



Genetic selection for spatial and temporal characteristics of root growth in plants using a high-throughput phenotyping platform

Alison Kelly, Valeria Paccapelo
Jack Christopher, Karine Chenu
Patrick Thaon, Xavier Draye

International Biometrics Conference, 24 August 2020

The HTP Platform

- A robotic platform where plants are grown in a controlled environment
- Individual seeds are germinated and the plants grow under aeroponic conditions
- Roots dangle freely – water and nutrients are delivered as a mist to promote plant growth
- The platform can house a large number of plants and is automated to cycle past cameras capturing digital images of the root systems on a 2 hourly cycle for up to 21 days



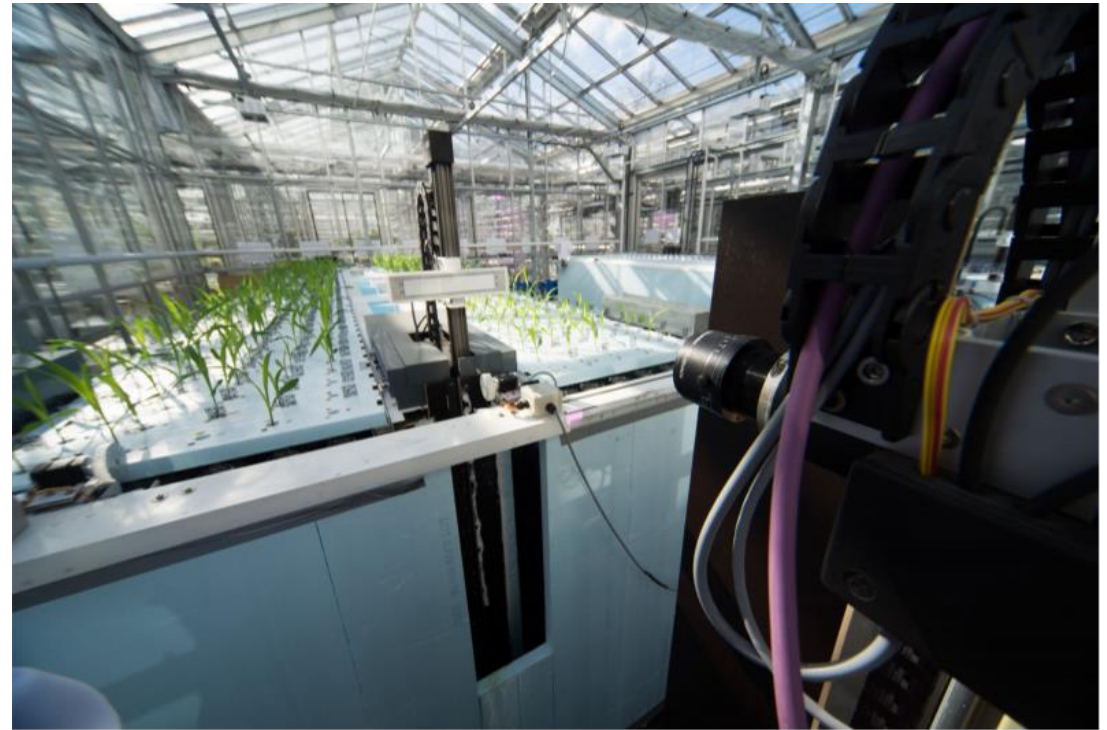
Why HTP for roots?

- Roots are the hidden component of a plant, making phenotyping for root architecture and root growth challenging
- Most field-based methods are destructive and labour intensive
- Root traits, like most soil-based traits, are notoriously variable



