Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide association mapping of eye size in Penaeus vannamei

Mingxuan Teng
Ocean University of China
Zewei Hu
Ocean University of China
Jie Ma
Ocean University of China
Mingyang Zhao
Ocean University of China
Kaixiang Gao
Ocean University of China
Hao Wang
Ocean University of China
Mengran Liu
Ocean University of China
Qifan Zeng
zengqifan@ouc.edu.cn

Ocean University of China

Research Article

Keywords: Penaeus vannamei, eye morphology, genome-wide association study, SNP association, positional annotation

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11540288/v1

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

Additional Declarations: No competing interests reported.

Genome-wide association mapping of eye size in *Penaeus vannamei*

Mingxuan Teng[a,#], Zewei Hu[a,#], Jie Ma[a], Mingyang Zhao[a], Kaixiang Gao[a], Hao Wang[a], Mengran Liu[a,b,*], Qifan Zeng[a,b,c,*]

[a] Hainan Key Laboratory of Tropical Aquatic Germplasm (Hainan Seed Industry Laboratory), Sanya Oceanographic Institution, Ocean University of China, Sanya 572024, China

[b] MOE Key Laboratory of Marine Genetics and Breeding, Shandong Key Laboratory of Marine Seed Industry (preparatory), and Qingdao Institute of Maritime Silk Road (Qingdao Institute of Blue Seed Industry), Ocean University of China, Qingdao 266003, China

[c] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

[#] These authors contributed equally to this work.

* Corresponding authors: Qifan Zeng (E-mail: zengqifan@ouc.edu.cn), Mengran Liu (E-mail: lmr@ouc.edu.cn)

Abstract

Eye morphology can be readily quantified in *Penaeus vannamei*, yet the genetic basis of eye-size variation remains poorly understood. We conducted a genome-wide association study in 350 individuals using low-coverage whole-genome resequencing and approximately 12.67 million imputed SNPs. Five correlated eye measurements were combined into a composite phenotype using the first principal component, which explained 99.957% of their total variance. Linear mixed models accounting for sex, population structure and genomic relatedness identified associations exceeding permutation-derived genome-wide significance thresholds in two intervals: chromosome 3 at 8.28-9.05 Mb and chromosome 6 at 12.59-12.86 Mb. After body length was included as an additional covariate, associations in the chromosome 3 interval remained genome-wide significant, whereas the chromosome 6 signals were substantially attenuated. An additional signal at chromosome 3:30,808,740 approached, but did not reach, the adjusted genome-wide threshold. Mediation analyses suggested differing contributions of body-length-associated variation across loci, without establishing causal pathways. Positional annotation identified *SLC22*, *Jus*, *BBIP1*, *AHR* and *G2B* near the leading SNPs. Haplotype analysis revealed differences in eye size among haplotype groups, with the strongest evidence in the chromosome 3 interval near 9 Mb. These findings prioritize genomic regions for independent replication and functional investigation and highlight the importance of accounting for body size when analysing eye morphology.

Keywords: *Penaeus vannamei*, eye morphology, genome-wide association study, SNP association, positional annotation

1. Introduction

Penaeus vannamei (Pacific white shrimp) is the most widely cultivated shrimp species globally, with an annual production reaching 6.8 million tonnes in 2022 (FAO 2024). Genetic improvement of economically important traits therefore matters for the sustainable development of the shrimp aquaculture industry. Among these traits, harvest body weight has low-to-moderate heritability and has potential for genetic improvement. A genetic evaluation across two generations and 24,072 individuals reported a heritability of 0.335 ± 0.087 and a selection response of 2.30% for harvest body weight (Sui et al. 2016). Body length and body weight are commonly used as indicators of growth performance, but repeated measurements require animal handling and can be affected by body curvature, appendage loss, and variation in posture. Image-based phenotyping offers a potential means of obtaining non-invasive and standardized morphological measurements. Eye dimensions are particularly suitable for image analysis because the eye contour is readily identifiable under controlled imaging conditions. Chen et al. reported a power-law relationship between eye axis length and body weight in *P. vannamei*, suggesting that eye measurements may be useful for image-based estimation of shrimp growth (Chen et al. 2019). However, this evidence was phenotypic, and the genetic basis of interindividual variation in eye size remains poorly characterized.

Eye morphology is also of biological interest because crustacean compound eyes are important sensory structures. Genomic studies in *P. vannamei* have identified expansions of rhodopsin and other vision-related gene families, which have been discussed in the context of adaptation to benthic environments (Zhang et al. 2019). These species-level genomic observations provide a broad biological context for studying eye-related traits but do not establish which genetic variants contribute to variation in eye size among individuals. Characterizing SNP associations with eye morphology may therefore provide a starting point for investigating this source of phenotypic variation.

Genome-wide association studies provide a marker-based approach for identifying genomic loci statistically associated with quantitative traits (Yu et al. 2025). High-density SNP data can facilitate the detection of both coding and non-coding association signals (Yu et al. 2025; Peña-Martínez and Rodríguez-Martínez 2024). In *P. vannamei*, genome-wide association study (GWAS) and related association analyses have been used to investigate growth traits and have prioritized SNPs and genomic regions associated with body weight and other growth measurements (Wang et al. 2019; Wang et al. 2020; Zhang et al. 2025; Lv et al. 2025). Nevertheless, no genome-wide association analysis of eye-size traits has yet been reported in this species.

In this study, we conducted an exploratory GWAS of five eye-size measurements in *P. vannamei*: eye area, fitted ellipse area, major axis length, minor axis length, and perimeter. We first estimated phenotypic and genetic correlations among the eye-size traits and between eye size and body length and weight. We then identified SNPs exceeding the study-defined genome-wide or suggestive thresholds and performed an additional GWAS with body length as a covariate to assess the sensitivity of the eye-trait signals to adjustment for overall body size. Genes overlapping the associated regions were reported only as

supplementary positional information. The primary objective was to prioritize SNPs and genomic regions for subsequent replication, rather than to infer causal genes or propose markers for immediate use in selection.

2. Materials and methods

Ethics approval Not applicable; the study was carried out on an invertebrate species (Penaeus vannamei), for which no specific ethical approval is required.

2.1 Experimental animals and phenotype measurement

The *Penaeus vannamei* broodstock used in this study were obtained from Lanyin Aquatic Technology Co., Ltd. (Wenchang, Hainan Province, China). Before sampling, all shrimp were maintained at the same facility under common rearing conditions. A total of 350 individuals were selected for phenotyping (body length, 197.58 ± 22.86 mm; body weight, 49.19 ± 14.83 g). The recorded variables were eye area (Area), fitted ellipse area (O-Area), major axis length (Feret), minor axis length (MinFeret), eye perimeter (Perim) (Fig. 1a), body length (BL), body weight (BW), and sex.

BL and BW were quantified using a proprietary, non-invasive, high-throughput multiview image-segmentation method (patent no. ZL202210298060.0). Three imaging replicates were obtained for each individual, and the arithmetic mean was used in subsequent analyses. High-resolution morphological images were captured at a constant magnification for measurement of the eye traits.

