論文題目

Evaluation of the potential of genomic selection in plant breeding using simulations and field trials

(シミュレーションと栽培試験を用いた植物育種におけるゲノミックセレクションの可能性評価)

Genomic selection (GS) is a promising method for animal and plant breeding. Using a model relating phenotypes of target traits to marker genotypes, GS enables breeders to predict the genotypic potential of selection candidates based on their marker genotypes. Although GS is expected to streamline and accelerate plant breeding, studies on GS may not be enough to achieve the routine use of GS in breeding programs. In plant breeding, the performance of GS is difficult to evaluate generally because it is affected by among-species variations in mating systems, ways to propagation, population structures, and conventional breeding strategies. In this dissertation, I evaluated the performance of GS through simulation studies and field trials in consideration of the among-species variations. In addition, I developed a novel simulation language for breeders to evaluate the potential of planned breeding schemes.

1.　Simulation study of genomic selection in allogamous plants

Mass selection is an important method for the breeding of allogamous crops. This method, however, has drawback, i.e., low efficiency of genetic improvement, because it requires a single-plant evaluation. GS enables higher reliability of a single plant evaluation than phenotypic selection (PS) because GS can be performed on marker genotypes. On the other hand, GS may not work well in an allogamous species with a large effective population size, which results in a low level of linkage disequilibrium (LD), because GS

utilizes LD between QTLs and markers. In this study, on the assumption that an annual allogamous plant species has a very low level (i.e., close to linkage equilibrium) of LD, I conducted breeding simulations for two types of target traits, a trait expressed before pollination and a trait expressed after pollination. Especially for a trait expressed after pollination, in which pollen parents cannot be selected before crossing in PS, GS had a larger genetic gain than PS. For a trait expressed before pollination, I compared GS with PS and conventional marker-assisted selection (MAS) in the simulations and evaluated the performance of GS under various scenarios. Results showed that GS attained higher genetic gain than either PS or MAS. GS with a larger population size and more cycles attained higher genetic gain except when the population size was small. The cost efficiency of GS was higher than that of PS only when the genotyping cost was lower than about one-fourth of the phenotyping cost. To evaluate the performance of GS in a trait expressed after pollination, I compared GS and PS in traits expressed before and after pollination. Results show that GS showed almost identical genetic gain genetic gain in both traits except when GS was conducted once per year (i.e., same as PS), while PS in a trait expressed after pollination showed much lower genetic gain than PS in a trait expressed before pollination. High efficiency of GS in a trait expressed after pollination was attributable to the high selection accuracy of chromosomes derived from a pollen parent at GS steps immediately after model updating steps, at which pollen parents were not selected because model updating requires phenotype evaluation. It resulted in increased population size and prevented depletion of genetic variation in a breeding population. This study indicated that GS has a great potential to improve the efficiency of mass selection of allogamous crops in particular when a target trait is expressed after pollination.

2.   Simulation evaluation of island-model genomic selection in an autogamous plant

In the breeding of autogamous crops, population breeding and pedigree method, which utilize inbred lines, are commonly used in breeding programs. This situation results in the issue of a lack of new combinations of genes in a breeding population. Recurrent selection can be used to create recombination in a population, but it requires single-plant evaluation, which is generally inaccurate. GS may have high reliability of single-plant evaluation and would be effective in recurrent selection of an autogamous species. Additionally, the concept of "island model" inspired from population genetics and evolutionally algorithms may be useful to maintain genetic variation through the breeding process. I conducted GS simulations using a real marker genotypic data of rice cultivars to evaluate the efficiency of recurrent selection and the island model in an autogamous species. Results suggested that recurrent selection could attain higher gain than a conventional method using inbred lines. In the recurrent selection, an initial population derived from multiple bi-parental crosses showed larger genetic gain than a population derived

from a single bi-parental cross, suggesting the importance of genetic variation in an initial population. The island-model GS could attain higher gain than the bulked GS in later generations because the island-model GS could maintain larger genetic variation than the bulked GS and improve the genetic potential of the whole subpopulations. Because of the ability of the bulked GS to attain gain rapidly in early generation, it is suggested that breeders should choose a suitable breeding scheme according to their required time.

