# Semi-automated image analysis of root architecture and early root development in faba bean and white clover and genomic estimation of breeding values and correlations

Istvan Nagy[1,*,†], Peter Skov Kristensen[1,†], Marta Malinowska[2], Linda Kærgaard Nielsen[3], Andrea Schiemann[4], Niels Rolund[2], Stig Uggerhøj Andersen[5], and Torben Asp[1]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Forsøgsvej 1, DK-4200 Slagelse, Denmark
[2]DLF Seeds A/S, Ny Østergade 9, DK-4000 Roskilde, Denmark
[3]Sejet Plant Breeding, Nørremarksvej 67, DK-8700 Horsens, Denmark
[4]Nordic Seed, Kornmarken 1, DK-8464 Galten, Denmark
[5]Department of Molecular Biology and Genetics, Aarhus University, Universitetsbyen 83, DK-8000 Aarhus, Denmark
[†]These authors contributed equally
[*]Correspondence: Istvan.Nagy@qgg.au.dk

## Abstract

Protein-rich leguminous plants, like faba bean and white clover are prospectively interesting crops in the North-European countries for reducing dependence on soybean import. Significant expansion of the production area of protein crops is challenged by the sub-optimal climatic conditions in this region, especially by the increasing probability of year-to-year fluctuation of extreme weather conditions due to global climate changes. To overcome these challenges, development of new climate-resilient varieties suitable for growing under Northern-European conditions are needed. Root architecture and early root development, as well as the availability of efficient root phenotyping technologies are crucial factors of advancing in breeding of adequate varieties. Multivariate models are valuable for integrating multiple correlated traits into the breeding process. Estimation of genetic correlations between traits can potentially improve the accuracy of genomic estimated breeding values.To facilitate quantitative genetics studies in this field, we conducted greenhouse rhizobox experiments using faba bean and white clover genotypes, including standard varieties and breeding lines. We report a study of semi-automated, image-based root phenotyping technology in connection with a relatively simple and affordable technology for screening early root development and provide a conceptual pipeline for genomic estimation of breeding values and correlating greenhouse- and field yield data.

## Keywords

Root System Architecture, early root development, image analysis, multivariate models, faba bean, white clover

## Introduction

Agriculture and food industry of Northern European countries, including Denmark are greatly depending on import of crop products providing protein resources - mainly soybean and processed soybean products. In order to diminish dependence on soybean import, there is an increasing interest in the Nordic countries for expanding local production of protein-rich leguminous crops, like faba bean (*Vicia faba* L.) and white clover (*Trifolium repens* L.). However, significant expansion of the production area of protein crops is challenged by the sub-optimal climatic conditions in these regions, especially by the increasing

probability of year-to-year fluctuation of extreme weather conditions due to global climate changes. To overcome these challenges, development of new climate-resilient varieties suitable for growing under Northern-European conditions are needed. Root architecture and early root development, as well as the availability of efficient root phenotyping technologies are crucial factors of advancing in breeding of adequate varieties.

Root system architecture (RSA) increasingly attracts the interest of scientist in the era of climate change. RSA comprises the shape and spatial arrangement of the root system within the soil[1] and is determined by genetic- as well as environmental factors, like soil temperature and water- and nutrient supply, and roots exhibit a high level of phenotypic plasticity in response of environmental conditions. Root architecture is a primarily important factor of efficient nutrient and water uptake, and a 'second green revolution' has been envisioned that deploys crops with improved below ground traits.[2,3] The development of new plant varieties with improved root phenotypes requires advances in the characterization of root development and morphology, as well as in the relationships between genetic factors and RSA. As being a complex, polygenic trait, studying of RSA should be conducted by parameterization of its elements, such as root depth, total root length, root angle, root thickness, root density and root surface area.[4,5] A wide range of root phenotyping technologies have been developed and applied during the last decades, ranging from excavating root crowns on the fields ("showelomics"[6]) to semi-field based micro-rhizothron facilities with automated image acquisition.[7] Non-destructive 3D root phenotyping using tomographic techniques, like X-ray computed tomography, magnetic resonance imaging (MRI) and positron emission tomography (PET) also become achievable during the last decades (for a review refer[3]). Soil-filled slab chambers with transparent lids (rhizoboxes) offer a reasonable and space-efficient approach for studying early root development under controlled conditions. Investigations on early root development are of fundamental importance, as a vigorous early root system is essential for healthy plant growth and establishment, especially under sub-optimal conditions[8] and recent quantitative genetics studies demonstrated that early root development traits can act as proxies for field yield across multiple years and locations.[9] Earlier studies identified relations between early root traits and crop productivity and/or nutrient- and water use efficiency also in other species, like maize[10,11] and rice[12] as well. Such relations have been confirmed by the co-occurrence of quantitative trait loci (QTLs) between early root traits and yield components.[11,13,14]

The decreasing genotyping costs and the availability of high-quality genomic references in most important crop species provide opportunities for breeders to obtain abundant genotyping information at affordable costs and to utilize large-scale genotype data for genomic predictions along with phenotpype data collected by multiple technologies across multiple years and locations. However, for the successful implementation of such efforts, the development of appropriate statistical models and methodologies is necessary. For polygenic quantitative traits, like RSA components, genomic selection offers opportunities for breeders to increase the rate of genetic gain, by reducing the time needed to complete a breeding cycle, increasing the selection intensity and by utilizing within-family variation that can be captured using molecular markers.[15] Genomic selection is a form of Marker Assisted Selection that simultaneously estimates all locus, haplotype or marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs[16]).

Multivariate models are valuable for integrating multiple correlated traits into the breeding process and estimation of genetic correlations between traits and can potentially improve the accuracy of genomic estimated breeding values.[17–19] To facilitate quantitative genetics studies to correlate early root development traits and yield components, we conducted greenhouse rhizobox experiments using faba bean and white clover genotypes, including standard varieties and breeding lines.

