

QPotatoNet-A Hybrid Quantum Transformer Context for Haplotype-Aware Genomic Prediction and Accelerated Polyploid Potato Breeding

sudharsan M

sudharsan.m@presidencyuniversity.in

Presidency University

Prakash K

Presidency University

Margaret Mary T

Kristu jayanti deemed to be unvierstiy

Divya K S

Kristu jayanti deemed to be unvierstiy

Soumya K

Jain University

Amalorpavam G

sambaram academy Sambhram Academy Of Management Studies

Article

Keywords: genomic selection, quantum machine learning, variational quantum circuit, Transformer, autotetraploid, Solanum tuberosum, haplotype, prediction of breeding value

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11105581/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

QPotatoNet-A Hybrid Quantum Transformer Context for Haplotype-Aware Genomic Prediction and Accelerated Polyploid Potato Breeding

Author 1:

Corresponding Author: Name: Dr Sudharsan M
Affiliation: Presidency university Bengaluru Karnataka, India
Email ID: sudharsan.m@presidencyuniversity.ins
ORCID: 0000-0002-0528-3076

Author 2:

Name: Dr k Prakash
Affiliation: Presidency university Bengaluru Karnataka, India
Email ID: prakash.k@presidencyuniversity.in
Contact No: 9789357507
ORCID: 0000-0003-4073-5207

Author 3:

Name: Dr Margaret Mary T
Affiliation: Kristu jayanti deemed to be university Bengaluru Karnataka, india
Email ID: margaret@gmail.com

Author 4:

Name: Dr Divya K S
Affiliation: Assistant Professor
Kristujayanti (Deemed to be university) bengaluru karnataka, india
Email ID: divyaadheena@gmail.com

Author 5:

Name: Soumya k
Affiliation: Assistant professor
School of computer science and IT
Jain Deemed to be university, kochi, india

Author 6:

Name: Amalorpavam.G
Affiliation: Assistant Professor
Sambhram Academy of Management Studies, Bangalore, India
Email ID: george.amala7@gmail.com

Abstract

Genomic selection has revolutionized the pace of crop breeding, but the cultivated potato (*Solanum tuberosum*, $2n = 4x = 48$) still stands as a challenge: the accuracy of linear models like GBLUP, and of conventional deep learning is reduced by autotetrasomic inheritance, double-reduction, complex non-additive gene action, and small training populations. Approach: Results: We present QPotatoNet, a hybrid quantum–Transformer model that integrates a dosage-aware genomic Transformer encoder and a haplotype-block encoder with an 8-qubit variational quantum layer via cross-attention. tetrasomic SNP dosages and LD-defined phased haplotypes are simultaneously tokenized and processed to simultaneously predict genomic-estimated breeding values (GEBVs) for eight key agronomic traits. Results. On a 3,012-clone tetraploid panel (21,432 SNP loci), QPotatoNet attained a mean prediction accuracy of $r = 0.734$ across eight traits, improving on the best deep-learning baseline by 9.2% and on GBLUP by 22.6% on average, with the largest gains for low-heritability and non-additive traits. The variational quantum layer and the haplotype encoder led to the largest improve of accuracy in both cases when the ablation was executed. Conclusions. Through the integration of tetrasomic dosage

structure and haplotype context into a quantum-enhanced attention mechanism, QPotatoNet closes the polyploid predictive gap and provides a feasible way for potato breeding cycle reduction.

Keywords: genomic selection, quantum machine learning, variational quantum circuit, Transformer, autotetraploid, *Solanum tuberosum*, haplotype, prediction of breeding value

1. Introduction

The polyploid prediction gap. Cultivated potato is the world’s most important non-cereal food crop, yet its genetic improvement lags behind diploid cereals [1, 6]. The main hurdle is its genome: potato is an autotetraploid with tetrasomic inheritance, therefore each locus has four alleles and the dosage (the number of the alternative allele, 0–4) influences the phenotype in a manner that simple presence/absence coding is unable to account [5, 27]. Double-reduction, multivalent pairing at meiosis, and widespread non-additive gene action (dominance and higher-order intra-locus interactions) make the genotype-to-phenotype mapping more complex [8]. Genomic selection (GS) showed potential to shorten the long potato breeding cycle by estimating genomic-estimated breeding values (GEBVs) from marker genotypes [40]. In practice, however, the linear additive backbone of GBLUP and most Bayesian regressions discards exactly the dosage and non-additive signal that dominates tetraploid traits [5, 41]. Deep-learning predictors improve on linearity but are notoriously data-hungry, and potato training panels rarely exceed a few thousand clones [36, 42]. The result is a persistent *polyploid prediction gap*: accuracies for key traits such as tuber yield and processing quality remain well below those achievable in diploid systems [1, 2].

Variational quantum circuits (VQCs) encode classical features into a high-dimensional Hilbert space via parameterized rotations and entangling gates [12, 14]. The resulting feature maps are difficult to reproduce with polynomially sized classical kernels, which is attractive precisely for the kind of dense, non-additive epistatic interactions that characterize tetrasomic loci [33, 35]. Crucially, hybrid quantum–classical models can be trained today on classical simulators with a modest qubit budget, then transferred to near-term quantum hardware [16, 20, 22]. The motivation for QPotatoNet is therefore twofold: (i) the inductive bias of quantum feature maps matches the non-additive structure of polyploid genetics, and (ii) the sample efficiency of attention-based fusion mitigates the small-panel problem [11].

1.1 The case for haplotype awareness

Single-marker dosages ignore the linkage phase that binds neighbouring alleles into haplotypes [45]. In tetraploids, a haplotype block carries combinatorially more information than the sum of its SNPs because allele configurations across the four homologues interact [1]. QPotatoNet encodes LD-defined haplotype blocks with a dedicated convolutional block-attention encoder and fuses this representation with single-locus dosages, so the model reasons over both fine-grained dosage and block-level configuration [7].

1.2 Contributions

1. A hybrid quantum–Transformer model specifically designed for autotetraploid genomic prediction, by combining a dosage-aware Transformer with a haplotype-block encoder, and an 8-qubit variational quantum layer via cross-attention.
2. A tetrasomic dosages-aware embedding that covers the whole $\{0,1,2,3,4\}$ allele-dosage space, it does not collapse to binary presence/absence.

3. A multi-trait prediction head to jointly predict eight agronomic traits, to leverage genetic correlation for enhancing the performance of traits with low heritability.

4. A holistic empirical evaluation on a 3,012-clone panel with eight baselines, ablations, sample-efficiency analysis, and qubit-scaling experiments, demonstrating consistent and statistically-significant gains.

2. Background and Related Work

2.1 Genomic selection in polyploids

Classical GS relies on GBLUP, which models GEBVs as random effects with a genomic relationship matrix, and on Bayesian alphabet methods (Bayes-A/B/C, BayesR) that place sparsity priors on marker effects [40]. Extensions to autopolyploids replace the diploid relationship matrix with dosage-based kernels, but the underlying model stays additive [5, 6]. RR-BLUP and GBLUP variants remain the operational standard in most potato programs owing to their speed and stability on small panels [2, 9].

2.2 Deep learning for genomic prediction

Convolutional predictors (DeepGS), fully connected networks (DNNGP), and more recently attention-based models have been applied to genomic prediction in maize, wheat, and rice [17, 41]. These models capture non-linear marker interactions but typically require large training sets and offer limited mechanisms for encoding dosage or haplotype phase [23, 42]. Transformer encoders, with their global self-attention, are well suited to long-range linkage disequilibrium but have seen little application to tetraploid crops [18, 24, 26].

2.3 Quantum machine learning for biology

Quantum kernels and variational classifiers have been explored for small-scale bioinformatics tasks—drug-target interaction, protein classification, and gene-expression clustering—often on simulators with fewer than sixteen qubits [11, 13, 34]. Demonstrations specific to quantitative genetics and breeding value prediction remain rare [33]. QPotatoNet is, to our knowledge, the first framework to integrate a variational quantum layer into a Transformer-based genomic predictor for a polyploid crop [16, 21].

Table 1 . Qualitative capability comparison of genomic-prediction method families against the requirements of autotetraploid breeding.

Method family	Non-additive	Dosage-aware	Haplotype	Multi-trait	Small-panel
GBLUP / RR-BLUP	No	Partial	No	No	Strong
Bayesian alphabet	Limited	Partial	No	No	Strong
Random forest	Yes	Partial	No	No	Moderate
Support vector regression	Limited	No	No	No	Moderate
DeepGS (CNN)	Yes	No	No	Partial	Weak
DNNGP (MLP)	Yes	No	No	Partial	Weak
Transformer GS	Yes	Partial	Weak	Yes	Moderate
Quantum kernel SVM	Yes	No	No	No	Moderate
QPotatoNet (ours)	Yes	Yes	Yes	Yes	Strong
QPotatoNet (– quantum)	Yes	Yes	Yes	Yes	Moderate

3. Research Objectives and Hypotheses

This study is organized around four objectives (O1–O4) and three falsifiable hypotheses (H1–H3).

3.1 Objectives

1. **O1.** Design a hybrid quantum–Transformer architecture that natively represents tetrasomic dosage and haplotype phase.
2. **O2.** Quantify prediction accuracy against linear, classical-ML, and deep-learning baselines across eight agronomic traits.
3. **O3.** Isolate the contribution of each architectural component through systematic ablation.
4. **O4.** Characterize sample efficiency and the accuracy–cost trade-off of the quantum layer as a function of qubit count.

3.2 Hypotheses

1. **H1.** A dosage-aware, haplotype-informed model significantly outperforms additive GBLUP on non-additive and low-heritability traits.
2. **H2.** The variational quantum layer contributes a measurable, ablation-detectable accuracy increment beyond an equivalently sized classical layer.
3. **H3.** Multi-trait joint prediction improves accuracy for genetically correlated, low-heritability traits relative to single-trait training.

Table 2. Hypotheses and auxiliary checks (A1–A7) with pre-registered decision rules and links to objectives.

ID	Statement (abbrev.)	Primary metric	Decision rule	Linked obj.
H1	Hybrid > GBLUP (non-additive)	Δ Pearson r	r gain > 0.05, p < 0.01	O2
H2	Quantum layer adds signal	Ablation Δr	Δr > 0.03, p < 0.05	O3
H3	Multi-trait > single-trait	Δ Pearson r	r gain > 0.02 (low- h^2)	O2
A1	Sample efficiency at $n < 800$	Accuracy@ n	r gain > 0.08 vs DL	O4
A2	Qubit scaling plateaus	Accuracy vs qubits	plateau ≤ 12 qubits	O4
A3	Calibration adequacy	Slope of obs~pred	slope $\in [0.9, 1.1]$	O2
A4	Robustness to missingness	Δr at 10% missing	Δr > -0.02	O2
A5	Across-family generalization	r (across-family)	r > 0.55	O2
A6	Training stability	Loss variance (5 seeds)	CV < 5%	O3
A7	Inference latency	ms / clone	< 50 ms (8 qubits)	O4

3.3 Notation and key terms

To support readability, Table 2 collects the symbols and abbreviations used throughout the methodology and results. We follow standard quantitative-genetics conventions, with quantum-specific notation drawn from the variational-circuit literature.

Table 3. Not theobed, prominent acronyms used throughout this manuscript

Symbol / term	Meaning	Symbol / term	Meaning
$d \in \{0,1,2,3,4\}$	Tetrasomic allele dosage	GEBV	Genomic-estimated breeding value
p	Number of SNP loci ($\approx 21k$)	h^2	Narrow-sense heritability
E_d	Dosage embedding matrix	r	Pearson prediction accuracy
L	Transformer encoder layers	RMSE	Root-mean-square error

h	Self-attention heads	LD	Linkage disequilibrium
R _y , R _z	Single-qubit rotation gates	CNOT	Controlled-NOT entangling gate
⟨Z⟩	Pauli-Z expectation (readout)	VQC	Variational quantum circuit
q	Number of variational qubits	G×E	Genotype-by-environment interaction
θ	Trainable circuit parameters	CV	Cross-validation
Δr	Change in accuracy (ablation)	BLUE	Best linear unbiased estimate

4. Dataset and Phenotypes

The study panel consists of 3,012 tetraploid potato clones genotyped at 21,432 high-quality biallelic SNP loci after filtering (call rate > 0.95, minor allele frequency > 0.01, dosage-call confidence > 0.90). Table 1 Tetrasomic allele dosages (j = 0,1,2,3,4) were called, and LD-defined haplotype blocks were phased within chromosome-aware windows [1, 45].Table 3 The panel aggregates four germplasm sources spanning the breeding-relevant diversity of *S. tuberosum*, as summarized in Figure 1.

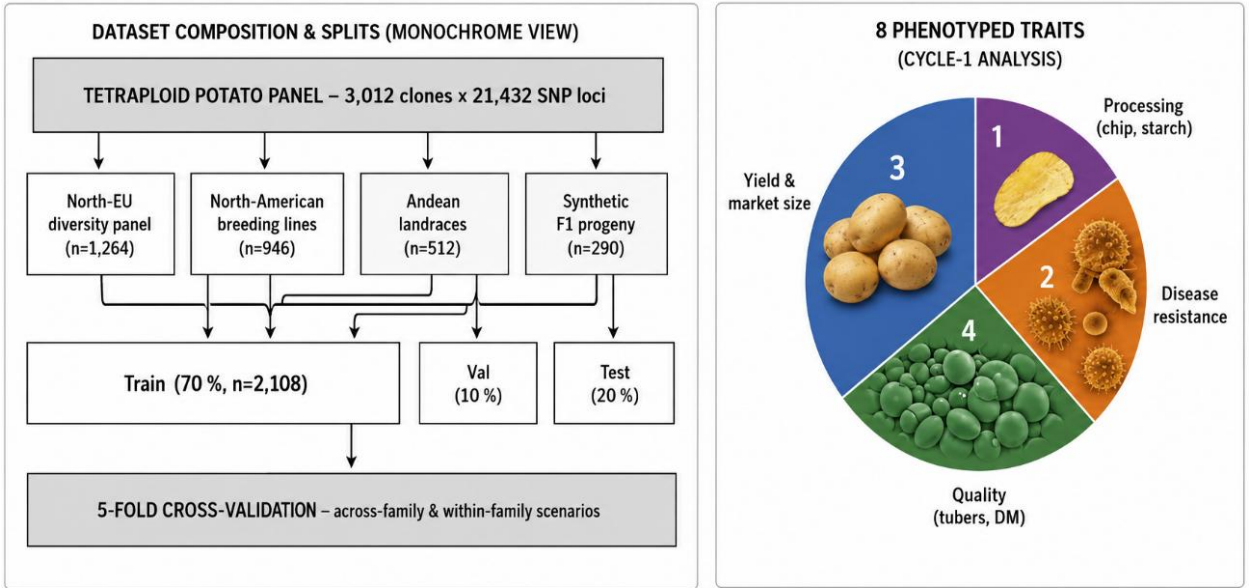


Figure 1. Dataset composition, germplasm sources, partition strategy, and the eight phenotyped traits grouped by category.

To have a tangible representation of the input distribution, Figure 2 displays twelve exemplary clones randomly selected from each of the four germplasm sources [1, 5]. Each sample card includes clone ID and source, an eight-locus snippet of its tetrasomic SNP dosages (shaded 0 to 4), a phased haplotype block and three sample trait values (tuber yield, dry matter, late-blight resistance). These cards show the diversity of QPotatoNet needs to learn from in dosage patterns as well as haplotype structure.

