

Model based design of QTL, GP and GWAS phenotyping experiments using genetic relatedness.

Aidan McGarty¹

Brian Cullis¹, Ahsan Asif² and Kristy Hobson²

November 24, 2022



Centre for Biometrics and Data Science for Sustainable Primary Industries (CBADS-SPI)¹

National Institute for Applied Statistics Research Australia

University of Wollongong

amcgarty@uow.edu.au

Chickpea Breeding Australia²

NSW Department of Primary Industries | Agriculture

Aim: Improvement of design efficiency through inclusion of genetic relatedness

Model based design with genetic relatedness

Model based design with genetic relatedness

- Important to closely match the linear mixed model (LMM) underpinning the design search with that which will be used in the proceeding analysis [4]

Model based design with genetic relatedness

- Important to closely match the linear mixed model (LMM) underpinning the design search with that which will be used in the proceeding analysis [4]
- Equally important to match the design with the aim of the experiment while considering any practical constraints

Model based design with genetic relatedness

- Important to closely match the linear mixed model (LMM) underpinning the design search with that which will be used in the proceeding analysis [4]
- Equally important to match the design with the aim of the experiment while considering any practical constraints
- Cullis et al. (2020) [4] demonstrates the gain in accuracy in partially replicated field trials with compared to without genetic relatedness via ancestral (pedigree) information

Model based design with genetic relatedness

- Important to closely match the linear mixed model (LMM) underpinning the design search with that which will be used in the proceeding analysis [4]
- Equally important to match the design with the aim of the experiment while considering any practical constraints
- Cullis et al. (2020) [4] demonstrates the gain in accuracy in partially replicated field trials with compared to without genetic relatedness via ancestral (pedigree) information
- Aim is to demonstrate similar improvements in the context of molecular marker information

Model based design with genetic relatedness

- Important to closely match the linear mixed model (LMM) underpinning the design search with that which will be used in the proceeding analysis [4]
- Equally important to match the design with the aim of the experiment while considering any practical constraints
- Cullis et al. (2020) [4] demonstrates the gain in accuracy in partially replicated field trials with compared to without genetic relatedness via ancestral (pedigree) information
- Aim is to demonstrate similar improvements in the context of molecular marker information
- Design software; **odw** [2] allows for inclusion of such information in the design process through the genomic relationship matrix (K)

Motivating example

Motivating example

- Disease screening hydroponic experiment for Chickpeas (*Cicer arietinum*) assessing Phytophthora root rot (PRR)

Motivating example

- Disease screening hydroponic experiment for Chickpeas (*Cicer arietinum*) assessing Phytophthora root rot (PRR)
- Germplasm to be assessed (185 chickpea lines) was genotyped thus allowing for inclusion of genetic relatedness in the model based design process

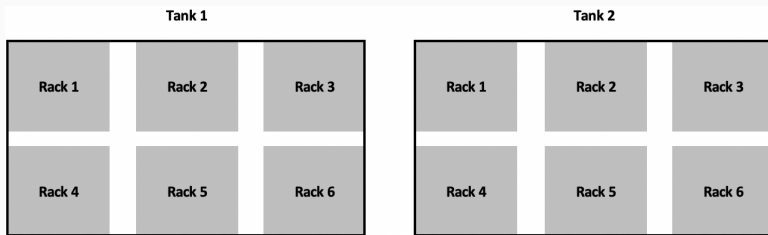
Motivating example

- Disease screening hydroponic experiment for Chickpeas (*Cicer arietinum*) assessing Phytophthora root rot (PRR)
- Germplasm to be assessed (185 chickpea lines) was genotyped thus allowing for inclusion of genetic relatedness in the model based design process
- Plot structure: **2 Tanks**, each with **6 Racks** and with each rack containing **56 Holes = 672 Holes**

