Check for updates

Received: 27 November 2025 | Accepted: 11 June 2026

DOI: 10.1002/tpg2.70277

The Plant Genome

ORIGINAL ARTICLE

Genomic prediction in quinoa across contrasting environments using statistical and machine learning models

Clara S. Stanschewski[1] | **Mark Warmington**[2,3] | **Irfan Afzal**[4,5] | **Elodie Rey**[1,6] | **Gabriele Fiene**[1] | **Evan Craine**[7,8] | **Kevin Murphy**[7] | **Mark Tester**[1] | **Jesse Poland**[1]

[1]Plant Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[2]Department of Primary Industries and Regional Development, Agriculture and Food Division, Kununurra, Western Australia, Australia

[3]Northern Australia Crop Research Alliance (NACRA), Kununurra, Western Australia, Australia

[4]Department of Agronomy, University of Agriculture Faisalabad, Faisalabad, Pakistan

[5]African Sustainable Agriculture Research Institute, Mohammed VI Polytechnic University, Laayoune, Morocco

[6]Crop Genetics Team, BASF Belgium Coordination Center CommV, Gent-Zwijnaarde, Belgium

[7]Department of Crop and Soil Sciences, Washington State University, Pullman, Washington, USA

[8]The Land Institute, Salina, Kansas, USA

Correspondence
Jesse Poland, Plant Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955–6900, Saudi Arabia.
Email: jesse.poland@kaust.edu.sa

Assigned to Associate Editor Diego Jarquin.

Funding information
King Abdullah University of Science and Technology, Grant/Award Numbers: CRG5-2966, baseline funding; United States Department of Agriculture, Grant/Award Number: 2016-680004-2470; National Science Foundation Graduate Research Fellowship, Grant/Award Number: 1842493

Abstract

Quinoa (*Chenopodium quinoa* Willd.) is gaining global importance for its nutritional value and adaptability; however, breeding progress remains limited. Genomic selection (GS), combined with rapid generation cycles, offers a strategy to accelerate genetic improvement. We conducted whole-genome resequencing of 610 accessions and present the first evaluation of genomic prediction in quinoa evaluated across six field trials in Australia and Pakistan for seven phenological and yield-related traits. Using ∼1.8 million single-nucleotide polymorphisms, we compared four models—genomic best linear unbiased prediction, reproducing kernel Hilbert space, BayesC, and light gradient boosting machine—for genotype ranking under four cross-validation schemes: predicting new genotypes (CV1), sparse testing (CV2), leave-one-location–year-out (CV0), and across locations. Model performance was evaluated using Pearson's correlation for overall accuracy and normalized discounted cumulative gain (NDCG@10) for ranking top performers. The four models were similar, with no method dominating across traits. NDCG@10 scores revealed that

Abbreviations: BLUE, best linear unbiased estimator; DCG@10, discounted cumulative gain;; DCG@10, discounted cumulative gain; FS, Fisher strand bias; GBLUP, genomic best linear unbiased prediction; GS, genomic selection; GWAS, genome-wide association study; G × E, genotype-by-environment interactions; MAF, minor allele frequency; NDCG@10, normalized discounted cumulative gain; QD, quality by depth; QTL, quantitative trait locus; RKHS, reproducing kernel Hilbert space; SNP, single-nucleotide polymorphism; SOR, strand odds ratio.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2026 The Author(s). *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

predictions remained useful for selecting superior genotypes even for difficult traits. Prediction accuracy was strongly associated with heritability and trait correlations across and within- location environments. Accuracy was highest for developmental traits and lowest for seed yield, while seed traits showed location-specific responses with higher accuracy in Australia. These findings support GS as a promising tool for quinoa breeding and provide benchmarks for global implementation.

Plain Language Summary

Quinoa is a nutritious crop that can grow in diverse environments. With rising global demand, there is need for improved varieties, but developing new varieties through traditional breeding is slow. This study evaluates genomic prediction, a method for predicting plant traits using genetic markers and supporting earlier selection in breeding programs. To assess how well top candidate varieties for seven key traits could be predicted, 551 quinoa varieties grown in field trials in Australia and Pakistan were used. Prediction accuracy varied by trait and location. In Australia, models predicted the performance of untested varieties with moderate to good accuracy for most traits, showing they captured useful genetic information. In Pakistan, predicting seed-related traits was less accurate, likely due to stronger environmental effects and trait measurement differences, but the models still supported selection decisions. Overall, genomic prediction is a promising tool for quinoa breeding.

1 | INTRODUCTION

Quinoa (*Chenopodium quinoa* Willd.), a semidomesticated grain crop, is native to the Andes in South America. Traditional quinoa production and breeding efforts remain concentrated in Ecuador, Bolivia, and Peru. However, recent cultivation and research have expanded quinoa production to over 120 countries globally (Alandia et al., 2020). Growing global interest in quinoa reflects recognition of its exceptional nutritional value, environmental adaptability, abiotic stress tolerance, and extensive genetic diversity (Alandia et al., 2020; Chaudhary et al., 2023; Jacobsen, 2003; Murphy & Matanguihan, 2015). Successful adaptation to diverse environments requires identifying optimal ideotypes for different climatic clusters (Lagergren et al., 2023). This process depends on coordinated multi-environment evaluations supported by standardized phenotyping protocols and comprehensive metadata collection (Stanschewski et al., 2021), alongside targeting key quinoa breeding objectives. Breeding targets include improved yield potential, harvest index, days-to-maturity, disease resistance, saponin reduction, and photoperiod insensitivity (Murphy et al., 2019; Nakamichi, 2015; Zurita-Silva et al., 2014).

Recent genomic advances have created unprecedented opportunities for quinoa improvement. Genome assemblies are now available for multiple quinoa accessions representing the broad phenotypic and genetic diversity of both lowland and highland ecotypes (Bodrug-Schepers et al., 2021; Rey et al., 2024; Yasui et al., 2016; Zou et al., 2017). Major developments include an improved chromosome-scale reference genome for accession PI-614886 (QQ74) (Jarvis et al., 2017), with 90.5% of the 1.33-Gb assembly organized into 18 chromosome-scale scaffolds using high-throughput chromosome conformation capture technology (Rey et al., 2023). Whole-genome resequencing data from diverse germplasm generated in this study, along with these genomic resources, establish a foundation for genetic analyses, including genomic prediction, association mapping, comparative genomics, and pan-genome analyses.

With the availability of robust and cost-effective genomic resources, including genome-wide skim sequencing for profiling large breeding populations, there is considerable value in accelerating crop improvement through leveraging genomic data. A central component of genomics-assisted breeding involves connecting phenotypic and genotypic data and developing predictive models for early-generation selection prior to phenotypic evaluation. Molecular tools have enhanced each component of the genetic gain equation, including access to genetic variation, precision of heritability estimates, selection intensity, and breeding cycle duration (Xu et al., 2017). Further improvements in genetic gain and reductions in the gap between experimental yield potential and on-farm real-

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ized yield under increasing climatic pressures require modern breeding technologies and a deeper understanding of the genetic architecture of target traits (Brummer et al., 2011; Daetwyler et al., 2010; Xu et al., 2017).

Genome-wide association studies (GWASs) provide valuable insights into trait architecture; however, their utility for crop improvement remains limited by trait complexity, population structure, and genetic heterogeneity driven by local adaptation (Fournier-Level et al., 2011; Hayes, 2013; Korte & Farlow, 2013). These limitations are particularly pronounced in quinoa, which exhibits strong population structure between highland and lowland ecotypes (F_{ST} divergence = 0.36) and distinct linkage disequilibrium decay patterns (6.5 and 49.8 kb, respectively). A large GWAS identified 1480 significant marker–trait associations across 17 traits after controlling for population structure and multiple testing (Patiranage et al., 2022). While this study provides valuable insights into quinoa genetics, the exceptionally high number of detected associations reflects the highly polygenic architecture of agronomically important traits and likely includes false-positive signals, limiting the practical utility of individual markers for breeding applications. Quantitative trait locus (QTL) mapping in biparental populations typically identifies a more limited and tractable number of loci; for example, Maldonado-Taipe et al. (2022) report 15 QTLs associated with 10 traits, including seven with pleiotropic effects. However, QTL effect estimates for complex traits are often inconsistent across populations and environments, QTL numbers are frequently underestimated, and marker effect sizes are often overestimated (Bernardo, 2008). For complex traits controlled by many genes with small effect sizes, genomic prediction may therefore offer a more effective approach for harnessing molecular markers in breeding programs (Bernardo, 2008; Meuwissen et al., 2001).

Genomic selection (GS) represents a fundamentally different strategy from association mapping by utilizing genome-wide marker information to predict breeding values rather than testing the statistical significance of individual loci. This approach relies on a reference population with both genotypic and phenotypic data to train prediction models that generate genomic estimated breeding values (GEBVs) for selection candidates carrying only marker data, thereby facilitating selection without direct phenotypic evaluation (Heffner et al., 2009; Meuwissen et al., 2001). The widespread availability of low-cost genomic data has shifted the primary bottleneck in genetic analyses and crop improvement from genotyping to phenotyping (Cobb et al., 2013). Integrating molecular markers into selection decisions can reduce evaluation costs compared to phenotypic assessment, support off-season selection to accelerate breeding cycles beyond traditional constraints, and improve the accuracy of GEBVs.

The effectiveness of GS depends on several factors, including trait heritability (Calus et al., 2008), and genetic architecture (Morgante et al., 2018). In contrast to GWAS, where population structure can generate false-positive associations (Hayes, 2013), research in wheat demonstrates that while closely related training and validation populations achieve higher prediction accuracies due to shared haplotype effects, prediction accuracy can be maintained across genetically distant populations by incorporating greater genetic diversity in training sets and utilizing higher marker densities (Norman et al., 2018). Fundamental to quantitative genetics analyses is that statistical power for detecting and estimating genetic parameters increases with sample size (Poland, 2015).

Combining phenotypic data from field trials across multiple environments increases the total number of observations and improves the statistical power for estimating genetic effects and genotype-by-environment interactions (G × E). Thus, multi-environmental trials are used to evaluate genotypes across diverse locations and years. Incorporating environmental data from trial networks into genomic prediction models enables breeders to predict genotype performance for both new genotypes and previously untested environments (Van Eeuwijk et al., 2016). This capability is particularly relevant for quinoa because of its limited breeding history and expanding cultivation into new production regions. Larger quinoa field trial networks could support the identification of optimal germplasm subsets for new breeding programs without extensive preliminary screenings requirements.

In this study, we use field-trial data from six environments (location–year combinations) generated from evaluations conducted across two locations over 3 years in Australia and Pakistan to evaluate for the first time the applicability of genomic prediction in quinoa. While various modeling approaches are available for GS, no single model consistently performs best across all species and traits (Azodi et al., 2019). We assessed four complementary models that cover the main classes of genomic prediction models. Genomic best linear unbiased prediction (GBLUP) is a linear mixed model that uses a genomic relationship matrix and is the established baseline for GS (VanRaden, 2008). BayesC is a Bayesian model that estimates marker effects directly while assuming most markers have zero effect (Habier et al., 2011). Reproducing kernel Hilbert spaces (RKHSs) regression with Gaussian kernels is a semiparametric approach that captures additive and nonadditive genetic effects (De Los Campos et al., 2010; Gianola et al., 2006). Light gradient boosting machine (LightGBM), a gradient boosting machine learning algorithm optimized for high-dimensional data (Ke et al., 2017), captures nonlinear interactions without distributional assumptions and can implicitly account for G × E when environment identifiers are included as features (Fernandes et al., 2024; Westhues et al., 2021). This study aims to assess the ability of statistical and machine learning approaches for predicting the ranking of new quinoa genotypes in specific target environments using multi-environmental trial data.

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | MATERIALS AND METHODS

2.1 | Field trial design and germplasm

A total of 551 quinoa accessions from a diversity panel of 610 whole-genome resequenced quinoa accessions (described in Section 2.4) were assessed across six field trials conducted over 3 years at two locations in Australia and Pakistan. The Australian field trials were conducted at the Kimberley Research Station in Kununurra, Western Australia (15°38′60.0″ S, 128°42′24.8″ E), during the May–September dry season, when heat and humidity are lower and average temperatures are approximately 32°C. Average annual rainfall is approximately 800 mm, with August as the driest month, recording 0 mm of precipitation (Bureau of Meteorology, 2025). The Pakistani field trials were conducted in central Pakistan at Faisalabad, within the Agronomic Research Area of the University of Agriculture (31°26′29.0″ N, 73°04′19.5″ E). The research farm is located in a semiarid climate zone bordering a humid subtropical climate, characterized by very hot, humid summers and dry, cool winters. Therefore, the growing season was set from November to March, when average temperatures can decline to approximately 12.5°C during January. Average annual rainfall in this region is only approximately 600 mm, with November and December representing the driest months (Weather Atlas, 2025).

These two locations provided contrasting environmental conditions. Kununurra represents a tropical dry-season environment characterized by minimal rainfall and moderate temperatures, whereas Faisalabad represents a subtropical winter/spring environment with cool temperatures and very low precipitation during the growing season. In Australia, photoperiods during the growing season range from approximately 11.2 h at the June winter solstice to approximately 11.8 h by late August/early September. In Pakistan, photoperiods exhibit greater seasonal variation, ranging from approximately 10.1 h in December (winter solstice) to approximately 12 h by the end of March.

In Pakistan, plots received three flood irrigations at 30 days, 75 days, and 120 days after sowing, while in Australia, fields were flood irrigated 9–11 times per growing season. Plot dimensions varied across locations and years. In Australia, plots comprised one to two rows, each 5.9 m in length during the 2017–2019 growing seasons. In Pakistan, plots measured 1.5 m × 3 m across all three field trials conducted between 2019 and 2021, with ∼10 cm plant spacing within rows and ∼ 30 cm between rows. Both locations used a sowing density of approximately 5 kg/ha.

