



More than 1000 genotypes are required to derive robust relationships between yield, yield stability and physiological parameters: a computational study on wheat crop

Tien-Cheng Wang^{1,2} · Pierre Casadebaig³ · Tsu-Wei Chen¹

Received: 27 June 2022 / Accepted: 10 October 2022 / Published online: 10 March 2023
© The Author(s) 2023

Abstract

Key message Using in silico experiment in crop model, we identified different physiological regulations of yield and yield stability, as well as quantify the genotype and environment numbers required for analysing yield stability convincingly.

Abstract Identifying target traits for breeding stable and high-yielded cultivars simultaneously is difficult due to limited knowledge of physiological mechanisms behind yield stability. Besides, there is no consensus about the adequacy of a stability index (SI) and the minimal number of environments and genotypes required for evaluating yield stability. We studied this question using the crop model APSIM-Wheat to simulate 9100 virtual genotypes grown under 9000 environments. By analysing the simulated data, we showed that the shape of phenotype distributions affected the correlation between SI and mean yield and the genotypic superiority measure (P_i) was least affected among 11 SI. P_i was used as index to demonstrate that more than 150 environments were required to estimate yield stability of a genotype convincingly and more than 1000 genotypes were necessary to evaluate the contribution of a physiological parameter to yield stability. Network analyses suggested that a physiological parameter contributed preferentially to yield or P_i . For example, soil water absorption efficiency and potential grain filling rate explained better the variations in yield than in P_i ; while light extinction coefficient and radiation use efficiency were more correlated with P_i than with yield. The high number of genotypes and environments required for studying P_i highlight the necessity and potential of in silico experiments to better understand the mechanisms behind yield stability.

Abbreviations

APSIM Agricultural production systems sIMulator
GxE Genotype by environment interaction
LAI Leaf area index
SI Stability index

SPG Sampled population of genotypes
SPE Sampled population of environments
TPE Target population of environments

Communicated by Peter Langridge.

✉ Tien-Cheng Wang
wangtien@student.hu-berlin.de

✉ Tsu-Wei Chen
tsu-wei.chen@hu-berlin.de

¹ Present Address: Section of Intensive Plant Food Systems, Albrecht Daniel Thaer-Institute of Agricultural and Horticultural Sciences, Humboldt Universität zu Berlin, Berlin, Germany

² Institut für Gartenbauliche Produktionssysteme, Leibniz Universität Hannover, Hannover, Germany

³ INRAE, UMR AGIR, Université de Toulouse, 31320 Castanet-Tolosan, France

Introduction

To ensure global food security, it is not only important to increase yield gain but also yield stability. Developing stable crop cultivars is therefore crucial maintaining the yield level and adapting to ever-changing weather schemes (Powell et al. 2012; Dwivedi et al. 2016; Macholdt and Honermeier 2017; Bocci et al. 2020; van Frank et al. 2020). Breeding stable plants requires profound crop physiological knowledge and empirical experiences to identify target traits. However, our physiological understanding of yield stability is still scarce (Pedro et al. 2011) since assessing the yield stability of a genotype requires field experiments across multiple years, locations, agriculture practices and comparisons with other genotypes (e.g. 440 progenies in

16 environments in Wang et al. 2015; 191 cultivars under 43 environments in Voss-Fels et al. 2019; 720 lines in 36 environments in Sehgal et al. 2017). Therefore, identifying cultivars with stable yields is time-consuming and labour intensive, which significantly restricts the speed of our knowledge gain in the eco-physiological mechanisms and their genetic controls resulting in yield stability. Furthermore, there is no consensus in the literature about (1) the adequacy of a stability index (SI) to quantify yield stability and (2) the minimal number of sampled environments and sampled genotypes required for evaluating the yield stability (Reckling et al. 2021). In other words, the minimal size of sampled populations of genotypes and of environments for assessing yield stability is unclear. Also, if a population of genotypes is selected, it is unknown how the yield stability of an individual genotype in the population is affected by the phenotypic distribution (e.g. yield distribution) of this population.

In the past, breeders discovered performance and stability related-traits based on their physiological knowledge and practical experience in the field (Bolaños and Edmeades 1993; Pfeiffer et al. 2001; Pedro et al. 2011). Nowadays, crop modelling and simulation can complement such empirical knowledge by generating thousands of virtual genotypes by subtle changes in structural and physiological parameters (Chen et al. 2015; Casadebaig et al. 2016; Perez et al. 2018) and ultimately help to identify structural and physiological traits of an ideotype. This approach allows us to quantify the potential contributions of a physiological parameter to the performance of new cultivars in test environments, accounting for a large climatic variability, the so-called target population of environments (Quilot-Turion et al. 2012; Senapati and Semenov 2020). For example, using the Sirius crop model, stay green and flag leaf area are identified as crucial parameters of wheat (*Triticum aestivum* L.) under drought and heat stress. It suggests a potential to increase the yield of current wheat cultivars in Europe by 3.5–5.2 t ha⁻¹ (Senapati and Semenov 2019). Since crop modelling predicts crop performance in response to given management or climatic regimes (Chenu et al. 2011; Barillot et al. 2014; Kouadio et al. 2015; Casadebaig et al. 2016; Sun et al. 2016; Parent et al. 2018; Leakey et al. 2019; Wu et al. 2019), simulated yield data obtained from crop models may be also used to analyse yield stability and to estimate the minimal population size of sampled environments and genotypes required for evaluating yield stability.

Here we reviewed and compared 11 stability indices for their adequacy to inform plant breeding for both crop yield and stability. We first developed an *R* package (with 11 stability indices including static, dynamic, probabilistic, parametric and nonparametric indices; Wang and Chen 2022), which facilitates the study of stability analysis and well-integrated with other packages for further analysis in

R environment. Second, we reused an *in silico* experiment conducted with the APSIM-Wheat crop model to analyse yield performance of 9100 virtual genotypes grown under 9000 environments (Casadebaig et al. 2016). Data from the *in silico* experiments enabled (1) to demonstrate the analysis pipeline; (2) to determine the minimal number of genotypes and environments to assess yield stability; (3) to identify the contribution of physiological traits on yield stability and (4) to propose physiological mechanisms to achieve stable yield.

Material and methods

Dataset obtained from *in silico* experiment with the APSIM-wheat

Crop model APSIM-Wheat (www.apsim.info) was used to simulate a dataset (<https://doi.org/10.5281/zenodo.4729636>; for details, see Casadebaig et al. 2016) with 9100 virtual genotypes ($N_{\text{gen}} = 9100$) grown under 9000 environments ($N_{\text{env}} = 9000$). In short, virtual genotypes were created by varying the value of 90 independent physiological parameters in a range of $\pm 20\%$ from the reference cultivar *Hartog*, which represents the default parameter values in the APSIM-Wheat. Environments in the dataset contain historical climate data of 125 years (1889–2013) in four locations in Australia (Emerald, Narrabri, Yanco and Merredin, see also Table 1 from Casadebaig et al. 2016) in Australia, in combination with two CO₂ levels (380 and 555 ppm), three nitrogen levels (low: 50%, control: 100% and high fertilization: 100% plus 50 kg ha⁻¹) and three sowing dates (early, control and late). Eight integrated model outputs (grain and straw yields, grain size, grain number, grain protein, leaf area index (LAI), maturity date and flowering date) were used for trait stability analysis. Straw yield was calculated by subtracting grain yield from biomass.

Computation of stability indices of the virtual genotypes with three sampling methods

All analyses were implemented in the *R* environment (R Core Team 2020) where 11 stability indices (SI) were calculated in a customized package *toolStability* (Wang and Chen 2022; <https://github.com/Illustration/toolStability>). The SI in *toolStability* include static and dynamic concepts of stability (Becker and Léon 1988). Under the static concept, the trait of a stable cultivar stays relatively unchanged across different environments. In contrast, dynamic concept takes the environmental mean into account and considers the interactions between genotypes and environments. Furthermore, each concept can be further classified as parametric or nonparametric. In *toolStability*, there are two parametric SI of static concept: environmental variance (Römer 1917)

Table 1 Mean tendency of a parameter ($T_{parameter}$), $|r|$ to yield and to $P_{i,yield}$ and parameter (R^2 and slope) of linear regression of $|r|$ to yield versus $|r|$ to $P_{i,yield}$ from six physiological parameters from 100 SPG in Fig. 4

Sampling method	Physiological parameter	$T_{parameter}$	$ r $ to yield	$ r $ to $P_{i,yield}$	R^2	Slope
Even	<i>y_rue</i>	1.01 ± 0.01	0.76 ± 0.04	0.77 ± 0.03	0.96	0.96
Random	<i>y_rue</i>	1.16 ± 0.08	0.39 ± 0.08	0.45 ± 0.08	0.95	0.94
Top 20	<i>y_rue</i>	1.55 ± 0.93	0.20 ± 0.09	0.27 ± 0.10	0.72	0.88
Even	<i>y_sla</i>	1.01 ± 0.01	0.64 ± 0.05	0.65 ± 0.05	0.98	0.96
Random	<i>y_sla</i>	1.28 ± 1.44	0.24 ± 0.10	0.26 ± 0.10	0.94	0.96
Top 20	<i>y_sla</i>	1.28 ± 0.77	0.20 ± 0.09	0.22 ± 0.09	0.75	0.91
Even	<i>ll_modifier</i>	0.77 ± 0.10	0.27 ± 0.07	0.21 ± 0.07	0.99	1.02
Random	<i>ll_modifier</i>	0.82 ± 0.10	0.47 ± 0.09	0.39 ± 0.10	0.91	1.11
Top 20	<i>ll_modifier</i>	8.52 ± 65.02	0.13 ± 0.08	0.10 ± 0.07	0	- 0.05
Even	<i>potential_grain_filling_rate</i>	0.98 ± 0.02	0.56 ± 0.06	0.55 ± 0.06	0.98	1.03
Random	<i>potential_grain_filling_rate</i>	0.87 ± 0.12	0.29 ± 0.10	0.26 ± 0.10	0.96	0.98
Top 20	<i>potential_grain_filling_rate</i>	0.85 ± 0.12	0.48 ± 0.06	0.41 ± 0.08	0.52	0.96
Even	<i>node_no_correction</i>	0.78 ± 0.20	0.18 ± 0.06	0.15 ± 0.06	0.96	0.98
Random	<i>node_no_correction</i>	0.93 ± 0.90	0.21 ± 0.09	0.18 ± 0.09	0.95	0.96
Top 20	<i>node_no_correction</i>	0.67 ± 0.14	0.43 ± 0.07	0.29 ± 0.09	0.63	1.05
Even	<i>tfac_slope</i>	1.01 ± 0.03	0.35 ± 0.06	0.35 ± 0.06	0.98	1.02
Random	<i>tfac_slope</i>	2.15 ± 6.94	0.09 ± 0.07	0.09 ± 0.07	0.8	0.82
Top 20	<i>tfac_slope</i>	1.71 ± 0.59	0.23 ± 0.08	0.37 ± 0.08	0.73	0.82