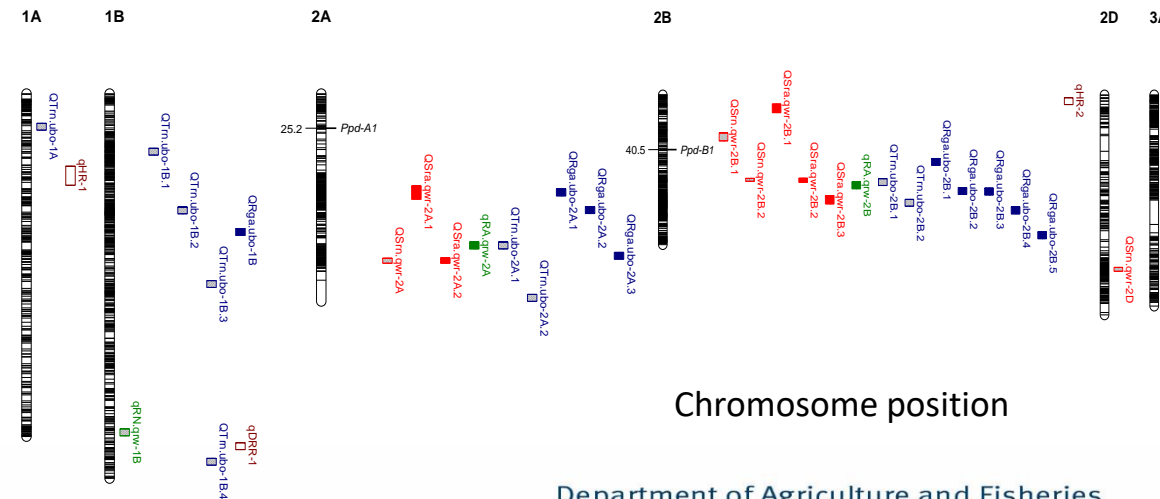
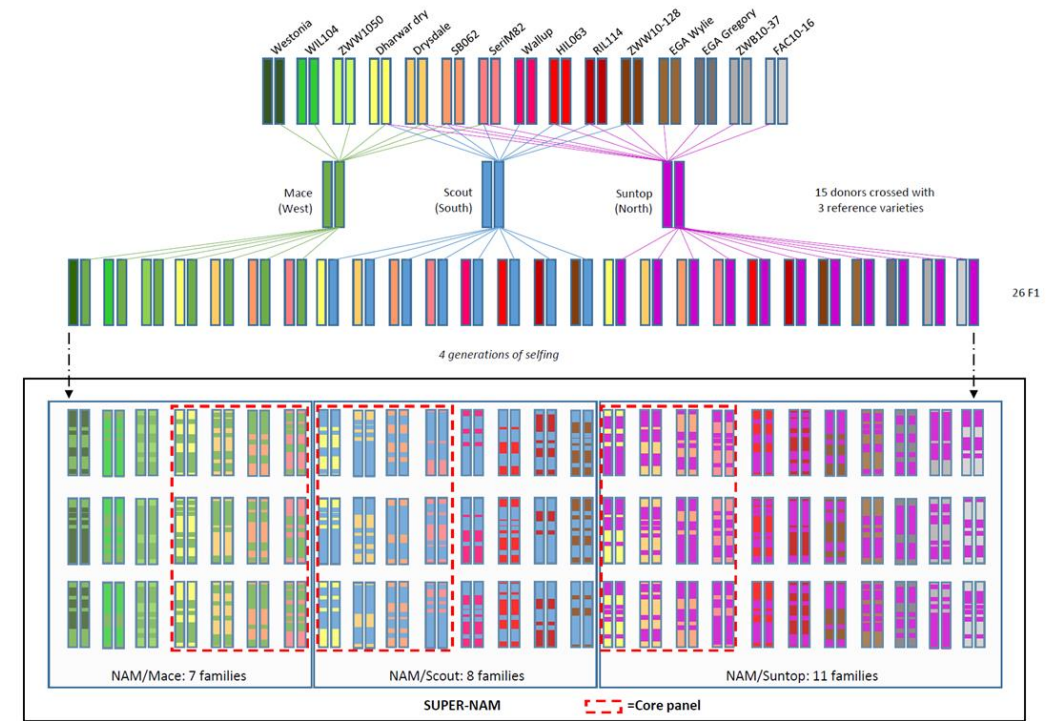
Phenotyping

- Image-based **phenotyping** allows for non-destructive assessment of plant structure and plant growth
- Commonly there is both a spatial and a temporal component to the data based on where or how the plants are grown/growing
- **The phenotype:** measured through high dimensional data over space and time
- **The aim:** to describe plant traits and growth patterns from these high dimensional data



Genotyping

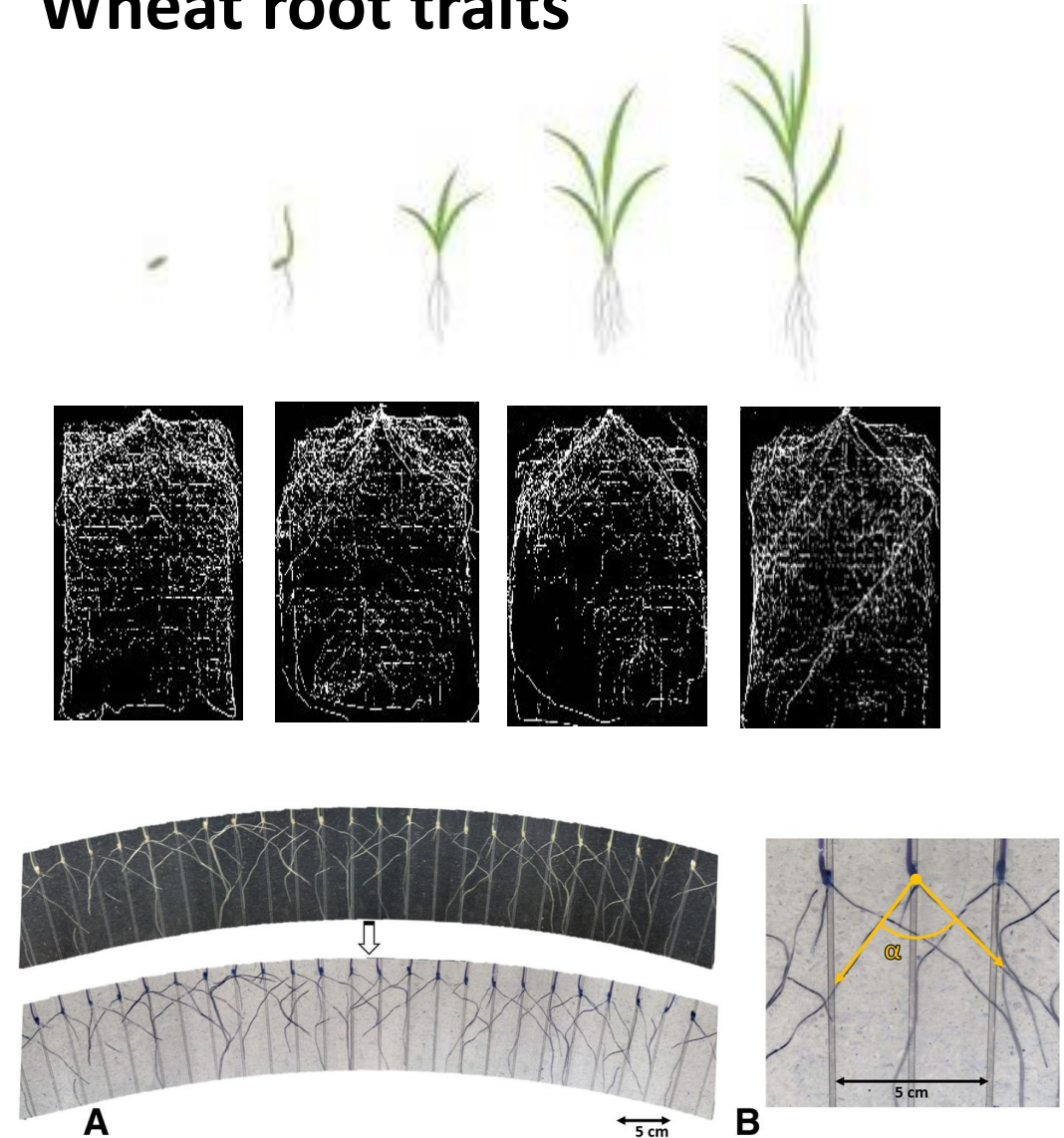
- HTP platforms are typically associated with the assessment of a large number of plants and are often utilised in genetic improvement programs to improve plant characteristics eg plant breeding
- There can be a complex **genetic component** to the plant populations that are assessed using these HTP platforms
 - Complexity can arise from the genetic structure in the plant material under assessment
 - Further complexity can arise from the genetic profiles available for each plant produced via sequencing platforms eg molecular markers
- The aim:** to determine genetic regions associated with favourable traits and growth patterns from these high dimensional data