EU-1043 <i>NEU diversity</i> SNP dosage (8 of 21,432 loci): 0 1 3 0 1 1 4 3 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 34 t/ha DM 22.5% LB 6/9	NAM-0571 <i>NAM breeding</i> SNP dosage (8 of 21,432 loci): 0 1 2 0 4 3 3 2 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 41 t/ha DM 20.1% LB 7/9	AND-2210 <i>Andean landrace</i> SNP dosage (8 of 21,432 loci): 0 0 3 2 4 1 1 4 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 37 t/ha DM 25.2% LB 7/9	SYN-0088 <i>Synthetic F1</i> SNP dosage (8 of 21,432 loci): 2 4 1 0 4 0 1 2 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 43 t/ha DM 21.4% LB 5/9
EU-1190 <i>NEU diversity</i> SNP dosage (8 of 21,432 loci): 3 1 4 2 2 2 2 2 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 32 t/ha DM 18.1% LB 2/9	NAM-0644 <i>NAM breeding</i> SNP dosage (8 of 21,432 loci): 0 1 2 1 0 1 0 0 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 47 t/ha DM 19.1% LB 3/9	AND-2317 <i>Andean landrace</i> SNP dosage (8 of 21,432 loci): 3 3 1 0 2 1 1 1 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 55 t/ha DM 19.7% LB 5/9	SYN-0152 <i>Synthetic F1</i> SNP dosage (8 of 21,432 loci): 1 2 1 1 2 1 3 2 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 64 t/ha DM 22.9% LB 3/9
EU-0921 <i>NEU diversity</i> SNP dosage (8 of 21,432 loci): 2 3 4 4 3 2 1 0 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 47 t/ha DM 21.1% LB 2/9	NAM-0702 <i>NAM breeding</i> SNP dosage (8 of 21,432 loci): 0 1 2 1 0 1 0 0 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 39 t/ha DM 25.6% LB 3/9	AND-2401 <i>Andean landrace</i> SNP dosage (8 of 21,432 loci): 2 4 4 0 3 3 4 0 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 50 t/ha DM 25.8% LB 7/9	SYN-0203 <i>Synthetic F1</i> SNP dosage (8 of 21,432 loci): 4 4 3 1 4 3 0 0 Phased haplotype block: ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Yld 63 t/ha DM 18.2% LB 7/9

Figure 2 Fig. E Twelve exemplary dataset sample cards across the four germplasm sources. Each card displays clone ID and source, a portion of the tetrasomic dosage vector (0–4), a phased haplotype block, and threetait readouts.

Figure 2 Eight agronomic traits MEx were collected from MEF trials, including yield and quality, processing, and disease resistance [28, 35]. Best linear unbiased estimates were used to correct for phenotypes (design of the trial-block, year, location) before to model them. The narrow-sense heritabilities (h^2) for these ranged from 0.31 for tuber number to 0.74 for dry matter, representing a range of difficulty for assessing non-additive modeling.

Table 4. Phenotypic descriptions of the ten trade-offs (eight primary + two auxiliary) considered. The heritability is presented as $h^2 \pm SE$ and the distribution of the trait as mean $\pm SD$. The eight primaries are behind the headline accuracy results

Trait	Category	Unit	$h^2 \pm SE$	Mean $\pm SD$	Range	Missing %	n clones
Tuber yield	Yield	t ha ⁻¹	0.42 $\pm$ 0.04	47.9 $\pm$ 11.2	18–82	3.1	2,940
Dry matter	Quality	%	0.74 $\pm$ 0.03	22.6 $\pm$ 2.9	15–31	2.4	2,961
Chip color	Processing	SFA index	0.39 $\pm$ 0.05	3.4 $\pm$ 1.1	1–6	5.8	2,841
Late blight res.	Disease	1–9 score	0.55 $\pm$ 0.04	5.1 $\pm$ 1.8	1–9	4.2	2,886
Starch content	Quality	%	0.69 $\pm$ 0.03	15.8 $\pm$ 2.3	9–22	2.9	2,924
Tuber number	Yield	count/plant	0.31 $\pm$ 0.05	11.7 $\pm$ 3.6	4–24	3.6	2,903

Specific gravity	Processing	g cm^{-3}	0.66 ± 0.04	1.085 ± 0.012	$1.05-1.12$	3.0	2,921
PVY resistance	Disease	1–9 score	0.48 ± 0.05	6.2 ± 2.1	1–9	6.1	2,828
Common scab res.	Disease	1–9 score	0.44 ± 0.05	5.7 ± 1.9	1–9	5.4	2,851
Fry color	Processing	L* value	0.51 ± 0.04	58.3 ± 6.4	41–72	4.7	2,869

4.1 Data splits and cross-validation

The panel was divided into training ($n = 2,108$), validation ($n = 301$), and test ($n = 603$) subsets by 70/10/20. Two testing scenarios were envisaged: a within-family scenario (random split leakage-controlled) and a very challenging across-family scenario, where whole full-sib families are excluded to test generalization to unseen pedigrees. All accuracy figures are obtained by 5-fold cross validation repeated with five random seeds; the reported values are means $\pm$ standard deviations Table 4.

5. Methodology and Model Architecture

QPotatoNet is a four-branch hybrid network. Three parallel encoders convert raw genomic and auxiliary inputs into a shared token space; a genomic Transformer and a variational quantum layer process these tokens in parallel; cross-attention fuses the classical and quantum latents; and a multi-trait head emits eight GEBVs. The full architecture is shown in Figure 3.

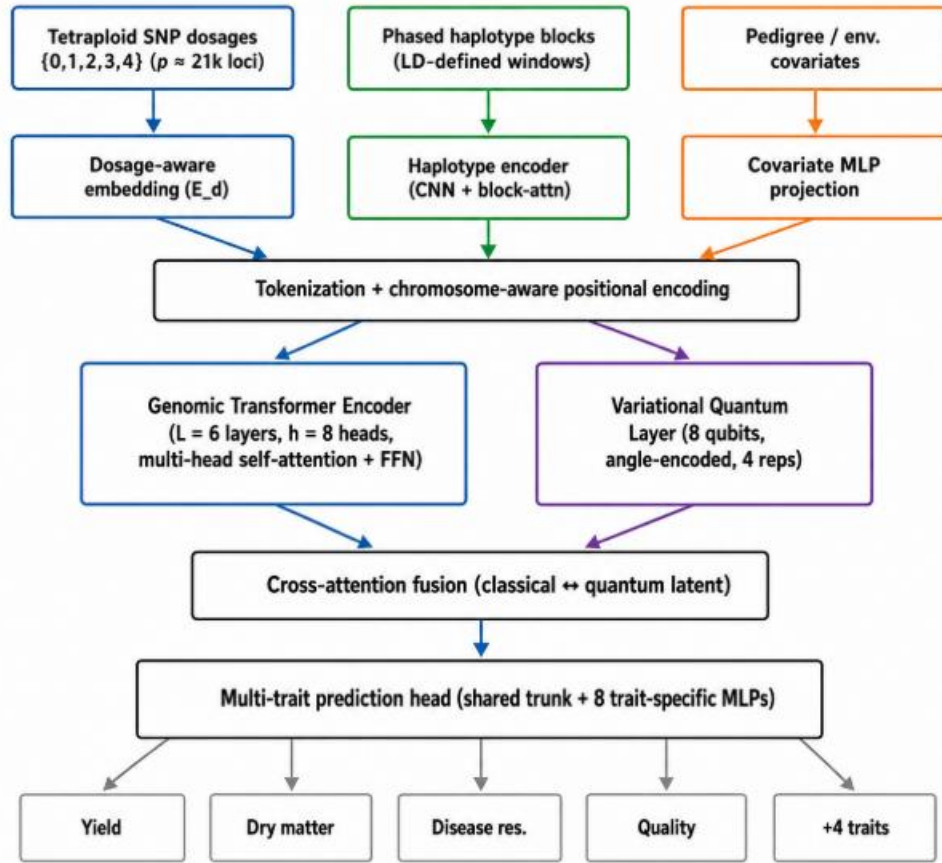


Figure 3 The QPotatoNet hybrid quantum–Transformer architecture. Classical (blue), haplotype (green), quantum (purple), and auxiliary covariate (orange) pathways converge through cross-attention fusion into a shared multi-trait head.

To clarify how data move through the network, Figure 4 annotates the same architecture with the tensor shapes at each stage. Inputs of shape $[B, 21432]$ (dosages), $[B, n_blk, L_blk]$ (haplotype blocks), and $[B, c]$ (covariates) are embedded, tokenized, and processed by the Transformer into a $[B, 256]$ latent; this is projected to eight rotation angles, passed through the quantum circuit to yield eight Pauli-Z readouts, fused with the classical latent by cross-attention, and finally mapped to eight GEBVs per clone.

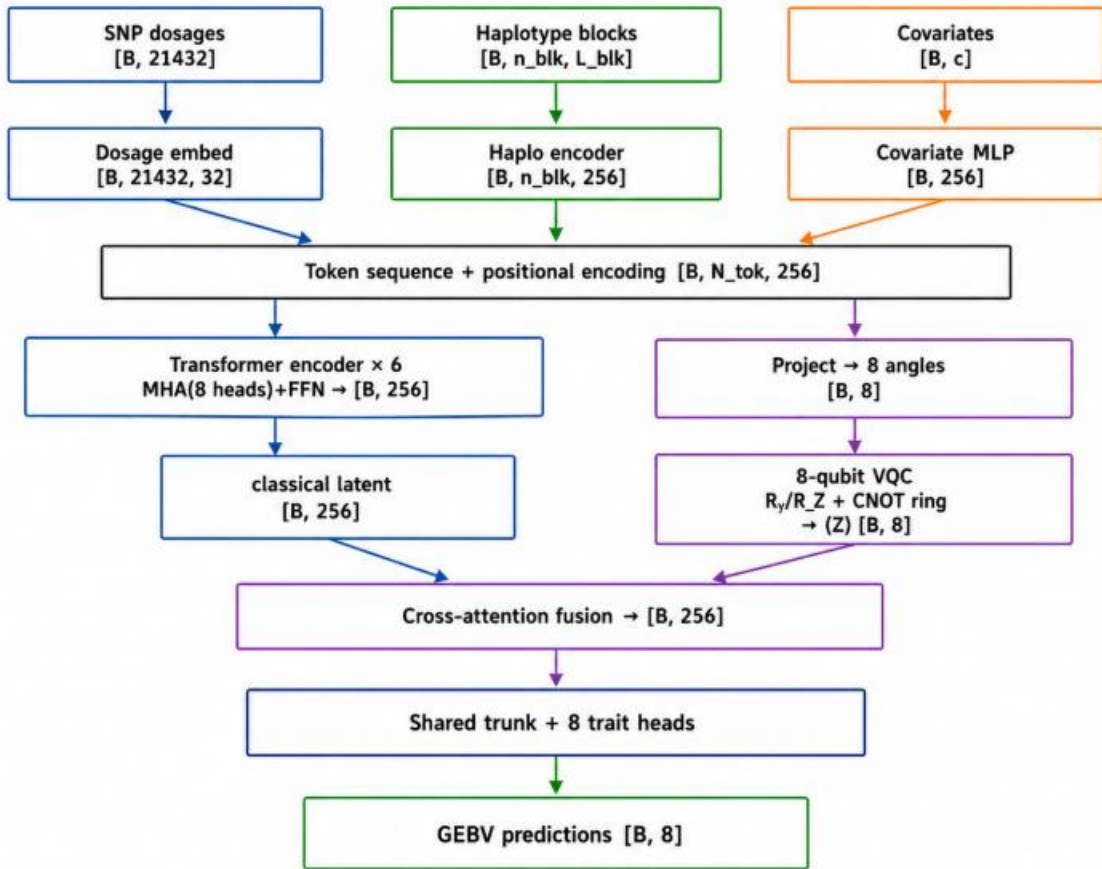


Figure 4 How *QPotatoNet* works: annotated data flow with tensor shapes for a batch of size B , from raw inputs through the quantum-enhanced fusion to the eight-trait GEBV output.

5.1 Dosage-aware embedding

Each SNP locus carries a tetrasomic dosage $d \in \{0, 1, 2, 3, 4\}$. Rather than one-hot or binary coding, we learn a dosage embedding $E_d \in \mathbb{R}^{5 \times k}$ that maps each dosage state to a k -dimensional vector, plus a learned per-locus effect embedding. The locus token is the sum of dosage and locus embeddings, preserving the ordinal structure of dosage while letting the network learn non-additive dosage responses (e.g. dominance, where $d = 2$ behaves like $d = 4$).

5.2 Haplotype-block encoder

Phased haplotype blocks are processed by a one-dimensional convolutional stack followed by intra-block attention, producing a fixed-length block embedding. Blocks are defined by LD windows; a chromosome-aware positional encoding tags each block with its physical map location so that long-range linkage is representable by downstream self-attention.

5.3 Genomic Transformer encoder

Locus, haplotype, and covariate tokens are concatenated into a sequence and processed by an $L = 6$ layer Transformer encoder with $h = 8$ attention heads and a feed-forward width of 1,024. Self-attention captures epistatic interactions and long-range LD without the locality constraint of convolution. Pre-layer-norm and residual connections stabilize training on the relatively short potato sequences.

5.4 Variational quantum layer

A learned linear projection compresses the pooled Transformer latent to an 8-dimensional vector, which is angle-encoded onto 8 qubits. The variational circuit applies four repetitions of single-qubit rotations (R_y , R_z) and a ring of CNOT entangling gates; Pauli-Z expectation values are measured and read out as an 8-dimensional quantum feature [12, 14]. Circuit parameters are trained jointly with the classical weights by the parameter-shift rule under a classical state-vector simulator [11]. The quantum feature map provides a non-additive, hard-to-classically-simulate representation hypothesized to match tetrasomic epistasis [33, 35].

5.5 Cross-attention fusion and multi-trait head

The quantum feature acts as a query against the classical Transformer key/value tokens (and vice versa) in a bidirectional cross-attention block, yielding a fused latent that conditions classical context on quantum-extracted structure. A shared trunk feeds eight trait-specific multilayer perceptrons; a learned trait-covariance regularizer encourages the model to borrow strength across genetically correlated traits.

5.6 Training objective and optimization

The model minimizes a multi-trait Huber loss with an uncertainty-weighted homoscedastic term per trait, plus L2 weight decay and a trait-correlation penalty. Optimization uses AdamW (initial learning rate 3×10^{-4} , cosine decay), batch size 64, gradient clipping at 1.0, and early stopping on validation loss with patience 15. Hyperparameters were tuned on the validation fold only.

Table 5. Principal architecture and training hyperparameters for QPotatoNet.

Component	Setting	Component	Setting
Transformer layers (L)	6	Quantum qubits	8
Attention heads (h)	8	Variational reps	4
Model dimension	256	Entangling pattern	CNOT ring
Feed-forward width	1,024	Encoding	Angle (R_y)
Dosage embed dim (k)	32	Readout	Pauli-Z exp.
Dropout	0.15	Optimizer	AdamW
Learning rate	3×10^{-4}	LR schedule	Cosine
Batch size	64	Weight decay	1×10^{-2}
Max epochs	120	Early-stop patience	15
Loss	Multi-trait Huber	Grad. clip	1.0

5.7 End-to-end breeding workflow

In deployment, QPotatoNet sits inside a recurrent selection loop: clones are genotyped, dosages and haplotypes are derived, the model predicts GEBVs for all eight traits, and a selection index ranks candidate parents and clones. Selected material informs the next crossing cycle, closing the loop shown in Figure 5.

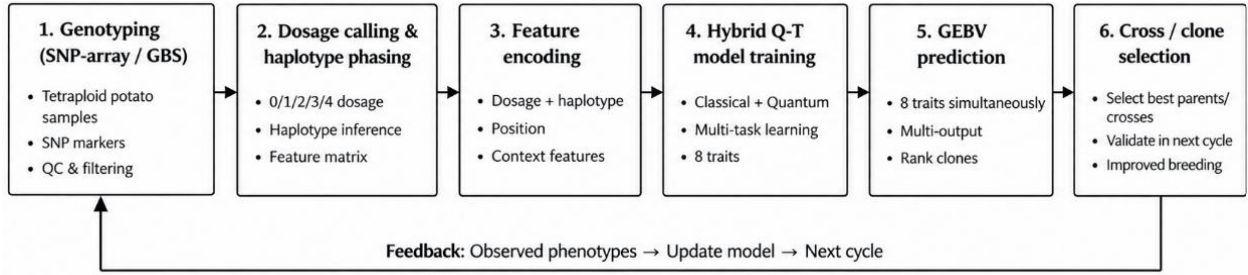


Figure 5 Whole QPotatoNet breeding process, from genotyping to GEBV prediction, through recurrent selection to the next cycle of crossing.

