

RESEARCH

Open Access



Enhancing genomic prediction for key production traits in chickens through ultrasound phenotyping and multi-model comparative analysis

Ranran Zhu^{1,2}, Yuxiang Jiang^{1,2}, Wanyi Xiong^{1,2}, Yu Zhang^{1,2}, Ziyi Lian^{1,2}, Danni Gou^{1,2}, Zhandeng Li³, Xiuping Wang³ and Xuemei Deng^{1,2*}

Abstract

Background Growth performance and carcass traits are economically vital in poultry breeding. In Wenchang chickens, reducing excessive abdominal fat represents a critical breeding objective. However, as a typical carcass trait, abdominal fat thickness has traditionally been measurable only post-slaughter, resulting in inefficient and costly selection processes that hinder genetic progress for these traits. To overcome this limitation, we developed an integrated approach combining non-invasive ultrasound phenotyping and multi-model genomic selection to evaluate growth and fat-related traits in Wenchang chickens.

Results We genotyped 3,737 chickens using the “Jingxin No.1” 55K SNP array and performed longitudinal measurement of abdominal fat thickness (AFT) via ultrasound imaging. A comprehensive evaluation of genomic prediction models revealed that WGBLUP (informed by wssGWAS), and GBLUP models based on LD-pruned whole-genome sequencing (WGS) data significantly outperformed standard GBLUP, with accuracy gains of 5.25% and 6.58%–15.30%, respectively. Among the machine learning algorithms tested, kernel ridge regression (KRR) and support vector regression (SVR) achieved the highest predictive improvement (3.00%–4.15%) while maintaining superior computational efficiency, whereas ensemble methods provide no consistent advantage.

Conclusions Our work established ultrasound imaging as a scalable, non-invasive phenotyping platform for poultry breeding. Results demonstrated that integrating wssGWAS-derived biological priors with WGS data substantially improves genomic prediction accuracy for complex traits. This integration, enhanced by computationally efficient machine learning algorithms, provides a powerful and practical strategy to accelerate genetic gain.

Keywords Genomic prediction, Machine learning, Ultrasound phenotyping, Wenchang chicken, Whole-genome sequencing

*Correspondence:

Xuemei Deng
deng@cau.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Genomic selection (GS) [1] has become a cornerstone of modern animal and plant breeding, enabling the prediction of genetic values for complex traits using genome-wide markers [2–4]. By leveraging reference populations with both genotypic and phenotypic data, GS facilitates the selection of superior candidates, significantly accelerating genetic gain [5]. The accuracy of genomic prediction, however, is influenced by multiple factors, including trait heritability, reference population size, marker density, and statistical methods employed [6, 7].

The evolution of statistical methodologies has been central to advancing GS. Standard approaches such as genomic best linear unbiased prediction (GBLUP) assume an infinitesimal genetic architecture [8]. To better capture trait-specific genetic architectures, more sophisticated methods have been developed. These include weighted GBLUP (WGBLUP); genomic feature BLUP (GFBLUP), which focuses on functionally informed marker subsets; and various Bayesian methods that allow for flexible prior distributions of marker effects [9–11]. More recently, machine learning (ML) algorithms such as kernel ridge regression (KRR) and random forest (RF) have been introduced to GS, offering the potential to model complex non-linear relationships without pre-specified genetic assumptions [9, 12].

Evidence from previous studies underscores the context-dependent nature of these methods. In chicken, Romé et al. [13] demonstrated that WGBLUP improved the prediction accuracy for body weight traits by 2%–7% over GBLUP, and further gains were achieved by including top GWAS hits as fixed effects. Similarly, Li et al. [14] reported that ML methods outperformed traditional methods for certain traits in laying hens, with integration of GWAS prior boosting accuracy by up to 27%. In yellow-feathered broilers, Yang et al. [15] showed that incorporating significant QTLs into GFBLUP increased prediction accuracy for body weights by over 13%. Despite these advances, the performance of prediction models remains highly variable across traits and populations, particularly for complex characteristics like fat deposition. This inconsistency highlights a critical challenge: optimal GS strategy appears to be contingent on the underlying genetic architecture of the trait, yet a systematic framework for matching models to traits is still lacking.

Conventional measurement of carcass traits such as abdominal fat thickness relies on post-slaughter assessment, which is not only time-consuming and costly but also precludes selection in live animals. To accelerate breeding progress, researchers have developed various *in vivo* measurement techniques for predicting carcass

traits. In poultry research, ultrasound scanning has been successfully combined with live body measurements to predict abdominal fat content [16, 17] as well as other key carcass traits, such as breast muscle weight [18] and skeletal characteristics [19]. This methodological approach has also been effectively extended to other livestock species, including body fat prediction in Pekin ducks [20] and carcass chemical composition assessment in sheep [21]. These studies demonstrate that B-mode ultrasonography combined with live measurement provides a reliable and efficient means for the *in vivo* evaluation of economically important traits in poultry and livestock. Building on this technical foundation, this study aims to apply ultrasound-based *in vivo* detection to measure abdominal fat thickness in Wenchang chickens, with the objectives of evaluating its potential application in the breeding of this indigenous breed and supplying accurate phenotypic data for the establishment of an efficient genomic selection program.

To address this gap and to overcome the high costs associated with phenotyping carcass traits, we conducted a comprehensive genomic analysis in Wenchang chickens. We genotyped 3,737 individuals from two consecutive generations using the “Jingxin No.1” 55K SNP chip, followed by genotype imputation to whole-genome sequencing (WGS) density. A key innovation of our study is the integration of a novel, ultrasound-derived phenotype for abdominal fat thickness (AFT)—a trait typically measured postmortem—with a multi-faceted GS evaluation. We systematically compared the performance of: (1) conventional linear mixed models (GBLUP, ssGBLUP); (2) a prior-informed model (WGBLUP) incorporating weights from weighted single-step GWAS (wssGWAS); (3) five ML algorithms (KRR, SVR, RF, LightGBM, and AdaBoost.R2) optimized via grid search; and (4) GBLUP models based on GBLUP LD-pruned WGS data to assess the impact of marker density.

Through rigorous 10-replicates, five-fold cross-validation (CV), this study aims to: (1) assess whether integrating GWAS-derived prior information with ML algorithms enhances predictive performance; (2) compare the predictive efficiency of conventional, prior-informed, and machine learning models; and (3) evaluate the impact of marker density by comparing SNP chip data with imputed WGS data under varying LD pruning thresholds.

By systematically comparing these diverse methodologies, this study establishes a trait-informed decision framework to guide breeders in selecting optimal genomic prediction strategies, thereby enhancing both the efficiency and accuracy of breeding programs for economically important traits in poultry.

Materials and methods

Experiment animals, phenotypes and pedigree

A total of 3,737 M-line hens were sourced from the Hainan Wenchang Chicken Breeding Farm, including 2,304 from the 19th generation and 1,433 from the 20th generation, all reared under uniform environmental conditions. Phenotypic traits were recorded at three developmental stages (22, 32, and 45 weeks of age), including body weight (BW) and a novel ultrasound-derived trait—AFT. As a result of culling and slaughtering during the rearing process, the number of sampled birds decreased with age, leading to missing phenotypic records for some individuals. Prior to statistical analysis, quality control procedures and normality tests were rigorously performed. Descriptive statistics of the phenotypic data are presented in Table 1.

AFT was quantitatively measured using a DCU10 digital color Doppler ultrasound system (Xuzhou Kaixin Electronic Instrument Co., Ltd., China) equipped with a 7.5-MHz linear-array probe operating in B-mode grayscale imaging. Prior to scanning, chickens were gently removed from their cages and manually restrained in a dorsal recumbent position on the operator's left leg, after which feathers in the abdominal region were removed. An appropriate amount of medical coupling gel was applied to the probe surface to ensure effective acoustic transmission. During ultrasonography, the probe was positioned 3–5 cm posterior to the keel bone with an imaging depth of 5.5 cm. All instrument parameters were maintained within the safety range (thermal index, TIS=0.2; mechanical index, MI=0.8), ensuring no adverse effects on the examined birds. Considering the heterogeneity of abdominal fat distribution, three independent measurements were taken from the left, middle, and right abdominal regions. The saved ultrasound images were later analyzed using Image-Pro Plus 6.0 software, where clear upper and lower boundaries

of the fat layer were manually demarcated for measurement. The mean value of these triplicate measurements was calculated and used as the final AFT phenotype for each individual.

Initial parent–offspring assignments from farm records were verified against genotype data, leading to the correction of 246 erroneous assignments using genotype information; a further 54 pairs were annotated as having unknown parental status in the final pedigree file, which included 4,472 individuals from generations 18 to 20. All phenotypic measurements were conducted on-site by trained personnel to ensure data consistency and measurement accuracy.

Genotyping by SNP chip and genotype imputation

Genomic DNA was extracted from whole blood samples of 3,737 Wenchang chickens collected via wing vein puncture, and genotyping was conducted using the Jinxin 55K SNP chip, with SNP positions mapped to the GRCg6a reference genome. Quality control was conducted using PLINK v1.90b6.21 [22], which included the following steps: (1) SNPs with a call rate of less than 98% were removed; (2) SNPs with a minor allele frequency (MAF) of less than 0.05 were excluded; (3) SNPs significantly deviating from Hardy–Weinberg equilibrium ($P < 10^{-6}$) were eliminated; (4) individuals with a genotype call rate of less than 95% were excluded; and (5) SNPs located on chromosomes with unknown physical positions were removed. A total of 41,086 high-quality SNPs from 3,737 individuals were retained for subsequent genetic evaluation and wssGWAS analysis.

To further enhance genomic resolution, 312 Wenchang chickens closely related to the chip-genotyped population were selected for whole-genome sequencing. Sequencing libraries were constructed from high-quality DNA (fragmented to 300–500 bp) and sequenced on the DNBSEQ-T7 platform (MGI) to an average depth of 15×.

