

## Introduction

Data generated across multiple spatial and temporal scales requires **robust analytical pipelines** capable of managing complex **phenotypic datasets**. The application of machine learning to phenomic data offers a powerful framework for uncovering patterns and extracting biologically meaningful insights. In this study, we used a **high-throughput phenotyping** (HTP) platform, PlantScreen™ Modular System (PSI, Czechia) equipped with multiple imaging sensors to assess morphological and physiological response of barley (*Hordeum vulgare*) lines subjected to drought stress, a key challenge in future climate scenarios. Our objective was to develop an **advanced data analysis pipeline** for the generated phenotypic data. The applied modeling approach enabled the identification of the **predictive traits** with high accuracy even at the early reproductive stage, thereby supporting researchers and breeders in efficiently selecting tolerant genotypes for future climate resilience.

## Methods

**Genotypes:** Six barley lines (267 plants in total).  
**Treatment:** Control and drought at 25–20%.  
**Duration:** From tillering to maturity stage.  
**Phenotyping approach per week:**

- 3x <sup>1</sup>FC morning measurement, Thermal IR, RGB1, RGB2
- 2x <sup>2</sup>FC night measurements
- 1x <sup>3</sup>FC chlorophyll content, Hyperspectral (HS)
- Daily weighing & watering

**Dataset:** 145 image-based traits over 70 time points + harvest traits

**Analysis pipeline** includes multiple complementary approaches:

- K-means clustering for genotypes based on phenotypic traits and treatments
- Drought tolerance ranking of lines using permutational multivariate analysis of variance (PERMANOVA).
- Variance decomposition of traits using mixed effects linear models.
- Temporal phenomic classification of treatment to identify distinct traits under drought.
- Temporal phenomic prediction of harvest traits using LASSO model.

Used models were trained using all traits measured at all time point and only early time points on control, drought and pooled dataset.

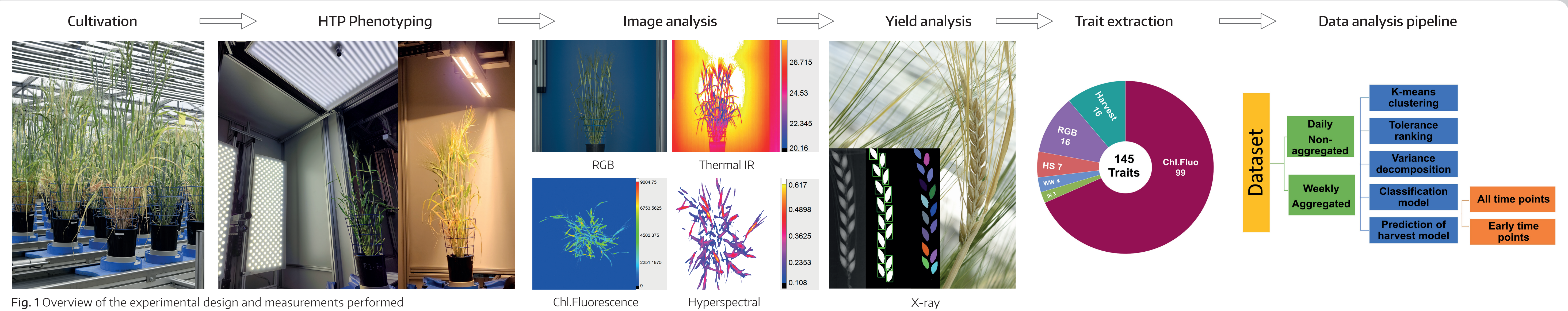


Fig. 1 Overview of the experimental design and measurements performed

## Results

### Genotypic clustering and drought tolerance ranking of barley lines

To visualize the genotypic and treatment-specific grouping patterns, k-mean clustering (with  $k = 2$ ) was applied on PCA projected data. The results showed that plants grown under control conditions were separated from drought-stressed plants, highlighting that treatment effects contribute to variance in the data (Fig. 2).

