Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation

Lucas Busemeyer¹, Arno Ruckelshausen¹, Kim Möller¹, Albrecht E. Melchinger², Katharina V. Alheit³, Hans Peter Maurer³, Volker Hahn³, Elmar A. Weissmann⁴, Jochen C. Reif²* & Tobias Würschum³

¹Competence Centre of Applied Agricultural Engineering COALA, University of Applied Sciences Osnabrück, 49076 Osnabrück, Germany, ²Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany, ³State Plant Breeding Institute, University of Hohenheim, 70599 Stuttgart, Germany, ⁴Saatzucht Dr. Hege GbR Domäne Hohebuch, 74638 Waldenburg, Germany.

To extend agricultural productivity by knowledge-based breeding and tailor varieties adapted to specific environmental conditions, it is imperative to improve our ability to assess the dynamic changes of the phenotype of crops under field conditions. To this end, we have developed a precision phenotyping platform that combines various sensors for a non-invasive, high-throughput and high-dimensional phenotyping of small grain cereals. This platform yielded high prediction accuracies and heritabilities for biomass of triticale. Genetic variation for biomass accumulation was dissected with 647 doubled haploid lines derived from four families. Employing a genome-wide association mapping approach, two major quantitative trait loci (QTL) for biomass were identified and the genetic architecture of biomass accumulation was found to be characterized by dynamic temporal patterns. Our findings highlight the potential of precision phenotyping to assess the dynamic genetics of complex traits, especially those not amenable to traditional phenotyping.

Results

Precision phenotyping platform. We have developed a novel precision phenotyping platform that enables high-throughput and high-dimensional phenotyping of small grain cereals in the field (Fig. 1). Our precision phenotyping platform permits the collection of data from complementary types of sensors and incorporates light curtains, laser distance sensors, 3D-Time-of-Flight cameras, and high-resolution hyperspectral imaging. In addition, it allows for sensor fusion to accurately predict biomass under field conditions (Fig. 1b,c). We conducted...
a detailed calibration experiment for our precision phenotyping platform based on 1,200 biomass yield plots of triticale. The technical repeatability of the different single sensor measurements was high (Table S1). The platform enables high-throughput screening of more than 2,000 plots per day which in combination with the nearly fully automated subsequent data analysis facilitates rapid and economic trait determination. The technical repeatability of the sensor fusion calibration model for biomass was very high (Fig. 1d) and we obtained excellent prediction accuracies for biomass at all three developmental stages (Fig. 1e). Transferability of the calibration models across environments was evaluated by predicting biomass yields of the year 2012 based on the data from 2011 and vice versa. This yielded high cohort validated $R^2$ values of 0.93 and 0.84 (Fig. S3).

**Temporal genetic patterns of biomass accumulation.** Using the precision phenotyping platform and the developed calibration models we predicted biomass at three developmental stages (BM1–BM3) in a large mapping population of 647 doubled haploid triticale lines (Fig. 2a, S4a). The predicted biomass data were highly heritable with heritabilities ranging from 0.78 to 0.84 (Table 1, S3). We used a genome-wide association mapping scan to identify QTL associated with biomass at each of the three developmental stages (Table 1, Fig. 2, S5, S6a). A detailed analysis of the linkage disequilibrium structure in the genomic regions associated with biomass (Fig. S7–S10) and the collinearity among markers (Table S4) revealed two major QTL regions on chromosomes 5A and 5R with effects on biomass (Fig. S11). The markers most closely linked to the two major QTL and thus candidates for a marker-assisted selection are wPt-2329 on chromosome 5A and rPt-399681 on chromosome 5R. Together, all identified QTL explained 40.14, 31.55, and 28.52% of the genetic variance of predicted BM1, BM2, and BM3, respectively.

A multivariate analysis incorporating BM1, BM2, and BM3, jointly identified the same two major QTL on chromosomes 5A and 5R that have been identified in the analyses of the single developmental stages (Fig. 2, S4b, S6b, Table S5, S6). In addition, growth curve parameters were estimated for each line and used for QTL mapping (Fig. S13, Table S7, S8). To dissect the genetic architecture underlying biomass accumulation, we performed full 2-dimensional epistasis scans. Epistatic QTL for biomass were detected for all three developmental stages (Fig. 3).

