**Materials and Methods**

**2.1. Plant materials and field experiment**

Plant materials were previously described (Lin et al. 2021, Lin et al., 2023). In brief, a diverse panel of 200 varieties and accessions from industry and government was used. Collectively, S&W Seed Co., Alforex Seed Co., Legacy Seed Co., and Blue River Hybrid Seed Co. The diverse panel of germplasm was planted in three locations using an augmented randomized complete block design (ARCBD) with 11 blocks. Each block contained 20 plots with a plot size of 1 meter by 4.6 meters. Twenty entries were planted per block, including two controls (Vernal and Hi-Gest360) and 18 new entries. The controls were replicated in each block. The diverse panel of germplasm was planted in three locations: Kimberly, Idaho (ID) (42°33'0.16"N 114°20'17.33"W), Union, Oregon (OR) (45°12'21.4"N 117°52'35.6"W), and Prosser, Washington (WA) (46°17'31.45"N 119°44'10.26"W).

**2.2 Phenotypic data collection**

**2.3 Single-trial analysis**

Fourteen phenotypic traits (A, A2, ADIPCP, ANDForm, Ash, B, B1, B2, C, CP, ERD6, kd, NIDCP, and SP) were fitted using a mixed linear model (MLM) in a single-trial approach to obtain the best linear unbiased estimate (BLUE) values for all entries. Control varieties were considered as fixed effects, block positions as random effects, and column and row field positions were used as residuals.

where is a vector () of response variable observed in accession at the block, is the overall mean, is a vector of accessions as fixed effects, is the fixed effect of the control accessions nested in the plot , is the random effect of the block, and is the vector () of residuals for the trial. Auto-regressive model of order one (AR1× AR1) was used to improve the estimation of residuals field position coordinates of plots described as follows:

where and are the and correlation matrices for the row and column dimensions, and is the Kronecker product of correlation matrices to obtain a single autocorrelation parameter.

**2.4. Stagewise analysis**

Stagewise analysis uses BLUE values from single-trial analysis and fits the response using multiple environments to obtain best linear unbiased prediction (BLUP) values. The environment was defined as the unique interaction of location by phenotyping year: ID\_2020, OR\_2020, OR\_2021, WA\_2020, and WA\_2021. BLUPs were obtained with a variance-covariance matrix for multi-environment datasets as random effects using ASReml-R software (Butler et al. 2018). The correlation structure for the effects of a common accessions across multiple environments can be specified in a nested model as

where is the response variable observed in accession at the environment, is the overall mean; is the fixed effect of the environment , is a vector of random genetic effects of accession nested in the environment ; assuming a distribution of identical and independently distributed, . We assume that , where is a variance-covariance matrix across environments and is the identity matrix for accessions.

Factor analytic (FA-*k*) approach was used to estimate the variance-covariance matrix across environments. FA-*k* models were specified in ASReml-R software (Butler et al. 2018) where the variance was given as the direct product of an FA covariance matrix for environments () and an accession effect correlation matrix parameterized as:

where is the Kronecker product, is the identity matrix for accessions, and is described as

where, is the matrix of environment loadings with dimension ; is an off-diagonal matrix with site-specific genetic variances with dimension . Factor loadings are part of the outcome from FA, which serves as a data reduction method designed to explain the correlation between observed variables using a smaller number of factors.

**2.5. Genome-wide association study and marker annotation**

GWAS were performed using Q+K MLM approach with the R package GWASpoly (Rosyara et al. 2016). The Q+K MLM incorporates the population structure information as principal components (Q) and the correlations among individuals as kinship matrix (K) to reduce the number of false positive associations described as

where is a vector of observed phenotypes (); is an incidence matrix of fixed effects (); is a vector of fixed effects (); is an incidence matrix mapping observed phenotypes to genotypes (accessions) (); is a structure incidence matrix (); is a vector of SNPs effects (); is an incidence matrix for population size (); is a vector of fixed effects for the principal components (); is a vector of random effects of the mixed model (), with a covariance proportional to K matrix, ; is the () kinship matrix inferred from genotypes and is a residuals vector (Rosyara et al. 2016). GWAS were performed with eight different models: general, diplo-general, additive, 1-dom-ref, 1-dom-alt, 2-dom-ref, 2-dom-alt, and diplo-additive (Rosyara et al. 2016). The p-values were used for identifying SNPs significantly associated with the traits with the Bonferroni correction at the cutoff value of 5%. Significantly associated markers were annotated against the Uniprot100 database (Bateman 2019) using the reference transcriptome dataset for *M. sativa* (Medina et al. 2021).