GigaScience

Predicting plant biomass accumulation from image-derived parameters

--Manuscript Draft--

Manuscript Number:	GIGA-D-17-00225R1
Full Title:	Predicting plant biomass accumulation from image-derived parameters
Article Type:	Research
Funding Information:	Robert Bosch Stiftung (32.5.8003.0116.0) — Dr. Christian Klukas Federal Agency for Agriculture and Food (15/12-13, 530-06.01-BiKo CHN) — Dr. Christian Klukas Bundesministerium für Bildung und Forschung (0315958A and 031A053B) — Dr. Christian Klukas European Plant Phenotyping Network (284443) — Dr. Christian Klukas
Abstract:	Background: Image-based high-throughput phenotyping technologies have been rapidly developed in plant science recently and they provide a great potential to gain more valuable information than traditionally destructive methods. Predicting plant biomass is regarded as a key purpose for plant breeders and ecologist. However, it is a great challenge to find a predictive biomass model across experiments. Results: In the present study, we constructed four predictive models to examine the quantitative relationship between image-based features and plant biomass accumulation. Our methodology has been applied to three consecutive barley (Hordeum vulgare) experiments with control and stress treatments. The results proved that plant biomass can be accurately predicted from image-based parameters using a random forest model. The high prediction accuracy based on this model, in particular the cross-experiment performance, will contribute to relieve the phenotyping bottleneck in biomass measurement in breeding applications. The relative contribution of individual features for predicting biomass was further quantified, revealing new insights into the phenotypic determinants of plant biomass outcome. Furthermore, the methods could also be used to determine the most important image-based features related to plant biomass accumulation, which would be promising for subsequent genetic mapping to uncover the genetic basis of biomass. Conclusions: We have developed quantitative models to accurately predict plant biomass accumulation from image data. We anticipate that the analysis results will be useful to advance our views of the phenotypic determinants of plant biomass outcome, and the statistical methods can be broadly used for other plant species.
Corresponding Author:	Dijun Chen GERMANY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Dijun Chen
First Author Secondary Information:	
Order of Authors:	Dijun Chen Rongli Shi Jean-Michel Pape

	Kerstin Neumann
	Daniel Arend
	Andreas Graner
	Ming Chen
	Christian Klukas
Order of Authors Secondary Information:	
Response to Reviewers:	We would like to thank both reviewers for their time to evaluate our manuscript and for their helpful comments/suggestions on our manuscript. It is our belief that this revised manuscript is significantly improved as a result of the changes suggested in the previous round of review. We explained point-by-point the changes made in response to the comments, and highlighted all the changes throughout the revised manuscript in blue. Our replies start with [Response]. Reviewer #1 Image datasets are available and are a valuable community resources. The code is available, which is great. While I definitely appreciate the authors work, I don't think the data support some of the statement throughout the paper, especially when it comes to the wording regarding MLR vs other models, unless further clarification can be provided (Figure 3). In some of the conditions (stress for example) MLR looks better than the other models. The inclusion of color, NIR, and Fluor traits into models is interesting. [Response] We appreciate the reviewer's assessment of our work and his/her comments on our manuscript. We realized that some of the sentences in the manuscript might be overstated by reading the questions raised below. In the revised manuscript, we changed some parts and gave the statement more carefully (Response 1.7). Lines 14-15: I think this statement needs to be qualified by saying that it is a challenge to find a predictive biomass model across experiments, not that it is a challenge to find a biomass model 'in the context of high-throughput phenotyping', which is vague and I don't think accurate without further clarification considering the number of previous papers that model biomass from images with high correlation to ground truth measurements. [Response 1.1] We thank the reviewer for this valuable suggestion. We rewrote the statement according to the suggestion in the 'Abstract' section. Lines 34 to 40: lacking in citations of literature. Introduction in general needs improvement in terms of the previous literature that it cites. [Response 1.2] We thank the reviewer for pointing out this. We have added some new references and re-organized some text in the section 'Introduction' (line 36-44). The second paragraph of the intro is a very limited short review of the literature but there are a number of papers that model biomass using ht-phenotyping that are not represented including Yang et al 2014 (nature communications), Montest et al. 2011 (Field Crops Research), Fahlgren et al. 2015 (Molecular Plant) to name a few. [Response 1.3] We appreciate the reviewer's nice suggestion. We have added some related references in the revised manuscript (line 44-50). Line 45: "On the other hand, to produce reliable assessments, suitable model types needs to be established and model construction requires integration of many components such as efficient mathematical analysis and representative data." Very vague. Line 58: Please clarify this statement: "Another concern is that the number of traits used in these studies were quite limited and perhaps not representative enough. Therefore, a more effective and powerful model is needed to overcome these limitations and to allow better utilization of the image-based plant features which are obtained from non-invasive phenotyping approaches." Not sure what this means exactly, very vague considering that the papers mentioned do have models of biomass that are not 'perfect' but do have high heritability and correlation with ground truth measurements. [Response 1.4/1.5] We have rephrased related sentences in the second paragraph of

	the ‘Introduction’ section (line 62-65). I think the authors need to adjust the justification of their research to stress that there needs to be biomass models that can be used across experiments/environment/treatments, which they do say, but needs to be stated more clearly. In general, many of the justification statements, which are pointed out in points 3 and 4 above are obscure to the point that they lose meaning. [Response 1.6] We rephrased these statements in this revised version (line 69-73). Line 146: "Although the performance of these models was roughly similar, RF, SVR and MARS methods had better performance than the MLR method for prediction of both FW (Fig. 3B) and DW (Fig. 3D), implying a nonlinear relationship between image-based phenotypic profiles and biomass output." This doesn't seem accurate, it looks like MLR has just as good predictive power in many of the situations presented. I don't think you can say that MLR and the others are roughly similar and then say that this implies a nonlinear relationship. Can this conclusion be clarified? It seems like there are only small differences between the models. [Response 1.7] We thank the reviewer for cautioning us to avoid overstatements in our manuscript. We have revised the manuscript in the ‘Results’ section so as not to overstate our observations (line 153-165). Regardless of whether or not random forest is the 'best' model, the data doesn't seem to support the statement that the RF model 'largely' outperformed the other models. This only seems accurate under the control condition, can this be clarified? [Response 1.8] We agree with the reviewer that the statement wasn’t proper here. We have corrected this point in the revised manuscript. Line 238: "Although previous attempts have been made to estimate plant biomass from image data, most of these studies consider only a single image-based feature or very few features in their models which are often linear-based, ignoring the fact that the phenotypic components underlying biomass accumulation are presumably complex. Accurately predicting biomass from image data requires efficient mathematical models as well as representative image-derived features." I disagree with the authors on this point, if biomass can be modeled with a few features with high correlation why does it matter if they presume that it is complex? Their more complex models were still decreased in R2 with environmental differences and between experiments and I don't find the data suggesting that RF model outperforming other models (particularly MLR) convincing without further clarification. [Response 1.9] We agree with the reviewer’s comment that it would be good that if biomass can be modelled with a few features with high correlation. What we meant or afraid is that using too less feature might lead to under-estimated or over-estimated results since other features were not considered and evaluated. But nevertheless, we removed this sentence in this revised manuscript. Reviewer #2: The authors investigate the ability of deriving plant biomass (both fresh and dry mass) from 2D image-based features acquired with visible, fluorescent and NIR multi-view imaging systems operating on an automated high throughput phenotyping platform. In a first part, several multivariate statistical models are compared for their ability at predicting biomass for two treatments within a single experiment, on three independent datasets, detailed results being presented for one experiment. One of the best model, the random forest, is then further investigated for its capacity at making prediction across experiments, being trained on one experiment at a time or on one treatment of one experiment at a time. Finally, the relative importance of individual image-based traits in the prediction of either fresh or dry weight is presented for two treatments of one dataset. Models and methods for model evaluation are clearly presented, and the overall quality of the text and Figure makes the paper easy to follow. The inclusion of other than visible images, the objective selection of image-based traits, the comparison of models and the use of 3 independent datasets clearly distinguish this paper from previous publications on the same subject. It provides the reader very valuable information on the current prediction capacity of the approach, together with a consistent methodology for analyzing other related practices. [Response] We thank the reviewer for his/her assessment of our work and appreciate

	that the reviewer recognizes the advantages of our approaches and analysis. However, I have two major concerns on the current version of this manuscript. First, I think that some conclusions highlighted in the abstract or in the text are not completely in line (or at least sufficiently tempered) with what is demonstrated in the text or shown on the figures. In the abstract (line 19-20), it is highlighted that 'The results proved that plant biomass can be accurately predicted from image-based parameters using a random forest model'. To me this conclusion is clearly supported by data in the case of within experiment predictions, but not fully in the case of the cross experiment test (i.e. quite opposite to what is stressed line 21). My impression, given results presented Figure 5, is that in one case out of two, a model trained on one experiment alone could not accurately (or at least with not the same accuracy) predict the biomass, despite a repeated protocol. This result is per se very interesting, as it demonstrates an important limitation of the approach. It can however not be summarized by what is written line 19-21, 201-202, 209-210 or 253-257. On another occasion (line 148 and line 248), I found the conclusion ('the RF model largely outperformed other models') a bit exaggerated, as, on Figure 3, depending on the criteria, RF model performs very similar to MARS model for example. [Response 2.1] We thank the reviewer for raising these points. 1) we agree with the reviewer that the prediction accuracy of the model across experiments is still lower in some case (mostly due to growth conditions changing over seasons). We changed relevant description in the text and added some text to discuss this point. 2) We agree with the reviewer on this point. We clarified this point in the revised version (see line 153-165). Second, I did not manage to test the models, nor to reproduce the analysis with the provided data and source code. Concerning the data, image traits are provided for all experiments, but manual measurement on Dry Weight are missing. Concerning the code, the R-script provided does not fit to the provided dataset, thus making it difficult to test. More important, model code runs with errors at runtime ('not defined' errors). I also think, but this is only a suggestion, that, in addition to raw image files, providing binary masks of plants, that are of high importance for all traits analyzed here, could improve the re-use of this nice dataset. [Response 2.2] We thank the viewer for raising this point. We now provided a separated R script ‘run.R’to test the models. The test data used in our analysis were now deposited in the Github repository (https://github.com/htpmod/HTPmod). We also thank for the reviewer for this nice suggestion. Unfortunately, the files for the binary masks of plants were not kept anymore due to storage limitation. But the report files (generated by the IAP software) that include all trait values are still kept. We uploaded these report files alongside the raw image datasets for re-use of these data (see the section ‘Availability of supporting data and materials’). Other minor points or comments for specific parts of the texts are provided bellow: Line 72-74: I think this sentence would be better be placed in the Potential application section [Response 2.3] We have moved the sentence to the section ‘Potential Implications’. Line 85: Do you mean that some image traits are more sensitive to physiological traits? I do not see why Fig 1B is illustrative for this point. [Response 2.4] The reviewer is correct that some image traits are more sensitive to physiological traits. We agree with the reviewer that the citation of Fig. 1B is not proper here. We corrected this accordingly. Line 98: In the context of phenotyping, it might also be useful to add Spearman rank correlation to the assessment [Response 2.5] We agree with the reviewer’s opinion. In principle, the assessment based on these two kinds of correlation coefficients is similar for good models. We prefer to choose Pearson correlation coefficient based on the assumption that the relationship of observed values and predicted values is linear. Line 108: Fig 1B is only a heatmap image. May be a list of traits should be provided, or a reference to the supplementary data should be added here. [Response 2.6] We want to thank the reviewer for bringing up this point. We added a reference to the Supplementary Data S1.