Table 1 Descriptive statistics and genetic parameter estimates of traits

Traits ^a	22w-BW	32w-BW	45w-BW	22w-AFT	32w-AFT	45w-AFT
Ne ^b	3,434	3,153	2,297	2,365	2,628	1,630
Max	2.372	2.894	2.972	1.298	1.858	1.913
Min	1.062	1.225	1.413	0.197	0.197	0.199
Mean	1.774	1.905	1.969	0.750	0.698	0.762
Std	0.172	0.228	0.276	0.210	0.234	0.213
σ_g^2	0.004	0.013	0.014	0.007	0.013	0.010
σ_e^2	0.022	0.036	0.057	0.032	0.038	0.048
$h^2(\text{SE})^c$	0.155(0.005)	0.264(0.007)	0.200(0.008)	0.175(0.007)	0.254(0.008)	0.174(0.009)

^a 22w-AFT, 32w-AFT, 45w-AFT, 22w-BW, 32w-BW, and 45w-BW represent abdominal fat thickness (ultrasound-derived novel traits) and body weight measured at 22, 32, and 45 weeks of age in the experimental population

^b Ne: Effective number of records for each trait

^c $h^2(\text{SE})$: Heritability estimates of traits with standard errors in parentheses

After alignment to the GRCg7b reference genome using BWA, the resulting data served as a reference panel for genotype imputation. Throughout the process, stringent quality control was applied, including assessments of DNA concentration and integrity, library preparation monitoring, and post-sequencing QC with FastQC and Samtools, to ensure data reliability. Reference-based imputation was conducted using Beagle version 5.4 [23] on the 55K SNP chip data. Only SNPs with consistent positions between the chip and sequencing datasets and a post-imputation dosage R-squared (R^2) ≥ 0.95 were retained. The imputed data were then subjected to quality control using the same criteria as the chip data, resulting in 6,360,109 high-quality SNPs for downstream analysis. Subsequently, LD pruning was performed using PLINK v1.90b6.21 [22] with the command “--indep-pairwise 50 5 $r^2 >$ ” under LD thresholds (r^2) of 0.2, 0.4, and 0.6, yielding 554,660, 1,463,395, and 2,425,353 SNPs, respectively, for subsequent genomic prediction analyses.

Derivation of corrected phenotypes

To avoid redundant accounting of parental information, the adjusted phenotypic values (y_c) were used as the response variable in genomic prediction. For each trait, estimated breeding values (EBVs) were obtained using pedigree-based ABLUP under a single-trait animal model, where the numerator relationship matrix (A matrix) was constructed from the complete pedigree information to quantify additive genetic relationships among individuals.

$$y = Xb + \gamma W + Za + e \quad (1)$$

where y represents the phenotypic observations of the traits in Wenchang chickens; b is the vector of fixed effects (e.g., batch, effect); W is a vector of covariate of the phenotypic value (e.g., pc1, pc2, pc3); γ is the regression coefficient associated with W ; a is the random additive genetic effect of the individual, assuming a normal distribution: $a \sim N(0, A\sigma_a^2)$, where σ_a^2 is the additive genetic variance; e is the vector of random residuals with distribution of $N(0, I\sigma_e^2)$, σ_e^2 is the residual variance, I is the identity matrix. X and Z are incidence matrices associating b and a with y , respectively. EBVs were calculated using the blupf90 program within the BLUPF90 software suite [24]. y_c were obtained by adding the estimated residuals to the corresponding EBVs [25].

Genomic prediction models

In this study, multiple genomic prediction approaches were employed to estimate genomic estimated breeding values (GEBVs) for target traits. These included three conventional models—GBLUP, ssGBLUP, and WGBLUP—as well as five machine learning regression algorithms: KRR, SVR,

RF, LightGBM, and AdaBoost.R2. To assess the impact of SNP marker density on prediction accuracy, SNP chip data were first imputed to WGS level. LD-based pruning was then applied to generate SNP subsets under three LD thresholds (LD=0.2, 0.4, and 0.6). GBLUP models were constructed for each SNP subset to perform prediction analyses. All models were evaluated using five-fold cross-validation repeated ten times to ensure robustness and stability of the prediction performance. The GBLUP, ssGBLUP, and WGBLUP models were implemented using the blupf90 software package [24].

Genomic best linear unbiased prediction

For genomic prediction, we first employed the standard genomic best linear unbiased prediction model. This model estimates individual genomic breeding values by constructing a genomic relationship matrix (G -matrix) based on genome-wide markers, which replaces the traditional pedigree-based relationship matrix.

$$y_c = \mu + Zg + e \quad (2)$$

where y_c is the vector of adjusted phenotypic values, μ is the overall mean, Z is the incidence matrix relating observations to individuals, g is the vector of additive genetic values assumed to follow a normal distribution $g \sim N(0, G\sigma_g^2)$, and e is the vector of random residuals with $e \sim N(0, I\sigma_e^2)$. Here, G denotes the genomic relationship matrix (G matrix), while σ_g^2 and σ_e^2 represent the additive genetic variance and residual variance, respectively.

Single-step genomic best linear unbiased prediction

Unlike GBLUP, the ssGBLUP model constructs a unified H matrix by integrating the pedigree-based relationship matrix (A matrix) with the genomic relationship matrix (G matrix) [26]. It was assumed that g followed a normal distribution $N(0, H\sigma_g^2)$. The inverse of matrix H was:

$$H^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & G_w^{-1} - A_{22}^{-1} \end{bmatrix} + A^{-1} \quad (3)$$

where A is the numerator relationship matrix based on pedigree for all animals, A_{22}^{-1} is the inverse of the pedigree-based relationship matrix for the genotyped animals, and G_w^{-1} is the inverse of the genomic relationship matrix, $G_w = 0.95G + 0.05A_{22}$.

Weighted genomic best linear unbiased prediction

The WGBLUP model is conceptually similar to the conventional GBLUP model, with the key distinction that it incorporates marker-specific weights when constructing the genomic relationship matrix. In WGBLUP, all SNPs are assigned weights, which are iteratively

updated based on their estimated contributions to the trait of interest. The formulas below outline the computational procedures for deriving weighted SNPs through the WSSGWAS approach:

- Step 1: let $t = 0, \mathbf{D}_{(t)} = \mathbf{I}, \mathbf{G}_{(t)} = \mathbf{Z}\mathbf{D}_{(t)}\mathbf{Z}'\lambda$, and $\lambda = \sum_{i=1}^M 2p_i(1 - p_i)$;
 Step 2: Compute $\hat{\mathbf{a}}_g$ by ssGBLUP;
 Step 3: Calculate $\hat{\mathbf{u}}_{(t)} = \lambda \mathbf{D}_{(t)} \mathbf{Z}' \mathbf{G}_{(t)}^{-1} \hat{\mathbf{a}}_g$;
 Step 4: Calculate $d_{i(t+1)}^* = \hat{u}_{i(t+1)}^2 / 2p_i(1 - p_i)$ for all SNPs;
 Step 5: Normalize $\mathbf{D}_{(t+1)} = \frac{\hat{\mathbf{u}}_{(t+1)}^2}{\text{tr}(\mathbf{D}_{(t+1)})} \mathbf{D}_{(t+1)}$;
 Step 6: Calculate $\mathbf{G}$ for the next iteration, $\mathbf{G}_{(t+1)} = \mathbf{Z}\mathbf{D}_{(t+1)}\mathbf{Z}'\lambda$;
 Step 7: let $t = t + 1$;
 Step 8: Loop to step 2 or 3.

As suggested by Wang et al. [27], the procedure was run for two iterations. Each non-overlapping window included 20 consecutive SNPs, and the genetic variances explained by these windows were calculated following the method of Wang et al. [27].

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{D}\mathbf{Z}'}{\sum_{i=1}^M 2p_i(1 - p_i)} \quad (4)$$

where $\mathbf{Z}$ is a matrix of gene content adjusted for allele frequencies (); $\mathbf{D}$ is a diagonal matrix of weights for SNP variances (initially $\mathbf{D}=\mathbf{I}$); M is 0, 1, or 2 for *aa*, *Aa*, and *AA*, respectively the number of SNPs, and p_i is the minor allele frequency of the i^{th} SNP. This iterative weighting approach allows WGBLUP to prioritize SNPs based on their trait contributions, enhancing both prediction accuracy and model interpretability.

Kernel ridge regression

KRR, as a nonparametric extension of ridge regression, is effective in capturing complex nonlinear relationships in the data [28]. KRR maps the input data into a high-dimensional kernel space using nonlinear kernel functions, where a linear ridge regression model is then constructed to enable linear separability in this transformed space. The optimal linear function in the kernel space is selected by minimizing the ridge-regularized mean squared error (MSE) loss function. The final predictive model in KRR can be expressed as:

$$\mathbf{y}(x_i) = \mathbf{k}'(\mathbf{K} + \lambda \mathbf{I})^{-1} \hat{\mathbf{y}} \quad (5)$$

where λ is the regularization parameter, $\mathbf{K}$ is the Gram matrix defined by $\mathbf{K}_{ij} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) * \phi(\mathbf{x}_j)^T$, and $\mathbf{I}$ is the identity matrix. For a training dataset with n samples, the kernel matrix $\mathbf{K}$ is defined as:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}(\mathbf{x}_1, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \mathbf{K}(\mathbf{x}_1, \mathbf{x}_n) \\ \mathbf{K}(\mathbf{x}_2, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_2, \mathbf{x}_2) & \cdots & \mathbf{K}(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}(\mathbf{x}_n, \mathbf{x}_1) & \mathbf{K}(\mathbf{x}_n, \mathbf{x}_2) & \cdots & \mathbf{K}(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}_{n \times n} \quad (6)$$

In this study, a grid search strategy combined with internal five-fold CV was employed to optimize both the kernel function and the regularization parameter λ , aiming to achieve optimal model performance.

Support vector regression

Support vector machines (SVM), based on statistical learning theory, are widely applied in both classification and regression tasks. SVR extends SVM to regression problems by using linear or nonlinear kernel functions to map input data into a high-dimensional feature space, where a linear regression model is constructed to fit continuous response variables [29, 30]. The SVR model can be formulated as:

$$f(\mathbf{x}) = \beta_0 + \mathbf{h}(\mathbf{x})^T \boldsymbol{\beta} \quad (7)$$

where $\mathbf{h}(\mathbf{x})^T \boldsymbol{\beta}$ represents the kernel-induced linear combination, $\boldsymbol{\beta}$ is the weight vector, and β_0 is the intercept. SVR aims to minimize the following objective function:

$$\min_{\beta_0, \boldsymbol{\beta}} \frac{1}{2} \|\boldsymbol{\beta}\|^2 = C \sum_{i=1}^n V_\varepsilon(y_i - f(\mathbf{x}_i)) \quad (8)$$

where $V_\varepsilon(r)$ is the ε -insensitive loss function defined as:

$$V_\varepsilon(r) = \begin{cases} 0, & \text{if } |r| < \varepsilon \\ |r| - \varepsilon, & \text{otherwise} \end{cases} \quad (9)$$

The regularization parameter C controls the trade-off between the model's complexity and prediction error. $\|\cdot\|$ denotes the norm in the reproducing kernel Hilbert space. After optimization, the final form of the SVR model can be written as:

$$f(\mathbf{x}) = \sum_{i=1}^m (\hat{a}_i - a_i) \mathbf{k}(\mathbf{x}, \mathbf{x}_i) \quad (10)$$

where $\mathbf{k}(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x}_j)^T \phi(\mathbf{x}_i)$ is the kernel function, and $\hat{a}_i$ and a_i are the Lagrange multipliers obtained from the dual optimization [31]. In this study, an internal five-fold cross-validation strategy combined with grid search was applied to select the optimal kernel function and tune the hyperparameters C and γ to enhance model generalization.