Following previously described procedures (Ferreira and Rasband 2012), the acquired images were imported into ImageJ v1.54 image analysis software for scale calibration and morphological measurement. To calibrate pixel-to-physical dimensions, a 1-cm standard ruler was photographed under the same fixed conditions (including illumination, camera focal length, and shooting distance) as the shrimp samples. The ruler image was imported into ImageJ, and a straight line was drawn along the 1-cm mark using the Straight Line tool. The pixel-to-physical scale was then calibrated via the "Analyze/Set Scale" command by entering the known length (1 cm) and the corresponding pixel distance, thereby establishing the correspondence between image pixels and actual physical length for all images acquired under identical conditions. After calibration, the complete contour of the shrimp eyeball was manually delineated using the Freehand Selection tool (Collins 2007). The area and perimeter of the delineated contour were then recorded (Fig. 1a). Subsequently, the "Fit Ellipse" function (Schneider et al. 2012) was applied to the contour to obtain the major axis length of the fitted ellipse and the fitted ellipse area. Each sample was measured three times, and the average value was used in subsequent analyses. To improve normality and stabilize variance, the square roots of the eye area (Area) and fitted ellipse area (O-Area) were calculated. Principal component analysis (PCA) was then performed on the five eye-size-related phenotypes, and the first principal component (PC1) was extracted as the composite eye-size phenotype. The PC1 values were subsequently rank-based inverse normal transformed to generate the final eye-size metric, and body length (BL) and body weight (BW) were normalized in the same way before the downstream analyses. Sex was assigned based on visual inspection of external reproductive organs at the time of measurement.

2.2 Library construction and whole-genome resequencing

We collected antennae from each shrimp and immediately preserved the tissue in liquid nitrogen.

Genomic DNA was extracted using the phenol-chloroform method (Sambrook and Russell 2001) and sheared with a Covaris ultrasonicator to 300–400 bp fragments. After purification and size selection using AMPure XP beads, sequencing libraries were prepared following standard procedures (end-repair, A-tailing, and adapter ligation). Library concentration and fragment-size distribution were assessed using a Qubit fluorometer and an Agilent 2100 Bioanalyzer. The libraries were then subjected to rolling-circle amplification and high-throughput sequencing on the DNBSEQ platform.

2.3 Sequencing data processing and genotype imputation

Raw sequencing reads from each sample were subjected to quality control using fastp to remove low-quality bases and sequences (Chen 2025). After quality control, the reads were aligned to the reference genome using the Sentieon BWA module, producing SAM files (Kendig et al. 2019; Freed et al. 2017; Liao et al. 2025). The resulting SAM files were converted to BAM format, sorted, and indexed using the Sentieon utils sort tool. Duplicate reads introduced during PCR amplification were marked and removed with the Sentieon util dedup module. Variant discovery was performed using the Sentieon Genotype module, with all-mode configuration applied to call variants across all covered positions. Variant calls were filtered using a minimum quality threshold of 30 for both variant calling (--call_conf=30) and variant emission (--emit_conf=30). A reference haplotype database was used during joint genotyping.

Genotype imputation was performed using GLIMPSE (Rubinacci et al. 2021), with a haplotype reference panel constructed from high-depth sequencing data for *Penaeus vannamei* (Zeng et al. 2022) and the chromosome-level reference genome. Genotype likelihoods were imputed, and low-confidence genotypes were set to missing using a posterior probability threshold of ≤0.9. The imputed genotype dataset was refined with Beagle 5 (Browning et al. 2018) to improve consistency. After imputation, rare variants with a minor allele frequency (MAF) <0.05 were excluded from downstream analyses.

To evaluate imputation performance, 91 samples were selected for 10× high-depth whole-genome sequencing (WGS). The imputed genotypes were subsequently compared against the high-depth sequencing data to assess imputation accuracy. Two metrics were used to evaluate the imputation performance: genotype concordance (GC) and allele concordance (AC).

Genotype concordance measures the proportion of perfectly matching genotypes at individual loci, calculated as:

$$Genotype\ Concordance = (N_{g,match} / N_{g,valid}) \times 100\%$$

where $N_{g,match}$ represents the number of loci where the imputed genotype exactly matches the true genotype, and $N_{g,valid}$ denotes the total number of assayed loci.

Allele concordance evaluates the accuracy at the single-allele level by calculating the proportion of matching alleles among all valid alleles:

$$Allele\ Concordance\ =\ (N_{a,match}\ /\ N_{a,total})\ \times\ 100\%$$

where $N_{a,match}$ is the total number of matching alleles between the imputed and true genotypes, and $N_{a,total}$ is the total number of assayed alleles. For diploid organisms, each valid locus contributes two alleles. Therefore, $N_{a,total}\ = 2 \times\ N_{g,valid}$.

To ensure the reliability of the genetic variants for downstream analysis, only SNPs with GC greater than 0.9 were retained for the subsequent genome-wide association study.

2.4 Linkage disequilibrium and population structure analysis

To characterize the extent of linkage disequilibrium (LD) across the genome, pairwise r^2 values between SNPs were estimated using PopLDdecay (Zhang et al. 2019). For population structure assessment, we first performed principal component analysis on the SNP set with PLINK (Purcell et al. 2007). We then used GCTA to construct the genetic relationship matrix (GRM) for all individuals based on the genome-wide SNP data (Yang et al. 2011). A maximum-likelihood phylogenetic tree was also constructed from the SNP data using IQ-TREE (Minh et al. 2020).

2.5 Correlation analysis

Phenotypic correlation was calculated with

$$Cor(p) = \frac{V_{p12}}{\sqrt{V_{p11} \times V_{p22}}}\ ,$$

where V_{p12} is the covariance between the two traits and $V_{p11} \times V_{p22}$ is the product of their phenotypic standard deviations. To calculate genetic correlation, we constructed a GRM with GCTA using genome-wide SNP data to quantify the genetic relatedness between individuals, using the formula:

$$A_{jk} = \frac{1}{N}\sum_{i=1}^{N} \frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}\ ,$$

where A_{jk} is the genetic relationship between individuals j and k, N is the number of SNPs, p_i is the minor allele frequency of SNP i, x_{ij} and x_{ik} are the genotypes of individual j and k at SNP i. Bivariate restricted maximum-likelihood (REML) analysis in GCTA was used to estimate the additive genetic variance for each trait (V_{g11}, V_{g22}) and covariance (V_{g12}) between two traits (Yang et al. 2011). The genetic correlation was then calculated as

$$Cor(g) = \frac{V_{g12}}{\sqrt{V_{g11} \times V_{g22}}}$$

2.6 Genome-wide association analysis

Genome-wide association analyses were conducted using the linear mixed model (LMM) implemented in GEMMA (0.98.5) (Zhou and Stephens 2012). To account for sex and population structure, we included sex and the first ten principal components (PC1–PC10) as fixed covariates in the association model (Price

et al. 2010). The statistical model can be expressed as:

$$y = W\alpha + x\beta + u + \epsilon$$

in which y denotes an $n \times 1$ phenotype vector representing trait measurements across n individuals, W is an $n \times p$ design matrix containing fixed-effect explanatory variables, including SNP genotypes and the specified covariates (e.g., sex), and α represents a $p \times 1$ vector of fixed-effect coefficients. For the random-effect component, x is an $n \times 1$ vector for marker genotypes, β is the effect size of the marker. u is an n-vector of random effects and ϵ is an $n \times 1$ residual vector. The phenotypic variance explained (PVE) for each locus was calculated using the following formula:

$$PVE = \frac{2\beta^2 MAF(1-MAF)}{2\beta^2 MAF(1-MAF) + \left(se(\beta)\right)^2 2NMAF(1-MAF)}$$

where β is the effect size from the GWAS, MAF represents the minor allele frequency, N is the total number of individuals, and $se(\beta)$ is the standard error of the β, calculated as $se(\beta) = |\beta/z|$. The z-score was obtained from the corresponding P-value using the inverse cumulative distribution function of the standard normal distribution, i.e., $z = \Phi^{-1}(1 - P/2)$, $z = \Phi^{-1}(1 - P - value\ /\ 2)$.