Overview of this study

- HTP provides the platform for bridging the phenotype-to-genotype gap in the study of root architecture
- The statistical challenge is dealing with dimensionality in the response variable, as well as the independent variables under study
- In this talk we present a novel statistical approach for dealing with the dimensionality problem in **genetic improvement of root architecture**

Wheat root traits



The Motivating Experiment

Experimental Structure

- 2 HTP platforms arranged in a 2D array of 5 plant positions by 99 rows per platform
- Multiple measurements taken of each plant over time (18 days with up to 163 images taken for each plant)

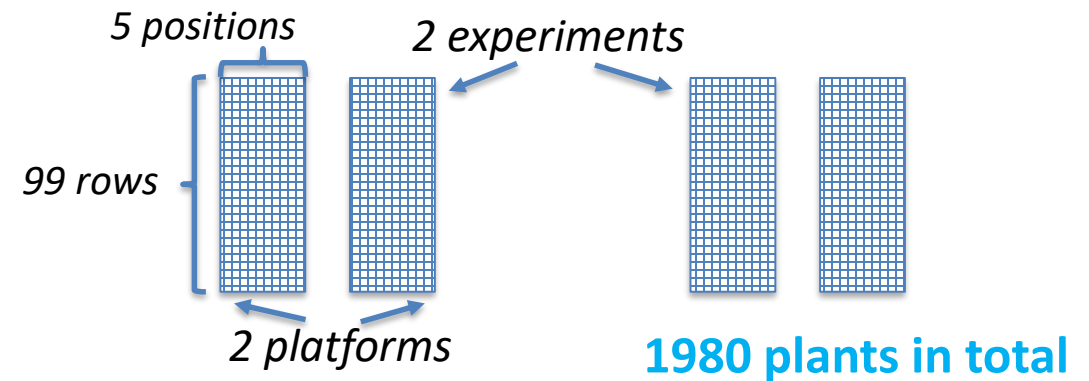
Genetic structure

- 520 genotypes from a **nested association mapping (NAM)** population
- Check varieties, Parents plus 11 NAM families with 40-52 lines per family, forming a total of 417 NAM lines
- 4300 SNP markers available for the 417 NAM lines (after quality control)

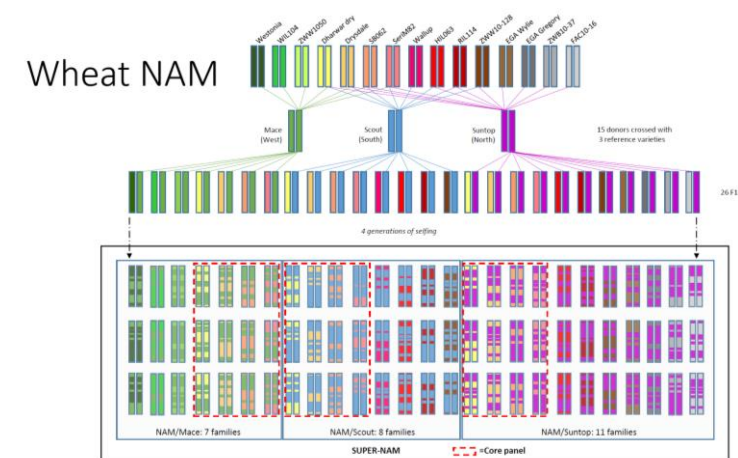
Richard (2017). Breeding wheat for drought adaptation: Development of selection tools for root architectural traits. *PhD Thesis, The University of Qld.*