Random forest

RF is an ensemble-based machine learning algorithm that constructs multiple decision trees and integrates their outputs by voting (for classification) or averaging (for regression) to enhance predictive accuracy and reduce overfitting risk [32]. Each tree in the forest is trained on a different random subset of the data, enabling the model to maintain strong generalization ability while reducing variance [31]. The RF regression model is expressed as:

$$y = \frac{1}{M} \sum_{m=1}^M t_m(\psi_m(x)) \quad (11)$$

where y represents the predicted value, $t_m(\psi_m(x))$ denotes the output of the m^{th} decision tree, and M is the total number of trees in the forest. Each tree processes the input features through a hierarchical structure, and the predicted value is assigned based on the output of the terminal leaf node. Internal fivefold cross-validation was employed in combination with grid search to optimize the number of trees (M) and the maximum tree depth.

Genotype vectors (encoded as 0, 1, and 2) were employed as predictor variables, while y_c served as the response in all five machine learning models. The Scikit-learn package (version 1.6.0) in Python was used for model training and evaluation. Grid search was used to optimize hyperparameters, and the best-performing model within each cross-validation fold was identified based on the highest Pearson correlation between predicted and actual phenotypes.

Cross-validation and genome prediction effect

To evaluate the accuracy of genomic prediction, a repeated five-fold CV scheme was employed, involving ten repetitions for a total of 50 validation rounds. Prediction accuracy was defined as the Pearson correlation coefficient between y_c and the corresponding GEBVs within the validation subsets. This metric directly measures the model's predictive performance in the target population. Although standardized accuracy (r/h) is useful for comparing traits with differing heritabilities, the unstandardized correlation (r) is more widely adopted in the literature, enabling direct comparisons with previous studies. In each repetition, the 3,737 individuals were randomly partitioned into five subsets of approximately 747 birds each. Four subsets were used to train the prediction model, while the remaining one served as the validation set.

Prediction bias was assessed by regressing y_c on GEBVs in the validation population, with the resulting regression coefficient reflecting the degree of inflation or deflation in the genomic predictions. Both prediction accuracy and

bias were averaged across all 50 CV rounds to obtain the final estimates.

To further assess the performance of the regression models, MSE and mean absolute error (MAE) were also calculated. MSE incorporates both variance and bias, where a lower value indicates better model fit to the observed data. MAE, on the other hand, offers a more intuitive measure of the average prediction error. These metrics were computed using the following formulas:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (f_i - y_i)^2, \text{ and } \text{MAE} = \frac{1}{m} \sum_{i=1}^m |f_i - y_i| \quad (12)$$

where m represents the number of individuals in the validation set for each CV round, f denotes the predicted GEBV vector, and y is the vector of y_c . The final MSE and MAE values were averaged across all cross-validation rounds to represent overall model performance.

Results

This study established a precision breeding system for chicken abdominal fat traits by integrating ultrasound imaging and multi-omics data (Fig. 1). First, high-throughput phenotyping and standardization of abdominal fat thickness were achieved through multi-stage dynamic ultrasound monitoring (Fig. 1a). Second, by combining SNP arrays, whole-genome sequencing, and multiple prediction models, we developed an integrated genotype–phenotype analysis pipeline encompassing GBLUP, WGBLUP, and machine learning algorithms (Fig. 1b). Finally, the optimal prediction model was selected through repeated cross-validation and applied to breeding practice, thereby completing a closed-loop technical pathway from accurate phenotyping to efficient genomic selection (Fig. 1c). This system provides a scalable methodological framework for genetic improvement of fat-related traits in poultry.

Phenotypic variation and genetic parameters underpin predictive challenges

Prior to genomic prediction, descriptive statistics were computed for the target traits. BW increased progressively from 22 to 45 weeks of age, with mean values of 1.774 kg, 1.905 kg, and 1.969 kg, respectively. In contrast, AFT showed a fluctuating trend, declining initially from 0.750 cm at 22 weeks to 0.698 cm at 32 weeks, before rising to 0.762 cm at 45 weeks. Considerable phenotypic variation was observed, as reflected by standard deviations ranging from 0.172 to 0.276 for BW and from 0.210 to 0.234 for AFT. Heritability estimates (with standard errors) for BW across the three ages were 0.155 (0.010), 0.264 (0.012), and 0.200 (0.013), while those for AFT were 0.175 (0.012), 0.254 (0.012), and 0.174 (0.015), with the

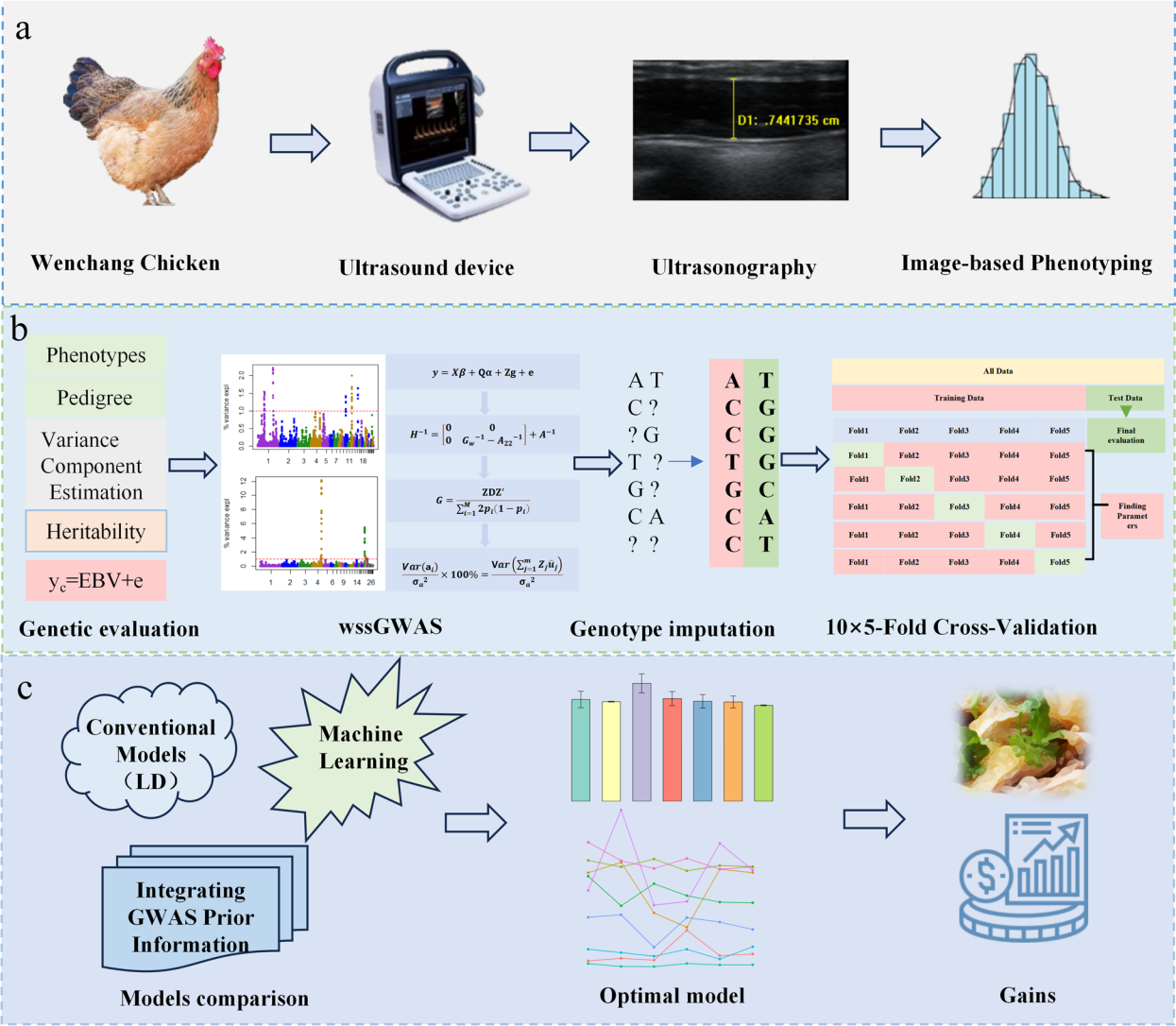


Fig. 1 An integrated framework for precision breeding of chicken abdominal fat traits by combining ultrasound phenotyping and multi-omics prediction

highest heritability observed at 32 weeks for both traits. Strong pairwise genetic correlations were observed for BW (0.84–0.87) and AFT (0.83–0.98) across the three measurement ages (22, 32, and 45 weeks). Concurrently, moderate positive genetic correlations were evident between BW and AFT at different developmental stages (Fig. 2b).

To evaluate the stability of population genetic structure under continuous artificial selection, principal component analysis (PCA) was conducted on the 19th (M19) and 20th (M20) generations of the Wenchang chicken M-line. Based on 40,988 SNP markers, the first two principal components explained 0.65% and 0.59% of the total genetic variation, respectively (Fig. 2c).

The two consecutive generations showed high overlap in the PC1–PC2 space, with 95% confidence ellipses nearly fully overlapping, indicating that population genetic structure remained stable under ongoing selection and that genetic drift was limited between generations. These results provide a favorable and consistent genetic background for implementing cross-generational genomic selection in this chicken line. Following imputation of the 55K SNP array data to whole-genome sequence level, a total of 41,195 SNPs common to both platforms were retained. The distribution of SNP density across chromosomes and the overlap of shared loci are visualized in Fig. 2d.