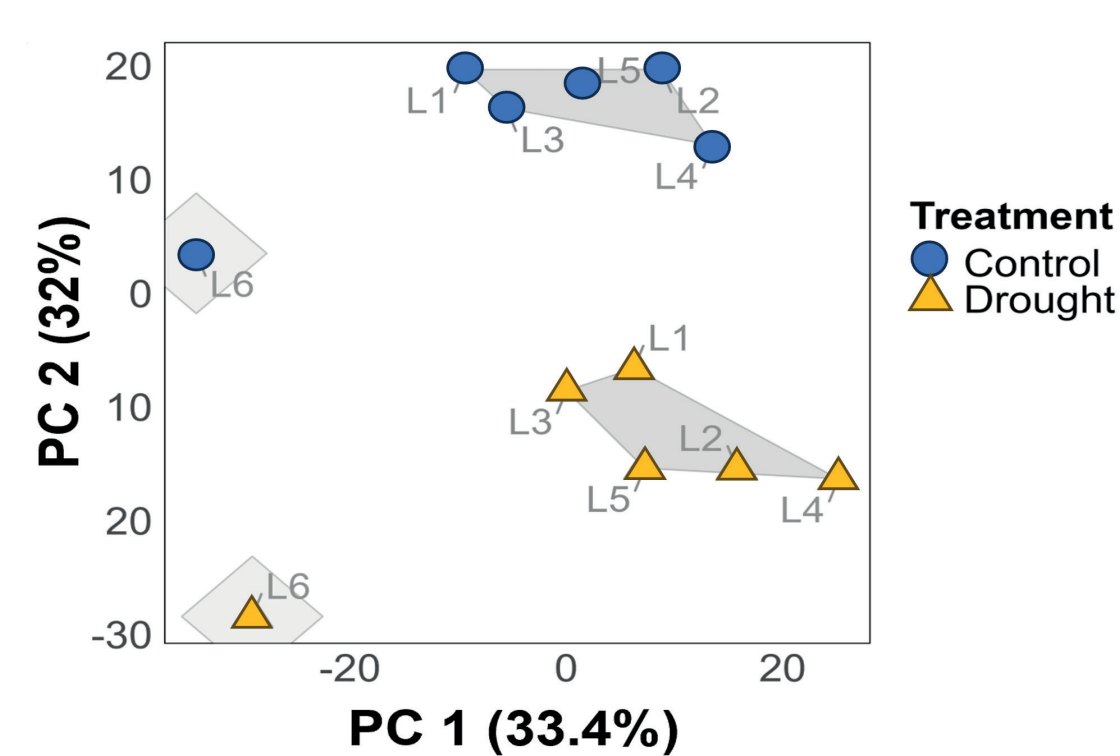


Figure 2 Treatment separation using k-means clustering based on all traits across all time points

Tolerance ranking, i.e. quantifying the effect of drought on temporal or harvest traits, found that drought has a highly significant effect on both temporal and harvest traits (Fig. 3).

