Work plan for the MSc. thesis on

Offspring simulation for plant breeding

in cooperation between

Dr. Baumdicker and Computomics

1. Objective

The main goal of the project is to develop a new method for offspring simulation that must be capable to simulate effectively the development of plant breeding populations across a wide range of mating designs and selection strategies for major crop species, including maize. This includes the simulation of genetic variation based on species-specific meiosis-related parameters, the prediction of the phenotypic variation for traits of interest in agriculture based on the simulated genetic variation, and the validation of both simulated genotypes and phenotypes with publicly available data and population genetics theory. In addition, the method will be tested in its capacity to unlock new strategies to advance genetic gain in breeding programs using machine learning prediction methods trained on the simulated data.

2. Background

While real plant breeding programs demand a high-effort in time and resources, the large-scale field-testing needed to develop new breeding strategies is commonly impractical, unaffordable, or nearly impossible to perform by seed companies. Therefore, simulation tools offer an efficient and affordable alternative to design new breeding strategies as well as to evaluate existing programs.

Many softwares for plant breeding simulations are currently available: Plabsoft, AlphaSimR, MoBPS, genomicSimulation, simcross, OffspringSimulator, and XSim. Although the mentioned programs present limitations in terms of flexibility to simulate different breeding scenarios, accuracy in the simulation of genotypes and phenotypes, or computation efficiency, the fine-tuning and manipulation of their functions can still be useful to develop a new tool capable to overcome such caveats. In addition, there are available programs in the field of natural population genetics that can be explored to increase the accuracy in the offspring simulation for plant breeding purposes. Such programs utilize models either based on coalescent theory or the so-called forward-in-time approaches. Some popular software in population genetics are: simuPOP, FREGENE, AdmixSim, and stdpopsim. The later is a public library maintained by the population genetics research community world-wide that aims to establish standard benchmarks by which new tools for population genetics simulation might be tested.

The improvement of the existing simulation programs and the integration of population genetics models into plant breeding pipelines, plus their respective validation with real data, have the potential to generate a new tool to effectively and efficiently simulate different breeding scenarios in order to help breeders with making accurate decisions to maximize return on investment.

3.1 Work plan

The project will be organized in mutual agreement among Dr. Baumdicker, the MSc. Student and Computomics through regular meetings. The developed software for offspring simulation plus the data generated during the project will be shared among all parties. The project will cover offspring simulation and plant breeding strategies for the crop species maize (*Zea mays*) and (optionally) also sunflower (*Helianthus annuus*). The related work will be divided into sections, each of which will be managed by a project partner under the following structure:

* **Data collection and processing:**

Computomics will provide both genotypic and phenotypic publicly available data sources to develop and validate the method. Dr. Baumdicker and the MSc. Student will curate and process the collected data for its proper use.

* **Testing available software:**

Computomics will test the plant breeding-related software mentioned above while Dr. Baumdicker will test population genetics models and programs, including stdpopsim.

* **Genomic simulation and validation:**

The MSc. Student under the supervision of Dr. Baumdicker and Computomics will design and code the new simulation tool and validate it with real data. Computomics will help in providing information for meiosis parameter optimization. Dr. Baumdicker will help in the validation with population genetics theory.

* **Phenotype simulation and validation:**

The MSc. Student and Dr. Baumdicker will simulate phenotypes with publicly available simulation methods and simple scripts. We will focus on the case where two parental genomes are crossed followed by a few selfing generations. If time permits, we will also consider the scenario of a cross of two parental genomes followed by a single backcrossing and selfing.  
Based on the simulated genotypes they will validate the performance of the predicition tools on these simulated datasets as well as on publicly available phenotype data. Computomics will discuss and validate the results with them.  
 In addition, Computomics will do the phenotypic prediction of the simulated genotypes with in house programs and validate it with real data and discuss the approach with Dr. Baumdicker and the MSc. Student.

* **Testing of different breeding strategies:**

The MSc. Student will simulate different plant breeding scenarios to test the developed offspring simulation method. We will try to predict the best pair of parents in the two parents breeding scenario and compare it to more complex scenarios if time permits. Computomics will provide advice in the designing of such scenarios.

* **Incorporation of selfing in stdpopsim and msprime:**

The MSc. Student and Dr. Baumdicker will aim add the feature of selfing plants into the msprime simulation framework and the stdpopsim catalog in general.

* **Manuscript writing:**

The MSc. Student will write the manuscript for the MSc. thesis while Dr. Baumdicker and Computomics will help in the review process.

3.2 Work plan schedule

| Activity | 2023 | | 2024 |
| --- | --- | --- | --- |
| Q3 | Q4 | Q1 |
| Organization - meetings |  |  |  |
| Data collection and processing |  |  |  |
| Testing available software |  |  |  |
| Genomic simulation and validation |  |  |  |
| Phenotype simulation and validation |  |  |  |
| Testing of different breeding strategies |  |  |  |
| Manuscript writing |  |  |  |

**3.3 Publication & IP**

Computomics supports the publication of this work in an academic journal and the release of the developed code under OSI approved open source software license.