# Materials and methods

## Plant material

For faba bean, 180 lines from the breeding program of the Danish breeding companies Nordic Seeds (75 lines) and Sejet (75 lines), as well as 10 variety standards and 20 core lines belonging to the IMFABA Consortium (https://projects.au.dk/fabagenome/genomics-data) were used in the study. For white clover, a panel consisting of 174 lines selected from 20 commercial varieties with diverse agronomic characters was used. Initial white clover plant material was maintained in the greenhouse as clonally propagated stocks in 17cm pots.

## Rhizobox setup and growing conditions

Plants were grown in custom-made plastic rhizoboxes (internal dimensions: 36x18x2.5cm). Boxes were filled with 1.8L substrate (a mix of turf, local topsoil and sand) and supplemented with 300ml water. In case of faba bean, two seeds were sown in each box. For white clover, 8 to 12 cm long cuttings of young lateral shoots were collected from the middle part of the pots. It was attempted that cuttings have a nodus above the cutting site and fully developed leaflets below the shoot tip. During processing, cuttings were kept between wet filter paper sheets. Boxes were placed in the greenhouse on racks that ensured to incline them at 60° to facilitate root growth along the visible side. During growing, the transparent plexiglass front side of the boxes was covered with a black foil. For both plant species, experiments were conducted in three subsequent batches in randomized blocks throughout a 12 month period in the same greenhouse compartment equipped with an automatic shadowing system during summer and additional lighting and heating (16/8h day/night at 21/18 °C) during winter. In each replicate, plants were grown for 25 to 28 days, by applying manual watering with equal amounts (100 to 200ml) of tap water in 5 to 10 days periods, depending on humidity and outer temperature.

## Phenotype data collection

### Root imaging

On the final day of growing period, in each batch experiment root images were taken by a horizontal office scanner (Epson Perfection V700) with a modified lid. Root imaging was conducted by the RootPainter software.[20] The analysis procedure involved training with manual annotation of clear foreground and background regions, and corrective annotations to refine the segmentation model. The final model was used to automatically segment the entire root dataset. In addition, segmented root images were exported and further processed by the RhizoVision software[21] to quantify and record the following root phenotype

components: total root length, number of root tips, number of branching points and branching frequency, root diameter and perimeter, root volume and surface area.

### White clover leaf phenotyping

For leaf phenotyping, in each genotype, three fully developed trifoliate leaves were collected from the source plants maintained in the greenhouse. Separated leaflets (9 leaflets for each genotype) were fixed on white paper sheets by slight glueing their back side and scanned in full-color mode at the highest possible resolution. The scanned images were used for digital image analysis to determine leaf size, leaf shape and leaf color. Leaf size (average surface area of leaflets) and leaf shape parameters (circularity and solidity) were determined by ImageJ (v1.54g).[22] For leaf color analysis, 20x20 pixel squares were cut out from the middle part of the leaflet images and saved to separate files using the image editing software Inkscape.[23] Average pixel RGB values were calculated from the 400px image files by the *convert* tool of the ImageMagick software[24] using a single-line command including the following formula:

$$\begin{aligned} RGB = convert\ [400px_image_file]\ \text{-}resize\ 1x1\backslash! \\ \text{-}format\ \text{"}\%[fx:int(255*r+.5)],\%[fx:int(255*g+.5)],\%[fx:int(255*b+.5)]\text{"} \end{aligned} \tag{1}$$

For each line, RGB values were converted to perceived brightness values using the following formula (Luma conversion for perceived brightness, in which the green component has the main effect):

$$Y = 0.2126 * R + 0.7152 * G + 0.0722 * B \tag{2}$$

### Faba bean field phenotype data

Faba bean breeding lines of two Danish breeding companies Nordic Seed (75 lines) and Sejet (75 lines), as well as 10 variety standards were tested in replicated multi-year field trials in 3 years (2022, 2023, 2024) at 3 locations (Dyngby, Svinø and Gamborg, Denmark) with 3 replicates. Plant height, Thousand Kernel Weight (TKW), grain yield and grain protein content were recorded, resulting in ca. 4000 observations in total for each trait.

### White clover field phenotype data

A two year field experiment was conducted in Haldrup, Denmark, with 12.5 $m^2$ plots in two replicates. Fresh and dry biomass, as well as Near Infrared Spectroscopy based quality values were recorded after four cuts in the first year and five cuts in the second year.

### White clover greenhouse biomass data

Image-based biomass measurements were carried out on a panel of white clover lines overlapping with the panels of rhizobox- and field experiments. For each line, 10 individual plants were monitored using stationary cameras to estimate cumulated biomass after 10 days of growth.[25]

## Genetic variant data

### Faba bean variant data

Variant data based on pseudo-chromosome references of the latest version of the *Vicia faba* genome sequences (Hedin/2 v.2),[26] containing 122291 bi-allelic SNPs was provided by the IMFABA consortium (https://projects.au.dk/fabagenome/genomics-data). Of the 178 lines used for the rhizobox experiments, 157 were represented in the VCF file.

### White clover variant data

Genetic variant data based on Genotyping by Sequencing (GBS) short-read sequences mapped onto the *Trifolium repens* v.4.0 genome references.[27] Of the 195 genotypes represented in the original variant discovery project, 167 were common with the line set used in the root phenotyping experiments, therefore variant data for the not matching lines were removed from the VCF file and genomic analysis studies was further on restricted to these 167 lines). The VCF file was further on filtered for a minimum Minor Allele Frequency of 3%, maximum number of missing data of 30% across lines and a minimum Read Depth of 5, keeping bi-allelic variants on 280265 loci.

# Genomic analysis conception

### Calculating variance components and heritabilities

Variance components of all models were estimated by restricted maximum likelihood using the software package DMU for analyzing Multivariate Mixed Models (provided by the Center of Quantitative Genetis and Genomics, Aarhus University[28]).
Heritabilities at field plot or rhizobox level were calculated as narrow sense using the following formula:

$$h^2 = \frac{\sigma_g^2}{\sigma_p^2} \tag{3}$$

and as broad sense using the following formula:

$$H^2 = \frac{\sigma_g^2 + \sigma_i^2}{\sigma_p^2} \tag{4}$$

where $\sigma_p^2$ is the sum of all estimated variance components within each of the models.

Correlations between traits were calculated as Pearson correlations between estimated additive genomic breeding values for the respective traits.