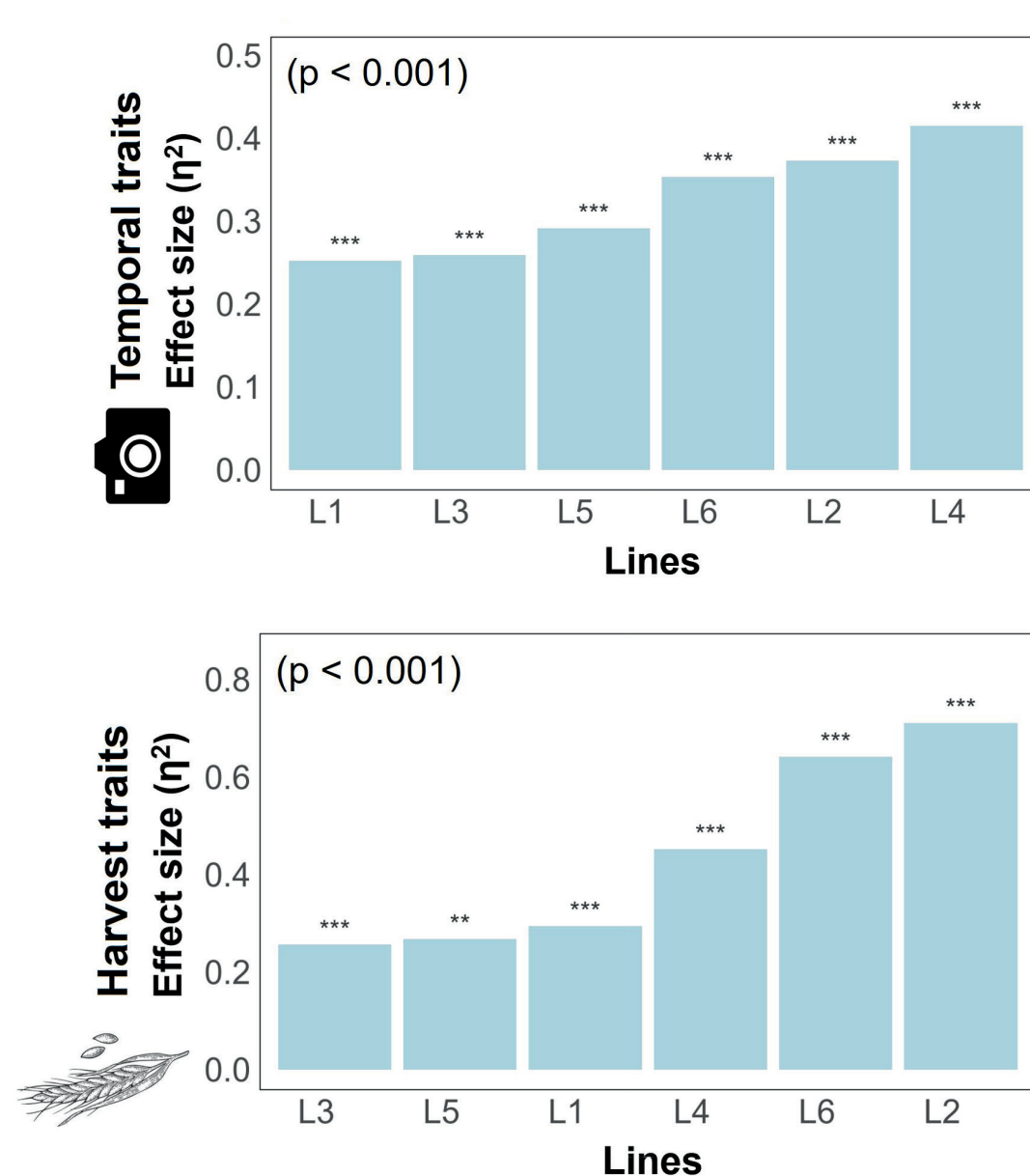


Figure 3 Tolerance ranking of genotype to drought stress using PERMANOVA quantifying significance and effect size of treatment on traits

### Variance decomposition of traits

Variance decomposition was performed to partition trait variance into genetic, time or environmental components and resulted showed that canopy temperature and plant biomass were affected by treatment (Fig. 4).

