Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Bio-economic Modeling for Swine Genomic Selection Index Design and Multi-Generational Response Simulation

Caiming Zhang[1], Ximing Dong[1], Haiyang Lu[1], Peiyuan Wang[1], Le Sun[1], Xinyun Li[1], Guojian Ma[2,3*] and Tao Xiang[1*]

[1] Key Laboratory of Agricultural Animal Genetics Breeding and Reproduction of Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China

[2] Breeding Department, Wuhan COFCO Meat Co., Ltd., Wuhan 430074, China

[3] Hubei Hongshan Laboratory, Wuhan 430070, China

* Correspondence: Tao Xiang, Tao.Xiang@mail.hzau.edu.cn; Guojian Ma, maguojian@cofco.com

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Abstract

The selection index is related to the efficiency and direction of genetic improvement for livestock. Nevertheless, specialized software that employs multiple bio-economic models to design and compute the selection index and predict the multi-generational selection response is still lacking. In this study, to meet the needs of pig breeding programs, visualized multivariate model-based software for pig selection index design and computation, CSIP, was developed to rapidly construct a selection index that is applicable to different pig breeding systems, such as farrow-to-finish herds, single pure-line, two-way cross, and three-way cross breeding programs, and to predict multi-generational selection response. The main advantages of CSIP include full-process visualization for the design of a selection index, which increases transparency and user comprehension; incorporation of multi-generational selection response, which facilitates the proactive mitigation of antagonistic trait response and suboptimal genetic gain; integration of gene flow and economic weight sensitivity analysis, which increases robustness and prevents errors in economic weight calculation; and an accessible, workflow-based interface that supports flexible calculation strategies to accommodate diverse practical breeding scenarios. These features position CSIP as an innovative tool for increasing the precision, efficiency, and economic viability of pig breeding programs.

Keywords: Gene flow; Multi-generational selection response; Bio-economic model; Selection index

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Introduction

In pig breeding, the simultaneous improvement of multiple traits is often required to meet production and economic goals. To address this requirement, the selection index method has been widely adopted, as it integrates information from various traits into a single numerical value by combining genetic and economic parameters [1,2]. A selection index is typically a linear combination of estimated breeding values (EBVs) and their corresponding economic weights, where the economic weights reflect the relative economic importance of each trait [3-5]. The ultimate goal of animal breeding is to maximize the overall selection response in alignment with a defined economic objective [6,7]. Therefore, accurately determining the economic weight of each trait is central to the construction of an effective selection index.

The estimation of economic weights for different traits in the pig selection index has been explored extensively since the concept of the selection index was proposed [8-15]. The simple profit–function approach, which uses a simple function to calculate the partial derivative of traits to obtain economic weights [10,16,17], is often inadequate for complex pig breeding systems; subsequently, bio-economic models can be established to derive economic weights [8,9,11,18-20]. At the initial stage of using the selection index, marginal economic values (MEVs) were regarded as the economic weights for different traits [1,9]. However, real pig production systems involve distinct population structures (nucleus, multiplier, and commercial herds), time delays in trait expression between parents and offspring, and varying gene expression across breeds, rendering the use of simple MEVs as economic weights insufficient [21]. Therefore, economic weights in bio-economic models can be calculated as the product of a trait's MEVs and its corresponding number of discounted gene expressions (NDEs) to capture

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

both the trait's economic impact and its transmission dynamics across generations [22,23]. The NDEs reflects the accumulation of selected parental genes in progeny across generations and production tiers, with trait-expressing genes discounted back to the year of parental selection using time-adjusted discount rates [23,24]. For compute NDEs, a gene flow method [24,25], which has been successfully applied to optimize pig breeding programs [26-28], was applied to quantify gene expression. To expand the scenarios for the use of bio-economic models, Wolfová *et al.* [29] developed a flexible bio-economic model to calculate the MEVs of traits in diverse pig breeding systems, and subsequently, a computation program was developed to calculate economic weights [23].

After a selection index is formulated, it is essential for breeders to evaluate whether the index maximizes the multi-generational selection response [30]. This evaluation not only verifies the alignment between the index and the breeding objectives but also facilitates the refinement of breeding strategies and the optimization of index design. However, studies of the integration of multi-generational selection response prediction with the formulation of selection indices are still lacking. Moreover, there is currently no software that supports the full workflow—from index formulation to multi-generational response prediction and visualization—which limits the effective and practical application of the selection index in pig breeding.

The goal of this study is thus to develop an intuitive, computationally efficient, and fully visualized tool to support the formulation of a pig selection index in diverse breeding scenarios while enabling the prediction of multi-generational selection response. To achieve this goal, we constructed bio-economic models for farrow-to-finish (commercial) herds and different pig breeding systems (including pure-line, two-way cross, and three-way cross breeding systems)

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

to derive MEVs; integrated the gene flow approach into distinct pig breeding systems to calculate NDEs and the corresponding economic weights; incorporated predictions of multi-generational selection response to validate the rationality of the formulated index; and streamlined the index formulation workflow for user accessibility and flexibility in its implementation. CSIP features built-in sensitivity analysis for economic weights, standardized index computations, and comprehensive visualization tools to meet both research and industry demands.

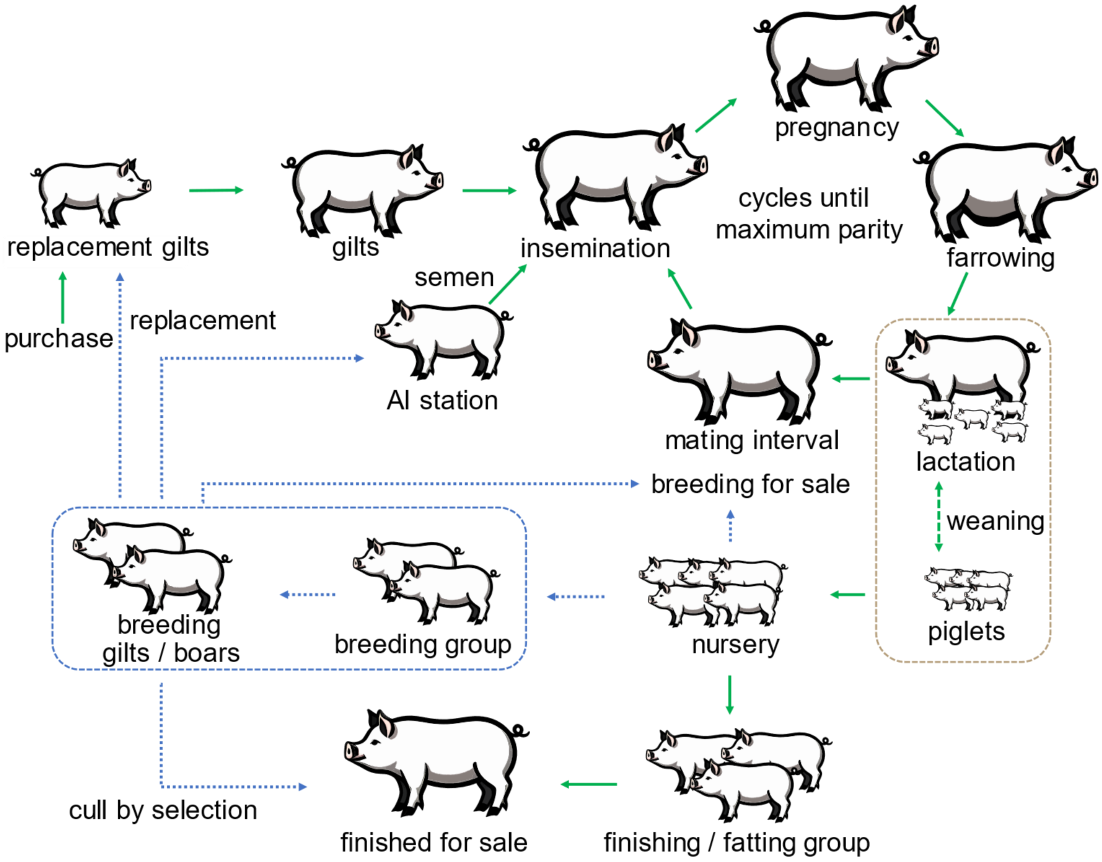

Figure 1. Basic framework of the bio-economic models for farrow-to-finish and pure-line herds. The diagram illustrates the integrated production system, encompassing the life-history of breeding sows (reproduction and growth) and the complete rearing process of the offspring.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

The part connected by the blue box and the dotted line is included only in the pure-line model, and the remaining parts are contained in both models.

Methods

In this study, several bio-economic models were established to derive economic values for different pig traits. Among these models, the commercial farrow-to-finish herd model was adapted from De Vries [31]. In addition, we modeled the standard pure-line, two-way cross, and three-way cross breeding systems. These structures were selected as they represent the most widely adopted production systems in the global swine industry, where artificial insemination (AI) is the standard practice. Within the CSIP, these models were parameterized to construct deterministic profit functions, integrating twenty-six specific traits to derive precise marginal economic values.

Farrow-to-finish (commercial) herd bio-economic model framework

The farrow-to-finish (commercial) herd bio-economic model (Fig. 1 and Supplementary Fig. S1A) manages the complete lifecycle of purchased gilts within a single farm, commencing with gilt acquisition and rearing to breeding age, followed by systematic artificial insemination implementation. Sows progress cyclically through successive reproductive phases—insemination, pregnancy, farrowing, lactation, and mating intervals—until they reach maximum productive longevity, with standardized feeding procedures dynamically adjusted to physiological stage-specific nutritional demands and health status. Throughout this cycle, sows face culling risks due to mortality, disease, no heat, or suboptimal litter performance. Progeny

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0

development occurs through three sequential phases: lactation, where piglets derive primary nutrition from suckling while progressively incorporating creep feed; nursery, which follows weaning and facility transfer; and finishing, where pigs advance to target slaughter weight. To reflect operational realities, the nursery and finishing stages may be subdivided into one-to-three management intervals aligned with specific feeding regimens to accommodate evolving nutritional requirements. Slaughter pigs are sold on the basis of live weight, carcass weight, or quality grading according to actual market conditions.