	Line 117: Figure 2B is poorly informative as traits are not identified. This figure is also not commented in the text, I suggest removing it. [Response 2.7] We agree to the reviewer's suggestion and removed Figure 2B in this revised version. Line 144: I would find useful to make here perfectly clear that all the models were trained on the control + stress plants, to avoid any confusion with the 'cross treatment test' later on (Figure 6) [Response 2.8] We thank the reviewer for pointing this out. We have corrected this in the revised manuscript. Line 146-151: I found the analysis a bit confusing as, in the details, the ranking of the different methods varies, and I do not clearly see why RF 'largely outperforms' other methods (especially MARS). [Response 2.9] We agree with the reviewer that this statement is overstated. We have corrected this issue in the revised manuscript (see line 153-165). Line 152-155: The comparison with the widely used 'single feature' method is very interesting. Can you consider to add its score/line on the R2 and RMSRE? [Response 2.10] This is indeed a good suggestion. We now added this value in Figure 3B and D. Line 178: May be it is also worth noting in the text that geometric + color traits trust 13 out of 15 (FW) and 15 out of 15 (DW) first places, as these two types of data are widely available among phenotyping platform and yet not so often used in biomass predictions. [Response 2.11] We appreciate the reviewer for this valuable comment. We have included this point in the revised manuscript. Line 201 - 211: The text seems to me a bit too optimistic regarding the cross experiment predictions. Exp3 clearly shows a non-conservation of the relationship obtained in Exp1 or 2, and a clear loss of predictive power compared to within experiment training. [Response 2.12] We admit that we are a bit overstate here. We have changed it in the revised version. Line 281: typo: sophisticated [Response 2.13] We appreciate the reviewer for pointing out this mistake. We have corrected this word in the revised manuscript. Line 349: could you give an idea of the amount of such filled missing values? [Response 2.14] After the feature selection step (e.g., outlier detection, reproducibility analysis and redundancy removal), the missing values for the remaining features are quite rare (much less than 1%). Line 400: the formulation is a bit strange as it sounds like a conclusion already. [Response 2.15] We have improved this sentence in the revised version. Line 426: DW data are missing. [Response 2.16] We have added DW data in the Supplemental Data S1. Line 535: legend of figure 5 did not really apply to these figures. A complete legend should be added. [Response 2.17] A new legend was added for Figure 5.
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics	Yes

Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?	Yes
Availability of data and materials All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript. Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?	Yes

Predicting plant biomass accumulation from image-derived parameters

Dijun Chen[1,§,*], Rongli Shi[1], Jean-Michel Pape[1], Kerstin Neumann[1], Daniel Arend[1], Andreas Graner[1], Ming Chen[2], and Christian Klukas[1]

[1]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany.*

[2]*Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China.*

[§]Current address: *Department for Plant Cell and Molecular Biology, Institute for Biology, Humboldt-Universität zu Berlin, 10115 Berlin, Germany.*

[*]Correspondence should be addressed to D.C. (chendijun2012@gmail.com)

Abstract

Background:

Image-based high-throughput phenotyping technologies have been rapidly developed in plant science recently and they provide a great potential to gain more valuable information than traditionally destructive methods. Predicting plant biomass is regarded as a key purpose for plant breeders and ecologist. However, it is a great challenge to find a predictive biomass model across experiments.

Results:

In the present study, we constructed four predictive models to examine the quantitative relationship between image-based features and plant biomass accumulation. Our methodology has been applied to three consecutive barley (*Hordeum vulgare*) experiments with control and stress treatments. The results proved that plant biomass can be accurately predicted from image-based parameters using a random forest model. The high prediction accuracy based on this model, in particular the cross-experiment performance, will contribute to relieve the phenotyping bottleneck in biomass measurement in breeding applications. The relative contribution of individual features for predicting biomass was further quantified, revealing new insights into the phenotypic determinants of plant biomass outcome. Furthermore, the methods could also be used to determine the most important image-based features related to plant biomass accumulation, which would be promising for subsequent genetic mapping to uncover the genetic basis of biomass.

Conclusions:

We have developed quantitative models to accurately predict plant biomass accumulation from image data. We anticipate that the analysis results will be useful to advance our views of the phenotypic determinants of plant biomass outcome, and the statistical methods can be broadly used for other plant species.

Keywords: Barley; High-throughput phenotyping; Phenomics; Biomass; Modeling.

Introduction

Biomass accumulation is an important indicator of crop final product and plant performance. It is thus considered as a key trait in plant breeding, agriculture improvement and ecological applications. The conventional approach of measuring plant biomass is very time consuming and labour intensive since plants need to be harvested destructively to obtain the fresh or dry weight [1]. Moreover, the destructive method makes multiple measurements of the same plant over time impossible. With the development of new technology, digital image analysis has been used more broadly in many fields, as well as in plant research [2-4]. It allows faster and more accurate plant phenotyping and has been proposed as an alternative way to infer plant biomass [2, 3, 5].

In recent years, plant biomass has been subject to intensive investigation by using high-throughput phenotyping (HTP) approaches in both controlled growth chambers [2-3, 6-11] and field environments [5, 12-17], demonstrating that the ability of imaging-based methods to infer plant biomass accumulation. For example, significant genotypic and environmental effects on plant biomass and some other traits in Setaria were revealed by the Bellwether Phenotyping Platform under controlled-environmental condition [10]. Yang et al [11] showed that predicted rice biomass (including shoot fresh and dry weight) based on image-derived morphological and texture features provided a relatively more complete representation than manual measurements in dissecting its genetic architecture. In this regard, optimized models plus image-derived features from HTP systems will improve the power of dissecting genetic architecture of complex traits. Although there are some developed models for predicting plant biomass, most of them have certain limitations. For example, Golzarian *et al.* (2011) modelled the plant biomass (dry weight) in wheat (*Triticum aestivum* L.) as a linear function of projected area, assuming plant density was constant. However, this

method under-estimated dry weight of salt stressed plants and over-estimated that of control plants. Even though the authors argued that the bias was largely related to plant age and the model might be improved by including the factor of plant age [3], the differences in plant density between stressed and control plants may be caused by different physiological properties of plants rather than plant age. In another study, Busemeyer *et al.* (2013) developed a calibrated biomass determination model for triticale (x *Triticosecale* Wittmack L.) under field conditions based on multiple linear regression analysis of a diverse set of parameters, considering both, the volume of the plants and their density. Indeed, this model largely improved the prediction accuracy of the calibration models based on a single type of parameters and can precisely predict biomass accumulation across environments [15]. However, they used very limited traits for the model and make it a question whether it could be applied broadly in other cases. As mentioned by Yang et al. (2014), noticeable improvement was achieved by adding morphological features or texture features to the biomass-predicting model [11]. This suggests that adding more information/traits could improve the predictive performance of models. Therefore, a more effective and powerful model is needed to overcome these limitations and to allow better utilization of the image-based plant features which are obtained from non-invasive phenotyping approaches.

Individual studies have recently shown that the prediction accuracy of plant biomass based on image-derived features is relatively high even using the simplest linear regression models [3,10,18]. However, the performance of nonlinear predictive models has not been well evaluated. Besides, it is still challenging to apply these models across experiments that are performed in different environmental conditions or with different treatments due to lack of representative datasets so far to make reliable assessments. In this study, we present a general framework for investigating the relationships between plant biomass (referred to as shoot biomass hereafter) and image-derived parameters. We applied a multitude of supervised and unsupervised statistical methods to investigate different aspects of biomass determinants by a list of representative phenotypic traits in three consecutive experiments in barley. The results showed that image-based features can accurately predict plant biomass output and collectively reflect large proportions of the variation in biomass accumulation. We elucidated the relative importance of different feature categories and of individual features in prediction of biomass accumulation. The differences in the contribution of the image-based features for prediction of two types of biomass measurements, fresh weight and dry weight were

compared as well. Furthermore, our models were tested for the possibility of predicting plant biomass in different experiments with different treatments.

Results

Development of statistical models for modelling plant biomass accumulation using image-based features

In the previous studies [19,20], we have shown that a single phenotypic trait -- the three-dimensional digital volume, which is a derived feature from projected side and top areas -- can be reasonably predictive to estimate plant biomass accumulation. We expect that the predictive power could be improved when multiple phenotypic traits are combined in a prediction model since plant biomass is determined not only by their structural features but also by their density (physiological properties). To further investigate the relationship between image-derived parameters and plant biomass accumulation, deep phenotyping data which contain both structural (e.g., geometric traits) and physiological traits (e.g., plant moisture content as reflected by near-infrared [NIR]-related traits) were analysed (**Fig. 1**, **A** and **B**). Pot weights of the plants were not included for the analysis although they were weighed regularly. It might reflect the growth tendency of the whole plants (shoots and roots) where herein we focused mainly on shoots.

Models were constructed to quantify the ability of imaging-based features to statistically predict the biomass accumulation. The models were developed by using four widely used machine-learning methods (**Fig. 1C**): multivariate linear regression (MLR), multivariate adaptive regression splines (MARS), random forest (RF) and support vector regression (SVR), which have extensively been used in accurate prediction of gene expression [21-25] and DNA methylation levels [26-29]. We combined the biomass measurements (fresh weight [FW] and/or dry weight [DW]) with image-based features and then divided them into a training data set and a test data set. A model was trained on the training data set and has then been applied to the test data set to predict the plant biomass. The relationship between plant biomass accumulation and image-based features was assessed based on the criterion of the Pearson correlation coefficient (r) between the predicted values and the actual values, or the coefficient of determination (R^2; the percentage of variance of biomass

explained by the model; **Fig. 1D**).