### Correlated response to selection

The genetic response to selection per generation of a trait can be calculated using the breeder's equation:

$$R = ih\sigma_A \tag{5}$$

where ***i*** is selection intensity, ***h*** is the square root of the narrow-sense heritability, $\sigma_A$ is the genetic standard deviation of the trait (i.e. the square root of the genetic variance). The genetic gain of a trait can be improved by incorporating information from another genetically correlated trait. To determine how much genetic gain can be achieved for a primary trait (***Y***), when selection is made based on a secondary trait (***X***), the correlated response to selection can be calculated as:

$$CR_Y = ih_X\gamma_A\sigma_{AY} = ih_X h_Y\gamma_A\sigma_{PY} \tag{6}$$

where $h_X h_Y \gamma_A$ is the co-heritability of the two traits ***X*** and ***Y***, $\gamma_A$ is the genetic correlation between traits ***X*** and ***Y***, the genetic and phenotypic standard deviations of trait ***Y*** are $\sigma_{AY}$,$\sigma_{AP}$ and $\sigma_{PY}$ respectively, according to Falconer and Mackay.[29]

Thus, the genetic gain of trait ***Y***, when selecting indirectly based on trait ***X***, will better in cases where the heritability of trait ***X*** is high, where the genetic correlation between the traits is high, or where the selection intensity can be increased.

Yield is often an important breeding goal, but it is relatively expensive and time-consuming to get phenotypic observations for yield, since varieties ideally should be tested in replicated field trials in several locations and several years due to potential spatial variation within fields and due to Genetics by Environment (GxE) interactions across different environments. Therefore, yield cannot directly be measured very accurately early in breeding programs, when only one or few plants are available per variety. However, if other traits can be identified that are genetically correlated to yield and that can be phenotyped more easily and/or in shorter time, e.g. in controlled conditions, using few seeds or plants per variety, then such traits would be useful to implement in early selection stages of breeding programs in order to increase the genetic gain of yield.

GBLUP (Genomic Best Linear Unbiased Prediction) models were used in order to partition environmental, spatial, and genetic effects from each other for the different traits. Here, genomic relationships between lines based on SNP markers were included in order to estimate breeding values (additive genetic effects) for each trait, heritabilities (i.e. the ratio between genetic variance and total phenotypic variance), and to estimate the genetic correlations between traits.[29]

Dominance and epistatic genomic relationships were not explicitly included in the models, since large data sets are required to estimate such non-additive genetic effects. Instead, the line variance was included in order to capture non-additive genetic variance as well as any additive genetic variance that was not captured by the SNP markers used in the genomic relationship matrix. The line variance can only be estimated for the traits, where phenotypic observations are available for more than one replicate, since the line variance will be confounded with the residual variance if only one observation is available per line.

## Models

### Faba bean models

**Yield data from field trials was modelled by:**

$$y_1 = X_1 b_1 + Z_1 t + Z_2 g_1 + Z_3 i_1 + e_1 \tag{7}$$

where $\boldsymbol{y}_1$ is a vector of phenotypic observations, $\boldsymbol{X}_1$ is a design matrix for fixed effects, $\boldsymbol{b}_1$ is a vector of effects of year-location-replicate, $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are design matrices for random effects, $\boldsymbol{t}$ is a vector of effects of blocks within year-location-replicate with $\boldsymbol{t}\sim N(0,\boldsymbol{I}_1\sigma_t{}^2)$, $\sigma^2{}_t$ is the variance of block effects, $\boldsymbol{g}_1$ is a vector of additive genomic breeding values with $\boldsymbol{g}_1\sim N(0,\boldsymbol{G}_1\sigma^2_{g_1})$, $\boldsymbol{G}_1$ is a genomic relationship matrix calculated using method one proposed by VanRaden[30], $\sigma^2_{g_1}$ is the additive genetic variance, $\boldsymbol{i}_1$ is a vector of line effects with $\boldsymbol{i}_1\sim N(0,\boldsymbol{I}_2\sigma^2_{i_1})$, $\sigma^2_{i_1}$ is the variance of line effects, and $\boldsymbol{e}_1$ is a vector of residual effects with $\boldsymbol{e}_1\sim N(0,\boldsymbol{I}_3\sigma^2_{e_1})$, $\sigma^2_{e_1}$ is the variance of residual effects, and $\boldsymbol{I}_n$ are identity matrices.

**Root length data from rhizoboxes was modelled by:**

$$y_2 = X_2 b_2 + cw + Z_4 g_2 + Z_5 i_2 + e_2 \tag{8}$$

where $\boldsymbol{y}_2$ is a vector of phenotypic observations, $\boldsymbol{X}_2$ is a design matrix for fixed effects, $\boldsymbol{b}_2$ is a vector of effects of experiment, $\boldsymbol{c}$ is a fixed regression coefficient for observed root lengths regressed on thousand seed weights, and $\boldsymbol{w}$ is a vector of thousand seed weights of seeds sown in rhizoboxes, $\boldsymbol{Z}_4$ and $\boldsymbol{Z}_5$ are design matrices for random effects, $\boldsymbol{g}_2$ is a vector of additive genomic breeding values with $\boldsymbol{g}_2\sim N(0,\boldsymbol{G}_1\sigma^2_{g_2})$, $\sigma^2_{g_2}$ is the additive genetic variance, $\boldsymbol{i}_2$ is a vector of line effects with $\boldsymbol{i}_2\sim N(0,\boldsymbol{I}_4\sigma^2_{i_2})$, $\sigma^2_{i_2}$ is the variance of line effects, and $\boldsymbol{e}_2$ is a vector of residual effects with $\boldsymbol{e}_2\sim N(0,\boldsymbol{I}_5\sigma^2_{e_2})$, $\sigma^2_{e_2}$ is the variance of residual effects, and $\boldsymbol{I}_n$ are identity matrices.