**Discussion**

Genomic approaches to dissect the genetic architecture of complex traits rely on both high-quality genotypic and phenotypic data. While the power of current genotyping technologies enables high-density genotyping or even resequencing of entire genomes, our ability to measure traits in the field is limited. Advances in sensor technologies, however, now offer an array of tools that can be deployed to overcome this phenotyping bottleneck and bring a revolution to the assessment of the plant phenome. The precision phenotyping platform presented here combines various sensors for a non-invasive assessment of small grain cereals under field conditions.
With a capacity of more than 2,000 plots per day for data collection and post-processing of the collected raw data, it outperforms the phenotyping capacity of a person and, most importantly, permits the collection of data for novel traits not amenable to traditional phenotyping (e.g., biomass, stress tolerance, or primary and secondary metabolites).

The limited prediction accuracies of the calibration models based on single sensors (Table S1) clearly emphasized the need to integrate data from multiple sensors to capture complementary information on plant characteristics (Table S2). This is, for example, illustrated by the improvement in biomass prediction accuracy by considering dry matter content which can accurately be predicted with hyperspectral imaging (Fig. S1, S2). The high technical repeatability (Fig. 1d), the high prediction accuracies (Fig. 1e) and the high heritability (Table 1) illustrate that the sensors and the sensor fusion calibration models as employed here are well suited for the non-invasive assessment of biomass yield. A parameter of eminent relevance for plant breeding is the transferability of the established calibration models across environments. The high accuracies obtained for calibration models established in one year and predicting in the other year demonstrate the robustness of the approach (Fig. S3). In conclusion, the precision phenotyping platform facilitates the non-invasive collection of high-quality, multi-dimensional, and high-throughput phenotypic data under field conditions.

A large mapping population of 647 doubled haploid triticale lines derived from four crosses was employed to determine the heritable portion of variation for biomass accumulation and dissect the underlying genetic architecture. The biomass data predicted based on the developed calibration models were highly heritable at all three developmental stages (Table 1, S3). These high heritabilities in combination with the high prediction accuracies illustrate the great potential of the precision phenotyping platform for genomics approaches. Genome-wide association mapping scans at the three developmental stages identified two major QTL regions associated with biomass yield (Fig. 2). The moderate predictive power for the proportion of explained genotypic variance obtained with the identified QTL (Table 1) suggests that biomass must be regarded as highly complex trait with many small effect QTL that escape detection in QTL mapping approaches. Consistent with the complex nature of the trait, grain yield, heading time, spikes per square meter, 1000-kernel weight, and early plant height have recently been identified as key contributors to early biomass. Our results illustrate the great importance of non-invasively assessing different time points in plant development as a prerequisite for knowledge-based adaptation breeding to tailor cultivars to local climatic conditions (Fig. S12).

Table 1 | Detection of main effect QTL

|          | BM1  | BM2  | BM3  |
|----------|------|------|------|
| $h^2$    | 0.78 | 0.84 | 0.79 |
| QTL      | 23   | 25   | 17   |
| $pG$     | 40.14| 31.55| 28.52|

Heritability ($h^2$), number of significantly associated markers (QTL), and the proportion of explained genotypic variance ($pG$ in%) for biomass (BM) at three developmental stages.
genotypic variance (BM) at the three developmental stages and their proportion of explained variance in biomass accumulation (Fig. 2). Whereas the major QTL on chromosome 5R is active throughout plant development, the other major QTL on chromosome 5A shows a temporal pattern of activity. It contributes strongly to biomass at the early stage, then its activity ceases and by the last developmental stage it has completely discontinued its contribution to biomass accumulation. This may also be caused by the temporal contribution of one of the above-mentioned traits with a causal relation to biomass, for example heading time. The observed temporal activity of QTL motivated us to perform a QTL scan accommodating the multivariate nature of the developing trait which confirmed the need for a temporal assessment because many QTL would remain undetected by the traditional static examination of only a single time point in development (Fig. 2, S4b, S6b, Table S5, S6). The growth of organisms has recently been shown to follow a sigmoid curve based on fundamental principles for the allocation of metabolic energy between maintenance of existing tissue and the production of new biomass14. As an alternative approach we therefore determined growth curve parameters based on a logistic growth function, a common sigmoid function, and used these for QTL mapping1. This approach confirmed that biomass accumulation is to a large extent genetically determined (Table S7) and the identified QTL further substantiate the dynamic genetics underlying the trait (Fig. S13, Table S8). In conclusion, a temporal assessment of the phenome, as facilitated by the developed precision phenotyping platform, is of paramount importance to unravel the complex genetic patterns underlying the expression of dynamic traits. This knowledge can in turn assist the selection of lines with complementary pattern types as crossing parents.

