12: e0188839; Lehermeier et al. 2017 Genetics 207.4 : 1651-1661

Master or Bachelor thesis: Development of models for conversion of physical maps to linkage maps in wheat

For many applications in population genetics such as simulation studies and cross planning, recombination frequencies between genotypic loci are required. Information on the recombination frequency is usually presented in the form of a linkage map in centiMorgans (cM). However, while physical marker positions are often available for many DNA marker assays, linkage maps are rare.

In your thesis, you will collect different publically available consensus maps of the wheat genome, prepare the data and fit different prediction models (KNN, sliding windows, splines etc.) that take physical positions in bP as inputs and return a linkage map suitable for population genetical simulation.

Target group: Bachelor students of Agricultural Science, and master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

Supervisor: Dr. Eva Herzog (eva.herzog@agrar.uni-giessen.de)

Requirements: Bachelor level: BP-041 Biostatistics, master level: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN, or MP-240-EN completed successfully, all: R programming skills, willingness to learn use of new R packages.

Work focus: Data management, R programming, algorithm development

Literature to start with: Qu et al. 2021 Frontiers in Plant Science 12:727077 and supplementary material (get an idea of the type of data set), potentially chapters on *k nearest neighbors* and splines in the following book: <https://www.statlearning.com/>

Master thesis: Representative subset selection in an elite winter wheat panel

From a panel of 191 elite winter wheat lines representing 50 years of breeding progress, a representative subset is to be selected as a training set for genomic selection in a cross-validation study. The main objective of this thesis is to compare different methods of representative subset selection such as partitioning around medoids (PAM) and maximum connectedness and diversity (MaxCD) and others, and their effects on prediction accuracy of genomic selection. Scenarios which will serve as benchmark for the performance of the new methods are a) random selection and b) an established selection method based on principal component analysis.

Your task will be to implement the different methods of representative subset selection in R code, mostly using available functions from R packages, to program and run the cross-validation routine for different sizes of the training and the validation set, and to evaluate and compare the methods by different criteria (e.g. prediction accuracy, mean square error, ranking of selection candidates).

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, Bioinformatik und Systembiologie, or Agrobiotechnology

Supervisor: Dr. Eva Herzog (eva.herzog@agrار.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN or MP-240-EN completed successfully, R programming skills, willingness to learn use of new R packages.

Work focus: R programming, application of unsupervised learning methods

Literature to start with: Voss-Fels et al. 2019 Nature Plants 5(7) 706-714 (dataset), Guo et al. 2019 Molecular Plant 12, 390-401 (method)

Master thesis: Validating the accuracy of published Deep Learning architectures for genomic prediction

In the recent years, a considerable number of papers have been published presenting novel ideas and architectures for applying Neural Networks to genomic prediction in different plant species. However, validation of these models by extensive cross validation is often lacking. In this thesis, the objective will be to recreate a neural network as described in a paper that will be provided to you. Then, a cross validation study based on this Neural Network will be conducted to compare it to an industry standard method, GBLUP.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

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

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-163-EN-DI, MP-236-EN, or MP-240-EN completed successfully. Ideally a solid background in programming (Python).

Focus: Programming

Master thesis: The Influence of Phenotypic Data Curation Method on Genomic Prediction Accuracy in Winter Oilseed Rape Hybrid Breeding

This thesis examines how different phenotyping approaches affect genomic prediction accuracy in winter oilseed rape (*Brassica napus* L.) hybrids. Different agronomic traits will be evaluated based on data collected by technical staff, the breeder, and drone-based digital phenotyping. The student will evaluate data consistency, estimate trait heritability, and perform genomic prediction analyses to quantify the effect of measurement method on prediction accuracy. Additionally, the development of an integrated index combining information from all phenotyping methods holds potential to improve prediction accuracy and selection decisions. Your work aims to identify reliable, efficient phenotyping strategies that optimize prediction accuracy in hybrid breeding programs.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

Supervisor: Lars Erik Thomsen (lars.e.thomsen@agrar.uni-giessen.de)

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN, or MP-240-EN completed successfully. R programming skills and interest in applied plant breeding.

Focus: R programming

Literature to start with: Walter et al. 2015 Plant Methods 11(14)

Master thesis: Construction of a genetic map from a non-biparental breeding population

In the past, genomic selection was not possible for neglected crops due to a lack of genotyping capabilities. Fortunately, genotyping by sequencing now allows us to genotype crops without the need for a specific SNP chip. While SNP chips only recognize specific markers for which the recombination frequencies in genetic maps are known, genotyping by sequencing recognizes random SNPs for which we only know the position in base pairs (physical map). If we are interested in a possible recombination of a specific cross, we can approximate the recombination frequency from the physical map. However, recombination events around the centromere are rare (Fernandes et al. 2024) and occur more frequently in other parts of the genome. The calculated variance of a potential offspring can be misleading if we do not take this fact into account.

