Coursework guidelines

Plant Genetic Resources (3502-470)

06 Jun 2025

Table of contents

1 General information

Students are expected to perform analyses of a simulated data set and write a short report about it. The assignment of the data set to each student is provided in a PDF file and the numbered datasets are combined in a single ZIP file that needs to be downloaded and unpacked.

The objective of the coursework is to exercise the analysis of a dataset that is typical in the genetic analysis of plant genetic resources, i.e., that was generated by whole genome resequencing or genotyping with genome-wide markers. The goal of the analysis is to describe the genetic structure of a sample, and this is usually the first step in the analysis before either conducting a genome-wide association study or a genomic prediction analysis.

If you have any questions during analysis or writing, please post them in the ILIAS forum!

2 Description and analyses of the data set

The data sets contain SNP data in VCF format that were generated with coalescent simulations and were then converted to a VCF file. The parameters were number of populations (two to five), migration (with or without) and population growth (with or without).

Download the dataset and the assignment of each student from this link.

Your tasks are:

1. to estimate the **number of populations** that is consistent with the data,
2. to analyse whether there is **gene flow between populations**,
3. to test for evidence of **population growth**,
4. and to **construct a core collection**.

To achieve these objectives, you are expected to carry out the following analyses:

1. A basic description of the data (number of individuals, number of SNPs).
2. An estimate of genetic diversity (the sequence length can be obtained from the VCF file) and Tajima's D for the complete data set but also for the subpopulations (if there are any).
3. A population structure analysis with a PCA, a model-based analysis (with LEA), and a discriminant analysis of principal components (DAPC, with the R package adegenet).
4. A phylogenetic tree with the neighbor joining method.
5. A core collection and a subsequent PCA to evaluate the core collection (and compare with the full data).

For almost all analyses you can use the R codes from the computer labs. As a novel approach, you will use DAPC, for which there are several tutorials online.

Analysis report

The analysis results are summarized in a report that needs to be submitted as Quarto Markdown and a PDF report. To produce the PDF report, use the `tinytex` package. Instructions for the installation of this package are here.

The analysis report should consist of the following sections:

- ☐ A header with your immatriculation number (do not include your name to avoid any bias during grading) and the dataset number.
- ☐ A short description of the data (number of individuals, number of polymorphisms, etc.)
- ☐ One section for each analysis which should include
    - ☐ A short description of the analysis, i.e., what the analysis is about and what you did (including choice of parameters etc)

  - ☐ The R code that you used in proper code chunks. The code chunks should also be displayed in proper context in the PDF (or HTML).
  - ☐ A brief description of the results which summarises which information can be retrieved from the results and how it can be interpreted.
  - ☐ Figures and/or tables that show the results, properly referenced in the text sections and with proper caption and correct annotation (e.g., have axis labels). Avoid redundancy.
  - ☐ Each analysis should be fully repeatable with the given information.
- ☐ A discussion section, in which you summarise all your results and discuss them in context with each other. In particular make statements whether your data supports a certain number of populations, whether you think there is admixture or whether there was population growth in one or all of the subpoplations. If your analyses disagree in one or more aspect, disuss this as well. Discuss also whether your core collection represents your data well. Justify all your conclusions with your analysis results.

The writing should be concise but consist of full sentences, not just bullet points with key points. It should be understandable by someone who is not familiar with this particular dataset or analysis, but it can be short. You may use references to the lecture or scientific literature, but you need to properly reference them in the text and at the end. For further instructions how to cite literature in Quarto Markdown, see here.

To make the report readable and concise, make sure that only the relevant output from the calculations is included in the report. You can set this with the execution options as described here.

3 Evaluation criteria

The Markdown file needs be executable without errors when knitting it to PDF and it should produce the exact same file you submitted. Any data imports should assume that the data file is in the same directory as the Quarto Markdown file.

If you need any R packages, provide the loading of them as a separate code chunk at the beginning of the respective analysis section. If you use an R package that was not used in the computer lab, also add the codes to install it (but outcommented).

Both the Quarto Markdown file and the PDF file rendered from the respective `.qmd` file have to be submitted via upload to ILIAS. The files have to be named `123456-report.Rmd` (or `123456-report.qmd`) and `123456-report.pdf`, where `123456` is your immatriculation number. If any additional files are required for your Quarto Markdown file to be executable, please upload them as well (and name them correspondingly with your immatriculation number).

You should demonstrate in the term paper

- that you understand the conceptual basis of the methods

- that you are able to critically think about the conditions of your analyses (including choice of e.g. parameters, settings, data subsets, …)
- that you are able to discuss your own results
- that you are able to write a report with a good flow of arguments and high readability

The analysis needs to be repeatable with the information given in the analysis report.