Experimental design

- Replicated design where each platform was treated as a replicate block

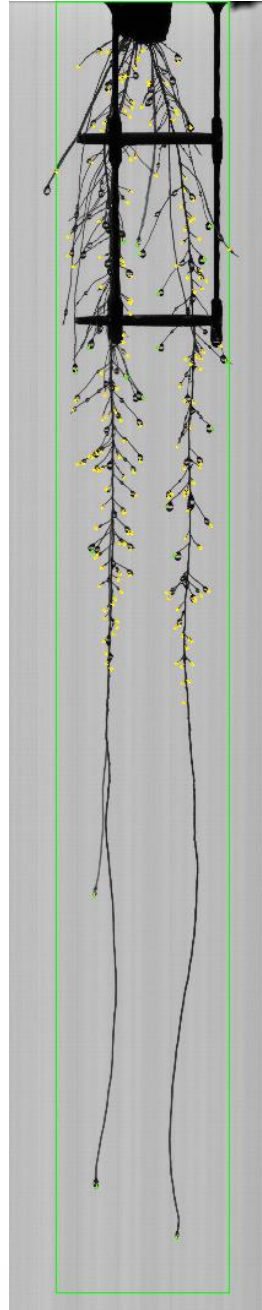


Genetic structure



The Data

1. **Imaging** – 2D linescan digital camera at 720 DPI (dots per inch)
2. **Images** – backlight illumination used to produce high-contrast greyscale images, with a spatial resolution of 3.5 μm
3. **Tip detection** – tips are filtered out and recorded as the x, y coordinates of root tips for each image at each time point
4. **Image data** – 163 images per plant, with up to 8000 tips per plant in total. 570Gb of data.



Root system architecture

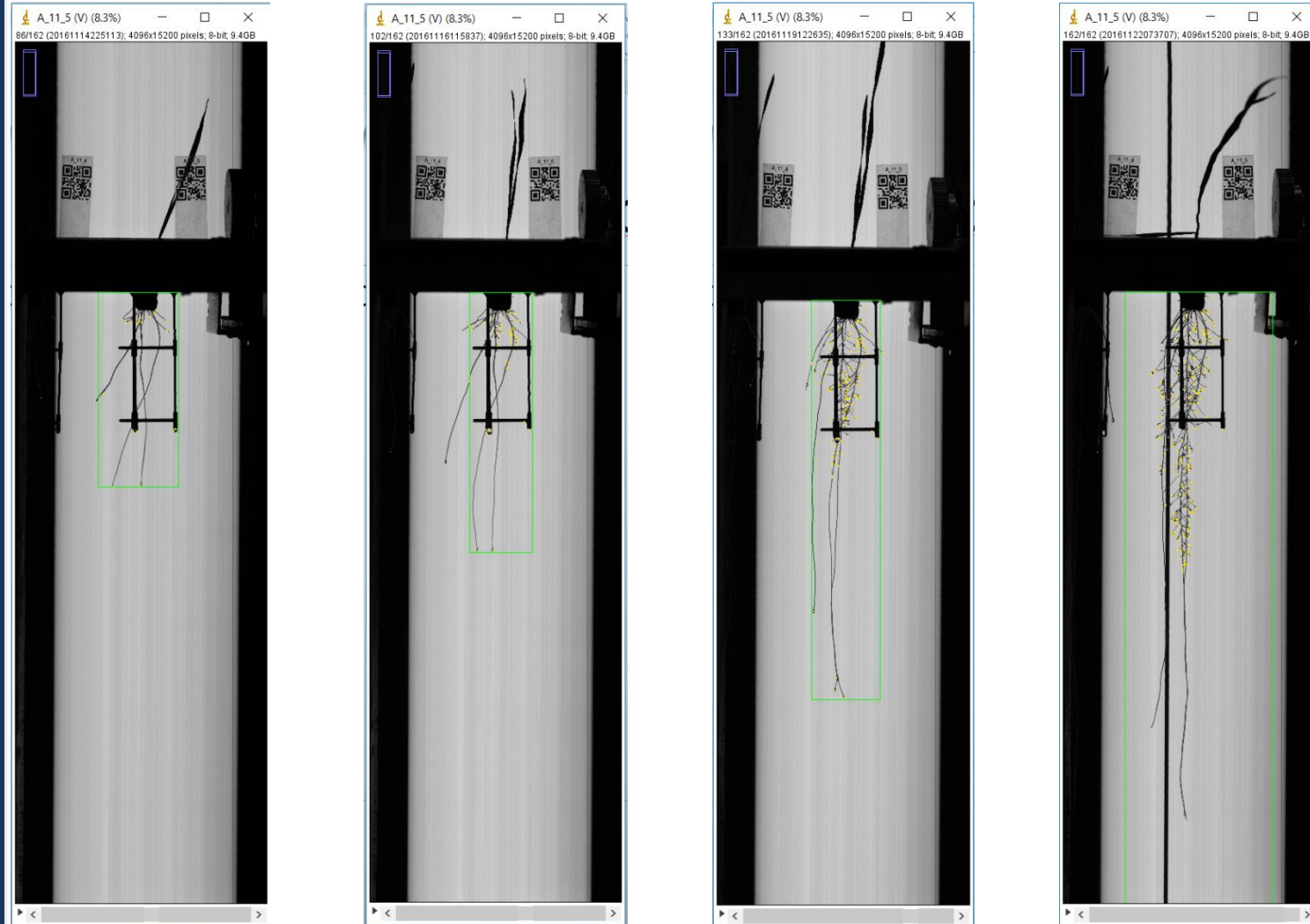
- Root tips can be filtered and chunks of trajectories used to reconstruct growth rates, root numbers and branching frequencies using machine learning techniques (Thaon, 2018)
- Simple traits such as root number and root depth can be derived and analysed (Royer et al., 2020)