The total number of evaluated accessions differed among location–year combinations, ranging from 107 to 370 accessions evaluated across ≥2 location–years (Table 1). Individual trial designs also varied, with most designed as augmented designs, with only check lines replicated throughout the field to maximize the number of genotypes evaluated per year. In contrast, the PAK 2021 environment (Pakistan, 2021) was conducted using a randomized complete block design with replication. Although the trials were conducted independently and not all accessions were evaluated in every location–year (environment), the overlap across environments provided adequate connectivity to enable robust estimation of genotypic effects by allowing predictions to leverage information across environments. This connectivity also ensured that each cross-validation fold maintained a sufficient test set size.

2.2 | Phenotypic data

Seven phenotypic traits were selected for this study based on their agronomic importance and availability across environments: days to flowering, days to harvest, plant height (cm), panicle length (cm), seed yield (g), seed length (mm), and thousand-grain weight (g). Phenotypic measurements were primarily conducted as previously described (Stanschewski et al., 2021), with environment- and trait-specific modifications detailed below.

All seven phenotypic traits were recorded across all location–year environments except AUS 2017, where only plant height and seed yield were available. Seed yield required additional preprocessing prior to analysis. Plot sizes differed among locations and between years within the Australian trials, while yield measurements were recorded on a plot basis in Australia and a plant basis in Pakistan. To enable joint analysis of data collected at different scales, values were standardized using z-score within each location–year environment before estimating best linear unbiased estimators (BLUEs) or accession-level phenotypic means. This approach preserved relative performance differences among accessions across environments.

Seed length measurements for the Australian dataset were obtained using the image-based seed phenotyping pipeline previously described in Stanschewski et al. (2021). In the Pakistan dataset, a curved-surface seed caliper designed for measuring seed dimensional assessment was used. For each sample, three to five seeds were measured and averaged, whereas the imaging method enabled measurement of hundreds of seeds per sample. While this introduces methodological heterogeneity, it reflects practical conditions in breeding programs, where standardized protocols across global sites are challenging.

Another trait with differing methodological approaches between both locations was the days to harvest, defined as the number of days required to reach harvest maturity, the senescence growth stage at ∼ BBCH95 (Stanschewski et al., 2021). This trait was recorded in Pakistan, while for the

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Cross-tabulation matrix showing the number of shared accessions among location–year environments.

Location and year	PAK 2019	PAK 2020	PAK 2021	AUS 2017	AUS 2018	AUS 2019
PAK 2019	**224**	209	181	181	210	107
PAK 2020	209	**284**	242	230	268	135
PAK 2021	181	242	**244**	208	228	118
AUS 2017	181	230	208	**413**	370	192
AUS 2018	210	268	228	370	**494**	215
AUS 2019	107	135	118	192	215	**240**

Note: Bold diagonal values represent the total number of accessions evaluated within each environment, while off-diagonal values show the number of accessions shared between the corresponding row and column environments. Environments are designated by location and trial year.
Abbreviations: AUS, Australia; PAK, Pakistan.

TABLE 2 Fixed and random effects included in the models for the Pakistan and Australia datasets

Model	Accession	Year	Block (Year)[a]	Residual
BLUE (PAK)	Fixed	Random	Random	Random
BLUE (AUS)	Fixed	Random	–	Random
Heritability (PAK)	Random	Random	Random	Random
Heritability (AUS)	Random	Random	–	Random

[a]Block nested within year.
Abbreviations: AUS, Australia; BLUE, best linear unbiased estimator; PAK, Pakistan.

Australian datasets, days to actual harvest were used. This was less specific than the measurements in Pakistan, as plants were harvested in batches based on similar maturity stages.

2.3 | Statistical analysis

Statistical analyses were performed separately for the Pakistani and Australian datasets, using linear mixed models implemented in the R package ASReml-R (Butler et al., 2023). For each trait, two complementary models were fit to address different analytical objectives.

Location-level BLUEs were used for phenotypic data exploration, including trait distribution summaries and inter-location trait correlations. BLUEs were estimated using separate mixed linear models for each location (Australia and Pakistan). Within each location, data from all 3 years were jointly analyzed in a single model, with accession treated as a fixed effect and year as a random effect. The general model was specified as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is the vector of phenotypic observations; $\mathbf{X}$ is the design matrix for fixed effects, with $\boldsymbol{\beta}$ containing the intercept (overall mean), and accession effects; $\mathbf{Z}$ is the incidence matrix for random effects; $\mathbf{u}$ is the vector of random effects assumed to follow $\mathbf{u} \sim N(0, \sigma_u^2\, \mathbf{I})$; and $\mathbf{e}$ is the vector of residual effects, assumed to follow $\mathbf{e} \sim N(0, \sigma_e^2\, \mathbf{I})$. The specific random terms included in $\mathbf{u}$ varied between the locations and model objectives (Table 2). The PAK dataset included block nested within year (for 2021) as a random effect to account for the replicated block structure present in that experimental year. The AUS field trials followed an augmented design with unreplicated test lines and replicated check lines. No formal block structure was imposed; therefore, block effects were excluded from the AUS model. In these trials, check lines contributed to the estimation of year effects through their replication within and across years. BLUEs were then estimated for each accession evaluated in this location.

Generalized heritability was estimated using separate models with accession fit as random effects with $\mathbf{u} \sim N(0, \sigma_g^2\mathbf{I})$ (Table 2), following the method of Cullis et al. (2006), which is particularly suitable for the unbalanced experimental designs common in plant breeding trials because it accounts for the experimental design and data structure without requiring balanced data assumptions. Differing from the standard broad-sense heritability described by Falconer and Mackay (1996), which is based on the ratio of genetic to total variance, it incorporates the prediction error variances from the mixed-model analysis. It was calculated as follows:

$$h_g^2 = 1 - \frac{A_{tt}}{2\sigma_g^2} \quad (2)$$

where A_{tt} is the average pairwise prediction error variance of accession effects and σ_g^2 is the genetic variance component estimated from the mixed model (Table 2).

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Input data for the genomic prediction models consisted of location-level phenotypic means. Although these provided less precise genetic estimates than the BLUEs, trial-level means were used instead of BLUEs to retain environment-specific information for cross-validation. Owing to the largely unreplicated trial structure with formal replication present only within one of the six location–year environments, insufficient information was available to estimate reliable within-environment BLUEs. Even in the trial containing two replicates, BLUEs provided a limited advantage over unadjusted means in the absence of additional covariates.

2.4 | Genotypic data

2.4.1 | Diversity panel and sequencing

A quinoa diversity panel comprising 610 quinoa accessions was assembled from multiple germplasm collections, including the United States Department of Agriculture and the Leibniz Institute for Plant Genetics and Crop Plant Research. The panel represented both native cultivation regions in South America (Peru, Bolivia, and Chile) and introduced cultivation regions, including the United States, Australia, and Europe.

Leaf tissue was collected from 3- to 6-week-old plants and dried on silica gel. DNA was extracted using either the Genomic Micro AX Blood Gravity Kit (A&A Biotechnology) or the DNeasy Plant Kit following the manufacturer's protocols. Sequencing libraries were prepared using the Illumina Nextera protocol (Baym et al., 2015) or standard Novogene procedures. All samples were sequenced on an Illumina NovaSeq 6000 platform at Novogene Co., Ltd., generating 150 bp paired-end reads with an average coverage of 6–8×.

2.4.2 | Read processing and variant calling

Raw sequencing reads were quality filtered using Trimmomatic v0.38 (Bolger et al., 2014) to remove low-quality bases (LEADING:20, TRAILING:20) and short reads (MINLEN:50 bp), applying a sliding-window approach (SLIDINGWINDOW:5:20). Filtered paired-end reads were aligned to the updated quinoa reference genome assembly for accession PI-614886 (QQ74) (Rey et al., 2023) using the Burrows–Wheeler Aligner Maximal Exact Matches (BWA-MEM) algorithm v0.7.17 (Li & Durbin, 2009) and a mapping quality (MQ) threshold of 30. Post-alignment processing included duplicate marking, read-group assignment, and Binary Alignment/Map (BAM) file indexing using Picard Tools and Sequence Alignment/Map tools (SAMtools) v1.8 (Li et al., 2009).

Variant calling was performed using an automated pipeline (Abrouk et al., 2020) based on the HaplotypeCaller algorithm implemented in the Genome Analysis Toolkit v4.1.8.0 (Depristo et al., 2011; McKenna et al., 2010). This step generated individual genomic Variant Call Formats for each sample. Joint genotyping across all 610 accessions was subsequently performed using GenomicsDB and GenotypeGVCFs, with chromosomes processed in segmented chunks to improve computational efficiency. Quality filtering was applied using the following exclusion criteria for single-nucleotide polymorphisms (SNPs): quality by depth (QD) < 2.0, Fisher strand bias (FS) > 60.0, root mean square MQ < 40.0, MQ rank sum < −12.5, read position rank sum (ReadPosRankSum) < −8.0, and strand odds ratio (SOR) > 3.0. Additionally, clustered SNPs were removed using a window size of 10 bp and a cluster size of 3. Indels were filtered using QD < 2.0, FS > 200.0, ReadPosRankSum < −20.0, and SOR > 10.0. Quality filtering retained approximately 38 million SNPs and 20 million indels across the 610-accession panel.

2.4.3 | Single-nucleotide polymorphism preparation for genomic prediction

SNP data were further processed for genomic prediction analyses using bcftools (Danecek et al., 2021), applying quality filters that preserve population genetic structure while maintaining data quality. From the initial dataset, samples were first subset to include only the 551 accessions used in this study. Subsequent filtering retained only biallelic SNPs located on the 18 quinoa chromosomes (Cq1A–Cq9B), excluding organellar sequences. A relatively permissive missing data threshold of 20% per SNP (F_MISSING < 0.2) was applied to avoid excessive removal of presence–absence variation, which can appear as missing data in high-depth sequencing datasets (~8–10× coverage on inbred lines in this study) but represents genuine biological variation. The minor allele frequency (MAF) threshold was set at 0.01 (MAF > 0.01) to retain rare alleles relevant for kinship estimation in genomic prediction models. Mean read depth filters were set between 5× and 30× across samples (MEAN(FORMAT/DP) > 5 and < 30), with the upper threshold chosen to maintain high-coverage variants while excluding repetitive genomic regions that typically exhibit ~100× coverage. Following filtering, approximately 1.8 million high-quality biallelic SNPs were retained for genomic prediction analyses.

2.5 | Genomic prediction methods

Genomic prediction was evaluated using four complementary approaches: the statistical mixed-model approaches GBLUP, an RKHSs regression with Gaussian kernel averaging, Bayesian SNP regression (BayesC), and a machine learning method (LightGBM). Models were trained using

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data from all locations and years, with test genotypes, observations, or locations/years masked depending on the cross-validation scheme (described in Section 2.6). Prediction accuracy was evaluated separately within each location–year to avoid artificial inflation of correlation-based prediction accuracy estimates.

To ensure prediction accuracy reflected the ranking of genotypes within locations rather than environmental mean differences, phenotypes were z-score standardized within each location–year environment (mean = 0, standard deviation = 1) prior to model training. This transformation removes between-environment differences in trait means and scales, and ensures that the models learn relative genotypic performance rather than absolute values influenced by environmental main effects. Accurate prediction of absolute genotypic performance across environments would require the explicit modeling of genotype-by-environment interactions, which was not feasible given the limited replication in only six location–year environments and two locations.

2.5.1 | Genomic best linear unbiased prediction model

The GBLUP model was implemented in ASReml-R (Butler et al., 2023) with the model specification as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{\mathrm{g}}\mathbf{u}_{\mathrm{g}} + \mathbf{Z}_{\mathrm{y(loc)}}\mathbf{u}_{\mathrm{y(loc)}} + \mathbf{e} \quad (3)$$

where $\mathbf{y}$ is the vector of z-score standardized phenotypic observations; $\mathbf{X}$ is the design matrix for fixed effects relating observations to $\boldsymbol{\beta}$, the vector of fixed location effects; $\mathbf{Z}_{\mathrm{g}}$ is the incidence matrix relating observations to genotypes, with $\mathbf{u}_{\mathrm{g}}$ representing additive genetic random effects following $\mathbf{u}_{\mathrm{g}} \sim N(0, \sigma^2\mathbf{G})$, where $\mathbf{G}$ is the VanRaden genomic relationship matrix and accessions share information based on genetic similarity; $\mathbf{Z}_{\mathrm{y(loc)}}$ is the incidence matrix for year effects nested within location, with $\mathbf{u}_{\mathrm{y(loc)}}$ representing location-specific year effects modeled with separate variance components for each location, that is, $\mathbf{u}_{\mathrm{y(loc)}} \sim N(0, \sigma^2_{y(\mathrm{loc})}\mathbf{I})$ for locations $\in$ {AUS, PAK}; and $\mathbf{e}$ represents the vector of residual effects modeled with location-specific residual variances, $\mathbf{e} \sim N(0, \mathrm{diag}(\sigma^2_{e,\mathrm{AUS}}, \sigma^2_{e,\mathrm{PAK}})$, to account for the significant differences in signal-to-noise ratios between both locations.

To benchmark the global model against within-location models and estimate location-specific genetic variance components, separate GBLUP models were also fitted independently for each location. These models included an intercept-only fixed effect, a random year effect, and a homogeneous residual variance structure. Heterogeneous residual variances across years were not modeled in the per-location models due to frequent convergence failures owing to the limited number of years per trait (2–3), the absence of within-year replication, and the presence of year as a random effect.

The additive genomic relationship matrix was constructed from biallelic SNP data using the VanRaden method (ploidy = 2) implemented in the AGHmatrix R package (Amadeu et al., 2023). It was used as an input feature in the LightGBM model (Section 2.5.4), and its sparse inverse form was incorporated into the GBLUP model.