The observed temporal activity of individual QTL suggested that also interactions among loci may show patterns of genetic regulation during plant development. Epistasis refers to the dependency of the effect of an allele at one genetic locus on the allele status at one or multiple other loci15. Recent work at the level of proteins has provided evidence that epistasis is a key player regulating evolution and potentially also fitness levels of individuals within a population16. In accordance with this, epistasis has been shown to contribute to the genetic architecture of complex traits in crops17,18. Employing full 2-dimensional epistasis scans we detected epistatic QTL for biomass at all three developmental stages (Fig. 3). The proportion of explained genotypic variance by single epistatic QTL was small, but given their high number we speculate that epistasis contributes substantially to the heritability of biomass in small grain cereals. Our results on the complex trait biomass thus corroborate those from a recent theoretical approach showing that the contribution of epistasis increases with the biological complexity of the trait19. Interestingly, we found that the main effect QTL detected on chromosome 5R is also involved in a high number of epistatic interactions with loci throughout the genome. Such epistatic master regulators have recently been described in Arabidopsis, wheat and sugar beet20–22. A possible molecular mechanism mediating such associations has been provided by the discovery that the human transcription factor *KLF14* acts as *trans* master regulator of adipose gene expression23 suggesting that transcriptional regulation may in part explain the observed epistatic interactions. As for the main effect QTL, we observed temporal patterns also for the epistatic interactions. Our results thus illustrate that the entire genetic architecture of biomass accumulation in triticale is under dynamic control.

In a wider context, our discoveries on biomass accumulation in triticale are likely to have broad relevance to other crops as well as other traits showing dynamic genetic patterns of regulation. We anticipate that with the incorporation of information from additional sensors other agronomic important traits like for example disease resistances can be dynamically phenotyped. On the basis of our results we conclude that precision phenotyping platforms may become the method of choice to assess the genetics of complex dynamic traits under field conditions.
Methods

Plant material and field trials. For this study we used triticate as model species for small grain cereals. Two experiments were conducted: one for the calibration of the precision phenotyping platform and a second experiment aimed at dissecting the genetic architecture of biomass accumulation. The calibration experiment was based on 25 diverse tritite line sets that were grown at one location (Stuttgart-Hohenheim, Germany) in two years under two plant densities (280 plants per m² as optimum and reduced to 150 plants per m² as suboptimal density) with two replicates per treatment combination. The plants were phenotyped with the precision phenotyping platform and subsequently harvested with a field chopper at approximately BBCH stage 49 (awns visible), 69 (late flowering), and 81 (very early dough development) to determine the reference fresh weight. Every sample of every plot was dried to determine dry matter content and dry biomass yield.

The second experiment was based on a mapping population of 647 doubled haploid tritite line sets derived from four families designated AxB (131), AxC (120), DxE (200), and DxF (196) which have been described by Alheit et al.20 as populations DH6, DH7, EAW74 and EAW78. The lines from the done populations were in partially replications including common checks with 960 plots per location at a plant density of 280 plants per m². The mapping population was grown at two locations (Germany: Stuttgart-Hohenheim, 48.77° latitude, 9.18° longitude; Bohlingen, 47.72° latitude, 8.9° longitude) in two years (2011 and 2012). All plants were assessed with the phenotyping platform between 9 AM and 6 PM at approximately BBCH stage 49, 69, and 81 to predict biomass based on the calibration models established in the calibration experiment. The growth rate parameters were determined with the R package grofit28. The growth rate is expressed by the maximum biomass yield.

Association mapping. The plants were genotyped with 1710 DArT markers and the map positions of a consensus map were used for the analysis of linkage disequilibrium (LD) was measured as r² [ref 29] and calculated with software package Plabso. The association mapping was done with a mixed model approach incorporating kinship information30,31,32. For main effect QTL and for epistatic QTL, the Bonferroni-Holm procedure33 was applied to correct for multiple testing with P < 0.05. In addition, a multivariate mixed model approach with different models for the variance structure was used to allow for correlations among the traits. All QTL were phenotyped with the software package ASReml 3.0.34. The proportion of genotypic variance (p²) explained by the detected QTL was calculated by fitting all QTL simultaneously in the order of the strength of their association with the trait in a linear model including a family effect to obtain the sums of squares of the QTL (S(qT)). Thus, p² = SSqT/SS, where SS refers to the heritability.35 The circular plots illustrating the epistatic interactions were created with Circos.35 The results from the QTL mapping are available as supplementary data.