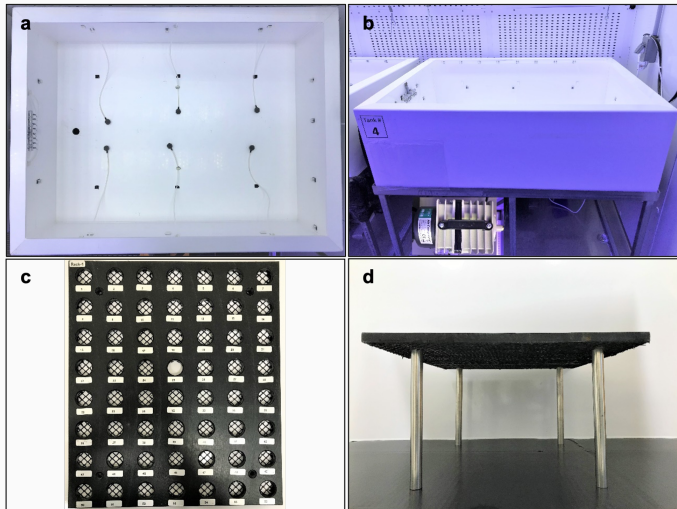
Motivating example

- Disease screening hydroponic experiment for Chickpeas (*Cicer arietinum*) assessing Phytophthora root rot (PRR)
- Germplasm to be assessed (185 chickpea lines) was genotyped thus allowing for inclusion of genetic relatedness in the model based design process
- Plot structure: **2 Tanks**, each with **6 Racks** and with each rack containing **56 Holes = 672 Holes**
- Holes are the both the smallest unit on which an observation can be made (observational unit) and the smallest unit to which a treatment can be applied (experimental unit)

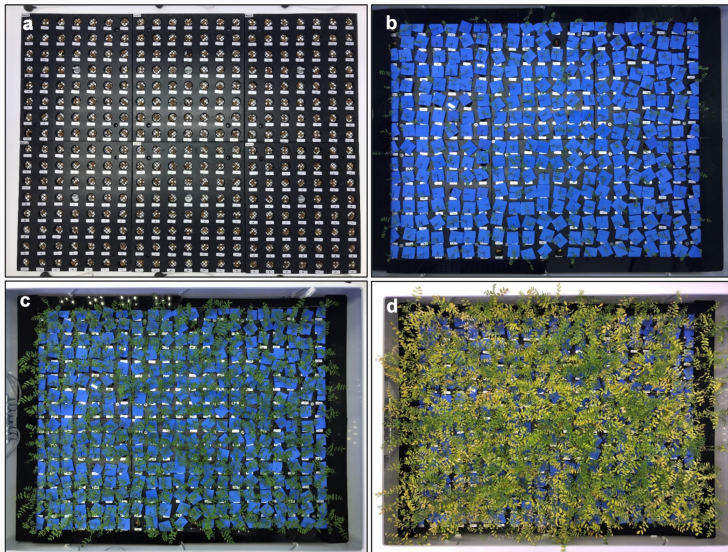
PRR Hydroponic Experimental Layout



PRR Hydroponic Experimental Layout



PRR Hydroponic Experimental Layout



Plant and root systems after PRR exposure



- Utilise LMM with effects aligning with both genetic and non-genetic terms in the design construction

- Utilise LMM with effects aligning with both genetic and non-genetic terms in the design construction
- Genetic effects = **additive effects** + **non-additive effects**

- Utilise LMM with effects aligning with both genetic and non-genetic terms in the design construction
- Genetic effects = **additive effects** + **non-additive effects**
- Non-genetic effects = **Tank effects**
 - + **Rack effects** (within tanks)
 - + **Row effects** (within racks within tanks)
 - + **Column effects** (within racks within tanks)

- Utilise LMM with effects aligning with both genetic and non-genetic terms in the design construction
- Genetic effects = **additive effects** + **non-additive effects**
- Non-genetic effects = **Tank effects**
 - + **Rack effects** (within tanks)
 - + **Row effects** (within racks within tanks)
 - + **Column effects** (within racks within tanks)
- 660 (672-12) holes available, hence $660/185 = 3.57 \notin \mathbb{Z}$

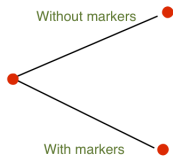
Design process

- Two step design process

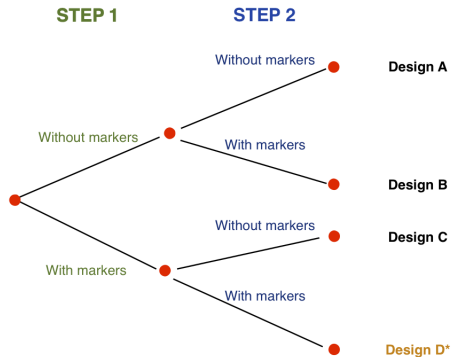
- Two step design process
- Step 1 - Determine which lines will receive an extra replicate (lines to packets)

- Two step design process
- Step 1 - Determine which lines will receive an extra replicate (lines to packets)
- Step 2 - Determine allocation of packets to holes

STEP 1



Designs for Comparison



Step 1: Determining Replication

Step 1: Determining Replication

- Option 1 - Randomly allocate lines to 3 or 4 replicates (packets)