3. Simulation of the impact of mis-labeling on genomic selection in cassava

In actual plant breeding, humans make mistakes unlike simulations. Especially in GS breeding, humans would tend to make mistakes because GS involves more steps than PS does. To implement GS breeding in actual, the effect of human mistake should be taken into account to consider the level of restriction to prevent human mistakes. In plant breeding, controlling mistakes too strictly may not be cost effective if the mistakes do not have a large impact. I evaluated the impact of mis-labeling, in which a plant happens to be swapped for another one, in cassava breeding using simulation. As simulation results, all scenarios with six levels of mis-labeling (5, 10, 20, 30, 40, and 50%) attained a certain genetic gain because of the relationship between the genetic variance and the prediction accuracy. The higher mis-labeling rate became in a breeding population, the lower selection intensity the breeding population experienced at selection cycles. This situation made the genetic variance in a population with mis-labeling high, and made the response to selection high. The increased genetic variance observed under mis-labeling led to sufficiently improve the accuracy, at least for low mis-labeling rate (10% or less). It is suggested that the large scale of mis-labeling should be prevented, but that preventing small scale of mis-labeling is not cost effective in plant breeding.

4. Field trial of genomic selection using common buckwheat

A field experiment of GS breeding was performed with a real breeding population in common buckwheat. I compared the efficiency of GS with that of PS for improvement of seed yield per unit area in the two years of field trial. To select seed yield per unit area, which cannot evaluated in a single plant, I built a selection index that predict performance of each plant in seed yield based on other seven traits (main stem length, number of nodes, flowering of the first flower, number of flower clusters, number of primary branches, 1000 seed weight, and test weight) that can be evaluated in a single plant. This index was used throughout the selection cycles. In GS breeding, two selection cycles were conducted in each year, and the prediction model was updated every year by using 14,598 to 50,000 markers. In PS breeding, selection was conducted once per year. To verify the difference in performance between GS breeding and

PS breeding, a field test was conducted in 2013 after the two years of breeding. In the test, 48 plants from each generation were cultivated. The selection index, seven traits composing the selection index, the number of seed set of a plant, and the number of secondary branches were evaluated in the field test. Through two years of selection, GS breeding attained 49% higher gain in number of flower clusters and number of seed set than the base population. For the selection index, which was used in selection directly, GS breeding attained 15% higher value than the base population. In PS breeding, selection index increased 4% from the base population, while it was not statistically significant. These results show that GS has higher performance than PS in the genetic improvement in common buckwheat. I compared two prediction models built at the first and second years through the evaluation of their prediction accuracy at the second year, and found that the former showed lower accuracy than the latter, suggesting the importance of model updating in GS breeding. The superiority of GS over PS might be resulted from the effects of accelerating generations using offseason nursery. In the field trial, the efficiency of GS based on the selection index was suggested to improve yield related traits simultaneously.

5. Development of a simple language to script and simulate breeding schemes: the breeding scheme language

It is difficult for plant breeders to determine the optimal scheme under conditions of a target species and target traits because there are a number of possible breeding schemes. Although simulation study is useful to help choose a better (or the best) breeding scheme, it is difficult for breeders to take the first step in conducting breeding simulation because of the complexity to build a simulation platform or even to using a simulation tools. In the present study, I developed a simple and flexible simulation platform, breeding scheme language. Users can define their target species and breeding schemes by utilizing the language. This language might be useful for breeders to evaluate breeding schemes and to choose a breeding scheme among a number of possible schemes.

I demonstrated the high potential of GS by using simulations and field trials. Through these studies, it is suggested that there are factors that affect on GS gain as well as factors that have less effects on GS gain. Update of a prediction model is essential for GS breeding even with a cost especially when a breeding population has low levels of LD. The result of the simulations was coincided with that of the field trial, suggesting the properness of the simulations performed in this study. In the future, the collateral implementation of breeding simulations and a field trial might enable us to improve the efficiency of plant breeding by reflecting the current situation to the simulations and choosing suitable selection strategy at each step of breeding on the basis of the results of the simulations.