2.5.2 | Reproducing kernel Hilbert spaces model

The RKHS model is a semiparametric regression approach that captures both additive and nonlinear genetic effects, including epistatic interactions, through the use of kernel functions (De Los Campos et al., 2010; Gianola et al., 2006). The model was implemented using the R package BGLR (Pérez & De Los Campos, 2014) with the following linear predictor:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \mathbf{e} \quad (4)$$

where $\mathbf{X}\boldsymbol{\beta}$ represents to the fixed effects of location and year nested within location (flat prior), $\mathbf{u}_k$ ($k = 1, 2, 3$) denotes random genetic effects each following $\mathbf{u}_k \sim N(0, \sigma^2_{uk}\mathbf{K}_k)$, and $\mathbf{e} \sim N(0, \sigma^2_e\mathbf{I})$ represents the residual term. Unlike BayesC (Section 2.5.3), the BGLR package does not support heterogeneous residual variances for the RKHS model type; consequently, a single residual variance was estimated across all location–year combinations.

Each $\mathbf{K}_k$ represents a Gaussian kernel matrix that quantifies pairwise genomic similarity among individuals based on the squared Euclidean distance between their marker profiles, computed from the complete marker set comprising ~ 1.8 million markers. Three kernels with distinct bandwidth parameters were employed to capture genomic similarity at multiple scales, following the kernel-averaging approach described by De Los Campos et al. (2010) and subsequently applied by Cuevas et al. (2016). The BGLR package was used independently to estimate the variance component associated with each kernel, thereby employing the data to identify the scale of genomic similarity that was most informative for each trait. Posterior inference was conducted using Gibbs sampling with 15,000 iterations, including a burn-in period of 5000 iterations, and a thinning interval of 5.

2.5.3 | BayesC model

The BayesC model is a Bayesian variable-selection regression approach that estimates individual marker effects while assuming that only a fraction of markers exerts nonzero effects

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

on the trait (Habier et al., 2011). The model was implemented using the BGLR package (Pérez & De Los Campos, 2014) with the following linear predictor:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{M\alpha} + \mathbf{e} \quad (5)$$

where $\mathbf{X\beta}$ represents fixed effects for location and year nested within location (flat prior), $\mathbf{M}$ denotes the matrix of mean-imputed SNP genotypes (556,604 markers retained after LD pruning at $r^2 < 0.5$), $\boldsymbol{\alpha}$ represents the vector of marker effects, and $\mathbf{e}$ denotes the residual term. The residual variance was modeled separately for each location–year combination to account for differences in experimental precision across environments. Each marker effect was assigned a spike-and-slab prior, allowing the model to constrain a proportion of marker effects to exactly zero while assigning normally distributed effects to the remaining markers (Habier et al., 2011). Hyperparameter settings included 15,000 iterations, a burn-in period of 5000 iterations, and a thinning interval of 5.

2.5.4 | Machine learning model

The LightGBM algorithm is a computationally efficient and memory-optimized implementation of the gradient boosting decision tree machine learning algorithm (Ke et al., 2017). Gradient boosting adds predictors as an additive ensemble of decision trees, in which each successive tree is fitted to the residual errors of the preceding ensemble (Géron, 2022). The input features for the machine-learning approach comprised the genomic relationship matrix used in the GBLUP model (Section 2.5.1), along with one-hot encoded location and location–year variables, alongside trait data. Model hyperparameters were optimized independently for each trait using randomized search (RandomizedSearchCV; Pedregosa et al., 2011) over 50 parameter combinations, with performance evaluated via fivefold cross-validation (GroupKFold; Pedregosa et al., 2011) stratified by genotype. The hyperparameter search space included the learning rate, maximum tree depth, number of estimators, subsample ratio, column sample ratio, minimum number of child samples, and regularization parameters. The optimal hyperparameter configuration for each trait was selected based on the mean cross-validation performance and was subsequently used for all subsequent prediction runs across the cross-validation schemes.

2.6 | Evaluation of model performance

Genomic prediction accuracy was assessed using four cross-validation schemes, each representing a distinct breeding scenario as described in Alemu et al. (2024).

CV1 was used to evaluate the prediction of new, untested accessions within known location–year environments. The 551 accessions were randomly partitioned into five nonoverlapping groups, with each fold containing approximately 110 accessions (20% of the total). In each iteration, four folds (440 accessions; 80%) were used for model training, while one fold was reserved for testing. All phenotypic observations for the test accessions were masked across all location–years. This procedure was repeated five times until each accession served as a test case exactly once. To obtain robust estimates of prediction accuracy and account for variability arising from random fold partitioning, the fivefold cross-validation was repeated 15 times using different random seeds.

CV2 simulates a sparse testing scenario in which accessions are evaluated in some location–years but not in others. Within each location–year, the observed accession × location–year cells were stratified into five folds. Each fold was held out in turn, while the remaining cells were used for model training. This procedure was repeated for 15 iterations using different random seeds.

CV0 was employed to evaluate prediction in untested environments by holding out one entire location–year combination while training on the remaining five. The *cross-location* scheme was used to evaluate the transferability of predictions between locations by training the model on all years from one location and predicting all years from the other. CV0 and cross-location schemes were implemented deterministically because the held-out unit (location–year or location) was fixed by design rather than randomly sampled. Under the cross-location scenario, a simplified genomic prediction model was required relative to those for CV1, CV2, and CV0 because the location-level fixed effects were not estimable when all observations from one location were absent. Consequently, for GBLUP, RKHS, and BayesC, the fixed effects for location were omitted, whereas for LightGBM, the location and location–year encoded variables were excluded during cross-location prediction.

Model performance was evaluated within each cross-validation fold using two complementary metrics: the Pearson correlation between GEBVs and observed phenotypic values, and normalized discounted cumulative gain at rank 10 (NDCG@10) (Blondel et al., 2015). Both metrics were calculated separately for each location–year combination rather than being pooled across environments. Pooling predictions across environments before computing correlation may artificially inflate accuracy estimates when models accurately predict environmental mean differences but fail to correctly rank genotypes within environments, because the between-environment variance in trait means may dominate the correlation structure. To mitigate this effect, observed phenotypic values were standardized using z-scores prior to model fitting (Section 2.5), thereby removing environmental main effects such that prediction accuracy primar-

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ily reflected within-environment ranking ability. Computing correlations separately for each location–year combination provided a complementary safeguard against inflated accuracy estimates, whereas grouping prediction accuracies by location in the results enabled assessment of location-specific model performance under differing environmental conditions.

NDCG@10, as described by Blondel et al. (2015), was used to evaluate the ability of the model to correctly rank the top 10 accessions within each cross-validation fold (~15% selection intensity on average, depending on fold size). NDCG@10 was calculated as follows: first, accessions were ranked according to their predicted values, and the top 10 were selected. Each selected accession was then assigned a gain value. A linear gain function was used; therefore, the gain assigned to each accession was equal to its observed phenotypic value. The gains were subsequently summed across the top 10 ranking using a position-based discount function that down-weights contributions from lower-ranked positions giving a discounted cumulative gain (DCG@10):

$$\mathrm{DCG@10} = \sum_{i=1}^{10} \frac{g\left(y_{(i)}\right)}{\log_2(i+1)}$$

where $y_{(i)}$ denotes the observed phenotypic value of the accession ranked at position i according to the predicted values, and $g(y) = y$ under the linear gain function. To normalize the metric, the same calculation was performed using the ideal ranking, defined as the top 10 accessions ordered according to their observed phenotypic values, thereby yielding the maximum achievable DCG@10. NDCG@10 was calculated as the following ratio:

$$\mathrm{NDCG@10} = \frac{\mathrm{DCG@10}}{\text{Ideal DCG@10}}$$

NDCG@10 scores range from 0, representing the poorest possible ranking, to 1, representing a perfect ranking in which the model selected the same top 10 accessions as the top 10 from observed phenotypic values in the identical order. Because gains enter the formula as magnitudes rather than ranks, NDCG@10 is defined only for non-negative relevance values; negative gains would reverse the sign of the discount-weighted sum and invalidate the normalization to the [0,1] interval. Since phenotypic values were standardized using z-score within each location–year combination, observed and predicted values included negative values. Therefore, both sets of values were shifted by a common constant to ensure non-negativity prior to NDCG@10 calculation. This transformation preserved the ranking and relative spacing of accessions and did not alter the interpretation of the metric. For traits in which higher values were not considered desirable breeding targets (days to flowering, days to harvest, and plant height), lower-ranked positions were instead rewarded. However, some traits may not be subject to directional selection.

2.7 | Software and packages

The GBLUP approach was implemented in R using ASReml-R (Butler et al., 2023) for mixed model fitting and the dplyr package for data manipulation (Wickham et al., 2025). The machine learning approach was implemented in Python using LightGBM for gradient boosting (Ke et al., 2017), scikit-learn for cross-validation and preprocessing (Pedregosa et al., 2011), and the pandas (The Pandas Development Team, 2024) and NumPy (Harris et al., 2020) libraries for data handling. Custom implementations of NDCG@10 calculations were developed in both R and Python environments (https://github.com/queenoa/Quinoa-genomic-prediction).

3 | RESULTS AND DISCUSSION

The seven agronomic traits evaluated revealed substantial phenotypic variation and distinct distribution patterns between Australia and Pakistan (Figure 1). Phenological traits showed the most pronounced location-specific responses. Days to flowering demonstrated a 41-day mean difference between locations, with trials in Australia flowering in approximately half the time (~46.5 days after sowing) compared to those in Pakistan (~87.1 days after sowing). Despite this substantial difference in location means, days to flowering showed a moderate correlation between the two locations ($r = 0.41$; central band of Figure 2). Days to harvest exhibited a similar pattern, with a moderate correlation across locations ($r = 0.39$). However, the population reached harvest maturity on average approximately 29 days earlier in Australia (mean = 115.4 days) than in Pakistan (mean = 144.0 days), indicating a consistent environmental acceleration of developmental processes under Australian environmental conditions. Days to flowering and harvest showed moderate within-location trait correlations (0.38 and 0.48 in Australia and Pakistan, respectively; Figure 2), indicating variation in the interval between these two phenological stages.

3.1 | Phenotypic variation across locations

The observed environmental responses likely reflect complex interactions among multiple environmental factors, including temperature, photoperiod, water availability, soil conditions, biotic stressors, and management practices, rather than the influence of any single factor. The 41-day flowering difference between locations is likely owing to photoperiod variation (with both growing seasons occurring within short-day conditions but differing by 1–2 h); beyond responses to

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

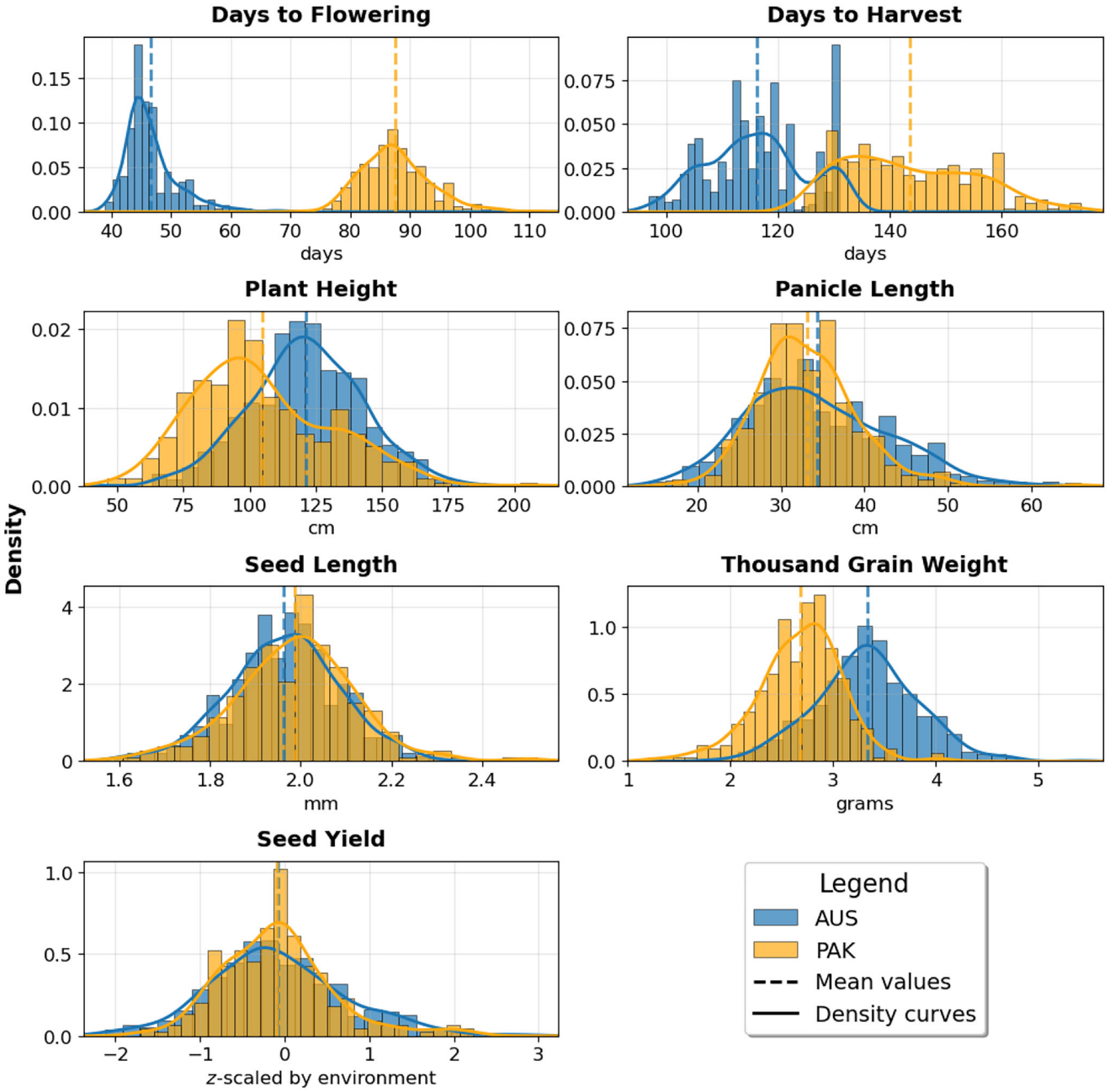

FIGURE 1 Distribution of phenotypic trait values (best linear unbiased estimators [BLUEs]) across locations. Histogram density plots showing the distribution of seven agronomic traits measured in Australia (AUS, blue) and Pakistan (PAK, yellow). Each panel presents the frequency distribution (histogram bars) and density estimate (solid curves) for each trait prior to *z*-scaling as used for genomic prediction. Dashed vertical lines show mean values for each location. Traits include days to flowering, days to harvest, plant height (cm), panicle length (cm), seed yield (g), seed length (mm), and thousand-grain weight (g). Seed yield data are shown already as *z*-scaled at the location–year to ensure comparability across trials conducted using different plot sizes or measurement methods (total yield per plot in Australia vs. yield per plant in Pakistan). Overlapping distributions facilitate visual comparison of trait ranges, phenotypic variation, central tendencies, and distributional shapes between the two geographic locations.

temperature (García-Parra et al., 2020), frost exposure (Bois et al., 2006; Jacobsen et al., 2005), water availability (Geerts et al., 2008), and nitrogen fertilization (Deng et al., 2024). Field observations in the Pakistani environments revealed that temperatures exceeding 35°C during grain filling substantially reduced both seed yield and seed quality, leading to lower germination rates in the subsequent season, thereby highlighting the importance of heat tolerance for quinoa adaptation in subtropical production systems.

