# Impact of pedigree depth in the Australian strawberry breeding program: variance component estimation, prediction accuracies and estimation of clonal and breeding values

Ky Mathews, Katie O’Connor and Jodi Neal

Blurb introducing ASBP and different regions bred for.

The Australian Strawberry Breeding Program (ASBP) uses pedigree information in the routine analysis of all traits to estimate breeding values for parental selection. The total genetic effects for each genotype, the so-called “clonal values”, are also estimated from these models.

The analyses are performed in the linear mixed model framework and follow Oakey et al 2006 and Oakey et al 2007 which uses the numerator relationship matrix generated from pedigree information. The ASBP pedigree records are in good order with genotype name aliases regularly updated and information back to founders mostly complete. Currently, the breeding program uses a standard pedigree depth of 10 generations. It is of interest to determine what an optimal pedigree depth could be and the effect of changing pedigree depth on the genetic variance estimates, the breeding and clonal value estimates and the prediction accuracies these estimates.

Aitken et al (2009) performed a similar study in sugar cane where they also investigated the number of years of trialling that make a difference to the estimates. In the case of the ASBP, they have implemented the contemporary group technology described in Smith et al 2021 to determine that for the subtropical program 5 years was sufficient, and for the temperate program 4 years.

Contemporary groups

Yield and average fruit weight are key traits. How other programs model this – repeated measures?

Strawberries are an open-pollinated crop and so crosses are made by…. Description of breeding program here…

Strawberries are considered as diploid for all intents and purposes in the realm of genomics and breeding. However, it is actually a ?? species and the present calculation of the A matrix is based on Henderson (???) and Wright (1922) which is for diploid species.

What is the current total number of generations in the breeding program pedigree?

The purpose of this paper is to investigate the effect of changing the number of generations used to calculate the numerator relationship matrix on the additive and non-additive genetic effects, the breeding and clonal value estimates and their respective accuracies.

## Data set

The proposed datasets are the:

* Subtropical 2020-2024 yldpp, avfrtwt, cumulative yld (cyldpp) and cavfrtwt
* Temperate 2020-2024 yldpp, avfrtwt, cumulative yld (cyldpp) and cavfrtwt
* Pedigree information from Katmandoo, then subset for the following and generate A matrices for each
  + A2: 2 generations (individuals and parents)
  + A3: 3 generations (individuals, parents and grandparents)
  + A4: 4 generations (individuals and parents)
  + A5: 5 generations (individuals and parents)
  + A6: 6 generations (individuals and parents)
  + AF: Back to founders
  + AI: AISweep generated from AF

## Proposed Analyses

1. No pedigree: Single YearWeek analyses for yldpp and avfrtwt
   1. Investigate spatial effects as a side task
2. No pedigree: Single Year analyses of cumulative traits, cyldpp and cavfrtwt
   1. Not sure about spatial for these because they are accumulated over the season– but not difficult to include
3. Develop repeated measures model for yldpp and avfrtwt
4. Investigate different ways of calculating an A matrix for non-diploid species
5. Best model from 1, 2 and 3 above with all 6 x t A matrices, where t = number of types of A matrices (no more than 3), if there are any!)
6. Contemporary groups
7. Overall performance and stability – FAST

## Proposed Results

1. Spatial vs non-spatial models
2. Repeated measures versus cumulative – BLUPS
3. Report on number of parents added with each generation
4. Additive vs non-additive genetic variances
5. Clonal and breeding values with accuracies
6. Correlation between CV and BVs for each trait and each matrix type

## Other thoughts

1. Link to simulation of missingness
2. This may not be the place for the repeated measures
3. Find out if the animal model is appropriate for this species, rather than the parental model we use from Oakey et al 2006, 2007.

QChat suggested structure:

Title: Optimal Pedigree Depth for Genetic Variance and Value Estimation in the Australian Strawberry Breeding Program

Abstract: (Summarise the objectives, methodology, key results, and major conclusions of the study. Highlight the novelty and implications of optimal pedigree depth investigation on breeding and clonal value estimates, as well as the prediction accuracies in the context of the ASBP.)