Significance thresholds were determined by permutation using permGWAS2 (John et al. 2024), which preserves population structure during permutation and uses a block matrix decomposition to reduce redundant computation. For each trait, 1,000 permutations of the phenotype were performed under the same linear mixed-model specification as the association analysis (identical genotype matrix, sex and PC1–PC10 as fixed covariates, and the same genomic relationship matrix). The genome-wide significance threshold was defined as the 5% quantile of the empirical distribution of the maximum test statistic across all SNPs (family-wise error rate = 0.05), corresponding to $-\log_{10}P = 7.64$ ($P = 2.30 \times 10^{-8}$) for the composite eye-size trait and $-\log_{10}P = 7.71$ ($P = 1.96 \times 10^{-8}$) for the analysis with body length as an additional covariate. A nominal threshold of $P = 1 \times 10^{-6}$ ($-\log_{10}P = 6.0$) was used to report suggestive associations.

2.7 Mediation analysis

To distinguish direct effects of the candidate SNPs on eye size from effects mediated by overall body size, we performed causal mediation analysis with body length (BL) as the mediator (Imai et al. 2010). For each of the six significant SNPs, two linear models were fitted to the same 350 individuals:

$$M_i = \alpha_0 + \alpha_1 X_i + \alpha_2^{\top} W_i + \varepsilon_{M,i}$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 M_i + \beta_3^{\top} W_i + \varepsilon_{Y,i}$$

where X_i is the genotype dosage of the SNP (coded 0, 1 or 2 as the number of copies of the tested allele), M_i is the normalized body length, Y_i is the normalized composite eye-size phenotype, and W_i includes sex and PC1–PC10 as covariates. Both models were fitted by ordinary least squares. The average causal mediation effect (ACME) was computed as $\alpha_1\beta_2$, the average direct effect (ADE)

as β_1, the total effect as $\beta_1 + \alpha_1\beta_2$, and the proportion mediated as ACME/(total effect). Because individuals within the study population are not independent, confidence intervals and P values were obtained by block bootstrap with 2,000 resamples, in which the five genetic clusters defined in the population structure analysis (Section 2.4, Fig. 4) were resampled as blocks instead of individuals, so that the dependence among related individuals within a cluster was preserved. Individual-level resampling was also performed for comparison, using the same number of resamples and the same percentile intervals, so that the two schemes differ only in the resampling unit. Point estimates were verified to match those obtained with the R mediation package (version 4.5.0) (Imai et al. 2010).

2.8 Haplotype analysis

To examine whether the associated regions carry haplotype-level signals, a haplotype analysis was performed for each significant SNP. For each index SNP, SNPs within ±50 kb that were in linkage disequilibrium with it (pairwise $r^2 \geqslant 0.2$, both upstream and downstream) were extracted to define a locus-specific SNP set. The 50-kb window was chosen according to the extent of LD decay in this population (Fig. 4b). Genotypes of the selected SNPs in the 350 individuals were extracted with PLINK v1.9 and converted into haplotypes with geneHapR version 1.2.6 (Zhang et al. 2023). Associations between haplotypes and the normalized composite eye-size phenotype were tested with hapVsPheno, per-site genotype effects across each gene model were estimated with siteEFF and visualized with plotEFF. Associations at individual sites were further examined with hapVsPhenoPerSite.

2.9 Functional annotation of positional candidate genes

Genes overlapping the associated regions were identified from the chromosome-level reference-genome annotation. The protein sequences of the resulting positional candidate genes were functionally annotated with InterProScan 6 against the Pfam, PANTHER, CDD, NCBIFAM, SUPERFAMILY and CATH-Gene3D member databases. Matches at overlapping positions from different member databases were merged into a single domain entry, and entries without a functional description were not reported. The resulting domain architecture of each candidate gene is summarized in Online Resource 6.

3. Results

3.1 Phenotypic distributions and genetic parameters

Descriptive statistics for the five eye-size traits and the two body-size traits (BL and BW), measured in 350 *Penaeus vannamei* individuals, are summarized in Table 1. The Shapiro-Wilk test rejected normality for all raw traits (Online Resource 4), although the distributions and Q-Q plots indicated only modest departures from normality (Fig. 1b–h). Area and fitted-ellipse area (O-Area) were square-root transformed to reduce skewness and stabilize variance, whereas Feret, MinFeret and Perim were used on their original scale.

Phenotypic and genetic correlations among the five eye-size traits were positive and generally high (Table 2): genetic correlations ranged from 0.857 (O-Area with Feret) to 0.999 (Feret with Perim) and phenotypic correlations from 0.761 to 0.967, indicating that the five measurements capture overlapping aspects of eye morphology. Heritability estimates ranged from 0.312 (MinFeret) to 0.493 (Feret). Correlations between the eye-size traits and the two body-size traits were positive but consistently weaker. Genetic correlations were 0.622–0.863 with BL and 0.668–0.877 with BW, and the corresponding phenotypic correlations were 0.636–0.719 and 0.495–0.584 (Table 2).

Because of this redundancy, the five traits were not analysed individually. Principal component analysis was applied to the five eye-size traits without additional standardization (Fig. 2a), and the first principal component, which explained 99.957% of the total variance, was adopted as the composite eye-size phenotype (Online Resource 5). The remaining components explained 0.027%, 0.015%, 0.0002% and approximately 0%, respectively. The first component was normalized to a standard normal distribution by a rank-based inverse normal transformation (Fig. 2b), and BL and BW were normalized in the same way. The normalized composite eye-size phenotype and body-size traits were used in all subsequent analyses.

3.2 Genotyping, genotype imputation and variant calling

The 350 *Penaeus vannamei* individuals were sequenced at an average depth of approximately 1×. In total, 6,950.36 million raw reads were generated (19.86 million reads per individual), and 5,613.21 million clean reads remained after quality filtering, corresponding to 768.85 Gb of sequence data (2.20 Gb per individual). The mean alignment rate to the reference genome was 86.52%.

After retaining variants with a genotype concordance above 0.9 and a minor allele frequency above 0.05, 12,674,256 SNPs were used for the association analysis, with a mean genome-wide density of 5.07 SNPs per kilobase and a transition/transversion ratio of 1.56 (Fig. 4a). Imputation accuracy was evaluated against 91 individuals with high-depth sequencing, where the mean genotype concordance was 96.5% and the mean allele concordance was 97.9%. For the variant set retained for the association analysis, accuracy was summarized over the seven MAF bins spanning 0.05–0.50, genotype concordance was 98.9% in the lowest bin and 97.2% in the highest bin, and allele concordance was

99.4% and 98.5%, respectively (Fig. 3a–c).

3.3 Population genetic structure analysis

The linkage disequilibrium coefficient (r^2) declined to half of its initial value at approximately 16 kb, indicating relatively rapid LD decay in this population (Fig. 4b). The genomic relationship matrix separated the 350 individuals into five genetic clusters (Fig. 4c), containing 61, 77, 73, 55 and 84 individuals, respectively. Clusters 1 and 2 were the most distinct, whereas clusters 3–5 appeared more closely related. Principal component analysis and the maximum-likelihood phylogeny recovered the same clustering (Fig. 4d–f). To account for this structure in the association analyses, sex and the first ten principal components (PC1–PC10) were included as fixed covariates.