Meaning of the physiological parameters (in *italic*) follow the order (left to right in Fig. 4): radiation use efficiency (*y_rue*), potential leaf specific area (*y_sla*), efficiency of roots to extract soil water (*ll_modifier*), potential grain growth rate at grain filling (*potential_grain_filling_rate*), number of growing leaves in the sheath (*node_no_correction*) and temperature effect on biomass accumulation (*tfac_slope*).

and adjusted coefficient of variation (Reckling et al. 2018). A dynamic concept has nine SI, eight parametric and one nonparametric. Parametric dynamic SI are: coefficient of determination (Pinthus 1973), coefficient of regression (Finlay and Wilkinson 1963), deviation mean squares (Eberhart and Russell, 1966), ecovalence (Wricke 1962), genotypic stability (Hanson 1970), genotypic superiority measure (Lin and Binns 1988), safety first index (Eskridge 1990) and stability variance (Shukla 1972). Depending on the value, the coefficient of regression can be static or dynamic (Becker and Léon 1988). The only nonparametric SI in *toolStability* is the variance of rank (Nassar and Hühn 1987). Each SI represents a specific way of describing a kind of interaction between genotypes and environments. The choice of SI depends on the research question. In this study, we focus on the SI that highly correlates with the genotypic mean yield from all environments to target high and stable yield for crop breeding.

To be consistent between the dimensions of trait (e.g. yield, t ha⁻¹) and stability indices, indices which with squared units of trait were square-rooted to avoid artificial nonlinear relationship between trait and SI (e.g. genotypic superiority index, P_i , Lin and Binns 1988, ecovalence, W_i , Wricke 1962 and variance of rank, S_i^2 , Nassar and Hühn 1987). We noticed that the value of ecovalence (W_i) depends on the number of environments. To ensure the comparability of W_i between calculations with different number of

environments, a modified ecovalence (W_i') is proposed as dividing original ecovalence (W_i) with the number of environments. The dimension-less indices remained unmodified (e.g. b_i , Finlay and Wilkinson 1963).

To calculate a SI of a genotype, data of multiple genotypes (referred to as sampled population of genotypes, SPG, selected from 9100 genotypes pool) grown under multiple environments (referred to as sampled population of environments, SPE, selected from 9000 environments pool) are required. If a SI is highly correlated with the mean yield from all environments in the studied genotypes, this SI indicates a high and stable yield at the same time. As a first step, we investigated whether the shape of phenotypic distribution (e.g. yield distribution in the SPG) affects the relationship between mean yield and SI (Fig. S1k). For this, three sampling methods were used: (1) “random” sampling method that resulted in a population with normal distribution, which is commonly found in real breeding programs (Powell and Rutten 2013); (2) “even” sampling method with a flat and even distributed population (Bresseghele et al. 2009), which is created by first dividing the whole population into ten deciles based on the mean genotypic yield in 9000 environments and then randomly sampled 10% of total sampling number in each decile; and 3) “top 20” method representing the population of elite cultivars that had yield values larger than 80% of genotypes from the whole population (Longin and Reif 2014). For each sampling method, 100 virtual genotypes

(number of genotypes in each SPG, $N_{\text{gen}} = 100$) grown under 100 environments (number of randomly selected environments in each SPE, $N_{\text{env}} = 100$) were first selected to test how the shape of phenotype distribution affects the relationship between mean yield and 11 SI.

Estimating the minimal number of environments required to estimate yield stability

To represent both high and stable performance of yield, genotypic superiority measure of yield ($P_{i,\text{yield}}$) was selected to demonstrate the minimal required N_{env} for reliable estimation of a SI. We first calculated $P_{i,\text{yield}}$ of 100 virtual genotypes ($N_{\text{gen}} = 100$) in 100 SPE, in combination with different N_{env} ranging from 3–600 using “random” sampling method. Secondly, the coefficient of variation of $P_{i,\text{yield}}$ ($CV_{P_{i,\text{yield}}}$) was calculated for each N_{env} between 100 SPG. An arbitrary low threshold value (i.e. 5 or 10% $CV_{P_{i,\text{yield}}}$, Piepho 1998) was used to determine the minimal required N_{env} to estimate yield stability. Moreover, to test the effect of sampling methods on $CV_{SI,\text{yield}}$ of other 10 SI, the same setting ($N_{\text{gen}} = 100$, $N_{\text{env}} = 10$ –600 and $SPE = 100$) was applied.

Analysis of the correlation network between plant traits, crop performance and stability

A network analysis (node and edge graph) was performed to illustrate the Pearson correlation coefficient (r , referred to edge in a network) between yield, $P_{i,\text{yield}}$ and physiological parameters (referred to node) of genotypes ($N_{\text{gen}} = 100$, $SPG = 100$) selected by three sampling methods (i.e. “random”, “even” and “top 20”) in a SPE. For each SPG, a table listing mean yield, $P_{i,\text{yield}}$ and 90 physiological parameters of each genotype was created (Supplementary Fig. S2a). The Pearson correlation coefficients between all columns in this table were calculated to produce a r -matrix (Supplementary Fig. S2b), which was further transformed into a linear vector format (r -vector, Supplementary Fig. S2c) with 181 values (total number of all combinations, $C_2^{92} = 4186$, minus number of correlations between physiological parameters, $C_2^{90} = 4005$). To identify whether a physiological parameter tends to explain yield or $P_{i,\text{yield}}$ more, a tendency index ($T_{\text{parameter}}$) in each SPG was quantified by the ratio of absolute r value ($|r|$) for $P_{i,\text{yield}}$ ($|r|$ between $P_{i,\text{yield}}$ and physiological parameter) to $|r|$ for yield ($|r|$ between yield and physiological parameter). If $T_{\text{parameter}} > 1$, this parameter is more related to $P_{i,\text{yield}}$ than yield. On the other hand, a parameter explains yield more than $P_{i,\text{yield}}$ when $T_{\text{parameter}} < 1$.

Estimating the minimum number of genotypes for robust correlation networks between plant traits, crop performance and stability

To acquire the minimum N_{gen} and N_{env} that produce the robust and representative correlation between yield, $P_{i,\text{yield}}$ and physiological parameters, we evaluated the overall strength of correlation network by four steps: (1) 100 SPG in combination with nine genotype numbers ($N_{\text{gen}} = 5, 50, 100, 200, 300, 500, 700, 900$ and 1100) and six environment numbers ($N_{\text{env}} = 5, 50, 100, 300, 500$ and 700, $SPE = 1$) were sampled to obtain 5400 r -vectors (Supplementary Fig. S2d); (2) a table listing r -vectors of 100 SPG was created for each combination of genotype and environment numbers (Supplementary Fig. S2e); (3) since the similarity between different r -networks can be represented by calculating r between two r -vectors in this table, an “edge- r -matrix” representing r between 100 r -vectors from 100 SPG in this table was calculated (Supplementary Fig. S2f); (4) the edge- r -matrix was squared (to represent power of explanation for correlation between nodes) and averaged to obtain an indicator S representing the similarity between networks of 100 SPG (Supplementary Fig. S2g). If S is close to one, networks between SPG are similar and if $S = 0$, networks between SPG are completely different.

Results

Relationships between mean yield and yield stability were affected by the sampling methods.

To facilitate and reproduce our yield stability analysis, we developed “*toolStability*”, which is an *R* package (Wang and Chen 2022) available on a public repository providing a wide range of functions to calculate 11 stability indices (SI). From the 9100 virtual genotypes created by the APSIM-Wheat, 100 of them were selected (number of genotypes, $N_{\text{gen}} = 100$) by three sampling methods (i.e. “random”, “even” and “top 20”) for 100 times (sampled population of genotype, $SPG = 100$) in 100 environments (number of environments, $N_{\text{env}} = 100$, sampled population of environment, $SPE = 1$), resulting in different shape of phenotype distributions (Supplementary Fig. S1k) which represent different strategies or steps in the breeding program.

Pearson correlation coefficient (r) between trait (e.g. mean yield) and SI was used to identify the SI that represent stable and high trait performance simultaneously. Three SI correlated positively to mean yield (Fig. 1a–c): environmental variance ($W_{i,\text{yield}}$), coefficient of regression ($b_{i,\text{yield}}$) and genotypic stability ($D_{i,\text{yield}}^2$). Two SI negatively correlated with mean yield (Fig. 1d–e): genotypic superiority measure ($P_{i,\text{yield}}$) and safety first index. Other five SI