Morphological traits exhibited more moderate but distinct environmental responses. Plant height showed the greatest environmental stability with the highest cross-location corre-

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

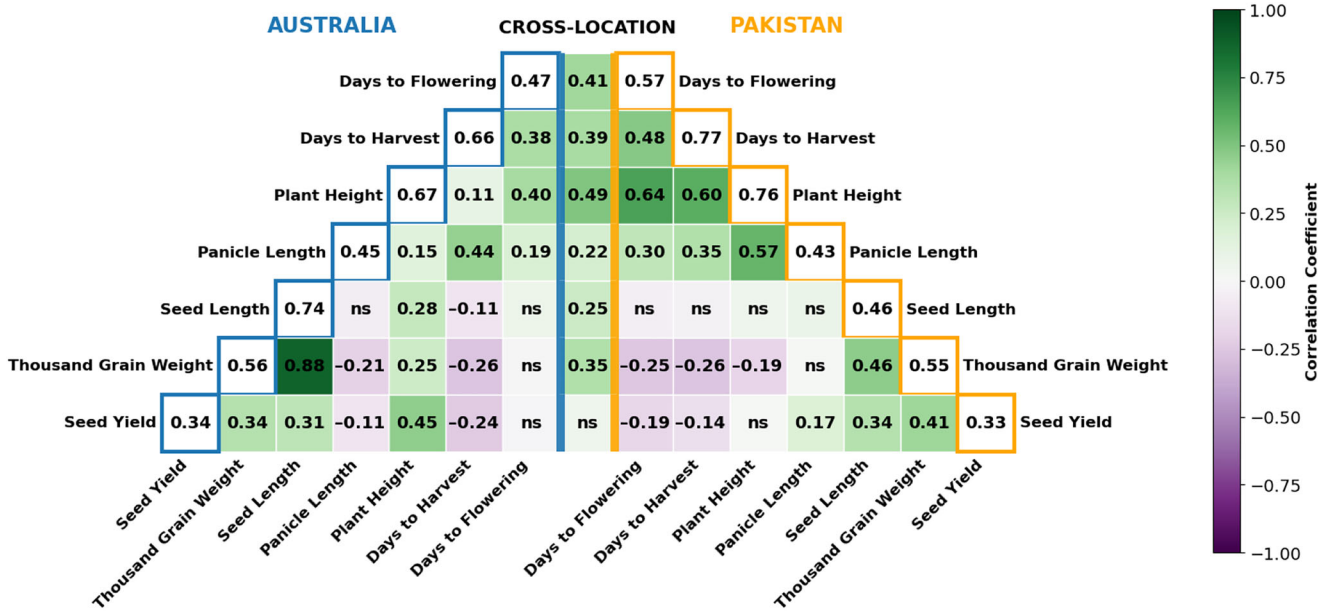

FIGURE 2 Integrated correlation matrix showing within-location trait relationships, generalized heritabilities, and cross-location correlations for seven agronomic traits evaluated in Australia and Pakistan. The pyramid-shaped heatmap displays Pearson correlation coefficients among accession best linear unbiased estimators (BLUEs), with correlations from Australia trials shown on the left side of the triangle, correlations from Pakistan trials on the right side, and cross-location correlations in the central vertical band ($n = 265–298$ overlapping genotypes per trait). Cell colors indicate correlation strength and direction according to the scale bar, with green indicating positive correlations and purple indicating negative correlations. Correlation coefficients are displayed for statistically significant relationships ($p < 0.05$), while nonsignificant correlations are labeled as "ns." Diagonal cells outlined in blue and orange represent generalized heritability estimates (h^2) for each trait at each location, indicating the proportion of phenotypic variance attributable to genetic factors (range: 0–1).

lation among all traits ($r = 0.49$). Despite mean differences between Australia (124.5 cm) and Pakistan (101.1 cm), both locations exhibited similar wide ranges of approximately 50–210 cm. Conversely, panicle length demonstrated nearly identical mean values across locations (Australia: 34.2 cm; Pakistan: 33.0 cm) but a substantially lower cross-location correlation ($r = 0.22$), suggesting differential genotypic performance under contrasting environmental conditions. The moderate heritability of panicle length (0.45 in Australia and 0.43 in Pakistan) suggests nongenetic factors influencing this trait. Notably, although plant height showed high heritability in both locations (0.67–0.76), its correlation with panicle length differed markedly between Australia ($r = 0.15$) and Pakistan ($r = 0.57$), suggesting that environmental conditions modulate the expression of genetic associations between these morphological traits.

Seed-related traits showed varying degrees of environment-specific responses. Thousand-grain weight exhibited the strongest environmental differentiation, with mean seed weight in Australia averaging 3.26 g in Australia compared to 2.65 g in Pakistan, showing minimal distributional overlap. Despite this substantial difference, thousand-grain weight showed the highest cross-location correlation among seed traits ($r = 0.35$) and moderate-to-high heritability ($h^2 = 0.55–0.56$), suggesting that environmental factors primarily influence mean differences rather than G × E.

Seed length showed contrasting location-specific patterns despite similar mean values (Australia: 1.96 mm, Pakistan: 1.99 mm). In Australia, seed length was highly heritable ($h^2 = 0.74$) and strongly correlated with thousand-grain weight ($r = 0.88$), whereas in Pakistan, it showed lower heritability ($h^2 = 0.46$), a more moderate correlation with thousand-grain weight ($r = 0.46$), and no significant correlations with other traits. This discrepancy likely reflects differences in measurement precision, as Australia used automated seed scanning while Pakistan relied on manual measurements. The lower heritability and weaker trait correlations observed in Pakistan are consistent with increased measurement error reducing the proportion of genetic variance.

Yield components (seed yield and thousand-grain weight) were negatively correlated with phenological traits ($r = −0.14$ to $−0.26$), indicating that earlier flowering and maturity are associated with higher yields and larger seed size in these environments. Thousand-grain weight is particularly valuable because it is associated with both overall yield and seed size, an important end-use quality trait for quinoa marketed as a whole grain. This favorable relationship facilitates breeding objectives by indicating that selection for shorter crop cycles and increased productivity can be pursued simul-

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

taneously without creating conflicting selection pressures. This pattern contrasts with many crops where later maturity extends the growing period and typically leads to higher yield, creating challenges in breeding early- and high-yielding lines. In the present quinoa trials, however, this correlation likely reflects the adaptation context of the tested germplasm. Late-flowering accessions are more likely to experience terminal heat stress during grain filling, while accessions with poorly matched photoperiod or thermoperiod requirements often flower late or incompletely. These factors likely penalize late phenology through reproductive failure rather than any inherent disadvantage of a longer growth cycle. Consequently, the association between earliness and yield may differ in environments better suited to later-maturing accessions. Nevertheless, in subtropical and similar environments, the ability to select for early maturity while improving both yield and seed quality represents a significant advantage for breeding programs targeting commercial markets where large seed size is highly desirable.

3.2 | Prediction of new accessions (CV1)

The moderate-to-high heritability observed across traits (h^2 = 0.33–0.77) and the distinct environmental responses described above provided a foundation for evaluating genomic prediction approaches. One of the most important scenarios for evaluating GS as a viable tool in quinoa breeding programs is the CV1 cross-validation strategy, which evaluates the ability to predict new breeding lines. Under CV1, unseen accessions are predicted by grouping cross-validation folds based on accession, such that when an accession is assigned to the testing fold, all observations across all locations and trials are excluded from the training data. In this study, genomic prediction was applied to within-location accession ranking, which was facilitated by standardizing trait data within field trials.

Given the varying consistency of trait distributions and cross-location correlations, we expected prediction performance to differ among traits and methods. Developmental traits showed comparable prediction performance across locations. Days to flowering and days to harvest showed strong ranking performance, with NDCG@10 of 0.71–0.82 across all four models (Figure 3). Since NDCG@10 scores reflect the accurate selection of accessions with high breeding value, these results are promising. However, days to flowering showed lower prediction accuracy across the entire distribution (Pearson: 0.19–0.27) than days to harvest (Pearson: 0.26–0.37). For days to harvest, Australia showed higher prediction accuracy based on Pearson correlations than Pakistan, while the opposite trend was observed for selection ranking (NDCG@10), indicating that top-performing accessions could be selected more reliably in Pakistan across all models. One explanation for the greater predictability of early-maturing plants may be a truncated phenotypic distribution near the end of the growing season. Nevertheless, these findings suggest that early-maturing accessions can be reliably selected for despite year-to-year variations and genotype-by-environment interactions affecting the ranking accuracy among remaining tested accessions.

Plant height showed the most similar mean prediction accuracy across locations (Pearson: 0.30–0.35 in Australia and 0.30–0.34 in Pakistan), consistent with it having the highest cross-location correlation among all traits (r = 0.49). Selection ranking for this trait was slightly higher in Pakistan (NDCG@10: ∼ 0.70 in AUS; ∼ 0.78 in PAK). In contrast, panicle length showed near-zero prediction accuracy in Pakistan (Pearson: 0.03–0.06), while prediction accuracies in Australia were comparable to those for plant height (Pearson: 0.26–0.32). Notably, the two traits were weakly correlated in Australia but strongly correlated in Pakistan (r = 0.15 vs. 0.57). These findings suggest that plant height and panicle length are governed by different genetic factors and that panicle length may exhibit environmental plasticity that affects its expression differently in Pakistan. The similar generalized heritability estimates across locations (h^2 = 0.45 in Australia and 0.43 in Pakistan) argue against inaccurate phenotyping and instead support a true genotype-by-environment interaction or differences in polygenic architecture between locations. In Australia, panicle length was also negatively correlated with thousand-grain weight and seed yield (r = −0.21 and −0.11), suggesting that this trait may represent a useful breeding target in this environment. The opposite trait correlations across locations would require different selection criteria, with shorter favored in Australia and longer favored in Pakistan, as well as consideration of ranking-based evaluation when selecting the best-performing model for making selection decisions.

The largest location-dependent differences in prediction accuracy were observed for seed traits. In Australia, both seed length and thousand-grain weight showed robust prediction accuracies (Pearson: ∼ 0.41 and 0.37), while prediction accuracy for these traits was very low in Pakistan (Pearson: ∼ 0.09 and 0.11). This difference is likely caused by variation in phenotyping precision, as manual seed measurements were used in Pakistan compared to image-based phenotyping in Australia. This pattern highlights the importance of high-precision and standardized phenotyping methods, as described by Stanschewski et al. (2021). The performance gap between locations is narrower when models are evaluated for selecting accessions with high breeding value (NDCG@10: 0.73–0.77 in Australia and 0.62–0.65 in Pakistan), suggesting that these models may still be useful for guiding breeding decisions in Pakistan.

The differential performance patterns across traits and locations highlight the complex interplay among genetic architecture, environmental sensitivity, and phenotyping precision in determining genomic prediction success in quinoa