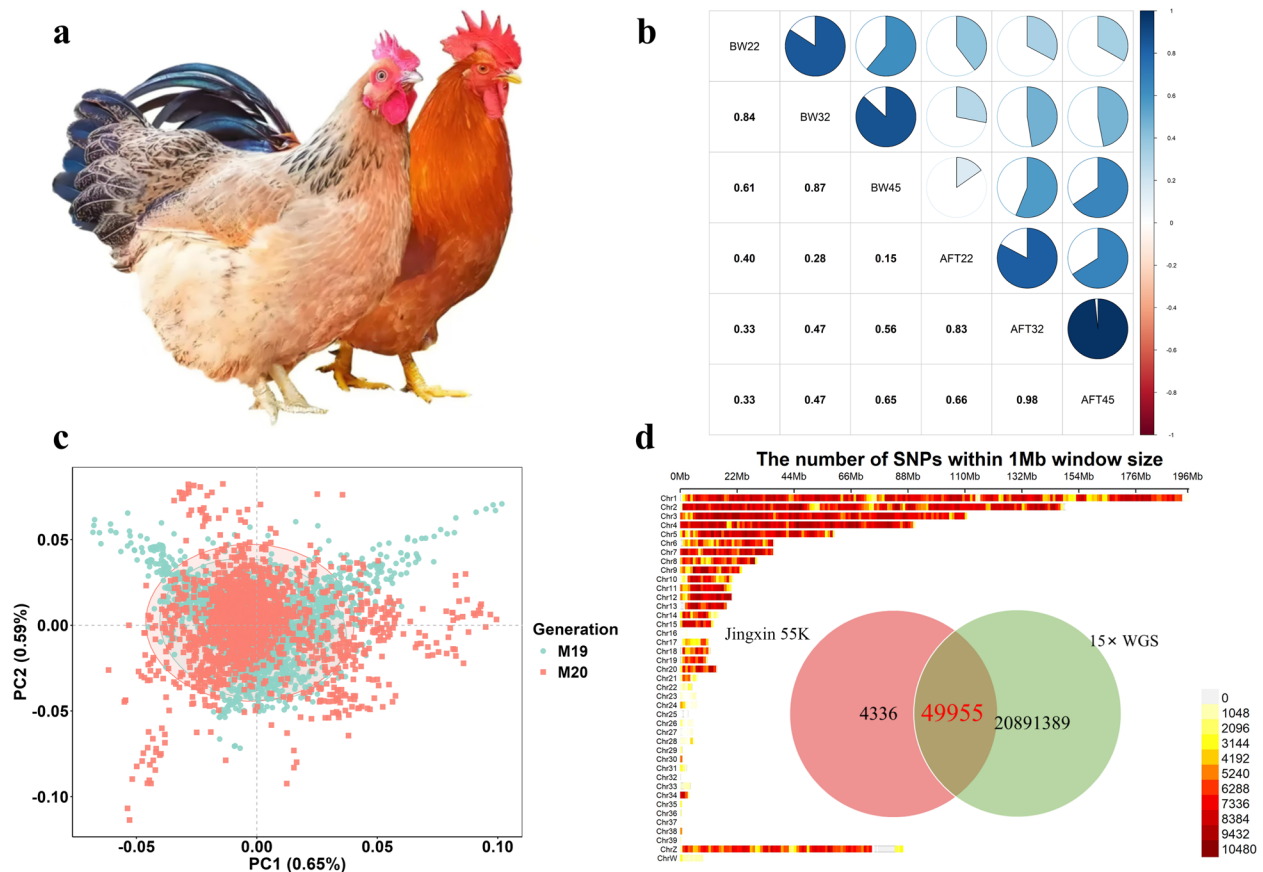


Fig. 2 Population overview and genomic characteristics of Wenchang chicken. **a** Images of male and female Wenchang chickens. **b** Visualization of genetic correlation matrix among production traits. Traits abbreviations are displayed on the diagonal, the lower triangle presents numerical values of genetic correlations, while the upper triangle uses sector size to visually represent the correlation strength. **c** PCA plot of the studied population. Ellipses represent 95% confidence intervals for each generation (M19: 19th generation; M20: 20th generation). **d** SNP density distribution on chromosomes after imputation to whole-genome sequencing level, along with a Venn diagram illustrating the overlap between imputed and chip SNP loci

A slaughter validation on 320 chickens confirmed the accuracy of ultrasound-derived AFT. Strong correlations were observed with post-slaughter abdominal fat weight ($r=0.760$, $P<0.001$) and percentage ($r=0.764$, $P<0.001$) (Fig. S1A and C). Bland–Altman analysis showed excellent agreement, with 94.4% (fat weight) and 95.0% (fat percentage) of points within the 95% limits of agreement and no systematic bias (Fig. S1B and D). These results validate ultrasound as a reliable in vivo phenotyping tool for abdominal fat in Wenchang chickens.

Incorporating prior information enhances genomic prediction

Figure 3a illustrates the prediction accuracies obtained from ten replicates of five-fold cross-validation for the GBLUP, ssGBLUP, and WGBLUP models, where

WGBLUP incorporates SNP weights derived from wssGWAS results. For body weight traits, the GBLUP model based on the 55K SNP chip yielded prediction accuracies of 0.1619, 0.2407, and 0.1903 at 22, 32, and 45 weeks of age, respectively. Overall, WGBLUP—which incorporates prior weights from wssGWAS—consistently achieved the highest accuracy for both body weight and abdominal fat thickness, with average improvements of 2.18% and 0.55%, respectively, over GBLUP. The ssGBLUP model also consistently outperformed GBLUP. Notably, the advantage of WGBLUP was most pronounced at specific time points: it improved prediction accuracy by 5.25% for body weight at 22 weeks and by 4.47% for abdominal fat thickness at 45 weeks. These findings underscore that integrating prior biological information enhances genomic prediction, particularly for traits under strong selection or stage-specific regulation.

Machine learning models offer targeted gains for complex traits

Figure 3b shows the average prediction accuracies of various ML methods and GBLUP for BW and AFT traits, based on ten replicates of five-fold cross-validation. Using default parameters, only SVR demonstrated a modest improvement over GBLUP for the 32-week AFT trait, while the other models showed no enhancement. After hyperparameter tuning, KRR achieved the first significant gains, followed by SVR; in contrast, RF failed to improve prediction accuracy under any tested conditions. Specifically, relative to GBLUP, KRR increased the average prediction accuracy for BW by 1.92%, 3.00%, and 1.37% across different time points. SVR exhibited a modest improvement of approximately 0.20% for BW at the first two time points. For AFT measured at three time points, KRR enhanced prediction accuracy by 0.68%, 4.15%, and 0.20%, while SVR improved accuracy by only 2.1% at 32 weeks. Among the ML approaches, RF consistently performed the poorest for the traits evaluated. The optimal hyperparameters for KRR, SVR, and RF, determined via grid search, are summarized in Table 2. Additionally, ensemble learning methods AdaBoost.R2 and LightGBM were assessed. Despite their theoretical advantages in modeling nonlinear relationships, these methods generally yielded lower prediction accuracies than GBLUP and other ML models in this study (detailed results are provided in Table S1) and thus are not discussed in detail in the main text.

Optimal marker density is trait-specific, reflecting underlying genetic architecture

Figure 3c presents bar plots of prediction accuracies from five-fold cross-validation repeated ten times under varying SNP marker densities. The compared models include the original GBLUP based on the 55K SNP chip and three GBLUP_LD models derived from WGS-imputed genotypes pruned at different LD

thresholds (LD0.2, LD0.4, and LD0.6). For body weight traits, the prediction accuracies of GBLUP_LD0.2 at 22, 32, and 45 weeks were 0.1835, 0.2474, and 0.2017, respectively. The respective prediction accuracies for GBLUP_LD0.4 were 0.1822, 0.2457, and 0.2012, while those for GBLUP_LD0.6 were 0.1722, 0.2442, and 0.2010, respectively. GBLUP_LD0.2 consistently achieved the highest prediction accuracy across all time points, suggesting that this LD threshold offers an optimal marker density for genomic prediction of body weight. On average, it improved prediction accuracy by 6.58% compared to the original GBLUP model. For abdominal fat thickness, prediction accuracies at 22 weeks were 0.1976, 0.1979, and 0.1980 for GBLUP_LD0.2, GBLUP_LD0.4, and GBLUP_LD0.6, respectively; at 32 weeks, they were 0.2583, 0.2589, and 0.2591; and at 45 weeks, all three models produced identical values of 0.1759. GBLUP_LD0.6 exhibited the best overall performance across the three stages, with an average improvement of 15.30% over the original GBLUP. These findings highlight that appropriate LD pruning of WGS-level SNP data can optimize marker distribution and substantially enhance the accuracy of genomic predictions. Moreover, genetic evaluations using SNPs obtained after pruning from WGS data resulted in improved estimated heritability for all traits. Detailed heritability values for each trait can be found in Table S2.

Bias level

The regression coefficient of y_c on predicted values (PV) was used to evaluate the prediction bias level of each model. As shown in Fig. 3d, the average bias level coefficients for GBLUP across three time points were 0.9863 for BW and 0.9989 for AFT, indicating no evidence of dispersion bias. For BW traits, WGBLUP (1.0142) and ssGBLUP (1.0228) showed bias level closer to the ideal value of 1 compared to GBLUP, whereas the opposite trend was observed for AFT traits. Among GBLUP models with

(See figure on next page.)