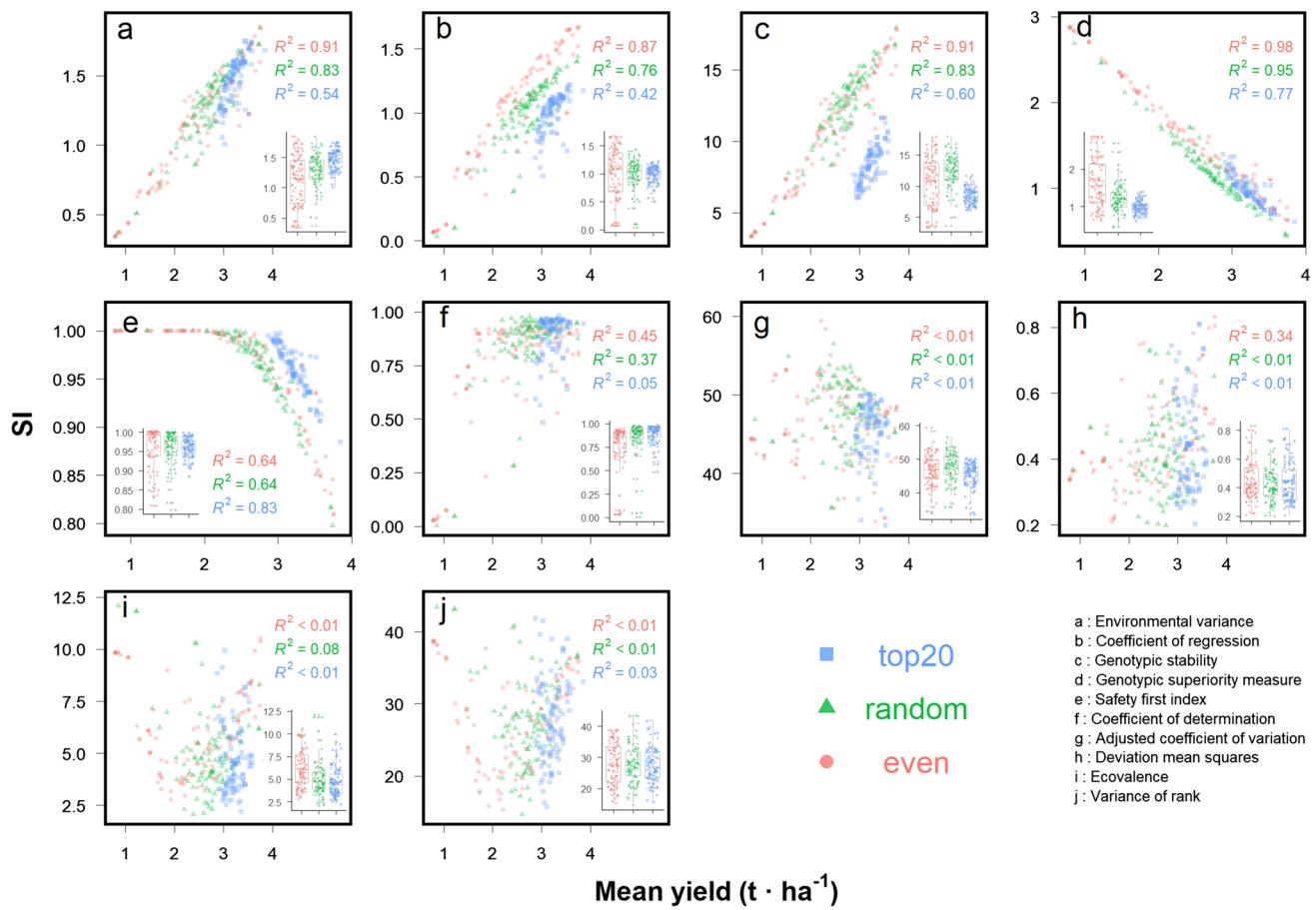


Fig. 1 Comparison of relationship between mean yield and ten stability indices (SI) of randomly selected 100 virtual genotypes (SPG=1) under 100 environments (SPE=1) with three sampling methods. Each point represents a genotype. Red, green and blue colours indicate sampling method “even”, “random” and “top 20”, respectively. Subfigures show the boxplots of SI. The “even” sampling method divides population into ten groups based on the mean yield value,

with ten virtual genotypes randomly selected in each yield group. The “top 20” method randomly selects virtual genotypes with mean yield above 80th percentile of whole population. Ten SI are (a) environmental variance, (b) coefficient of regression, (c) genotypic stability, (d) genotypic superiority measure, (e) safety first index, (f) coefficient of determination, (g) adjusted coefficient of variation, (h) deviation mean squares, (i) ecovalence and (j) variance of rank (colour figure online)

showed low correlations ($R^2 < 0.5$) with mean yield and were not suitable for selecting high and stable yield at the same time (Fig. 1f–j). Low correlation between $W_{i,yield}$ and genotypic mean yield (Fig. 1i, $R^2 < 0.08$) was expected due to the orthogonal relationship between genotypic mean yield and the effect of interaction of genotype by environment (Mohammadi and Amri 2008). Another SI, $S_{i4,yield}$, was highly correlated with $W_{i,yield}$ and expected to have also low correlation (Fig. 1j $R^2 < 0.03$) to genotypic mean yield, as reported in the literature (Piepho and Lotito 1992).

Sampling method affected the correlation between SI and mean yield (Fig. 1) and the shape of SI distribution (Supplementary Fig. S1). In general, the ranking of R^2 between mean yield and four SI was the highest in “even” selection method, followed by “random” selection and the lowest in “top 20” selection method. Taking $b_{i,yield}$

(Fig. 1c) as example, the effect of sampling method on correlation between mean yield and $b_{i,yield}$ was the highest in “even” ($R^2 = 0.87$), followed by “random” ($R^2 = 0.76$) and the lowest in “top 20” ($R^2 = 0.42$). Among the 11 studied SI, the linear correlation between $P_{i,yield}$ and mean yield was least affected by the sampling methods (Fig. 1d), with R^2 of 0.98, 0.95 and 0.77 in methods “even,” “random” and “top 20”, respectively. This is also a reason to use P_i as the representative stability index in the following analysis in this study.

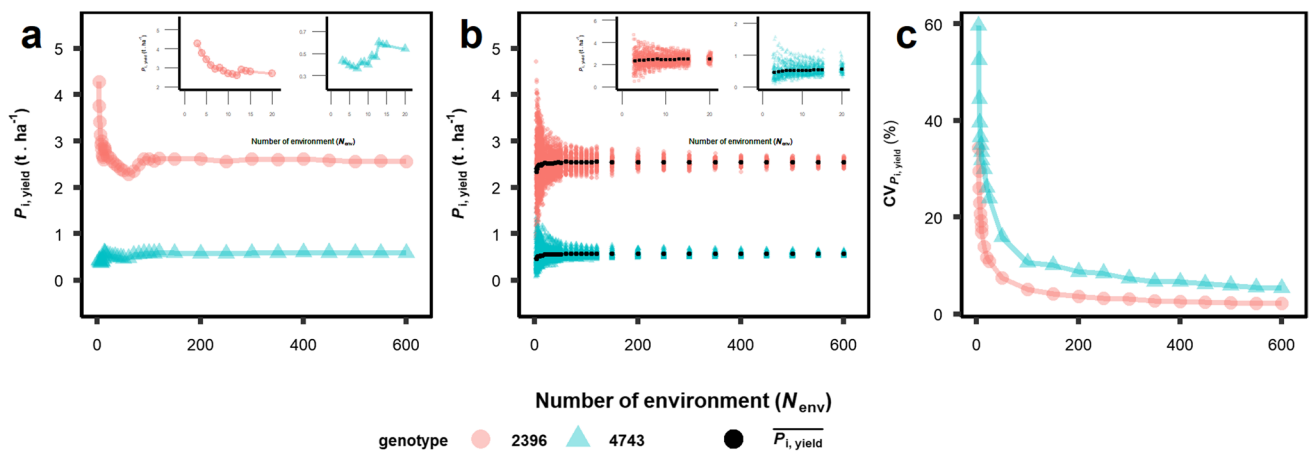


Fig. 2 Relationships between $P_{i,yield}$ and environment number (N_{env} , sampled with 100 virtual genotypes). Two genotypes with contrasting $P_{i,yield}$ values were chosen: red circle for unstable genotype 2396 and blue triangle for stable genotype 4743. **a** Results from single sam-

pled population of environment (SPE = 1). **b** Results from SPE = 100. Black circles stand for the averaged $P_{i,yield}$ value from 100 SPE. **c** Coefficient of variation of $P_{i,yield}$ ($CV_{P_{i,yield}}$) between 100 SPE (colour figure online)

More than 150 environments are required to estimate genotypic yield stability robustly using 100 genotypes

To test the minimum N_{env} required for robust estimation of $P_{i,yield}$ of a genotype, 100 random virtual genotypes ($N_{gen} = 100$, SPG = 1) created by APSIM-Wheat were first selected by “random” method, then their yields were simulated from 3 to 600 random environments ($N_{env} = 3-600$). The selections of environments were repeated 100 times (sampled population of environments, SPE = 100), and $P_{i,yield}$ of genotypes in each SPE was calculated. Within one SPE (Fig. 2a), the range of $P_{i,yield}$ of an unstable genotype (represented by genotype 2396) between different N_{env} varied from 2.28 to 4.28 t ha⁻¹, with coefficient of variation (CV) of 14.1%. In comparison with genotype 2396, a stable genotype (represented by genotype number 4743) under the same SPE had similar CV of $P_{i,yield}$ (14.5%) but a smaller range of $P_{i,yield}$ (from 0.36 to 0.6 t ha⁻¹). Irregular variations in $P_{i,yield}$ in both genotypes (subfigures in Fig. 2a) indicated strong effects of SPE on the estimation of $P_{i,yield}$. Using 100 SPE, the potential bias of estimated $P_{i,yield}$ from the sampling of environments were quantified (Fig. 2b). $P_{i,yield}$ estimated from 100 SPE with three environments ($N_{env} = 3$) varied largely in stable and unstable genotypes (0.08–1.56 and 0.47–4.73 t ha⁻¹, respectively), indicating unreliable estimation of $P_{i,yield}$ at low N_{env} . Standard deviation of $P_{i,yield}$ between 100 SPE decreased with the increase of N_{env} , while the mean of $P_{i,yield}$ from 100 SPE increased with N_{env} slightly in an asymptotic manner.

Based on the results shown in Fig. 2b, coefficient of variation ($CV_{P_{i,yield}}$) between SPE was calculated for each N_{env} (Fig. 2c). The minimal N_{env} was defined as N_{env} at which

$CV_{P_{i,yield}}$ became lower than the predefined thresholds (5 or 10%). Stable genotype 4743 was found with larger $CV_{P_{i,yield}}$ than unstable genotype 2396. The relationship between $CV_{P_{i,yield}}$ and N_{env} was fitted by an exponential function $CV_{P_{i,yield}} = \alpha * N_{env}^{\beta}$, where α adjusts the range of $CV_{P_{i,yield}}$ and β controls the curvature. For a stable genotype 4743, $CV_{P_{i,yield}} = 0.93 * N_{env}^{-0.44}$ ($R^2 = 0.993$, p -value < 0.001, $se_{\alpha} = 0.06$, $se_{\beta} = 0.012$). For an unstable genotype 2396, $CV_{P_{i,yield}} = 0.52 * N_{env}^{-0.50}$ ($R^2 = 0.999$, p -value < 0.001, $se_{\alpha} = 0.03$, $se_{\beta} = 0.004$). According to these equations, at least 151 and 28 environments were required to reach the threshold of $CV_{P_{i,yield}} = 10\%$ for stable and unstable genotypes, respectively. If the threshold = 5%, minimal N_{env} for stable and unstable genotypes is 718 and 111, respectively. This suggested that minimal N_{env} for robust estimation of $P_{i,yield}$ is genotype and threshold dependent. Under the threshold of 10% $CV_{P_{i,yield}}$, more than 150 environments are required. We expanded this analysis to three different sampling methods and 11 SI (Supplementary Fig. S3). In general, the choice of SI, rather than the sampling method, determined the minimal required N_{env} for robust estimation of stability.