Our methodology was applied to three consecutive experiments (**Fig. 2A**; **Supplemental Table S1** and **Data S1**), which were designed to investigate vegetative biomass accumulation in response to two different watering regimes under semi-controlled greenhouse conditions in a core set of barley cultivars by non-invasive phenotyping [20, 30]. There were 312 plants with 18 genotypes for each experiment. Plants were monitored using three types of sensors (visible, fluorescence [FLUO] and near-infrared [NIR]) in a LemnaTec-Scanalyzer 3D imaging system. An extensive list of phenotypic traits ranging from geometric (shape descriptors) to physiological properties (i.e., colour-, FLUO- and NIR-related traits) could be extracted from the image data **(Supplemental Data S1)** using our image processing pipeline IAP [19]. A representative list of traits for each plant in the last growth day were selected to test their ability to predict plant biomass.

Coordinated patterns of plant-image-based profiles and their relation to plant biomass

We extracted a list of representative and non-redundant phenotypic traits for each plant from image datasets for each experiment (see **Materials and Methods**; **Fig. 1B**). In common for these experiments, overall thirty-six high-quality traits which describe plant growth status in the last growth day were obtained. As a result, each dataset was assigned a matrix whose elements were the signals of different features in different plants (**Fig. 1C**). Principal component analysis (PCA; **Fig. 2B**) was applied to these datasets. We found that plants from different experiments with different treatments showed clearly distinct patterns of phenotypic profiles. For instance, stressed plants and control plants were separated using PCA by their first principal component (PC1) and also by the top clusters obtained in HCA, while plants from different experiments were distinguished by PC2 and PC3 in PCA or subordinate clusters in HCA. Accordingly, it could be observed that biomass (e.g., FW) of plants from different experiments with different treatments was significantly different (two-way ANOVA, p-value < 2e-16; **Fig. 2C**). The relationship was reflected by a dendrogram from cluster analysis based on the means of FW over genotypes (**Fig. 2D**). Furthermore, the overall phenotypic patterns of these plants were similar to their biomass output (**Fig. 2**, **B-D**), revealing that these image-based features were potential factors reflecting the accumulation of plant biomass. We thus explored the relationship between the signals of these image-based features and the level of plant biomass output. We calculated the

correlation coefficients for each dataset. The correlation patterns were consistent for different datasets and more than half of the features revealed high correlation coefficients ($r > 0.5$; **Fig. 2E**). Interestingly, both structural features (such as digital volume, projected area and the length of the projected plant area border) and density-related features (such as NIR and FLUO intensities) were involved in the top ranked features.

Relating image-based signals to plant biomass output

The above analyses suggest that plant biomass can at least be partially inferred from image-based features. To examine which model has the best performance and to select an appropriate model for biomass prediction, we then applied our regression models (**Fig. 1C**) to predict plant biomass using image-based features. Our analyses were focused on the first experiment (i.e., Exp 1), since the phenotypic traits of the corresponding dataset have been intensively investigated in our previous study [20]. In this experiment, plant biomass was quantified in two forms: FW and DW. We selected a collection of 45 image-derived parameters from this dataset that were non-redundant and highly representative.

We next tried to predict FW and DW based on this set of image-derived features using four different regression models (MLR, RF, SVR and MARS; **Fig. 3**). The models were respectively tested on control plants, stressed plants and the whole set of plants (**Fig. 3**, **A** and **C**). The prediction accuracy of our models (the correlation coefficients between the predicted biomass and the actual biomass) was firstly compared with the ability of individual features to predict biomass. It was found that our models generally showed better prediction power than the single digital volume-based prediction (**Fig. 3**, **B** and **D**), indicating that additional features improved the predictive power. Then the performance of these models was compared and evaluated. Overall, the performance of all the tested models showed roughly similar for the prediction of both FW (**Fig. 3B**) and DW (**Fig. 3D**) under stressed conditions. The prediction accuracy of our models is still comparable to the results from previous studies [3, 6, 18] based on MLR models, even though much more features were considered in our study. The RF model slightly outperformed other models in predicting biomass of control plants, accounting for the most variance ($R^2 = 0.85$ for FW and $R^2 = 0.62$ for DW; **Fig. 3**, **B** and **D**, left panels) and showed the best prediction accuracy (Pearson's correlation $r = 0.93$ for FW and $r = 0.80$ for DW; **Fig. 3**, **B** and **D**, middle panels). Of note, RF is the only model showing better performance

than single digital volume-based prediction (**Fig. 3D**). In this study, we focused on the results from the RF method in the rest of analysis, although results from different methods were highly consistent and led to the same conclusions.

Relative importance of different image-based features for predicting plant biomass

As mentioned above, the image-based features could be classified broadly into four categories: plant structure properties, colour-related features, NIR signals, and FLUO-based traits (**Fig. 1B**). The last three types of features reflect plant physiological properties and can be considered as plant density-related traits and are thus related to their fresh or dry matter content. For each individual feature or each type of features, we constructed a degenerate model for biomass prediction using the corresponding feature(s) as the predictor(s). We compared the capability of each individual or type of feature for predicting biomass accumulation in the first experiment (i.e., experiment 1). Geometric features showed the most predictive power among the four categories for prediction of both FW and DW, but were slightly less predictive than all features in a full model (**Fig. 4**, **A** and **B**). Strikingly, the predictability of other types of features (such as colour-related and FLUO-based traits) was substantial, indicating that these traits may act as unforeseen factors in biomass prediction. In addition, the NIR-based features showed higher predictive capability for FW than for DW in control and stressed plants, revealing NIR signals were import factors in determining FW accumulation.

Next, we investigated the relative importance (RI) of each feature for predicting biomass using a full model in the whole set of plants (i.e., "control + stressed plants"; **Fig. 4**, **C** and **D**, upper panels). In a RF model, the RI of a feature is calculated as the increase of prediction error (%IncMSE) when phenotypic data for this feature is permuted [31], and thus indicates the contribution of the feature after considering its intercorrelation in a model. We found that the top ten most important features in the full model for predicting FW and DW included both structure and density-related traits. As expected, projected area (from side or top view) and digital volume were the top ranked features, which have individually been considered as proxies of shoot biomass in previous studies [3, 20, 18, 32-37]. However, several geometric and colour-related features that are top ranked in the prediction have not been used in biomass predictions in previous analysis although they are widely available among phenotyping platforms.

In principle, we would expect that highly important features in the full model would be related to a high predictive power in a degenerate model. Surprisingly, there was no clear correlation observed between the feature importance and their predictive power (**Fig. 4**, **C** and **D**). For example, several colour-related and NIR-based features which were in the top ten list of the most important features revealed insubstantial predictive power in individual models. This observation implies that the relation of the underlying biomass determinants is extremely complex and not a linear combinations of the investigated features.

Furthermore, we compared the relative importance of each feature in predicting FW and DW (**Fig. 4E**). Although a positive correlation ($r = 0.88$) between the feature importance for FW and DW could be observed, several features showed large differences in their ability to interpret FW or DW, including "nir.intensity" (derived from side view images), "compactness.01" (top), "hull.pc1" (top), "leaf.count" (side), "hsv.h.average" (top) and "lab.a.mean" (top). For instance, NIR intensity and plant compactness (top view) may be important for predicting FW but not for DW. We also performed the above analyses by using only control (**Supplemental Fig. S1**) or stressed plants (**Supplemental Fig. S2**), respectively. We found that the patterns of feature importance were distinct between these two groups of plants. For example, NIR intensity was ranked as the top fifth feature for predicting FW for stressed plants but was not substantially important for control plants. These findings suggest that there are differences in underlying plant biomass determinants in these kinds of treatment situations that are also reflected by their image-based phenotypic traits.

Image-based features are predictive of plant biomass across experiments with similar conditions or treatments

In order to explore whether our models were generalizable across different experiments, we applied our models trained in one experiment to predict biomass (herein FW) in other experiments using a common set of features. Examples of such cross-experiment predictions are shown in **Figure 5A**. We tested and illustrated all possibilities for cross prediction using the whole set of plants in the corresponding experiment. In general, the prediction accuracy within individual experiments remained high ($r > 0.97$ and $R^2 > 0.93$ for all three experiments; **Fig. 5B**), revealing that our models were effectively predicting plant biomass based on image-

derived feature signals among different experiments. Moreover, the prediction accuracy for cross-experiment prediction, especially between the first two experiments ($r > 0.97$ and $R^2 > 0.94$), was still relatively high, implying that our models generally captured the relationships among the various image-based features. However, the third experiment had relative weaker correlations with the other two experiments for predicting biomass (with $r > 0.81$ and $R^2 > 0.65$; **Fig. 5A**). This might be mainly due to seasonal (temperature and illumination) differences which caused different plants behaviours, namely lower biomass for both control and stressed plants in experiment 3 [30]. This suggests that different plant growth conditions might cause some variation for cross-experiment prediction.

At the same time, we tested cross predictability of our models using treatment-specific data in the experiments (**Fig. 6**). Similar results were obtained as above using the whole dataset (**Fig. 5B**). The weak predictive power for cross-prediction involving control plants from the third experiment was most clearly observable in the low accuracy in the biomass prediction of this particular subset of plants. Generally, control and stressed plants were found to have very weak predictive power when related to each other (**Fig. 6**), as also supported by the distinct patterns of relative feature importance between these two plant groups (**Supplemental Figs. S1** and **S2**). For each experiment, the prediction accuracy was higher for stressed plants compared to control plants. This might resulted from the imaging analysis process. Relatively small plants, stressed plants in this case, would gain more clear images due to less overlapping or less area of out range. Therefore, image quality would be an important variation source for our modelling and should be taking into consideration for any application.

Discussion

Biomass is a complex but important trait in functional ecology and agronomy for studying plant growth, crop productive potential and plant regeneration capabilities. Many different techniques, either destructive or non-destructive, have been used to estimate biomass [1, 2-3, 5-17]. Compared with the traditional destructive methods for measuring biomass, non-destructive imaging methods provide a faster, more accurate approach for plant phenotyping. In recent years, more and more high-throughput plant phenotyping platforms have

been set up and applied worldwide. Accordingly, it becomes a current challenge to establish models utilizing the big datasets gained from high-throughput imaging systems. Accurately predicting biomass from image data requires efficient mathematical models as well as representative image-derived features.

