



Peter Jaksons¹
 Peter Alspach²
 Carmel Woods¹
 peter.jaksons@plantandfood.co.nz



Maximising the benefits of a plant breeding database

¹ The New Zealand Institute for Plant & Food Research Limited (PFR), Private Bag 4704, Christchurch Mail Centre, Christchurch 8140, New Zealand
² The New Zealand Institute for Plant & Food Research Limited (PFR), 55 Old Mill Road, RD 3, Motueka 7198

Intro and Problem

Plant breeding is an important part of PFR's work; recently we have started using an SQL plant breeders' database application (E-Brida) in most of our breeding programs. We identified three challenges for people using this database and the accompanying user interface that we tried to address as part of the 'Phenotypic data checking project'. These 3 challenges are:

1. It is crucial that the data is entered in the right format and with the least number of (measurement or data entry) errors as possible
2. The data in the database should be easily accessible to all parties (to breeders and data analysts)
3. Methods should be available that allow for easy exploration of the data and which enhance the selection of genotypes.

Goal

To maximise the benefit of this breeders' database for both the breeders as well as for the data analysts involved, we have been developing an R package, called PFRBreedR. This R-package has three main purposes:

1. Immediate database automated daily data checking.
2. Increasing the accessibility of the database.
3. Creating R-functions to facilitate the fitting of appropriate models to the data collected in the database to enhance genotype selection and to help displaying the results.

Outcome

1. Based on historical data, typos and measurement error are being detected each time data is added to the database and descriptive warning emails are sent out in case a possible error is detected. The data checking is ideally done in a univariate and bivariate way. This should increase the quality of the data input.
2. Additionally to the default user interface to extract data, faster and more versatile R functions have been created. In the future, some of these R functions will be turned into R shiny apps to improve the user accessibility.
3. Several R functions have been created to explore the data and to improve the (mainly visual) selection of genotypes. For some of these functions R shiny apps have been developed.

Conclusion

The developed R package, PFRBreedR, is a collection of R functions which aim to increase the quality of the data entered in E-Brida, facilitate data accessibility (which is especially useful for when reproducible results are needed), and provide tools that help breeders with the selection of genotypes.