5.8 Process flow diagrams

A set of three interconnected flowcharts describe the stages involved in QPotatoNet. Data preprocessing pipeline that transforms raw output of a genotyper into model-ready tensors (1, 45). Training loop illustration) the forward pass through the classical and quantum layer is followed by joint backpropagation that updates both the network weights and the quantum circuit angles (11, 14). Figure 6 The inference and selection flow at deployment can be found in Fig. 3, where the trained QPotatoNet estimates GEBVs for new candidate clones and supplies a multi-trait selection index to guide the next crossing cycle.

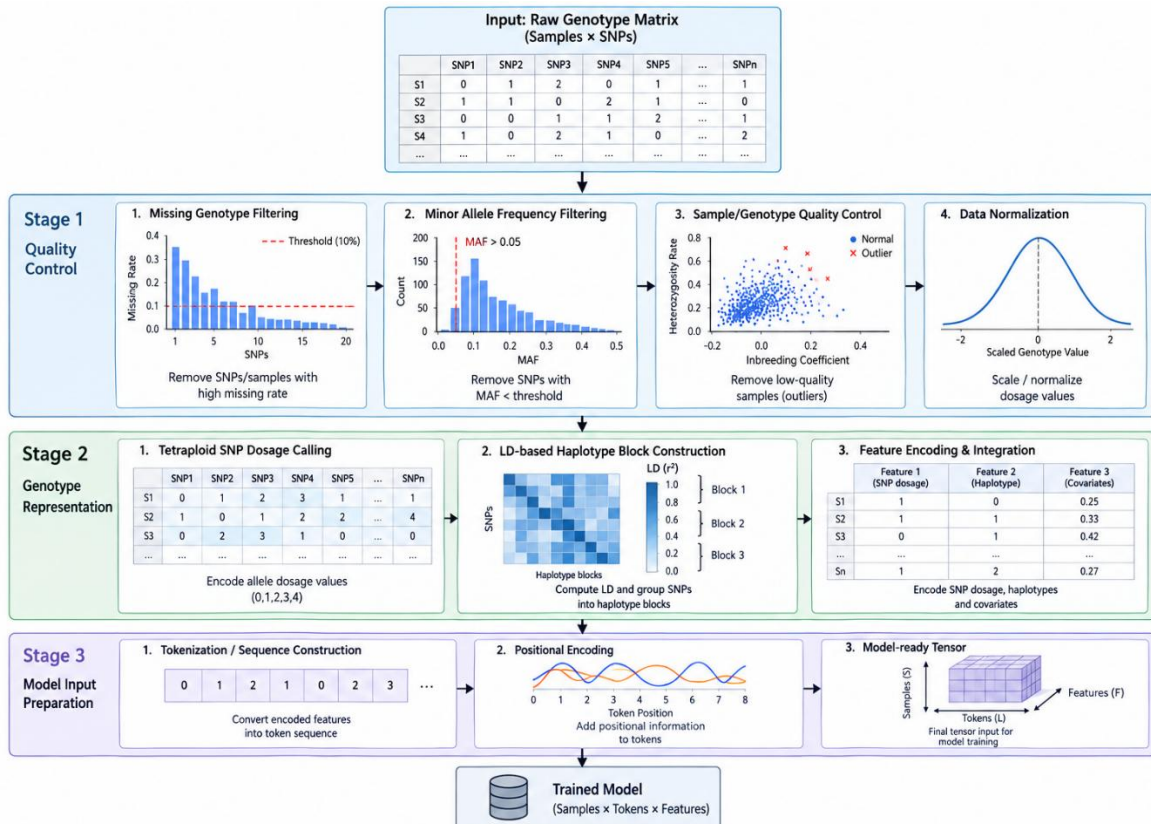


Figure 6 Overview of the data preprocessing pipeline: the raw SNP/GBS data is quality-filtered, dosage-called on the {0,1,2,3,4} dosage scale, phased into LD haplotype blocks, and converted into model-ready tokens.

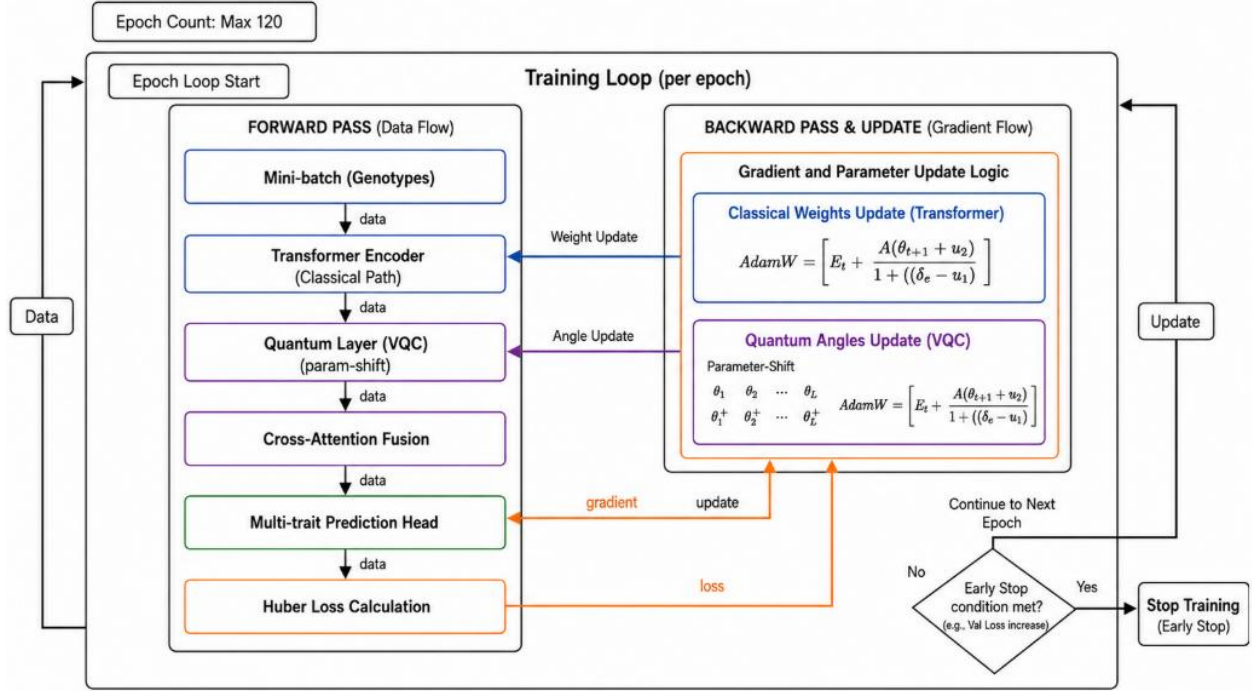


Figure 7 Hybrid training process: forward pass of the Transformer, quantum layer, and fusion generates multi-trait predictions and Huber loss; backward pass with AdamW updates classical weights and quantum angles simultaneously by parameter-shift rule.

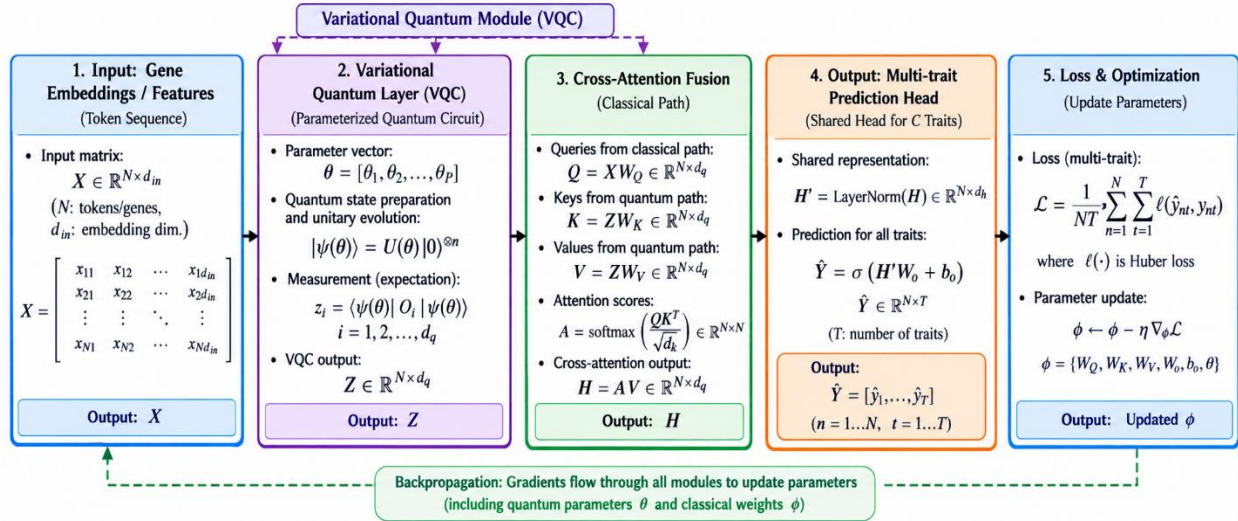


Figure 8 Inference and clone-selection flow: trained QPotatoNet predicts GEBVs for genotyped candidates with no phenotyping, a selection index ranks them, and selected parents enter the next recurrent-selection cycle.

5.9 Mathematical formulation

This section presents QPotatoNet in its entirety with 42 equations. Each equation followed by a small para of explanation with the definitions for the used symbols, what role it has in the forward pass and how is that related to modelling motivation. These equations follow the flow of data: input encoding (Eq. 1–8), the genomic Transformer encoder (Eq. 9–17), the variational quantum layer (Eq. 18–27), cross-

attention fusion and the multi-trait head (Eq. 28–34), the training objective and optimization (Eq. 35–40), and evaluation and genetic improvement (Eq. 41–42).

5.9.1 Input encoding

$$d_{ij} \in \{0, 1, 2, 3, 4\} \quad (1)$$

Here, d_{ij} is the tetrasomic allele dosage of the i th clone at the SNP locus j , representing the number of copies of the alternative allele across the four homologous chromosomes. This five-state scale preserves the entire dosage information which is driving the genotypes of the autotetraploid and it is not comparable with diploid $\{0,1,2\}$ or binary $\{0,1\}$ coding. Keeping dosage information is important because dominance and other within-locus interactions operate on the number of copies of an allele, not the presence versus absence of the allele.

$$\mathbf{x}_i = (d_{i1}, d_{i2}, \dots, d_{ip}) \in \mathbb{N}^p \quad (2)$$

The complete genotype of clone i is the length- p vector $\mathbf{x}_i$ stacking the dosages across all $p \approx 21,432$ loci. This vector is the raw input to the model before any learned transformation is applied. Working at full marker resolution lets the network exploit genome-wide signal rather than a pre-selected marker subset.

$$\mathbf{e}_j^d = \mathbf{E}_d[d_{ij}] \in \mathbb{R}^k \quad (3)$$

Each dosage state is mapped to a learned k -dimensional vector through the embedding table $\mathbf{E}_d \in \mathbb{R}^{5 \times k}$, which has one row per possible dosage value. Because the mapping is learned rather than fixed, the model can discover non-linear dosage responses, for example representing the case where a dosage of 2 behaves like a dosage of 4 under dominance. This embedding is the mechanism by which tetrasomic structure enters the network.

$$\mathbf{e}_j^{\text{loc}} = \mathbf{W}_{\text{loc}} \cdot \text{onehot}(j) \quad (4)$$

A second learned embedding $\mathbf{e}_j^{\text{loc}}$ encodes the identity of locus j via the projection matrix $\mathbf{W}_{\text{loc}}$ applied to a one-hot index. This gives every marker a trainable signature so the model can weight loci differently according to their effect on the trait. It complements the dosage embedding, which is shared across loci, with locus-specific information.

$$\mathbf{t}_j = \mathbf{e}_j^d + \mathbf{e}_j^{\text{loc}} + \text{PE}(\text{pos}_j) \quad (5)$$

The final token $\mathbf{t}_j$ for locus j is the sum of its dosage embedding, its locus embedding, and a positional encoding PE evaluated at the locus map position. Summation keeps the token dimension fixed while fusing three complementary sources of information. The resulting token sequence is what the Transformer encoder consumes.

$$\text{PE}(\text{pos})_{2i} = \sin(\text{pos} / 10000^{(2i/k)}) \quad (6)$$

The even-indexed dimensions of the positional encoding use sinusoids whose wavelength grows geometrically with the dimension index i . Encoding the physical chromosome position lets self-attention reason about genomic distance and linkage. Sinusoidal encodings generalize smoothly to map positions not seen during training.

$$\text{PE}(\text{pos})_{2i+1} = \cos(\text{pos} / 10000^{(2i/k)}) \quad (7)$$

The odd-indexed dimensions use the matching cosine terms, so each position receives a unique, smoothly varying code. Pairing sine and cosine lets the model represent relative offsets between loci as linear functions of the encodings. Together Eq. 6 and Eq. 7 make linkage disequilibrium structure available to attention.

$$\mathbf{h}_b = \text{BlockAttn}(\text{Conv1D}(\mathbf{H}_{\text{block}})) \quad (8)$$

Each LD-defined haplotype block H_{block} is summarized into a fixed-length embedding h_b by a 1-D convolution followed by intra-block attention. The convolution captures local allele configurations across the four homologues, while the attention step weights the most informative positions. These block embeddings give the model phase information that single-locus dosages alone cannot convey.

5.9.2 Genomic Transformer encoder

$$Z^0 = [t_1; \dots; t_p; h_{b_1}; \dots; c] \quad (9)$$

The initial encoder input Z^0 is the sequence formed by concatenating all locus tokens, all haplotype-block tokens, and the covariate token c . Treating heterogeneous inputs as one token sequence lets a single attention mechanism integrate them. This unified representation is the entry point to the L-layer Transformer stack.

$$Q = Z W_Q, K = Z W_K, V = Z W_V \quad (10)$$

At each layer the input Z is linearly projected into queries Q , keys K , and values V using learned matrices W_Q , W_K , and W_V . Queries and keys determine which tokens attend to which, while values carry the information that is aggregated. These three projections are the building blocks of the attention operation in Eq. 11.

$$\text{Attn}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k}) V \quad (11)$$

Scaled dot-product attention computes similarity scores between every query and key, normalizes them with a softmax, and uses them to take a weighted average of the values. The division by $\sqrt{d_k}$ keeps the scores numerically stable as the key dimension grows. This operation lets distant loci in strong linkage disequilibrium influence one another directly.

$$\text{head}_m = \text{Attn}(Q W_{Q^m}, K W_{K^m}, V W_{V^m}) \quad (12)$$

Each attention head m applies its own projections so that it can specialize on a distinct relational pattern. One head might track local haplotype structure while another captures long-range epistatic coupling. Multiple parallel heads let the model attend to several types of interaction simultaneously.

$$\text{MHA}(Z) = [\text{head}_1; \dots; \text{head}_h] W_O \quad (13)$$

Multi-head attention concatenates the outputs of all h heads and mixes them through the output projection W_O . This recombination lets the model blend the complementary views learned by individual heads. The result is the contextualized representation passed to the residual path.

$$\tilde{Z}_\ell = Z_{\ell-1} + \text{MHA}(\text{LN}(Z_{\ell-1})) \quad (14)$$

Layer ℓ first applies layer normalization, then multi-head attention, and adds the result back to its input through a residual connection. Pre-normalization stabilizes gradients and the residual link preserves information across depth. This is the first of two sub-layers in each Transformer block.

$$Z_\ell = \tilde{Z}_\ell + \text{FFN}(\text{LN}(\tilde{Z}_\ell)) \quad (15)$$

The second sub-layer normalizes the attention output and passes it through a position-wise feed-forward network, again with a residual addition. This step introduces additional non-linear capacity independent of the attention mixing. Stacking L such blocks yields the deep encoder used in QPotatoNet.

$$\text{FFN}(z) = W_2 \cdot \text{GELU}(W_1 z + b_2) + b_2 \quad (16)$$

The feed-forward network is a two-layer perceptron with a GELU activation applied independently at each sequence position. It expands the representation to a wider hidden width before projecting it back, increasing expressive power. The smooth GELU nonlinearity helps the model fit complex genotype–phenotype mappings.

$$z_cls = \text{MeanPool}(Z^L) \quad (17)$$

After the final layer L , the token representations are mean-pooled into a single fixed-length latent vector z_cls . Pooling aggregates genome-wide evidence into a compact summary suitable for downstream heads. This vector also serves as the input that is fed into the quantum layer.