In this study, we have presented a systematic analysis of relationships between plant biomass accumulation and image-derived signals, to confirm the assumption that biomass can be accurately predicted from image-based parameters. We built a random forest model of biomass accumulation using a comprehensive list of representative image-based features. The comparison between a random forest model and alternative regression models indicated that the RF model outperforms other models in terms of (1) better predictive power – especially in comparison with the linear model, confirming the complex phenotypic architecture of biomass, (2) better outperformance than a single-feature-prediction model – arguing the complex phenotypic makeup of biomass, and (3) feasible biological interpretability – the ability to readily extract information about the importance of each feature in prediction. The high prediction accuracy based on this model, in particular the cross-experiment performance, is promising to relieve the phenotyping bottleneck in biomass measurement in breeding applications. For example, based on an established small reference dataset which is used to train a RF model, it is possible to predict biomass in several large plant populations within one experiment or across several experiments using image data by taking advantage of high-throughput phenotyping technologies. Alternatively, the model can be trained from a much larger reference panel of plants that are grown in diverse environmental conditions which is then applied to a diverse set of experiments. The first evidence for this notion is the observation that our model showed more predictive power in plants with two treatments than with a single treatment (**Fig. 3**, **B** and **D**). Indeed, when applying our model to the combined dataset from all the three experiments, we found the prediction accuracy remains very high ($R^2 =$ 0.96 and $r = 0.98$, average values from ten times of ten-fold cross-validation). To keep the high prediction accuracy in other application, there are some points should be take caution. Considering the environmental effects on biomass accumulation, the application of our model will require the testing experiments showing similar conducted conditions with that of the reference experiments. This means the plant cultivation conditions should be standardized and any noise which might lower image quality should be avoided. Another approach to improve applicability of models, which could not be tested in this study, would be to

improve the data base for the training, by acquiring data from additional environment sensors. Temperature, humidity, and illumination data would certainly help to explain differences in the growth patterns among experiments, performed in different growth seasons. To this end, we expect that our approach is extensible by incorporating such sensor data in the data matrices. Furthermore, our results can provide suggestive hints for biologists to setup phenotyping infrastructures for investigation of plant biomass. For instance, a visible light imaging system would be sufficient to accurately predict fresh weight based on the observation that geometric features alone show high prediction accuracy (**Fig. 4A**). However, to investigate dry weight, it would be helpful to include an additional near-infrared camera system under normal growth conditions and an additional fluorescence camera system under drought stress conditions (**Fig. 4B**).

In contrast to previous studies [2-3, 6-7, 18, 32-37], in which biomass was investigated using only single image-derived parameter (such as projected area) or several geometric parameters, our analyses extended these studies by incorporating more representative features that cover both structural and physiological-related properties into a more sophisticated model. Although the predictive power of our model is roughly higher than that of single feature-based prediction, such as the digital volume (**Fig. 3**) [20], our model also reveals the relative contribution of individual feature in prediction of biomass. The information regarding the importance of each feature will offer new insights into the phenotypic determinants of plant biomass outcome. Interestingly, we found that several top ranked features, such as digital volume and NIR intensity, showed genetic correlations with biomass of fresh weight (**Fig. 4C**) [20], implying these top ranked features may represent the main "phenotypic components" of biomass outcome and can be further used to dissect genetic components underlying biomass accumulation. As image-based high-throughput phenotyping in plants developed mainly in recent years and therefore few corresponding modelling studies have been performed, we believe that our model could be further improved when new types of cameras and/or newly defined features are available.

In summary, we have developed a quantitative model for dissecting the phenotypic components of biomass accumulation based on image data. Apart from predicting biomass outcome, the methods can be used to determine the most important image-based features related to plant biomass accumulation, which are

promising for subsequent genetic mapping to uncover the genetic basis of biomass.

Potential Implications

As high-throughput plant phenotyping is a technique which is becoming more and more widely used for automated phenotype in plant research, especially in plant breeding, we anticipate that the methodologies proposed in this work will have various potential applications. We anticipate that the analysis results will be useful to advance our views of the phenotypic determinants of plant biomass outcome, and the statistical methods can be broadly used for other plant species and therefore assist plant breeding in the context of phenomics.

Materials and Methods

Germplasm and experiments

Barley plant image data were obtained as described previously [20, 30]. Briefly, a core set of 16 two-rowed spring barley cultivars (*Hordeum vulgare* L.) and two parental cultivars of a double haploid (DH) were monitored for vegetative biomass accumulation. Three independent experiments with identical setup were performed in a (semi-) controlled greenhouse at IPK by using the automated phenotyping and imaging platform LemnaTec-Scanalyzer 3D. Experiments were performed consecutively from May to November 2011 over a period of 58 days each (**Supplemental Table S1**). The greenhouse setup enabled sowing for the next experiment already 2 days before the old experiment ended. For this, new pots were placed in the middle of the greenhouse, while the old experiment was still on the conveyer belts.

Each experiment consisted of two treatments: well-watered (control treatment) and water limited (drought stress treatment). In each treatment, nine plants per core set cultivar as well as six plants per DH parent were tested. This resulted in a total of 312 plants per experiment, corresponding to the maximal capacity of the phenotyping platform. Watering and imaging were performed daily. Drought stress was imposed by intercepting water supply from 27 days after sowing (DAS 27) to DAS 44. Stressed plants were re-watered

at DAS 45. In total, for each of the experiments about 100 GB of raw (image) data was accumulated. At the end of experiments (DAS 58), plants were harvested to measure above-ground biomass in form of plant fresh weight (FW; for all experiments) and/or dry weight (DW; for experiment 1).

Image analysis

Image datasets were processed by the barley analysis pipelines in the IAP software (version v1.1.2) [19]. Analysed results were exported in the csv file format via IAP functionalities, which can be used for further data inspection. The result table includes columns for different phenotypic traits and rows as plants are imaged over time. The corresponding metadata is included in the result table as well.

Each plant was characterized by a set of phenotypic traits also referred to as features, which were grouped into four categories: geometric features, fluorescence-related (FLUO-related) features, colour-related features and near-infrared-related (NIR-related) features. These traits were defined by considering image information from different cameras (visible light, fluorescence and near infrared) and imaging views (side and top views). See the IAP online documentation (http://iapg2p.sourceforge.net/documentation.pdf) for details about trait definition.

Feature selection

Feature selection was performed with the same procedure as described in [20]. We applied the feature selection technique to each dataset. Generally, we captured almost identical subset features from different datasets. We manually added several representative traits due to removal by variance inflation factors. For example, the digital volume and projected area are highly correlated with each other but we kept both of them, because we would investigate the predictive power of both features. Moreover, the regression models we used are insensitive to collinear features. We thus kept as much representative features as possible. To apply the prediction models among different datasets, a common set of features supported by all the datasets was used.

Data transformation

Each plant can be presented by a representative list of phenotypic traits, resulting in a matrix $X_{n\times m}$ for each experiment, where n is the number of plants and m is the number of phenotypic traits. Missing values were filled by mean values of other replicated plants. To make the image-derived parameters from diverse sources comparable, we normalized the columns of X by dividing the values with the maximum value of each column across all plants. Plants with empty values of manual measurements (FW and DW) were discarded for analysis. These transformed data sets were subjected to regression models.

Hierarchical clustering analysis and PCA

Hierarchical clustering analysis (HCA) and principle component analysis (PCA) were performed on the transformed data matrix $X_{n\times m}$ in the same way as described in [20]. We also performed HCA using the genotype-level mean value of FW data to check the similarity of overall plant growth patterns in different experiments.

Models for predicting plant biomass

To understand the underlying relationship between image-derived parameters and the accumulated biomass (such as FW and DW), we constructed predictive models based on four different machine-learning methods: multivariate linear regression (MLR), multivariate adaptive regression splines (MARS), random forest (RF) and support vector regression (SVR). In these models, the normalized phenotypic profile matrices $X_{n\times m}$ for a representative list of phenotypic traits were used as predictors (explanatory variables) and the measured DW/FW as the response variable Y.

All these models were implemented in R (http://www.r-project.org/; release 2.15.2). To assess the relative contribution of each phenotypic trait to predicting the biomass. We also calculated the relative feature importance for each model. Specifically, for the MLR model, we used the "lm" function in the base installation packages. The relative importance of predictor variables in the MLR model was estimated by a heuristic method [38] which decomposes the proportionate contribution of each predictor variable to R^2. For MARS, we used the "earth" function in the *earth* R package. The "number of subsets (nsubsets)" criterion (counting the number of model subsets that include the variable) was used to calculate the variables feature

importance, which is implemented in the "evimp" function. For the RF model, we used the *randomForest* R package which implements Breiman's random forest algorithm [31]. We chose the "%IncMSE" (increase of mean squared error) to represent the criteria of relative importance measure. For SVR, we utilized the *e1071* R package which provides functionalities to use the *libsvm* library [39]. The absolute values of the coefficients of the normal vector to the "optimal" hyperplane can be considered as the relative importance of each predictor variable contributing to regression [40, 41].

Evaluation of the prediction models

To evaluate the performance of the predictive models, we adopted a 10-fold cross-validation strategy to check the prediction power of each regression model. Specifically, each dataset was randomly divided into a training set (90% of plants) and a testing set (10% of plants). We trained a model on the training data and then applied it to predict biomass for the testing data. Afterwards, the predicted biomass in the testing set was compared with the manually measured biomass. The predictive accuracy of the model can be measured by

1) the Pearson correlation coefficient (PCC; *r*) between the predicted values and the observed values;
2) the coefficient of determination (R^2) which equals to the fraction of variance of biomass explained by the model, defined as

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where SS_{res} and SS_{tot} are the sum of squares for residuals and the total sum of squares, respectively, $\hat{y}_i$ the predicted and y_i the observed biomass of the ith plant, $\bar{y}$ is the mean value of the observed biomass; and

3) the root mean squared relative error of cross-validation, defined as

$$\text{RMSRE} = \sqrt{\frac{\sum_{i=1}^{s}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}{s}}$$

where s denotes the sample size of the testing dataset.

We repeated the cross-validation procedure ten times. The mean and standard deviation of the resulting R^2 and RMSRE values were calculated across runs.