Trait derivation

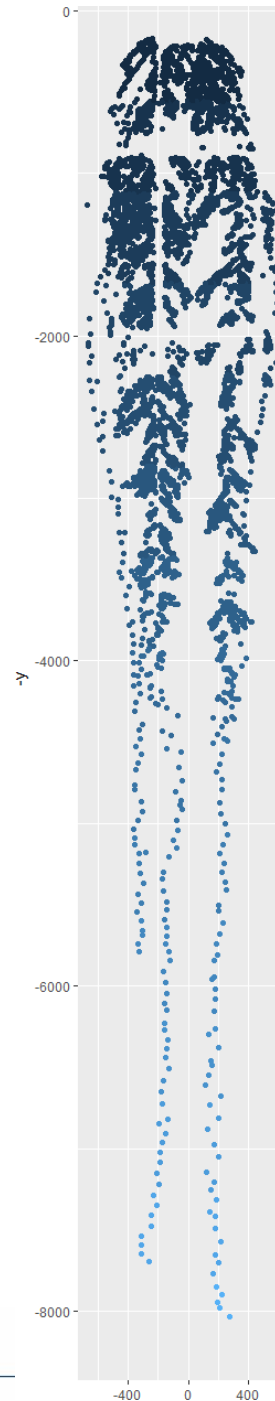
- **Trait derivation** – simple traits can be derived from the point data by counting, tracking and through summary statistics
- Root tips can be filtered and chunks of trajectories used to reconstruct growth rates, root numbers and branching frequencies using machine learning techniques (Thaon, 2018)
- Simple traits such as root number and root depth can be derived and analysed (Royer et al., 2020)

Images of root growth for one plant over time



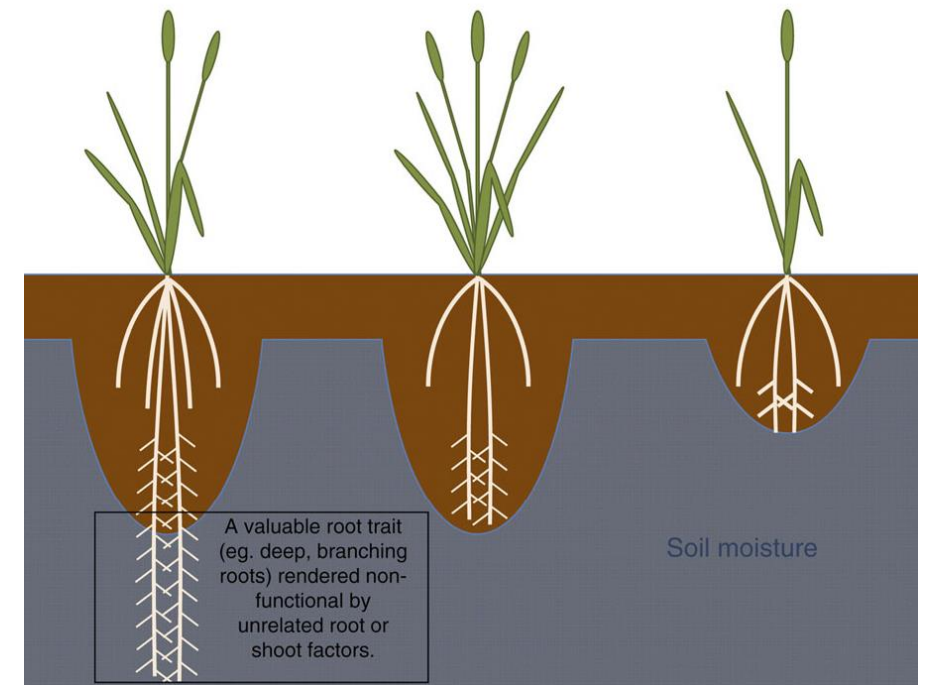
The response variable

- The response forms a spatial point pattern as the coordinates of root tips form part of the response
- The spatial arrangement of points is of interest
- The response is also generated over time, as images are captured as the root grows
- Here, we will just consider the spatial configuration of root patterns at the final time point