Fig. 3 Comprehensive evaluation of genomic prediction models for body weight and abdominal fat thickness. The figure compares the prediction accuracy and bias level of nine genomic prediction methods for body weight (BW) and abdominal fat thickness (AFT) based on 10 repetitions of fivefold cross-validation. **a** Prediction accuracy of GBLUP, ssGBLUP, and WGBLUP across different traits. **b** Comparative prediction accuracy of GBLUP, SVR, KRR, and RF for each trait. **c** Accuracy of GBLUP under three LD-pruning thresholds (LD = 0.2, 0.4, 0.6). **d** Comparison of prediction bias among the nine methods for BW and AFT, where values on the y-axis closer to 1 indicate better bias level. For panels **a–c**, the numerical values above each bar represent the corresponding prediction accuracy, and the error bars denote the standard error. Different lowercase letters above bars indicate significant differences based on Tukey's HSD test ($P < 0.05$) following one-way ANOVA of 10 replicated fivefold cross-validation runs. GBLUP: Genomic Best Linear Unbiased Prediction; ssGBLUP: Single-Step Genomic Best Linear Unbiased Prediction; WGBLUP: Weighted Genomic Best Linear Unbiased Prediction; GBLUP_LD0.2, GBLUP_LD0.4, GBLUP_LD0.6: GBLUP models fitted using SNP subsets pruned at linkage disequilibrium thresholds of $r^2 < 0.2$, 0.4, and 0.6, respectively. KRR: Kernel Ridge Regression; SVR: Support Vector Regression; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; AdaBoost.R2: AdaBoost for Regression

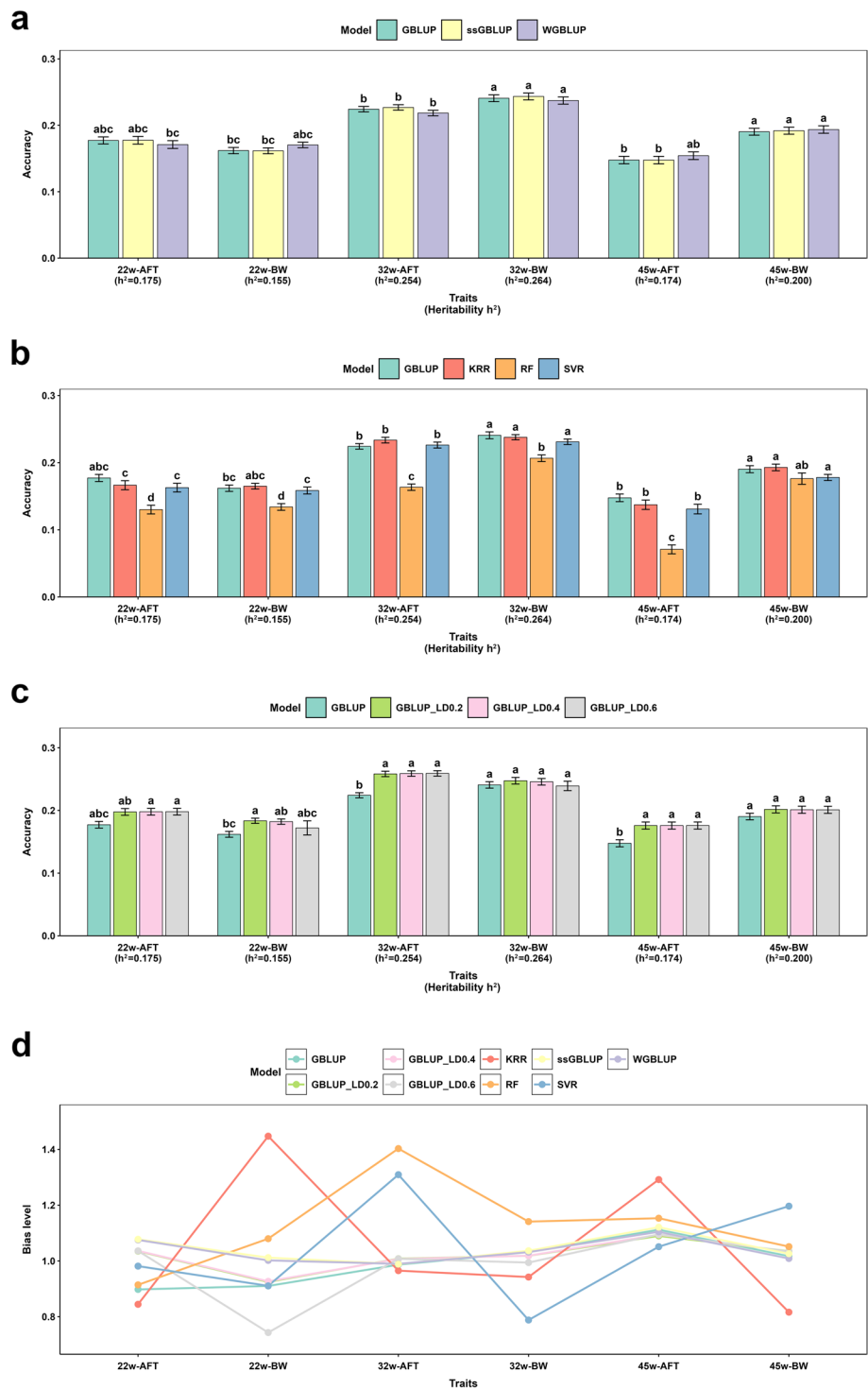


Fig. 3 (See legend on previous page.)

different LD pruning thresholds (LD0.2, LD0.4, LD0.6), the prediction bias level for BW traits initially improved and then declined, with GBLUP_LD0.4 performing best and surpassing the standard GBLUP model. For AFT

traits, GBLUP_LD0.2 exhibited bias level closer to 1 but did not exceed GBLUP. ML models such as KRR, SVR, and RF demonstrated relatively poor bias level with default parameters. After hyperparameter tuning via

Table 2 The optimal hyperparameters of each ML model obtained through a grid search for BW and AFT traits in 10 replicates of fivefold CV

Model ¹	Optimal hyperparameters ²	
	BW	AFT
KRR	kernel='rbf', $\lambda = 0.1$, gamma = 0.0001	
SVR	kernel='rbf', C = 1.0, gamma = 0.0001	kernel='rbf', C = 0.1, gamma = 0.0001
RF	n_estimators = 200, max_depth = 10, min_samples_split = 5	
LightGBM	learning_rate = 0.01, max_depth = 5, n_estimators = 200, num_leaves = 15	
AdaBoost.R2	learning_rate = 0.01, n_estimators = 100	learning_rate = 0.01, n_estimators = 50

¹ KRR: Kernel Ridge Regression; SVR: Support Vector Regression; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; AdaBoost.R2: AdaBoost for Regression

² Optimal hyperparameters: The optimal hyperparameters of each machine learning method obtained by using a grid search

grid search, their bias level approached 1; however, RF and KRR still underperformed relative to GBLUP, while only SVR outperformed GBLUP in bias level. Overall, the WGBLUP model showed the closest bias level to 1 for 22-week BW (1.0019), whereas GBLUP_LD0.4 was closest to 1 for 32-week AFT (1.0071).

MSE and MAE

Model performance was further evaluated using MSE and MAE. As presented in Table 3 and Fig. 4, both WGBLUP and ssGBLUP consistently outperformed GBLUP in terms of lower MAE and MSE across all evaluated traits. For instance, for 22-week BW, GBLUP yielded higher prediction errors (MAE: 0.1262; MSE: 0.0254) compared to WGBLUP (MAE: 0.1258; MSE: 0.0252) and ssGBLUP (MAE: 0.1260; MSE: 0.0253). Additionally, for BW traits, the MSE and MAE performances of GBLUP_LD0.2, GBLUP_LD0.4, and GBLUP_LD0.6 followed their prediction accuracy trends, with GBLUP_LD0.2 achieving the lowest MAE values (0.1255 ± 0.0005 , 0.1675 ± 0.0007 , and 0.2055 ± 0.0010) and MSE values (0.0251 ± 0.0002 ,

0.0450 ± 0.0004 , and 0.0684 ± 0.0007). However, these three models performed similarly for AFT traits. ML methods with default parameters exhibited higher MAE and MSE than GBLUP; following hyperparameter optimization, KRR and SVR generally outperformed GBLUP on these metrics and performed comparably to ssGBLUP and WGBLUP overall. Among ML methods, KRR produced the lowest average MSE (0.0397 for BW; 0.0470 for AFT) and MAE (0.1615 for BW; 0.1704 for AFT) across time points, whereas RF performed the worst, with MSEs of 0.0471 (BW) and 0.0478 (AFT), and MAEs of 0.2004 (BW) and 0.1963 (AFT). Taken together, among the nine methods evaluated, those based on WGS-imputed data combined with LD pruning consistently achieved lower MAE and MSE values across all traits. Table 3 presents MSE values as means with corresponding standard errors (mean $\pm$ SE).

Table 3 Mean squared error (MSE) of nine methods for BW and AFT as assessed with 10 replicates of fivefold CV (mean $\pm$ SE)

Model ¹	22w-BW	32w-BW	45w-BW	22w-AFT	32w-AFT	45w-AFT
GBLUP	0.0254 $\pm$ 0.0002	0.0452 $\pm$ 0.0004	0.0687 $\pm$ 0.0007	0.0376 $\pm$ 0.0003	0.0475 $\pm$ 0.0004	0.0563 $\pm$ 0.0006
SSGBLUP	0.0253 $\pm$ 0.0002	0.0452 $\pm$ 0.0004	0.0687 $\pm$ 0.0007	0.0375 $\pm$ 0.0004	0.0475 $\pm$ 0.0004	0.0563 $\pm$ 0.0006
WGBLUP	0.0252 $\pm$ 0.0002	0.0452 $\pm$ 0.0004	0.0686 $\pm$ 0.0007	0.0376 $\pm$ 0.0004	0.0477 $\pm$ 0.0004	0.0562 $\pm$ 0.0006
GBLUP_LD0.2	0.0251 $\pm$ 0.0002	0.0450 $\pm$ 0.0004	0.0684 $\pm$ 0.0007	0.0372 $\pm$ 0.0004	0.0467 $\pm$ 0.0004	0.0558 $\pm$ 0.0006
GBLUP_LD0.4	0.0251 $\pm$ 0.0002	0.0451 $\pm$ 0.0004	0.0684 $\pm$ 0.0007	0.0372 $\pm$ 0.0004	0.0467 $\pm$ 0.0004	0.0558 $\pm$ 0.0006
GBLUP_LD0.6	0.0259 $\pm$ 0.0003	0.0454 $\pm$ 0.0004	0.0684 $\pm$ 0.0007	0.0372 $\pm$ 0.0004	0.0467 $\pm$ 0.0004	0.0558 $\pm$ 0.0006
KRR	0.0253 $\pm$ 0.0002	0.0451 $\pm$ 0.0003	0.0686 $\pm$ 0.0006	0.0375 $\pm$ 0.0003	0.0474 $\pm$ 0.0004	0.0561 $\pm$ 0.0009
SVR	0.0254 $\pm$ 0.0002	0.0453 $\pm$ 0.0003	0.0689 $\pm$ 0.0006	0.0375 $\pm$ 0.0003	0.0479 $\pm$ 0.0005	0.0570 $\pm$ 0.0009
RF	0.0256 $\pm$ 0.0002	0.0456 $\pm$ 0.0003	0.0701 $\pm$ 0.0007	0.0379 $\pm$ 0.0003	0.0478 $\pm$ 0.0004	0.0576 $\pm$ 0.0009

¹ GBLUP Genomic: Best Linear Unbiased Prediction; ssGBLUP: Single-Step Genomic Best Linear Unbiased Prediction; WGBLUP: Weighted Genomic Best Linear Unbiased Prediction; GBLUP_LD0.2, GBLUP_LD0.4, GBLUP_LD0.6: GBLUP models fitted using SNP subsets pruned at linkage disequilibrium thresholds of $r^2 < 0.2$, 0.4, and 0.6, respectively. KRR: Kernel Ridge Regression; SVR: Support Vector Regression; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; AdaBoost.R2: AdaBoost for Regression