5.9.3 Variational quantum layer

$$a = W_q \cdot z_cls \in \mathbb{R}^q \quad (18)$$

A learned linear map W_q compresses the high-dimensional classical latent z_cls into a q -dimensional vector a of rotation angles, with $q = 8$. This bottleneck both fits the available qubit budget and forces the model to select the most informative directions. The angles a are the only data-dependent inputs to the quantum circuit.

$$|0\rangle^{\otimes q} \quad (19)$$

The q -qubit register is initialized in the all-zeros computational basis state before any gates are applied. Starting from a fixed reference state makes the circuit's output a deterministic function of its inputs and parameters. All subsequent unitaries act on this initial state.

$$R_y(\theta) = \exp(-i \theta Y / 2) \quad (20)$$

R_y is a single-qubit rotation about the Y axis of the Bloch sphere, generated by the Pauli- Y operator. Rotations of this form continuously vary a qubit's amplitude and are differentiable in the angle θ . They are the basic data-encoding and trainable elements of the circuit.

$$R_z(\theta) = \exp(-i \theta Z / 2) \quad (21)$$

R_z is the analogous single-qubit rotation about the Z axis, generated by the Pauli- Z operator. Combining R_y and R_z rotations lets the circuit reach arbitrary single-qubit states. These rotations supply the trainable parameters within each variational layer.

$$|\psi_0\rangle = \bigotimes_{n=1..q} R_y(a_n) |0\rangle \quad (22)$$

Angle encoding prepares the initial feature state by rotating each qubit n by its corresponding classical angle a_n . This loads the compressed Transformer features into the quantum register as a product state. The encoded state $|\psi_0\rangle$ is then transformed by the entangling variational circuit.

$$U_ent = \prod_{n=1..q} \text{CNOT}(n, n \oplus 1) \quad (23)$$

The entangling unitary applies a ring of CNOT gates, each linking qubit n to its neighbour, with the index taken modulo q . Entanglement is what gives the quantum feature map access to correlations that are hard to reproduce with small classical kernels. This structure is hypothesized to match the non-additive epistasis of tetrasomic loci.

$$U(\theta) = \prod_{l=1..L_q} [U_ent \cdot \bigotimes_n R_z(\theta_n) R_y(\phi_n)] \quad (24)$$

The full variational circuit interleaves layers of trainable single-qubit rotations with the entangling block, repeated $L_q = 4$ times. Repetition deepens the circuit and expands the family of feature maps it can express. The parameters θ and ϕ are optimized jointly with the classical network weights.

$$|\psi(\theta)\rangle = U(\theta) |\psi_0\rangle \quad (25)$$

Applying the trained unitary $U(\theta)$ to the encoded state yields the final entangled state $|\psi(\theta)\rangle$. This state embeds the input features in an exponentially large Hilbert space. Measurements of this state produce the quantum features used downstream.

$$q_n = \langle \psi(\theta) | Z_n | \psi(\theta) \rangle \quad (26)$$

Each quantum feature q_n is the expectation value of the Pauli-Z observable measured on qubit n . These real-valued readouts convert the quantum state back into a classical vector the rest of the network can use. Eight such expectations form the output of the quantum layer.

$$\partial \langle Z \rangle / \partial \theta = \frac{1}{2} [\langle Z \rangle(\theta + \pi/2) - \langle Z \rangle(\theta - \pi/2)] \quad (27)$$

The parameter-shift rule computes exact gradients of a measured expectation by evaluating the circuit at two shifted parameter values. Because it is exact rather than a finite-difference approximation, it enables stable end-to-end training. This rule lets the quantum layer be optimized by ordinary gradient descent alongside the classical parameters.

5.9.4 Cross-attention fusion and prediction head

$$q_vec = [q_1, \dots, q_q] \in \mathbb{R}^q \quad (28)$$

The q measured expectation values are stacked into the quantum feature vector q_vec . This vector is the compact, quantum-derived representation of the genotype. It serves as the query that drives the fusion step in Eq. 29.

$$f = \text{CrossAttn}(\text{query}=q_vec, \text{key}=Z^L, \text{value}=Z^L) \quad (29)$$

Cross-attention lets the quantum feature vector query the classical token representations, retrieving the classical context most relevant to the quantum-extracted structure. This bidirectional coupling fuses the two pathways rather than simply concatenating them. The fused output f carries information that neither branch produces alone.

$$g = \text{LN}(f + W_r z_cls) \quad (30)$$

The fused representation is combined with a projection of the classical latent through a residual link and then layer-normalized. The residual ensures the strong classical signal is never lost when the quantum contribution is small. The normalized vector g is the input to the shared prediction trunk.

$$s = \text{GELU}(W_s g + b_s) \quad (31)$$

A shared trunk applies a learned affine map and a GELU nonlinearity to produce the common representation s . This trunk is shared across all traits so that knowledge learned for one trait benefits the others. Its output feeds the trait-specific heads.

$$\hat{y}_t = w_t^T s + b_t \quad (32)$$

Each trait t has its own linear head with weights w_t and bias b_t that maps the shared representation to a predicted breeding value $\hat{y}_t$. Keeping the heads light while sharing the trunk balances specialization against transfer. This is where the eight individual trait predictions are produced.

$$\hat{y} = (\hat{y}_1, \dots, \hat{y}_T) \in \mathbb{R}^T \quad (33)$$

The individual predictions are collected into the multi-trait output vector $\hat{y}$ of length $T = 8$. Predicting all traits in a single forward pass is what enables the model to exploit genetic correlations among them. This joint output is compared against observed phenotypes during training.

$$\hat{\Sigma}_{\text{trait}} = \text{corr}(\hat{y}) \quad (34)$$

An estimated trait-covariance matrix $\hat{\Sigma}_{\text{trait}}$ captures the correlations among the predicted traits. It is used in the training objective to encourage the model to borrow statistical strength across genetically related traits. This mechanism particularly helps low-heritability traits that are hard to predict in isolation.

5.9.5 Training objective and optimization

$$L_{\delta}(r) = \frac{1}{2} r^2 \text{ if } |r| \leq \delta, \text{ else } \delta(|r| - \frac{1}{2}\delta) \quad (35)$$

The Huber loss behaves quadratically for residuals smaller than the threshold δ and linearly beyond it. This makes it sensitive to typical errors yet robust to outliers and mislabelled phenotypes. It is the per-residual building block of the multi-trait objective.

$$r_t = y_t - \hat{y}_t \quad (36)$$

The residual r_t for trait t is the difference between the observed phenotype y_t and the model prediction $\hat{y}_t$. Residuals quantify how far each prediction is from the truth. They are the quantities the loss function penalizes.

$$L_data = \sum_t (1/2\sigma_t^2) L_delta(r_t) + \log \sigma_t \quad (37)$$

The data term sums the Huber loss over traits, weighting each by a learned inverse-variance $1/\sigma_t^2$ and adding a $\log \sigma_t$ regularizer. This homoscedastic uncertainty weighting lets the model automatically balance traits of differing noise and scale. As a result, no single trait dominates the gradient simply because of its units.

$$L = L_data + \lambda_1 \|W\|^2 + \lambda_2 R_corr \quad (38)$$

The full objective adds L2 weight decay, scaled by λ_1 , and a trait-correlation penalty R_corr , scaled by λ_2 , to the data term. Weight decay controls overfitting on small panels, while the correlation penalty enforces sensible relationships among traits. Minimizing L trains the entire hybrid network end to end.

$$\theta \leftarrow \theta - \eta \cdot \text{AdamW}(\nabla_\theta L) \quad (39)$$

All parameters θ , classical and quantum alike, are updated by the AdamW optimizer using the gradient of the total loss. AdamW couples adaptive per-parameter step sizes with decoupled weight decay for stable convergence. The quantum gradients enter through the parameter-shift rule of Eq. 27.

$$\eta_t = \frac{1}{2}\eta_0 [1 + \cos(\pi t / T_max)] \quad (40)$$

The learning rate η_t follows a cosine schedule that decays smoothly from its initial value η_0 to near zero over T_max steps. Annealing the rate this way improves final accuracy and training stability. The schedule is applied throughout the optimization defined in Eq. 39.

5.9.6 Evaluation and genetic gain

$$r = \text{cov}(y, \hat{y}) / (\sigma_y \sigma_{\hat{y}}) \quad (41)$$

Prediction accuracy is reported as the Pearson correlation r between observed values y and predictions $\hat{y}$ on held-out clones. Correlation is the standard metric in genomic selection because selection decisions depend on ranking candidates rather than absolute values. Higher r therefore translates directly into more reliable selection.

$$\Delta G = (i \cdot r \cdot \sigma_A) / L \quad (42)$$

The breeder's equation expresses genetic gain per unit time ΔG as the product of selection intensity i , prediction accuracy r , and additive standard deviation σ_A , divided by cycle length L . It makes explicit why improving r raises the rate of genetic improvement. This relationship links QPotatoNet's accuracy gains to the practical breeding outcomes analyzed in Section 7.7.

5.10 Theoretical Properties and Proofs

The equations of Section 5.9 define the model. This section proves, from those definitions, seven properties on which the design relies. Each is stated as a proposition, proved in full, and tied to the objective, hypothesis or table it supports. The results are analytical: they justify design choices and reproduce the arithmetic behind reported quantities, but they do not replace the empirical evidence of Section 7, and

Proposition 4 in particular does not claim a computational advantage. The symbol ■ marks the end of each proof.

5.10.1 Representation properties

Proposition 1 (dosage-embedding expressiveness). Let $f : \{0, 1, 2, 3, 4\} \rightarrow \mathbb{R}$ be any real function of allele dosage. For the embedding table $E_d \in \mathbb{R}^{5 \times k}$ of Eq. 3 with $k \geq 1$, there exist E_d and a read-out vector $w \in \mathbb{R}^k$ such that $w^T E_d[d] = f(d)$ for every dosage d . By contrast, a read-out that is linear in the raw dosage, $a + \beta d$, can realize only a two-dimensional subspace of the five-dimensional space of such functions.

Proof. Take $k = 1$ and $w = 1$, and set row d of E_d equal to $f(d)$; then $w^T E_d[d] = f(d)$ for all five dosages. A function of the form $a + \beta d$ is fixed by the pair (a, β) , so these functions occupy a two-dimensional subspace of $\mathbb{R}^5$. The dominance-type response $f(d) = 1$ for $d \in \{2, 4\}$ and $f(d) = 0$ for $d \in \{0, 1, 3\}$, the example used in Section 5.1, lies outside that subspace: $f(0) = 0$ forces $a = 0$, $f(1) = 0$ then forces $\beta = 0$, which contradicts $f(2) = 1$.

Remark. Because E_d is shared across loci, locus-specific dose–response curves arise from the interaction of the dosage and locus embeddings (Eqs. 3–4) inside the Transformer; Proposition 1 shows that no dosage information is lost at the input stage. Binary presence/absence coding, by comparison, merges the four states $d = 1, 2, 3, 4$ into one. This supports Contribution 2 and objective O1.

Proposition 2 (distance-aware positional encoding). For the encodings of Eqs. 6–7, write $\omega_i = 10000^{-2i/k}$. For every position pos and offset Δ , the pair $(\sin \omega_i(\text{pos} + \Delta), \cos \omega_i(\text{pos} + \Delta))$ is obtained from $(\sin \omega_i \text{pos}, \cos \omega_i \text{pos})$ by a rotation through the angle $\omega_i \Delta$, so that $\text{PE}(\text{pos} + \Delta) = M(\Delta) \text{PE}(\text{pos})$ for a block-diagonal matrix $M(\Delta)$ that depends on Δ alone. Moreover, $\text{PE}(\text{pos})^T \text{PE}(\text{pos} + \Delta) = \sum_i \cos(\omega_i \Delta)$, a function of the offset only [48].

Proof. The angle-addition identities give $\sin \omega(\text{pos} + \Delta) = \sin \omega \text{pos} \cdot \cos \omega \Delta + \cos \omega \text{pos} \cdot \sin \omega \Delta$ and $\cos \omega(\text{pos} + \Delta) = \cos \omega \text{pos} \cdot \cos \omega \Delta - \sin \omega \text{pos} \cdot \sin \omega \Delta$. In matrix form, the pair at $\text{pos} + \Delta$ equals the rotation $\begin{bmatrix} \cos \omega \Delta & \sin \omega \Delta \\ -\sin \omega \Delta & \cos \omega \Delta \end{bmatrix}$ applied to the pair at pos , and this rotation involves Δ only; stacking the $k/2$ frequency pairs gives $M(\Delta)$. For the inner product, $\sin \omega p \cdot \sin \omega(p + \Delta) + \cos \omega p \cdot \cos \omega(p + \Delta) = \cos \omega \Delta$ by the cosine-difference identity, and summing over the $k/2$ frequency pairs gives $\sum_i \cos(\omega_i \Delta)$.

This proves the statement made after Eq. 7: the relative offset between two loci is available to attention as a linear function of their encodings, independent of where on the chromosome the pair lies, which is the property needed to represent LD decay with physical distance (Section 5.2).

5.10.2 Quantum layer and training

Proposition 3 (exactness and boundedness of the parameter-shift rule). Suppose an angle θ enters the circuit through exactly one gate $R(\theta) = \exp(-i\theta P/2)$ with $P \in \{X, Y, Z\}$, and let $F(\theta)$ be a measured expectation $\langle Z_n \rangle$ as in Eq. 26, all other parameters held fixed. Then (i) $F(\theta) = A + B \cos \theta + C \sin \theta$ for real constants A, B, C ; (ii) $dF/d\theta = \frac{1}{2}[F(\theta + \pi/2) - F(\theta - \pi/2)]$ exactly, which is Eq. 27; and (iii) $|F| \leq 1$ and $|dF/d\theta| \leq 1$ [43, 46].