Combining the two models above, a bivariate model was used for estimating genetic covariances between field yield and rhizobox root length traits:

$$\begin{bmatrix}\boldsymbol{g}_1\\ \boldsymbol{g}_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{g1} & \sigma_{g1g2}\\ \sigma_{g1g2} & \sigma^2_{g2}\end{bmatrix} \otimes \boldsymbol{G}_1\right), \begin{bmatrix}\boldsymbol{i}_1\\ \boldsymbol{i}_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{i1} & \sigma_{i_1 i_2}\\ \sigma_{i1i2} & \sigma^2_{i2}\end{bmatrix}\right), \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{e1} & 0\\ 0 & \sigma^2_{e2}\end{bmatrix}\right) \tag{9}$$

The remaining random effects were assumed normally and independently distributed with zero means and variances as described above.

The genetic correlation between traits, $\gamma_{g_1g_2}$, can be calculated from the estimated genetic variances and co-variances:

$$\gamma_{g_1g_2} = \frac{\sigma_{g_1g_2}}{\sqrt{\sigma^2_{g1} * \sigma^2_{g2}}} \tag{10}$$

**Clover models**

**Yield data from field trials was modelled by:**

$$y_3 = X_3 b_3 + Z_6 p + Z_7 g_3 + e_3 \tag{11}$$

where $\boldsymbol{y}_3$ is a vector of phenotypic observations, $\boldsymbol{X}_3$ is a design matrix for fixed effects, $\boldsymbol{b}_3$ is a vector of effects of cut number,$\boldsymbol{Z}_6$ , and $\boldsymbol{Z}_7$ are design matrices for random effects, $\boldsymbol{p}$ is a vector of effects of plot with $\boldsymbol{p}\sim N(0,\boldsymbol{I}_6\sigma^2_p)$, $\sigma^2_p$ is the variance of plot effects, $\boldsymbol{g}_1$ is a vector of additive genomic breeding values with $\boldsymbol{g}_3\sim N(0,\boldsymbol{G}_2\sigma^2_{g_3})$, $\boldsymbol{G}_2$ is a genomic relationship matrix, $\sigma^2_{g_3}$ is the additive genetic variance, and $\boldsymbol{e}_3$ is a vector of residual effects with $\boldsymbol{e}_3\sim N(0,\boldsymbol{I}_7\sigma^2_{e_3})$, $\sigma^2_{e_3}$ is the variance of residual effects, and $\boldsymbol{I}_n$ are identity matrices.

**Data from greenhouse was modelled by:**

$$y_4 = X_4 b_4 + Z_8 g_4 + e_4 \tag{12}$$

where $\mathbf{y}_4$ is a vector of phenotypic observations, $\boldsymbol{X}_4$ is a design matrix for fixed effects, $\boldsymbol{b}_4$ is a vector of effects of experiment, $\boldsymbol{Z}_8$ is a design matrix for random effects, $\boldsymbol{g}_4$ is a vector of additive genomic breeding values with $\boldsymbol{g}_4 \sim N(0, \boldsymbol{G}_2 \sigma^2_{g_4})$, $\sigma^2_{g_4}$ is the additive genetic variance, and $\boldsymbol{e}_4$ is a vector of residual effects with is the variance of residual effects with $\boldsymbol{e}_4 \sim N(0, \boldsymbol{I}_8 \sigma^2_{e_4})$, $\sigma^2_{e_4}$ is the variance of residual effects, and $\boldsymbol{I}_n$ identity matrices.

# Results and discussion

## Rhizobox experiments

For the two species used the study different setup was applied for the rhizobox experiments: direct sowing of seeds into the boxes in case of faba bean and transplantig shoot segments in case of white clover. The later approach represented more challenges concerning uniformity of the initial material, as the stock plants maintained in the greenhouse appeared to perform a certain level of inherent inhomogeneity. Furthermore, as the batch experiments were carried in a raw throughout the year in the same greenhouse compartment, seasonal differences certainly influenced the consistency of data obtained from different experiment batches. Roots systems developed the rhizoboxes were scanned when roots reached about two third of the hight of the boxes. This took 12 to 15 days in case of faba bean, and 15 to 20 days in case of white clover (Figure 1). The RootPainter software enabled a robust and reliable root image segmentation. Image-based quantification resulted in approximately normal distributions for most parameters (Figure 2), of which the total root length values proved to be the most robust and reliable ones, while we found that vertical measurements, like diameter or perimeter are tending to be more influenced by background noise and/or overlapping side roots and image-based quantification is generally less reliable for such parameters.

## Genetic variant data, diversity and population structure

In case of both species, high-quality reference genomes and large sets of genetic variant data were available for the lines included in the study (bi-allelic SNPs on 122291 loci for faba bean and on 280265 loci for white clover). In faba bean, the core plant material included 75-75 breeding lines from two Danish breeding companies. According to the PCA plot based on genetic variant data is showing that the majority of these faba breeding lines are separating to distinct clusters according to company origin, suggesting that the breeding programs of the companies is predominantly based in different germ plasm resources. resources (Figure 3). For white clover, a panel consisting of 174 lines selected from 20 commercial varieties (4 to 10 lines for each variety) was used. In both species, the available genetic variant data enabled to build robust and reliable Genomic Relationship Matrices for genomic estimation of breeding values and correlations (Figure 4).

Table 1: Genomic estimations and heritabilities - Faba bean

| | Variance component | | | | Heritability | |
|---|---|---|---|---|---|---|
| | **Block** | **Id$_g$** | **Id$_l$** | **e** | **h$^2$** | **H$^2$** |
| **Grain yield** | | | | | | |
| Estimate (*Standard error*) | 15.2 (*1.1*) | 8.8 (*2.3*) | 2.6 (*1.4*) | 23.1 (*0.6*) | 0.18 | 0.23 |
| **Protein content** | | | | | | |
| Estimate (*Standard error*) | 0.15 (*0.03*) | 1.35 (*0.26*) | 0.16 (*0.13*) | 1.33 (*0.04*) | 0.45 | 0.50 |
| **Total root length** | | | | | | |
| Estimate (*Standard error*) | - | 510.1 (*473.8*) | 405.0 (*529.1*) | 6356.3 (*516.3*) | 0.07 | 0.13 |

