

論文の内容の要旨

生産・環境生物学専攻
平成 30 年度博士課程進学
氏名 戸田悠介
指導教員名 岩田洋佳

論文題目 Time-series measurement of crop growth process and modeling of genetic and environmental effects
(作物成長過程の時系列計測とその遺伝的・環境効果のモデル化)

The growth process of crops is an important research topic in breeding science because traits, such as yield and quality, observed at harvest time are determined in a series of growth processes. By dissecting the formation process of a target trait, we can obtain a deeper understanding of its mechanism, which will improve the efficiency of genetic improvement of the target traits. The growth process, however, has been largely neglected in crop breeding research because obtaining data on the growth process requires time-series observations of plants. Since a large number of genotypes are tested in a breeding program, their time-series measurement is time-consuming and labor-intensive.

Recent developments in sensing technology have made high-throughput phenotyping possible. The application of sensing devices, such as unmanned aerial vehicles (UAVs), can reduce the time and labor required for the measurement of plant traits. It is expected that observation of the growth process of many genotypes will become more accessible for crop breeding. The development and application of appropriate modeling methods are necessary to maximize the benefits derived from such growth process data. In this dissertation, I proposed several modeling methods to evaluate the relationships among crop growth, genomic, and environmental data, and validated whether the models are useful for crop breeding.

1. Predicting biomass of rice with intermediate traits: Modeling method combining crop growth models and genomic prediction models

Genomic prediction (GP) is a method that uses genome-wide marker data and statistical models to predict genotypic values of a target trait. Although GP is a useful method for predicting the phenotype of a target trait, the standard GP model cannot take into account the effect of genotype-by-environment

interaction on the trait. In this study, I developed new prediction models of rice biomass that take into account manually measured growth process data.

‘Kinmaze’, ‘Koshihikari’, and their 123 recombinant inbred lines (RILs) were used as plant materials. The manually measured phenotypic data from field experiments conducted in Tsukuba, Japan in 2014–2015 were provided and used in the study. The phenotypic data included time-series data of leaf age (a continuous index of the emergence of i^{th} leaf number), the number of tillers, heading date, leaf length, and biomass. RILs were genotyped with 362 SNPs.

Simple growth models were applied to time-series growth data, and the estimated parameters were used as growth-related traits along with the other traits. Two-step models were developed to predict biomass. In the first step, growth-related traits except heading date were predicted by simple GP. The heading date was predicted with a model that takes into account the effect of the heading-date-related genes and a developmental rate model. In the second step, the predicted growth-related traits were used to predict the biomass using a crop growth model (CGM) or machine learning (ML). In CGM, the biomass was described as the summation of daily growth calculated from growth-related traits and environmental information. In ML, the relationship between the biomass and the growth-related traits was modeled using linear regression or random forest. As a result, the CGM based model worked better than the standard GP in both known and unknown environments. It was concluded that the efficient use of growth process data could increase the accuracy and robustness of genomic prediction in yield-related traits that are affected by environmental variations.

2. Genomic prediction modeling of soybean biomass using UAV-based remote sensing and longitudinal model parameters

The application of remote sensing (RS) to crop breeding can provide a wealth of information on plant growth processes in field trials. The inclusion of RS data in multivariate GP (MGP) models has been shown to improve the prediction accuracy of target traits, indicating that traits measured by RS were also beneficial for GP. However, the current MGP model cannot incorporate high-dimensional RS data due to the difficulty in estimating the covariance matrix among variables. In this study, I applied growth models to the time-series RS data and used several parameters to represent the growth pattern, and investigated whether the MGP model with these parameters could improve the prediction accuracy of soybean biomass.