3.4 Genome-wide association analysis of the composite eye-size trait

Study-defined multiple-testing control was performed by permutation (permGWAS2, 1,000 permutations), which yielded genome-wide thresholds of $-\log_{10}P = 7.64$ for the composite eye-size trait and $-\log_{10}P = 7.71$ for the analysis with BL as a covariate. Associations with $P < 1 \times 10^{-6}$ ($-\log_{10}P = 6.0$) were considered suggestive (Section 2.6). Three SNPs exceeded the genome-wide threshold (Table 3): 3_9054101 on chromosome 3 ($-\log_{10}P = 8.30$; PVE = 3.9%), 6_12640972 on chromosome 6 ($-\log_{10}P = 8.46$; PVE = 2.5%) and 6_12808861 on chromosome 6 ($-\log_{10}P = 7.97$; PVE = 2.7%). These SNPs defined two associated regions, chromosome 3 at 8.28–9.05 Mb and chromosome 6 at 12.59–12.86 Mb (Fig. 5a). Neither region reached the genome-wide threshold in the separate analyses of BL and BW (Fig. 5b, c).

To assess the sensitivity of these signals to adjustment for overall body size, the composite eye-size GWAS was repeated with BL included as an additional fixed covariate (Fig. 5d). After adjustment, both chromosome 6 signals fell below the genome-wide threshold (6_12640972: $-\log_{10}P = 0.46$; 6_12808861: $-\log_{10}P = 0.24$), whereas the chromosome 3 region remained significant (3_8284439: $-\log_{10}P = 8.00$; 3_9054098: 7.88; 3_9054101: 7.64). An additional signal appeared at 3_30808740, for which $-\log_{10}P$ rose from 2.14 in the primary analysis to 7.70, approaching the adjusted threshold of 7.71.

The six leading SNPs fell into three effect patterns (Table 3). The chromosome 3 SNPs at 8.28–9.05 Mb were significant in both analyses, with negative effects ($\beta = -0.525$ to -0.540) and PVE of 3.3–3.9% before adjustment, and remained significant after adjustment ($\beta = -0.280$ to -0.325; PVE = 1.1–1.2%). The two chromosome 6 SNPs were significant only before adjustment. Their effects decreased from $\beta = 0.607$–0.705 (PVE = 2.5–2.7%) to $\beta = 0.043$–0.092 (PVE = 0.01–0.04%), indicating an effect mediated by body length. At 3_30808740, the effect size and the explained variance were essentially unchanged after adjustment ($\beta = 0.239$ vs 0.243; PVE = 1.41% vs 1.46%), whereas $-\log_{10}P$ increased from 2.14 to 7.70. This locus therefore approached the adjusted threshold while its estimated effect remained stable, indicating that the increase in significance reflected greater detection power rather than a change in effect size, and the locus therefore warrants attention in follow-up analyses.

3.5 Positional annotation of the associated regions

All six leading SNPs were located in intergenic regions (Table 3). The three chromosome 3 SNPs were distributed across two regions: 3_8284439 (8.28 Mb) flanks *SLC22*, whereas 3_9054098 and 3_9054101 (9.01–9.05 Mb) flank *Jus*. 3_30808740 (30.81 Mb) lies between *BBIP1* and *AHR*, and the two chromosome 6 SNPs flank *G2B*.

The five candidate genes encode distinct domain architectures (Online Resource 6). *SLC22* contains a major facilitator superfamily (MFS) transporter domain (residues 110–530), the defining feature of solute carrier transporters. *Jus* contains a single EB module (40–98), and *BBIP1* contains a ciliary BBSome complex subunit 10 domain (5–65). *AHR* combines a basic helix-loop-helix (bHLH) DNA-binding domain (46–87) with two PAS domains (127–186 and 276–378), the architecture of bHLH-PAS transcription factors. *G2B* contains a glucosidase II beta subunit N-terminal region with an LDL-receptor class A repeat (16–175), an EF-hand domain (289–328), and a beta subunit-like C-terminal region with a mannose-6-phosphate receptor domain (440–578).

3.6 Mediation analysis

To separate direct effects of the associated loci on eye size from effects mediated by body length, causal mediation analysis was performed. The five loci analysed correspond to the six SNPs listed in Table 4: the two SNPs located 3 bp apart on chromosome 3 (3_9054098 and 3_9054101) were treated as a single locus, giving 3_8284439, 3_9054098/3_9054101, 3_30808740, 6_12640972 and 6_12808861. Confidence intervals and P values were obtained by block bootstrap resampling the five genetic groups identified in the population structure analysis (2,000 resamples).

Three effect patterns were observed (Table 4). For the two chromosome 6 SNPs, the indirect effect was significant and the direct effect was not, indicating that their association with eye size was fully mediated by body length: ACME = 0.808 (P = 0.007) and 0.573 (P = 0.008), with non-significant ADE of 0.056 (P = 0.533) and 0.028 (P = 0.653), corresponding to 93.5% and 95.4% of the total effect being mediated (both P < 0.001). For 3_30808740, the indirect effect was not significant (ACME = 0.082, P = 0.178), whereas the direct effect was significant (ADE = 0.247, P < 0.001), indicating a direct effect on eye size. For the three chromosome 3 SNPs, both the indirect and direct effects were significant, indicating partial mediation by body length. The ACME was −0.431 (P = 0.023) at 3_8284439 and −0.393 and −0.403 at 3_9054098 and 3_9054101 (both P < 0.001), and the proportion mediated was 57.6–62.9%. The corresponding direct effects were −0.288, −0.290 and −0.238 (all P < 0.001). Across the five loci, total effects ranged from −0.720 at 3_8284439 to 0.865 at 6_12640972 (Table 4). Online Resource 7 compares the two resampling schemes for the six SNPs analysed. Point estimates were identical, and only the confidence intervals and P values differed, with the ratio of interval widths between the two schemes ranging from 0.52 to 2.59. Neither scheme was uniformly more conservative, and the direction and significance of every direct and indirect effect were unchanged.

3.7 Haplotype analysis

Haplotype analysis was performed for the four gene regions as described. Each locus-specific SNP set was defined by linkage disequilibrium with the index SNP ($r^2 \geq 0.2$ within ±50 kb). Haplotype–phenotype associations were tested against the normalized composite eye-size phenotype, and genotypes at the index SNP of each region were compared separately (Fig. 6d; Online Resources 1d, 2d and 3d). The four regions differed in haplotype diversity. The *SLC22* region was resolved into seven haplotypes, the *Jus* region into ten haplotypes, the *BBIP1/AHR* region into eleven haplotypes and the *G2B* region into ten haplotypes (Fig. 6a; Online Resources 1a, 2a and 3a). In each region, haplotype groups differed significantly in eye size ($P < 0.05$ for several haplotype pairs) (Fig. 6b, c; Online Resources 1b, c, 2b, c and 3b, c).

The strongest haplotype-level evidence was observed in the chromosome 3 region at approximately 9 Mb, which was therefore chosen as the representative analysis (Fig. 6). Pairwise comparisons among the ten haplotypes resolved several groups that differed in eye size (Fig. 6b, c), and genotypes at the index SNP 3_9054101 differed significantly in all three pairwise comparisons (G versus A|G, $P = 2.3 \times 10^{-4}$; A versus A|G, $P = 1.4 \times 10^{-3}$; G versus A, $P = 1.7 \times 10^{-2}$) (Fig. 6d). Site-level effect estimates across the *Jus* gene model reached a maximum of $-\log_{10}P = 8.36$ (Fig. 6e). The remaining three regions are presented as supplementary analyses (Online Resources 1–3). At 6_12640972 in the *G2B* region, the A and A|G genotype classes differed significantly ($P = 4.9 \times 10^{-5}$) and the site-level signal reached $-\log_{10}P = 7.61$ (Online Resource 1d, e). At 3_8284439 in the *SLC22* region, two of the three genotype-class comparisons were significant ($P = 3.6 \times 10^{-6}$ and $P = 0.039$) and the site-level signal reached $-\log_{10}P = 6.72$ (Online Resource 3d, e). At 3_30808740 in the *BBIP1/AHR* region, two of the three comparisons were significant ($P = 1.4 \times 10^{-4}$ and $P = 6.3 \times 10^{-3}$), with a site-level maximum of $-\log_{10}P = 5.79$ (Online Resource 2d, e).