The general appearance of the analysis report will be taken into account for grading:

- Use of complete sentences instead of bullet points
- Orthography and grammar
- Quality of writing and plots (including figure captions)
- Do not write it like a tutorial or instructions for others.

Length of report: There is not a minimal or maximal length, but you should be quite concise and you do not have to write a long. A rough estimate is that the report has

Note: the report will be checked for plagiarism and also whether it has been created by AI models such as ChatGPT.

Submission of report

The report needs to be uploaded on ILIAS together with a scan of the declaration of originality to be found under https://www.uni-hohenheim.de/fileadmin/uni_hohenheim/PA/formulare/allgemein/declaration-of-originality-digital-thesis.pdf (Please check the box 'seminar paper'.)

Deadline for the submission of the report: **11 July 2025** at 24:00 (midnight).

The upload link is on ILIAS under 'Coursework upload'.

Schedule

Note: The schedule is subject to changes!

Date	Day	Time	Topic	Room	Lecturer
03-04-2025	Thu	08-10	Introduction	S09	Schmid
07-04-2025	Mon	08-10	Genetic diversity	S09	Schmid
07-04-2025	Mon		*No class*		
10-04-2025	Thu	08-10	Genomic variation: Genotyping and sequencing	S09	Daware
14-04-2025	Mon	08-10	Phylogenetic analysis	S09	Daware
14-04-2025	Mon	14-18	Computer lab: Data preparation, Genetic diversity, Phylogenetics	PC3	Daware

Date	Day	Time	Topic	Room	Lecturer
17-04-2025	Thu	08-10	Biodiversity & Crop diversity and systematics	S09	Schmid
21-04-2025	Mon		*Easter Monday - No class*		
24-04-2025	Thu	08-10	Crop domestication	S09	Daware
28-04-2025	Mon	08-10	Population structure & Gene flow	S09	Daware
28-04-2025	Mon	14-18	Computer lab	PC3	Daware
01-05-2025	Thu		*Public holiday - No class*		
05-05-2025	Mon	08-10	Coalescent Theory	S09	Schmid
05-05-2025	Mon	14-18	Tests of selection / Computer lab	PC3	Schmid
08-05-2025	Thu	08-10	Genetics of crop evolution	S09	Daware
12-05-2025	Mon	08-10	History of PGR, Legislation for PGR	S09	Schmid
12-05-2025	Mon	14-18	Demographic analysis / Computer lab	PC3	Schmid
15-05-2025	Thu	08-10	Conservation of plant genetic resources	S09	Daware
19-05-2025	Mon	08-10	Core collections	PC3	Daware
19-05-2025	Mon	14-18	Allele mining in PGR / Computer lab	S09	Schmid
22-05-2025	Thu	08-10	Genetic mapping of useful alleles	S09	Daware
26-05-2025	Mon	08-10	Analysis of phenotypic diversity	S09	Schmid
26-05-2025	Mon	14-18	Genetic resources in plant breeding / Computer lab	PC3	Schmid
29-05-2025	Thu		*Ascension day - No class*		
02-06-2025	Mon	08-10	Genetic resources in plant breeding - II / Data analysis project	S09	Schmid
11-06-2025	Wed		Excursion during Pentecost break: 11-13 June		

Date	Day	Time	Topic	Room	Lecturer
11-07-2025	Fri	23:55	Deadline submission of data analysis report		
17-07-2025	Thu	08-10	Written exam (Exam period 1)	S09	
22-09-2025	Mon	14-16	Written exam (Exam period 2)	S09	

Course organisation

- Syllabus: HTML
- Computer labs:
    - The computer labs will take place in PC Room 3 where we have access to windows PCs. You can either work with them or bring your own laptop.
    - We will use R and Rstudio for the computer exercises. If you want to install R and Rstudio on your own computer: R download and R studio
    - We will work with the same data set for many of the computer exercises. To avoid having to prepare the data each time, we will do it only once using these instructions and save the prepared data. You will need the following data files: genetic data, list with teosinte samples, list with landraces, and list with improved varieties. If you encounter any problems you can also download the prepared data here.
- Data analysis project: Download the datasets and the assignment of each student to the data here.