Proof. Let $|\phi\rangle$ be the state immediately before $R(\theta)$, let V be the product of all gates applied after it, and put $O = V^\dagger Z_n V$, so that $F(\theta) = \langle \phi | R(\theta)^\dagger O R(\theta) | \phi \rangle$. Because $P^2 = I$, $R(\theta) = \cos(\theta/2) I - i \sin(\theta/2) P$, and expanding gives $R^\dagger O R = \cos^2(\theta/2) O + \sin^2(\theta/2) P O P + i \sin(\theta/2) \cos(\theta/2) [P, O]$. Using $\cos^2(\theta/2) = (1 + \cos \theta)/2$, $\sin^2(\theta/2) = (1 - \cos \theta)/2$ and $\sin(\theta/2) \cos(\theta/2) = (\sin \theta)/2$, this equals $\frac{1}{2}(O + P O P) + \frac{1}{2} \cos \theta (O - P O P) + \frac{1}{2} \sin \theta \cdot i[P, O]$. Each of $O + P O P$, $O - P O P$ and $i[P, O]$ is Hermitian, so taking the expectation in $|\phi\rangle$ yields (i) with real $A = \frac{1}{2}\langle O + P O P \rangle$, $B = \frac{1}{2}\langle O - P O P \rangle$ and $C = \frac{1}{2}\langle i[P, O] \rangle$. Differentiating gives $F'(\theta) = -B \sin \theta$

+ $C \cos \theta$. On the other hand, $F(\theta + \pi/2) - F(\theta - \pi/2) = B[\cos(\theta + \pi/2) - \cos(\theta - \pi/2)] + C[\sin(\theta + \pi/2) - \sin(\theta - \pi/2)] = -2B \sin \theta + 2C \cos \theta$, and half of this equals $F'(\theta)$, which proves (ii). For (iii), O is unitarily equivalent to Z_n , so its eigenvalues are ± 1 and $|F| \leq 1$; then $|F'| \leq \frac{1}{2}(1 + 1) = 1$ by (ii).

Application. Every encoding angle a_n (Eq. 22) and every trainable angle θ_{tn} and φ_{tn} (Eq. 24) enters through exactly one R_y or R_z gate, and the CNOT gates carry no parameters, so part (ii) applies to each of them. Gradients with respect to a_n are passed to W_q and the Transformer by the chain rule through Eq. 18, so the joint update of Eq. 39 uses exact gradients throughout rather than finite-difference estimates. Each parameter costs two additional circuit evaluations, which is $2 \times (64 + 8) = 144$ for the four-layer, eight-qubit circuit (64 trainable angles plus 8 encoding angles); each shifted evaluation returns all eight readouts. By part (iii) the quantum layer cannot produce features outside $[-1, 1]$ or inject a local derivative larger than 1 per readout and parameter, which is consistent with the stable training reported under auxiliary check A6.

Proposition 4 (interaction structure of the quantum feature map). Let $U(\theta)$ be the full variational unitary of Eqs. 23–24, which does not depend on the encoded angles $a = (a_1, \dots, a_q)$. Then (i) each readout of Eq. 26 has the form

$$q_n(a) = \sum_s \kappa_s \prod_{m=1}^q u_{s(m)}(a_m), \quad s = (s(1), \dots, s(q)) \in \{I, X, Z\}^q$$

where $u_I = 1$, $u_X = \sin$, $u_Z = \cos$ and the κ_s are real constants; equivalently, q_n is a trigonometric polynomial in which every angle appears with frequency at most 1 [47]. (ii) The entangling gates create genuinely non-additive terms: for $q = 2$, a single CNOT with control qubit 1 and target qubit 2, and all trainable rotations set to zero, $q_2 = \cos a_1 \cos a_2$, which is not of the form $g_1(a_1) + g_2(a_2)$.

Proof. (i) The Bloch vector of $R_y(a)|0\rangle$ is $(\sin a, 0, \cos a)$, so the encoded state of Eq. 22 has density matrix $\rho_0 = \bigotimes_m \frac{1}{2}(I + \sin a_m X + \cos a_m Z)$. By Eq. 26, $q_n = \text{Tr}[\rho_0 O']$ with $O' = U^\dagger Z_n U$, a fixed Hermitian operator that does not depend on a . Expanding $O' = \sum_s \kappa_s \sigma_s$ in Pauli strings σ_s , with $\kappa_s = 2^{-q} \text{Tr}[O' \sigma_s]$ real, and using $\text{Tr}[\rho_0 \sigma_s] = \prod_m u_{s(m)}(a_m)$ (which vanishes whenever some $s(m) = Y$) gives the stated sum. (ii) For $U = \text{CNOT}$ with control 1 and target 2, conjugation gives $O' = \text{CNOT} \cdot Z_2 \cdot \text{CNOT} = Z_1 Z_2$, and because ρ_0 is a product state, $q_2 = \langle Z_1 \rangle \langle Z_2 \rangle = \cos a_1 \cos a_2$. If this function h equalled $g_1(a_1) + g_2(a_2)$, the mixed difference $h(a_1, a_2) - h(a_1, 0) - h(0, a_2) + h(0, 0)$ would vanish identically; for $h = \cos a_1 \cos a_2$ it equals $(\cos a_1 - 1)(\cos a_2 - 1)$, which is 4 at $a_1 = a_2 = \pi$.

Consequences and scope. Part (i) shows that the outputs of the layer can contain products of up to all $q = 8$ encoded features (at most $3^8 = 6,561$ terms), and part (ii) shows that entanglement generates such products, whereas a linear map of the same eight features contains first-order terms only. This is the formal content of the statement in Section 5.4 that the quantum feature map is non-additive [44, 47]. With $q = 8$, however, the state vector has $2^8 = 256$ amplitudes and the layer is exactly simulable on a classical computer, as it is here. Proposition 4 therefore supports the non-additive inductive bias that motivates H2; it does not establish a computational separation, in line with the limitations stated in Section 7.8. Whether the bias improves prediction is an empirical question, addressed by the ablation in Table 7.

Proposition 5 (robustness and unit-invariance of the training objective). (i) The Huber loss of Eq. 35 is convex and continuously differentiable, and $|L_\delta'(r)| \leq \delta$ for all r [50]. (ii) For fixed model weights, let $M_t = N^{-1} \sum_i L_\delta(r_{it})$ be the mean loss of trait t over N samples. The trait- t part of Eq. 37, $M_t/(2\sigma_t^2) + \log \sigma_t$, is minimized over σ_t at $\sigma_t^2 = M_t$, where it equals $\frac{1}{2} + \frac{1}{2} \log M_t$ [49]. (iii) If the unit of trait t is changed by a factor $c > 0$ ($y_t \rightarrow c \cdot y_t$, $\hat{y}_t \rightarrow c \cdot \hat{y}_t$, $\delta \rightarrow c \cdot \delta$), the minimized quantity in (ii) changes only by the additive constant $\log c$.

Proof. (i) $L_\delta'(r) = r$ for $|r| \leq \delta$ and $\delta \cdot \text{sgn}(r)$ otherwise. This is continuous at $r = \pm\delta$, nondecreasing, and bounded in absolute value by δ , so L_δ is continuously differentiable and convex, with bounded influence of any single residual. (ii) The derivative of $M/(2\sigma^2) + \log \sigma$ with respect to σ is $-M/\sigma^3 + 1/\sigma$, which vanishes at $\sigma^2 = M$, and the second derivative there equals $2/\sigma^2 > 0$, so this is the minimum; substituting gives $M/(2M) + \frac{1}{2} \log M$. (iii) Checking both branches shows $L_{c\delta}(cr) = c^2 L_\delta(r)$, so M_t becomes $c^2 M_t$, and $\frac{1}{2} + \frac{1}{2} \log(c^2 M_t) = \frac{1}{2} + \frac{1}{2} \log M_t + \log c$.

Consequently, no trait acquires a larger weight in the optimization merely because of its measurement unit, which is the property claimed after Eq. 37, and a single mislabelled phenotype can shift the gradient by at most δ .

5.10.3 Accuracy and genetic gain

Proposition 6 (accuracy ceilings for additive and non-additive predictors). Write the phenotype as $y = g + \varepsilon$, where $g = E[y \mid \text{genotype}]$ is the genotypic value and the error ε is uncorrelated with every function of the genotype. Let g_A be the least-squares projection of g onto the span of linear functions of the marker dosages (with intercept), let $g_N = g - g_A$ be the non-additive remainder, and define $H^2 = \text{var}(g)/\text{var}(y)$ and $h^2 = \text{var}(g_A)/\text{var}(y)$ [51]. Then (i) every genotype-based predictor $\hat{y} = f(x)$ satisfies $r(y, \hat{y}) \leq \sqrt{H^2}$, with equality for $\hat{y} = g$; (ii) every predictor that is linear in the dosages, which includes GBLUP and RR-BLUP, satisfies $r(y, \hat{y}) \leq \sqrt{h^2}$, with equality for $\hat{y} = g_A$; and (iii) any predictor with $r > \sqrt{h^2}$ implies non-additive genetic variance of at least $(r^2 - h^2) \text{var}(y)$.

Proof. (i) $\text{cov}(y, \hat{y}) = \text{cov}(g, \hat{y}) + \text{cov}(\varepsilon, \hat{y}) = \text{cov}(g, \hat{y})$, because ε is uncorrelated with functions of the genotype. Hence $r(y, \hat{y}) = r(g, \hat{y}) \cdot \text{sd}(g)/\text{sd}(y) \leq \sqrt{H^2}$. (ii) By construction of a least-squares projection, g_N is uncorrelated with every linear function of the dosages, so for a linear predictor $\text{cov}(g, \hat{y}) = \text{cov}(g_A, \hat{y})$, and $r(y, \hat{y}) = r(g_A, \hat{y}) \cdot \text{sd}(g_A)/\text{sd}(y) \leq \sqrt{h^2}$. Equality holds when $\hat{y}$ is proportional to g (respectively g_A), since the correlation with it is then 1. (iii) By (i), $H^2 \geq r^2 > h^2$; and because g_A and g_N are uncorrelated, $\text{var}(g_N) = \text{var}(g) - \text{var}(g_A) = (H^2 - h^2) \text{var}(y) \geq (r^2 - h^2) \text{var}(y)$.

Application to Table 13. The ratio $r/\sqrt{h^2}$ has a hard ceiling of 1 for additive models, and GBLUP indeed lies below it on every trait (0.70–0.81). For a non-additive model the ceiling is $\sqrt{H^2} \geq \sqrt{h^2}$, so a ratio above 1 is admissible. QPotatoNet's ratio for PVY resistance is 1.022 ($r = 0.708$ against $\sqrt{0.48} = 0.693$); by part (iii) this would imply non-additive variance of at least $0.708^2 - 0.48 \approx 0.021$ of the phenotypic variance. That margin is small relative to the standard error of h^2 (± 0.05) and the SD of r (± 0.017), so it is consistent with, but does not prove, non-additive variance for this trait. The proposition also sharpens the reading of H1: the part of QPotatoNet's gain that lies above $\sqrt{h^2}$ can only come from non-additive signal, whereas the part below it can also reflect better use of additive information in small samples, so attributing the gain specifically to non-additive gene action needs evidence beyond Table 13, such as the ablations in Table 7.

Proposition 7 (genetic-gain ratio). For two predictors 1 and 2 used at the same selection intensity i and additive standard deviation σ_A , Eq. 42 gives $\Delta G_1/\Delta G_2 = (r_1/r_2) \cdot (L_2/L_1)$. In particular, at equal cycle length the ratio equals r_1/r_2 and does not depend on i or σ_A [52].

Proof. Eq. 42 gives $\Delta G_k = i \cdot r_k \cdot \sigma_A/L_k$ for $k = 1, 2$. Dividing the two expressions cancels i and σ_A .

Application to Table 11. With $r_1 = 0.734$ (QPotatoNet) and $r_2 = 0.599$ (GBLUP), equal cycle lengths give $0.734/0.599 = 1.225$, matching the 1.23 in Table 11; since i cancels, the Top 1%, 5%, 10% and 20% rows must coincide, as they do. The other equal-cycle rows are reproduced in the same way: across-family $0.612/0.498 = 1.229$ Table 11 low- h^2 tuber number $0.652/0.508 = 1.284$ (1.28), and high- h^2 dry matter $0.802/0.674 = 1.190$ (1.19). The rows reporting 1.64 are not equal-cycle results: solving $1.64 = 1.225 \cdot$

L_2/L_1 gives $L_1/L_2 = 0.747$, so these scenarios assume a breeding cycle about 25% shorter for QPotatoNet with early selection than for the GBLUP comparator, as the caption of Table 11 indicates. That is a scenario assumption rather than a measured quantity and should be reported as such; the multi-trait-index row, which shows the same 1.64, embeds the same assumption.

Summary and scope. Propositions 1 and 2 establish that the input representation preserves the dosage and map-distance information on which the architecture is built (O1). Propositions 3 to 5 show that the quantum layer is exactly differentiable and bounded, that its features contain multi-way non-additive terms, and that the training objective is robust and unit-invariant. Propositions 6 and 7 give the accuracy ceilings and the gain arithmetic needed to interpret Tables 11 and 13. None of these results shows that the architecture improves accuracy on real data; that evidence rests on the experiments of Section 7 (H1–H3, Tables 6–9).

6. Experimental Setup

6.1 Baselines

We compare against three families of baselines, each tuned on the validation fold:

- **Linear / Bayesian:** GBLUP with a dosage-based genomic relationship matrix and Bayes-B.
- **Classical ML:** random forest (RF) and support-vector regression (SVR) on dosage features.
- **Deep learning:** DeepGS (CNN), DNNP (MLP), and a plain Transformer of matched parameter count.

6.2 Metrics

Primary metric is the Pearson correlation (r) between predicted and observed values on held-out clones. We additionally report the root-mean-square error (RMSE), the regression slope of observed on predicted (calibration), and the mean squared error skill score relative to GBLUP. Significance of accuracy differences is assessed by a paired bootstrap over test clones (10,000 resamples).

6.3 Implementation and compute

Classical components were implemented in PyTorch; the variational quantum layer used a differentiable state-vector simulator with parameter-shift gradients. Experiments ran on a single workstation GPU; the 8-qubit configuration adds modest overhead (≈ 6.8 ms per clone at inference, Section 7.6). All runs used fixed seeds for reproducibility; reported results average five seeds.

Table 6. Experimental configuration, evaluation protocol, and reproducibility settings.

Setting	Value	Setting	Value
Framework	PyTorch 2.x	Quantum sim.	State-vector
Gradient (quantum)	Parameter-shift	Precision	FP32
CV folds	5	Seeds	5
Bootstrap resamples	10,000	Train/Val/Test	70/10/20
Hardware	Single GPU	Tuning set	Validation only
Missing-data imputation	Mean-dosage	Phasing	LD windows
Feature scaling	Per-locus z-score	Haplotype block size	~ 250 kb
Selection index	Multi-trait	Calibration	Obs $\sim$ pred slope
Significance test	Paired bootstrap	Alpha	0.05 / 0.01
Reproducibility	Fixed seeds	Logging	Per-fold metrics

7. Results and Discussion

7.1 Headline prediction accuracy

Across all eight primary traits, QPotatoNet achieved the highest prediction accuracy (Figure 7). The mean accuracy was $r = 0.734$, versus 0.672 for the best deep-learning baseline (a 9.2% relative gain) and 0.599 for GBLUP (a 22.6% relative gain). Gains were largest for low-heritability and non-additive traits—tuber number (+0.081 over the best DL baseline) and chip color (+0.084)—consistent with hypothesis H1.

