A method for identifying environmental stimuli and genes responsible for genotypeby-environment interactions from a large-scale multi-environment data set

Supplementary Figures

Akio Onogi^{*1}, Daisuke Sekine², Akito Kaga³, Satoshi Nakano⁴, Tetsuya Yamada³, Jianming Yu⁵, Seishi Ninomiya⁶

¹ Department of Plant Life Science, Faculty of Agriculture, Ryukoku University, Otsu, Shiga, 520-2194, Japan

²Institute of Vegetable and Floriculture Science, National Agriculture and Food Research Organization, Tsu, Mie 514-2392, Japan

³Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8518, Japan

⁴Institute for Agro-Environmental Sciences, National Agriculture and Food Research Organization,

Tsukuba, Ibaraki 305-8604, Japan

⁵Department of Agronomy, Iowa State University, Ames, IA 50011, USA

⁶Graduate School of Agricultural and Life Science, The University of Tokyo, Nishitokyo, Tokyo 188-0002, Japan

Corresponding author:

Akio Onogi Postal address: 1-5, Yokotani, Oe-cho, Seta, Otsu, Shiga, 520-2194, Japan Phone: +81-(0)77-599-5719 E-mail: onogiakio@gmail.com

Abbreviations

Trait

YI	yield
SW	seed weight
DTF	days to flowering
DTM	days to maturity
SL	stem length
PR	protein content;

Meteorological factor

Т	mean temperature	u
Tmax	maximum temperature	u10m
Tmin	minimum temperature	Ν
Pr	precipitation	Ss
е	vapour pressure	EP
VPD	vapour pressure deficit	Ph
RH	relative humidity	
RHmin	minimum relative humidity	

u	wind speed
u10max	maximum wind speed
Ν	hours of sunshine
Ss	solar radiation
EP	potential evapotranspiration
Ph	photoperiod.



Figure S1 Results of the clustering of environmental covariates. The diagonal boxes of the triangles correspond to the 1st to 30th growth stages, from the lower left to the upper right. The off-diagonal elements correspond to the growth periods that span multiple stages, where the x and y axes denote the start and end of the periods, respectively. Note that, although neighbouring clusters are coloured using different colours, the same colour does not indicate the same cluster. The number of clusters is shown in the plots.

(A)



Figure S2 *P* values obtained from the genome-wide association mapping of the slopes. (A) Distribution of *P* values. (B) QQ plot of -log10P values. The broken line is the 1:1 line.



Figure S3 Distributions of estimates of genetic correlations between environments



Figure S4 Associations of environmental covariates with genetic correlations between environments. The heat maps represent the $-\log 10 P$ values for the correlation coefficients (r^2) of off-diagonal elements between the similarity matrix of each environmental covariate and the genetic correlation matrix. The diagonal boxes of the triangles correspond to the 1st to 30th growth stages, from the lower left to the upper right. The off-diagonal elements correspond to the growth periods that span multiple stages, where the x and y axes denote the start and end of the periods, respectively. The broken lines indicate flowering time.



Start stage of growth period



Figure S5 Results of the genome-wide association mapping for the slopes obtained for seed weight. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The $-\log 10 P$ values are illustrated using 7 colours. The blue triangles indicate significant SNPs with a false discovery rate <0.05.



Figure S6 Results of the genome-wide association mapping for the slopes obtained for yield. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The $-\log 10 P$ values are illustrated using colours. The blue triangles indicate significant SNPs with a false discovery rate <0.05.



Figure S7 Results of the genome-wide association mapping for the slopes obtained for stem length. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The -log10 P values are illustrated using colours. The blue triangles indicate significant SNPs with a false discovery rate < 0.05.



Figure S8 Results of the genome-wide association mapping for the slopes obtained for protein content. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The $-\log 10 P$ values are illustrated using colours. The blue triangles indicate significant SNPs with a false discovery rate <0.05.



Figure S9 Results of the genome-wide association mapping for the slopes obtained for days to flowering. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The $-\log 10 P$ values are illustrated using colours. The blue triangles indicate significant SNPs with a false discovery rate <0.05.



11



factors and

(start_end)

Figure S10 Results of the genome-wide association mapping for the slopes obtained for days to maturity. The x and y axes indicate the chromosomal positions and the environmental covariates, respectively. The -log10 P values are illustrated using colours. The blue triangles indicate significant SNPs with a false discovery rate < 0.05.



12



Figure S11 Distributions of meteorological factor/growth stage combinations where significant associations were detected. Red dots indicate the combinations with significant associations. The diagonal elements of the triangles correspond with the 1st to 30th growth stages, from the lower left to the upper right. The off-diagonal elements correspond to the growth periods that span multiple stages, where the x and y axes denote the start and end of the periods, respectively.



Start stage of growth period

Photoperiod at the 4–5th growth stage

Maximum temperature at the 5–14th growth stage



Figure S12 Allele substitution effects of flowering genes (E2, E3 and E4) on DTF. The effects of loss-offunction alleles (e2-ns, e3-tr and e4-SORE-I) were estimated for each environment (Supplementary Methods) and plotted against two environmental covariates (photoperiod at the 4th to 5th growth stages and maximum temperature at the 5th to 14th growth stages). For both environmental covariates, E2 showed the greatest slopes, suggesting that *E*2 can affect the $G \times E$ interactions of DTF. The red lines are the regression lines estimated using the least squares method.