Single pure-line breeding system bio-economic model framework

The single pure-line breeding system bio-economic model (Fig. 1 and Supplementary Fig. S1B) achieves within-breed genetic improvement through controlled self-breeding, commencing with gilt acquisition via internal selection or external purchase and progressing through sequential reproductive cycles until maximum productive longevity is reached. The structure of this model is further optimized on the basis of the farrow-to-finish herd model. The difference is that after the nursery period is completed, the pigs are separated into breeding group and fattening group. Within the breeding group, breeders or farmers evaluate pig performance and select the best offspring as replacements for boars and sows, and the superior boars are transferred to AI stations for semen production, while suboptimal candidates join commercial fattening streams. Notably, during the nursery stage and in the breeding group, a certain number of breeding pigs can be sold outside the system for purchase and use by other herds. The unselected gilts/boars are fattened and sold as commercial pigs along with the fattening group, collectively constituting the system's dual-output revenue structure.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Two-way crossbreeding system bio-economic model framework

The two-way crossbreeding system bio-economic model integrates dam breed A and sire breed B within a tiered production structure, synthesizing frameworks from both farrow-to-finish herd and pure-line breeding models (Supplementary Fig. S2A). Nucleus herds maintain purebred populations (A and B), whereas multiplier herds utilize semen transferred from breed B to inseminate breed A dams, generating hybrid progeny (A×B). To account for practical variations in productive lifetime and rearing environments when breed A is utilized for different purposes within breeding programs, individuals designated exclusively for hybrid progeny production were segregated into distinct reproduction A. Within this integrated system, breeds A and B retain their respective reproductive functionalities, with critical directional flows: purebred female offspring of breed A either replenish nucleus herds or transfer to multiplier herds for hybrid production, while male offspring of breed B supply semen for intra-breed artificial insemination or crossbreeding in multiplier herds. Hybrid offspring (A×B) are exclusively used for fattening and sale, which establishes a clear genetic and economic separation between purebred maintenance streams and commercial output streams.

Three-way crossbreeding system bio-economic model framework

The three-way crossbreeding system bio-economic model integrates breed A, B and C through a hierarchical genetic pyramid that includes the nucleus, multiplier and commercial herds. This model extends integrated frameworks from farrow-to-finish herd and pure-line system to five distinct populations (Supplementary Fig. S2B): nucleus herds maintaining pure breeds A, B, and C; reproduction A groups producing A×B hybrids using breed A gilts (sourced from the

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

nucleus) and breed B semen; and reproduction AB groups where elite A×B females (selected from reproduction A outputs) are crossed with breed C semen to produce exclusively (A×B)×C finishers. This process establishes a unidirectional genetic flow from nucleus conservation to commercial terminal cross production.

Bio-economic model sow herd and offspring structure

On the basis of the framework of each model, the population counts of sows and their offspring at different production stages serve as the fundamental basis for determining total profit. Initially, the sow population is assumed to remain constant and is distributed in dynamic equilibrium across reproductive cycles. In the farrow-to-finish herd model, to minimize the number of model parameters and enhance applicability, sow numbers at subsequent stages are calculated on the basis of stage-specific mortality and culling rates. Using parameters for litter size and stage-specific offspring mortality rates, the model subsequently computes the number of offspring surviving each stage.

For other breeding systems, owing to the complexity inherent in breeding systems, the steady-state distribution of sows across various physiological states within different reproductive cycles is calculated using the Markov chain method [32,33]. We adopted the method of constructing the Markov chain transition matrix $\boldsymbol{T}$ by Wolfová *et al.* [29], which has been successfully applied to calculate the structure of the sow herd. The transition matrix $\boldsymbol{T}$ represents the sequential transfer process among all parities of the sow before the maximum service life is reached. For each parity, sow states are categorized into mutually exclusive scenarios: sows reaching the maximum number of allowed matings and culled due to

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

conception failure; sows dying during pregnancy or lactation; sows culled during the mating interval for health reasons; and sows advancing to the next breeding cycle. The probabilities of these mutually exclusive scenarios are regarded as elements in the matrix $\boldsymbol{T}$. Then, an initial vector $\boldsymbol{I}$ is established to represent the probability ratio at the time of population establishment, with dimensions equal to the number of rows in matrix $\boldsymbol{T}$:

$$\boldsymbol{I} = [1\ 0\ 0\ \cdots 0]$$

Thus, the state distribution $\boldsymbol{P}$ of the number of sows at time t is expressed as follows:

$$\boldsymbol{P}_t = \boldsymbol{I}\boldsymbol{T}^t$$

According to the properties of the Markov chain, $\boldsymbol{P}$ remains stable as time t increases. We set the default value of t to 30 to obtain the probability distribution of sows in the stable state as well as the number of sows in different productive states. The parameters for litter size and stage-specific offspring mortality rates were subsequently used to compute the number of offspring surviving at each stage. These models integrate the growth and reproductive processes of sows and offspring into a unified system, regardless of whether offspring are raised on separate finishing farms.

Costs and revenues in bio-economic model

Building upon these population frameworks, the costs and revenues of all the components of the bio-economic models are considered, which are included in the models as follows:

Costs include the following:

- Feed costs for gilts and sows at different stages (e.g., pregnancy, lactation, and the mating interval).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

- Feed costs for piglets during lactation (creep feed).
- Feed costs for weaned pigs in the nursery and fattening pigs in the feeder-to-finish stage (up to three stages can be considered, depending on actual feeding needs).
- Feed costs for the breeding group, separated by sex (boars and replacement gilts during the feeder-to-finish stage, with up to three stages considered).
- Fixed costs for gilts and sows (e.g., medical costs, housing costs, water, electricity, environmental costs, and labor costs).
- Artificial insemination costs and farrowing costs for gilts and sows.
- Costs for the disposal of dead animals.
- Fixed costs for sow offspring at different stages (e.g., testing and evaluation of breeding animals, medical costs, housing costs, water, electricity, environmental costs, and labor costs).
- Costs for transporting and selling pigs.

Revenue includes the following:

- Revenue from culling sows and selling breeding group pigs.
- Revenue from selling finished pigs.

The data required in the calculation process of these costs and revenues are used as parameters input to the model and must be input by the user.

Production efficiency

In farrow-to-finish herds, production efficiency is measured as total profit per finished pig, expressed as follows:

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

$$Prof_f = \frac{Trevenues_f - Tcosts_f}{number\ of\ finished\ pigs}$$

where $Prof_f$ is the total profit per finished pig and where $Trevenues_f$ is the total revenue, including revenue from culling sows and finished pigs for sale. $Tcosts_f$ is the total cost, which includes all feed costs and non-feed costs for the sows and their offspring.

In other breeding systems, production efficiency is expressed as the total profit per sow per parity as follows:

$$Prof_s = \frac{Trevenues_s - Tcosts_s}{number\ of\ sows}$$

where $Prof_s$ is the profit per sow in this system and where $Trevenues_s$ is the total revenue, including revenue from culling sows, breeding group pigs and finished pigs for sale. $Tcosts_s$ is the total cost, which includes all feed costs and non-feed costs for the sows and their offspring.

Calculation of NDEs by using the gene flow method

Conceptually, the gene flow method is designed to mirror the biological lifecycle of the swine herd by tracking gene transmission through two fundamental processes: aging (growth) and reproduction. To operationalize this, the population was stratified into distinct sex-age groups with a discrete time step of half a year, corresponding to the average reproductive cycle of sows. According, a gene transmission matrix **P** was constructed based on the breeding system's structure, where rows represent sex–age groups at time t and columns represent groups at $t - 1$.

For the pure-line bio-economic model (Fig. 2A), **P1** is partitioned into four quadrants to capture specific transmission pathways: the upper-left and lower-right quadrants represent within-sex transmission (sire-to-sire and dam-to-dam), while the off-diagonal quadrants

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

represent cross-sex transmission (sire-to-dam and dam-to-sire). Each quadrant integrates two matrices: reproduction ($\mathbf{P2}$) and ageing ($\mathbf{P3}$). Within the reproduction matrix $\mathbf{P2}$, the elements in the first row (represented by asterisks, *) quantify the proportion of genes transmitted from parental group 1-x at $t-1$ to the offspring (age group 1) at time t; these values range from 0 to 0.5, representing the Mendelian inheritance probability scaled by the reproductive contribution of the respective age group. Conversely, the growth matrix $\mathbf{P3}$ contains identity elements ("1") on the sub-diagonal, representing the survival and transition of animals to the subsequent age class. This framework extends to crossbreeding systems by integrating multiple pure-line dynamics. For the two-way crossbreeding system (Fig. 2B), the matrix tracks gene flow through maternal line A and paternal line B, explicitly modeling the production of hybrid offspring. Owing to age-dependent utilization differences in maternal line A, the reproduction A gene transmission process is modeled separately. The complete gene flow matrix integrates submatrices $\mathbf{P1}$, $\mathbf{P2}$, $\mathbf{P3}$, and $\mathbf{P4}$, with all remaining non-functional elements set to 0. Similarly, in the three-way crossbreeding system (Fig. 2C), reproduction AB is modeled independently to account for age-specific usage, with the global transmission matrix constructed to strictly align with the artificial insemination protocols and the hierarchical flow of the crossbreeding program.