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

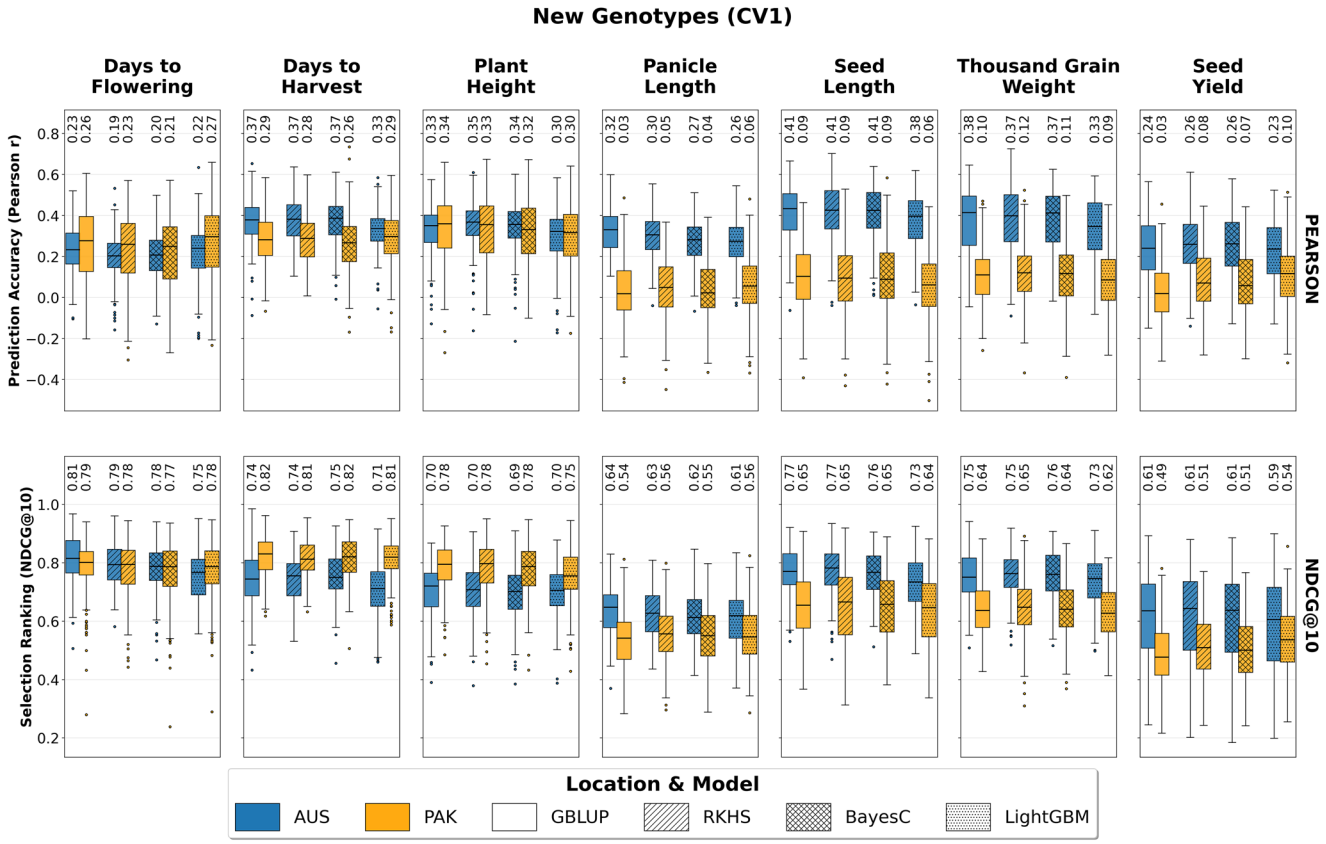

FIGURE 3 Comparison of genomic prediction accuracy for new genotypes under CV1. Boxplots show the distribution of prediction accuracy for seven agronomic traits evaluated separately in Australia (AUS) (blue) and Pakistan (PAK) (yellow) using four prediction models: genomic best linear unbiased prediction (GBLUP) (solid fill), reproducing kernel Hilbert space (RKHS) (diagonal hatching), BayesC (cross hatching), and light gradient boosting machine (LightGBM) (dotted fill). Rows represent Pearson correlation (top) and normalized discounted cumulative gain at rank 10 (NDCG@10; bottom), whereas columns represent traits. Within each panel, boxplots are grouped by model, with the two locations shown side by side within each model group. Mean values are displayed above each boxplot. Each boxplot summarizes prediction accuracy distributions from cross-validation (5 folds × 15 iterations), evaluated separately per location–year, with genotypes assigned to folds. Pearson correlation and NDCG@10 are presented on separate *y*-axes and should be interpreted using different scales (−1 to 1 and 0 to 1, respectively).

breeding programs. The prediction accuracies achieved in this study establish GS as a viable tool for quinoa improvement. Although direct comparisons across species are challenging due to differences in populations, environments, and genomic resources, the observed prediction accuracy for seed yield (Pearson correlations: 0.03–0.26) fall within ranges reported for major cereals: wheat (0.13–0.50; Gill et al., 2021; Poland et al., 2012), maize (0.10–0.36; Washburn et al., 2025), and rice (0.01–0.31; Spindel et al., 2015). This alignment suggests that quinoa breeding programs be able to realize commensurate value from genomic prediction as in established cereal crops.

3.3 | Sparse testing and leveraging environmental connectivity (CV2)

Sparse testing is a highly relevant scenario for breeding applications in which not all accessions are grown in every year or location, as in our trials (Table 1). Genomic prediction models enable the integration of information across location–years to produce more accurate genetic estimates (e.g., breeding values) for tested accessions. This strategy reduces resource requirements while leveraging information across environments to improve prediction accuracy (Jarquin et al., 2020).

As expected, under the CV2 scenario (Figure 4), we observed similar trait- and location-specific patterns in prediction accuracy as in CV1, but with generally higher accuracies across most traits and models. Notably, the two poorest-performing traits in both scenarios, panicle length and seed yield, showed only marginal improvement from the additional information provided via sparse testing. Prediction accuracies were therefore broadly similar between the two schemes for both evaluation methods. For example, Pearson correlations for seed yield ranged from 0.03 to 0.26 across models under CV1 and 0.09 to 0.26 under CV2, while panicle length in Australia ranged from 0.26 to 0.32 under CV1 compared to 0.30–0.33 under CV2. Only Pakistan showed an improvement in prediction accuracy for panicle

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

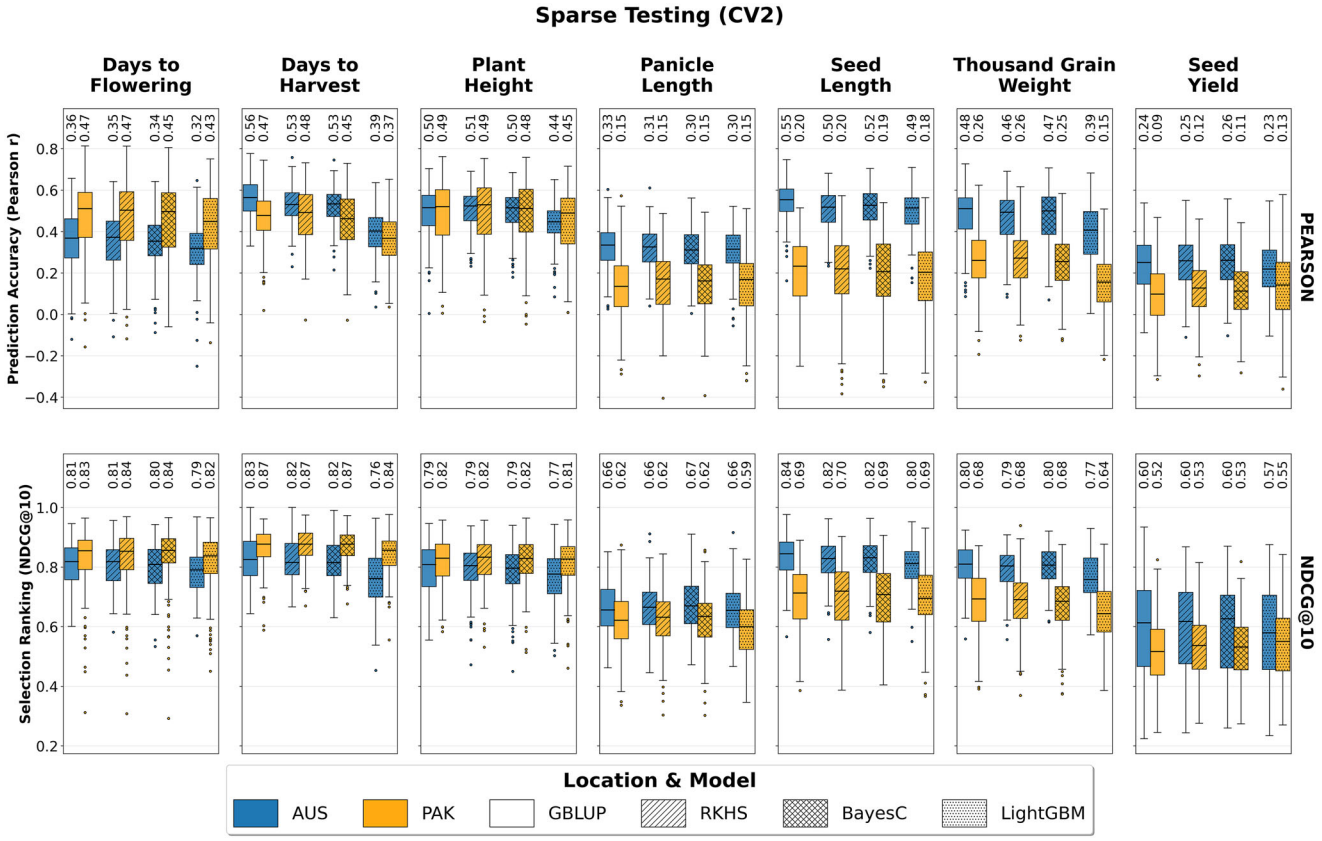

FIGURE 4 Genomic prediction accuracy under sparse testing (CV2). Figure layout, models, and visual encoding follow the same conventions as the CV1 (Figure 3). Each boxplot summarizes accuracy estimates from fivefold cross-validation over 15 iterations, evaluated per location–year, with individual genotype × location–year observations randomly masked (stratified by location–year) to simulate incomplete phenotyping across environments. AUS, Australia; GBLUP, genomic best linear unbiased prediction; PAK, Pakistan; RKHS, reproducing kernel Hilbert space.

length under CV2, increasing from 0.03–0.06 to 0.15. In contrast, selection based on the highest GEBVs for seed yield showed slightly lower accuracy under CV2 when comparing the best-performing models (NDCG@10: 0.49–0.61 in CV1 vs. 0.52–0.60 in CV2). Consistent with the low heritability of this trait in both locations (h^2 = 0.33 and 0.34), this suggests that seed yield is strongly influenced by nongenetic factors to the extent that performance of the same accession in another trials adds limited additional information. To better leverage information from multiple environments, modeling genotype-by-environment interaction is crucial, as models incorporating G × E consistently outperform others by explaining a substantial portion of otherwise unexplained variance and enabling information sharing across similar environments (Jarquin et al., 2020; Jarquín et al., 2017; Jighly et al., 2021). As additional multi-environment quinoa trials become available (germinate.quinoadb.org; Stanschewski et al., 2021), incorporating genotype-by-environment modeling and environmental covariates will be a priority for improving prediction accuracy and supporting more reliable selection decisions.

Generally, the largest improvements under CV2 were observed in prediction accuracy across the full trait distribution based on Pearson correlations, while the ranking of individuals with the highest breeding values was less affected. This suggests that, within locations, genomic relationships captured under CV1 already extract most of the available genetic signal, and that additional information from observing the same accession in another trial provides limited benefit when within-location trial-to-trial correlations are modest (Figure S1). The two traits showing the least improvement under sparse testing (seed yield and panicle length) exhibited low correlations between years within locations (ns–0.27), while the traits with the greatest differences between CV1 and CV2 schemes, such as days to flowering (0.34–0.54), days to harvest (0.40–0.81), and plant height (0.24–0.59), showed higher between-year correlations. This pattern aligns with the broader observation that sparse phenotyping yields the greatest gains when expansion environments are positively correlated with the fully phenotyped environments, but provides limited benefit when between-environment correlations are weak (He et al., 2021).

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

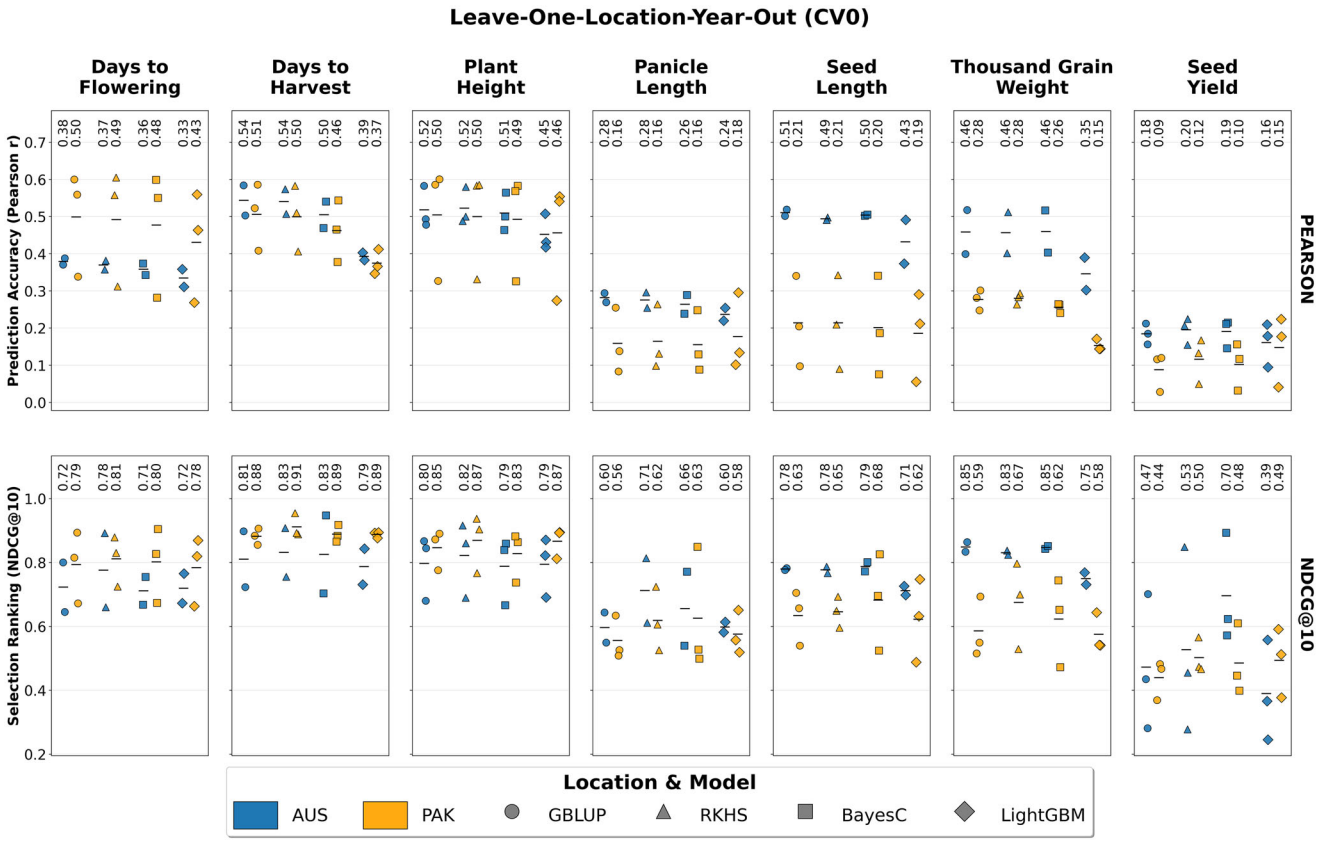

FIGURE 5 Genomic prediction accuracy under leave-one-location–year-out cross-validation. Dot plots show prediction accuracy for seven traits evaluated in Australia (AUS) (blue) and Pakistan (PAK) (yellow) across three different years using four models: genomic best linear unbiased prediction (GBLUP) (circles), reproducing kernel Hilbert space (RKHS) (triangles), BayesC (squares), and light gradient boosting machine (LightGBM) (diamonds). Rows represent Pearson correlation (top) and NDCG@10 (bottom), whereas columns represent traits. Each point represents accuracy when one location–year was held out for prediction, and the remaining five were used for training. Horizontal lines and values above each group indicate the overall accuracy.