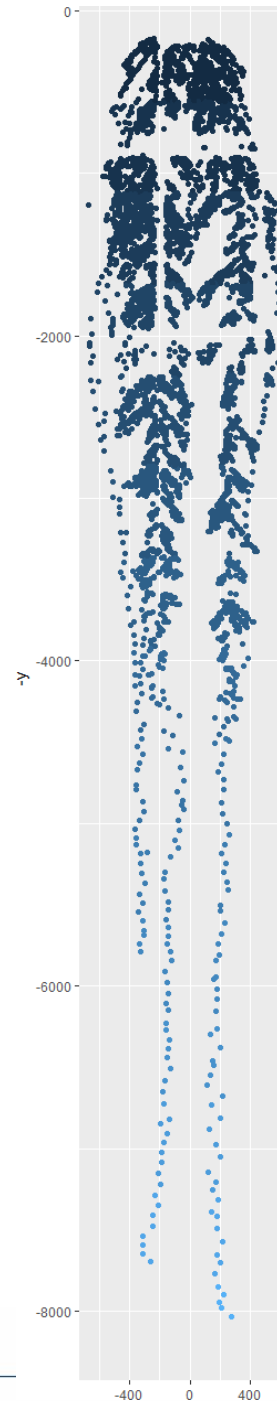
Root system architecture

Root systems of cereals access moisture and nutrients through soil contact, so it is pragmatic to model the spatial pattern of root coverage and estimate how this changes over time



Spatial point process

- Inhomogeneous process, where intensity varies with spatial location
- Time is a numeric variable and can be treated as a mark in the point process (not considered here)
- We have replicates of the point patterns from our designed experiment
- Predictions from this model are the **log(intensity)** and the change in intensity over space



For one point pattern

$$\lambda(u) = \exp(\beta_0 + \beta_1 x + \beta_2 y)$$

where u is the point with spatial coordinates (x, y)

Log intensity: $\sim \text{abs}(x) + y$

Fitted trend coefficients:

(Intercept)	abs(x)	y
-4.6550395191	-0.0046586344	0.0004569481

	Estimate	S.E.
(Intercept)	-4.6550395191	1.913053e-02
abs(x)	-0.0046586344	5.173826e-05
y	0.0004569481	5.544517e-06

- All point process models were fitted in R using the *spatstat* function

Baddeley A, Rubak E, Turner R (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London.

The genetic model

- Single-stage analysis of the derived trait from the point process model, molecular markers and population design
- Population structure captured through the probability of inheriting founders' alleles
- Perform forward marker selection to build a multi-marker model
- An iterative process based on a linear mixed model

Paccapelo et al. (2018) A whole genome QTL analysis for NAM populations based on

Verbyla et al. (2014) Whole-genome QTL analysis for MAGIC, *Theoretical and Applied Genetics*.

A whole genome NAM model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_g\mathbf{g} + \boldsymbol{\epsilon}$$

\mathbf{y} : trait data, taken as β_i where $i=0,1,2$ from the point pattern model

$\mathbf{X}, \mathbf{Z}_u, \mathbf{Z}_g$: incidence matrices

$\boldsymbol{\tau}$: fixed effects parameters

\mathbf{u} : non-genetic random effects $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}_u)$

$\boldsymbol{\epsilon}$: residual $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$

\mathbf{g} : genetic random effects

The multi-marker model with probabilities to account for the NAM structure

$$g_i = \sum_{j=1}^{n_m^*} \mathbf{p}_{ij}^T \mathbf{a}_j + u_{p_i}$$

g_i : genetic effect of individual i .

$\mathbf{p}_{ij} = (p_{ijk})$: $n_f \times 1$ vector of founder probabilities

p_{ijk} is the probability for individual i that the allele at marker j is coming from founder k ($k = 1, 2, \dots, n_f$).

$\mathbf{a}_j = (a_{jk})$: $n_f \times 1$ vector of effects of each founder allele at marker j .

u_{p_i} : polygenic effect for individual i with $\mathbf{u}_p \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$.

The marker selection method

An iterative process with the following steps

1. Fit the model for random genetic effects, g
2. Test $H_0: \sigma_a^2 = 0$. If rejected, select the most important marker (outlier statistic)
3. Update the probability matrix excluding the selected marker $\mathbf{P}_{A,-1}$
4. Update the model in Step 1 and include the selected marker as a random effect ($+\mathbf{P}_{A,1}\mathbf{a}_1$)
5. Repeat from step 2. until H_0 is not rejected

Verbyla et al. (2012) Rwgaim: an efficient high-dimensional random whole genome average (qtl) interval mapping approach. *Genetics Research*.

The multi-marker model with probabilities to account for the NAM structure

$$\mathbf{g} = \mathbf{P}_A \mathbf{a} + \mathbf{u}_p$$

\mathbf{P}_A contains the allele probabilities of each founder by marker for each individual and is of size $(n_g \times n_f n_m)$