To calculate the gene expression of the selected breeding pigs, a gene probability vector $\boldsymbol{m}$ (dimensions matching the number of age–sex groups) was defined. At $t = 0$, the element corresponding to the initial age–sex group of the selected pigs was assigned a value of 1, with all other elements initialized to 0. The temporal probability distribution of genes across age–sex groups was derived iteratively using the following equation:

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

$$\boldsymbol{m_t} = \boldsymbol{P}\boldsymbol{m_{t-1}}$$

where $\boldsymbol{m_t}$ represents the gene probability vector at time t. Traits expressed earlier in the breeding period carry greater economic weight than those expressed later. A discount factor r was therefore applied to account for the temporal value of breeding investments. For each trait, a realization vector **h** (dimensions matching the number of age–sex groups) was defined to represent trait expression levels across sex–age groups. Its elements denote the proportion of pigs expressing this trait in each age–sex group of pigs relative to the total number of sows. This vector **h** was assumed to be constant throughout the investment period. The NDEs for a trait are calculated as follows:

$$n_{ij} = \sum_{t=1}^{D} \boldsymbol{h'}_i \boldsymbol{m_{jt}} (1+r)^{-t}$$

where n_{ij} represents the NDEs of trait i in sex group j; D represents the breeding investment period; $\boldsymbol{h_i}$ represents the realization vector for trait i; $\boldsymbol{m_{jt}}$ represents the gene probability vector for sex group j at time t; r represents the discount factor; and t represents the time of trait expression.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

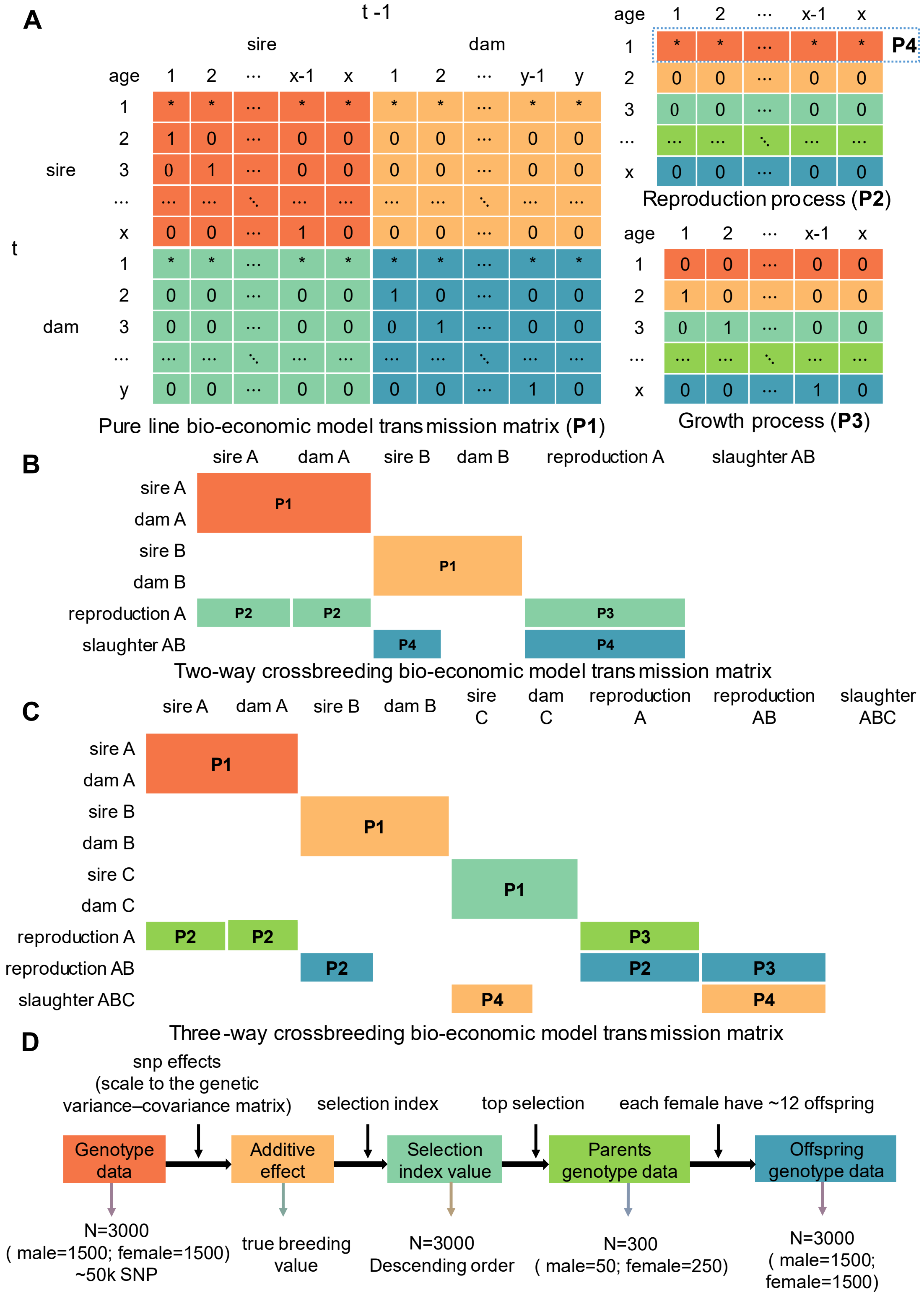

Figure 2. Structure of gene transmission matrices and the simulation flowchart. (A) The transmission matrix (**P1**) for the pure-line bio-economic model. The matrix models the gene flow between sex-age groups over a time step of half a year (representing one age class).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Columns represent the population state at time $t-1$, and rows represent time t. The matrix is divided into four quadrants representing genetic transmission pathways (sire to sire, sire to dam, dam to sire, dam to dam). The first row of each quadrant (marked with asterisks, *) represents the genetic contribution of parents to offspring (values 0–0.5, reflecting Mendelian inheritance). The sub-diagonal elements marked with "1" represent the aging process, where surviving individuals move to the next age class. (B) The transmission matrix for the two-way crossbreeding system. This matrix integrates the pure-line dynamics for Dam A and Sire B, modeling the flow into Reproduction A and the terminal Slaughter AB group. (C) The transmission matrix for the three-way crossbreeding system. This extends the model to include Sire C, producing the final Slaughter ABC population through sequential crossbreeding steps. (D) Flowchart of the multi-generational simulation. The process illustrates the calculation steps from initial genotype data to the calculation of additive effects, selection index values, and the selection of top parents to generate the next generation's offspring genotype data.

Economic weights and sensitivity analysis

The MEV of a trait is calculated as the change in profit per unit change in the trait mean (TM), corresponding to the partial derivative of the profit function. In this study, this value is approximated via a central finite difference method:

$$MEV = \frac{\partial Prof}{\partial TM} \approx \frac{Prof_{(1+0.1\%)TM} - Prof_{(1-0.1\%)TM}}{0.2\% TM}$$

where $Prof_{(1+0.1\%)TM}$ and $Prof_{(1-0.1\%)TM}$ represent profits when the trait mean is increased or decreased by 0.1%, respectively. A variation of $\pm 0.1\%$ was chosen to minimize discretization error, ensuring the result accurately approximates the tangent slope of the profit

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

function at the population mean. In farrow-to-finish herds, the MEV serves as a direct proxy for trait weight in the selection index. For other breeding systems, the economic weight (EW_i) of trait i is derived by scaling its MEV by the NDE (n_{ij}) associated with the selected sex group, following $EW_i = MEV_i \times n_{ij}$.

To explore the sensitivity of economic weights to such fluctuations, three simulation scenarios (feed price fluctuations, non-feed price fluctuations, and sale price fluctuations) were set up to predict the effects of different types of price fluctuations on economic weights. In each simulated scenario, the economic weights of target traits under incremental price increases of 10%, 20%, 30%, 40%, and 50% can be iteratively calculated via a bio-economic model.

Relative economic weights and standardized selection index

To compare the relative importance of different traits, standardized economic weights (SEWs) were first calculated by multiplying each trait’s economic weight (EW) by its genetic standard deviation. The relative economic weight (REW) was expressed as the percentage contribution of a trait’s SEW to the total SEWs of all the traits:

$$SEW_i = EW_i \times \sigma_i$$

$$REW_i = \frac{SEW_i}{\sum_1^n SEW_i} \times 100\%$$

where σ_i is the genetic standard deviation of trait i and where n is the total number of traits.

In the context of modern breeding programs, the selection index (I) is first calculated as the linear combination of EW and EBV:

$$I = \sum_{i=1}^{n} (EW_i \times EBV_i)$$

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Unlike the classical index which relies on phenotypic to infer genetic merit, this approach utilizes pre-estimated EBVs directly. Since EBVs are now directly computable via upstream evaluations (e.g., Pedigree-BLUP or GBLUP), the construction of a phenotype-based index (and its associated covariance estimation) is rendered unnecessary. Subsequently, to facilitate practical decision-making, the mean and variance of this selection index are standardized (H**)** and expressed as follows:

$$H = \mu + \frac{I}{\sqrt{\boldsymbol{EW}'(\boldsymbol{V})\boldsymbol{EW}}} \times STD$$

where μ is the average of the standardized selection index (e.g., 100), $\boldsymbol{V}$ is the variance–covariance matrix of the EBVs, and STD is the standard deviation of the standardized selection index (e.g., 25).