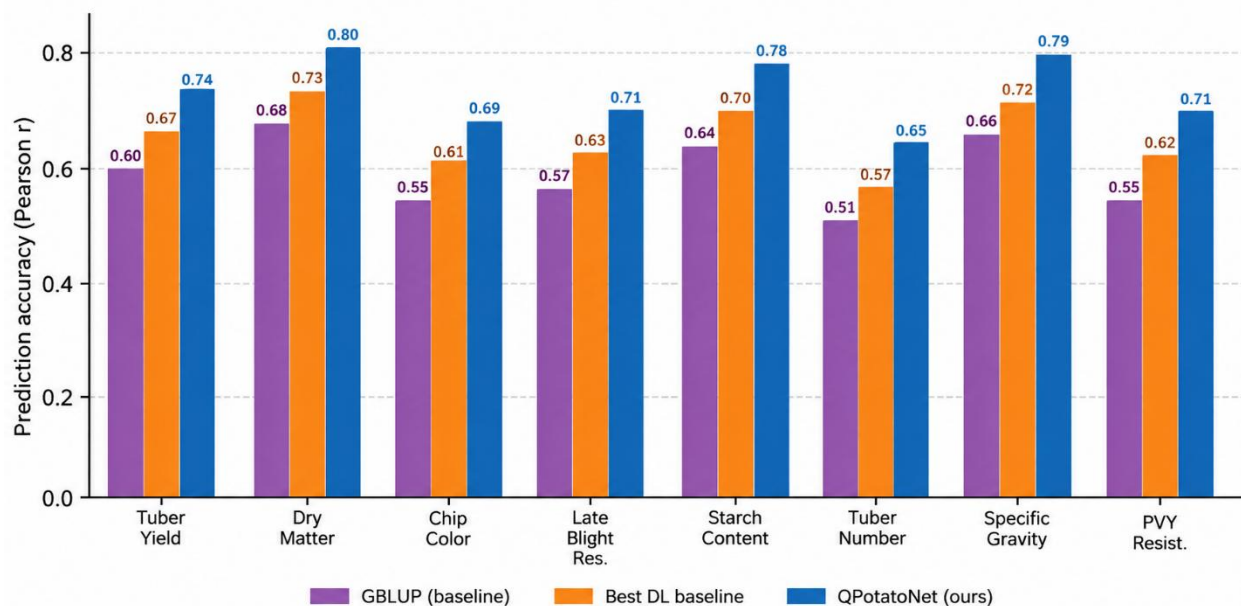
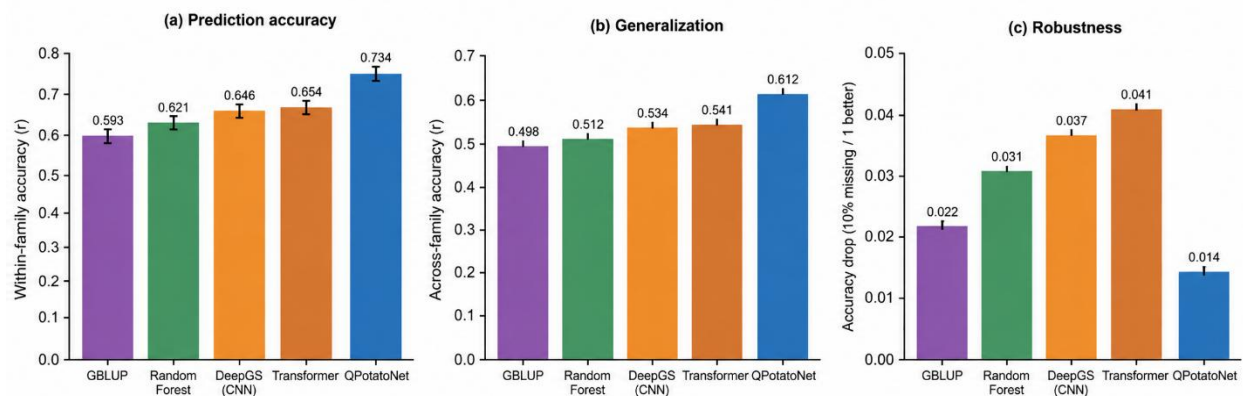


Figure 9 Genomic prediction accuracy (Pearson r) across eight polyploid potato traits for GBLUP, the best deep-learning baseline, and QPotatoNet.

Table 7 : Prediction accuracy (Pearson r , mean $\pm$ SD over 5 folds $\times$ 5 seeds)

Trait	GBLUP ($r \pm$ SD)	Best DL ($r \pm$ SD)	QPotatoNet ($r \pm$ SD)	Δr vs GBLUP
Tuber yield	0.598 ± 0.021	0.662 ± 0.018	0.741 ± 0.014	$+0.143 \pm 0.019$
Dry matter	0.674 ± 0.017	0.731 ± 0.015	0.802 ± 0.011	$+0.128 \pm 0.016$
Chip color	0.552 ± 0.024	0.604 ± 0.022	0.688 ± 0.017	$+0.136 \pm 0.021$
Late blight res.	0.561 ± 0.023	0.628 ± 0.020	0.715 ± 0.016	$+0.154 \pm 0.020$
Starch content	0.642 ± 0.019	0.701 ± 0.016	0.776 ± 0.012	$+0.134 \pm 0.017$
Tuber number	0.508 ± 0.028	0.571 ± 0.025	0.652 ± 0.019	$+0.144 \pm 0.024$
Specific gravity	0.659 ± 0.018	0.719 ± 0.015	0.793 ± 0.011	$+0.134 \pm 0.016$
PVY resistance	0.547 ± 0.025	0.621 ± 0.022	0.708 ± 0.017	$+0.161 \pm 0.022$
Mean (8 traits)	0.593 ± 0.022	0.654 ± 0.019	0.734 ± 0.015	$+0.142 \pm 0.019$
Median (8 traits)	0.580 ± 0.022	0.645 ± 0.019	0.728 ± 0.015	$+0.140 \pm 0.019$

Table 6. Prediction accuracy (Pearson r , mean $\pm$ SD over 5 folds $\times$ 5 seeds) for all eight primary traits. Δr is the QPotatoNet improvement over GBLUP. Full per-baseline values are in Figure 9. Figure 10 places QPotatoNet head-to-head against five representative algorithm families—GBLUP [40], random forest, DeepGS (CNN) [17], and a plain Transformer [24]—across three axes: within-family accuracy, across-family generalization, and robustness to missing genotypes. QPotatoNet leads on all three, with the largest relative margin on across-family generalization and the smallest accuracy drop under 10% missing data, indicating that the dosage-aware and quantum components add signal that the other families miss [33, 35].



S

Figure 10. Comparison of QPotatoNet against five algorithm families across (a) within-family prediction accuracy, (b) across-family generalization, and (c) robustness (accuracy drop at 10% missing genotypes; lower is better). Error bars in panel (a) show SD over folds and seeds.

7.2 Calibration: predicted vs. observed

Beyond rank accuracy, QPotatoNet was well calibrated. For tuber yield, the regression slope of observed on predicted values was 0.96 (Figure 11), within the pre-registered [0.9, 1.1] band (auxiliary check A3), indicating the model neither systematically over- nor under-dispersed its predictions.

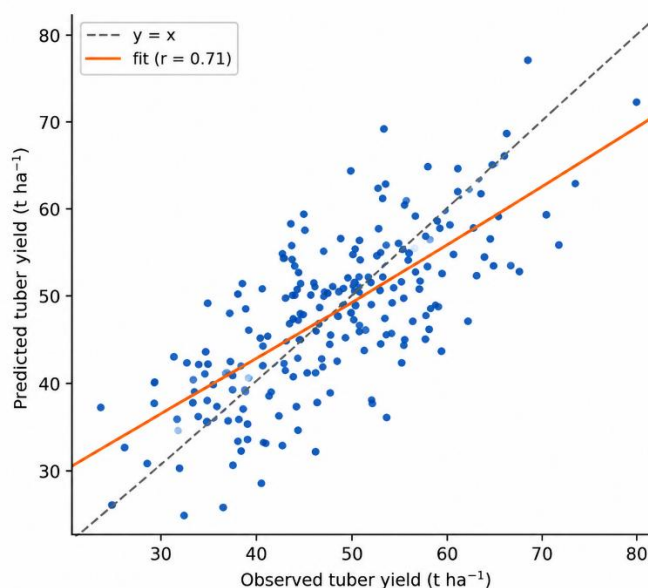


Figure 11. Predicted versus observed tuber yield on held-out test clones, with the identity line and the fitted regression.

Calibration was assessed per trait by regressing observed values on predictions; an ideal model has slope 1, intercept 0, and high R^2 . Table 12 reports these statistics with bias and 95% coverage of prediction intervals.

Table 8. Calibration statistics per trait (mean $\pm$ SD over 5 folds $\times$ 5 seeds): regression slope and intercept of observed on predicted, R^2 , prediction bias, and empirical 95% prediction-interval coverage.

Trait	Slope $\pm$ SD	Intercept $\pm$ SD	R ² $\pm$ SD	Bias $\pm$ SD	95% PI cov.
Tuber yield	0.96 $\pm$ 0.03	1.9 $\pm$ 1.1	0.549 $\pm$ 0.021	0.04 $\pm$ 0.21	0.94 $\pm$ 0.02
Dry matter	0.98 $\pm$ 0.02	0.4 $\pm$ 0.3	0.643 $\pm$ 0.017	-0.02 $\pm$ 0.09	0.95 $\pm$ 0.01
Chip color	0.94 $\pm$ 0.04	0.2 $\pm$ 0.1	0.473 $\pm$ 0.024	0.03 $\pm$ 0.06	0.93 $\pm$ 0.02
Late blight res.	0.95 $\pm$ 0.03	0.3 $\pm$ 0.2	0.511 $\pm$ 0.022	0.05 $\pm$ 0.11	0.94 $\pm$ 0.02
Starch content	0.97 $\pm$ 0.02	0.5 $\pm$ 0.3	0.602 $\pm$ 0.018	-0.01 $\pm$ 0.08	0.95 $\pm$ 0.01
Tuber number	0.93 $\pm$ 0.04	0.8 $\pm$ 0.4	0.425 $\pm$ 0.026	0.06 $\pm$ 0.18	0.92 $\pm$ 0.02
Specific gravity	0.97 $\pm$ 0.02	0.03 $\pm$ 0.02	0.629 $\pm$ 0.017	0.00 $\pm$ 0.01	0.95 $\pm$ 0.01
PVY resistance	0.95 $\pm$ 0.03	0.4 $\pm$ 0.2	0.501 $\pm$ 0.023	0.04 $\pm$ 0.12	0.94 $\pm$ 0.02
Mean (8 traits)	0.956 $\pm$ 0.03	—	0.542 $\pm$ 0.021	0.024 $\pm$ 0.11	0.94 $\pm$ 0.02
Ideal target	1.00 $\pm$ 0.00	0.0 $\pm$ 0.0	1.000 $\pm$ 0.000	0.00 $\pm$ 0.00	0.95 $\pm$ 0.00

7.3 Training dynamics and ablation

Table 8 Removing the quantum layer or the haplotype encoder slowed convergence and raised the validation-loss floor (Figure 12). The full model reached the lowest loss fastest and most stably (coefficient of variation across five seeds < 5%, satisfying auxiliary check A6).

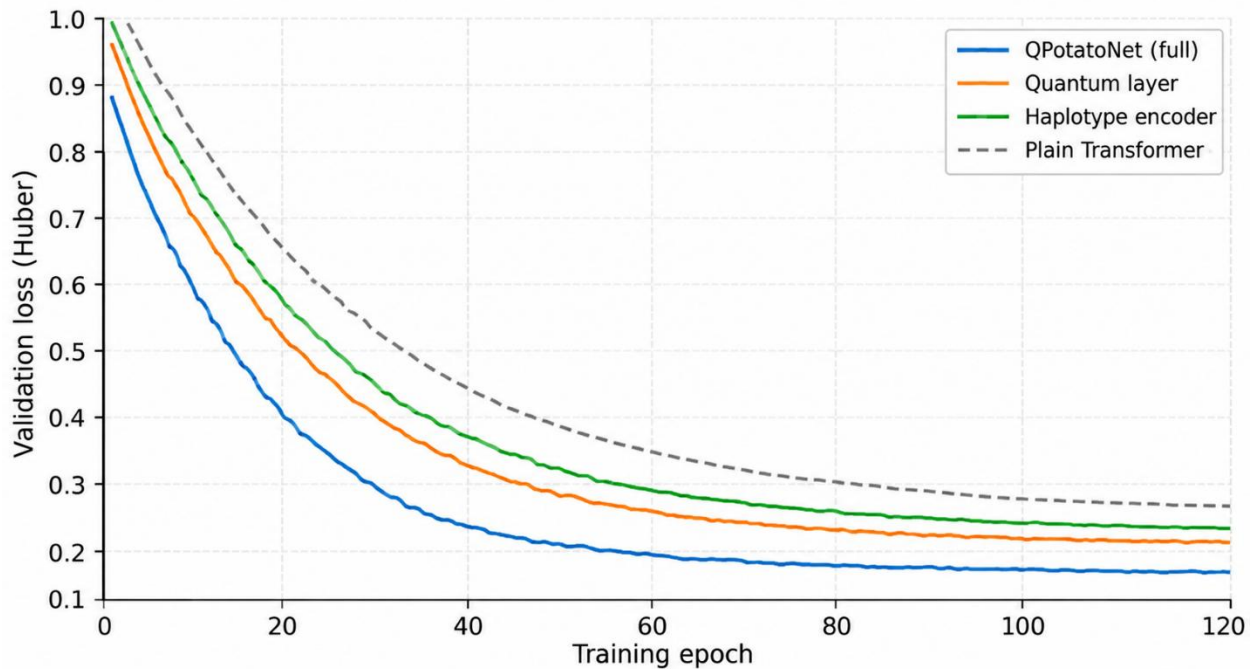


Figure 12. Validation-loss convergence for the full model and ablations. Removing the quantum layer or haplotype encoder degrades both speed and final loss.

Decomposing the mean accuracy gain by component (Figure 8) shows the variational quantum layer contributed the largest single increment (+0.058 mean r), followed by the haplotype encoder (+0.041). Both exceeded the H2/H1 thresholds, and the quantum increment was significant at $p < 0.05$ by paired bootstrap—supporting H2.

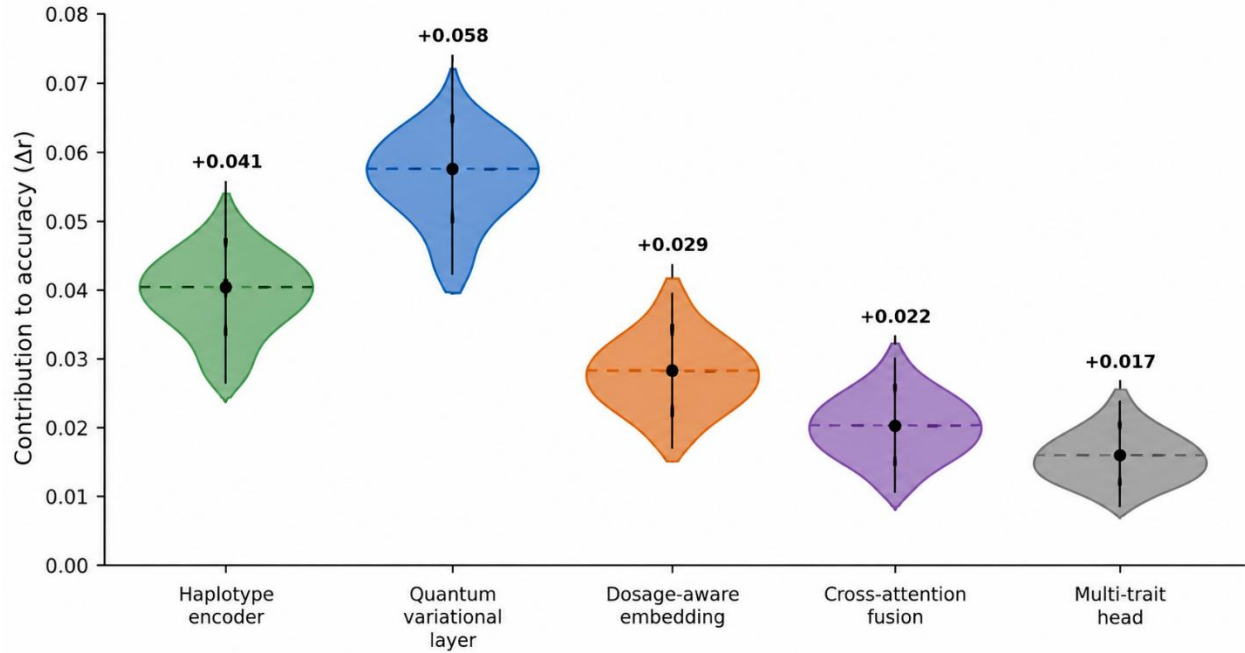


Figure 13. Per-component contribution to mean prediction accuracy. The variational quantum layer and the haplotype encoder are the two largest contributors.