## Genomic estimations and correlations

### Faba bean

The narrow-sense heritability at plot level for faba bean grain yield was 0.18, and the broad-sense heritability was 0.23, which indicates that the genetic control of yield is largely due to additive genetic effects and to a smaller extent due to non-additive genetic effects. On the other hand, the narrow-sense and broad-sense heritabilities for total root length of 0.07 and 0.13, respectively, indicate that additive and non-additive genetic effects were approximately equally important for root length, and that root length could also be strongly affected by other factors, such as environmental effects or GxE interactions. Based on the bivariate model for grain yield and total root length, the estimated additive genetic correlation between the two traits was 0.83 (SE = 0.47), and the estimated correlation of the line effects was 0.53 (SE = 0.45). So these two traits seem to be highly genetically correlated, although the standard errors of the estimated correlations were quite large. Larger datasets should preferably be used to estimate the different genetic parameters of the bivariate models more accurately. Due to the high correlations, phenotyping and selection of early root length could potentially be useful in breeding programs to increase the genetic gain for grain yield (Table 1).

### White clover

Moderately positive genetic correlations of 0.26 between estimated breeding values of yield from field cuts and of both leaf size and of leaf solidity were identified. For leaf size and leaf solidity, moderate to high narrow-sense heritabilities of 0.46 and 0.63, respectively, were estimated. Both traits had higher narrow-sense heritabilities than 0.22 that was estimated for field yield, and they could therefore be suitable indicator traits in breeding programs for improving yield by indirect selection. The leaf traits can be measured on few, young plants grown in controlled greenhouse conditions, and selection based on these traits might thereby be used at low costs in early stages of breeding programs to increase the genetic gain of yield. Furthermore, phenotyping such traits on selection candidates may improve genomic prediction accuracies for yield in cases where larger training sets are available to use for multivariate models (Table 2, Figure 5).

Table 2: Genomic estimations and heritabilities - White clover

| | **Heritability**[a] | **KGGM** | **TRL** | **LS** | **Lsize** | **iSize_GEBV**[b] | **gpdCor_GEBV**[c] |
|---|---|---|---|---|---|---|---|
| Drymatter, kg (KGGM) | 0.22 | 1.00 | 0.17 | 0.26 | 0.26 | 0.21 | -0.13 |
| Total root length (TRL) | 0.02 | 0.17 | 1.00 | 0.08 | 0.11 | -0.06 | 0.13 |
| Leaf solidity (LS) | 0.63 | 0.26 | 0.08 | 1.00 | 0.61 | 0.27 | -0.07 |
| Leaf size (Lsize) | 0.46 | 0.26 | 0.11 | 0.61 | 1.00 | 0.40 | -0.01 |
| iSize_GEBV | 0.83 | 0.21 | -0.06 | 0.27 | 0.40 | 1.00 | -0.06 |
| gpdCor_GEBV | 0.24 | -0.13 | 0.13 | -0.07 | -0.01 | -0.06 | 1.00 |

[a]Narrow sense
[b]Image-based biomass measurements calculated from pixel counts in the first 10 days of growth
[c]Dry weight per plant corrected for the full effect of iSize

# Figures

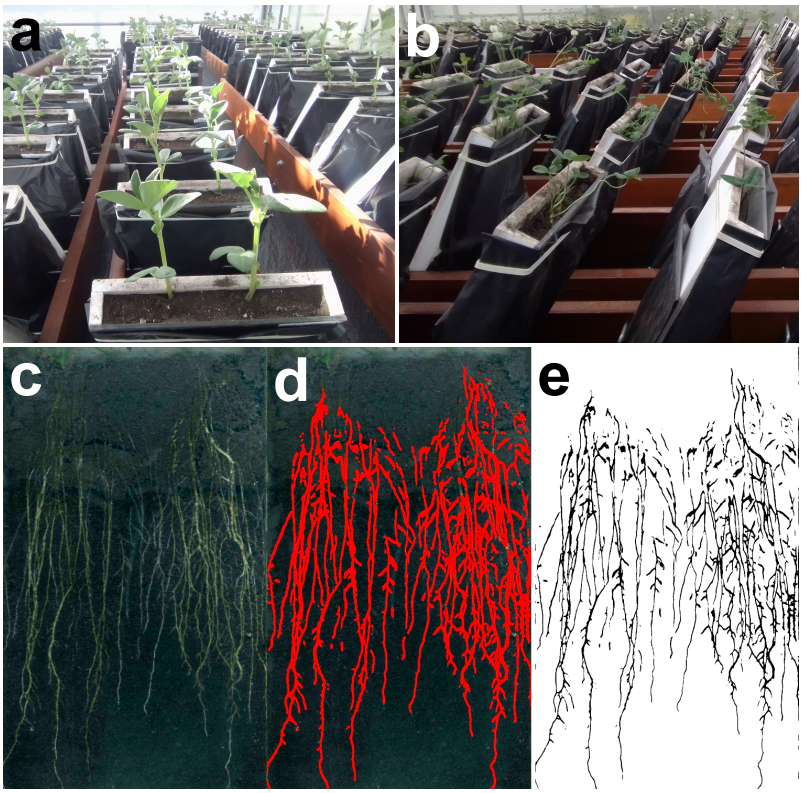



Figure 1: Greenhouse rhizobox experiments and root image processing
**(a)** Faba bean seedlings in rhizoboxes
**(b)** White clover explants in rhizoboxes
**(c)** Scanned root image
**(d)** Segmentation image made by RootPainter
**(e)** Segmentation image exported to analysis by RhizoVision