Simulation of multi-generational selection response

The selection response module is validated using a built-in simulated pig population containing 3,000 genotyped individuals (~50K SNP density) with full pedigree records. This scale effectively represents a nucleus breeding herd and allows for rapid simulation of genetic gain. This synthetic population, adapted from our previous work [34], replicates realistic genetic architectures, trait correlations, and population structures observed in real pig herds. To predict multi-generation genetic gain, the simulation workflow proceeded as follows (Fig. 2D): it was assumed that target traits were affected by a certain number of QTLs per chromosome, and additive genetic effects for QTLs were sampled from a multivariate normal distribution scaled to match the user-defined genetic variance–covariance matrix. For each generation, candidate pigs were ranked by the calculated selection index, derived from true breeding values (TBVs)

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0

to validate the theoretical efficiency of the index. The top 50 males and 250 females were selected as parents for the next generation, and each female produced 12 offspring (sex ratio of 0.5), maintaining a population size of 3,000 per generation. For each generation, pedigrees, genotypic additive effects, and indices were recorded, and the means of additive effects and selection indices were visualized as line graphs to track the multi-generation selection response throughout the future breeding cycles/generations. This simulation is explicitly designed to allow users to rapidly validate the expected genetic gain of the index on standard personal computers, balancing statistical representativeness with operational efficiency.

QTL allelic effect values

The built-in genotype data encompass 18 chromosomes, with each chromosome associated with 40 quantitative trait locus (QTL) positions. First, the genotype data are read to extract the genomic positions of both the QTLs and the markers. The effect value for each QTL locus is subsequently simulated using a multivariate normal distribution. Let the vector $\boldsymbol{X}$ represent the effect of each QTL on n traits, following a multivariate normal distribution $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{u}$ represents the mean vector and $\boldsymbol{V}$ represents the genetic variance–covariance matrix of n traits. Then, N $(QTL \times chromosome\ number)$ independent and identically distributed random samples are generated, and the locus effect values matrix $\boldsymbol{A}$ is obtained:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X}_1^T \\ \boldsymbol{X}_2^T \\ \vdots \\ \boldsymbol{X}_N^T \end{bmatrix}_{N \times n}$$

where $\boldsymbol{X}_i \sim N_n\ (0, \boldsymbol{V})$ and $i = 1, 2, \cdots, N$. Finally, the TBVs for the traits in the base population are computed by multiplying the genotype dosage matrix (sum of alleles 1 and 2 at each QTL)

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

by the QTL effect values matrix, with these effects subsequently standardized to align with the target population's genetic variance–covariance matrix. Scaling is performed as follows:

$$Scale\ = \frac{sqrt(Input\ variance\ matrix)}{sqrt(TBVs\ variance\ matrix)}$$

then, the QTL allelic effect at each locus is scaled as $(allelic\ effect\ -\ QTL\ mean) \times Scale$. The effect values after scaling are used to calculate the subsequent additive genetic effect.

Software architecture and system requirements

The software was developed using a multi-tier architecture. The frontend was implemented with *Vue.js*, which is a responsive and intuitive interface that facilitates seamless interaction with various modules and visualizations. The backend computational engine was developed in *Python*, with *NumPy* and *Pandas* used for efficient numerical processing and data manipulation and *SciPy* used for advanced scientific computations. *Electron* was used to package the application, and excellent cross-platform capabilities were ensured.

Selection index for pig production in the farrow-to-finish (commercial) herd and three-way crossbreeding system

We use the farrow-to-finish (commercial) herd model to demonstrate the functions of the CSIP software. A dam selection index with five traits (age at target weight (ATW), finishing-phase feed conversion ratio (F-FCR), backfat depth (BFD), loin muscle depth (LMD), and number of born alive (NBA)) was formulated, and the multi-generational genetic response was simulated. Moreover, a three-way crossbreeding system model, incorporating genetic data from three

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

nucleus populations (Yorkshire, Landrace, and Duroc) and economic parameters derived from typical commercial production systems and market surveys, was employed to calculate economic weights and design indices to demonstrate that the software can calculate the economic weights of up to 26 traits and the selection index of the crossbreeding system. The input parameters that generate these results can be found in Supplementary Table S1–6.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

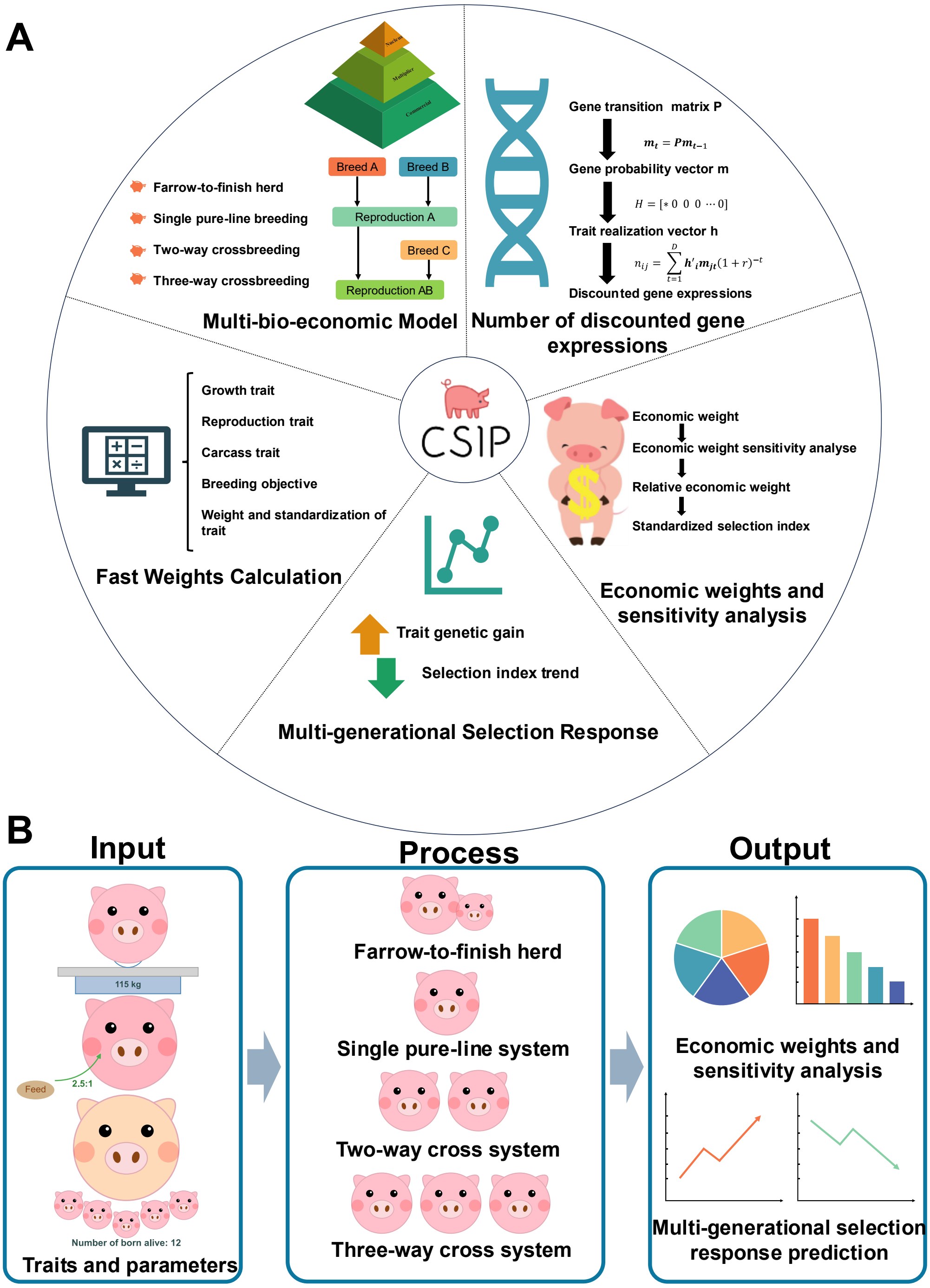

Figure 3. Overview of the CSIP software. (A) Core functional modules. The diagram illustrates the five interconnected computational components integrated within CSIP: the multi-bio-economic model, number of discounted gene expressions, economic weights and sensitivity analysis, multi-generational selection response, and fast weights calculation. (B) Operational

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

workflow. The diagram presents the data flow from the input of traits and parameters, through the Process of breeding system, to the output of visualized analysis results.

Results

Overview of the CSIP software

CSIP is specialized software for formulating pig selection index via the integration of bio-economic models and multi-generational prediction of selection response within a streamlined, practical, and fully visualized workflow. Its core comprises five interconnected modules (Fig. 3A): the multi-bio-economic model, number of discounted gene expressions, economic weights and sensitivity analysis, multi-generational selection response, and fast weights calculation. The multi-bio-economic model module provides selectable frameworks (e.g., farrow-to-finish (commercial) herd, single pure-line, two-way cross, and three-way cross breeding systems) for calculating the MEVs of traits. Each model integrates 26 economic productive traits, including growth, reproduction, and carcass traits. The number of discounted gene expressions module uses a gene flow approach to compute NDEs. The obtained MEVs and NDEs are then integrated into the economic weights and sensitivity analysis module to derive economic weights and analyze the sensitivity of these weights to fluctuations in economic parameters. This module also supports the calculation of relative economic weights alongside the formulation of a standardized selection index. The multi-generational selection response module allows the prediction of multi-generational selection response when the designed selection index is applied in the breeding system. This approach allows a comparison of the consequences of genetic gain when a suboptimal selection index is applied in a breeding project. The fast weights calculation

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

module provides alternative flexible index calculation methods. In this approach, economic weights are derived on the basis of the desired genetic gains or simplified economic values calculated by using few input parameters. These modules provide a comprehensive solution for pig selection index formulation.