3.4 | Prediction across environments (CV0 and cross-location)

To test the prediction of new environments, we implemented CV0, which holds out one entire location–year while training on the remaining five (Figure 5), and a cross-location scheme, which holds out an entire location and trains on the other (Figure 6). Since the held-out environments in both CV0 and cross-location contain a mix of accessions with and without phenotypic records in the training environments, these schemes also include some more challenging cases of unseen genotypes in unseen environments. This structure reflects the operational reality faced by breeders deploying genomic prediction in target environments. While CV1 and CV2 evaluate within-environment ranking ability, which is the key target for selecting among test lines within a target environment, CV0 and cross-location evaluate whether accession rankings transfer across environments. Since phenotypes were z-score standardized within each location–year, none of these schemes evaluate the ability to predict absolute performance in a new environment, which would require modeling approaches beyond the scope of this study.

Plant height showed the highest cross-location correlation ($r = 0.49$, Figure 2) and the highest cross-location prediction accuracy (Pearson: 0.36 – 0.43; NDCG@10: 0.78 – 0.86). Days to flowering and days to harvest, which had the next highest cross-location correlations, also showed moderate prediction accuracies across locations (Pearson: 0.23–0.36 and 0.29–0.31, respectively). Seed yield showed the lowest prediction accuracies and was also the trait with a non-significant cross-location correlation. These trait patterns were consistent across both cross-validation schemes. One location–year in Pakistan, the 2019–2020 trial, showed consistently lower prediction accuracy than other environments for most traits, including days to flowering, days to harvest, plant height, seed length, and seed yield, compared to other years at the same location. This pattern suggests substantial year-to-year variation that could be better captured with additional years of evaluation. Seed length in Pakistan under CV0 showed particularly large variability in prediction accuracy (GBLUP Pearson correlation: AUS mean = 0.51, SD = 0.01;

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

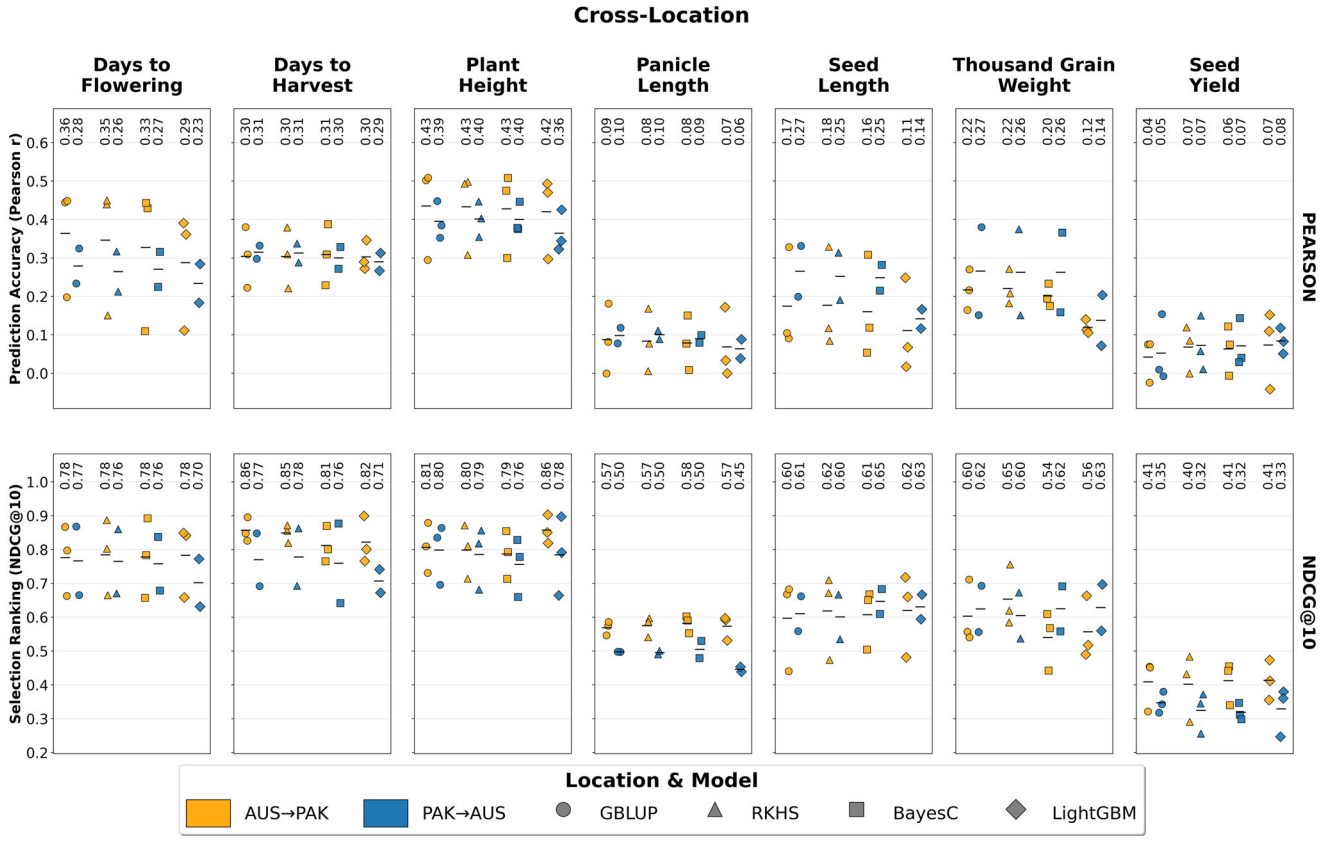

FIGURE 6 Genomic prediction accuracy for cross-location transfer. Dot plots follow the same conventions as the CV0 figure but are grouped based on transfer direction: Australia (AUS)→Pakistan (PAK) (yellow, trained in Australia, predicted in Pakistan) and PAK→AUS (blue, trained in Pakistan, predicted in Australia). Each point represents prediction accuracy for one predicted location–year at the target location. GBLUP, Genomic Best Linear Unbiased Prediction; RKHS, reproducing kernel Hilbert space.

PAK mean = 0.21, SD = 0.12). This pattern likely reflects differences in within-location trait correlations (0.67 in AUS and nonsignificant–0.22 in PAK; Figure S1).

Seed yield presented the most challenging prediction scenario across environments. After *z*-score standardization to account for differences in measurement scales, both locations exhibited comparable phenotypic variation. The near-zero cross-location correlation ($r = 0.06$) of standardized yield, combined with low within-location heritability ($h^2 = 0.33$–0.35), indicates strong location specificity in genotype ranking and differing genetic control across sites. The connectivity between trials, with 107–268 shared accessions across location–years (Table 1), provides a sufficient basis for cross-location comparison; therefore, the near-zero value is unlikely to reflect low statistical power. Per-location genetic variance estimates from independent within-location GBLUP models (Table 3) showed similarly low additive genetic variance for seed yield at both locations ($V_g = 0.05$). This finding indicates that the weak cross-location correlation is not attributed to strong location-specific genetic effects, but rather the limited genetic variance at both sites. With only two locations and six location–year combinations in total, formal estimation of genetic correlations and G × E variance components was not feasible, preventing reliable quantification of interaction effects. Consistent with the phenotypic patterns, training a genomic prediction model on one location failed to predict genotype ranking in the other location for this trait (Figure 6). Much larger datasets spanning multiple locations, the use of phenotypic values that retain between-environment

TABLE 3 Per-location genomic best linear unbiased prediction (GBLUP) variance component for seven traits in Australia (AUS) and Pakistan (PAK).

Trait	V_g (AUS)	V_g (PAK)
Days to flowering	0.16	0.16
Days to harvest	0.31	0.18
Plant height	0.16	0.15
Panicle length	0.10	0.06
Seed length	0.23	0.06
Thousand grain weight	0.17	0.08
Seed yield	0.05	0.05

Note: Models were fitted independently per location. V_g denotes additive genetic variance.

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variance rather than within-environment z-scores, the inclusion of environmental covariates, and modeling of G × E would be required before cross-location extrapolation from a global model would be feasible for complex traits like yield.

The contrasting prediction success between environments for seed-related traits illustrates how both biological and methodological factors influence genomic prediction outcomes. Beyond environmental effects on seed filling processes, differences in phenotyping precision between automated and manual measurements likely contributed to variation in model accuracy. This emphasizes that genomic prediction performance depends not only on genetic architecture and environmental stability but also on phenotyping quality and standardization (Stanschewski et al., 2021). As most trials were augmented and unreplicated, the evaluation of phenotypic data was limited to the extent that location-level BLUEs could not be estimated, and phenotypic means were used as input data for GS models. Partially replicated designs with 25% replication improve the precision of phenotypic estimates (Cullis et al., 2006) and thereby strengthen genomic prediction reliability.

Per-location variance components from per-location GBLUP models (Table 3) were used to quantify these differences in genetic signal across environments. Seed length showed a 3.8-fold higher genetic variance in Australia ($V_g = 0.23$) than that of Pakistan ($V_g = 0.06$), consistent with differences in phenotyping methods (automated image-based phenotyping in Australia vs. manual caliper measurements in Pakistan). Thousand-grain weight also showed higher genetic variance in Australia (2.1-fold difference). In contrast, days to flowering, plant height, and seed yield showed similar genetic variances between locations, indicating consistent genetic signal detection across environments for these traits.

3.5 | Within-location vs. pooled training

Low cross-location phenotypic correlation (Figure 2), variation in per-location variance components (Table 3), and cross-location prediction results (Figure 6) indicated G × E effects. Consequently, CV1 prediction accuracy from a global GBLUP model (pooled observations across locations) was compared to location-specific GBLUP models trained only on data from the target location (Figure 7).

Prediction accuracies from within-location and global models were largely similar. For most traits, within-location models showed marginally higher prediction accuracy than the global model, suggesting a limited benefit from additional cross-location training under the assumption of homogeneous breeding values across locations. Two exceptions were observed: plant height showed slightly higher Pearson correlations under the global model in both locations (+0.01), consistent with its relatively high cross-location correlation ($r = 0.49$). For thousand-grain weight in Pakistan, the global model increased NDCG@10 by 0.02, indicating that pooling improved identification of top-performing genotypes, despite no improvement in overall ranking correlation.

Collectively, these findings indicate that with our limited sampling of unique locations, pooling data across locations does not significantly improve prediction accuracy through increased sample size under a model assuming a single set of breeding values across locations. When phenotypic correlations are modest, within-location models may be preferred. Increasing the number of evaluated locations, particularly locations with similar environments, would likely lead to an increased value of global models using all locations over location-specific models.

3.6 | Implications for quinoa breeding and future directions

Here, we show the first application of genomic prediction in quinoa with strong evidence that the use of GS will have value for accelerating quinoa improvement. Discussed in more detail below, the findings of this study show strong support that the use of genomic prediction within a location has accuracy commensurate with the overall heritability of a given trait and will support selection decisions. Second, is that we found overall consistency with different models tested, which has important implications that the simpler models can be effectively implemented.

Across all four models and four CV schemes, prediction accuracies were broadly similar (median between-model range of 0.04 in Pearson; 0.03 in NDCG@10; >80% of trait × location × scheme cells within 0.10). No single method consistently outperformed the others across traits. RKHS and GBLUP produced nearly identical results across most trait-location combinations. BayesC achieved comparable performance despite using an LD-pruned marker set (~557k SNPs after pruning at $r^2 < 0.5$), indicating that the residual linkage disequilibrium after pruning is sufficient to recover the genetic value for these traits. LightGBM showed similar performance to the parametric methods when combined with hyperparameter tuning and kinship-matrix features, although with a slight tendency toward lower accuracy.