In 2016, 198 genotypes of soybean germplasm were grown in experimental fields in Tottori, Japan. Unmanned aerial vehicle (UAV)-RS was used to measure longitudinal changes in canopy height and area continuously. Growth parameters were estimated by applying simple growth models and

incorporated into the GP of biomass. The assessment of genomic heritability and correlation of the parameters indicated that the estimated growth parameters adequately represented the observed growth curves. The incorporation of these growth parameters into the MGP model partially contributed to the accuracy of biomass prediction. It was concluded that these growth models could represent genetic variations in the growth pattern as a function of multiple growth parameters measured by RS. These dimensionally reduced growth models are essential for extracting useful information from RS data and using it for GP and plant breeding.

3. Longitudinal growth analysis of soybean using UAV-based remote sensing and its application on genomic prediction

In this study, I developed models to predict the growth process. With the widespread use of high-throughput phenotyping systems, it is expected that the acquisition of the growth process data will become much easily available. Therefore, by applying GP models to growth data, it is expected to be able to predict the growth of untested genotypes. The growth prediction will be useful for crop breeding because it is essential for cultivation to take into account the growth process. In this study, I implemented several prediction models combining GP and a growth model. The accuracy of the models was validated using growth process data measured with UAV-RS.

In 2017–2019, 198 genotypes of soybean germplasm were grown in experimental fields in Tottori, Japan. The longitudinal changes in their canopy area were measured using UAV-RS. A growth model was applied to the canopy area, with growth expressed as a logistic function and senescence expressed as an exponential function. Next, I developed a two-step GP (TGP) model and tested whether the growth model contributed to the improvement of growth prediction. In the TGP model, the growth process was predicted by first predicting the parameters of the growth model with GP and then substituting the predicted values for the parameters of the growth model. The prediction accuracy was compared with GP and MGP under three prediction schemes. As a result, TGP showed higher prediction accuracy than the other models in the scheme of future growth prediction, in which the second half of the growth period was predicted from the data of the first half of the growth period. It was concluded that the TGP was useful for future growth prediction using data from the early growth period. This prediction method could be applied to the selection at an early growth stage in crop breeding, and could reduce the cost and time of field trials.

4. Prediction of soybean growth curves by modeling genetic and environmental effects on daily growth

It is well known that genotype-by-environment (G×E) interaction has non-negligible effects on crop traits. Compared to G×E on harvest traits, G×E for growth process has been less discussed due to the difficulty of measuring growth. The analysis of G×E on growth process is essential to clarify and understand the mechanism of G×E. In this study, I developed a statistical model to describe G×E on the growth process measured by UAV-RS.

Phenotypic data of canopy area and height measured using UAV-RS in the same field trials were used. Prior to the modeling of G×E, it is necessary to estimate daily phenotypic values of the canopy area and height. Because the measurement data of UAV-RS were affected by certain noises, I developed a model to distinguish between growth and noise. The model was validated to be proper by comparing estimates of canopy height with the plant height data measured manually. Next, I developed models of the environmental response of daily growth using statistical and machine learning methods. In the machine learning model, environmental data and marker genotype data were included as inputs. On the other hand, in the statistical model, marker genotype data were included as input to account for the similarity among genotypes in the shape of environmental response curves. As a result, the estimated daily response of the soil moisture explained the drought stress on the growth well, although no specific tendency was observed for the other environmental factors. The ability to adequately estimate the effect of soil moisture was supported by the experimental design using the multi-level watering treatment and abundant soil moisture data. By comparing the prediction accuracy, the machine learning model outperformed the statistical model in predicting the growth process, although the possibility of overfitting could not be ignored. The results indicated that attention should be paid to data structure when building statistical and machine learning models for G×E of crop growth.

In this dissertation, I proposed several modeling methods to include the growth process data in the prediction of harvest traits or to analyze the growth process. In these models, the relationships among the plant growth process data, genomic data, and environmental data were modeled using CGM, GP, and growth models. The main novelty of these studies was that the simultaneous measurement and modeling of the growth processes of a large number of genotypes. As high-throughput phenotyping systems become more widely used for crop breeding, such data will become more readily available. Further improvements in methods for modeling growth process data will be essential to extract useful insights and generate benefits for crop breeding.