Introduction:

Elucidate the critical role of pedigree analysis in plant breeding, particularly in parental selection and estimation of breeding values.

Offer a succinct summary of the Australian Strawberry Breeding Program (ASBP), its aims, and historical pedigree record-keeping.

Review existing literature on pedigree analysis techniques used in plant breeding, with an emphasis on Oakey et al (2006, 2007) and other relevant studies influencing the current methodology.

Describe how previous studies, such as Aitken et al (2009), have targeted similar issues in other crops, revealing this study’s potential for broad applicability.

Clearly state the purpose of this paper: to evaluate the impact of changing the pedigree depth on genetic variance, breeding and clonal value estimates, and the accuracy of these estimates in the context of the ASBP.

Materials and Methods:

Give a comprehensive introduction to the ASBP, emphasising the attention to detail in maintaining accurate and up-to-date pedigree records and genotype aliases.

Detail the linear mixed model framework applied in the program and the role of the numerator relationship matrix.

Outline the standard pedigree depth used in the ASBP and the reasoning behind investigating different depths.

Explain the strategies employed and data analysed to determine the optimal pedigree depth, including the specific statistical models and software.

Incorporate the theoretical underpinnings for considering the strawberry as a diploid in genetic evaluations while indicating its true ploidy level and the potential ramifications.

Touch on the breeding strategy used for strawberries, underscoring how the propagation method informs genetic evaluation and the calculation of the A matrix as per Henderson and Wright.

Proposed Analyses Additions:

1. Cross-validation techniques to assess the predictive accuracy and robustness of models 1, 2, and 3.
2. Sensitivity analysis for the choice of pedigree depth, considering alternative depths beyond the standard 10 generations and comparing the impact on genetic evaluations.
3. Examination of genotype by environment interaction (GxE) through multi-environmental trials to better understand the stability of genetic expression across conditions.
4. Pedigree completeness index analysis to quantify the quality and connectivity of pedigree data at different depths.

Results:

Report the analysis outcomes on genetic variance estimates at varying pedigree depths.

Present the refined breeding and clonal value estimates and compare these to current values based on the standard pedigree depth.

Discuss the accuracies of the genetic effect predictions at different pedigree depths as revealed by the models.

Analyse the ideal trial years needed as per the contemporary group technology, comparing subtropical versus temperate programs.

Provide the number of generations in the existing ASBP pedigree records, setting the baseline for study comparisons.

Proposed Results Additions:

1. Analysis of temporal trends in genetic gain or loss when adjusting pedigree depth, assessing the long-term impact on the breeding program.
2. Impact assessment of various A matrix configurations on the selection differential for key traits.
3. Model goodness-of-fit statistics for each tested model to ensure the selection of the most appropriate one.
4. Discussion of implications for breeding program design, including recommendations for optimal trial structuring and data collection strategies.
5. In-depth exploration of significant non-additive genetic effects, if found, with a special focus on dominance and epistasis which are important in polyploid species.
6. Presentation of potential limitations in the current analysis methods for assessing non-diploid species and how such results guide future improvement of genetic evaluation tools.

Discussion:

Discuss the significance of the results in defining the optimal pedigree depth for the ASBP, drawing parallels with plant breeding practices and genetic theory.

Debate how changes in pedigree depth might alter breeding strategies and the long-term implications on the ASBP’s genetic progress.

Relate findings to those obtained by Aitken et al (2009) and the practical applications of contemporary group methods as seen in Smith et al (2021).

Contemplate the implications of the complex genomics of strawberries, referencing the ASBP’s treatment of the crop as diploid, and hypothesise on unresolved questions related to ploidy and its influence on genetic evaluations.

Acknowledge limitations in current research methodologies and propose areas for future investigation.

Conclusion:

Recapitulate the principal findings, emphasising the importance of optimising pedigree depth for accurate genetic variance and breeding value estimations.

Offer recommendations for the ASBP and potentially other breeding programs aiming to refine their genetic evaluation methods.

Declare the broader impact of such studies on improving genetic assessment protocols within plant breeding disciplines.