4. Discussion

In this study, the five eye-size measurements behaved as a single phenotypic complex. Genetic correlations among them ranged from 0.857 to 0.999 and phenotypic correlations from 0.761 to 0.967, with moderate heritability estimates (0.312–0.493; Table 2). Correlations of the eye-size traits with the two body-size traits were also positive but consistently lower, ranging from 0.622 to 0.877 at the genetic level and from 0.495 to 0.719 at the phenotypic level. Near-unity among measurements of the same organ, together with a weaker but still substantial coupling to body size, fits a model in which eye dimensions are set mainly by a common growth programme of the eye primordium. An allometric component scaling eye size to body size would then be superimposed on that programme. A phenotypic relationship between eye dimensions and body weight has been reported previously in *P. vannamei* (Chen et al. 2019). The genetic correlations observed here indicate that this allometric coupling is not only phenotypic but also has a heritable component, so that alleles increasing body size tend to increase eye size.

Two consequences follow for the design of the association analysis. First, the five measurements are not

independent traits, and analysing them separately would multiply nearly redundant tests without adding biological resolution. The first principal component explained 99.957% of the variance, so the composite axis captured nearly all of the shared signal and defined the quantity the five measurements jointly estimate, namely overall eye size. Second, because eye size shares genetic variation with body size, an unadjusted analysis cannot distinguish variants that act on eye development from those that act on growth and carry eye size along as a consequence. Fitting BL as a covariate partitions the two components: retention of a signal after adjustment points to an eye-intrinsic component, whereas attenuation indicates a component shared with overall body size. The adjusted GWAS analysis therefore answers a distinct question, and the two analyses were interpreted together, with the mediation analysis used to quantify the partition.

The partition made visible by this design is first evident in the direction and stability of the effects. Six SNPs defined the two associated regions, with small effects (PVE 1.4–3.9%) and consistent directions within each region (Table 3). The chromosome 3 SNPs at 8.28–9.05 Mb carried negative effects and retained significance after adjustment (PVE 1.1–1.2%), whereas the chromosome 6 SNPs carried positive effects of larger magnitude that collapsed after adjustment (PVE 0.01–0.04%). At 3_30808740 the effect was small but stable (PVE 1.41% before and 1.46% after adjustment), while its significance rose from $-\log_{10}P = 2.14$ to 7.70. The opposite signs of the two regions make a single shared pathway unlikely, and the small individual PVE values fit a polygenic architecture in which each locus contributes a modest fraction of the variance.

Mediation analysis resolved the mode of action of these effects. At the two chromosome 6 SNPs the indirect effect was significant (ACME = 0.808, P = 0.007 and 0.573, P = 0.008), whereas the direct effect was not (ADE = 0.056 and 0.028). For these two SNPs, 93.5% and 95.4% of the total effect were mediated by body length, indicating that the variants influence eye size through overall growth rather than through the eye itself. At 3_30808740 the pattern was reversed: the indirect effect was not significant (ACME = 0.082, P = 0.178) whereas the direct effect was (ADE = 0.247, P < 0.001), and the proportion mediated (24.9%) was not significant. The three chromosome 3 SNPs at 8.28–9.05 Mb showed both significant indirect and direct effects (ACME = −0.431 to −0.393, ADE = −0.288 to −0.238), with 57.6–62.9% of the total effect mediated. This region therefore carries both an eye-intrinsic and a growth-dependent component. The behaviour of 3_30808740 after adjustment illustrates this pattern. Its effect size was essentially unchanged, while its significance increased by more than five orders of magnitude. The gain is therefore one of precision rather than of effect size. Body length accounts for a large fraction of eye-size variance, so removing it lowers the residual variance against which the direct effect is tested. Signals attenuated by adjustment (chromosome 6) are more likely to mark variants of the growth axis, whereas signals retained or revealed by adjustment (chromosome 3) are better candidates for eye-intrinsic regulation.

The associated regions are intergenic, so the genes discussed below are positional candidates rather than established targets. *SLC22* flanks the chromosome 3 SNPs at 8.28 Mb and encodes a major facilitator

superfamily organic cation transporter. Members of this family are functional in ocular tissues: *SLC22A5/OCTN2* is expressed in lens epithelial cells (Ochiai et al. 2014), and the blood-aqueous humour barrier expresses SLC transporters that shape aqueous humour composition and hence intraocular pressure (Lee and Pelis 2016). L-carnitine also protects lens epithelial cells against oxidative damage (Li et al. 2019). A transporter variant could therefore act on lens growth and aqueous humour dynamics, both of which contribute to eye size. *Jus* flanks the chromosome 3 SNPs at 9.01–9.05 Mb and encodes a single EB-module protein whose *Drosophila* orthologue defines a neuronal population underlying epileptogenesis (Horne et al. 2017). Its family is ecdysone-inducible, and ecdysteroid signalling is required for furrow progression in the developing compound eye (Brennan et al. 1998) and acts in moulting and growth in penaeid shrimp (Zhou et al. 2019). An effect through the timing of ecdysteroid-dependent growth is therefore plausible, but the evidence for *Jus* remains inferential.

BBIP1 and *AHR* flank 3_30808740, the locus that retained a direct effect. *BBIP1* is a BBSome subunit: a null mutation causes Bardet-Biedl syndrome with retinal dystrophy (Scheidecker et al. 2014), primary cilia are required for optic vesicle morphogenesis (Fiore et al. 2020), and ciliary assembly restricts Hedgehog signalling during early eye development (Burnett et al. 2017). A ciliary mechanism for an eye-intrinsic effect on eye size is therefore credible. *AHR* encodes a bHLH-PAS transcription factor orthologous to *Drosophila spineless* (Emmons et al. 1999). Activation by phenanthrene causes ocular developmental toxicity with up-regulation of AhR and MITF (Huang et al. 2013), and *ahr2* is required for craniofacial development in the absence of exogenous ligand (Souder and Gorelick 2019). A developmental or MITF-related route is thus plausible for this locus.

The chromosome 6 loci occupy the opposite end of the same contrast. The two SNPs there flank *G2B*, which encodes the beta subunit of glucosidase II (PRKCSH), a component of endoplasmic reticulum N-glycosylation and folding quality control. PRKCSH and GANAB are effectors of polycystin-1 and are mutated in autosomal dominant polycystic liver disease (Besse et al. 2017), and common variants in the N-glycosylation gene PMM2 are associated with primary open-angle glaucoma (Chen et al. 2014). Congenital disorders of glycosylation frequently include ocular phenotypes (Ng et al. 2024). This evidence is pathway-level rather than gene-specific. The mediation results support this reading: a housekeeping glycosylation enzyme is a plausible contributor to systemic growth and matches the body-length-mediated effect of this locus, whereas the locally acting candidates match the direct component of the chromosome 3 loci. This correspondence remains a hypothesis generated by the present data.