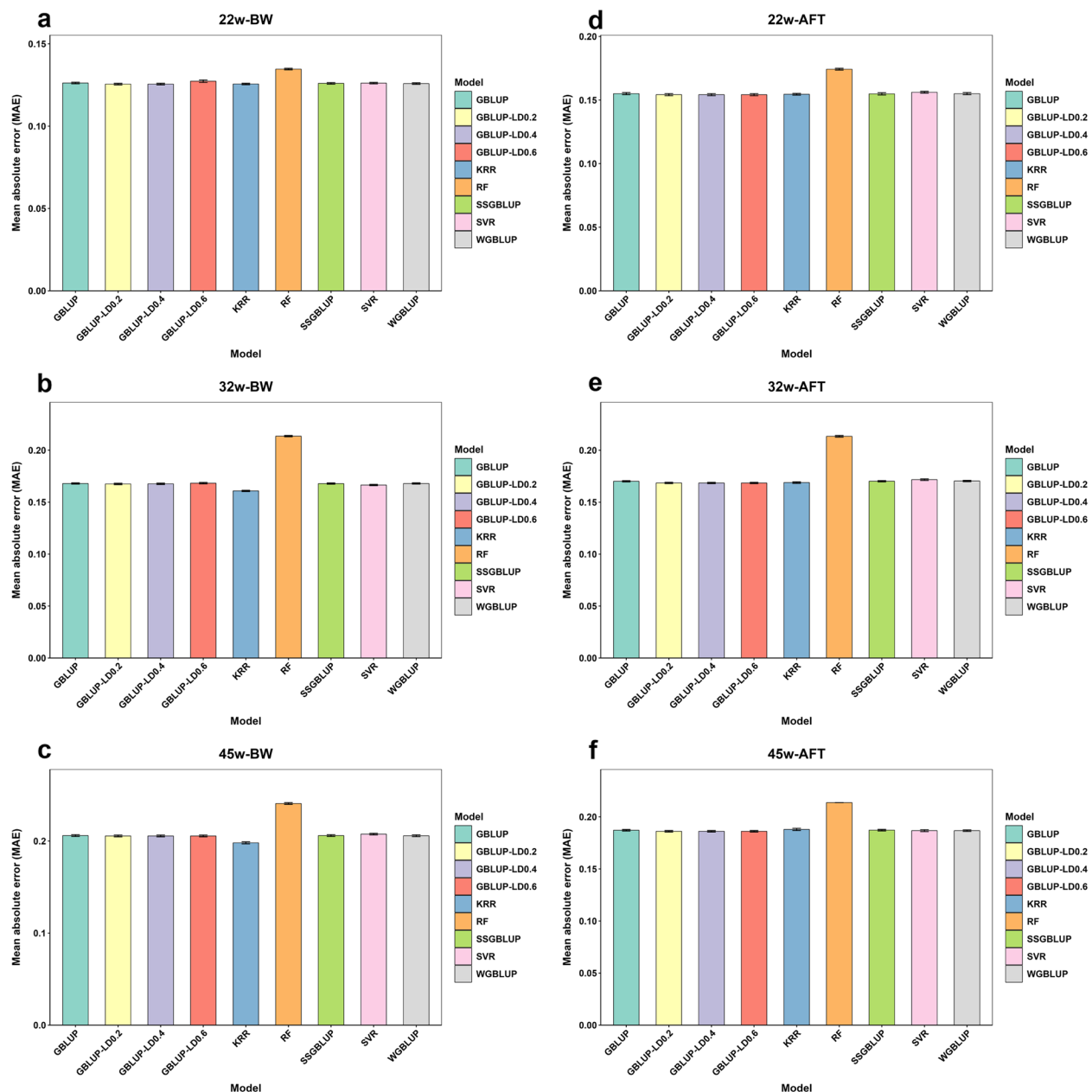


Fig. 4 Comparison of mean absolute error across models for body weight and abdominal fat thickness at three time points. **a–c** Body weight (BW) traits measured at 22, 32, and 45 weeks of age respectively. **d–f** Abdominal fat thickness (AFT) traits at the corresponding time points. Each panel presents the average mean absolute error (MAE) derived from ten replicates of five-fold cross-validation across various prediction models. Error bars indicate the standard error. Model: GBLUP Genomic: Best Linear Unbiased Prediction; ssGBLUP: Single-Step Genomic Best Linear Unbiased Prediction; WGBLUP: Weighted Genomic Best Linear Unbiased Prediction; GBLUP_LD0.2, GBLUP_LD0.4, GBLUP_LD0.6: GBLUP models fitted using SNP subsets pruned at linkage disequilibrium thresholds of $r^2 < 0.2, 0.4$, and 0.6 , respectively. KRR: Kernel Ridge Regression; SVR: Support Vector Regression; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; AdaBoost.R2: AdaBoost for Regression

Computing time

Table 4 summarizes the average computational time per fold for each genomic prediction method during CV in this study. The analyses were conducted on a high-performance computing cluster running Rocky Linux

5.6, consisting of two types of compute nodes: c01/c02 series CPU nodes, each equipped with dual Intel Xeon Platinum 8458P processors (2.7 GHz base frequency). The cluster includes 640 nodes with 1 TB DDR4 RAM and 396 nodes with 512 GB DDR4 RAM. Computation

times for all methods are reported in minutes. Among the methods tested, the scikit-learn-based KRR exhibited the highest computational efficiency, requiring only 0.02 to 0.04 min per fold on average. This was substantially faster than the mixed-model methods implemented in BLUPF90: GBLUP (1.27–1.91 min), WGBLUP (1.54–2.05 min), and ssGBLUP (1.62–2.04 min). SVR demonstrated moderate efficiency, with runtimes between 0.39 and 1.21 min, outperforming the GBLUP variants but slower than KRR. RF, however, was the slowest among the machine learning methods, with fold times ranging from 22.74 to 65.78 min. Additionally, GBLUP models based on WGS-imputed SNPs with linkage disequilibrium (LD) pruning showed computation times increasing with marker density: GBLUP_LD0.2 took 7.56–7.85 min, GBLUP_LD0.4 required 19.13–20.74 min, and GBLUP_LD0.6 demanded 20.80–32.52 min per fold. Importantly, the differences in computation time are attributable not just to the algorithms, but also to a major factor: the efficiency of their software implementations (BLUPF90 vs. scikit-learn).

Discussion

Our study delivers two pivotal contributions to poultry genomics: the establishment of ultrasound imaging as a scalable platform for in vivo phenotyping, and a systematic dissection of how genetic architecture dictates the performance of genomic prediction models. The moderate heritability observed for both BW and AFT, peaking at 32 weeks of age, underscores their genetic basis and highlights a critical developmental window for selection. This pattern is consistent with previous reports [33, 34], suggesting that this time point may represent a critical window for implementing effective selection strategies. The novel application of non-invasive ultrasound

measurement of AFT in poultry successfully transitions a traditionally postmortem trait into a longitudinal, animal welfare phenotype, thereby overcoming a major bottleneck in breeding for carcass composition traits. While our ultrasound-based phenotyping protocol proved practical and efficient, potential operator or device effects were not modeled statistically due to lack of systematic metadata tracking during high-throughput data collection. Future applications should implement standardized recording of operator and device information. Including these factors as covariates in sensitivity analyses—especially in multi-operator or multi-center breeding programs—would strengthen the robustness and generalizability of genomic predictions using ultrasound-derived traits.

The consistent superiority of ssGBLUP over traditional GBLUP reaffirms the value of integrating pedigree and genomic relationships, thereby enhancing the accuracy of genetic evaluations through a comprehensive representation of the genetic architecture[1, 11, 35, 36]. To further improve predictive performance, we implemented the WGBLUP. Unlike the traditional GBLUP model, which assumes equal contributions of all SNPs to the genetic variance [35], WGBLUP assigns differential weights to individual SNPs based on their estimated associations with the target traits [27], better capturing the underlying genetic complexity. We posit that this improvement stems from the model's ability to recalibrate the genomic relationship matrix by upweighting SNPs in or near QTL regions identified in our prior wssGWAS [34]. These findings are further supported by multiple independent studies, which have confirmed the effectiveness of incorporating functional information or association-based weights into genomic prediction models [36, 37]. This mechanism allows WGBLUP to transition from assuming a purely polygenic background to approximating a mixed

Table 4 Average computing time to complete each fold of fivefold CV according to different genomic prediction methods

Model ¹	22w-BW	32w-BW	45w-BW	22w-AFT	32w-AFT	45w-AFT
GBLUP	1.43 min	1.27 min	1.44 min	1.67 min	1.36 min	1.91 min
SSGBLUP	2.04 min	1.64 min	1.62 min	1.69 min	1.63 min	1.80 min
WGBLUP	1.54 min	1.72 min	1.64 min	2.05 min	1.70 min	1.79 min
GBLUP_LD0.2	7.76 min	7.56 min	7.57 min	7.77 min	7.60 min	7.85 min
GBLUP_LD0.4	20.74 min	19.13 min	20.28 min	20.74 min	19.13 min	20.28 min
GBLUP_LD0.6	20.80 min	32.49 min	31.86 min	31.66 min	31.64 min	32.52 min
KRR	0.04 min	0.04 min	0.03 min	0.02 min	0.03 min	0.02 min
SVR	1.06 min	1.21 min	0.74 min	0.74 min	1.05 min	0.39 min
RF	1 h 6 min	30.27 min	45.78 min	22.74 min	37.45 min	31.81 min

¹ GBLUP Genomic: Best Linear Unbiased Prediction; ssGBLUP: Single-Step Genomic Best Linear Unbiased Prediction; WGBLUP: Weighted Genomic Best Linear Unbiased Prediction; GBLUP_LD0.2, GBLUP_LD0.4, GBLUP_LD0.6: GBLUP models fitted using SNP subsets pruned at linkage disequilibrium thresholds of $r^2 < 0.2, 0.4$, and 0.6, respectively. KRR: Kernel Ridge Regression; SVR: Support Vector Regression; RF: Random Forest; LightGBM: Light Gradient Boosting Machine; AdaBoost.R2: AdaBoost for Regression

architecture of large and small-effect loci. Overall, our results support the broader application of WGBLUP as a promising strategy for enhancing the accuracy of genetic evaluations in poultry and other livestock species.