To evaluate the applicability of our methods across seasons (thus different growth environments) and

treatments (e.g., control versus drought stress) in the same season, we applied the models in different contexts with cohort validation. Specifically, we trained the biomass prediction models under one specific context and predicted biomass in another different context and *vice versa*. The predictive accuracy of the model was evaluated based on the measures R^2 and RMSRE as described above. Furthermore, the predictive power was reflected by the bias μ between the predicted and observed values, defined as

$$\mu = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{\hat{y}_i - y_i}{y_i}$$

where n denotes the sample size of the dataset. This bias indicates over- ($\mu > 0$) or under-estimation ($\mu < 0$) of biomass.

Availability of source code and requirements

- Project name: Modeling of plant biomass accumulation with HTP data
- Project home page: https://github.com/htpmod/HTPmod
- Operating system(s): Windows, Linux and Mac OS.
- Programming language: R
- License: open source under GNU GPL v3.0.

Availability of supporting data and materials

The raw image data sets as well as analysed data supporting the results of this article are available in the PGP repository [42] under XXXX (please use the following links for review: https://doi.ipk-gatersleben.de/DOI/aee46b58-628f-4f8b-9097-0c87cdc2fb39/e281580f-58a8-4a95-9b16-a89e22bba55e/2/1847940088, https://doi.ipk-gatersleben.de/DOI/269b0f6b-2bf9-4d31-b6b0-70639a8416a2/2c368112-3f49-467f-9cc1-71b33323b2a0/2/1847940088, and https://doi.ipk-gatersleben.de/DOI/d87676ef-9327-4675-99a2-55a2bd0d95fa/8dbaf3cb-b644-4162-95b1-2f925fe9dfba/2/1847940088), according to the ISA-Tab format and the recommendations of the MIAPPE (Minimum Information About a Plant Phenotyping Experiment) standard [43]. The selected data for modelling are available in the **Supplemental Data S1**.

Declarations

List of abbreviations

DAS: Days After Sowing

DW: Dry Weight

FLUO: Fluorescence

FW: Fresh Weight

HCA: Hierarchical Clustering Analysis

HTP: High-Throughput Phenotyping

MLR: Multivariate Linear Regression

MARS: Multivariate Adaptive Regression Splines

NIR: Near-Infrared

PCA: Principal Component Analysis

PCC: Pearson Correlation Coefficient

RF: Random Forest

RMSRE: Root Mean Squared Relative Error

SVR: Support Vector Regression

Consent for publication

Not applicable.

Funding

This work was supported by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), the Robert Bosch Stiftung (32.5.8003.0116.0) and the Federal Agency for Agriculture and Food (BEL, 15/12-13, 530-06.01-BiKo CHN) and the Federal Ministry of Education and Research (BMBF, 0315958A and 031A053B). This research was furthermore enabled with support of the European Plant Phenotyping Network (EPPN, grant agreement no. 284443) funded by the FP7 Research Infrastructures Programme of the European Union.

Competing interests

The authors declare that they have no competing interests.

Author contributions

D.C. designed the research. C.K. and M.C. supervised the project. K.N. and G.A. performed the LemnaTec experiments. D.A. created the ISA-Tab formatted description and uploaded data records in the PGP repository. J.M.P. and C.K. analyzed image data. D.C. implemented the methods, analyzed data, interpreted the results, and wrote the manuscript with contribution from R.S.. All authors read and approved the final version of the article.

Acknowledgements

We would like to thank Ingo Mücke for his management of the LemnaTec system operations. We thank Michael Ulrich for performing software tests and helping in data analysis. We would like to thank two anonymous referees for their helpful comments and suggestions.

Figure Legends

Figure1. Modeling pipeline for predicting plant biomass accumulation based on image-derived parameters. (**A**) Input data, including high-throughput image data and manually measured biomass data. Plants were phenotyped using various cameras such as visible (or color), fluorescence (FLUO) and near-infrared (NIR) sensors. Image analysis was performed with IAP software [10] for feature extraction. The same plants were harvested and measured at the end of growth. Generally, two types of biomass were measured: fresh weight (FW) and dry weight (DW). (**B**) Trait processing. All the phenotypic traits were grouped into four categories: geometric, color-related, FLUO-related and NIR-related traits. Phenotypic data were subjected to quality check to remove low-quality data. (**C**) Each plant was described by a list of traits, resulting in a predictor matrix whose rows represent plants and columns represent image-based traits. This matrix was used to predicted plant biomass accumulation by MLR (multivariate linear regression), MARS (multivariate adaptive regression splines), RF (random forest) and SVR (support vector regression) models. The right panel represents the schema of model validation. In the first schema, a dataset (Dataset 1) was divided into training set and testing set in a ten-fold cross-validation manner. In the second schema, the whole of one dataset (Dataset 1) was used for training and another dataset (Dataset 2) was used for testing. (**D**) Model selection, evaluation and result interpretation. The correlation of the predicted values and measured values was used to assess the overall performance of the model.

Figure 2. Predictability of image-based traits to plant biomass.

(**A**) Schema depicting three consecutive high-throughput phenotyping experiments in barley. Plants in each experiment were harvested for biomass measurements: fresh weight (FW; for all experiments) and dry weight (DW; only for experiment 1). (**B**) Scatter plots showing projections of the top four Principal components (PCs) based on PCA of image-based data. The component scores (shown in points) are colored and shaped according to the experiments (as legend listed in the box). The component loading vectors (represented in lines) of all traits (as colored according to their categories) were superimposed proportionally to their contribution. (**C**) Boxplot showing the distribution of FW across different experiments. (**D**) A dendrogram from cluster analysis based on the means of FW data over genotypes. (**E**) Pearson's correlation (mean values in the three datasets) between image-based traits and FW. Traits with the largest mean correlations values are

labeled: 1 -- sum of leaf length (side view), 2 -- sum of FLUO intensity (side), 3 -- plant area border length (side), 4 -- sum of NIR intensity (top), 5 -- sum of FLUO intensity (top), 6 -- projected area (top), 7 -- projected area (side) and 8 -- digital volume.

Figure 3. Quantitative relationship between image-based features and plant biomass.

(**A**) and (**C**) Scatter plots of manually measured plant biomass (fresh weight [FW] and dry weight [DW]) versus predicted biomass values using four prediction models: multivariate linear regression (MLR), multivariate adaptive regression splines (MARS), random forest (RF) and support vector regression (SVR). The red line indicates the expected prediction ($y = x$). The quantitative relationship between image-based features and biomass was evaluated by Pearson's correlation coefficient (PCC r and its corresponding p-value), RMSRE (root mean squared relative error) and the percentage of variance explained by the models (the coefficient of determination R^2). (**B**) and (**D**) Summary of the predictive power of each regression model. The results were based on ten-fold cross-validation with ten trials. Models were evaluated based on control plants, stressed plants and the whole set of plants. The solid lines represent the predictive performance based on the single "digital volume" feature.

Figure 4. The relative importance of image-based features in prediction of plant biomass.

The capabilities of different types of image-based features to predict plant biomass based on evaluation of either fresh weight (FW) (**A**) or dry weight (DW) (**B**). The overall predictive accuracies of each type of features are indicated. Grey bar denote the predictive accuracy using all features. The relative importance of each feature in the Random Forest model (upper panel) and the predictive accuracy of each individual feature as the single predictor (lower panel) based on investigation of either FW (**C**) or DW (**D**). The calculation was based on the whole set of plants (control and stressed plants). Note that feature labels are shared in the upper and lower panels. Features are shown in numbers as ordered by their names. The three features highlighted in the red dash box are digital volume, projected side area and projected top area. (**E**) Comparison of the relative importance of features in prediction of FW and DW. The top six most different features are highlighted and labeled.

Figure 5. Comparison of prediction accuracy across different experiments. (**A**)Biomass prediction across experiments. Models were trained using data from one experiment and were applied to another experiment for prediction. The whole set of plants (i.e., "control + stressed" plant) were used in the analysis. Brown triangles denote stressed plants and green circles control plants. Red box indicates that the prediction accuracy is relatively high between experiments 1 (Exp. 1) and 2. (**B**) Boxplots of coefficient determination (R^2, left), Pearson's correlation coefficients (r, middle) and the root mean squared relative error (RMSRE, right) for different comparisons. "Within" denotes a model trained and tested on data from the same dataset with specific treatments (control, stress or both), and "Cross" represents a model trained on one dataset and tested on another dataset. "Control → stress" denotes a model trained on data with control treatment and tested on data with stress treatment, and vice versa for "stress → control". The number of possible analyses for each category was shown above the boxes.

Figure 6. Comparison of prediction accuracy across different treatments. Refer to **Figure 5A** for legend. The analysis was performed for control and stressed plants separately.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. The relative importance of image-based features in prediction of biomass in control plants. Refer to **Figure 4** for legend. The calculation was based on control plants.

Supplemental Figure S2. The relative importance of image-based features in prediction of biomass in stressed plants. Refer to **Figure 4** for legend. The calculation was based on stressed plants.

Supplemental Table S1. Overview of three high-throughput phenotyping experiments in barley.

Experiment	#plants/#genotypes[1]	Date of sowing	Date of harvesting	Biomass[2]
Exp. 1 (1121KN)	310/18	27.05.2011	24.07.2011	FW & DW
Exp. 2 (1130KN)	310/18	22.07.2011	18.09.2011	FW

Exp. 3 (1137KN)	309/18	16.09.2011	13.11.2011	FW & DW

[1] Number of plants or genotypes used in analysis (filtered data).

[2]Types of biomass measurement. FW: fresh weight; DW: dry weight.

Supplemental Data S1. Manual data and image-derived data in the three experiments.