To ensure an enhanced user experience, the software is built using Electron, Vue.js and Python, enabling excellent cross-platform capabilities. CSIP streamlines the operation into a straightforward 'input–process–output' process (Fig. 3B). Each analytical module is guided through a step-by-step workflow, effectively eliminating the need for traditional command line interfaces and complex file parsing procedures. Demonstrating high computational efficiency on standard personal computers (Intel Core Ultra 7 265K processor, 32 GB RAM), CSIP completes economic weights calculation in seconds and executes stochastic validation (simulating a typical 3,000-head nucleus herd over 10 generations) in approximately 8 minutes. Methodologically, the software covers diverse breeding structures and enhances the precision of economic weight calculations by synergistically integrating a gene flow approach for calculating NDEs directly within specialized breeding system bio-economic models and facilitates the computation of economic weights for sensitivity analysis, relative economic weights and standardized selection index. Furthermore, CSIP features an innovative module for the prediction of multi-generational selection response and is proactively designed to mitigate antagonistic trait response and suboptimal genetic gain arising from poorly designed breeding programs. Overall, this integrated platform represents the first full-process visualized software for the design and formulation of the pig selection index.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Results of the economic weight, selection index and sensitivity analysis for the farrow-to-finish herd

The relative economic weights are shown in Fig. 4A, where F-FCR was prioritized as the most critical trait (38.23%), followed by NBA (28.88%), ATW (17.11%), BFD (11.19%), and LMD (4.59%). In the current baseline scenario (designed selection index), economic weights were calculated to be \$0.53 per day for ATW, \$37.37 per unit for F-FCR, \$1.56 per piglet for NBA, \$1.07 per millimeter for BFD, and \$0.25 per millimeter for LMD (Supplementary Table S7). For the sensitivity analyses, when feed prices increased within a certain range, economic weights increased for ATW, F-FCR, and NBA (Fig. 4B-D); when the non-feed price increased, only ATW and NBA presented higher economic weights, whereas other traits remained unchanged (Fig. 4B, D); when the sale price increased, higher sale prices increased the importance of BFD and LMD (Fig. 4E, F) while decreasing the weight of NBA (Fig. 4D). These trends demonstrate the sensitivity of economic weights to market fluctuations, providing actionable insights for dynamic breeding program adjustments. The standardized selection index was scaled to a mean of 100 and a standard deviation of 25 and is expressed as follows:

$$H = 100 - 101.75 \times EBV_{F-FCR} - 1.44 \times EBV_{ATW} + 8.50 \times EBV_{NBA} - 2.91 \times EBV_{BFD} + 0.68 \times EBV_{LMD}$$

where EBV represents the estimated breeding value for the corresponding trait. This standardized selection index can be directly applied in actual production, significantly lowering the threshold for formulating the selection index.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

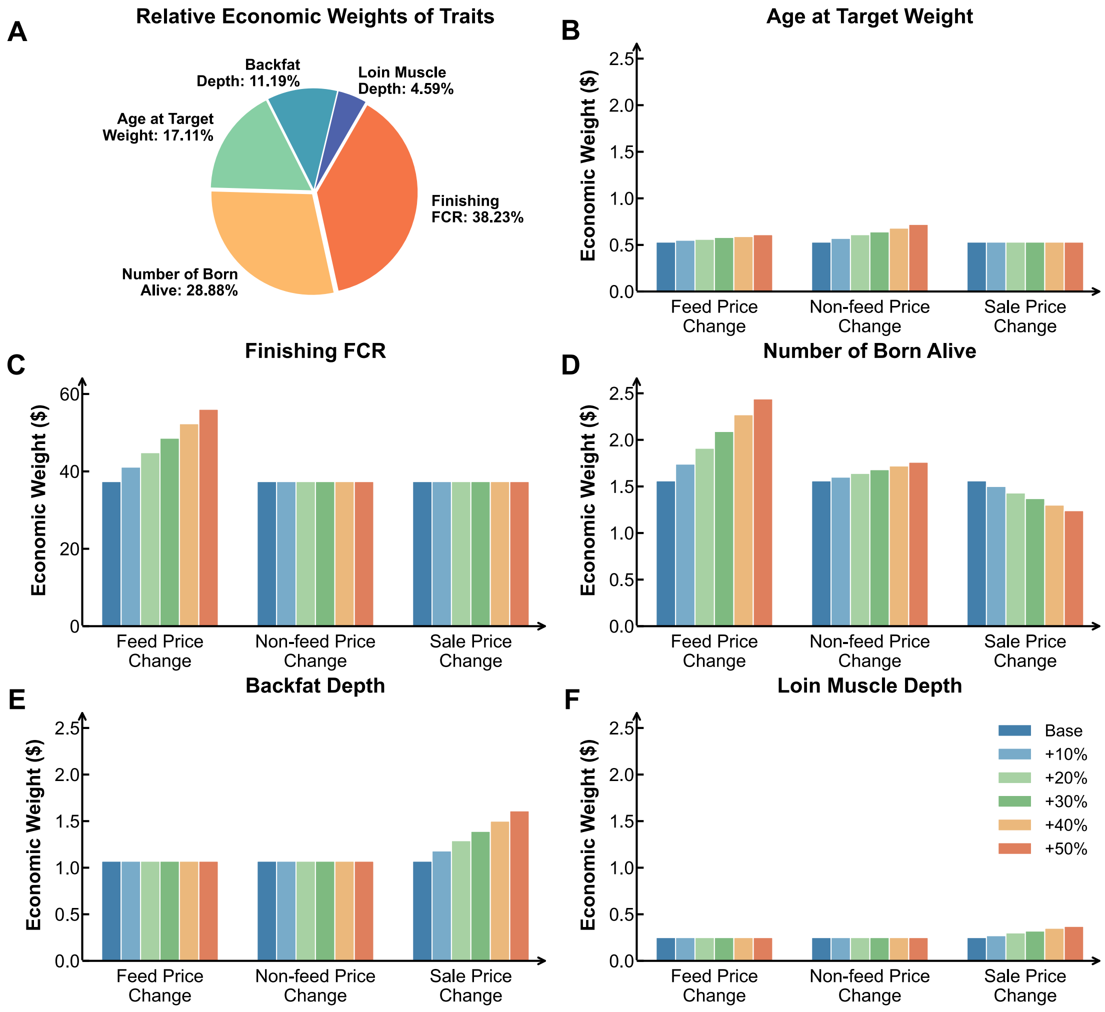

Figure 4. Relative economic importance and sensitivity analysis of traits in the farrow-to-finish system. (A) Pie chart showing the relative economic weights of the five evaluated traits. (B–F) Sensitivity analysis of economic weights for (B) Age at Target Weight, (C) Finishing FCR, (D) Number of Born Alive, (E) Backfat Depth, and (F) Loin Muscle Depth. The "Base" scenario represents the reference economic weights derived for the originally designed selection index (under current market parameters), while the gradient bars illustrate fluctuations resulting from stepwise increases (+10% to +50%) in feed prices, non-feed prices, and sale prices.

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Results of multi-generational selection response under the designed selection index and price variation scenarios for the farrow-to-finish herd

To assess the impact of economic fluctuations on economic weights and selection response, relative economic weights were calculated in the designed selection index scenario and scenarios with the price increases of 10%, 20% and 40% (Supplementary Table S8), and the corresponding genetic gains were subsequently computed (Supplementary Fig. S3-5). The results revealed that ATW, F-FCR and BFD tended to decrease, whereas NBA and LMD tended to increase (Fig. 5A–E). Market fluctuations had a limited effect on the selection response, as the trends of different lines are similar with different price variations. Although the genetic gains of certain traits increased slightly in the fluctuating price scenarios, others showed marginal decreases compared with the designed selection index. The overall trend of the selection index values remained largely consistent across scenarios with rapid increases in the early generations (generations 1–3), followed by a gradual plateau in generations 7–10, but the design selection index was the highest selection index in the later stage (Fig. 5F), indicating the robustness and stability of the formulated selection index in the face of moderate economic variability.

Results of the tenth-generation multi-generational selection response obtained with different selection strategies in the farrow-to-finish herd

To evaluate the effectiveness of the formulated selection index, multi-generational selection response simulations were performed under three selection strategies: best selection, random selection, and a worst-case selection scenario. The results from the 10th generation (Fig. 6)

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

demonstrated that the index formulated by the CSIP software (the best selection scenario) significantly improved the genetic value of the target traits and achieved high overall genetic gains in the desired direction. The worst selection strategy led to a pronounced decline in performance for all the traits. These results highlight the effectiveness of the software-formulated selection index in achieving targeted genetic improvement, underscoring its potential to increase both genetic gain and overall performance in pig breeding programs.