The main limitation is the sample size of 350 individuals, which limited power for small effects and inflated the effect sizes of the significant SNPs. The PVE estimates should therefore be treated as upper bounds for this discovery panel. In addition, the population is a single cultured stock, and the genotype set was imputed from low-coverage resequencing rather than directly genotyped, which further restricts the scope of inference from this panel. In the mediation models, the genomic relationship matrix was also not included as a random effect, resulting in confidence intervals that depend on principal-component adjustment for relatedness and that were obtained by resampling the five genetic clusters as

blocks. Moreover, mediation estimates rest on the sequential ignorability assumption, which cannot be verified in cross-sectional data, so the distinction between direct and body-length-mediated effects should be read as a quantitative decomposition rather than a causal proof. At the variant level, all variants reported here are intergenic, the causal variant and the target gene within each region remain unresolved, and no expression or functional data are available for validation. Because much of the functional evidence for these candidates comes from *Drosophila* and vertebrate models, and eye-development genetics in crustaceans is still largely descriptive, the gene-level interpretation here is positional rather than demonstrated in shrimp. Accordingly, these loci should be regarded as prioritized candidates for replication and functional work rather than as markers ready for application.

This study shows that eye size in *P. vannamei* is a highly integrated phenotype whose genetic architecture is organized partly around the growth axis and partly around eye-intrinsic regulation. Combining the five eye measurements into a single composite phenotype, and contrasting an unadjusted scan with one conditioned on body length, separated these two components. The signals could then be assigned to a direct, a body-length-mediated or a precision-limited class, a distinction that a single unconditional scan leaves unresolved. The positional candidates identified here provide concrete hypotheses for expression and functional studies, and the design offers a template for association analysis of image-derived morphological traits in aquaculture species.

Statements and Declarations

Competing interests The authors declare no competing financial or non-financial interests.

Funding The Hainan Province Science and Technology Talent Innovation Project (KJRC2025B14), Key R&D Project of Hainan Province (ZDYF2023XDNY176), Hainan Yazhou Bay Seed Industry Laboratory project (B24YQ0009, B24E10010, B25H10CI9), Blue Seed Industry Innovation Project of Qingdao Institute of Blue Seed Industry (QDLYY-2024014), and Fundamental Research Funds for the Central Universities (202341015).

Data availability The genotype data are available through Figshare: https://doi.org/10.6084/m9.figshare.32860241.

Author contributions Mingxuan Teng and Zewei Hu: Data curation, Investigation, Visualization. Jie Ma, Mingyang Zhao, Kaixiang Gao and Hao Wang: Methodology, Resources. Mingxuan Teng: Formal analysis, Writing - original draft. Mengran Liu and Qifan Zeng: Writing - review and editing. Qifan Zeng: Conceptualization, Supervision. All authors read and approved the final manuscript.

Ethics approval Not applicable; the study was carried out on an invertebrate species (Penaeus vannamei), for which no specific ethical approval is required.

Acknowledgements We acknowledge the support of the High-Performance Biological Supercomputing Center at the Ocean University of China and the High-Performance Computing Platform of YZBSTCACC for this research.

References

Besse W, Dong K, Choi J, Punia S, Fedeles SV, Choi M, et al (2017) Isolated polycystic liver disease genes define effectors of polycystin-1 function. *Journal of Clinical Investigation* 127(5):1772–1785.

Brennan CA, Ashburner M, Moses K (1998) Ecdysone pathway is required for furrow progression in the developing *Drosophila* eye. *Development* 125(14):2653–2664.

Browning BL, Zhou Y, Browning SR (2018) A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics* 103(3):338–348.

Burnett JB, Lupu FI, Eggenschwiler JT (2017) Proper ciliary assembly is critical for restricting Hedgehog signaling during early eye development in mice. *Developmental Biology* 430(1):32–40.

Chen Y, Lin Y, Vithana EN, Jia L, Zuo X, Wong TY, et al (2014) Common variants near *ABCA1* and in *PMM2* are associated with primary open-angle glaucoma. *Nature Genetics* 46(10):1115–1119.

Chen F, Xu J, Wang Y, Shi J (2019) Establishing an eyeball-weight relationship for *Litopenaeus vannamei* using machine vision technology. *Aquacultural Engineering* 87:102014.

Chen S (2025) fastp 1.0: an ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta* 4(5):e70078.

Collins TJ (2007) ImageJ for microscopy. *Biotechniques* 43(1 Suppl):25–30. https://doi.org/10.2144/000112517

Emmons RB, Duncan D, Estes PA, Kiefel P, Mosher JT, Sonnenfeld M, et al (1999) The Spineless-Aristapedia and Tango bHLH-PAS proteins interact to control antennal and tarsal development in *Drosophila*. *Development* 126(17):3937–3945.

FAO (2024) The State of World Fisheries and Aquaculture 2024. FAO, Rome

Ferreira T, Rasband W (2012) ImageJ User Guide: IJ 1.46r. National Institutes of Health, Bethesda (MD)

Fiore L, Takata N, Acosta S, Ma W, Pandit T, Oxendine M, et al (2020) Optic vesicle morphogenesis requires primary cilia. *Developmental Biology* 462(2):119–128.

Freed DN, Aldana R, Weber JA, Edwards JS (2017) The Sentieon genomics tools - a fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv* https://doi.org/10.1101/115717

Horne M, Krebushevski K, Wells A, Tunio N, Jarvis C, Francisco G, et al (2017) *julius seizure*, a *Drosophila* mutant, defines a neuronal population underlying epileptogenesis. *Genetics* 205(3):1261–1269.

Huang L, Wang C, Zhang Y, Wu M, Zuo Z (2013) Phenanthrene causes ocular developmental toxicity in zebrafish embryos and the possible mechanisms involved. *Journal of Hazardous Materials* 261:172–

180.

Imai K, Keele L, Tingley D (2010) A general approach to causal mediation analysis. *Psychological Methods* 15(4):309–334.

John M, Korte A, Todesco M, Grimm DG (2024) Population-aware permutation-based significance thresholds for genome-wide association studies. *Bioinformatics Advances*.

Kendig KI, Baheti S, Bockol MA, Drucker TM, Mainzer LS (2019) Sentieon DNASeq variant calling workflow demonstrates strong computational performance and accuracy. *Frontiers in Genetics* 10:736.

Lee J, Pelis RM (2016) Drug transport by the blood-aqueous humor barrier of the eye. *Drug Metabolism and Disposition* 44(10):1675–1681.

Li X, Meng F, Li H, Hua X, Wu LA, Yuan X (2019) L-carnitine alleviates oxidative stress-related damage via *MAPK* signaling in human lens epithelial cells exposed to H2O2. *International Journal of Molecular Medicine* 44(4):1515–1522.

Liao J, Lai Z, He C, Li G, Dong Z, Guo Y, et al (2025) A high-quality chromosome-level genome assembly of Pacific whiteleg shrimp (*Penaeus vannamei*). *Scientific Data* 12(1):340.

Lv T, Xia Y, Tan J, Fu Q, Luo K, Meng X, et al (2025) Multi-marker GWAS and variant-specific genomic prediction for growth traits in Pacific white shrimp. *Scientific Reports* 15(1):42103.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler AV, et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Ng BG, Freeze HH, Himmelreich N, Blau N, Ferreira CR (2024) Clinical and biochemical footprints of congenital disorders of glycosylation: proposed nosology. *Molecular Genetics and Metabolism* 142(1):108476.

Ochiai H, Kanemaki N, Sato R, Onda K (2014) Distribution, molecular structure and functional analysis of carnitine transporter (*SLC22A5*) in canine lens epithelial cells. *Experimental Animals* 63(4):467–473.

Peña-Martínez EG, Rodríguez-Martínez JA (2024) Decoding non-coding variants: recent approaches to studying their role in gene regulation and human diseases. *Frontiers in Bioscience (Scholar Edition)* 16(1):4. https://doi.org/10.31083/j.fbs1601004

Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11(7):459–463. https://doi.org/10.1038/nrg2813

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human*

Genetics 81(3):559–575.

Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O (2021) Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics* 53(1):120–126.

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)

Scheidecker S, Etard C, Pierce NW, Geoffroy V, Schaefer E, Muller J, et al (2014) Exome sequencing of Bardet-Biedl syndrome patient identifies a null mutation in the BBSome subunit *BBIP1* (*BBS18*). *Journal of Medical Genetics* 51(2):132–136.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9(7):671–675.

Souder JP, Gorelick DA (2019) *ahr2*, but not *ahr1a* or *ahr1b*, is required for craniofacial and fin development and TCDD-dependent cardiotoxicity in zebrafish. *Toxicological Sciences* 170(1):25–44.

Sui J, Luan S, Luo K, Meng X, Lu X, Cao B, et al (2016) Genetic parameters and response to selection for harvest body weight of Pacific white shrimp, *Litopenaeus vannamei*. *Aquaculture Research* 47(9):2795–2803.

Wang Q, Yu Y, Zhang Q, Zhang X, Yuan J, Huang H, et al (2019) A novel candidate gene associated with body weight in the Pacific white shrimp *Litopenaeus vannamei*. *Frontiers in Genetics* 10:520.

Wang Q, Yu Y, Zhang Q, Luo Z, Zhang X, Xiang J, et al (2020) The polymorphism of *LvMMD2* and its association with growth traits in *Litopenaeus vannamei*. *Marine Biotechnology* 22(4):564–571.

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88(1):76–82.

Yu X, Faggion S, Liu Y, Wang B, Zeng Q, Lu C, et al (2025) Role of multi-omics in aquaculture genetics and breeding: current status and future perspective. *Science China Life Sciences* 68(9):2591–2604.

Zeng Q, Zhao B, Wang H, Wang M, Teng M, Hu J, et al (2022) Aquaculture Molecular Breeding Platform (AMBP): a comprehensive web server for genotype imputation and genetic analysis in aquaculture. *Nucleic Acids Research* 50(W1):W66–W74.

Zhang X, Yuan J, Sun Y, Li S, Gao Y, Yu Y, et al (2019) Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications* 10(1):356.

Zhang C, Dong S, Xu J, He W, Yang T (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35(10):1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zhang R, Jia G, Diao X (2023) geneHapR: an R package for gene haplotypic statistics and visualization. *BMC Bioinformatics* 24(1):199.

Zhang L, Wang H, Teng M, Zhao Q, Zhao M, Huang Y, et al (2025) Integrative genomics reveals *PDE11A* and *SAAL* are key regulators of growth in Pacific white shrimp (*Penaeus vannamei*). *Water Biology and Security* 5(3):100474.

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* 44(7):821–824.

Zhou K, Zhou F, Jiang S, Huang J, Yang Q, Yang L, et al (2019) Ecdysone inducible gene *E75* from black tiger shrimp *Penaeus monodon*: characterization and elucidation of its role in molting. *Molecular Reproduction and Development* 86(3):265–277.

Figure legends

Fig. 1 Phenotypic characterization of eye-size and growth traits in Penaeus vannamei. (a) Representative image showing the measurement of eye traits. The eyeball contour was manually delineated (yellow outline), and the fitted ellipse (red dashed line) was generated in ImageJ. (b-h) Frequency distributions and Q-Q plots (insets) of phenotypic values for the five eye-size traits (Area, O-Area, Feret, MinFeret, Perim) and the two body-size traits (body length, BL; body weight, BW)

Fig. 2 Dimensionality reduction and normalization of the eye-size traits. (a) Histogram and Q-Q plot of the first principal component (PC1) of the five eye-size traits. PC1 explained 99.957% of the total variance. (b) Histogram and Q-Q plot of PC1 after rank-based inverse normal transformation

Fig. 3 Imputation accuracy of the variants retained for association analysis. (a) Genotype and allele concordance estimated in 91 individuals with high-depth sequencing. (b, c) Genotype and allele concordance summarized across minor allele frequency bins from 0.05 to 0.50

Fig. 4 SNP density, linkage disequilibrium and population structure. (a) Genome-wide SNP density. (b) LD decay curve showing r^2 as a function of physical distance. The half-decay distance is approximately 16 kb (red dashed line). (c) Heatmap of the genomic relationship matrix (GRM). (d, e) Principal component analysis plots of PC1 versus PC2 (d) and PC1 versus PC3 (e). (f) Maximum-likelihood phylogenetic tree. Colours denote the five genetic clusters identified in the population structure analysis

Fig. 5 Genome-wide association analysis of the composite eye-size phenotype. (a) Manhattan and Q-Q plots for the composite phenotype. (b, c) The same plots for body length (BL) and body weight (BW). (d) Manhattan and Q-Q plots for the composite phenotype with BL included as an additional covariate. The red line denotes the permutation-derived genome-wide threshold and the green dashed line the suggestive threshold ($P = 1 \times 10^{-6}$)

Fig. 6 Haplotype analysis of the chromosome 3 region at approximately 9 Mb. (a) Haplotype structure of the LD-defined SNP set flanking Jus. (b) Pairwise P-value matrix among haplotypes. (c) Distribution

of the composite eye-size phenotype across haplotypes. Asterisks denote significant pairwise differences. (d) Comparison of the phenotype among genotype classes at the index SNP 3_9054101. (e) Site-level effect estimates and $-\log_{10}P$ values plotted along the Jus gene model

Table legends

Table 1 Summary statistics for the phenotypic traits measured in the P. vannamei GWAS panel (n = 350)

Table 2 Heritability (diagonal), phenotypic (lower triangle) and genetic (upper triangle) correlations among the eye-size and body-size traits. Values are estimates (standard error). P values are from Wald tests

Table 3 Leading SNPs associated with the composite eye-size phenotype. Alleles, allele frequencies, effect sizes (beta), phenotypic variance explained (PVE), significance and genomic annotation are given for the primary analysis and for the analysis with body length as a covariate

Table 4 Causal mediation analysis with body length as the mediator. The average causal mediation effect (ACME), average direct effect (ADE), total effect and proportion mediated are shown with 95% confidence intervals obtained by block bootstrap resampling of the five genetic groups (2,000 resamples)

Online Resource figure legends

Online Resource 1 Haplotype analysis of the G2B region on chromosome 6. (a) Haplotype structure of the LD-defined SNP set flanking G2B. (b) Pairwise P-value matrix among haplotypes. (c) Distribution of the composite eye-size phenotype across haplotypes. (d) Comparison of the phenotype among genotype classes at the index SNP 6_12640972. (e) Site-level effect estimates and $-\log_{10}P$ values along the G2B gene model

Online Resource 2 Haplotype analysis of the BBIP1/AHR region at 30.81 Mb. (a) Haplotype structure of the LD-defined SNP set. (b) Pairwise P-value matrix among haplotypes. (c) Distribution of the composite eye-size phenotype across haplotypes. (d) Comparison of the phenotype among genotype classes at the index SNP 3_30808740. (e) Site-level effect estimates and $-\log_{10}P$ values along the BBIP1 and AHR gene models

Online Resource 3 Haplotype analysis of the SLC22 region at 8.28 Mb. (a) Haplotype structure of the LD-defined SNP set flanking SLC22. (b) Pairwise P-value matrix among haplotypes. (c) Distribution of the composite eye-size phenotype across haplotypes. (d) Comparison of the phenotype among genotype classes at the index SNP 3_8284439. (e) Site-level effect estimates and $-\log_{10}P$ values along the SLC22 gene model