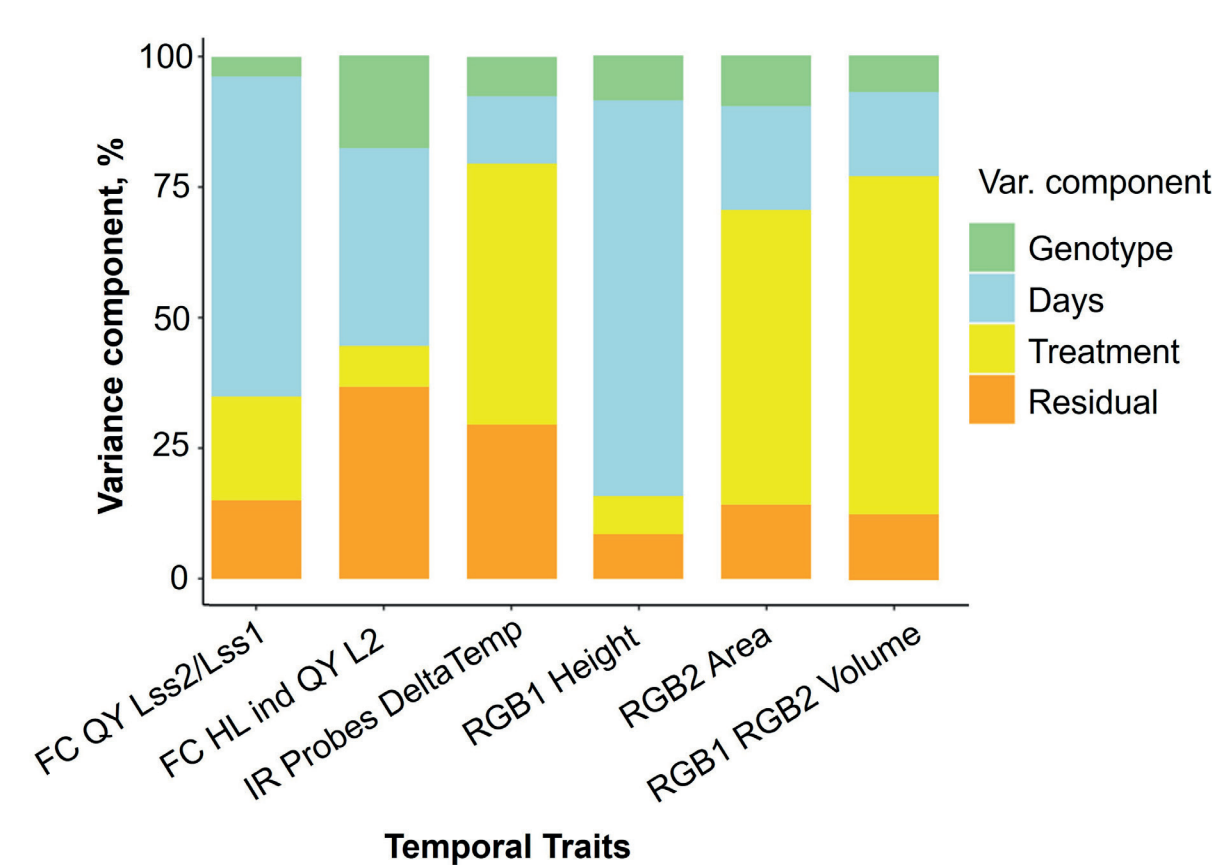


Figure 4 Partitioning of variances of selected temporal traits

### Temporal phenomic classification of treatment

Variable importance analysis revealed that the most influential predictors of treatment classification, regardless of daily or weekly aggregated data (model mean accuracy > 0.98), at early-stage, canopy temperature depression ( $\Delta T$ ) and at late-stage, RGB-based plant size estimates (RGB1 Area) and the quantum yield (FC1 HLL LQ Y) (Fig. 5).