Phenotyping platform. We implemented a sensor platform using a tractor pulled trailer equipped with two light curtains (Infrascan 3000, Sitromnic GmbH, Steyregg, Austria), the Laser distance sensors (LDS1: 96k/66-2300-S12, LDS2: 96k/66-2300-S12, LDS3: 96k/66-2300-S12, Baumer Holding AG, Frauenfeld, Switzerland), two 3D-Time-of-Flight cameras (Effector3D, IIM Electronic GmbH, Essen, Germany), and a hyperspectral imaging system (Helios Core NIR, EVK Di Kierscheggl GmbH, Raaba, Austria) equipped with a 12 W halogen lighting system (Figure 1). The sensors were mounted at the back of the trailer in a separate height adjustable sensor module which was shaded with a black canvas to avoid influence of direct solar radiation. A detailed technical overview of the developed precision phenotyping platform including descriptions of the mechanical design, the integrated sensor systems, the hard- and software design for plot based data collection and analysis and the phenotyping procedure are given in Busmeier et al.11.

Biomass prediction. The model to determine the biomass of the plots is based on the fusion of parameters with selectivity to (i) the volume of the plants and (ii) their density, both extracted from sensor raw data for each plot. These parameters were fused and related to the dry biomass of each plot of the calibration experiment by multiple linear regression analysis to generate a calibrated biomass determination model. The density of a plant is related to its dry matter content. Consequently, we used the hyperspectral imaging system to determine the dry matter content of the plants (which is physically based on the selectivity of the spectral imaging system to the plants’ moisture content) and used this information as an approximation for the density of the plants. The automated analysis of spectral imaging data was performed with an application developed with the software package MATLAB (The Math Works, Natick, USA) including all steps of data transformation between sensor raw data and dry matter content for each plot. In a first step noise occurring in the raw data coming from uneven spectral sensitivities of the sensor and a spatially unbalanced illumination was compensated by dividing all measured spectra by a function obtained with software package MATLAB. We computed maximum likelihood estimates with a limit of 100 iterations and the default settings of convergence tolerance for changes of beta and the objective function. We applied a forward model selection algorithm with function stepwisefit of software package MATLAB with an entrance tolerance of P < 0.01 and an exit tolerance of P > 0.05 to assess the contribution of the different parameters to the described variance in the model and to select a final sensor fusion calibration model of biomass.

Evaluation of the prediction models. We determined the quality of our calibrations estimating (i) the root mean squared relative error of calibration as a measure for the accuracy of the biomass predictions as

\[ \text{RMSE}_{cal} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2} \]

where y denotes the observed biomass for each plot in the field, the n parameter extracted out of one class of sensors but also across sensors, x, the ith parameter of the model, b, the regression coefficients, and c the error term. The logarithmic model was chosen, because approximation of volume is based on a multiplicative action of single parameters. Regression analysis was performed with function mregress of software package MATLAB. We computed maximum likelihood estimates with a limit of 100 iterations and the default settings of convergence tolerance for changes of beta and the objective function. We applied a forward model selection algorithm with function stepwisefit of software package MATLAB with an entrance tolerance of P < 0.01 and an exit tolerance of P > 0.05 to assess the contribution of the different parameters to the described variance in the model and to select a final sensor fusion calibration model of biomass.
of biomass yield. Furthermore, the root mean squared relative error (RMSRE$_w$) and the coefficient of determination were determined for the cohort validation ($R^2_w$).

**Technical repeatability.** To assess the precision of the developed precision phenotyping platform, every plot was recorded twice within a repetition time of less than 10 minutes except for BM3 in year 2011 where the repetition time was about 90 minutes. The technical repeatability was determined by comparing the two measurements with linear regression. To quantify the results, the following statistics were applied: (i) root mean squared relative error of repetition

$$RMSRE_w = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left( \frac{y_{i} - \bar{y}_w}{\bar{y}_w} \right)^2}$$

where $m$ denotes the number of repeated samples in the dataset, $y_i$ the first and $\bar{y}_w$ the second repetition of the same plot and (ii) the coefficient of determination of repetition ($R^2_w$).