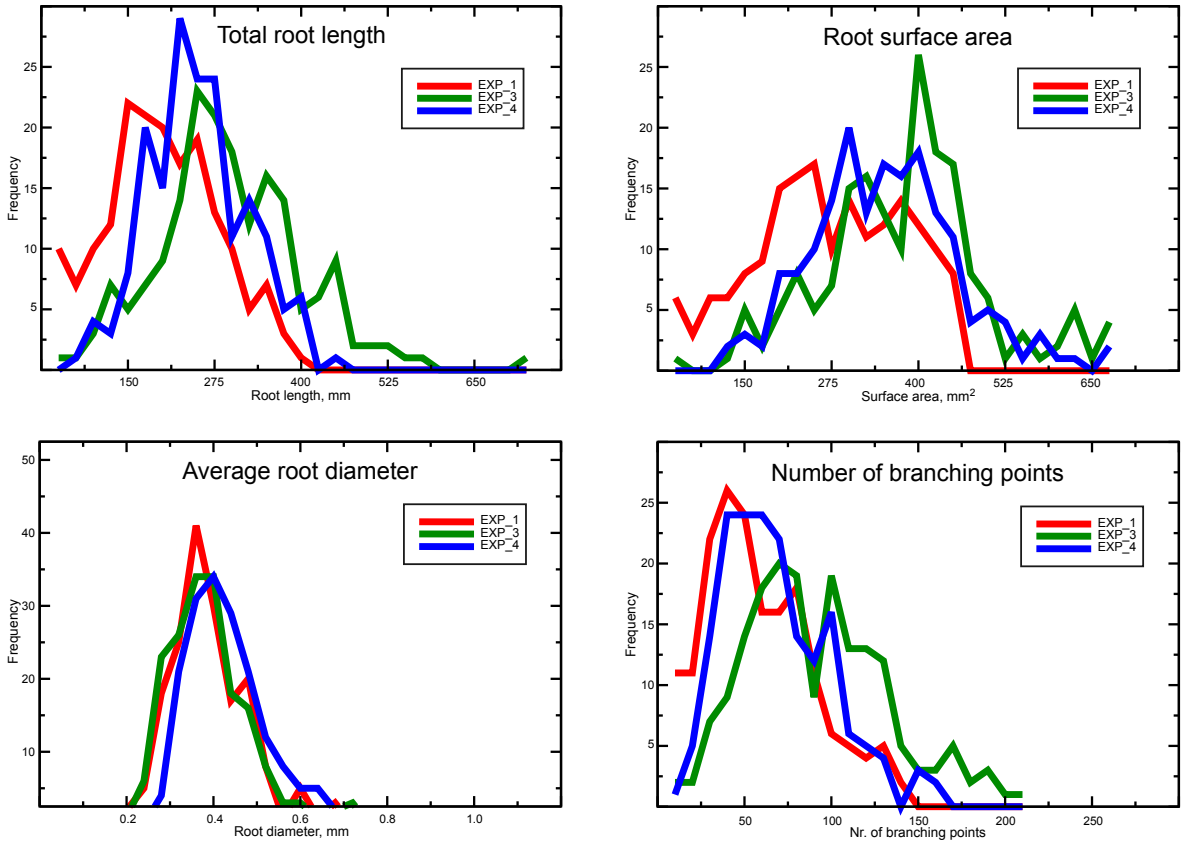



Figure 2: Distribution of quantitative root parameter values in three faba bean rhizobox batch experiments

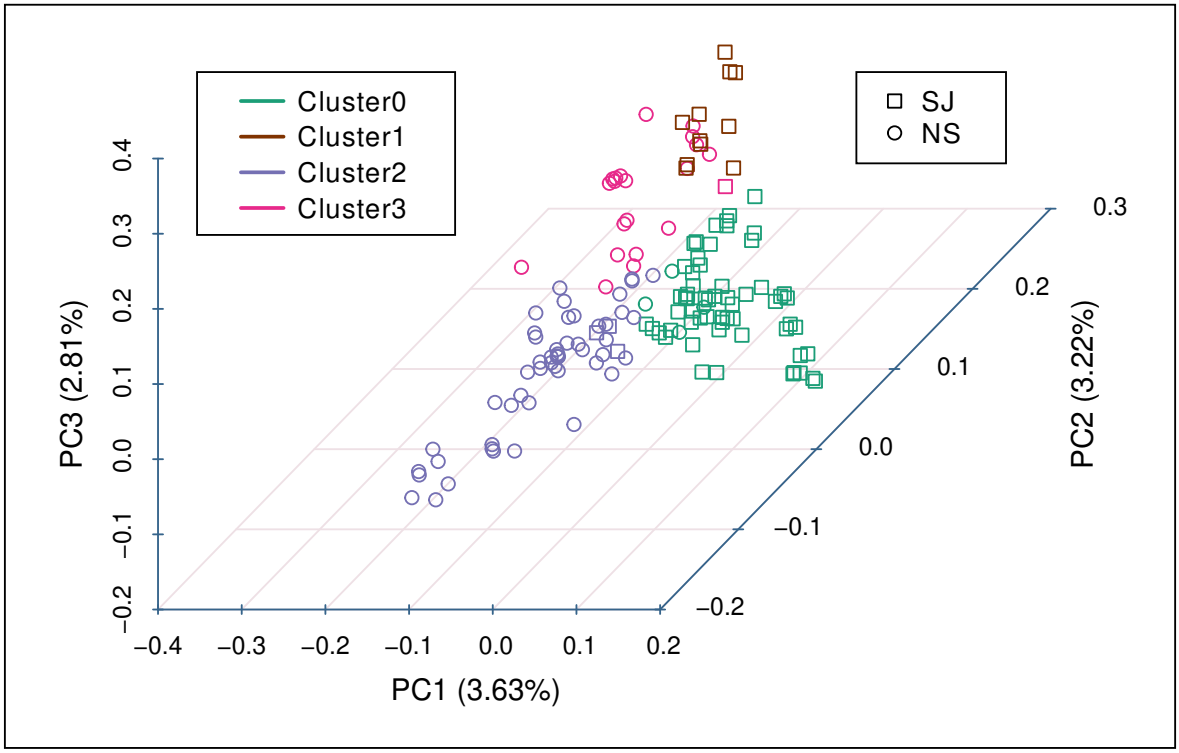



Figure 3: 3D clustered PCA results of Faba bean breeding lines based on bi-allelic SNP markers. **SJ** and **NS** are for the Danish breeding companies Sejet and Nordic Seed