More than 1000 genotypes are required to establish robust correlations between physiological parameters and yield stability

To illustrate the relationship between physiological parameters, yield and $P_{i,yield}$ (nodes), Pearson correlation coefficient, r (edges), between nodes were visualized by connecting the nodes with edges in a r -network (Supplementary Fig. S2c). Since the robustness of $P_{i,yield}$ is related to number of environments in an exponential manner (Fig. 2c), 9000

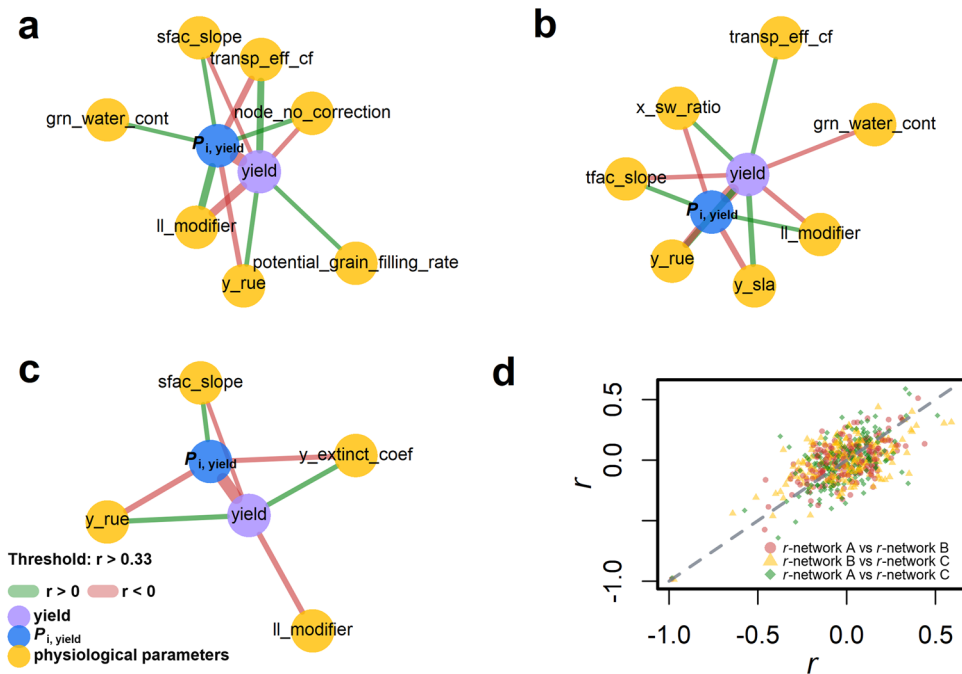


Fig. 3 Three r -network of yield, $P_{i,yield}$ and 90 physiological parameters from three different sampled populations of genotype (**a–c**, $N_{gen}=100$) grown under 9000 environments ($SPE=1$). Nodes (circle) represent mean yield (purple), $P_{i,yield}$ (blue) or physiological parameters (yellow). Correlation between two nodes presents as a line (edge), which's width is proportional to the absolute value of Pearson correlation coefficient ($|r|$). Green and red lines mark positive and negative r value, respectively. Nodes are shown with connected edge when $|r| > 0.33$. **d** Comparison of r -vector (Supplementary Fig. S2c) from **a–c**. Colour and shape stand for the combination of networks. Grey dashed line stands for the 1:1 line. Description of physiological parameters: water content of grain (*grn_water_cont*), number of

grains that are set depending on the stem dry weight (*grain_per_gram_stem*), efficiency of roots to extract soil water (*ll_modifier*), rate of node senescence on main stem (*node_sen_rate*), potential rate of grain growth at grain filling (*potential_grain_filling_rate*), sensitivity to photoperiod (*photop_sens*), soil water effect on biomass accumulation (*sfac_slope*), transpiration efficient (*transp_eff_cf*), temperature effect on biomass accumulation (*tfac_slope*), water availability affecting the stress factor for root depth growth (*x_sw_ratio*), fraction of dry matter allocated to rachis for specific stages (*y_frac_leaf*), radiation use efficiency (*y_rue*), potential leaf specific area (*y_sla*), Extinction coefficient of green leaves as a response to row spacing (*y_extinct_coef*) (colour figure online)

environments ($N_{env}=9000$, $SPE=1$) were used to ensure the robustness of $P_{i,yield}$ estimation. From the 9100 virtual genotypes, we first selected 100 virtual genotypes for three times ($N_{gen}=100$, $SPG=3$) using “random” sampling method to demonstrate the effects of genotype selection on the robustness of the edge in the r -network (Fig. 3).

Between three randomly selected SPG with “random” method, number and width (positively correlated to $|r|$) of edges and type of nodes varied between r -networks with a threshold of $|r| > 0.33$ for displaying the edges (Fig. 3a–c). Among these three r -networks, two physiological parameters related to efficiency of roots to extract soil water, linked to plant water status (*ll_modifier*, mean r to $P_{i,yield}=0.41 \pm 0.15$, mean r to yield = -0.49 ± 0.13), potential radiation use efficiency for biomass production (*y_rue*, mean r to $P_{i,yield}=-0.48 \pm 0.08$, mean r to yield = 0.43 ± 0.08) show medium correlation to yield and $P_{i,yield}$ (Supplementary Table S1). Interestingly, while $P_{i,yield}$ and yield were highly negatively correlated ($r=-0.98$, -0.98 and -0.97 in r -network a-c), a physiological parameter might tend to explain

yield better than $P_{i,yield}$, and vice versa. For example, the mean tendency of *ll_modifier* ($T_{ll_modifier}$) was 0.83 ± 0.11 ($SPG=3$), indicating *ll_modifier* is more correlated to yield. In contrast, the mean T_{y_rue} was 1.13 ± 0.01 , indicating *y_rue* is more correlated to $P_{i,yield}$.

To investigate the response of $T_{parameter}$ to three sampling methods (“random”, “even” and “top 20”), 100 virtual genotypes ($N_{gen}=100$) were selected for 100 times ($SPG=100$) and grown under 9000 environments ($N_{env}=9000$, $SPE=1$). Six physiological parameters having the highest $|r|$ with yield and with $P_{i,yield}$ and with varying $T_{parameter}$ were identified (Fig. 4): potential radiation use efficiency for biomass production (*y_rue*, g MJ⁻¹); potential leaf specific area (*y_sla*, unitless), which determines the final leaf size; efficiency of roots to extract soil water, linked to plant water status (*ll_modifier*, unitless); potential grain growth rate at grain filling (*potential_grain_filling_rate*, g day⁻¹); number of growing leaves in the sheath (*node_no_correction*, leaf), and temperature effect on biomass accumulation (*tfac_slope*, unitless). Interestingly, the range of $|r|$ between

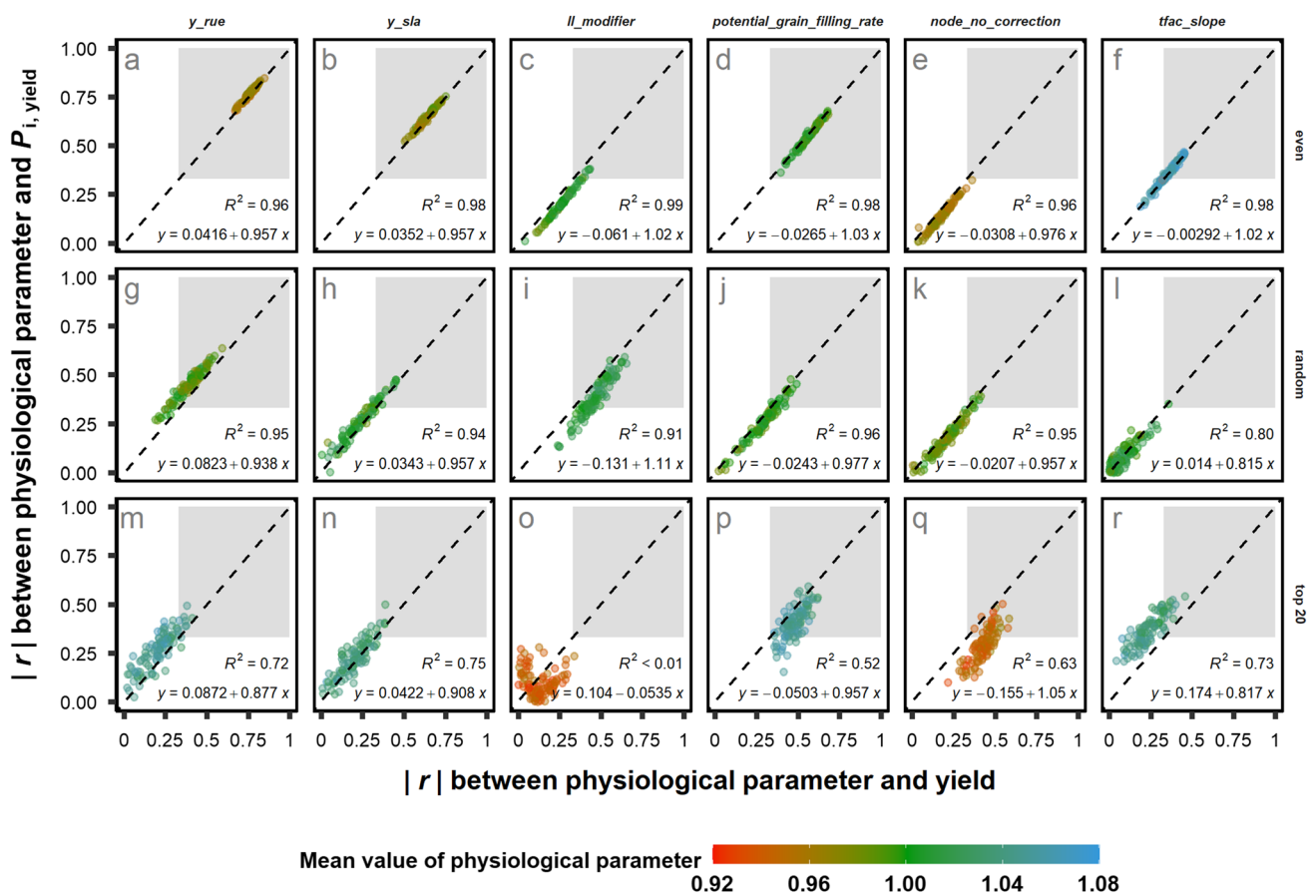


Fig. 4 Scatterplots of $|r|$ of six physiological parameters to $P_{i,yield}$ versus $|r|$ to yield from three selection methods (“random”, “even” and “top 20”) of 100 genotypes ($N_{gen}=100$, SPG=100) from 9000 environments ($N_{env}=9000$, SPE=1). Each point represents a SPG and colour of the point represents the mean values (normalized to one by the parameter values of calibrated cultivar *Hartog*) of physiological parameters from 100 genotypes in one SPG. The “even” sampling method divides population into ten groups based on mean yield value, with ten virtual genotypes randomly selected in each yield group. The “top 20” method randomly selects virtual genotypes with mean

yield above 80th percentile of the whole population. Grey rectangle area stands for the region of $|r|>0.33$. Black dashed line stands for 1:1 line. Meaning of the physiological parameters (in *italic*) follow the order (left to right) in the figure: radiation use efficiency (*y_rue*), potential leaf specific area (*y_sla*), efficiency of roots to extract soil water (*ll_modifier*), potential grain growth rate at grain filling (*potential_grain_filling_rate*), number of growing leaves in the sheath (*node_no_correction*) and temperature effect on biomass accumulation (*tfac_slope*)

physiological parameters and yield (or $P_{i,yield}$) depends on the sampling method. For example, $|r|$ between *y_rue* and $P_{i,yield}$ was the highest (0.77 ± 0.03) from method “even”, followed by “random” (0.45 ± 0.08) and “top 20” (0.27 ± 0.1). In “top 20”, *node_no_correction*, showed the highest $|r|$ with $P_{i,yield}$ (0.29 ± 0.09), followed by “random” (0.18 ± 0.09) and “even” (0.15 ± 0.06). This indicates the effects of sampling method on the explanatory power of a physiological parameter for yield and $P_{i,yield}$.