References

1. Catchpole WR and Wheeler CJ: **Estimating plant biomass: a review of techniques.** *Australian Journal of Ecology* 1992, **17**: 121–131.
2. Tackenberg O: **A new method for non-destructive measurement of biomass, growth rates, vertical biomass distribution and dry matter content based on digital image analysis**. *Annals of botany* 2007, **99**(4):777-783.
3. Golzarian MR, Frick RA, Rajendran K, Berger B, Roy S, Tester M, Lun DS: **Accurate inference of shoot biomass from high-throughput images of cereal plants.** *Plant Methods* 2011, **7**:2.
4. Fahlgren N, Gehan MA and Baxter I: **Lights, camera, action: high-throughput plant phenotyping is ready for a close-up**. *Current Opinion in Plant Biology* 2015, **24**:93–99.
5. Montes JM, Technow F, Dhillon BS, Mauch F, Melchinger AE: **High-throughput non-destructive biomass determination during early plant development in maize under field conditions.** *Field Crops Research* **121**(2): 268-273.
6. Feng H, Jiang N, Huang C, Fang W, Yang W, Chen G, Xiong L, Liu Q: **A hyperspectral imaging system for an accurate prediction of the above-ground biomass of individual rice plants.** *Review of Scientific Instruments* 2013, **84**(9):095107-095107.
7. Neumann K, Zhao Y, Chu J, Keilwagen J, Reif JC, Kilian B, Graner A: **Genetic architecture and temporal patterns of biomass accumulation in spring barley revealed by image analysis.** *BMC plant biology* 2017, 17(1):137.
8. Zhang X, Huang C, Wu D, Qiao F, Li W, Duan L, Wang K, Xiao Y, Chen G, Liu Q et al: **High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant**

Growth. *Plant Physiology* 2017, **173** (3): 1554-1564.

9. Muraya MM, Chu J, Zhao Y, Junker A, Klukas C, Reif JC, Altmann T: **Genetic variation of growth dynamics in maize (Zea mays L.) revealed through automated non-invasive phenotyping.** *The Plant Journal* 2017, **89**(2):366-380.
10. Fahlgren N., Feldman M., Gehan M.A., Wilson M.S., Shyu C., Bryant D.W., Hill S.T., McEntee C.J., Warnasooriya S.N., Kumar I., Ficor T., Turnipseed S., Gilbert K.B., Brutnell T.P., Carrington J.C., Mockler T.C., and Baxter I: **A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in Setaria.** *Molecular Plant* 2015, **8**: 1520–1535.
11. Yang WN, Zilong Guo ZL, Huang CL, Duan LF, Chen GX, Jiang N, Fang W, Feng H, Xie WB, Lian XM et al: **Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice.** *Nature Communications* 2014, **5**:5087.
12. Ehlert D, Horn H-J, Adamek R: **Measuring crop biomass density by laser triangulation**. *Computers and electronics in agriculture* 2008, **61**(2):117-125.
13. Ehlert D, Heisig M, Adamek R: **Suitability of a laser rangefinder to characterize winter wheat**. *Precision Agric* 2010, **11**(6):650-663.
14. Erdle K, Mistele B, Schmidhalter U: **Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars**. *Field Crops Research* 2011, **124**(1):74-84.
15. Busemeyer L, Ruckelshausen A, Moller K, Melchinger AE, Alheit KV, Maurer HP, Hahn V, Weissmann EA, Reif JC, Wurschum T: **Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation**. *Scientific reports* 2013, **3**:2442-2442.
16. Cao Q, Miao Y, Wang H, Huang S, Cheng S, Khosla R, Jiang R: **Non-destructive estimation of rice plant nitrogen status with Crop Circle multispectral active canopy sensor**. *Field Crops Research* 2013, **154**:133-144.
17. Fernandez MGS, Bao Y, Tang L, Schnable PS: High-throughput phenotyping for biomass crops. Plant Physiology 2017, DOI: 10.1104/pp.17.00707.
18. Neilson EH, Edwards AM, Blomstedt CK, Berger B, Møller BL, Gleadow RM: **Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal**

crop plant to nitrogen and water deficiency over time. *Journal of experimental botany* 2015.

19. Klukas C, Chen D, Pape JM: **Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping**. *Plant Physiol* 2014, **165**(2):506-518.
20. Chen D, Neumann K, Friedel S, Kilian B, Chen M, Altmann T, Klukas C: **Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis**. *Plant Cell* 2014, **26**:4636-4655.
21. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, Yan K-K, Dong X, Djebali S, Ruan Y *et al*: **Understanding transcriptional regulation by integrative analysis of transcription factor binding data**. *Genome research* 2012, **22**(9):1658-1667.
22. Cheng C, Gerstein M: **Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells**. *Nucleic Acids Research* 2012, **40**(2):553-568.
23. Cheng C, Yan K-K, Yip KY, Rozowsky J, Alexander R, Shou C, Gerstein M, others: **A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets**. *Genome Biol* 2011, **12**(2):R15-R15.
24. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, Gingeras TR, Gerstein M, Guigó R, Birney E *et al*: **Modeling gene expression using chromatin features in various cellular contexts**. *Genome Biol* 2012, **13**(9):R53-R53.
25. Karlić R, Chung H-R, Lasserre J, Vlahoviček K, Vingron M: **Histone modification levels are predictive for gene expression**. *Proceedings of the National Academy of Sciences* 2010, **107**(7):2926-2931.
26. Ma B, Wilker EH, Willis-Owen SAG, Byun H-M, Wong KCC, Motta V, Baccarelli AA, Schwartz J, Cookson WOCM, Khabbaz K *et al*: **Predicting DNA methylation level across human tissues**. *Nucleic acids research* 2014, **42**(6):3515-3528.
27. Zhang W, Spector T, Deloukas P, Bell J, Engelhardt B: **Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements**. *Genome Biology* 2015, **16**(1):14-14.
28. Das R, Dimitrova N, Xuan Z, Rollins RA, Haghighi F, Edwards JR, Ju J, Bestor TH, Zhang MQ: **Computational prediction of methylation status in human genomic sequences**. *Proceedings of the*

National Academy of Sciences 2006, **103**(28):10713-10716.

29. Zheng H, Wu H, Li J, Jiang S-W: **CpGIMethPred: computational model for predicting methylation status of CpG islands in human genome**. *BMC medical genomics* 2013, **6**(Suppl 1):S13-S13.
30. Neumann K, Klukas C, Friedel S, Rischbeck P, Chen D, Entzian A, Stein N, Graner A, Kilian B: **Dissecting spatiotemporal biomass accumulation in barley under different water regimes using high-throughput image analysis**. *Plant, cell & environment* 2015.
31. Breiman L: **Random forests**. *Machine learning* 2001, **45**(1):5-32.
32. Dietz H, Steinlein T: **Determination of plant species cover by means of image analysis**. *Journal of Vegetation Science* 1996, **7**(1):131-136.
33. Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F: **Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis**. *Plant Physiology and Biochemistry* 1999, **37**(9):671-678.
34. Paruelo JM, Lauenroth WK, Roset PA: **Estimating aboveground plant biomass using a photographic technique**. *Journal of Range Management* 2000:190-193.
35. Walter A, Scharr H, Gilmer F, Zierer R, Nagel KA, Ernst M, Wiese A, Virnich O, Christ MM, Uhlig B *et al*: **Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species**. *New Phytol* 2007, **174**(2):447-455.
36. Arvidsson S, Perez-Rodriguez P, Mueller-Roeber B: **A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects**. *New Phytol* 2011, **191**(3):895-907.
37. Hairmansis A, Berger B, Tester M, Roy SJ: **Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice**. *Rice* 2014, **7**(1):16-16.
38. Johnson JW: **A Heuristic Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression**. *Multivariate Behavioral Research* 2000, **35**(1):1-19.
39. Chang C-C, Lin C-J: **LIBSVM: a library for support vector machines**. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2011, **2**(3):27.
40. Loo LH, Wu LF, Altschuler SJ: **Image-based multivariate profiling of drug responses from single**

cells. *Nature methods* 2007, **4**(5):445-453.

41. Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, Harer J, Weitz JS, Benfey PN: **Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems**. *Plant physiology* 2010, **152**(3):1148-1157.
42. Arend D, Junker A, Scholz U, Schuler D, Wylie J, Lange M: **PGP repository: a plant phenomics and genomics data publication infrastructure**. *Database : the journal of biological databases and curation* 2016, **2016**.
43. Cwiek-Kupczynska H, Altmann T, Arend D, Arnaud E, Chen D, Cornut G, Fiorani F, Frohmberg W, Junker A, Klukas C *et al*: **Measures for interoperability of phenotypic data: minimum information requirements and formatting**. *Plant methods* 2016, **12**:44.