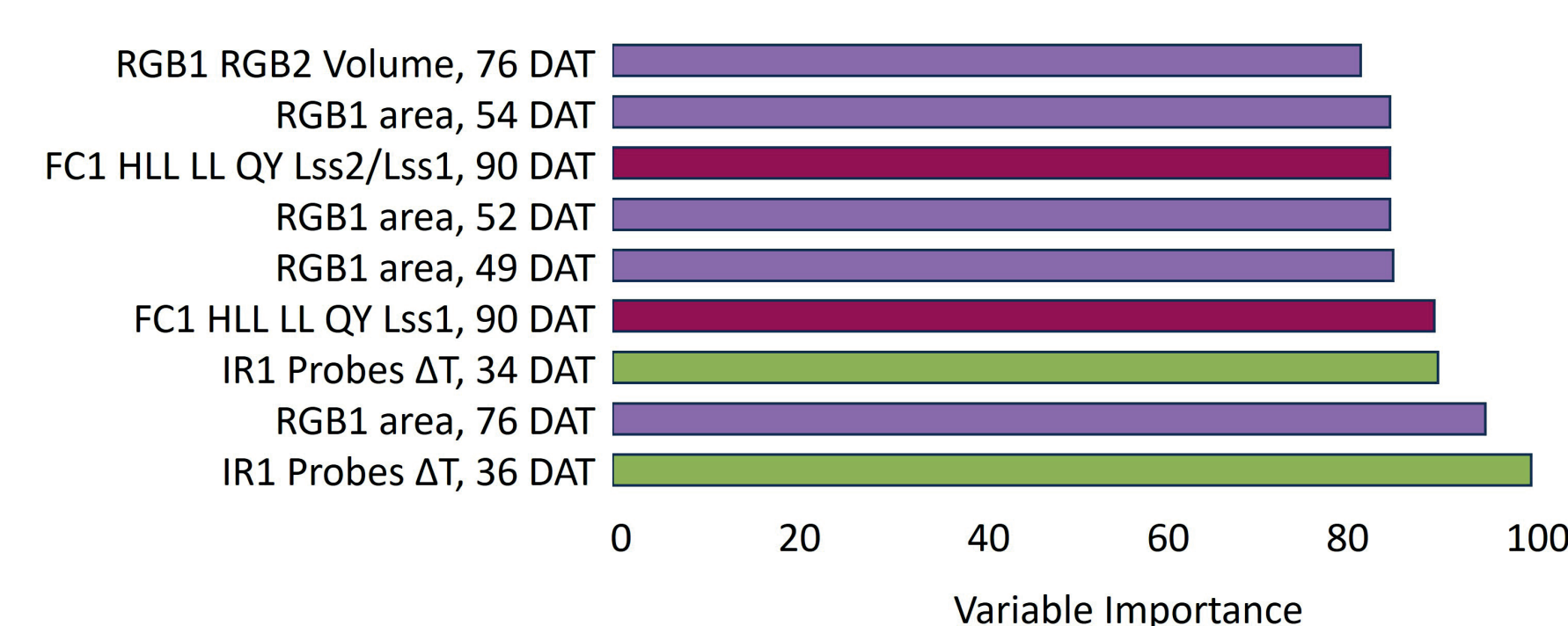


Figure 5 Variable importance for classification of treatment using a random forest model on non-aggregated data

### Temporal phenomic prediction of harvest traits

Temporal phenomic prediction (TPP) was also performed for all harvest traits using RF regression and LASSO models, with training data from the control treatment, drought treatment, or the pooled data set containing both treatments. Among the predicted traits, total biomass dry weight, spike number, total spike weight, and five spikes weight were generally the most predictable based on  $R^2$  values (Fig. 6).