In contrast to a “random” and “even” SPG, the most yield and $P_{i,yield}$ relevant parameter was not *y_rue* in a “top 20” SPG, but the parameters controlling grain filling (*potential_grain_filling_rate*), leaf expansion (*node_no_correction*) and temperature effect on biomass accumulation (*tfac_slope*). Among these six parameters, *ll_modifier*,

node_no_correction and *potential_grain_filling* explained yield better than $P_{i,yield}$. By contrast, *tfac_slope*, *y_rue* and *y_sla* explain $P_{i,yield}$ better than yield (Table 1). The effects of sampling methods on $T_{parameter}$ suggested that the importance of a target trait for yield or $P_{i,yield}$ depends on the shape of phenotype distribution in a SPG.

Relationships between physiological parameters and yield stability depend on the sampled population of genotypes but not affected by the population size of sampled environments

Since r -networks (Fig. 3) depend on the sampled genotypes in a relatively small population ($N_{gen}=100$), proper N_{gen} and N_{env} required for a robust estimation of r -networks were

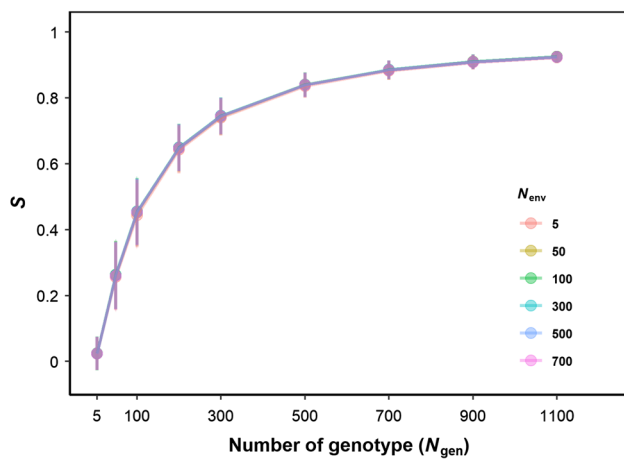


Fig. 5 Impact of genotype number ($N_{\text{gen}}=5, 100, 300, 500, 700, 900$ and 1100 , $\text{SPG}=100$) and environment number ($N_{\text{env}}=5, 50, 100, 300, 500$ and 700 , $\text{SPE}=1$) on similarity (S) between r -networks

tested using 100 SPG in combination with nine genotype numbers ($N_{\text{gen}}=5, 50, 100, 200, 300, 500, 700, 900$ and 1100) and six environment numbers ($N_{\text{env}}=5, 50, 100, 300, 500$ and 700). The similarity between networks (S) increased with N_{gen} but not with N_{env} (Fig. 5). Hence, S was fitted with N_{gen} using an asymptotic function $S = N_{\text{gen}} / (k + N_{\text{gen}})$ with $k = 111.71 \pm 5.43$. Using the asymptotic function, S reached 0.90 with $N_{\text{gen}} = 1006$.

Physiological network of multi-traits and their stability from random selected population

Since high number of genotypes should be used to obtain robust r -network (Fig. 5), we conducted network analysis for physiological parameters and eight model outputs (yield, straw yield, grain protein, grain number, grain size, flowering time, maturity time and LAI) and $P_{i,\text{Trait}}$ with $N_{\text{gen}} = 1000$ ($\text{SPG} = 1$) and $N_{\text{env}} = 9000$ ($\text{SPE} = 1$). Among all traits, grain number correlated most positively with yield ($r = 0.71$), followed by straw ($r = 0.64$) and LAI ($r = 0.60$). By contrast, grain protein correlated most negatively with yield ($r = -0.83$, Fig. 6a). In the same vein, $P_{i,\text{grain_number}}$ correlated most positively with $P_{i,\text{yield}}$ ($r = 0.78$), followed by $P_{i,\text{straw}}$ ($r = 0.69$) and $P_{i,\text{LAI}}$ ($r = 0.65$) and $P_{i,\text{grain_protein}}$ negatively correlated with $P_{i,\text{yield}}$ ($r = -0.82$). In general, correlations between $P_{i,\text{Trait}}$ were slightly higher than that between traits (Fig. 6 and Supplementary Table S2).

Grain yield was mostly explained by efficiency of roots to extract soil water ($ll_modifier$ to yield, $r = -0.44$, $T_{ll_modifier} = 0.81$), while the variations in $P_{i,\text{yield}}$ were best explained by radiation use efficiency (y_rue to $P_{i,\text{yield}}$, $r = 0.36$, $T_{y_rue} = 1.15$). Thermal time between plant emergence and end of juvenile stage ($tt_end_of_juvenile$)

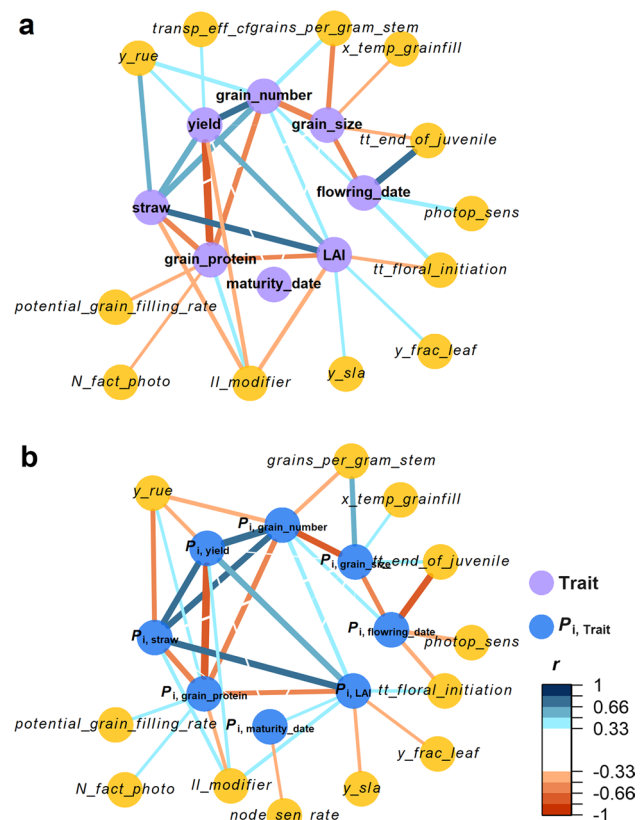


Fig. 6 r -network between 90 physiological parameters, $P_{i,\text{trait}}$ (a) and traits (b) of randomly selected 1000 virtual genotypes ($\text{SPG} = 1$) grown under 9000 environments ($\text{SPE} = 1$). Node (circle) represents mean trait (purple), $P_{i,\text{yield}}$ (blue) or physiological parameters (yellow). Description of physiological parameters: number of grains that are set depending on the stem dry weight ($grain_per_gram_stem$), efficiency of roots to extract soil water ($ll_modifier$), rate of node senescence on main stem ($node_sen_rate$), multiplier for nitrogen deficit effect on photosynthesis (N_fact_photo), potential rate of grain growth at grain filling ($potential_grain_filling_rate$), sensitivity to photoperiod ($photop_sens$), transpiration efficient ($transp_eff_cf$), thermal time between plant emergence and end of juvenile stage ($tt_end_of_juvenile$), thermal time between end of juvenile and floral initiation ($tt_floral_initiation$), temperature affecting stress index for the potential grain filling rate ($x_temp_grain_fill$), fraction of dry matter allocated to rachis for specific stages (y_frac_leaf), radiation use efficiency (y_rue), potential leaf specific area (y_sla) (colour figure online)

most correlated to flowering time and explained this trait and its stability equally ($tt_end_of_juvenile$ to flowering_date, $r = 0.73$, $T_{tt_end_of_juvenile} = 1$). Meanwhile, this physiological parameter also correlated with grain weight ($tt_end_of_juvenile$ to grain_size, $r = -0.37$, $T_{tt_end_of_juvenile} = 0.98$). LAI negatively correlated with thermal time before floral initiation ($tt_floral_initiation$ to LAI, $r = -0.39$, $T_{tt_floral_initiation} = 0.95$) and efficiency of roots to extract soil water ($ll_modifier$ to LAI, $r = -0.46$, $T_{ll_modifier} = 0.87$), while positively correlated with fraction of dry matter allocated to rachis for specific stage

(y_frac_leaf to LAI, $r=0.34$, $T_{y_frac_leaf}=1.05$) and potential leaf specific area (y_sla to LAI, $r=0.34$, $T_{y_sla}=1.07$). The edge of r -networks in Fig. 6 can be found in Supplementary Table S3.