Figure 1

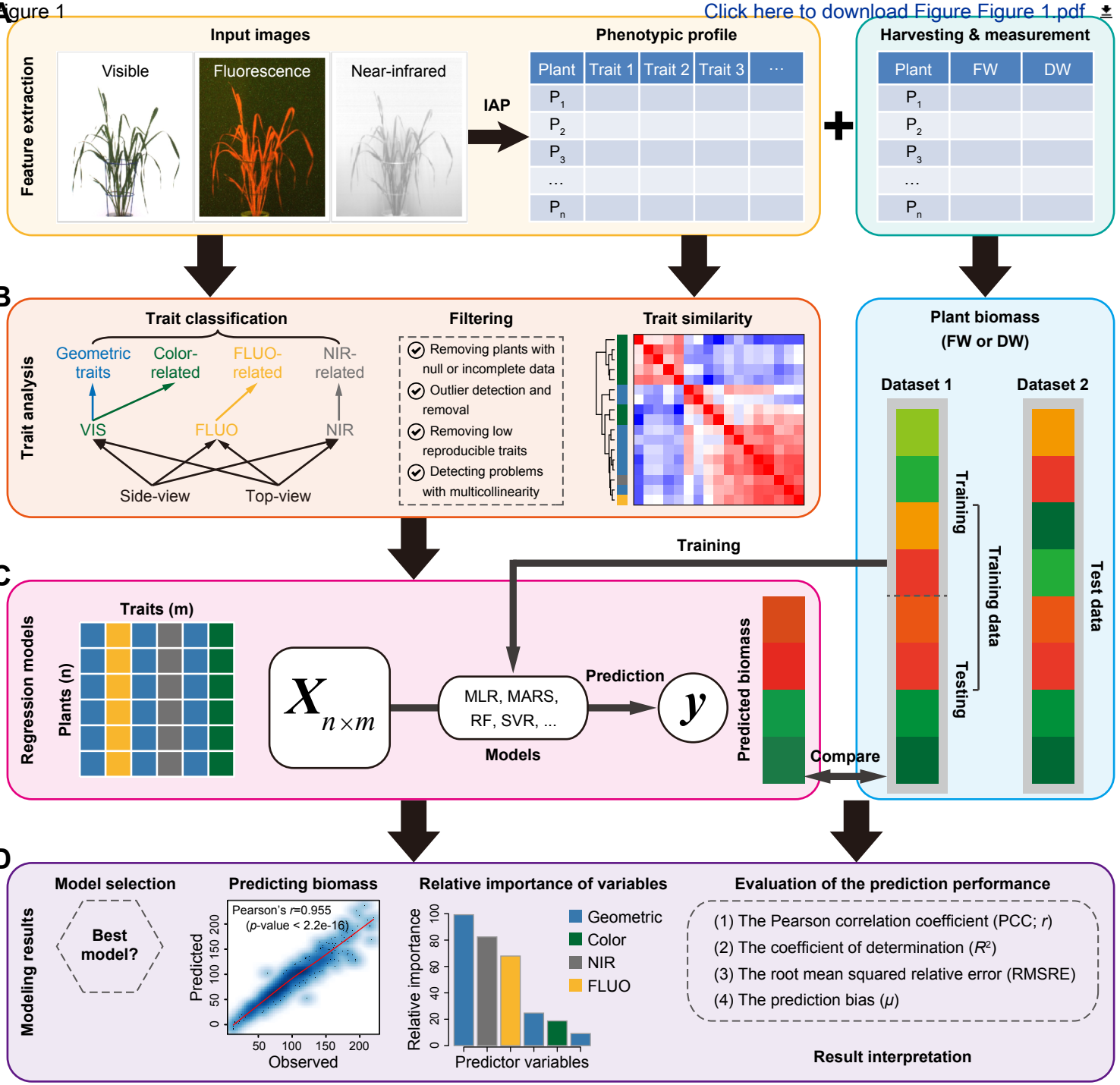


Figure 2

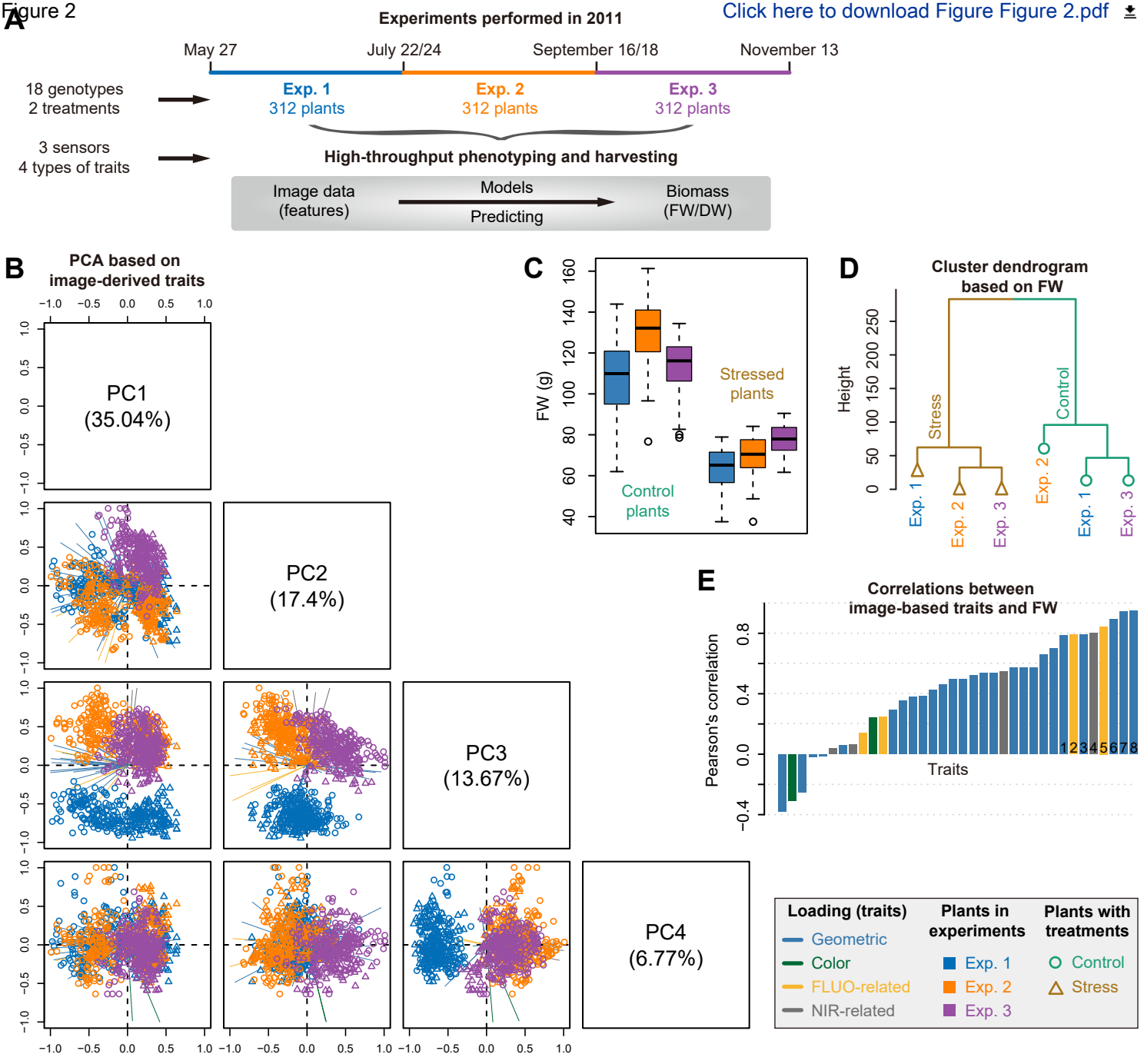


Figure 3

Click here to download Figure Figure 3.pdf

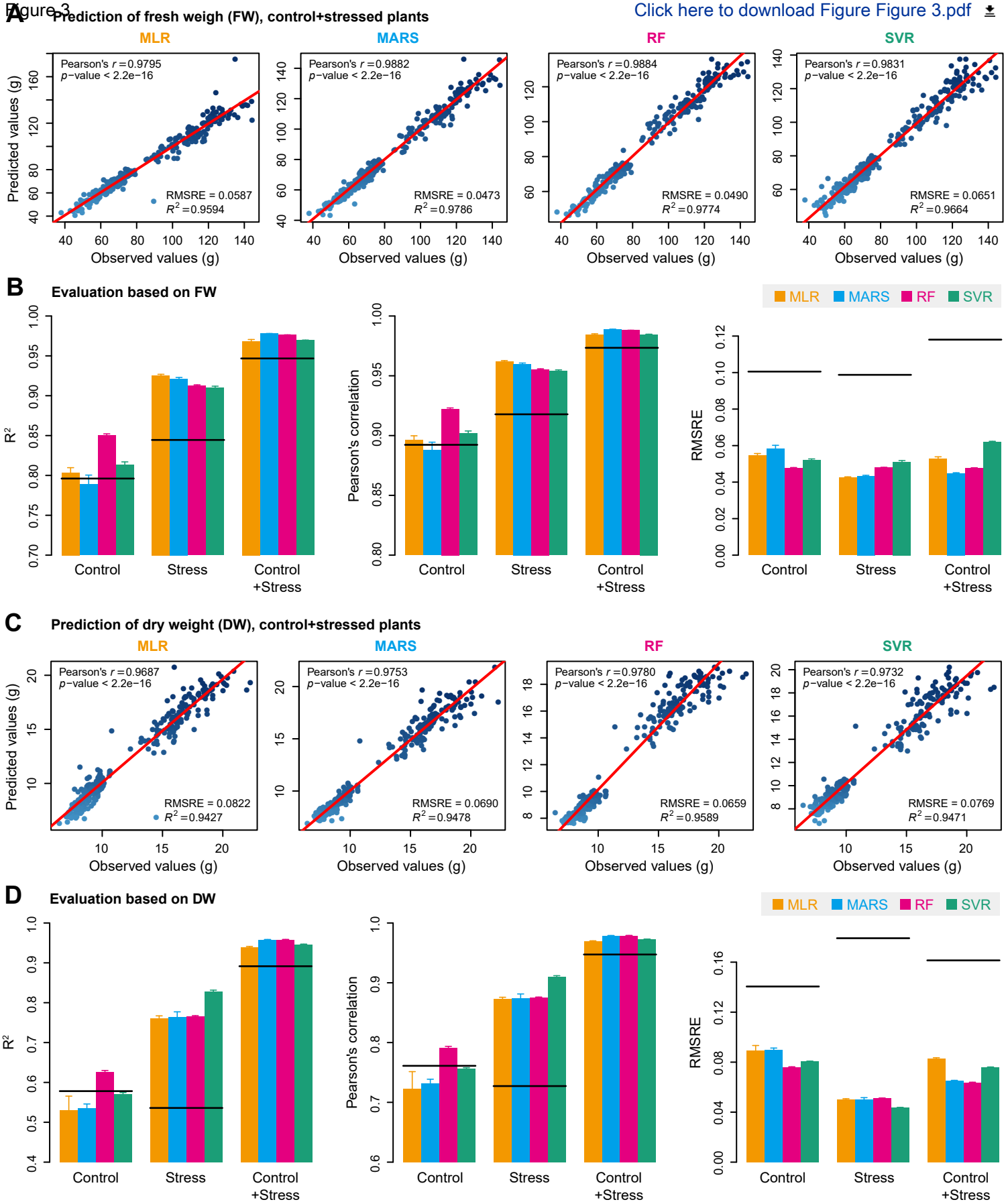


Figure 4

Click here to download Figure Figure 4.pdf