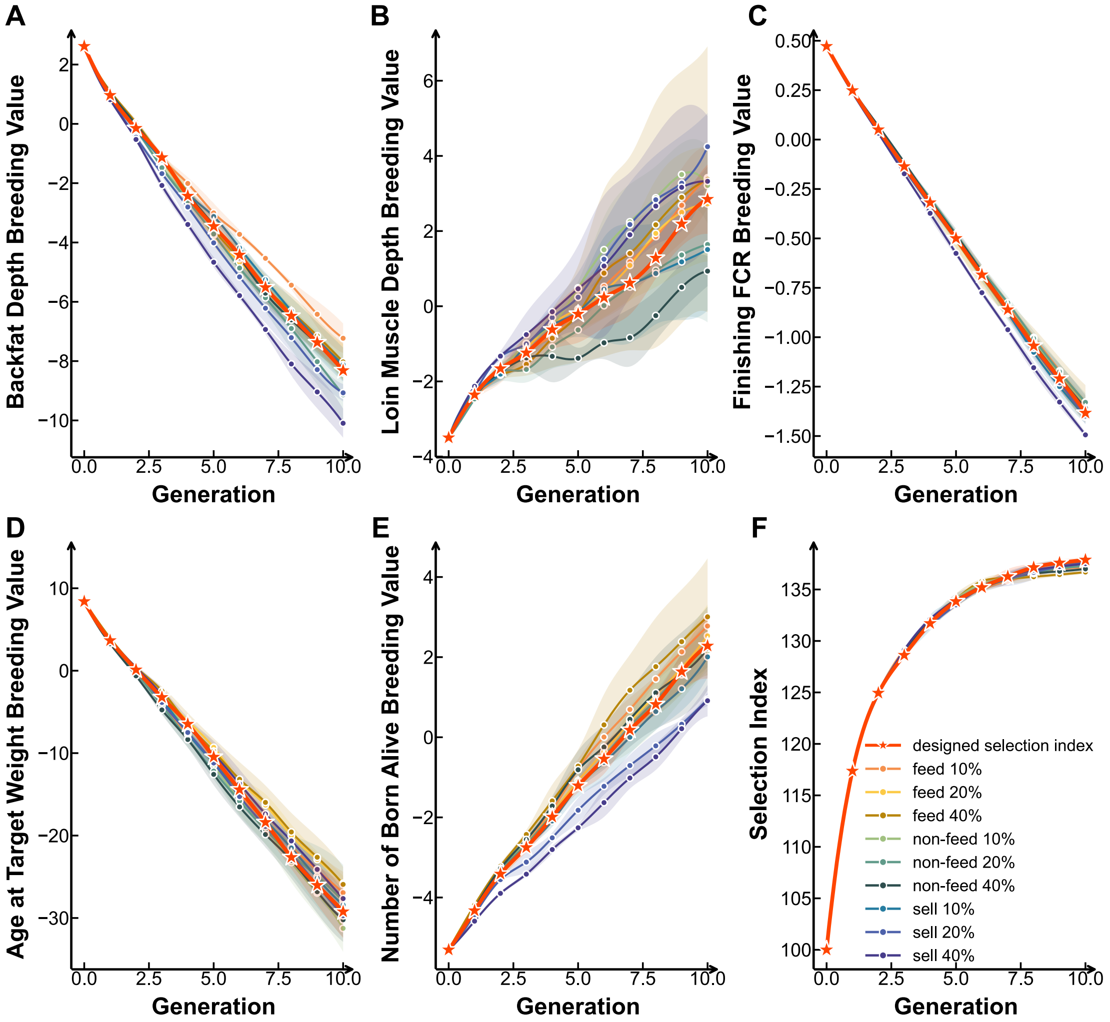

Figure 5. Predicted multi-generational genetic gain and selection index trends under varying economic scenarios. (A–E) Cumulative genetic gain for traits. The trajectories display the True

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Breeding Values (TBVs) over 10 generations for (A) Backfat Depth, (B) Loin Muscle Depth, (C) Finishing FCR, (D) Age at Target Weight, and (E) Number of Born Alive. Note that the occurrence of negative TBVs is inherent to the simulation design, as multi-trait SNP effects were initially sampled from a distribution with a mean of zero. (F) Trend of the Selection Index. This panel illustrates the overall response of the aggregate breeding objective. Scenarios: The orange line with star markers represents the baseline response using the designed selection index (under current economic parameters). The other colored lines represent sensitivity scenarios with varying price increments. Shaded regions indicate the standard deviation of the simulation.

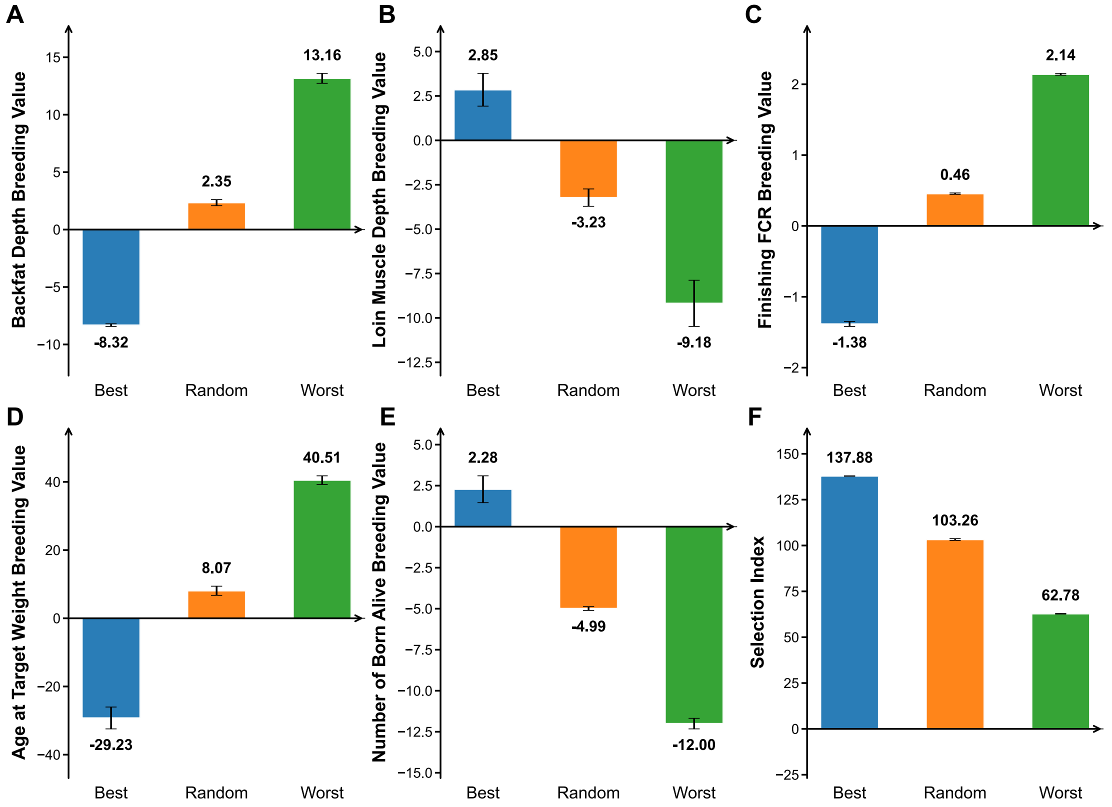

Figure 6. Comparative analysis of tenth-generation genetic response and selection index values under contrasting selection strategies. (A–E) Cumulative True Breeding Values (TBVs) for target traits. The bar charts display the final genetic status at generation 10 for (A) Backfat

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Depth, (B) Loin Muscle Depth, (C) Finishing FCR, (D) Age at Target Weight, and (E) Number of Born Alive. (F) Selection Index. Strategies: The analysis contrasts three selection strategies: Best (selection of top-ranking candidates), Random (random mating), and Worst (selection of lowest-ranking candidates). Note: Consistent with the simulation design (initial SNP effects sampled with a mean of zero), negative TBVs for Backfat Depth, Finishing FCR, and Age at Target Weight in the "Best" group indicate a favorable genetic improvement. Conversely, positive values for Loin Muscle Depth and Number of Born Alive represent favorable gains. Error bars indicate the standard deviation of the simulation.

Results of the multi-generational selection response and economic gain obtained with different economic weights in the farrow-to-finish herd

To evaluate the additional economic benefits of the designed selection index, simulations were conducted in several comparative scenarios, including a designed selection index, selection based on a single trait alone, and uniform relative economic weights assigned to all the traits. The results demonstrated that the targeted trait presented a higher genetic gain in single-trait selection than in the other scenarios (Fig. 7A–E); however, the genetic trends of the nontarget traits either slightly changed, fluctuated, or even were negative, indicating undesirable correlated responses. In contrast, the other scenarios achieved genetic gains in the expected direction across all the traits, although the magnitude of response varied due to differences in economic weights (Fig. 7A–E). In terms of additional economic gain per animal, the single-trait selection scenarios resulted in lower profits than the other strategies did (Fig. 7F). Notably, the designed selection index produced the highest cumulative profit in the later generations

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0

across all the evaluated strategies (Fig. 7F), underscoring its effectiveness in achieving balanced genetic improvement and maximizing multi-generational economic return in pig breeding programs.