Discussion

Using R package *toolStability* as a tool for reproducible analysis

To study yield stability systematically, we developed and shared an R package *toolStability* to analyse a virtual dataset containing ~82 million simulation outputs obtained from the APSIM-Wheat crop model. Our R package *toolStability* provides more indices in comparison with other published R packages (Branco 2015; Ajay et al. 2018; Yaseen et al. 2018) and online tool platform (Pour-Aboughadareh et al. 2019). Furthermore, *toolStability* adds genotypic superiority measure ($P_{i,yield}$, Lin and Binns 1988), a stability index which was not implemented before. While the characteristic of different SI and their pairwise correlations have been studied and reviewed in the past (Fasahat 2015; Mohammadi and Amri 2008; Piepho and Lotito 1992), there is no consensus in favour of a representative SI (Reckling et al. 2021). The main reason is due to each SI has its assumption and limitation (Lin et al. 1986; Becker and Léon 1988). For example, parametric SI has the advantage of using model that is easy for implementation and interpretation, while it is poor at describing the multivariate phenotypic response to environment or having risk of misleading when the assumption is wrong. Nonparametric SI can bypass this problem of parametric methods, while reference genotype may be required to compare genotype ranking. Multivariate methods are useful in finding extreme genotypes in phenotypic stability but usually hard to interpret. Here we want to emphasize that $P_{i,yield}$ was chosen in this study because it is an index characterizing high and stable yield at the same time, regardless of the population distribution of yield and among all 11 SI (Supplementary Fig. S1). Despite concerns about P_i (Fasahat 2015; Purchase et al. 2000), P_i is still useful for field studies (Mohammadi and Amri 2008; Sehgal et al. 2017) and suitable to demonstrate our analyses.

To calculate a stability index of a genotype, a population of genotypes and environments is always required (Tollenaar and Lee 2002; Sehgal et al. 2017). Therefore, it is essential to know how many genotypes and environments are necessary for an accurate estimation of a stability index and how the composition of a sampled population affects the SI of a genotype. This question can be only answered by investigating systematically with a substantial number of

genotypes and environments, which is experimentally difficult. Crop model can fulfil this requirement by simulating large numbers of genotypes, environments and their combinations (Casadebaig et al. 2016; Senapati and Semenov 2020). It has been suggested that more than 200 environments are required if the threshold of $CV_{S2xi,yield}$ is 10% (Piepho 1998). Their estimation was based on the assumption that this sample follows the scaled chi-squared distribution (Searle et al. 2010). In comparison to our simulation result (Supplementary Fig. S3a), only less than 50 environments is needed for reaching 10% of $CV_{S2xi,yield}$ for all three sampling methods. Under random sampling method, more than 150 environments were required to obtain robust estimation of yield stability $P_{i,yield}$ of two extreme genotypes (Fig. 2), indicating that the number of genotypes and environments in the published field trials for yield stability are insufficient (Wang et al. 2015; Sehgal et al. 2017; Voss-Fels et al. 2019). Considering this, in silico approaches could be used to assist breeding programs and pinpoint candidate mechanisms to be tested in the real world.

Target traits for yield stability depend on the types of breeding program

To our knowledge, this is the first study that brings the shape of phenotype distributions (the distribution of genotypic means of a trait in a SPG) into the context of analysing yield stability (Fig. 1), including the minimal number of environment and genotype (Fig. 2, Supplementary Fig. S3) and the relationship between physiological parameter to yield and yield stability (Fig. 4). An interesting finding from our analysis is the effect of sampling methods on the relationship between the trait and $P_{i,Trait}$ (e.g. for grain yield, Fig. 1 and Fig. 4) and the r -network between them and physiological parameters (Fig. 4), suggesting the differences in target traits between breeding programs. Our sampling methods (“even”, “random” and “top 20”) present three common shapes of phenotype distributions in genetic pools used in breeding programs.

Based on the central limit theorem (Laplace 1812), when the random sampling (method ‘random’) in combination with a large population size, a trait (e.g. yield) response will follow normal distribution (Juliana et al. 2020, Fig. 1a and Supplementary Fig. S4). Heterogeneous genetic background and a wide range of trait response make the “random” population (e.g. segregation population or evolutionary population) valuable for selecting favourable traits (Dwivedi et al. 2016; Bocci et al. 2020). In our simulation results, grain yield and $P_{i,yield}$ in random population distributes normally as expected (Supplementary Fig. S4u and v). However, the distributions of the most influential physiological parameters in random populations are relatively flat (Supplementary Fig. S4g–i), implying the random combinations of evenly

distributed physiological parameters might create a normal distribution of a complex trait (i.e. grain yield, similar to the method “*even*”). In contrast, trends and peaks can be observed in the distributions of parameters in “*top 20*” populations (Supplementary Fig. S4m–r).

Our “*top 20*” method represents the elite population with a high mean yield (Longin and Reif 2014; Ovenden et al. 2017). Compared with the random population, the elite population has a narrower and more homogeneous genetic background. Elite lines from the elite population are the result of selection methods like tail selection (Rebetzke et al. 2012) or recurrent selection (Vishwakarma et al. 2014; Rembe et al. 2019). Therefore, many traits of elite population are already optimized, for example, harvest index (HI; Zhu et al. 2010), nitrogen uptake (Cormier et al. 2013), or light interception (Rose and Kage 2019). The observation of optimized traits in elite population probably explains the observed effects of sampling methods on the correlations between *ll_modifier*, yield and $P_{i,yield}$ (Fig. 4c, i and o). Among three sampling methods, “*top 20*” has the lowest mean $|r|$ ($R^2 < 0.01$, Fig. 4o), suggesting that it is the parameter which has been optimized in the “*top 20*” population (see the distribution of *ll_modifier* in Supplementary Fig. 4c). Our results further suggested that potential grain growth rate at grain filling (*potential_grain_filling_rate*, Fig. 4p), number of growing leaves in the sheath (*node_no_correction*, Fig. 4q) and temperature effect on biomass accumulation (*tfac_slope*, Fig. 4r) could be the target traits for further improving elite cultivars.

Even distribution of traits can be found at the early stage of the breeding program (Brescghello et al. 2009) or in certain environment conditions (Mathews et al. 2007; Voss-Fels et al. 2019). Our results suggested that a physiological parameter in an “*even*” or a “*random*” population explains yield and $P_{i,yield}$ more equally (T close to one) and closely (R^2 close to one) than in an elite population (Fig. 4 and Table 1). Therefore, if a breeder selects a physiological parameter for yield in “*even*” and “*random*” population, $P_{i,yield}$ is also selected, while this is not guaranteed in an elite population. Our results emphasize that the shape of phenotype distribution is an important aspect in selecting target traits for improving yield stability.

Insights from physiological networks regulating stability

In APSIM-Wheat crop model, the interactions between physiological parameters, environment and crop management on canopy development (leaf area index, LAI), flowering time, grain yield, grain size and grain number were predicted as a function of physiological assumptions of the model. The simulated dataset provides us a chance to glance at the contour of the complex physiological network and its relation to the shape of phenotype distributions (Fig. 4

and Supplementary Fig. S4). Whereas, for the complex trait like grain yield, it is difficult to decipher the genetic and physiological regulations due to pleiotropic effect of genes and the minor contribution of each quantitative trait gene (Schulthess et al. 2017; Parent et al. 2018). Our model analyses suggested that the efficiency of roots to extract soil water (*ll_modifier*) and radiation use efficiency (*y_rue*) have the highest correlations with yield and yield stability in the random population. Although it is not especially surprising the close relation between root water extractability (*ll_modifier*) and yield from an eco-physiological view (Richards et al. 2010; Thorup-Kristensen et al. 2020), it is surprising that the explanatory power of root water extractability is higher for yield than for yield stability ($T < 1$), which was similar to the parameter “*potential_grain_filling_rate*” (Fig. 4j). In contrast, the explanatory power of radiation use efficiency (*y_rue*, Fig. 4g) is higher for yield stability than for yield ($T > 1$), which was similar to the parameter “*y_sla*” (Fig. 4h). □□□ This provides the first empirical proof that, despite of high correlation between mean yield and genotypic superiority measure (Fig. 1d), genetic and physiological regulations between them can still be different, as proposed in the previous genome-wide association study on yield stability (Sehgal et al. 2017). Our results from the model analysis showed the merits of in silico approach in associating physiological parameters differentially to closely related traits like yield and genotypic superiority measure for breeding programs (Hammer et al. 2019; Cooper et al. 2021).

The network between physiological parameters, model outputs and their stability (Fig. 6) suggests following physiological mechanisms regulating yield stability. Well-known mechanisms, including the trade-off between grain yield and grain protein (Slafer et al. 2014; Asseng et al. 2019) and the trade-off between grain number and grain size (Lichtardt et al. 2020; Voss-Fels et al. 2019), can be confirmed. Although the high correlation between model outputs (e.g. grain protein content and grain yield) is not always observed in the empirical datasets (Oury et al. 2003), a R^2 of 0.6 has been reported (Lollato and Edwards 2015; Voss-Fels et al. 2019). Highly positive correlations between the stability of LAI, straw yield and grain number in the *r*-network (Fig. 6b) suggested that stable canopy development during the vegetative phase ensures sufficient pre-anthesis nitrogen reserves for grain filling and thereby yield stability. Physiologically, stable and vigorous canopy development ensures radiation interception (Tian et al. 2011) and allows storage of nitrogen and water-soluble carbohydrates in the canopy at the end of the vegetative phase (referred to as pre-anthesis nitrogen and carbon reserves, respectively).

The pre-anthesis nitrogen and carbon reserves might contribute significantly to grain filling since wheat accumulates about 70% of the total biomass and takes up about 70–100% of total nitrogen before anthesis (Barraclough et al. 2014;

Wu et al. 2016). Under optimal nitrogen supply, the pre-anthesis nitrogen reserves in stems, sheathes and leaves contribute about 30%, 15% and 40% of the nitrogen content in wheat grains, respectively (re-calculated from Fig. 3 of Barraclough et al. 2014). These results indicate the importance of pre-anthesis nitrogen reserves on grain yield. Although forty years ago, the estimated contribution of pre-anthesis carbon reserves to grain weight ranged between 11 and 17% but is higher under stress conditions (up to 22–44%) due to the lower yield level. Since genetic variation of pre-anthesis carbon reserves in wheat exists (Ehdaie et al. 2006), together with the modern wheat cultivars have higher pre-anthesis carbon reserves than the old cultivars (Xiao et al. 2012), it is worth a revisit of the contribution of pre-anthesis carbon reserve to yield in the modern cultivars.

Deriving from the data of a recent study using 20 wheat cultivars suggests that, on average, biomass accumulation before anthesis may contribute up to 38–43% of the grain yield (Barraclough et al. 2014). High contribution to grain yield from pre-anthesis reserves indicates the potential role of pre-anthesis carbon reserve as a buffer to secure the yield. In other words, yield stability could be achieved by increasing the pre-anthesis carbon reserve pool that reduces the risk of insufficient photosynthate at the grain filling stage due to abiotic stress (Slewisinski 2012). This also explains the early observation that a wheat genotype with higher biomass accumulation until anthesis, a proxy of higher pre-anthesis nitrogen and carbon reserves, has a higher yield and less yield variation between experimental years (Damisch and Wiberg 1991). Furthermore, the size of the pre-anthesis carbon reserve pool is determined by carbon fixation, namely canopy photosynthesis, during the vegetative phase, as suggested by the correlations of radiation use efficiency (y_{rue}) with $P_{i,yield}$ and $P_{i,straw}$ (Fig. 6). Our *r*-network also suggests close relationship between stable canopy development (low $P_{i,LAI}$) and stable grain number (low $P_{i,grain_number}$), probably due to the effects of canopy condition at pre-anthesis stage on floral formation (Stockman et al. 1983) or carbon and nitrogen reserves that avoid pre-anthesis abortion (Sinclair and Jamieson 2008).