\mathbf{a} is the vector of effects of each founder by marker allele for each individual and is of size $(n_f n_m \times 1)$

$$\mathbf{a} \sim N(\mathbf{0}, I\sigma_a^2)$$

\mathbf{u}_p is the vector of polygenic effects and is of size $(n_g \times 1)$

$$\mathbf{u}_p \sim N(\mathbf{0}, I\sigma_p^2).$$

The whole genome NAM model

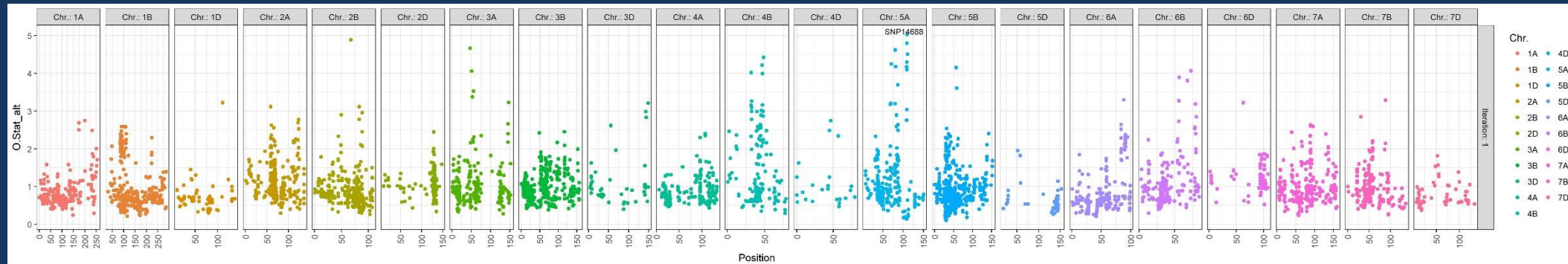


Forward selection of a multi-marker model for

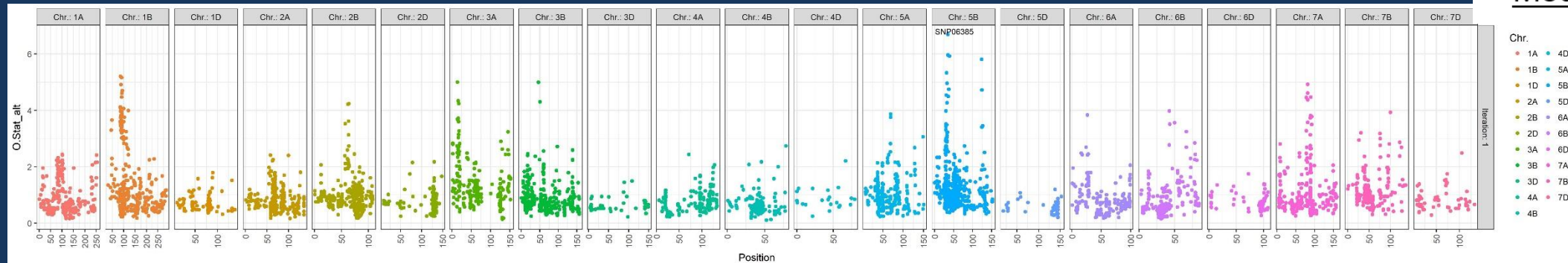
- 1) Root density – log(intensity)
- 2) Change in intensity over depth

An iterative process in R using the packages

- ASReml-R (Butler, 2009)
- mpwgaim (Verbyla et al. 2014)



Root
Intensity



Model 1: $g = P_A a + u_p$

Change in
Intensity
over depth

Summary and Future work

- A hierarchy of models has been utilised for phenotyping, including
 - A point process model to predict spatial intensity of root tips
 - Using replicated data in a linear mixed model to partition experimental design effects from genetic effects and residual errors for predicted intensity
- We have developed and implemented a novel method for QTL mapping in a NAM population to determine genetic regions associated with intensity of root tips
- Integration of the spatial point process into the linear mixed model can be improved
 - Use of weights for intensity estimates in the LMM
 - Incorporation of an appropriate model over time for root growth

Spatial point pattern

$$\lambda(u) = \exp(\beta_0 + \beta_1 x + \beta_2 y)$$

where u is the point with spatial coordinates (x,y)

Linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_u \mathbf{u} + \mathbf{Z}_g \mathbf{g} + \boldsymbol{\epsilon}$$

where \mathbf{y} is the vector of coefficients, $\boldsymbol{\beta}_i$, estimated in the point process model for each plant

partitioned for allele effects

$$\mathbf{g} = \mathbf{P}_A \mathbf{a} + \mathbf{u}_p$$

References

Baddeley A, Rubak E, Turner R (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London.

Paccapelo et al. (2018) A whole genome QTL analysis for NAM populations *EUCARPIA Biometrics in Plant Breeding*.

Paccapelo et al. (2020) WGNAM: whole-genome nested association mapping (*in prep*).

Richard C. (2017) Breeding wheat for drought adaptation: Development of selection tools for root architectural traits. *PhD Thesis, The University of Qld*.

Royer et al. (2020) High-throughput phenotyping and genetics of wheat root traits (*in prep*).