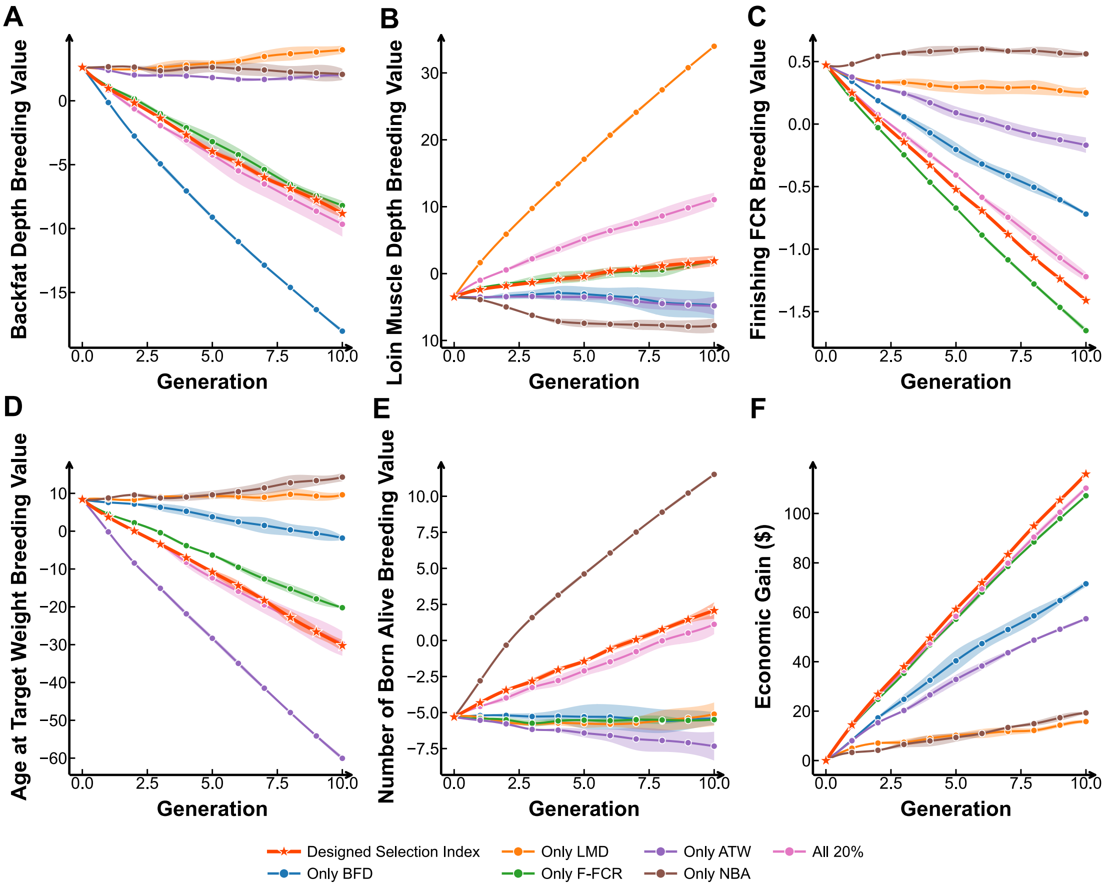

Figure 7. Comparative evaluation of multi-generational genetic response and economic gain under distinct selection strategies. (A–E) Trajectories of True Breeding Values (TBVs). The plots illustrate the genetic trends over 10 generations for (A) Backfat Depth, (B) Loin Muscle Depth, (C) Finishing FCR, (D) Age at Target Weight, and (E) Number of Born Alive. Note: Consistent with the simulation design, negative trends for Backfat Depth, Finishing FCR, and Age at Target Weight indicate favorable genetic improvement. (F) Economic Gain. This panel quantifies the additional monetary benefit per animal, derived from the change in the additive breeding values of the five target traits (assuming other non-target traits remain constant).

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Scenarios: The "Designed Selection Index" (orange line with star markers) represents the breeding objective set by the software. "Only [Trait]" (e.g., "Only BFD") refers to single-trait selection strategies. The "All 20%" scenario represents a balanced index strategy where equal relative economic weights (20%) are assigned to each of the five target traits. Shaded regions indicate the standard deviation of the simulation.

Results of the economic weights, relative economic weights and multi-generational selection response in the three-way crossbreeding system

The economic weights of all traits supported by the software were calculated using a three-way crossbreeding model for three different breeds. The results (Fig. 8A, C, E) indicate that there are notable differences in the economic weights of the same traits between breeds. These variations are attributed to differences in population size and the roles of the breeds within the crossbreeding system. This finding also suggests that considering only the economic values of traits within the commercial herd, without accounting for the structure of the breeding system, may lead to certain inefficiencies. Four production traits and four reproductive traits were subsequently selected from these economic weights to compute the relative economic weights and multi-generational selection response for the three breeds. Breed C, serving as the terminal sire, was considered for production traits only. The results show that relative economic weights differ across breeds within the same system (Fig. 8B, D, F). Furthermore, improper trait selection may result in a slow selection response for some traits and even lead to selection response in the opposite direction of the desired outcome (Supplementary Fig. S6). Therefore,

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

predicting multi-generational selection response while formulating a selection index is crucial to ensure effective genetic improvement.

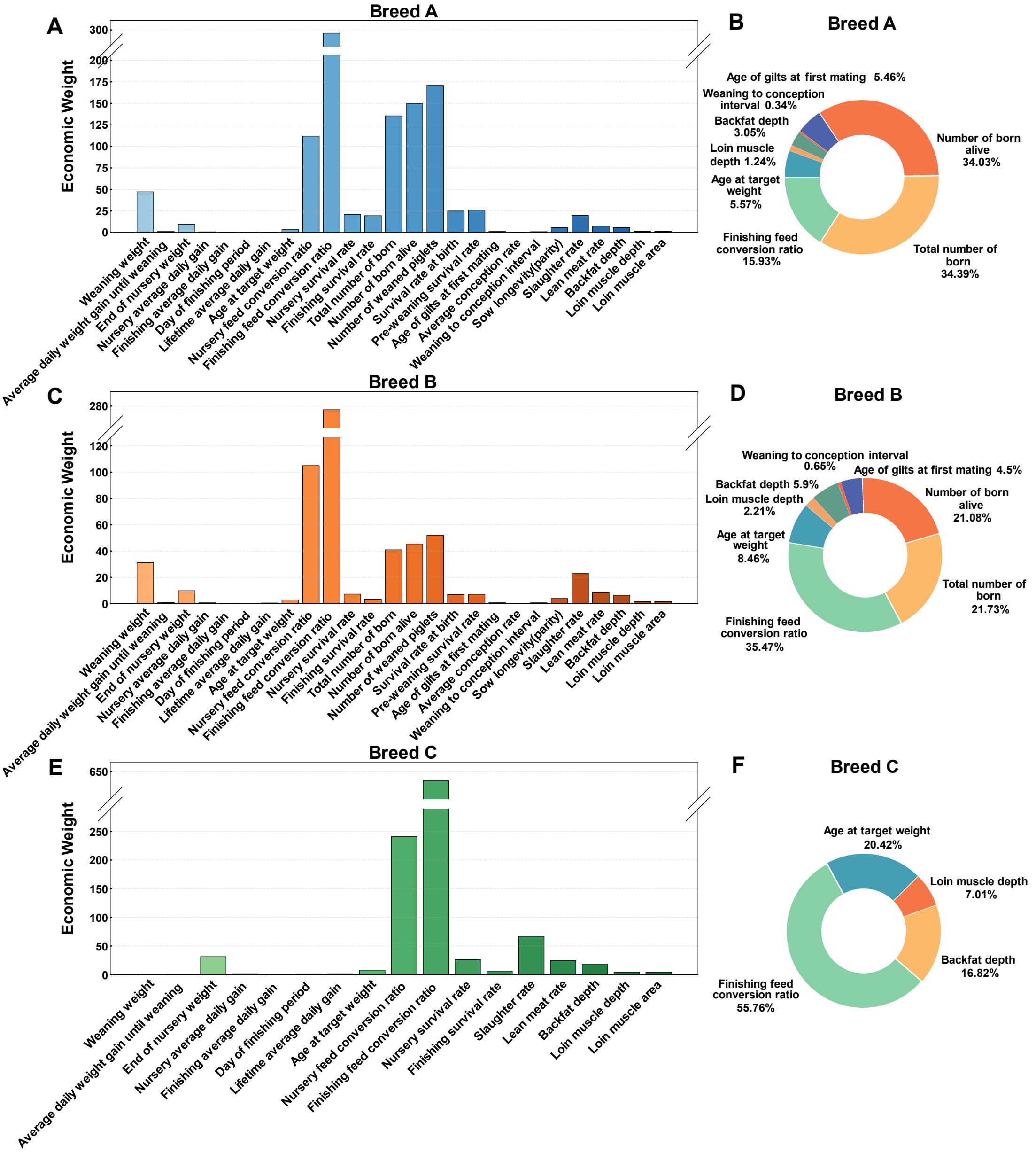

Figure 8. Economic weights and relative economic importance of breeding objective traits for the three-way crossbreeding system. The bar charts display the calculated economic weights for traits in (A) Breed A, (C) Breed B, and (E) Breed C. The pie charts illustrate the proportional

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

contribution of each trait to the overall breeding objective for (B) Breed A, (D) Breed B, and (F) Breed C.

Discussion

In this study, CSIP, software for calculating the pig selection index and predicting multi-generational selection response, was developed. The software integrates multiple bio-economic models to address the requirements for formulating a pig selection index and innovatively integrates a simulated selection process to predict multi-generational selection response, enabling users to adjust trait weights on the basis of predicted genetic trends. Furthermore, CSIP provides an intuitive workflow-driven interface lowers the learning barrier, facilitates rapid adoption by interdisciplinary users, and advances data-driven selection index design in swine breeding.

In developing bio-economic models for diverse swine breeding systems, our approach incorporates artificial insemination technology as a foundational element. This reflects standard practices in global breeding operations where large-scale breeding and disease control are paramount. To model the growth process of pigs, we utilize fundamental feed intake and price parameters to calculate production costs rather than implementing energy/protein deposition models. While some studies [31,35-37] have integrated detailed growth models into bio-economic frameworks—enabling precise nutritional requirement calculations—such approaches necessitate the development of specific growth models tailored to particular breeds and production conditions. This requirement can be challenging for the majority of farms,

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

thereby constraining the model's applicability. Consequently, our simplified method enhances practical adoption while maintaining economic relevance.