A fundamental assumption of SNP chip-based genomic selection is that each quantitative trait locus (QTL) is in LD with at least one nearby SNP marker [38, 39]. However, due to the limited density of commercial SNP arrays, some functional loci—particularly those with low minor allele frequencies or rare mutations—may not be adequately captured. This limitation can compromise the effectiveness of genomic prediction, especially for complex traits influenced by many small-effect variants. Conversely, WGS theoretically enables the detection of nearly all functional mutations across the genome, thereby reducing the reliance on LD and improving the resolution and accuracy of genetic evaluations [38, 39]. To balance predictive performance with economic feasibility, this study employed genotype imputation to elevate the marker density of SNP chip data to a WGS-like level. We then applied LD pruning to construct SNP panels of varying densities based on different LD thresholds. A pronounced contrast exists between the genetic foundations of BW and AFT, a finding consistently upheld by SNP panel analyses and our team's earlier wssGWAS [34]. For BW, the combination of accurate prediction from a sparse LD0.2 panel and a high proportion of genetic variance (19.05%) explained by the top three QTLs strongly suggests an architecture shaped by moderate-effect loci. For AFT, however, the congruent evidence—the need for a dense LD0.6 panel for optimal prediction and the low variance (5.87%) explained by the top QTLs—converge to support a highly polygenic model driven by the cumulative contribution of numerous small-effect variants. These observations are consistent with previous findings [7], reinforcing the idea that the integration of WGS-imputed genotypes with LD pruning offers a powerful and flexible strategy [10, 40] for improving genomic prediction accuracy in poultry. Such approaches not only enhance the capture of functional variation but also allow for customized marker panels tailored to specific traits and populations, supporting more efficient and cost-effective genomic selection in breeding programs.

Beyond this, this study systematically evaluated the performance of several mainstream ML methods in genomic prediction of body weight and abdominal fat thickness traits in chickens, and compared them with the conventional GBLUP model. The results showed that SVR and KRR achieved improvements of 4.15% and 2.10%, respectively, in predictive accuracy for abdominal fat thickness at 32 weeks of age, demonstrating clear advantages. Compared to GBLUP, ML methods are

better suited to handle high-dimensional data with relatively small sample sizes [41], and they do not rely on assumptions of additive genetic architecture. This allows them to flexibly capture potential nonlinear relationships between genotypes and phenotypes, particularly in the presence of non-additive effects such as epistasis. However, for traits predominantly governed by additive effects, linear models like GBLUP often demonstrate greater stability and reliability [39, 42]. These findings are consistent with previous studies. For instance, research in cattle, pigs, and crops has shown that ML approaches may outperform linear models for certain complex traits or under low LD conditions, while GBLUP remains more stable for traits primarily influenced by additive effects [43–45]. One study reported that the prediction accuracy of GBLUP for birth weight and yearling weight in cattle was significantly higher (by 10%–15%) than that of ML models such as RF, SVM, and Convolutional Neural Network (CNN), although RF outperformed GBLUP slightly for weaning weight after fine-tuning [43]. In pigs, Su et al. [44] found that different ML algorithms and GBLUP each had advantages for specific traits, with SVR, KRR, and ensemble methods like Stacking performing well across multiple traits. Similarly, Farooq et al. [45] reported that ML models outperformed rBLUP for disease resistance traits but showed less consistent performance for traits with simple additive architectures.

Notably, although some studies (e.g., Liang et al. [46]) have reported superior performance of ensemble methods such as AdaBoost in other species, Adaboost.R2 and LightGBM did not outperform GBLUP or the leading ML models (e.g., SVR and KRR) for the traits analyzed in this study (see Table S1). Several factors may account for this: (1) the traits studied are mainly controlled by additive genetic effects, limiting the advantage of boosting-based models; (2) the tuning process for Adaboost.R2 is complex, and full optimization was not feasible under current computational resources; and (3) LightGBM can be sensitive to hyperparameters in high-dimensional SNP datasets, increasing the risk of underfitting. Additionally, the performance of RF was likely affected by constraints in parameter tuning and computational efficiency. Taken together, the performance of ML methods further corroborates this architecture-dependent pattern. The significant gains from SVR and KRR for AFT, a trait governed by complex physiology, suggest their capacity to model subtle non-additive relationships that linear models may miss. Conversely, the robust performance of linear models for BW underscores the continued dominance of additive effects for this trait. The failure of ensemble methods like RF and LightGBM likely results from their susceptibility to overfitting in high-dimensional SNP

space without exceedingly large sample sizes, a known challenge in genomic ML.

Collectively, our findings advocate for a trait-aware genomic prediction framework. For traits with evidence of major QTLs, WGBLUP is highly effective. For highly polygenic complex traits, leveraging WGS data with relaxed LD pruning or non-linear ML models like KRR can yield significant gains. For standard additive traits, the computational simplicity of (ss)GBLUP remains compelling. This framework moves beyond one-size-fits-all solutions, guiding breeders to strategically match the analytical tool to the biological question.

Finally, while GS holds immense promise, its cost-effectiveness is crucial for widespread adoption. The combination of non-invasive ultrasound phenotyping with cost-efficient genotype imputation and tailored model selection, as demonstrated here, provides a practical and powerful pathway to accelerate genetic gain for complex traits in poultry breeding. Future efforts should focus on refining this trait-aware framework. The high-density genotypic data generated in this study provides a critical platform for the next phase: linking pivotal genomic signals to candidate genes and functional pathways to deepen mechanistic understanding. Building on this biological foundation, the integration of multi-omics data and the development of pre-hoc metrics to predict trait architecture will ultimately enable fully informed and dynamic genomic selection strategies.

Conclusions

Our study confirms that optimal genomic prediction strategies for Wenchang chickens are trait-specific, dictated by the underlying genetic architecture. We overcame a key barrier in carcass trait breeding by implementing ultrasound imaging as a scalable, non-invasive phenotyping platform. Our comprehensive analysis validated two superior approaches for enhancing prediction of complex traits like abdominal fat thickness: integrating prior biological knowledge via WGBLUP (informed by wssGWAS) and refining marker density from WGS data through LD pruning, both of which substantially outperformed standard GBLUP. Computationally efficient machine learning models, notably KRR and SVR, further delivered competitive accuracy with operational speed. Collectively, we provide a trait-informed framework that empowers breeders to maximize genetic gain efficiently in poultry breeding programs.

Abbreviations

AdaBoost.R2	Adaptive boosting for regression
CV	Cross-validation
DR2	Dosage R-squared measure
EBV	Estimated breeding values
GBLUP	Genomic best linear unbiased prediction
GEBV	Genomic estimated breeding value

GP	Genomic prediction
GS	Genomic selection
GWAS	Genome-wide association study
KRR	Kernel ridge regression
LD	Linkage disequilibrium
LightGBM	Light gradient boosting machine
MAE	Mean absolute error
ML	Machine learning
MSE	Mean squared error
RF	Random forest
SNP	Single nucleotide polymorphism
ssGBLUP	Single-step genomic best linear unbiased prediction
SVR	Support vector regression
WGBLUP	Weighted genomic best linear unbiased prediction
WGS	Whole genome sequence
wssGWAS	Weighted single-step genome-wide association study
y_c	Corrected phenotypes

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40104-026-01384-0>.

Additional file 1: Fig. S1. Validation of ultrasound-based abdominal fat thickness (AFT) measurement. (A) Scatter plot showing the correlation between ultrasound AFT and slaughter-measured abdominal fat weight (AFW). (B) Bland-Altman plot assessing agreement between ultrasound and slaughter measurements for AFW. (C) Scatter plot showing correlation between ultrasound AFT and abdominal fat percentage (AFP). (D) Bland-Altman plot assessing agreement for AFP. Solid lines represent regression lines (A, C) or mean differences (B, D); dashed lines indicate 95% limits of agreement. Statistical parameters including Pearson's correlation coefficient (r), coefficient of determination (R^2), and mean bias are displayed on each panel.

Additional file 2. Detailed methodologies of LightGBM and AdaBoost.R2 models.

Additional file 3: Table S1. Genomic prediction performance of LightGBM and AdaBoost.R2 for BW and AFT at different growth stages based on 10 repeats of 5-fold cross-validation.

Additional file 4: Table S2. Summary of genetic evaluations using LD-pruned WGS-level SNPs.

Acknowledgements

We would like to thank the referees and editor for their valuable comments and suggestions, as well as the careful corrections of our manuscript. The successful completion of this study owes much to the collaborative support from both the Hainan Luoniushan Wenchang Chicken Breeding Farm, whose staff assisted in sample collection, and the State Key Laboratory of Agrobiotechnology and the National Research Facility for Phenotypic and Genotypic Analysis of Model Animals, which provided the high-performance computing platform for high-depth sequencing and data analysis.

Authors' contributions

XMD conceived and designed the experiments. RRZ, YXJ, WYX, YZ, ZYL, and DNG collected the samples and recorded the phenotypes. ZDL and XPW contributed the materials. RRZ analyzed the data and wrote the manuscript. XMD revised the final manuscript. All authors reviewed and approved the manuscript.

Funding

This work is supported by National Key Research and Development Program of China (2023YFF1001100), Hainan Seed Industry Laboratory (B23CJ0521, B21HJ0506), Independent Research Project of State Key Laboratory of Animal Biotech Breeding (2023SKLAB1-7) and the PhD Scientific Research and Innovation Foundation of The Education Department of Hainan Province Joint Project of Sanya Ya Zhou Bay Science and Technology City (HSPHDSRF-2024-05-002). This work is also supported by the 2115 Talent Development program of China Agricultural. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

The datasets used or analyzed during the present study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

The animal study and all animal procedures were approved by the Animal Welfare Committee of China Agricultural University. The whole procedure for blood sample collection was carried out in strict accordance with the protocol approved by the Animal Care and Use Committee of China Agricultural University (Permit Number: DK996).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Sanya Institute, China Agricultural University, Sanya 572025, China. ²State Key Laboratory of Animal Biotech Breeding, Beijing Key Laboratory for Animal Genetic Improvement and Key Laboratory of Animal Genetics, Breeding and Reproduction of the Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. ³Hainan (Tanniu) Wenchang Chicken Co., Ltd., Haikou 570100, China.