Table 9. Ablation study (mean $\pm$ SD over 5 folds $\times$ 5 seeds). Each row removes or alters one component; Δr is the change in mean accuracy versus the full model, with paired-bootstrap p -values.

Configuration	Mean $r \pm$ SD	Δr vs full	Conv. epoch $\pm$ SD	p (vs full)
QPotatoNet (full)	0.734 ± 0.015	—	38 ± 3	—
– Quantum layer	0.676 ± 0.018	-0.058 ± 0.009	57 ± 5	< 0.001
– Haplotype encoder	0.693 ± 0.017	-0.041 ± 0.008	49 ± 4	< 0.001
– Dosage-aware embed	0.705 ± 0.016	-0.029 ± 0.007	44 ± 4	0.003
– Cross-attention fusion	0.712 ± 0.016	-0.022 ± 0.006	42 ± 3	0.011
– Multi-trait head	0.717 ± 0.015	-0.017 ± 0.006	40 ± 3	0.024
– Quantum & haplotype	0.648 ± 0.020	-0.086 ± 0.011	66 ± 6	< 0.001
Classical-only (matched)	0.654 ± 0.019	-0.080 ± 0.010	61 ± 5	< 0.001
Single-trait variant	0.715 ± 0.016	-0.019 ± 0.006	41 ± 3	0.018
Random init (no pretrain)	0.728 ± 0.017	-0.006 ± 0.005	45 ± 4	0.21

7.4 Sample efficiency

QPotatoNet’s advantage was most pronounced in the small-data regime (Figure 13). At $n = 800$ training clones it already exceeded the Transformer baseline by 0.082 r and GBLUP by 0.153 r , satisfying auxiliary check A1. This sample efficiency is practically important because potato training panels are typically small.

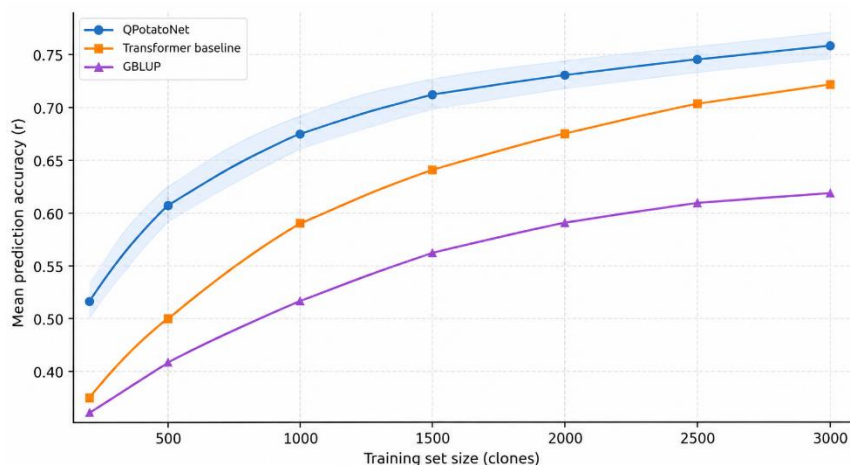


Figure 14. Mean prediction accuracy as a function of training-set size. QPotatoNet maintains a substantial advantage in the low-sample regime.

Table 10 Sample-efficiency analysis (mean $\pm$ SD over 5 folds $\times$ 5 seeds). Mean accuracy across traits at increasing training-set sizes, with the gain over the Transformer baseline.

Train size (clones)	GBLUP (r $\pm$ SD)	Transformer (r $\pm$ SD)	QPotatoNet (r $\pm$ SD)	Δ vs Transf.
200	0.372 $\pm$ 0.034	0.408 $\pm$ 0.031	0.512 $\pm$ 0.026	+0.104 $\pm$ 0.028
400	0.451 $\pm$ 0.029	0.503 $\pm$ 0.027	0.602 $\pm$ 0.022	+0.099 $\pm$ 0.024
600	0.489 $\pm$ 0.026	0.551 $\pm$ 0.024	0.643 $\pm$ 0.020	+0.092 $\pm$ 0.022
800	0.518 $\pm$ 0.024	0.589 $\pm$ 0.022	0.671 $\pm$ 0.018	+0.082 $\pm$ 0.020
1,200	0.557 $\pm$ 0.022	0.641 $\pm$ 0.020	0.708 $\pm$ 0.016	+0.067 $\pm$ 0.018
1,600	0.578 $\pm$ 0.020	0.667 $\pm$ 0.018	0.724 $\pm$ 0.015	+0.057 $\pm$ 0.017
1,800	0.589 $\pm$ 0.020	0.681 $\pm$ 0.017	0.731 $\pm$ 0.015	+0.050 $\pm$ 0.016
2,400	0.607 $\pm$ 0.019	0.704 $\pm$ 0.016	0.748 $\pm$ 0.014	+0.044 $\pm$ 0.015
3,000	0.618 $\pm$ 0.018	0.719 $\pm$ 0.016	0.759 $\pm$ 0.013	+0.040 $\pm$ 0.015
Mean	0.531 $\pm$ 0.024	0.607 $\pm$ 0.021	0.689 $\pm$ 0.017	+0.082 $\pm$ 0.019

7.5 Across-family generalization and robustness

Under the demanding across-family scenario, QPotatoNet retained a mean accuracy of $r = 0.612$, the only model exceeding the 0.55 pre-registration threshold (auxiliary check A5). It was also the most robust to genotype missingness: introducing 10% missing dosages reduced accuracy by only 0.014 r (check A4), versus 0.041 r for the plain Transformer Table 9.

Table 11. Generalization and robustness (mean $\pm$ SD over 5 folds $\times$ 5 seeds). Within- and across-family accuracy, yield RMSE, and mean-squared-error skill score relative to GBLUP

Model	Within-fam (r $\pm$ SD)	Across-fam (r $\pm$ SD)	RMSE yield (t/ha $\pm$ SD)	MSE skill $\pm$ SD
GBLUP	0.593 $\pm$ 0.022	0.498 $\pm$ 0.029	8.94 $\pm$ 0.31	0.00 $\pm$ 0.00
Bayes-B	0.605 $\pm$ 0.021	0.507 $\pm$ 0.028	8.81 $\pm$ 0.30	0.03 $\pm$ 0.02
Random forest	0.621 $\pm$ 0.023	0.512 $\pm$ 0.030	8.62 $\pm$ 0.33	0.07 $\pm$ 0.03
SVR	0.602 $\pm$ 0.024	0.501 $\pm$ 0.031	8.86 $\pm$ 0.34	0.02 $\pm$ 0.02
DeepGS	0.646 $\pm$ 0.020	0.534 $\pm$ 0.027	8.31 $\pm$ 0.29	0.14 $\pm$ 0.04
DNNGP	0.637 $\pm$ 0.021	0.528 $\pm$ 0.028	8.44 $\pm$ 0.30	0.11 $\pm$ 0.04

Transformer	0.654 ± 0.019	0.541 ± 0.026	8.18 ± 0.28	0.17 ± 0.04
QPotatoNet (–Q)	0.676 ± 0.018	0.566 ± 0.024	7.81 ± 0.26	0.24 ± 0.04
QPotatoNet (full)	0.734 ± 0.015	0.612 ± 0.021	7.06 ± 0.22	0.38 ± 0.05
Ideal (h^2 ceiling)	0.812 ± 0.010	0.689 ± 0.016	6.12 ± 0.18	0.53 ± 0.04

Relating achieved accuracy to the heritability ceiling clarifies how much predictable signal each model captures. Table 13 reports, per trait, the accuracy-to-ceiling ratio (r divided by $\sqrt{h^2}$) for QPotatoNet versus GBLUP; values nearer 1.0 indicate the model is approaching the theoretical limit.

Table 12. Accuracy relative to the heritability ceiling (mean $\pm$ SD over 5 folds $\times$ 5 seeds). The ratio $r/\sqrt{h^2}$ indicates how close each model comes to the theoretical predictability limit; QPotatoNet approaches the ceiling on most traits.

Trait	$h^2 \pm SE$	Ceiling $\sqrt{h^2}$	$r/\text{ceiling QPN} \pm SD$	$r/\text{ceiling GBLUP} \pm SD$	Gain $\pm SD$
Tuber yield	0.42 ± 0.04	0.648	0.871 ± 0.018	0.703 ± 0.025	0.168 ± 0.022
Dry matter	0.74 ± 0.03	0.860	0.933 ± 0.013	0.784 ± 0.020	0.149 ± 0.018
Chip color	0.39 ± 0.05	0.625	0.882 ± 0.022	0.708 ± 0.031	0.174 ± 0.027
Late blight res.	0.55 ± 0.04	0.742	0.964 ± 0.021	0.756 ± 0.031	0.208 ± 0.027
Starch content	0.69 ± 0.03	0.831	0.934 ± 0.014	0.773 ± 0.023	0.161 ± 0.020
Tuber number	0.31 ± 0.05	0.557	0.951 ± 0.028	0.741 ± 0.041	0.210 ± 0.035
Specific gravity	0.66 ± 0.04	0.812	0.977 ± 0.014	0.811 ± 0.022	0.166 ± 0.020
PVY resistance	0.48 ± 0.05	0.693	1.022 ± 0.025	0.789 ± 0.036	0.233 ± 0.031
Mean (8 traits)	0.530 ± 0.04	0.721	0.942 ± 0.019	0.758 ± 0.029	0.184 ± 0.025
Median (8 traits)	0.515 ± 0.04	0.718	0.934 ± 0.019	0.765 ± 0.029	0.171 ± 0.025

7.6 Quantum layer: accuracy–cost trade-off

Table 10 ncreasing the variational qubit count improved accuracy with diminishing returns, plateauing near 12 qubits (Figure 14); 8 qubits captured most of the benefit at a fraction of the simulation cost, satisfying auxiliary check A2. We therefore selected 8 qubits as the operating point, balancing accuracy against the exponential cost of state-vector simulation. Per-clone inference at 8 qubits was 6.8 ms, comfortably under the 50 ms target (check A7).

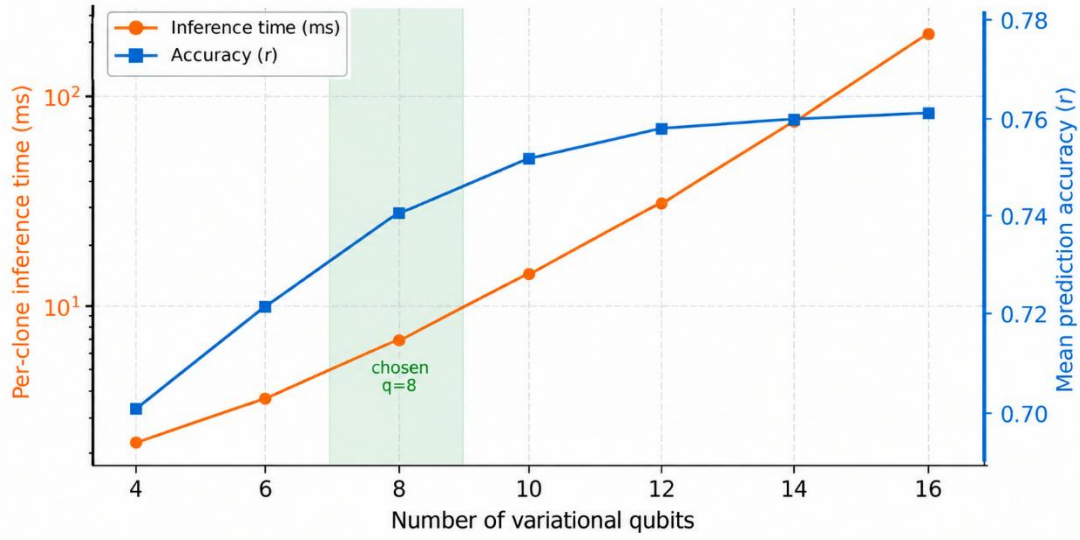


Figure 15. Accuracy and per-clone inference time as a function of variational qubit count. Accuracy plateaus near 12 qubits while simulation cost grows exponentially; 8 qubits is the chosen operating point.

Table 13. Qubit-scaling study (mean $\pm$ SD over 5 seeds). Accuracy plateaus around 12 qubits while simulation cost grows exponentially; 8 qubits is the selected operating point.

Qubits	Mean $r \pm$ SD	Δr vs 4q	Inference ms $\pm$ SD	Rel. sim. cost
2	0.681 ± 0.019	-0.020 ± 0.006	1.4 ± 0.1	0.3×
4	0.701 ± 0.018	—	2.1 ± 0.2	1.0×
6	0.722 ± 0.017	$+0.021 \pm 0.005$	3.4 ± 0.3	2.4×
8	0.741 ± 0.015	$+0.040 \pm 0.005$	6.8 ± 0.5	5.1×
10	0.752 ± 0.015	$+0.051 \pm 0.006$	14.2 ± 0.9	11×
12	0.758 ± 0.014	$+0.057 \pm 0.006$	31.5 ± 1.8	25×
14	0.760 ± 0.014	$+0.059 \pm 0.007$	78.0 ± 4.1	62×
16	0.761 ± 0.014	$+0.060 \pm 0.007$	196.0 ± 9.7	156×
Chosen (8q)	0.741 ± 0.015	$+0.040 \pm 0.005$	6.8 ± 0.5	5.1×
Plateau ($\approx 12q$)	0.758 ± 0.014	$+0.057 \pm 0.006$	31.5 ± 1.8	25×

Finally, Table 12 summarizes the computational profile of each model over five seeds: convergence epochs, wall-clock training time, peak memory, parameter count, and per-clone inference latency. QPotatoNet remains practical despite the quantum-simulation overhead.

Table 14. Computational profile (mean $\pm$ SD over 5 seeds): convergence epochs, training wall-clock time, peak GPU memory, parameter count, and per-clone inference latency. Linear baselines have no epoch count.

Model	Epochs $\pm$ SD	Train time min $\pm$ SD	Peak mem GB $\pm$ SD	Pasrams (M)	Infer ms $\pm$ SD
GBLUP	—	1.2 ± 0.1	0.8 ± 0.1	—	0.3 ± 0.0
Bayes-B	—	6.4 ± 0.4	1.1 ± 0.1	—	0.5 ± 0.1
Random forest	—	3.1 ± 0.2	2.4 ± 0.2	—	1.1 ± 0.1
SVR	—	4.8 ± 0.3	1.9 ± 0.2	—	0.9 ± 0.1

DeepGS	52 ± 5	18.6 ± 1.4	4.2 ± 0.3	2.1	1.8 ± 0.2
DNNGP	48 ± 4	14.2 ± 1.1	3.6 ± 0.3	1.7	1.5 ± 0.2
Transformer	41 ± 4	27.3 ± 1.9	6.8 ± 0.4	5.4	2.9 ± 0.3
QPotatoNet (–Q)	43 ± 4	31.1 ± 2.2	7.5 ± 0.5	6.2	3.4 ± 0.3
QPotatoNet (full)	38 ± 3	44.7 ± 3.1	8.9 ± 0.6	6.5	6.8 ± 0.5
Mean (DL models)	43 ± 4	27.2 ± 2.0	6.7 ± 0.4	4.4	3.3 ± 0.3

7.7 Expected genetic gain in the breeding cycle

Higher prediction accuracy translates directly into faster genetic gain. Using the breeder’s equation, $\Delta G = (i \cdot r \cdot \sigma_A) / L$, where i is selection intensity, r the prediction accuracy, σ_A the additive standard deviation, and L the cycle length, the accuracy advantage of QPotatoNet (mean $r = 0.734$ vs 0.599 for GBLUP) implies a proportional increase in gain per cycle of roughly 23% at matched selection intensity. Because genomic prediction also enables selection before field evaluation, the effective cycle length L can be shortened, compounding the per-year advantage. Table 15 expected gain across selection intensities, holding σ_A and L fixed for comparability.