The multi-generational selection response quantifies the expected genetic gain of a breeding program and serves as a critical metric for evaluating its design efficacy [38-40]. To model this response, we implemented a simulation framework that replicates multi-generational selection. This framework calculates TBVs on the basis of population genetic parameters, computes selection index values for all candidates, ranks individuals by these values, and selects the top performers. The process iterates over 10 simulated generations, projecting cumulative genetic change. This simulation-based approach—notably absent in conventional selection index methods—addresses a crucial practical gap in breeding program design. Integrating this predictive capability during index development increases formulation robustness by mitigating suboptimal genetic gain arising from antagonistic trait relationships [30].

Consequently, this framework offers a validated bio-economic basis for broader breeding scheme design. While established deterministic software (e.g., ZPLAN, SelAction) is widely utilized for optimizing population structures, these tools require reliable economic parameters as inputs. By internally calculating and verifying these weights through the simulation of cumulative genetic gain, our system ensures that the inputs for such deterministic optimizations are biologically sound. This enables a synergistic workflow where the breeding objective is first established, followed by structural refinement using deterministic tools.

In future developments, CSIP aims to transition towards a modular, plugin-based architecture, enabling the integration of user-defined algorithms and customizable components [41]. Concurrently, the software will be optimized to expand trait coverage and implement AI-driven

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

weight recommendations, particularly for non-economic traits. By combining architectural flexibility with intelligent modeling, CSIP bridges the gap between theoretical frameworks and practical application, thereby advancing precision livestock breeding.

Data availability

The software of CSIP can be research freely used at https://txiang-lab.github.io/CSIP.

Funding

This work was supported by the National Natural Science Foundation of China [NO. 32372840]; the Hubei Provincial Key Research and Development Program Funding [NO. 2023BBB175]; the Hubei Agricultural Core Technology Research Program [NO. HBNYHXGG2023-9-4]; the Hubei Modern Swine Breeding Technology R&D Program [NO. HBZY2023B006-01]; and the China Agriculture Research System of MOF and MARA [CARS-35].

Conflicts of interest

The authors declare no competing interests.

References

1. Kodak O, István N. Historical overview of the selection indices applied in pig breeding. Acta Agraria K 2019;23.

2. Foroutanifar S. Comparison of long-term effects of genomic selection index and genomic selection using different Bayesian methods. Livest Sci 2020;241.

3. Smith HF. A discriminant function for plant selection. Ann Eugen 1936;7(3):240–50.

4. Hazel LN. The genetic basis for constructing selection indexes. Genetics

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

1943;28(6):476–90.

5. Kosinska-Selbi B, Schmidtmann C, Ettema JF. *et al.* Breeding goals for conservation and active Polish dairy cattle breeds derived with a bio-economic model. Livest Sci 2022;255.

6. Hazel LN, Dickerson GE, Freeman AE. The Selection Index—Then, Now, and for the Future1. J Dairy Sci 1994;77(10):3236–51.

7. Wellmann R. Selection index theory for populations under directional and stabilizing selection. Genet Sel Evol 2023;55(1).

8. Vries AD. A model to estimate economic values of traits in pig breeding. Livest Prod Sci 1989;21:49–66.

9. Stewart TS, Bache DH, Harris DL. *et al.* A bioeconomic profit function for swine production: application to developing optimal multitrait selection indexes. J Anim Breed Genet 1990;107(1-6):340–50.

10. Quinton VM, Wilton JW, Robinson JAB. *et al.* Economic weights for sow productivity traits in nucleus pig populations. Livest Sci 2006;99:69–77.

11. Thekkoot DM, Kemp RA, Dekkers JCM. 481. A bio-economic model to estimate the economic value of sow feed intake during lactation. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) 2022. p. 1999–2002.

12. Krupa E, Krupová Z, Wolfová M. *et al.* Estimation of economic values for traits of pig breeds in different breeding systems: II. Model application to a three-way crossing system. Livest Sci 2017;205:70–8.

13. González-Peña D, Knox RV, MacNeil MD. *et al.* Genetic gain and economic values of selection strategies including semen traits in three- and four-way crossbreeding systems for

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

swine production. J Anim Sci 2015;93(3):879–91.

14. Houška L, Wolfová M, Nagy I. *et al.* Economic values for traits of pigs in Hungary. Czech J Anim Sci 2010;55(4):139–48.

15. Houška L, Wolfová M, Fiedler J. Economic weights for production and reproduction traits of pigs in the Czech Republic. Livest Prod Sci 2004;85(2):209–21.

16. Brascamp EW, Smith C, Guy DR. Derivation of economic weights from profit equations. Anim Prod 1985;40:175–9.

17. Hovenier R, Brascamp EW, Kanis E. *et al*. Economic values of optimum traits: the example of meat quality in pigs. J Anim Sci 1993;71 6:1429–33.

18. Tess MW, Bennett GL, Dickerson GE. Simulation of Genetic Changes in Life Cycle Efficiency of Pork Production. III. Effects of Management Systems and Feed Prices on Importance of Genetic Components. J Anim Sci 1983;56:369–79.

19. Tess MW, Bennett GL, Dickerson GE. Simulation of Genetic Changes in Life Cycle Efficiency of Pork Production, II. Effects of Components on Efficiency. J Anim Sci 1983;56:354–68.

20. Tess MW, Bennett GL, Dickerson GE. Simulation of Genetic Changes in Life Cycle Efficiency of Pork Production. I. A Bioeconomic Model. J Anim Sci 1983;56:336–53.

21. McClintock AE, Cunningham EP. Selection in dual purpose cattle populations: defining the breeding objective. Anim prod 1974;18:237–47.

22. Jiang X, Groen AF, Brascamp EW. Discounted expressions of traits in broiler breeding programs. Poultry Sci 1999;78(3):307–16.

23. Wolfová M, Krupa E, Krupová Z. *et al.* Economic weights of maternal and direct traits

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0
https://creativecommons.org/licenses/by-nc-nd/4.0

of pigs calculated by applying gene flow methods. Animal 2019;136:1127–36.

24. Danell Ö, Rönningen K, Ström H. *et al*. An Extension of the Discounted Gene Flow Method with Example in Pig Breeding. Acta Agr Scand 1976;26:203–10.

25. Hill WG. Prediction and evaluation of response to selection with overlapping generations. Anim prod 1974;18:117–39.

26. Täubert H, Reinhardt F, Simianer H. ZPLAN+ A new software to evaluate and optimize animal breeding programs. proceedings of the 9th World Congress on Genetics Applied to Livestock Production (WCGALP) 2010; Germany.

27. Lopez BM, Kang HS, Kim TH, *et al*. Optimization of Swine Breeding Programs Using Genomic Selection with ZPLAN. Asian-Australas J Anim Sci 2016;29(5):640–5.

28. Huang Q, Zhou L, Xue Y, *et al.* GOplan: an R package for animal breeding program design via integrating Gene Flow and Bayesian optimization methods. G3 2025;15(2).

29. Wolfová M, Wolf J, Krupová Z. *et al*. Estimation of economic values for traits of pig breeds in different breeding systems: I. Model development. Livest Sci 2017;205:79–87.

30. Mancin E, Mantovani R, Tuliozi B. *et al*. Economic weights for restriction of selection index as optimal strategy for combining multiple traits. J Dairy Sci 2022;105(12):9751–62.

31. De Vries AG. A model to estimate economic values of traits in pig breeding. Livest Prod Sci 1989;21(1):49–66.

32. Reinsch N, Dempfle L. Investigations on functional traits in Simmental: 3. Economic weights at the stationary state of a Markov chain. Archiv fur Tierzucht 1998;41(3):211–24.

33. Azzam SM, Azzam AM, Nielsen MK. *et al.* Markov Chains as a Shortcut Method to Estimate Age Distributions in Herds of Beef Cattle Under Different Culling Strategies. J Anim

Langtaosha (LTS) Preprint doi: https://doi.org/10.65215/LTSpreprints.2026.04.26.000203. This version posted April 27, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. Creative Commons license: CC Attribution-NonCommercial-NoDerivatives 4.0 https://creativecommons.org/licenses/by-nc-nd/4.0

Sci 1990;68(1):5–14.

34. Xiang T, Christensen OF, Vitezica ZG. *et al.* Genomic Model with Correlation Between Additive and Dominance Effects. Genetics 2018;209:711 – 23.

35. Hermesch S, Kanis E, Eissen JJ. Economic weights for feed intake in the growing pig derived from a growth model and an economic model. J Anim Sci 2003;81(4):895–903.

36. Lange CFMd, Marty BJ, Birkett S. *et al.* Application of pig growth models in commercial pork production. Canadian J Anim Sci 2001;81:1–8.

37. Schinckel AP, de Lange CFM. Characterization of growth parameters needed as inputs for pig growth models. J Anim Sci 1996;74 8:2021–36.

38. Hsu WL, Johnson RK. Analysis of 28 generations of selection for reproduction, growth, and carcass traits in swine. J Anim Sci 2014;92(11):4806–22.

39. De Beukelaer H, Badke Y, Fack V. *et al.* Moving Beyond Managing Realized Genomic Relationship in Long-Term Genomic Selection. Genetics 2017;206(2):1127-1138.

40. Zhang Z, Wang L. A look-ahead approach to maximizing present value of genetic gains in genomic selection. G3 2022;12(8):jkac136.

41. Zhang Z, Wang L. A simulation framework for reciprocal recurrent selection-based hybrid breeding under transparent and opaque simulators. Front Plant Sci. 2023 Jun 27;14:1174168.