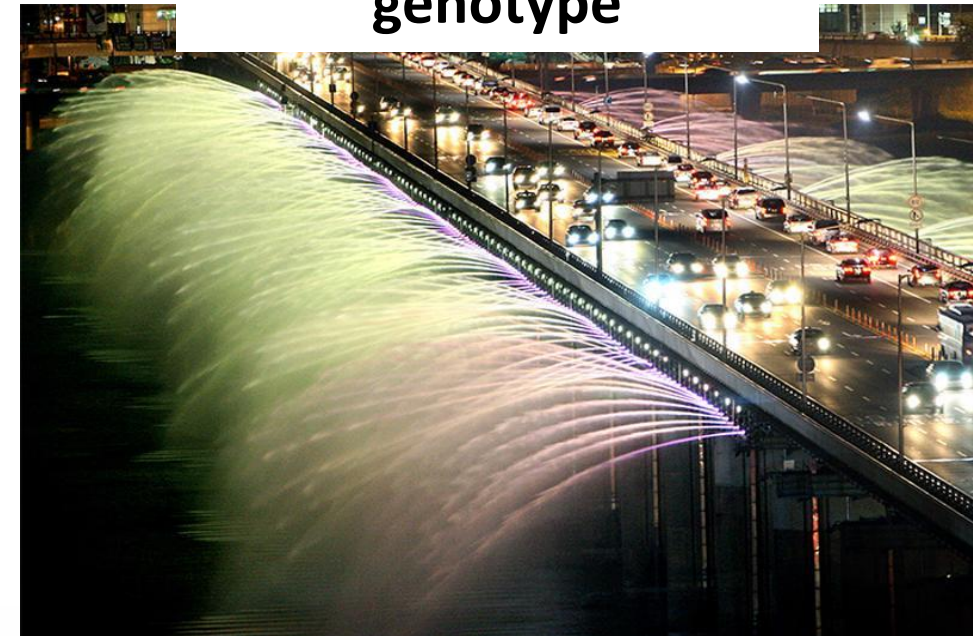
Thaon P. (2108) Genetic Analysis of Wheat Root Architecture in Aeroponics. *Masters Thesis, The University of Louvain*.

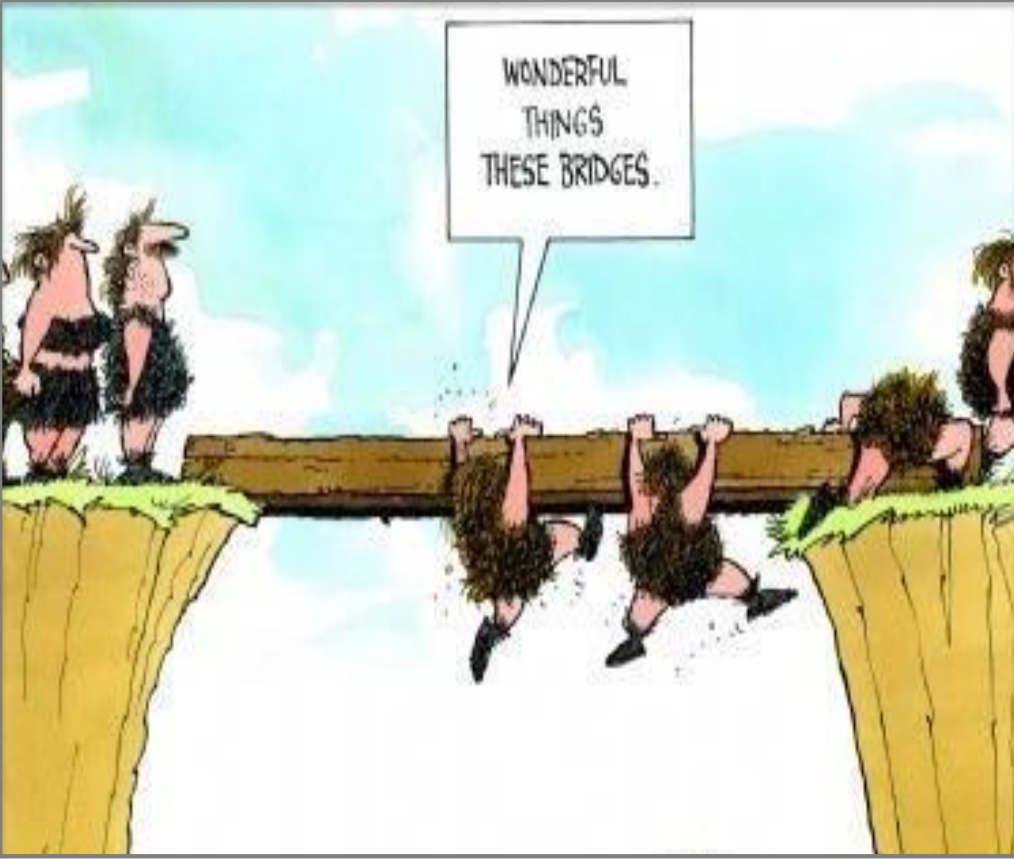
Verbyla et al. (2012) Rwgaim: an efficient high-dimensional random whole genome average (qtl) interval mapping approach. *Genetics Research*.

Verbyla et al. (2014) Whole-genome QTL analysis for MAGIC, *Theoretical and Applied Genetics*.



**Bridging the gap from
phenotype to
genotype**





Acknowledgements

Alison Kelly, Valeria Paccapelo, Michelle Adams-Beutel

Jack Christopher, Karine Chenu, Cecile Richard, Alice Royer

Patrick Thaon, Xavier Draye



© State of Queensland 2015



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance
for Agriculture & Food Innovation



GRDC
GRAINS RESEARCH &
DEVELOPMENT CORPORATION



SAGI-North
Statistics for the Australian Grains Industry – Northern Node

UCL
Université
catholique
de Louvain