Physiologically, it is noteworthy that not all traits (physiological parameters) have robust contributions to yield and yield stability and their contributions can be environment-dependent (Ferrante et al. 2017; Slafer et al. 2022). However, there are also traits (e.g. reproductive, phenological, photosynthetic and architectural traits) delivering stable and positive effects to yield formation and their contributions to yield are less environment-dependent (Welcker et al. 2022). To our opinion, these can be the traits showing significance within the network of yield and yield stability (Fig. 6; e.g. grain number, photoperiodic sensitivity and radiation use efficiency), as shown in the experimental findings in wheat (Voss-Fels et al. 2019;

Lichthardt et al. 2020) and in maize (Welcker et al. 2022) that these traits with stable effects on yield have been indirectly preferred under breeders' selections. Welcker et al. (2022) also clearly showed that physiological traits with different effects on yield between environments are phenotypically unchanged by selection. Therefore, we could speculate that the parameters showing importance in Fig. 6 are the parameters delivering stable effects on yield and can be the first target for breeders.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-023-04264-7>.

Acknowledgements The authors gratefully acknowledge the technical support by Dr. Katrin Leinweber for her help in developing the *R* package *toolStability* and setting up with GitHub. We also thank Magnus Alder from Leibniz Universität Hannover for maintaining the server for simulation. The study was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 419973621).

Author contribution statement TWC led the project; PC simulated the APSIM-Wheat dataset; TWC and TCW contributed to the development of analysis pipeline and wrote the paper; TCW developed the *R* package and conducted the analysis; all authors contributed to the editing of the paper.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 419973621).

Data availability All data supporting the findings of this study are available within the paper and within its supplementary data published online. An *R* package *toolStability* was published on CRAN (<https://cran.r-project.org/web/packages/toolStability/index.html>) and Zenodo (<https://doi.org/10.5281/zenodo.5804212>). An APSIM-Wheat dataset is available on Zenodo (<https://doi.org/10.5281/zenodo.4729636>). A repository for reproducing the figures in this publication is available on GitHub (https://github.com/Illustration/Wang_2023_TAAG) and Zenodo (<https://doi.org/10.5281/zenodo.7562420>).

Declarations

Conflict of interest The authors declare that they have no conflict of interest. The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ajay BC, Aravind J, Abdul R (2018) Ammistability: additive main effects and multiplicative interaction model stability parameters. <https://cran.r-project.org/package=ammistability>
- Asseng S, Martre P, Maiorano A et al (2019) Climate change impact and adaptation for wheat protein. *Glob Change Biol* 25(1):155–173. <https://doi.org/10.1111/gcb.14481>
- Barillot R, Escobar-Gutiérrez AJ, Fournier C et al (2014) Assessing the effects of architectural variations on light partitioning within virtual wheat-pea mixtures. *Ann Bot* 114(4):725–737. <https://doi.org/10.1093/aob/mcu099>
- Barraclough PB, Lopez-Bellido R, Hawkesford MJ (2014) Genotypic variation in the uptake, partitioning and remobilisation of nitrogen during grain-filling in wheat. *Field Crop Res* 156:242–248. <https://doi.org/10.1016/j.fcr.2013.10.004>
- Becker HC, León J (1988) Stability analysis in plant breeding. *Plant Breed* 101(1):1–23. <https://doi.org/10.1111/j.1439-0523.1988.tb00261.x>
- Bocci R, Bussi B, Petitti M et al (2020) Yield, yield stability and farmers' preferences of evolutionary populations of bread wheat: a dynamic solution to climate change. *Eur J Agron* 121:126–156. <https://doi.org/10.1016/j.eja.2020.126156>
- Bolaños J, Edmeades GO (1993) Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. *Field Crops Res* 31(3–4):233–252. [https://doi.org/10.1016/0378-4290\(93\)90064-T](https://doi.org/10.1016/0378-4290(93)90064-T)
- Branco LC (2015) Phenability: nonparametric stability analysis. <https://cran.r-project.org/package=phenability>
- Breseghello F, Morais OP, Castro EM et al. (2009) Recurrent selection resulted in rapid genetic gain for upland rice in Brazil. *International Rice Research Notes* 34. <https://doi.org/10.3860/irrn.v34i0.1069>
- Casadebaig P, Zheng B, Chapman S et al (2016) Assessment of the potential impacts of wheat plant traits across environments by combining crop modeling and global sensitivity analysis. *PLOS ONE* 11(1):e0146385. <https://doi.org/10.1371/journal.pone.0146385>
- Chen T-W, Nguyen TMN, Kahlen K, Stützel H (2015) High temperature and vapor pressure deficit aggravate architectural effects but ameliorate non-architectural effects of salinity on dry mass production of tomato. *Front Plant Sci* 6. <https://doi.org/10.3389/fpls.2015.00887>
- Chenu K, Cooper M, Hammer GL et al (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. *J Exp Bot* 62(6):1743–1755. <https://doi.org/10.1093/jxb/erq459>
- Cooper M, Powell O, Voss-Fels KP et al (2021) Modelling selection response in plant-breeding programs using crop models as mechanistic gene-to-phenotype (CGM-G2P) multi-trait link functions. in silico Plants. <https://doi.org/10.1093/insilicoplants/diaa016>
- Cormier F, Faure S, Dubreuil P et al (2013) A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Theor Appl Genet* 126(12):3035–3048. <https://doi.org/10.1007/s00122-013-2191-9>
- Damisch W, Wiberg A (1991) Biomass yield — a topical issue in modern wheat breeding programmes. *Plant Breed* 107(1):11–17. <https://doi.org/10.1111/j.1439-0523.1991.tb00523.x>
- Dwivedi SL, Ceccarelli S, Blair MW et al (2016) Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci* 21(1):31–42. <https://doi.org/10.1016/j.tplants.2015.10.012>
- Ehdaie B, Alloush GA, Madore MA, Waines JG (2006) Genotypic variation for stem reserves and mobilization in wheat: I. Postanthesis changes in internode dry matter. *Crop Sci* 46(2):735–746. <https://doi.org/10.2135/cropsci2005.04-0033>
- Eskridge KM (1990) Selection of stable cultivars using a safety-first rule. *Crop Sci* 30(2):369. <https://doi.org/10.2135/cropsci1990.0011183X003000020025x>
- Fasahat P (2015) An overview on the use of stability parameters in plant breeding. *BBIJ*. <https://doi.org/10.15406/bbij.2015.02.00043>
- Ferrante A, Cartelle J, Savin R, Slafer GA (2017) Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. *Field Crop Res* 203:114–127. <https://doi.org/10.1016/j.fcr.2016.12.028>
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. *Aust J Agric Res* 14(6):742–754. <https://doi.org/10.1071/AR9630742>
- Hammer G, Messina C, Wu A, Cooper M (2019) Biological reality and parsimony in crop models—why we need both in crop improvement! in silico Plants. <https://doi.org/10.1093/insilicoplants/diz010>
- Hanson WD (1970) Genotypic stability. *Theor Appl Genet* 40(5):226–231. <https://doi.org/10.1007/BF00285245>
- Juliana P, Singh RP, Braun H-J et al (2020) Genomic selection for grain yield in the CIMMYT wheat breeding program—status and perspectives. *Front Plant Sci* 11:1418. <https://doi.org/10.3389/fpls.2020.564183>
- Kouadio L, Newlands N, Potgieter A et al (2015) Exploring the potential impacts of climate variability on spring wheat yield with the APSIM decision support tool. *Agric Sci* 06(07):686–698. <https://doi.org/10.4236/as.2015.67066>
- Laplace P-S (1812) *Théorie analytique des probabilités*. Courcier
- Leakey ADB, Ferguson JN, Pignou CP et al (2019) Water use efficiency as a constraint and target for improving the resilience and productivity of C3 and C4 crops. *Annu Rev Plant Biol* 70:781–808. <https://doi.org/10.1146/annurev-arplant-042817-040305>
- Lichthardt C, Chen T-W, Stahl A, Stützel H (2020) Co-evolution of sink and source in the recent breeding history of winter wheat in Germany. *Front Plant Sci* 10:1771. <https://doi.org/10.3389/fpls.2019.01771>
- Lin CS, Binns MR (1988) A superiority measure of cultivar performance for cultivar × location data. *Can J Plant Sci* 68(1):193–198. <https://doi.org/10.4141/cjps88-018>
- Lin CS, Binns MR, Lefkovich LP (1986) Stability analysis: where do we stand? *Crop Sci* 26(5):894–900. <https://doi.org/10.2135/cropsci1986.0011183X002600050012x>
- Lollato RP, Edwards JT (2015) Maximum attainable wheat yield and resource-use efficiency in the southern great plains. *Crop Sci* 55(6):2863–2876. <https://doi.org/10.2135/cropsci2015.04.0215>
- Longin CFH, Reif JC (2014) Redesigning the exploitation of wheat genetic resources. *Trends Plant Sci* 19(10):631–636. <https://doi.org/10.1016/j.tplants.2014.06.012>
- Macholdt J, Honermeier B (2017) Yield stability in winter wheat production: a survey on German farmers' and advisors' views. *Agronomy*. <https://doi.org/10.3390/agronomy7030045>
- Mathews KL, Chapman SC, Trethowan R et al (2007) Global adaptation patterns of Australian and CIMMYT spring bread wheat. *Theor Appl Genet* 115(6):819–835. <https://doi.org/10.1007/s00122-007-0611-4>
- Mohammadi R, Amri A (2008) Comparison of parametric and non-parametric methods for selecting stable and adapted durum wheat genotypes in variable environments. *Euphytica* 159(3):419–432. <https://doi.org/10.1007/s10681-007-9600-6>
- Nassar R, Hühn M (1987) Studies on estimation of phenotypic stability: tests of significance for nonparametric measures of