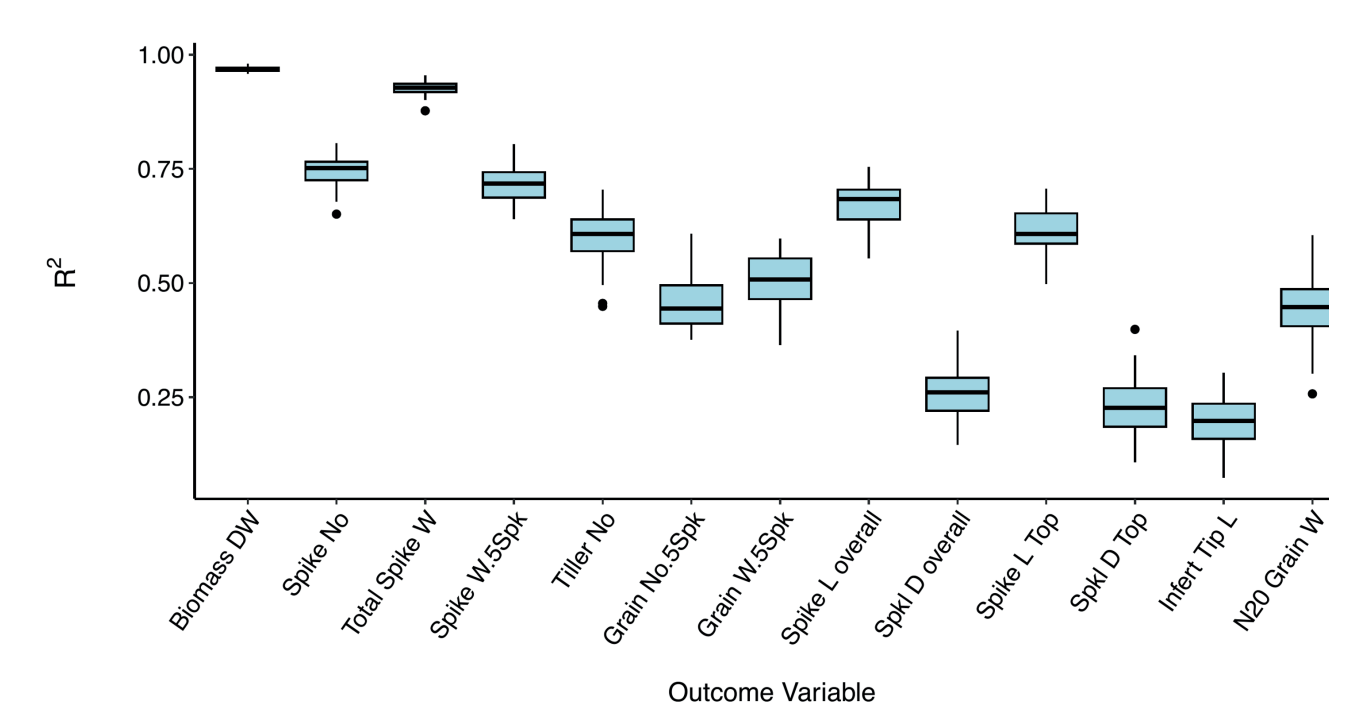


Figure 6 Showing the accuracy among all the harvested traits using a LASSO model trained on pooled

To assess the feasibility of early prediction of harvest traits and identify early predictors of importance, models were retrained using only measurements from the first half of the experiment. Predictions remained reasonably effective, particularly for biomass dry weight (mean  $R^2$  0.924 for pooled-treatment LASSO model), despite a slight reduction in accuracy (Fig. 7A). LASSO coefficients for either early or all time points in pooled treatment biomass prediction were generally associated with plant size estimates (e.g., plant area) derived from RGB imaging at late time points. These predictors also showed high variance contributions linked to the treatment (Fig. 7B, 4).