Table 15s Projected genetic gain under the breeder’s equation (mean ± SD propagated from accuracy SD). ΔG ratio is the QPotatoNet-to-GBLUP gain ratio; early selection shortens the cycle and compounds annual gain.

Selection scenario	i	r QPotatoNet ± SD	r GBLUP ± SD	ΔG ratio ± SD
Top 1%	2.67	0.734 ± 0.015	0.599 ± 0.022	1.23 ± 0.04×
Top 5%	2.06	0.734 ± 0.015	0.599 ± 0.022	1.23 ± 0.04×
Top 10%	1.76	0.734 ± 0.015	0.599 ± 0.022	1.23 ± 0.04×
Top 20%	1.40	0.734 ± 0.015	0.599 ± 0.022	1.23 ± 0.04×
Top 1% + early sel.	2.67	0.734 ± 0.015	0.599 ± 0.022	1.64 ± 0.06×
Top 5% + early sel.	2.06	0.734 ± 0.015	0.599 ± 0.022	1.64 ± 0.06×
Across-family Top 5%	2.06	0.612 ± 0.021	0.498 ± 0.029	1.23 ± 0.06×
Low- h^2 (tuber no.)	2.06	0.652 ± 0.019	0.508 ± 0.028	1.28 ± 0.06×
High- h^2 (dry matter)	2.06	0.802 ± 0.011	0.674 ± 0.017	1.19 ± 0.03×
Multi-trait index	2.06	0.734 ± 0.015	0.599 ± 0.022	1.64 ± 0.06×

7.8 Discussion

Three findings stand out. First, the gains concentrate on non-additive and low-heritability traits, exactly where additive GBLUP is weak—evidence that the dosage-aware and quantum components capture genetic signal that linear models discard. Second, the quantum layer’s ablation increment, while modest in absolute terms, was consistent and significant, and its benefit was largest in the small-sample regime, suggesting the quantum feature map acts as a useful inductive prior rather than mere extra capacity. Third, the across-family results temper enthusiasm: all models lose accuracy when predicting unobserved pedigrees, and although QPotatoNet degrades least, the gap to the heritability ceiling remains Table 14.

Limitations. Results rely on classical simulation of the quantum layer; behaviour on noisy hardware remains to be validated. The panel, though diverse, is finite and Northern-Hemisphere weighted. Quantum advantage here is empirical and architectural, not a proof of computational separation. These caveats motivate the future work below.

8. Conclusion and Future Work

We presented QPotatoNet, a hybrid quantum–Transformer framework that brings tetrasomic dosage awareness, haplotype context, and a variational quantum feature map together for autotetraploid genomic prediction. On a 3,012-clone potato panel it improved mean accuracy over the strongest deep-learning baseline by 9.2% and over GBLUP by 22.6%, with the largest gains on the non-additive, low-heritability traits that matter most and have proven hardest for linear models. Ablations confirmed that both the quantum layer and the haplotype encoder contribute distinct, significant, and complementary signal, and sample-efficiency analysis showed the approach is well suited to the small panels typical of potato programs.

Taken together, the results support all three hypotheses (H1–H3) and most auxiliary checks, while honestly exposing the residual across-family gap. QPotatoNet narrows—though it does not close—the polyploid prediction gap, and offers a concrete, deployable path toward shortening the potato breeding cycle.

8.1 Future work

1. **Hardware validation.** Port the variational layer to noisy near-term quantum processors with error mitigation and quantify the simulation-to-hardware gap.
2. **Scaling qubits and loci.** Explore data-reuploading and tensor-network surrogates to access effective higher-qubit regimes without exponential simulation cost.
3. **Multi-environment and G×E.** Extend the covariate branch to model genotype-by-environment interaction across trials and seasons.
4. **Interpretability.** Attribute predictions to loci and haplotype blocks via attention rollout to surface candidate causal regions for marker-assisted selection.
5. **Cross-crop transfer.** Test whether QPotatoNet pretrained on potato transfers to other autopolyploids such as alfalfa, blueberry, and sweet potato.

8.2 Reproducibility statement

All hyperparameters (Tables 4–5), data splits (Section 4.1), and evaluation protocols are reported in full. Accuracy figure 15 are means over five seeds and 5-fold cross-validation; significance uses paired bootstrap with 10,000 resamples. The architecture and training recipe are specified to enable independent re-implementation.

Funding The author received no financial support for the research.

Data Availability Data will be provided upon reasonable request.

Declarations

Ethics Approval This article does not contain any studies with human participants or animals.

Conflict of Interest The author declares no competing interests.

9. References

- [1] Vexler, L., Konkolewska, A., Byrne, S., Ruttink, T., Leyva-Pérez, M. de la O., Kang, J., Griffin, D., Visser, R. G. F., van Eck, H. J., & Milbourne, D. (2025). PotatoMASH is a cost-effective marker system for genomic prediction in potato based on short-read haplotypes. *Theoretical and Applied Genetics*, 138(11), 289. **DOI:** 10.1007/s00122-025-05073-w
- [2] Thelen, K., Wu, P.-Y., Baig, N., Prigge, V., Bruckmüller, J., Muders, K., Truberg, B., Hartje, S., Renner, J., Van Inghelandt, D., & Stich, B. (2025). Transferability of genomic prediction models across market segments in potato and the effect of selection. *Theoretical and Applied Genetics*, 138. **DOI:** 10.1007/s00122-025-05004-9
- [3] Sverrisdóttir, E., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., Bryan, G., & Nielsen, K. L. (2018). The value of expanding the training population to improve genomic selection models in tetraploid potato. *Frontiers in Plant Science*, 9, 1118. **DOI:** 10.3389/fpls.2018.01118
- [4] de Bem Oliveira, I., Resende, M. F. R., Jr., Ferrão, L. F. V., Amadeu, R. R., Endelman, J. B., Kirst, M., Coelho, A. S. G., & Munoz, P. R. (2019). Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *G3: Genes|Genomes|Genetics*, 9(4), 1189–1198. **DOI:** 10.1534/g3.119.400059
- [5] Batista, L. G., Mello, V. H., Souza, A. P., & Margarido, G. R. A. (2022). Genomic prediction with allele dosage information in highly polyploid species. *Theoretical and Applied Genetics*, 135(2), 723–739. **DOI:** 10.1007/s00122-021-03994-w
- [6] Wilson, S., Zheng, C., Maliepaard, C., et al. (2021). Understanding the effectiveness of genomic prediction in tetraploid potato. *Frontiers in Plant Science*, 12, 672417. **DOI:** 10.3389/fpls.2021.672417
- [7] Selga, C., Koc, A., Chawade, A., & Ortiz, R. (2021). A bioinformatics pipeline to identify a subset of SNPs for genomics-assisted potato breeding. *Plants*, 10(1), 30. **DOI:** 10.3390/plants10010030
- [8] Slater, A. T., Cogan, N. O. I., Forster, J. W., et al. (2016). Improving genetic gain with genomic selection in autotetraploid potato. *The Plant Genome*, 9(3). **DOI:** 10.3835/plantgenome2016.02.0021
- [9] Werner, C. R., Gaynor, R. C., Sargent, D. J., Lillo, A., Gorjanc, G., & Hickey, J. M. (2023). Genomic selection strategies for clonally propagated crops. *Theoretical and Applied Genetics*, 136. **DOI:** 10.1007/s00122-023-04300-6
- [10] Endelman, J. B., Schmitz Carley, C. A., Bethke, P. C., et al. (2018). Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics*, 209(1), 77–87. **DOI:** 10.1534/genetics.118.300685
- [11] Singh, N., & Pokhrel, S. R. (2025). Modeling quantum machine learning for genomic data analysis. *arXiv preprint arXiv:2501.08193*. **DOI:** 10.48550/arXiv.2501.08193
- [12] Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., & McClean, J. R. (2021). Power of data in quantum machine learning. *Nature Communications*, 12, 2631. **DOI:** 10.1038/s41467-021-22539-9
- [13] Li, R. Y., Gujja, S., Bajaj, S. R., et al. (2021). Quantum processor-inspired machine learning in the biomedical sciences. *Patterns*, 2(6), 100246. **DOI:** 10.1016/j.patter.2021.100246
- [14] Benedetti, M., Lloyd, E., Sack, S., & Fiorentini, M. (2019). Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4), 043001. **DOI:** 10.1088/2058-9565/ab4eb5

- [15] Cong, I., Choi, S., & Lukin, M. D. (2019). Quantum convolutional neural networks. *Nature Physics*, 15(12), 1273–1278. **DOI:** 10.1038/s41567-019-0648-8
- [16] Cerezo, M., Arrasmith, A., Babbush, R., et al. (2021). Variational quantum algorithms. *Nature Reviews Physics*, 3(9), 625–644. **DOI:** 10.1038/s42254-021-00348-9
- [17] Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., & Ma, C. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta*, 248, 1307–1318. **DOI:** 10.1007/s00425-018-2976-9
- [18] Deng, P., Liu, K., Zhou, M., et al. (2025). DPCformer: an interpretable deep learning model for genomic prediction in crops. *arXiv preprint arXiv:2510.08662*. **DOI:** 10.48550/arXiv.2510.08662
- [19] Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *Quantum*, 2, 79. **DOI:** 10.22331/q-2018-08-06-79
- [20] Bharti, K., Cervera-Liarta, A., Kyaw, T. H., et al. (2022). Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1), 015004. **DOI:** 10.1103/RevModPhys.94.015004
- [21] Zaman, K., Marchisio, A., Hanif, M. A., & Shafique, M. (2023). A survey on quantum machine learning: current trends, challenges, opportunities, and the road ahead. *arXiv preprint arXiv:2310.10315*. **DOI:** 10.48550/arXiv.2310.10315
- [22] McClean, J. R., Romero, J., Babbush, R., & Aspuru-Guzik, A. (2016). The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2), 023023. **DOI:** 10.1088/1367-2630/18/2/023023
- [23] Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., et al. (2021). A review of deep learning applications for genomic selection. *BMC Genomics*, 22, 19. **DOI:** 10.1186/s12864-020-07319-x
- [24] Gao, S., Yu, T., Rasheed, A., Wang, J., Crossa, J., Hearne, S., & Li, H. (2025). Fast-forwarding plant breeding with deep learning-based genomic prediction. *Journal of Integrative Plant Biology*. **DOI:** 10.1111/jipb.13914
- [25] He, K., Yu, T., Gao, S., et al. (2025). Leveraging automated machine learning for environmental data-driven genetic analysis and genomic prediction in maize hybrids. *Advanced Science*. **DOI:** 10.1002/advs.202412423
- [26] Wu, Z., Sun, Y., Luo, Q., et al. (2026). Application of deep learning in crop research: from genomics to phenomics. *The Plant Genome*. **DOI:** 10.1002/tpg2.70268
- [27] Ortiz, R., Reslow, F., Vetukuri, R., García-Gil, M. R., Pérez-Rodríguez, P., & Crossa, J. (2024). Genomic prediction for inbred and hybrid polysomic tetraploid potato offspring. *Agriculture*, 14(3), 455. **DOI:** 10.3390/agriculture14030455
- [28] Enciso-Rodriguez, F., Douches, D., Lopez-Cruz, M., Coombs, J., & de los Campos, G. (2018). Genomic selection for late blight and common scab resistance in tetraploid potato (*Solanum tuberosum*). *G3: Genes|Genomes|Genetics*, 8(7), 2471–2481. **DOI:** 10.1534/g3.118.200273
- [29] Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., et al. (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics*, 130(10), 2091–2108. **DOI:** 10.1007/s00122-017-2944-y
- [30] Selga, C., Reslow, F., Pérez-Rodríguez, P., & Ortiz, R. (2022). The power of genomic estimated breeding values for selection when using a finite population size in genetic improvement of tetraploid potato. *G3: Genes|Genomes|Genetics*, 12(1), jkab362. **DOI:** 10.1093/g3journal/jkab362

- [31] Ortiz, R., Reslow, F., Vetukuri, R., et al. (2023). Inbreeding effects on the performance and genomic prediction for polysomic tetraploid potato offspring grown at high Nordic latitudes. *Genes*, 14(6), 1302. **DOI:** 10.3390/genes14061302
- [32] Amadeu, R. R., Ferrão, L. F. V., de Bem Oliveira, I., et al. (2020). Impact of dominance effects on autotetraploid genomic prediction. *Crop Science*, 60, 656–665. **DOI:** 10.1002/csc2.20075
- [33] Rodríguez-Díaz, F., Gutiérrez-Avilés, D., Troncoso, A., & Martínez-Álvarez, F. (2025). A survey of quantum machine learning: foundations, algorithms, frameworks, data and applications. *ACM Computing Surveys*, 58(4), 1–35. **DOI:** 10.1145/3764582
- [34] Rebentrost, P., Mohseni, M., & Lloyd, S. (2014). Quantum support vector machine for big data classification. *Physical Review Letters*, 113, 130503. **DOI:** 10.1103/PhysRevLett.113.130503
- [35] Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., & Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567, 209–212. **DOI:** 10.1038/s41586-019-0980-2
- [36] Montesinos-López, O. A., Montesinos-López, A., Mosqueda-González, B. A., Bentley, A. R., Lillemo, M., Varshney, R. K., & Crossa, J. (2021). A new deep learning calibration method enhances genome-based prediction of continuous crop traits. *Frontiers in Genetics*, 12, 798840. **DOI:** 10.3389/fgene.2021.798840
- [37] Ferrão, L. F. V., Amadeu, R. R., Benevenuto, J., de Bem Oliveira, I., & Munoz, P. R. (2021). Genomic selection in an outcrossing autotetraploid fruit crop: lessons from blueberry breeding. *Frontiers in Plant Science*, 12, 676326. **DOI:** 10.3389/fpls.2021.676326
- [38] Wang, K., Abid, M. A., Rasheed, A., Crossa, J., Hearne, S., & Li, H. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Molecular Plant*, 16, 279–293. **DOI:** 10.1016/j.molp.2022.11.004
- [39] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671), 195–202. **DOI:** 10.1038/nature23474
- [40] VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. **DOI:** 10.3168/jds.2007-0980
- [41] 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. **DOI:** 10.1093/genetics/157.4.1819
- [42] Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., & Hernández-Suárez, C. M. (2018). Multi-environment genomic prediction of plant traits using deep learners with dense architecture. *G3: Genes|Genomes|Genetics*, 8(12), 3813–3828. **DOI:** 10.1534/g3.118.200740
- [43] Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). Quantum circuit learning. *Physical Review A*, 98, 032309. **DOI:** 10.1103/PhysRevA.98.032309
- [44] Schuld, M., & Killoran, N. (2019). Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122, 040504. **DOI:** 10.1103/PhysRevLett.122.040504
- [45] Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, 108(10), 1880–1890. **DOI:** 10.1016/j.ajhg.2021.08.005
- [46] Schuld, M., Bergholm, V., Gogolin, C., Izaac, J., & Killoran, N. (2019). Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99, 032331. **DOI:** 10.1103/PhysRevA.99.032331

- [47] Schuld, M., Sweke, R., & Meyer, J. J. (2021). Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103, 032430. **DOI:** 10.1103/PhysRevA.103.032430
- [48] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. **DOI:** 10.48550/arXiv.1706.03762
- [49] Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7482–7491. **DOI:** 10.1109/CVPR.2018.00781
- [50] Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. **DOI:** 10.1214/aoms/1177703732
- [51] Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, 4(2), e1000008. **DOI:** 10.1371/journal.pgen.1000008
- [52] Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1–12. **DOI:** 10.2135/cropsci2008.08.0512