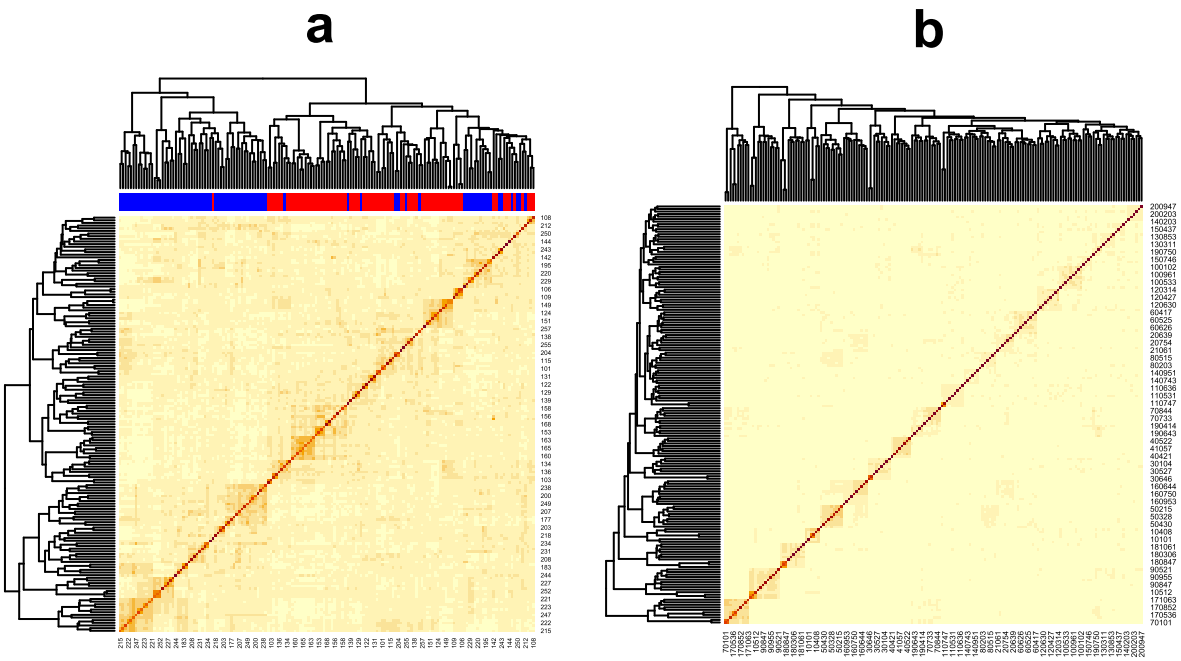
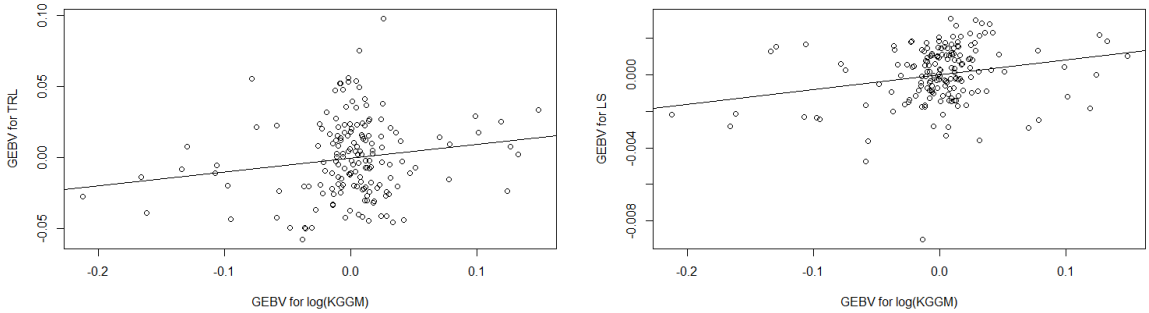



Figure 4: Genomic realationship matrices based on bi-allalic SNP markers
**(a)** Faba bean lines, Blue bars: **SJ**, Red bars: **SJ**
**(b)** White clover lines




| | KGGM | TRL | LS |
|---|---|---|---|
| **KGGM** | 1 | 0.166 | 0.233 |
| **TRL** | 0.166 | 1 | -0.001 |
| **LS** | 0.233 | -0.001 | 1 |

Figure 5: Correlations between Genomic Estimated Breeding Values in Faba bean breeding lines.
**KGGM**: Green matter, kg, **TRL**: Total root length, **LS**: Leaf solidity

# References


1. Rogers, E.D., and Benfey, P.N. (2015). Regulation of plant root system architecture: implications for crop advancement. Current Opinion in Biotechnology *32*, 93–98. doi: `10.1016/j.copbio.2014.11.015`.

2. Lynch, J.P. (2007). Roots of the second green revolution. Australian Journal of Botany *55*, 493–512. doi: `10.1071/bt06118`.

3. Atkinson, J.A., Pound, M.P., Bennett, M.J., and Wells, D.M. (2019). Uncovering the hidden half of plants using new advances in root phenotyping. Current Opinion in Biotechnology *55*, 1–8. doi: 10.1016/j.copbio.2018.06.002.

4. Lynch, J. (1995). Root architecture and plant productivity. Plant Physiology *109*, 7–13. doi: 10.1104/pp.109.1.7.

5. Maqbool, S., Hassan, M.A., Xia, X., York, L.M., Rasheed, A., and He, Z. (2022). Root system architecture in cereals: progress, challenges and perspective. The Plant Journal *110*, 23–42. doi: 10.1111/tpj.15669.

6. Trachsel, S., Kaeppler, S.M., Brown, K.M., and Lynch, J.P. (2010). Shovelomics: high throughput phenotyping of maize (zea mays l.) root architecture in the field. Plant and Soil *341*, 75–87. doi: 10.1007/s11104-010-0623-8.

7. Svane, S.F., Jensen, C.S., and Thorup-Kristensen, K. (2019). Construction of a large-scale semi-field facility to study genotypic differences in deep root growth and resources acquisition. Plant Methods *15*. doi: 10.1186/s13007-019-0409-9.

8. Wang, Y., Thorup-Kristensen, K., Jensen, L.S., and Magid, J. (2016). Vigorous root growth is a better indicator of early nutrient uptake than root hair traits in spring wheat grown under low fertility. Frontiers in Plant Science *7*. doi: 10.3389/fpls.2016.00865.