Given the near equivalence in prediction accuracy across models, computational efficiency becomes decisive. GBLUP and LightGBM converge within minutes on a laptop or workstation, whereas RKHS and BayesC implemented in BGLR were substantially slower. A single RKHS fit required approximately 15–20 min per trait, including the one-time computation of the Gaussian kernels from the full marker matrix (~1.8 million SNPs × 551 genotypes). A single BayesC fit in required approximately 18–23 h per trait, with a memory requirement of approximately 80 GB due to

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

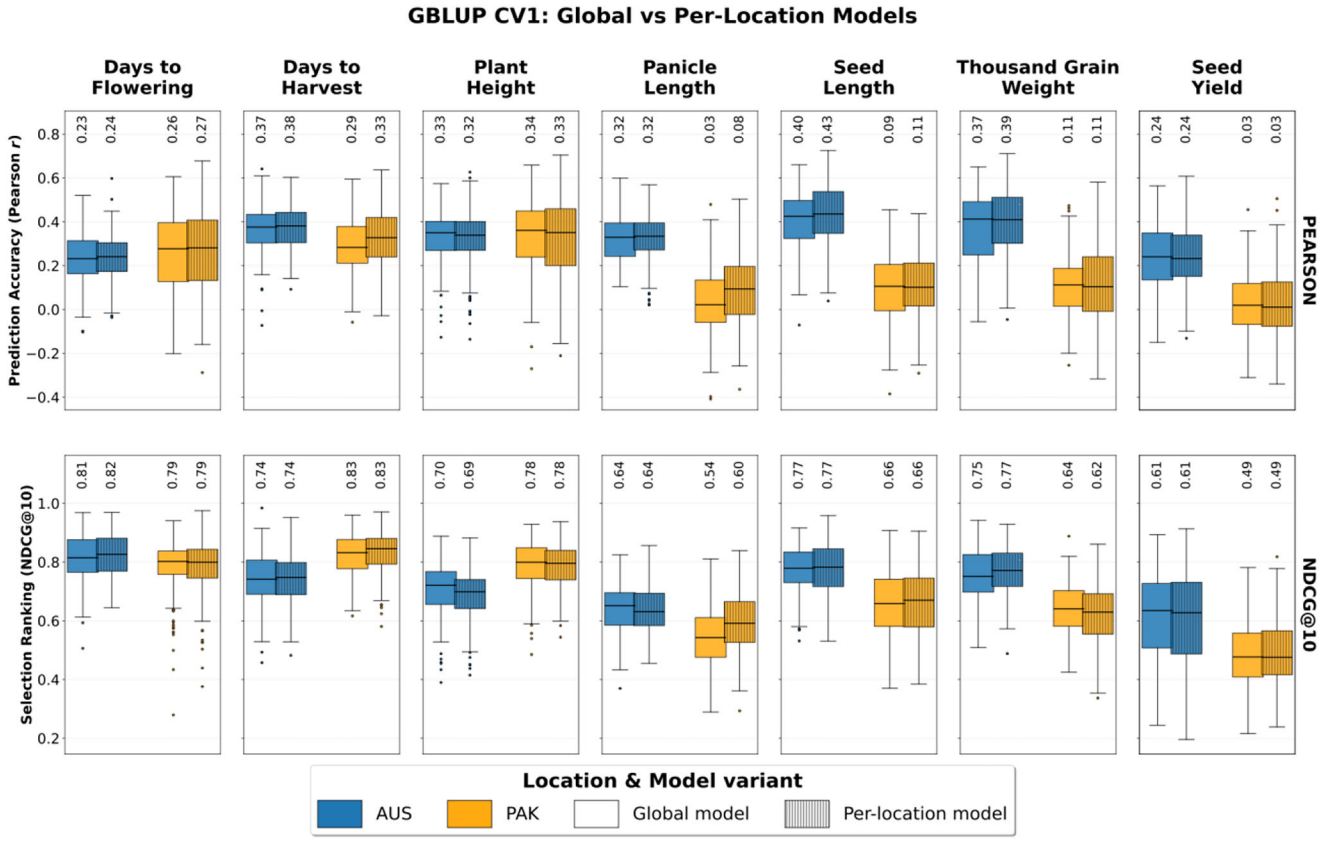

FIGURE 7 Comparison of genomic best linear unbiased prediction (GBLUP) CV1 prediction accuracy between the pooled global model and per-location models. Boxplots show the distribution of CV1 prediction accuracy for each trait evaluated separately in Australia (AUS, blue) and Pakistan (PAK, yellow) under two GBLUP modeling strategies: a pooled global model trained on combined data from both locations (solid fill) and location-specific models trained independently within each location (vertical hatching).

the dimensionality of the LD-pruned marker matrix (~557k SNPs). The validation experiment reported here involved 75 model fits per trait per CV scheme (5 folds × 15 iterations) plus the deterministic CV0 and cross-location schemes, which accumulated to approximately 3.5 h per CV1 scheme per trait for RKHS, and up to 5 days per iteration per trait for BayesC, distributed across SLURM jobs. Even on a per-fit basis, the BGLR-based models require high-performance computing infrastructure, whereas GBLUP and LightGBM can be fit on standard hardware. LightGBM was the fastest among the four models, but its prediction accuracy was consistently slightly lower than that of GBLUP across most traits and CV schemes. GBLUP therefore provides the best balance of accuracy, accessibility, and runtime for routine deployment in quinoa breeding programs, particularly in settings without high-performance computing infrastructure.

The modest Pearson correlation accuracies observed should be interpreted considering that correlations between GEBVs and phenotypic values are inherently limited because phenotypes capture genotype, environment, and G × E. Low correlations could stem from residual error in phenotypic values, while GEBVs may actually be closer to unknown true breeding values (Poland et al., 2012). This theoretical constraint is the reason why we chose to discard trial mean and scale differences entirely by z-score standardization of trait data within field trials. This emphasizes why rank-based evaluation methods like NDCG@10 provide complementary information for breeding decisions, focusing on the practical objective of identifying top-performing genotypes for selection decisions rather than precise phenotypic prediction (Blondel et al., 2015). Another recent paper has picked up the use of NDCG@10 as a complementary metric that is more relevant to breeding, in particular for the selection of best-performing models. Heinrich et al. (2025) propose a modification that removes the discounting factor for ranking within selected individuals, improves interpretability, and evaluates performance across all selection thresholds rather than relying on a pre-specified selection intensity. While this adaptation is promising, using a fixed selection intensity of approximately 15% by assessing the top 10 accessions per CV fold was appropriate for our experimental design, fold sizes, and study aims. NDCG@10 did not differentiate model performance more clearly than Pearson correlation because all models performed generally very similarly. This likely

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reflects that, after trial-wise standardization, Pearson correlation determined the accuracy of the phenotype prediction per se, essentially ranking of individuals across the entire population, while NDCG@10 focuses on how closely the selected individuals match the optimal ranking. This metric revealed for our data that the gap between accuracies in the two locations for the hardest to predict traits is narrower than appears under Pearson correlation. This finding suggests that accurately ranking the entire population remains challenging for some traits owing to the influence of the irrelevant ranking of low-performing individuals; however, the models still provide useful guidance for informing selection decisions involving those traits.

The present study establishes within-location genomic prediction as a viable selection tool across most traits in quinoa. Extending this framework requires three key developments. First, explicit modeling of G × E is needed to leverage information across environments, particularly for low-heritability traits such as seed yield, where within-location models alone show limited prediction accuracy. Second, with a sufficiently large number of locations, environmental covariates should be incorporated to support prediction performance in untested environments, consistent with envirotyping approaches that have improved cross-environment prediction accuracy in major crops (Costa-Neto et al., 2021; Winans et al., 2024; Xu, 2016). However, the choice of covariates requires careful evaluation, along with the identification of environments similar to the target environment, since additional environmental information does not guarantee to improve prediction accuracy (Widener et al., 2021). Prediction accuracy can also be improved by excluding environments with low predictability from the model (Heslot et al., 2013) or assigning different weights based on their relevance and correlation with the target population of environments (Podlich et al., 1999). Most importantly, these different strategies all depend on the availability of a large network of field trial data, alongside associated metadata and environmental information. Hence, expanding multi-environment trial data through coordinated efforts across the international quinoa research community (e.g., germinate.quinoadb.org, Stanschewski et al., 2021) will enable the identification of correlated environment networks and provide the environmental and genetic diversity required to fit and validate G × E models at scale, thereby strengthening the application of GS in quinoa breeding programs.

AUTHOR CONTRIBUTIONS

Clara S. Stanschewski: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Mark Warmington**: Investigation; resources. **Irfan Afzal**: Investigation; resources; writing—review and editing. **Elodie Rey**: Data curation; formal analysis; investigation; methodology; writing—review and editing. **Gabriele Fiene**: Investigation; methodology; project administration. **Evan Craine**: Resources; writing—review and editing. **Kevin Murphy**: Resources; writing—review and editing. **Mark Tester**: Conceptualization; funding acquisition; project administration; validation; writing—review and editing. **Jesse Poland**: Conceptualization; funding acquisition; supervision; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, including Competitive Research Grant CRG5-2966 and baseline funding to Mark Tester and Jesse Poland. We are grateful to the field research teams at Kimberley Research Station, Australia and the University of Agriculture in Faisalabad, Pakistan for their meticulous work in conducting the multi-environment trials and phenotypic data collection that made this study possible. For the Australian field trial in 2019, we additionally thank Gordon Wellman and Cristina Ocaña Gallegos for phenotyping support. We thank Karl Schmid, Sandra Schmöckel, Gordon Wellman, Noha Saber, and colleagues at the University of Hohenheim, Germany, and KAUST, Saudi Arabia for sample processing, DNA extraction, and library preparation. We thank Katharina Böndel for assistance with curation of the genomic dataset. Seed trait measurements for the AUS dataset were supported in part by the United States Department of Agriculture (award number 2016-680004-2470) and the National Science Foundation Graduate Research Fellowship (Grant No. 1842493). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Phenotypic data and genotypic data (two VCF files: full set and filtered for this study) are available at Dryad (https://doi.org/10.5061/dryad.brv15dvp1). Scripts that were used for variant calling are uploaded to https://github.com/queenoa/Quinoa-SNP-calling. Raw sequencing reads for the 610-accession quinoa diversity panel are deposited in the NCBI Sequence Read Archive under BioProjects PRJNA1357073 and the previously published subset PRJNA673789 (Patiranage et al., 2022). Detailed descriptions of germplasm, phenotyping protocols, environmental information, and metadata are available in the Quinoa Germinate Database (by the time of publication) (http://germinate.quinoadb.org) (Stanschewski et al., 2021). Seeds are available from public gene banks including IPK Gatersleben (https://www.ipk-gatersleben.de/en/genebank/) and the US National

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Plant Germplasm System (https://npgsweb.ars-grin.gov/grin-global/search). Custom scripts for data analysis and visualization are available at https://github.com/queenoa/Quinoa-genomic-prediction.

ORCID

Clara S. Stanschewski https://orcid.org/0000-0002-3596-9108
Irfan Afzal https://orcid.org/0000-0001-5942-7672
Evan Craine https://orcid.org/0000-0002-6252-8973
Kevin Murphy https://orcid.org/0000-0001-8982-3641
Mark Tester https://orcid.org/0000-0002-5085-8801
Jesse Poland https://orcid.org/0000-0002-7856-1399

REFERENCES

Abrouk, M., Ahmed, H. I., Cubry, P., Šimoníková, D., Cauet, S., Pailles, Y., Bettgenhaeuser, J., Gapa, L., Scarcelli, N., Couderc, M., Zekraoui, L., Kathiresan, N., Čížková, J., Hřibová, E., Doležel, J., Arribat, S., Bergès, H., Wieringa, J. J., Gueye, M., … Krattinger, S. G. (2020). Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-18329-4

Alandia, G., Rodriguez, J. P., Jacobsen, S. E., Bazile, D., & Condori, B. (2020). Global expansion of quinoa and challenges for the Andean region. *Global Food Security*, *26*, Article 100429. https://doi.org/10.1016/j.gfs.2020.100429

Alemu, A., Åstrand, J., Montesinos-López, O. A., Isidro y Sánchez, J., Fernández-Gónzalez, J., Tadesse, W., Vetukuri, R. R., Carlsson, A. S., Ceplitis, A., Crossa, J., Ortiz, R., & Chawade, A. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. *Molecular Plant*, *17*(4), 552–578. https://doi.org/10.1016/J.MOLP.2024.03.007

Amadeu, R. R., Garcia, A. A. F., Munoz, P. R., & Ferrão, L. F. V. (2023). AGHmatrix: Genetic relationship matrices in R. *Bioinformatics*, *39*(7), Article btad445. https://doi.org/10.1093/bioinformatics/btad445

Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., & Shiu, S. H. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3 Genes|Genomes|Genetics*, *9*(11), 3691–3702. https://doi.org/10.1534/G3.119.400498

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. K. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE*, *10*(5), Article e0128036. https://doi.org/10.1371/journal.pone.0128036

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, *48*(5), 1649–1664. https://doi.org/10.2135/CROPSCI2008.03.0131

Blondel, M., Onogi, A., Iwata, H., & Ueda, N. (2015). A ranking approach to genomic selection. *PLoS ONE*, *10*(6), Article e0128570. https://doi.org/10.1371/JOURNAL.PONE.0128570

Bodrug-Schepers, A., Stralis-Pavese, N., Buerstmayr, H., Dohm, J. C., & Himmelbauer, H. (2021). Quinoa genome assembly employing genomic variation for guided scaffolding. *Theoretical and Applied Genetics*, *134*(11), 3577–3594. https://doi.org/10.1007/S00122-021-03915-X/FIGURES/5

Bois, J. F., Winkel, T., Lhomme, J. P., Raffaillac, J. P., & Rocheteau, A. (2006). Response of some Andean cultivars of quinoa (*Chenopodium quinoa* Willd.) to temperature: Effects on germination, phenology, growth and freezing. *European Journal of Agronomy*, *25*(4), 299–308. https://doi.org/10.1016/J.EJA.2006.06.007

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bureau of Meteorology. (2025). *Climate statistics for Australian locations*. https://www.bom.gov.au/climate/averages/tables/cw_002038.shtml

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2023). *ASReml-R reference manual version 4.2*. VSN International. https://asreml.kb.vsni.co.uk/article-categories/asreml-r-resources/

Calus, M. P. L., Meuwissen, T. H. E., De Roos, A. P. W., & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, *178*(1), 553–561. https://doi.org/10.1534/GENETICS.107.080838

Charles Brummer, E., Barber, W. T., Collier, S. M., Cox, T. S., Johnson, R., Murray, S. C., Olsen, R. T., Pratt, R. C., & Thro, A. M. (2011). Plant breeding for harmony between agriculture and the environment. *Frontiers in Ecology and the Environment*, *9*(10), 561–568. https://doi.org/10.1890/100225

Chaudhary, N., Walia, S., & Kumar, R. (2023). Functional composition, physiological effect and agronomy of future food quinoa (Chenopodium quinoa Willd.): A review. *Journal of Food Composition and Analysis*, *118*, Article 105192. https://doi.org/10.1016/J.JFCA.2023.105192

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, *126*(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

Costa-Neto, G., Crossa, J., & Fritsche-Neto, R. (2021). Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. *Frontiers in Plant Science*, *12*, 717552. https://doi.org/10.3389/fpls.2021.717552

Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., Campos, G. D. L., Montesinos-López, O. A., & Burgueño, J. (2016). Genomic Prediction of Genotype × Environment Interaction Kernel Regression Models. *The Plant Genome*, *9*(3). https://doi.org/10.3835/PLANTGENOME2016.03.0024

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*(4), 381–393. https://doi.org/10.1198/108571106x154443

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, *185*(3), 1021–1031. https://doi.org/10.1534/GENETICS.110.116855

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), Article giab008. https://doi.org/10.1093/GIGASCIENCE/GIAB008

De Los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A., & Crossa, J. (2010). Semi-parametric genomic-enabled prediction

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, *92*(4), 295–308. https://doi.org/10.1017/S0016672310000285

Deng, Y., Zheng, Y., Lu, J., Guo, Z., Sun, X., Zhao, L., Guo, H., Zhang, L., & Wang, C. (2024). Effects of different N fertilizer doses on phenology, photosynthetic fluorescence, and yield of quinoa. *Agronomy*, *14*(5), Article 914. https://doi.org/10.3390/AGRONOMY14050914/S1

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Addison Wesley Longman.