Topics

4 Introduction and motivation

- Slides | Lecture notes
- Video: Introduction (30 min)

5 Genetic diversity

- Slides | Lecture notes
- Videos: 1-Types of genetic diversity (22 min) 2-Diversity in PGR (14 min) 3-Measuring diversity (18 min) 4-DNA sequence diversity (18 min) 5-Complex diversity (10 min)
- In class exercise
- Computer lab:
    - Data preparation: HTML | Markdown
    - Analysis of genetic diversity: HTML | Markdown

6 Genomic variation: Genotyping and sequencing

- Slides | Lecture notes
- Videos: 1-Genetic variation and genotyping (21 min) 2-Sequencing (19 min) 3-Bioinformatics (15 min)
- In class exercise
- Computer lab: No computerlab for this topic

Additional reading materials:

- *A field guide to whole-genome sequencing, assembly and annotation* by R Eckblom and JBW Wolf (2014)

7 Phylogenetic analysis

- Slides | Lecture notes
- Video: 1-Key concepts (14 min) 2-Phylogenetic trees (14 min) 3-Methods tree construction (8 min) 4-UPGMA clustering (10 min) 5-Further methods and PGR examples (20 min)
- In class exercise
- Computer lab: HTML | Markdown

8 Biodiversity

- Slides | Lecture notes
- Video: 1-Introduction (19 min) 2-Economic value (5 min) 3-Agrobiodiversity (13 min) 4-Changes in biodiversity (10 min)
- In class exercises

9 Crop diversity and systematics

- Slides | Lecture notes
- Video: 1-Plant phylogeny (15 min) 2-Plant architecture (20 min) 3-Crop phylogeny (17 min)
- In class exercise

10 Crop domestication

- Slides | Lecture notes
- Videos: 1-History (21 min) 2-Centres (23 min) 3-Old-World (17 min)
- In class discussion

11 Population structure

- Slides | Lecture notes | 3-D PCA of Amaranth domestication: HTML (loading may take a while)
- Videos: 1-Introduction and phylogeny (10 min) 2-PCA (14 min) 3-Modelbased inference (20 min)
- In class exercise:
- Computer lab: HTML | Markdown

12 Gene flow and reticulate evolution

- Slides | Lecture notes
- Videos: 1-Introduction (19 min) 2-Reticulate evolution (9 min) 3-Examples (28 min)
- In class discussion

13 Coalescent theory

- Slides | Lecture notes
- Videos: 1-Theory and genealogies (24 min) 2-Mutations (22 min) 3-Demography and applications (26 min)
- Computer lab: HTML | Markdown

14 Tests of selection

- Slides | Lecture notes
- Videos: 1-Introduction (20 min) 2-Concepts (13 min) 3-Selection tests (25 min)
- Computer lab: HTML | Markdown

15 Genetics of crop evolution

- Slides | Lecture notes
- Videos: 1-Domestication syndrome (15 min) 2-Genetics of maize domestication (15 min) 3-Molecular genetics of teosinte branched 1 (24 min)
- In class discussion (only questions 1 to 4).

16 Demographic analysis of crop evolution

- Slides | Lecture notes
- Videos: 1-Introduction (19 min) 2-Selection at tb1 (20 min) 3-Genome-wide selection detection (11 min)
- In class discussion (only questions 5 to 8).

17 History of plant genetic resources

- Slides | Lecture notes
- Videos: 1-Introduction (10 min) 2-World-changing plants (14 min) 3-Collection expeditions (29 min) 4-German genebank history (5 min) 5-International developments (20 min)

18 International legislation for plant genetic resources

- Slides | Lecture notes
- Videos: 1-Introduction (20 min) 2-International Treaty (9 min) 3-SMTA (14 min) 4-Nagoya-Protocol (13 min)
- Additional videos: ABS - Simply explained (5 min) What is ABS? (2.5 min) ABS monitoring (6 min)
- In class exercise: PDF

19 Conservation of plant genetic resources

- Slides | Lecture notes
- Videos: 1-Introduction (26 min) 2-Ex situ conservation (17 min) 3-In situ conservation (31 min)
- In class discussion

20 Core collections

- Slides | Lecture notes
- Videos: 1-Background (7 min) 2-Construction of core collections (24 min) 3-Examples (5 min)
- Computer lab: PDF, data

21 Allele mining in PGR

- Slides | Lecture notes
- Videos: 1-Introduction (19 min) 2-FIGS (16 min) 3-Popgen-based allele mining (6 min)
- In class discussion (part 1)

22 Genetic mapping of useful alleles

- Slides | Lecture notes
- Videos: 1-Introduction (16 min) 2-Basic principles (16 min) 3-Methods and caveats (18 min) 4-Genetic mapping in PGR (9 min)
- In class discussion (part 2)

23 Analysis of phenotypic diversity

- Lecture notes
- Videos: 1-Background (11 min) 2-Mapping populations (22 min) 3-Phenotyping technologies (24 min)
- Computer lab: Rmd, dataset

24 Genetic resources in plant breeding

- Slides | Lecture notes
- Video: 1-Introduction (12 min) 2-Prebreeding (25 min) 3-Breeding methods (17 min) 4-Genomic selection (45 min) 5-Introgression libraries (9 min) 6-Genetic engineering (8 min) *Note: No video on genome editing, check out the lecture notes*
- In class discussion

Course literature

- Jack Harlan: Crops and Man (1992) PDF