9. Malinowska, M., Kristensen, P.S., Nielsen, B., Fè, D., Ruud, A.K., Lenk, I., Greve, M., and Asp, T. (2025). The value of early root development traits in breeding programs for biomass yield in perennial ryegrass (lolium perenne l.). Theoretical and Applied Genetics *138*. doi: 10.1007/s00122-024-04797-5.

10. Tuberosa, R., Sanguineti, M.C., Landi, P., Michela Giuliani, M., Salvi, S., and Conti, S. (2002). Identification of qtls for root characteristics in maize grown in hydroponics and analysis of their overlap with qtls for grain yield in the field at two water regimes. Plant Molecular Biology *48*, 697–712. doi: 10.1023/a:1014897607670.

11. Li, P., Chen, F., Cai, H., Liu, J., Pan, Q., Liu, Z., Gu, R., Mi, G., Zhang, F., and Yuan, L. (2015). A genetic relationship between nitrogen use efficiency and seedling root traits in maize as revealed by qtl analysis. Journal of Experimental Botany *66*, 3175–3188. doi: 10.1093/jxb/erv127.

12. Uga, Y., Sugimoto, K., Ogawa, S., Rane, J., Ishitani, M., Hara, N., Kitomi, Y., Inukai, Y., Ono, K., Kanno, N., Inoue, H., Takehisa, H., Motoyama, R., Nagamura, Y., Wu, J., Matsumoto, T., Takai, T., Okuno, K., and Yano, M. (2013). Control of root system architecture by deeper rooting 1 increases rice yield under drought conditions. Nature Genetics *45*, 1097–1102. doi: 10.1038/ng.2725.

13. Atkinson, J.A., Wingen, L.U., Griffiths, M., Pound, M.P., Gaju, O., Foulkes, M.J., Le Gouis, J., Griffiths, S., Bennett, M.J., King, J., and Wells, D.M. (2015). Phenotyping pipeline reveals major

seedling root growth qtl in hexaploid wheat. Journal of Experimental Botany *66*, 2283–2292. doi: 10.1093/jxb/erv006.

14. Jia, Z., Liu, Y., Gruber, B.D., Neumann, K., Kilian, B., Graner, A., and von Wirén, N. (2019). Genetic dissection of root system architectural traits in spring barley. Frontiers in Plant Science *10*. doi: 10.3389/fpls.2019.00400.

15. Heslot, N., Jannink, J., and Sorrells, M.E. (2015). Perspectives for genomic selection applications and research in plants. Crop Science *55*, 1–12. doi: 10.2135/cropsci2014.03.0249.

16. Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics *157*, 1819–1829. doi: 10.1093/genetics/157.4.1819.

17. Thompson, R., and Meyer, K. (1986). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. Livestock Production Science *15*, 299–313. doi: 10.1016/0301-6226(86)90071-0.

18. Habier, D., Fernando, R.L., and Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. Genetics *177*, 2389–2397. doi: 10.1534/genetics.107.081190.

19. Jia, Y., and Jannink, J.L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics *192*, 1513–1522. doi: 10.1534/genetics.112.144246.

20. Smith, A.G., Han, E., Petersen, J., Olsen, N.A.F., Giese, C., Athmann, M., Dresbøll, D.B., and Thorup-Kristensen, K. (2022). RootPainter: deep learning segmentation of biological images with corrective annotation. New Phytologist *236*, 774–791. doi: 10.1111/nph.18387.

21. Seethepalli, A., Guo, H., Liu, X., Griffiths, M., Almtarfi, H., Li, Z., Liu, S., Zare, A., Fritschi, F.B., Blancaflor, E.B., Ma, X.F., and York, L.M. (2020). Rhizovision crown: An integrated hardware and software platform for root crown phenotyping. Plant Phenomics *2020*, 3074916. doi: 10.34133/2020/3074916.

22. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671–675. doi: 10.1038/nmeth.2089.

23. Inkscape. Inkscape project (2020). Tech. Rep. Retrieved from https://inkscape.org (2020).

24. ImageMagick. ImageMagick Studio LLC. Tech. Rep. Retrieved from https://imagemagick.org (2024).

25. Moeskjær, S., Skovbjerg, C.K., Tausen, M., Wind, R., Roulund, N., Janss, L., and Andersen, S.U. (2021). Major effect loci for plant size before onset of nitrogen fixation allow accurate prediction of yield in white clover. Theoretical and Applied Genetics *135*, 125–143. doi: 10.1007/s00122-021-03955-3.

26. Jayakodi, M., Zhang, H., Windhorst, A., Bornhofen, E., Tulpova, Z., Novak, P., Macas, J., Simkova, H., Nadzieja, M., Kim, J.M., Cram, D., Cao, Y., Konkin, D., Sass, O., Welna, G., Himmelbach, A., Mascher, M., Link, W., Kwon, S.J., Andersen, S., and Yang, T.J. (2024). Allelic variation at a single locus distinguishes spring and winter faba beans. doi: 10.21203/rs.3.rs-5356723/v1.

27. Griffiths, A.G., Moraga, R., Tausen, M., Gupta, V., Bilton, T.P., Campbell, M.A., Ashby, R., Nagy, I., Khan, A., Larking, A., Anderson, C., Franzmayr, B., Hancock, K., Scott, A., Ellison, N.W., Cox, M.P., Asp, T., Mailund, T., Schierup, M.H., and Andersen, S.U. (2019). Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. The Plant Cell *31*, 1466–1487. doi: 10.1105/tpc.18.00606.

28. Madsen, J.J., P. (2013). DMU: A user's guide. a package for analysing multivariate mixed models. release 5.6. In Proceedings, 10th World Congress of Genetics Applied to Livestock Production.

29. Falconer, D.S., and Mackay, T. (2009). Introduction to quantitative genetics. 4. ed., [16. print.] ed.. Harlow [u.a.]: Pearson, Prentice Hall. ISBN 9780582243026. Literaturverz. S. [437] - 457.

30. VanRaden, P. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science *91*, 4414–4423. doi: 10.3168/jds.2007-0980.