Step 1: Determining Replication

- Option 1 - Randomly allocate lines to 3 or 4 replicates (packets)
- Option 2 - Determine replication based on genotyping information, that is, allow **odw** to allocate lines to replication status (packets) such that genetic diversity is maximised across the two replication groups [3]

Step 2: Assigning lines to holes

Step 2: Assigning lines to holes

- Option 1 - Allocate packets within the experiment assuming independence between lines

Step 2: Assigning lines to holes

- Option 1 - Allocate packets within the experiment assuming independence between lines
- Option 2 - Allocate packets within the experiment such that genetic correlation between lines is considered in the (linear mixed) model-based design

Results

- **odw** utilises \mathcal{A} -optimality as the optimality criterion, which in the context of a comparative genetic evaluation experiment, an \mathcal{A} -optimal design is equivalent to minimising the average pairwise variance of all elementary treatment contrasts [1]

- **odw** utilises \mathcal{A} -optimality as the optimality criterion, which in the context of a comparative genetic evaluation experiment, an \mathcal{A} -optimal design is equivalent to minimising the average pairwise variance of all elementary treatment contrasts [1]
- We aim to quantify sub-optimality in the designs for comparison through differences in \mathcal{A} -values

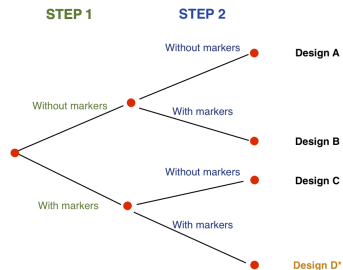
- **odw** utilises \mathcal{A} -optimality as the optimality criterion, which in the context of a comparative genetic evaluation experiment, an \mathcal{A} -optimal design is equivalent to minimising the average pairwise variance of all elementary treatment contrasts [1]
- We aim to quantify sub-optimality in the designs for comparison through differences in \mathcal{A} -values
- Namely, via evaluating the final design generated by Designs A, B and C under the model for Design D

- **odw** utilises \mathcal{A} -optimality as the optimality criterion, which in the context of a comparative genetic evaluation experiment, an \mathcal{A} -optimal design is equivalent to minimising the average pairwise variance of all elementary treatment contrasts [1]
- We aim to quantify sub-optimality in the designs for comparison through differences in \mathcal{A} -values
- Namely, via evaluating the final design generated by Designs A, B and C under the model for Design D
- Cullis et al. (2020) [4] indicates there exists a direct correlation between \mathcal{A} -values and response to selection gain

Results

	Design A	Design B	Design C	Design D
\mathcal{A} -values	0.195181	0.191557	0.193097	0.189527
Difference	0.005654	0.00203	0.00357	0

Table 1: Summary of \mathcal{A} -values for the different designs



	Design A	Design B	Design C	Design D
\mathcal{A} -values	0.195181	0.191557	0.193097	0.189527
Difference	0.005654	0.00203	0.00357	0

Table 2: Summary of \mathcal{A} -values for the different designs

	Design A	Design B	Design C	Design D
\mathcal{A} -values	0.195181	0.191557	0.193097	0.189527
Difference	0.005654	0.00203	0.00357	0

Table 2: Summary of \mathcal{A} -values for the different designs

- $A - B \approx C - D \approx 0.0036$

	Design A	Design B	Design C	Design D
\mathcal{A} -values	0.195181	0.191557	0.193097	0.189527
Difference	0.005654	0.00203	0.00357	0

Table 2: Summary of \mathcal{A} -values for the different designs





- $A - B \approx C - D \approx 0.0036$
- $A - C \approx B - D \approx 0.002$

Results

- Difference in \mathcal{A} -values allows for quantification of the effect of including genetic relatedness in the design process

- Difference in \mathcal{A} -values allows for quantification of the effect of including genetic relatedness in the design process
- Points to inclusion of markers in the final allocation as the slightly larger of the two effects

- Difference in \mathcal{A} -values allows for quantification of the effect of including genetic relatedness in the design process
- Points to inclusion of markers in the final allocation as the slightly larger of the two effects
- Effect of both markers in step 1 and step 2 appears to be additive (absence of interaction)

-  David G Butler.
On the optimal design of experiments under the linear mixed model, 2013.
-  David G Butler.
Optimal experimental design under the linear mixed model, 2022.
-  Brian Cullis and David Butler.
On model based design of comparative experiments in R [Unpublished].
Journal of statistical software, 2022.
-  Brian R Cullis, Alison B Smith, Nicole A Cocks, and David G Butler.
The design of early-stage plant breeding trials using genetic relatedness.
Journal of Agricultural, Biological, and Environmental Statistics, 25:553–578, 11
2020.