Since our population comes from a breeding population in which the genotypes are not the offspring of a single biparental cross, methods for developing a genetic map are rare (Zheng et al. 2019).

In your thesis, you will calculate recombination frequencies within cross families in a real oat dataset to create a genetic map. In addition, you will simulate marker data from a breeding population with different family sizes and a modified physical map to validate the accuracy of the genetic map you have created.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

Supervisor: Emanuel Grosch (emanuel.grosch-1@agrar.uni-giessen.de)

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

Focus: R programming, theory of population genetics

Literature to start with: Zheng et al. 2019 Genetics 212(4):1031-1044

Master thesis: Estimation of haplotype block effects

Haplotype blocks are commonly defined as groups of adjacent SNP markers on a chromosome which are either inherited together or are located within a certain physical or genetic distance on a chromosome (see Difabachew et al. 2023 for a more detailed explanation). They are said to capture local epistatic effects when they are used in genomic prediction; however, the challenge is to estimate effects for the different haplotype block alleles because ridge-regression BLUP, which is often used in genomic prediction, only works for biallelic markers. This thesis aims to investigate and compare different methods for estimating haplotype block effects, including neural networks.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

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

Requirements: MK-002 or MK-002-EN and either MK-119-EN, MP-236-EN, or MP-240-EN completed successfully.

Focus: R programming, theory of quantitative genetics

Literature to start with: Difabachew et al. 2023 *Frontiers in Plant Science* 14 (on haplotype blocks), Heilmann et al. 2025 *BMC Bioinformatics* (on effect estimation)

Master thesis: Cross planning with haplotype-block-based methods

The selection of parents and the planning of crosses are integral parts of any plant breeding program. In the past years, different methods for cross planning have been proposed. The focus of this thesis is on methods that employ haplotype blocks, *i.e.* groups of adjacent SNP markers on a chromosome. You will simulate a breeding program in order to compare the selection gain of a standard method (truncation selection) with genotype building (Kemper et al. 2012), the optimal haplotype value (Daetwyler et al. 2015), and the optimal population value (Goiffon et al. 2017).

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

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

Requirements: MK-002 or MK-002-EN and either MK-119-EN (would be ideal), MP-236-EN, or MP-240-EN completed successfully.

Focus: R programming, theory of quantitative genetics

Literature to start with: Difabachew et al. 2023 *Frontiers in Plant Science* 14 (on haplotype blocks), Kemper et al. 2012 *Journal of Dairy Science* 95, Daetwyler et al. 2015 *Genetics* 200, Goiffon et al. 2017 *Genetics* 206

Bachelor or Master thesis: Estimation of marker effects from two-stage field trial analyses

The R package **StageWise** (Endelman 2023) provides different methods for the two-stage analysis of field trials: First, each environment (year/location combination) is analysed separately. In the second stage, the resulting adjusted entry means are taken as the observations of a model with the environments as an independent variable in the model. The objective of the thesis is to conduct a simulation study that uses pre-defined marker effects. Based on these, field trials are simulated. The goal is then to investigate which of the two-stage evaluation methods for the field trials results in estimates of marker effects that resemble the original values as closely as possible.

Target group: Bachelor students of Agrarwissenschaften, Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

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

Requirements: For Bachelor students: BK-005 and BP-041 completed successfully. For master students: MK-002 or MK-002-EN and either MK-119-EN (would be ideal), MP-236-EN, or MP-240-EN completed successfully.

Focus: R programming, theory of population genetics

Literature to start with: Endelman 2023 Theoretical and Applied Genetics (2023) 136:65

Unbalanced/Unreplicated field trials in plant breeding programs

Since field trials are expensive, breeders and researchers aim to use experimental designs that deliver reliable results while minimizing the required numbers of plots in the field. Sometimes, seed availability is also an issue, particularly in trials that are conducted at early stages of breeding programs. In this thesis, you will analyse different wheat and oat datasets. The goal is to get an idea of how reliable the results of unbalanced/unreplicated field trials are and if reliability can be improved by using genotypic rather than phenotypic values for evaluation of selection candidates.

Target group: Master students of Crop Sciences (Nutzpflanzenwissenschaften), Informationstechnologie, or Agrobiotechnology

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

Requirements: MK-002 or MK-002-EN and either MK-119-EN (would be ideal), MP-236-EN, or MP-240-EN completed successfully.

Focus: R programming, theory of quantitative genetics

Literature to start with: Terrailon et al. 2022 Frontiers in Plant Science 13, Endelman 2023 Theoretical and Applied Genetics (2023) 136:65