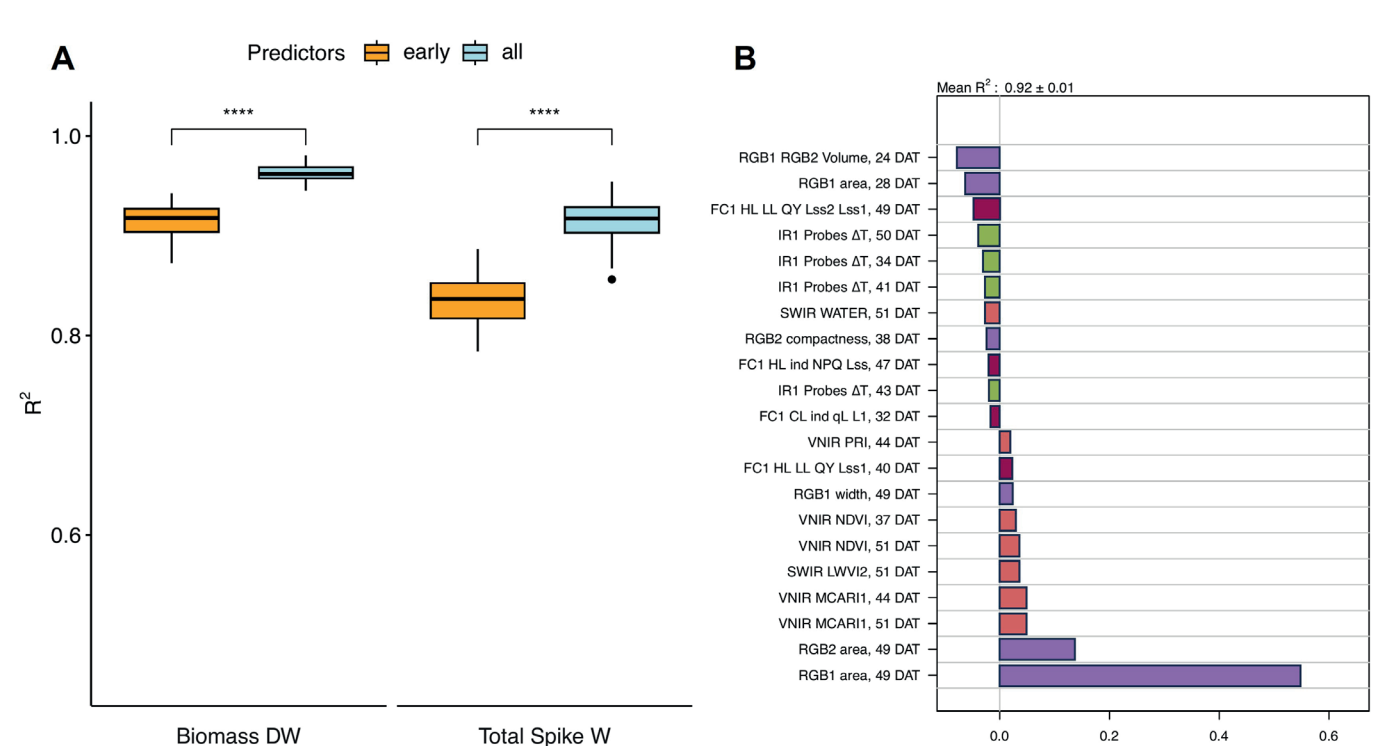
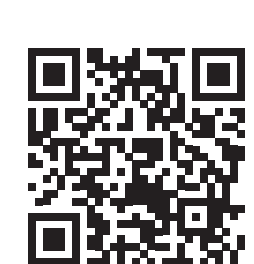


Figure 7 A) Comparison between the different early and all time points using the LASSO model. B) LASSO coefficients of early time point TPP models for biomass dry weight

## Conclusion

Development of a data integration pipeline that applies machine learning to highdimensional phenomic data, providing a powerful framework for uncovering patterns of drought-stressed plants. This approach enabled robust, unbiased analysis of drought stress responses in barley, facilitating both visualization and biological interpretation. Moreover, leverage statistics and machine learning to enable the extraction of meaningful morphological and physiological insights.



### Acknowledgment

CAPITALISE project is funded by the European Union's Horizon 2020 research and innovation program (grant no. 862201)

### References

Tietze, H., Abdelhakim, L., Pleskačová, B., Kurtz-Sohn, A., Fridman, E., Nikoloski, Z., & Panzarová, K. (2025). Prediction of harvest-related traits in barley using high-throughput phenotyping data and machine learning. *BioRxiv*, 2025.05.29.656856. <https://doi.org/10.1101/2025.05.29.656856>