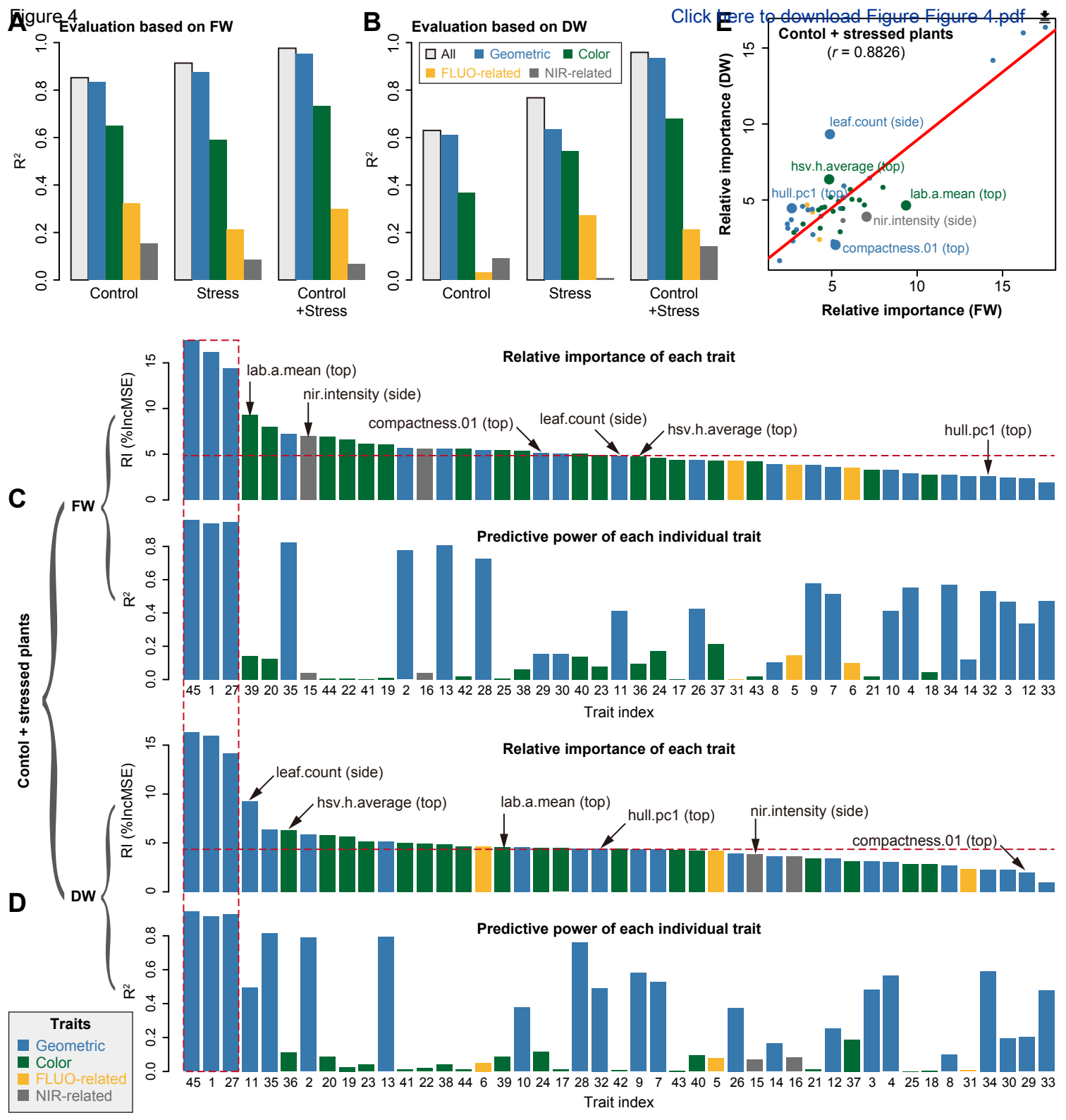


Figure 5

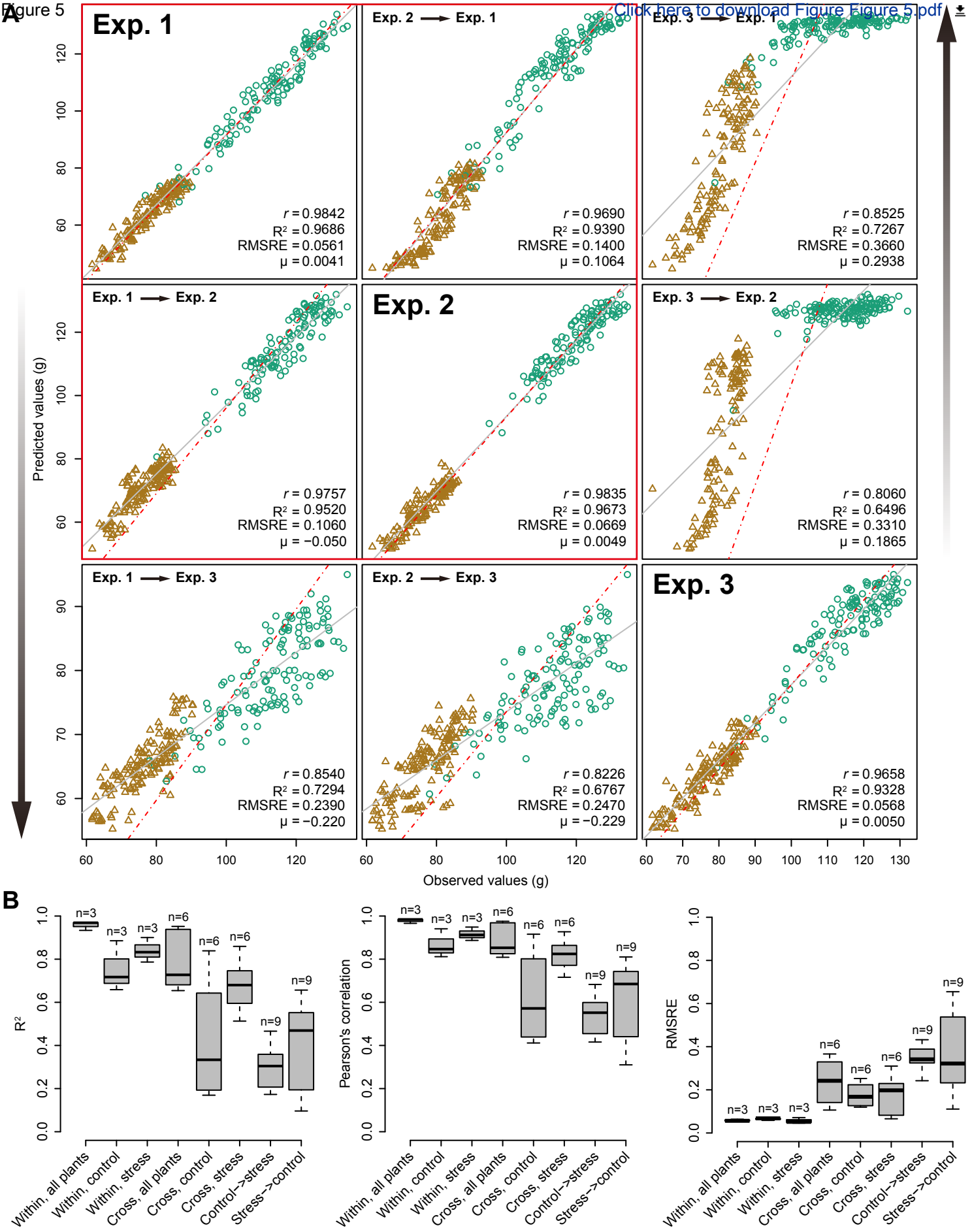
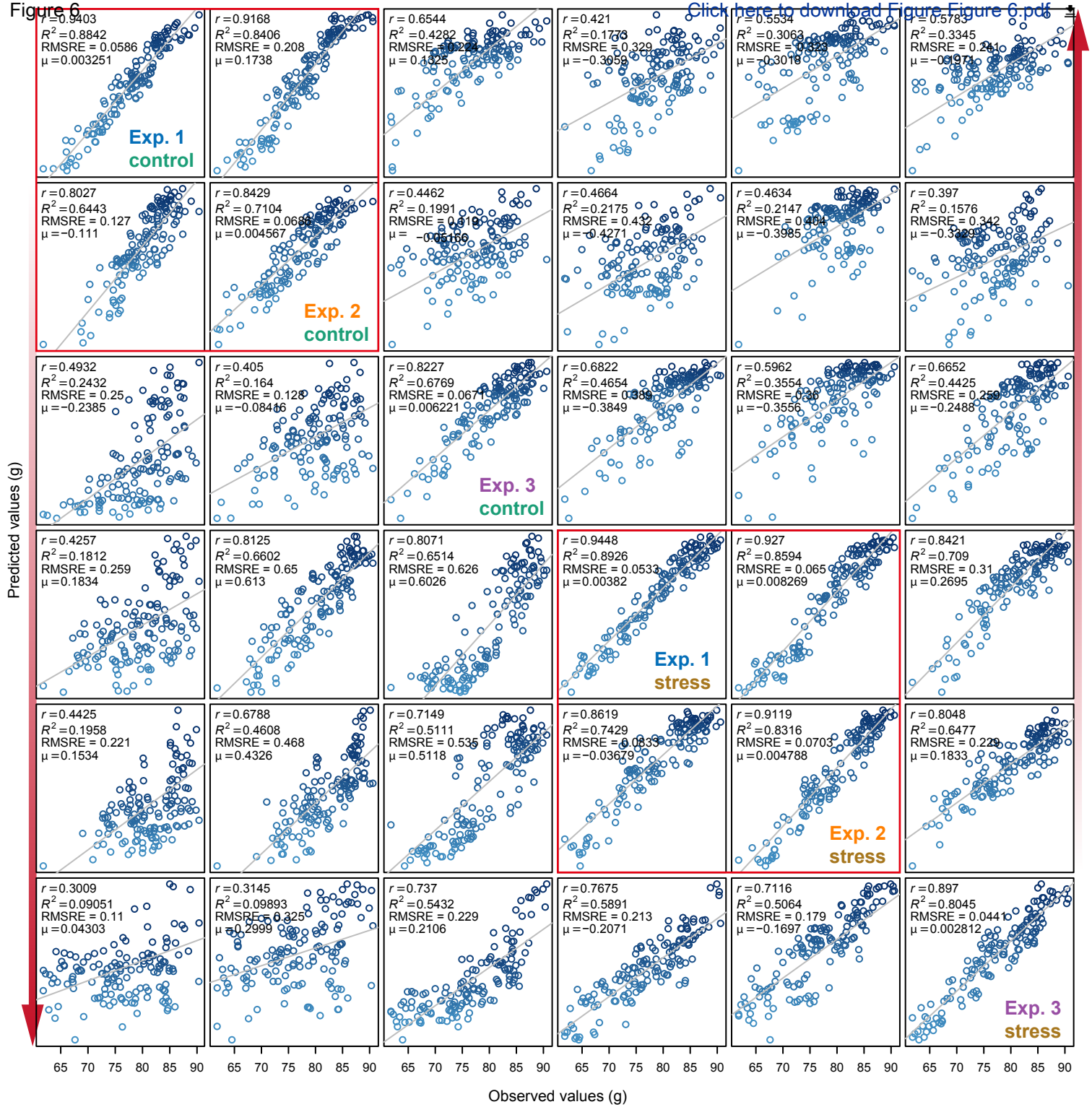


Figure 6

Click here to download Figure Figure 6.pdf



Supplemental Figures S1 and S2