Fernandes, I. K., Vieira, C. C., Dias, K. O. G., & Fernandes, S. B. (2024). Using machine learning to combine genetic and environmental data for maize grain yield predictions across multi-environment trials. *Theoretical and Applied Genetics*, *137*(8), 1–13. https://doi.org/10.1007/S00122-024-04687-W/FIGURES/4

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J., & Wilczek, A. M. (2011). A map of local adaptation in Arabidopsis thaliana. *Science*, *334*(6052), 86–89. https://doi.org/10.1126/SCIENCE.1209271

García-Parra, M. A., Roa-Acosta, D. F., Stechauner-Rohringer, R., García-Molano, F., Bazile, D., & Plazas-Leguizamón, N. (2020). Effect of temperature on the growth and development of quinoa plants (*Chenopodium quinoa* Willd.): A review on a global scale. *Sylwan*, *164*(5), 411–433.

Geerts, S., Raes, D., Garcia, M., Mendoza, J., & Huanca, R. (2008). Crop water use indicators to quantify the flexible phenology of quinoa (*Chenopodium quinoa* Willd.) in response to drought stress. *Field Crops Research*, *108*(2), 150–156. https://doi.org/10.1016/J.FCR.2008.04.008

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media, Inc. https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/

Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, *173*(3), 1761–1776. https://doi.org/10.1534/GENETICS.105.049510

Gill, H. S., Halder, J., Zhang, J., Brar, N. K., Rai, T. S., Hall, C., Bernardo, A., Amand, P. S., Bai, G., Olson, E., Ali, S., Turnipseed, B., & Sehgal, S. K. (2021). Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. *Frontiers in Plant Science*, *12*, Article 709545. https://doi.org/10.3389/fpls.2021.709545

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, *12*(1), Article 186. https://doi.org/10.1186/1471-2105-12-186

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/S41586-020-2649-2

Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). *Methods in Molecular Biology*, *1019*, 149–169. https://doi.org/10.1007/978-1-62703-447-0_6

He, S., Jiang, Y., Thistlethwaite, R., Hayden, M. J., Trethowan, R., & Daetwyler, H. D. (2021). Improving selection efficiency of crop breeding with genomic prediction aided sparse phenotyping. *Frontiers in Plant Science*, *12*, Article 735285. https://doi.org/10.3389/FPLS.2021.735285/TEXT

Heffner, E. L., Sorrells, M. E., & Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Science*, *49*(1), 1–12. https://doi.org/10.2135/CROPSCI2008.08.0512

Heinrich, F., Lange, T. M., Ramzan, F., Gültas, M., & Schmitt, A. O. (2025). Normalized cumulative gain as an alternative evaluation measure for genomic selection models. *Genetics Selection Evolution*, *57*(1), Article 70. https://doi.org/10.1186/S12711-025-01022-9

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2013). Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science*, *53*(3), 921–933. https://doi.org/10.2135/CROPSCI2012.07.0420

Jacobsen, S. E. (2003). The worldwide potential for quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International*, *19*(1–2), 167–177. https://doi.org/10.1081/FRI-120018883

Jacobsen, S. E., Monteros, C., Christiansen, J. L., Bravo, L. A., Corcuera, L. J., & Mujica, A. (2005). Plant responses of quinoa (*Chenopodium quinoa* Willd.) to frost at various phenological stages. *European Journal of Agronomy*, *22*(2), 131–139. https://doi.org/10.1016/J.EJA.2004.01.003

Jarquin, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W. R., Covarrubias Pazaran, G., Burgueño, J., Pacheco, A., Grondona, M., Wimmer, V., & Prasanna, B. M. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. *G3 Genes|Genomes|Genetics*, *10*(8), 2725–2739. https://doi.org/10.1534/G3.120.401349

Jarquín, D., Silva, C. L., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., Battenfield, S., & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *The Plant Genome*, *10*(2). https://doi.org/10.3835/PLANTGENOME2016.12.0130

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., Ohyanagi, H., Mineta, K., Michell, C. T., Saber, N., Kharbatia, N. M., Rupper, R. R., Sharp, A. R., Dally, N., Boughton, B. A., Woo, Y. H., Gao, G., Schijlen, E. G. W. M., Guo, X., … Tester, M. (2017). The genome of *Chenopodium quinoa*. *Nature*, *542*(7641), 307–312. https://doi.org/10.1038/nature21370

Jighly, A., Hayden, M., & Daetwyler, H. (2021). Integrating genomic selection with a genotype plus genotype x environment (GGE) model improves prediction accuracy and computational efficiency. *Plant, Cell & Environment*, *44*(10), 3459–3470. https://doi.org/10.1111/PCE.14145

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146–3154. https://github.com/Microsoft/LightGBM

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(1), Article 29. https://doi.org/10.1186/1746-4811-9-29

Lagergren, J., Cashman, M., Vergara, V. G. M., Eller, P. R., Gazolla, J. G. F. M., Chhetri, H. B., Streich, J., Climer, S., Thornton, P., Joubert,

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

W., & Jacobson, D. (2023). Climatic clustering and longitudinal analysis with impacts on food, bioenergy, and pandemics. *Phytobiomes Journal*, *7*(1), 65–77. https://doi.org/10.1094/PBIOMES-02-22-0007-R

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Maldonado-Taipe, N., Barbier, F., Schmid, K., Jung, C., & Emrani, N. (2022). High-density mapping of quantitative trait loci controlling agronomically important traits in quinoa (*Chenopodium quinoa* Willd.). *Frontiers in Plant Science*, *13*, Article 916067. https://doi.org/10.3389/FPLS.2022.916067

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/GR.107524.110

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/GENETICS/157.4.1819

Morgante, F., Huang, W., Maltecca, C., & Mackay, T. F. C. (2018). Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity*, *120*(6), 500–514. https://doi.org/10.1038/s41437-017-0043-0

Murphy, K. M., & Matanguihan, J. B. (2015). *Quinoa: Improvement and sustainable production*. John Wiley & Sons. https://doi.org/10.1002/9781118628041

Murphy, K. M., Matanguihan, J. B., Fuentes, F. F., Gómez-pando, L. R., Jellen, E. N., Maughan, P. J., & Jarvis, D. E. (2019). Quinoa breeding and genomics. *Plant Breeding Reviews*, *42*, 257–320. https://doi.org/10.1002/9781119521358.CH7

Nakamichi, N. (2015). Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant and Cell Physiology*, *56*(4), 594–604. https://doi.org/10.1093/PCP/PCU181

Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3 Genes|Genomes|Genetics*, *8*(9), 2889–2899. https://doi.org/10.1534/G3.118.200311

Patiranage, D. S. R., Rey, E., Emrani, N., Wellman, G., Schmid, K., Schmöckel, S. M., Tester, M., & Jung, C. (2022). Genome-wide association study in quinoa reveals selection pattern typical for crops with a short breeding history. *Elife*, *11*, Article e66873. https://doi.org/10.7554/ELIFE.66873

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pérez, P., & De Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, *198*(2), 483–495. https://doi.org/10.1534/GENETICS.114.164442

Podlich, D. W., Cooper, M., Basford, K. E., & Geiger, H. H. (1999). Computer simulation of a selection strategy to accommodate genotype environment interactions in a wheat recurrent selection programme. *Plant Breeding*, *118*(1), 17–28. https://doi.org/10.1046/J.1439-0523.1999.118001017.X

Poland, J. (2015). Breeding-assisted genomics. *Current Opinion in Plant Biology*, *24*, 119–124. https://doi.org/10.1016/J.PBI.2015.02.009

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., & Jannink, J.-L. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, *5*(3), 103–113. https://doi.org/10.3835/PLANTGENOME2012.06.0006

Rey, E., Abrouk, M., Dufau, I., Rodde, N., Saber, N., Cizkova, J., Fiene, G., Stanschewski, C., Jarvis, D. E., Jellen, E. N., Maughan, P. J., von Baer, I., Troukhan, M., Kravchuk, M., Hribova, E., Cauet, S., Krattinger, S. G., & Tester, M. (2024). Genome assembly of a diversity panel of Chenopodium quinoa. *Scientific Data*, *11*(1), 1–10. https://doi.org/10.1038/s41597-024-04200-4

Rey, E., Maughan, P. J., Maumus, F., Lewis, D., Wilson, L., Fuller, J., Schmöckel, S. M., Jellen, E. N., Tester, M., & Jarvis, D. E. (2023). A chromosome-scale assembly of the quinoa genome provides insights into the structure and dynamics of its subgenomes. *Communications Biology*, *6*(1), 1–14. https://doi.org/10.1038/s42003-023-05613-4

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J. L., & McCouch, S. R. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLOS Genetics*, *11*(2), Article e1004982. https://doi.org/10.1371/JOURNAL.PGEN.1004982

Stanschewski, C. S., Rey, E., Fiene, G., Craine, E. B., Wellman, G., Melino, V. J., Patiranage, D. S. R., Johansen, K., Schmöckel, S. M., Bertero, D., Oakey, H., Colque-Little, C., Afzal, I., Raubach, S., Miller, N., Streich, J., Buchvaldt Amby, D., Emrani, N., Warmington, M., … The Quinoa Phenotyping Consortium. (2021). Quinoa phenotyping methodologies: An international consensus. *Plants*, *10*(9), Article 1759. https://doi.org/10.3390/PLANTS10091759

The Pandas Development Team. (2024). *pandas-dev/pandas: Pandas (Version 2.2.3)*. Zenodo. https://doi.org/10.5281/ZENODO.13819579

Van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Science*, *56*(5), 2119–2140. https://doi.org/10.2135/CROPSCI2015.06.0375

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. https://doi.org/10.3168/JDS.2007-0980

Washburn, J. D., Varela, J. I., Xavier, A., Chen, Q., Ertl, D., Gage, J. L., Holland, J. B., Lima, D. C., Romay, M. C., Lopez-Cruz, M., de los Campos, G., Barber, W., Zimmer, C., Trucillo Silva, I., Rocha, F., Rincent, R., Ali, B., Hu, H., Runcie, D. E., … de Leon, N. (2025). Global genotype by environment prediction competition reveals that diverse modeling strategies can deliver satisfactory maize yield estimates. *Genetics*, *229*(2), Article 27. https://doi.org/10.1093/GENETICS/IYAE195

Weather Atlas. (2025). *December weather—winter 2025—Faisalabad, Pakistan*. Weather Atlas. https://www.weather-atlas.com/en/pakistan/faisalabad-weather-december

Westhues, C. C., Mahone, G. S., da Silva, S., Thorwarth, P., Schmidt, M., Richter, J. C., Simianer, H., & Beissinger, T. M. (2021). Prediction of maize phenotypic traits with genomic and environmental predictors

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

using gradient boosting frameworks. *Frontiers in Plant Science*, *12*, Article 699589. https://doi.org/10.3389/FPLS.2021.699589

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2025). *dplyr: A grammar of data manipulation (Version 1.1.4.9000)*. Comprehensive R Archive Network.

Widener, S., Graef, G., Lipka, A. E., & Jarquin, D. (2021). An assessment of the factors influencing the prediction accuracy of genomic prediction models across multiple environments. *Frontiers in Genetics*, *12*, Article 689319. https://doi.org/10.3389/FGENE.2021.689319/TEXT

Winans, N. D., Fonseca, J. M. O., Perumal, R., Klein, P. E., Klein, R. R., & Rooney, W. L. (2024). Envirotyping can increase genomic prediction accuracy of new environments in grain sorghum trials depending on mega-environment. *Crop Science*, *64*(5), 2519–2533. https://doi.org/10.1002/csc2.21213

Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, *129*(4), 653–673. https://doi.org/10.1007/s00122-016-2691-5

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B. M., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany*, *68*(11), 2641–2666. https://doi.org/10.1093/JXB/ERX135

Yasui, Y., Hirakawa, H., Oikawa, T., Toyoshima, M., Matsuzaki, C., Ueno, M., Mizuno, N., Nagatoshi, Y., Imamura, T., Miyago, M., Tanaka, K., Mise, K., Tanaka, T., Mizukoshi, H., Mori, M., & Fujita, Y. (2016). Draft genome sequence of an inbred line of *Chenopodium quinoa*, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Research*, *23*(6), 535–546. https://doi.org/10.1093/DNARES/DSW037

Zou, C., Chen, A., Xiao, L., Muller, H. M., Ache, P., Haberer, G., Zhang, M., Jia, W., Deng, P., Huang, R., Lang, D., Li, F., Zhan, D., Wu, X., Zhang, H., Bohm, J., Liu, R., Shabala, S., Hedrich, R., … Zhang, H. (2017). A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Research*, *27*(11), 1327–1340. https://doi.org/10.1038/cr.2017.124

Zurita-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S. E., & Schwember, A. R. (2014). Breeding quinoa (*Chenopodium quinoa* Willd.): Potential and perspectives. *Molecular Breeding*, *34*(1), 13–30. https://doi.org/10.1007/S11032-014-0023-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Stanschewski, C. S., Warmington, M., Afzal, I., Rey, E., Fiene, G., Craine, E., Murphy, K., Tester, M., & Poland, J. (2026). Genomic prediction in quinoa across contrasting environments using statistical and machine learning models. *The Plant Genome*, *19*, e70277. https://doi.org/10.1002/tpg2.70277

19403372, 2026, 3, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/tpg2.70277 by King Abdullah Univ. Of Science & Tech Kaust, Wiley Online Library on [20/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License