Received: 5 November 2025 Accepted: 26 February 2026

Published: 25 April 2026

References

- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819–29. <https://doi.org/10.1093/genetics/157.4.1819>.
- Wang K, Abid MA, Rasheed A, Crossa J, Hearne S, Li H. DNNP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Mol Plant*. 2023;16(1):279–93. <https://doi.org/10.1016/j.molp.2022.11.004>.
- Stewart-Brown BB, Song Q, Vaughn JN, Li Z. Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3 Genes|Genomes|Genetics*. 2019;9(7):2253–65. <https://doi.org/10.1534/g3.118.200917>.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23. <https://doi.org/10.3168/jds.2007-0980>.
- Zhang L, Duan Y, Zhao S, Xu N, Zhao Y. Caprine and ovine genomic selection-progress and application. *Animals (Basel)*. 2024. <https://doi.org/10.3390/ani14182659>.
- Gong J, Zhao J, Ke Q, Li B, Zhou Z, Wang J, et al. First genomic prediction and genome-wide association for complex growth-related traits in rock bream (*Oplegnathus fasciatus*). *Evol Appl*. 2022;15(4):523–36. <https://doi.org/10.1111/eva.13218>.
- Zhuang Z, Wu J, Qiu Y, Ruan D, Ding R, Xu C, et al. Improving the accuracy of genomic prediction for meat quality traits using whole genome sequence data in pigs. *J Anim Sci Biotechnol*. 2023;14:67. <https://doi.org/10.1186/s40104-023-00863-y>.
- Clark SA, van der Werf J. Genomic best linear unbiased prediction (gBLUP) for the estimation of genomic breeding values. *Methods Mol Biol*. 2013;1019:321–30. https://doi.org/10.1007/978-1-62703-447-0_13.
- Song H, Hu H. Strategies to improve the accuracy and reduce costs of genomic prediction in aquaculture species. *Evol Appl*. 2022;15(4):578–90. <https://doi.org/10.1111/eva.13262>.
- Zhang Y, Zhuang Z, Liu Y, Huang J, Luan M, Zhao X, et al. Genomic prediction based on preselected single-nucleotide polymorphisms from genome-wide association study and imputed whole-genome sequence data annotation for growth traits in duroc pigs. *Evol Appl*. 2024;17(2):e13651. <https://doi.org/10.1111/eva.13651>.
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12:186. <https://doi.org/10.1186/1471-2105-12-186>.
- Wang X, Shi S, Wang G, Luo W, Wei X, Qiu A, et al. Using machine learning to improve the accuracy of genomic prediction of reproduction traits in pigs. *J Anim Sci Biotechnol*. 2022;13:60. <https://doi.org/10.1186/s40104-022-00708-0>.
- Rome H, Chu TT, Marois D, Huang C, Madsen P, Jensen J. Accounting for genetic architecture for body weight improves accuracy of predicting breeding values in a commercial line of broilers. *J Anim Breed Genet*. 2021;138(5):528–40. <https://doi.org/10.1111/jbg.12546>.
- Li X, Chen X, Wang Q, Yang N, Sun C. Integrating bioinformatics and machine learning for genomic prediction in chickens. *Genes (Basel)*. 2024;15(6). <https://doi.org/10.3390/genes15060690>.
- Yang R, Xu Z, Wang Q, Zhu D, Bian C, Ren J, et al. Genome-wide association study and genomic prediction for growth traits in yellow-plumage chicken using genotyping-by-sequencing. *Genet Sel Evol*. 2021;53:82. <https://doi.org/10.1186/s12711-021-00672-9>.
- Souza C, Jaimes JJB, Gewehr CE. Equations of prediction for abdominal fat in brown egg-laying hens fed different diets. *Poult Sci*. 2017;96(6):1688–95. <https://doi.org/10.3382/ps/pew421>.
- Chen J, He P, Jiang J, Yang Y, Wang S, Pan C, et al. In vivo prediction of abdominal fat and breast muscle in broiler chicken using live body measurements based on machine learning. *Poult Sci*. 2023;102(1):102239. <https://doi.org/10.1016/j.psj.2022.102239>.
- Li Z, Zheng J, An B, Ma X, Ying F, Kong F, et al. Several models combined with ultrasound techniques to predict breast muscle weight in broilers. *Poult Sci*. 2023;102(10):102911. <https://doi.org/10.1016/j.psj.2023.102911>.
- Sallam M, Goransson L, Larsen A, Wall H, Alhamid W, Gunnarsson S, et al. Genetics of digital phenotypes of keel bone in layer chickens and correlations with keel bone fractures and deviations. *Genet Sel Evol*. 2025;57:69. <https://doi.org/10.1186/s12711-025-01016-7>.
- Lin F, Zhu F, Hao J, Yang F, Hou Z. In vivo prediction of the carcass fatness using live body measurements in Pekin ducks. *Poult Sci*. 2018;97(7):2365–71. <https://doi.org/10.3382/ps/pey079>.
- Barcelos SS, Vargas JAC, Mezzomo R, Gionbelli MP, Gomes DI, Oliveira LRS, et al. Predicting the chemical composition of the body and the carcass of hair sheep using body parts and carcass measurements. *Animal*. 2021;15(3):100139. <https://doi.org/10.1016/j.animal.2020.100139>.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009;84(2):210–23. <https://doi.org/10.1016/j.ajhg.2009.01.005>.
- Misztal I, Tsuruta S, Strabel T, Auvray B, Lee DH. BLUPF90 and related programs (BGF90). Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France. Montpellier: Institut National de la Recherche Agronomique (INRA); 2002.
- Guo X, Christensen OF, Ostensen T, Wang Y, Lund MS, Su G. Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method. *J Anim Sci*. 2015;93(2):503–12. <https://doi.org/10.2527/jas.2014-8331>.
- Brzakova M, Bauer J, Steyn Y, Splichal J, Fulinova D. The prediction accuracies of linear-type traits in czech holstein cattle when using ssGBLUP or wssGBLUP. *J Anim Sci*. 2022. <https://doi.org/10.1093/jas/skac369>.
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res Camb*. 2012;94(2):73–83. <https://doi.org/10.1017/S0016672312000274>.
- Müller AC, Guido S. Introduction to machine learning with Python: A guide for data scientists. Sebastopol, CA: O'Reilly Media, Inc.; 2016.
- Exterkate P, Groenen PJF, Heij C, Dijk DV. Nonlinear forecasting with many predictors using kernel ridge regression. *International Journal of Forecasting*. 2016;32(3):736–53. <https://doi.org/10.1016/j.ijforecast.2015.11.017>.
- Vapnik VN. An overview of statistical learning theory. *IEEE Trans Neural Netw*. 1999. <https://doi.org/10.1109/72.788640>.
- Gonzalez-Camacho JM, Ornella L, Perez-Rodriguez P, Gianola D, Dreisigacker S, Crossa J. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome*. 2018. <https://doi.org/10.3835/plantgenome2017.11.0104>.
- Breiman L. Random forests. *Machine Learning*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.

33. Cai K, Liu R, Wei L, Wang X, Cui H, Luo N, et al. Genome-wide association analysis identify candidate genes for feed efficiency and growth traits in wenchang chickens. *BMC Genomics*. 2024;25:645. <https://doi.org/10.1186/s12864-024-10559-w>.
34. Zhu R, Zhang Y, Jiang Y, Xu Z, Tai Y, Lian Z, et al. Weighted single-step genome-wide association study identified genomic regions and candidate genes for growth and reproductive traits in wenchang chicken. *Poult Sci*. 2025;104(5):104733. <https://doi.org/10.1016/j.psj.2024.104733>.
35. Tiezzi F, Maltecca C. Accounting for trait architecture in genomic predictions of US holstein cattle using a weighted realized relationship matrix. *Genet Sel Evol*. 2015;47:24. <https://doi.org/10.1186/s12711-015-0100-1>.
36. Haque MA, Iqbal A, Alam MZ, Lee Y, Ha J, Kim J. Estimation of genetic correlations and genomic prediction accuracy for reproductive and carcass traits in hanwoo cows. *J Anim Sci Technol*. 2024;66(4):682–701. <https://doi.org/10.5187/jast.2024.e75>.
37. de Las Heras-Saldana S, Lopez BI, Moghaddar N, Park W, Park J, Chung KY, et al. Use of gene expression and whole-genome sequence information to improve the accuracy of genomic prediction for carcass traits in Hanwoo cattle. *Genet Sel Evol*. 2020;52:54. <https://doi.org/10.1186/s12711-020-00574-2>.
38. Fernandes Junior GA, Carvalheiro R, de Oliveira HN, Sargolzaei M, Costilla R, Ventura RV, et al. Imputation accuracy to whole-genome sequence in nellore cattle. *Genet Sel Evol*. 2021;53:27. <https://doi.org/10.1186/s12711-021-00622-5>.
39. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92(2):433–43. <https://doi.org/10.3168/jds.2008-1646>.
40. Wang B, Li P, Hou L, Zhou W, Tao W, Liu C, et al. Genome-wide association study and genomic prediction for intramuscular fat content in suhuai pigs using imputed whole-genome sequencing data. *Evol Appl*. 2022;15(12):2054–66. <https://doi.org/10.1111/eva.13496>.
41. Raymond B, Bouwman AC, Schrooten C, Houwing-Duistermaat J, Veerkamp RF. Utility of whole-genome sequence data for across-breed genomic prediction. *Genet Sel Evol*. 2018;50:27. <https://doi.org/10.1186/s12711-018-0396-8>.
42. Wang X, Zhang Z, Du H, Pfeiffer C, Meszaros G, Ding X. Predictive ability of multi-population genomic prediction methods of phenotypes for reproduction traits in Chinese and Austrian pigs. *Genet Sel Evol*. 2024;56:49. <https://doi.org/10.1186/s12711-024-00915-5>.
43. Hay EH. Machine learning for the genomic prediction of growth traits in a composite beef cattle population. *Animals*. 2024;14(20):3014. <https://doi.org/10.3390/ani14203014>.
44. Su R, Lv J, Xue Y, Jiang S, Zhou L, Jiang L, et al. Genomic selection in pig breeding: comparative analysis of machine learning algorithms. *Genet Sel Evol*. 2025;57:13. <https://doi.org/10.1186/s12711-025-00957-3>.
45. Farooq M, van Dijk ADJ, Nijveen H, Mansoor S, de Ridr D. Genomic prediction in plants: opportunities for ensemble machine learning based approaches. *F1000Res*. 2022;11:802. <https://doi.org/10.12688/f1000research.122437.2>.
46. Liang M, Chang T, An B, Duan X, Du L, Wang X, et al. A stacking ensemble learning framework for genomic prediction. *Front Genet*. 2021;12:600040. <https://doi.org/10.3389/fgene.2021.600040>.