- phenotypic stability. *Biometrics* 43(1):45–53. <https://doi.org/10.2307/2531947>
- Oury FX, Bérard P, Brancourt-Hulmel M et al (2003) Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. *J Genet Breed* 57:59–68
- Ovenden B, Milgate A, Lisle C et al (2017) Selection for water-soluble carbohydrate accumulation and investigation of genetic \times environment interactions in an elite wheat breeding population. *Theor Appl Genet* 130(11):2445–2461. <https://doi.org/10.1007/s00122-017-2969-2>
- Parent B, Leclere M, Lacube S et al (2018) Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time. *Proc Natl Acad Sci USA* 115(42):10642–10647. <https://doi.org/10.1073/pnas.1720716115>
- Pedro A, Savin R, Habash DZ, Slafer GA (2011) Physiological attributes associated with yield and stability in selected lines of a durum wheat population. *Euphytica* 180(2):195–208. <https://doi.org/10.1007/s10681-011-0352-y>
- Perez RPA, Dauzat J, Pallas B et al (2018) Designing oil palm architectural ideotypes for optimal light interception and carbon assimilation through a sensitivity analysis of leaf traits. *Ann Bot* 121(5):909–926. <https://doi.org/10.1093/aob/mcx161>
- Pfeiffer WH, Sayre KD, Reynolds MP, Payne TS (2001) Increasing yield potential and yield stability in durum wheat. In: Bedő Z, Láng L (eds) *Wheat in a global environment. Proceedings of the 6th international wheat conference, 5–9 June 2000, Budapest, Hungary*. Springer Netherlands, Dordrecht, pp 569–577
- Piepho H-P (1998) Methods for comparing the yield stability of cropping systems. *J Agron Crop Sci* 180(4):193–213. <https://doi.org/10.1111/j.1439-037X.1998.tb00526.x>
- Piepho H-P, Lotito S (1992) Rank correlation among parametric and nonparametric measures of phenotypic stability. *Euphytica* 64:221–225. <https://doi.org/10.1007/BF00046052>
- Pinthus MJ (1973) Estimate of genotypic value: a proposed method. *Euphytica* 22(1):121–123. <https://doi.org/10.1007/BF00021563>
- Pour-Aboughadareh A, Yousefian M, Moradkhani H et al (2019) STABILITYSOFT: a new online program to calculate parametric and non-parametric stability statistics for crop traits. *Appl Plant Sci* 7(1):e01211–e01211. <https://doi.org/10.1002/aps3.1211>
- Powell JP, Rutten M (2013) Convergence of European wheat yields. *Renew Sustain Energy Rev* 28:53–70. <https://doi.org/10.1016/j.rser.2013.07.048>
- Powell N, Ji X, Ravash R et al (2012) Yield stability for cereals in a changing climate. *Funct Plant Biol* 39(7):539–552. <https://doi.org/10.1071/FP12078>
- Purchase JL, Hatting H, van Deventer CS (2000) Genotype \times environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa: II. Stability analysis of yield performance. *S Afr J Plant Soil* 17(3):101–107. <https://doi.org/10.1080/02571862.2000.10634878>
- Quilot-Turion B, Ould-Sidi M-M, Kadrani A et al (2012) Optimization of parameters of the ‘Virtual Fruit’ model to design peach genotype for sustainable production systems. *Eur J Agron* 42:34–48. <https://doi.org/10.1016/j.eja.2011.11.008>
- R Core Team (2020) R: a language and environment for statistical computing. <https://www.r-project.org/>
- Rebetzke GJ, Chenu K, Biddulph B et al (2012) A multisite managed environment facility for targeted trait and germplasm phenotyping. *Funct Plant Biol* 40(1):1–13. <https://doi.org/10.1071/FP12180>
- Reckling M, Ahrends H, Chen T-W et al (2021) Methods of yield stability analysis in long-term field experiments. *A Rev Agron Sustain Dev* 41(2):27. <https://doi.org/10.1007/s13593-021-00681-4>
- Reckling M, Döring TF, Bergkvist G et al (2018) Grain legume yields are as stable as other spring crops in long-term experiments across northern Europe. *Agron Sustain Dev* 38(6):63. <https://doi.org/10.1007/s13593-018-0541-3>
- Rembe M, Zhao Y, Jiang Y, Reif JC (2019) Reciprocal recurrent genomic selection: an attractive tool to leverage hybrid wheat breeding. *Theor Appl Genet* 132(3):687–698. <https://doi.org/10.1007/s00122-018-3244-x>
- Richards RA, Rebetzke GJ, Watt M et al (2010) Breeding for improved water productivity in temperate cereals: phenotyping, quantitative trait loci, markers and the selection environment. *Funct Plant Biol* 37(2):85–97. <https://doi.org/10.1071/FP09219>
- Römer T (1917) Sind die ertragreichen Sorten ertragssicherer? *Mitteilungen Der Deutschen Landwirtschaftlichen Gesellschaft* 32(1):87–89
- Rose T, Kage H (2019) The contribution of functional traits to the breeding progress of Central-European winter wheat under differing crop management intensities. *Front Plant Sci* 10:1521. <https://doi.org/10.3389/fpls.2019.01521>
- Schulthess AW, Reif JC, Ling J et al (2017) The roles of pleiotropy and close linkage as revealed by association mapping of yield and correlated traits of wheat (*Triticum aestivum* L.). *J Exp Bot* 68(15):4089–4101. <https://doi.org/10.1093/jxb/erx214>
- Searle SR, Casella G, McCulloch CE (2010) *Variance components*. Wiley, New York
- Sehgal D, Autrique E, Singh R et al (2017) Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci Rep* 7(1):41578. <https://doi.org/10.1038/srep41578>
- Senapati N, Semenov MA (2019) Assessing yield gap in high productive countries by designing wheat ideotypes. *Sci Rep* 9(1):5516. <https://doi.org/10.1038/s41598-019-40981-0>
- Senapati N, Semenov MA (2020) Large genetic yield potential and genetic yield gap estimated for wheat in Europe. *Glob Food Secur* 24:100340. <https://doi.org/10.1016/j.gfs.2019.100340>
- Shukla GK (1972) Some statistical aspects of partitioning genotype environmental components of variability. *Heredity* 29(2):237–245
- Sinclair TR, Jamieson PD (2008) Yield and grain number of wheat: A correlation or causal relationship?: Authors’ response to “The importance of grain or kernel number in wheat: A reply to Sinclair and Jamieson” by R.A. Fischer. *Field Crops Res* 105(1):22–26. <https://doi.org/10.1016/j.fcr.2007.07.003>
- Slafer GA, Savin R, Sadras VO (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crop Res* 157:71–83. <https://doi.org/10.1016/j.fcr.2013.12.004>
- Slafer GA, García GA, Serrago RA, Miralles DJ (2022) Physiological drivers of responses of grains per m² to environmental and genetic factors in wheat. *Field Crops Res* 285:108593. <https://doi.org/10.1016/j.fcr.2022.108593>
- Slewinski TL (2012) Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. *J Exp Bot* 63(13):4647–4670. <https://doi.org/10.1093/jxb/ers124>
- Stockman YM, Fischer RA, Brittain EG (1983) Assimilate supply and floret development within the spike of wheat (*Triticum aestivum* L.). *Funct Plant Biol* 10(6):585–594. <https://doi.org/10.1071/PP9830585>
- Sun H, Zhang X, Wang E et al (2016) Assessing the contribution of weather and management to the annual yield variation of summer maize using APSIM in the North China Plain. *Field Crop Res* 194:94–102. <https://doi.org/10.1016/j.fcr.2016.05.007>
- Thorup-Kristensen K, Halberg N, Nicolaisen M et al (2020) Digging deeper for agricultural resources, the value of deep rooting. *Trends Plant Sci* 25(4):406–417. <https://doi.org/10.1016/j.tplans.2019.12.007>

- Tian Z, Jing Q, Dai T et al (2011) Effects of genetic improvements on grain yield and agronomic traits of winter wheat in the Yangtze River Basin of China. *Field Crop Res* 124(3):417–425. <https://doi.org/10.1016/j.fcr.2011.07.012>
- Tollenaar M, Lee EA (2002) Yield potential, yield stability and stress tolerance in maize. *Field Crop Res* 75(2):161–169. [https://doi.org/10.1016/S0378-4290\(02\)00024-2](https://doi.org/10.1016/S0378-4290(02)00024-2)
- van Frank G, Rivière P, Pin S et al (2020) Genetic diversity and stability of performance of wheat population varieties developed by participatory breeding. *Sustainability* 12(1):384. <https://doi.org/10.3390/su12010384>
- Vishwakarma MK, Mishra VK, Gupta PK et al (2014) Introgression of the high grain protein gene Gpc-B1 in an elite wheat variety of Indo-Gangetic Plains through marker assisted backcross breeding. *Curr Plant Biol* 1:60–67. <https://doi.org/10.1016/j.cpb.2014.09.003>
- Voss-Fels KP, Stahl A, Wittkop B et al (2019) Breeding improves wheat productivity under contrasting agrochemical input levels. *Nat Plants* 5(7):706–714. <https://doi.org/10.1038/s41477-019-0445-5>
- Wang T-C, Chen T-W (2022) toolStability. Tool for Stability Indices Calculation. <https://cran.r-project.org/web/packages/toolStability/index.html>
- Wang Y, Mette MF, Miedaner T et al (2015) First insights into the genotype–phenotype map of phenotypic stability in rye. *J Exp Bot* 66(11):3275–3284. <https://doi.org/10.1093/jxb/erv145>
- Welcker C, Spencer NA, Turc O et al (2022) Physiological adaptive traits are a potential allele reservoir for maize genetic progress under challenging conditions. *Nat Commun* 13(1):3225. <https://doi.org/10.1038/s41467-022-30872-w>
- Wricke G (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z Für Pflanzenzücht* 47:92–96
- Wu A, Hammer GL, Doherty A et al (2019) Quantifying impacts of enhancing photosynthesis on crop yield. *Nat Plants* 5(4):380–388. <https://doi.org/10.1038/s41477-019-0398-8>
- Wu L, Yuan S, Huang L et al (2016) Physiological mechanisms underlying the high-grain yield and high-nitrogen use efficiency of elite rice varieties under a low rate of nitrogen application in China. *Front Plant Sci* 7:1024. <https://doi.org/10.3389/fpls.2016.01024>
- Xiao YG, Qian ZG, Wu K et al (2012) Genetic gains in grain yield and physiological traits of winter wheat in Shandong Province, China, from 1969 to 2006. *Crop Sci* 52(1):44–56. <https://doi.org/10.2135/cropsci2011.05.0246>
- Yaseen M, Eskridge KM, Murtaza G (2018) Stability: stability analysis of genotype by environment interaction (GEI). <https://cran.r-project.org/package=stability>
- Zhu X-G, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61(1):235–261. <https://doi.org/10.1146/annurev-arplant-042809-112206>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.