1. Evenson, R. E. & Golini, D. Assessing the impact of the green revolution, 1960 to 2000. Science 300, 758–762 (2003).
2. Takeda, S. & Matsuoka, M. Genetic approaches to crop improvement: responding to environmental and population changes. Nature Rev. Genet. 9, 444–457 (2008).
3. Edwards, D., Batley, J. & Snowdon, R. J. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 126, 1–11 (2013).
4. Wu, R. & Lin, M. Functional mapping – how to map and study the genetic architecture of dynamic complex traits. Nature Rev. Genet. 7, 229–237 (2006).
5. White, J. W. et al. Field-based phenomics for plant genetics research. Field Crops Res. 133, 101–112 (2012).
6. Montes, J. M., Melchinger, A. E. & Reif, J. R. Novel throughput phenotyping platforms in plant genetic studies. Trends Plant Sci. 12, 433–436 (2007).
7. Ehliert, D., Horn, H.-J. & Adamek, R. Measuring crop biomass density by laser triangulation. Comput. Electron. Agric. 61, 117–125 (2008).
8. Ehliert, D., Hessig, M. & Adamek, R. Suitability of a laser rangefinder to characterize winter wheat. Precision Agric. 11, 650–663 (2010).
9. Erdle, K., Mistele, B. & Schmidhalter, U. Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. Field Crop Res. 124, 74–84 (2011).
10. Ruckelshausen, A. Autonomous robots in agricultural field trials. In: Bleiholder, H., Piepho, H.-P. (Eds.), Proceedings of the International Symposium “Agricultural Field Experiments – Today and Tomorrow”, 190–197 (2007).
11. Bussemeyer, L. et al. BreedVision – A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. Sensors 13, 2830–2847 (2013).
12. Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. Nature Rev. Genet. 10, 565–577 (2009).
13. Gowda, M. et al. Potential for simultaneous improvement of grain and biomass yield in Central European winter triticale germplasm. Field Crops Res. 121, 153–157 (2011).
14. West, R. B., Brown, J. H. & Enquist, B. J. A general model for ontogenetic growth. Nature 413, 628–631 (2001).
15. Carlbom, G. & Haley, C. S. Epistasis: too often neglected in complex trait studies? Nature Rev. Genet. 5, 618–625 (2004).
16. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. Nature 490, 535–538 (2012).
17. Buckler, E. S. et al. The genetic architecture of maize flowering time. Science 325, 714–718 (2009).
18. Würschum, T., Maurer, H.-P., Dreyer, F. & Reif, J. C. Effect of inter- and intragenic epistasis on the heritability of oil content in rapeseed (Brassica napus L.). Theor. Appl. Genet. 126, 435–441 (2013).
19. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. PNAS 109, 1193–1198 (2012).
20. Wentzell, A. M., Boeye, J., Zhang, Z. & Klebentstein, D. J. Genetic networks controlling structural outcome of glucosinolate activation across development. PloS Genet. 4, e1000234 (2008).
21. Reif, J. C. et al. Mapping QTLs with main and epistatic effects underlining grain yield and heading time in soft winter wheat. Theor. Appl. Genet. 123, 283–292 (2011).
22. Würschum, T. et al. Genome-wide association mapping of agronomic traits in sugar beet. Theor. Appl. Genet. 123, 1121–1131 (2011).
23. Small, K. S. et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. Nature Genet. 43, 561–564 (2011).
24. Lancashire, P. D. et al. A uniform decimal code for growth stages of crops and weeds. Ann. Applied Biology 119, 561–601 (1991).
25. Würschum, T., Tucker, M. R., Reif, J. C. & Maurer, H.-P. Improved efficiency of doubled haploid generation in hexaploid triticale by in vitro chromosome doubling. BMC Plant Biology 12, 109 (2012).
26. Alheit, K. V. et al. Detection of segregation distortion loci in triticale (x Triticeae Wittmack) based on a high-density DARt marker consensus genetic linkage map. BMC Genomics 12, 380 (2011).
27. Williams, E., Piepho, H.-P. & Whitaker, D. Augmented p-rep designs. Biometrical Journal 53, 19–27 (2011).
28. Kahn, M. et al. Robotic fitting biological growth curves with R. Journal of Statistical Software Vol. 33, Issue 7 (2010).
29. Weir, B. S. Genetic data analysis II. 2nd edn. Sinauer Associates, Sunderland (1996).
30. Maurer, H. P., Melchinger, A. E. & Frisch, M. Population genetic simulation and data analysis with Plabsoft. Euphytica 161, 133–139 (2008).
31. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genet. 38, 203–208 (2006).
32. Würschum, T. et al. Comparison of biometrical models for joint linkage association mapping. Heredity 108, 332–340 (2012).
33. Holm, S. A Simple Sequentially Rejective Bonferroni Test Procedure. Scand. J. Stat. 6, 65–70 (1979).
34. Ehrich, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. ASReml User Guide, Release 3.0, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK (2009).
35. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Research 19, 1639–1645 (2010).
36. Park, B., Windham, W. R., Lawrence, K. C. & Smith, D. P. Contaminant classification of poultry hyperspectral imagery using spectral angle mapper algorithm. Biosystems Engineering 96, 323–333 (2007).