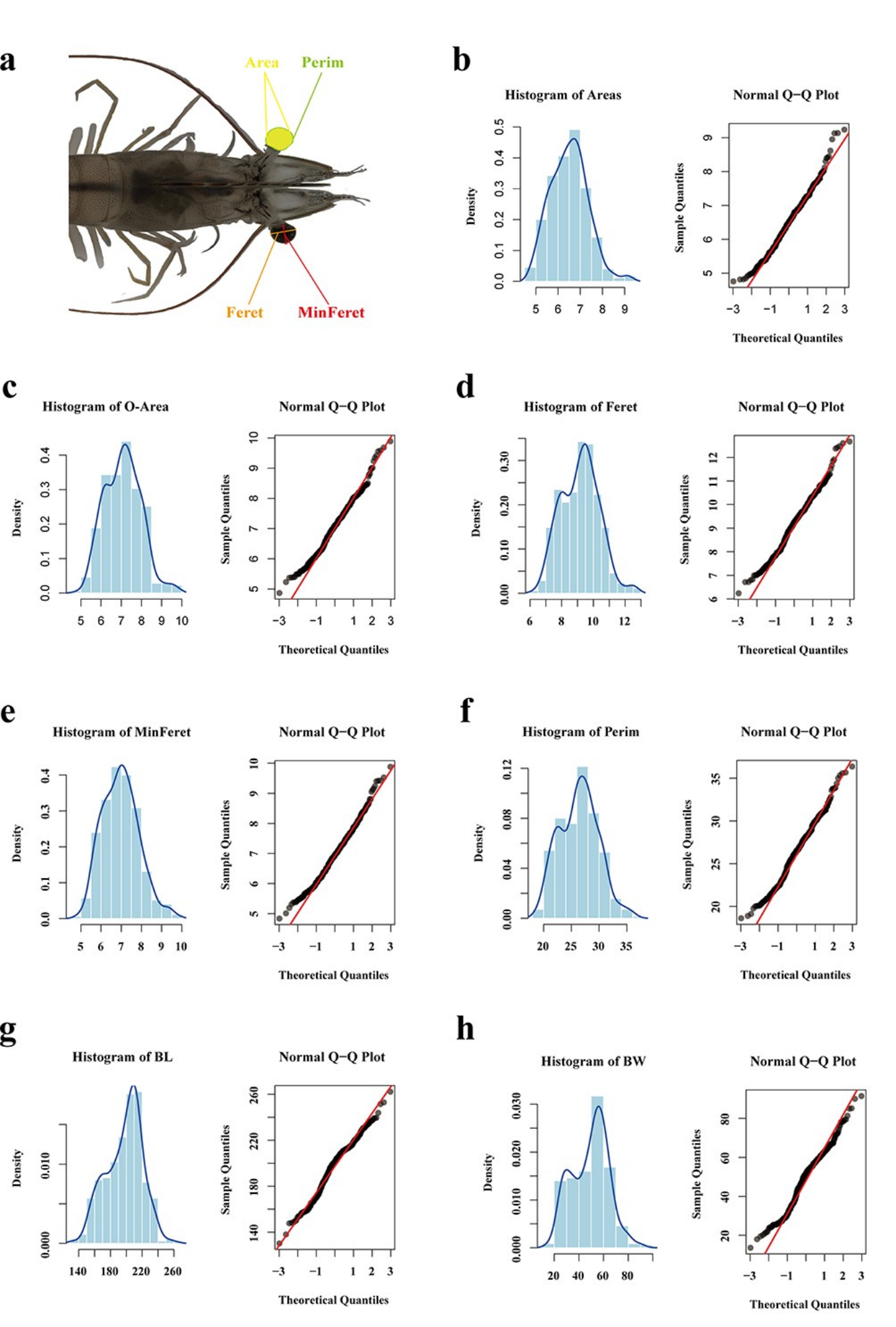
Online Resource table legends

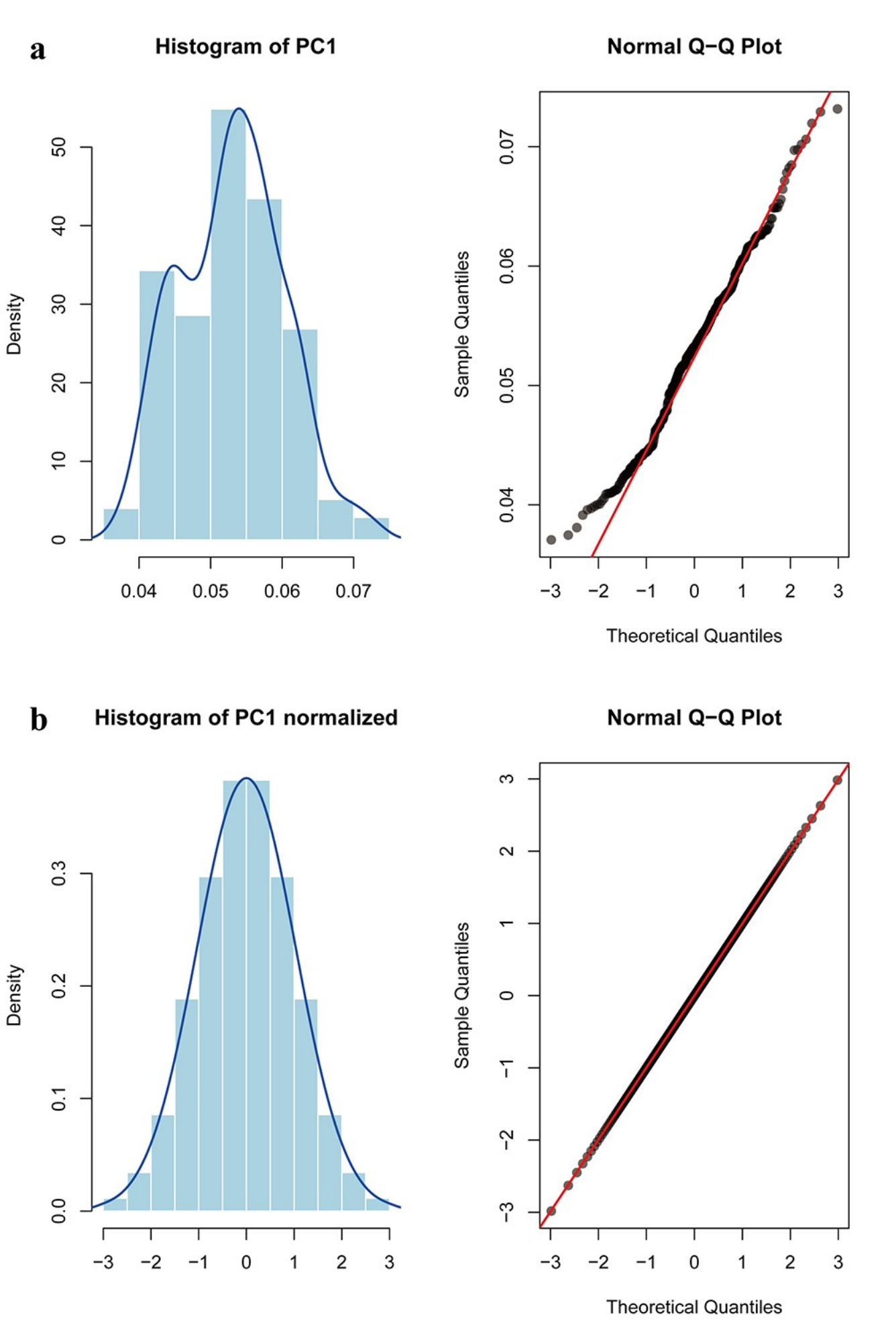
Online Resource 4 Shapiro-Wilk test for the five eye-size traits and the two body-size traits

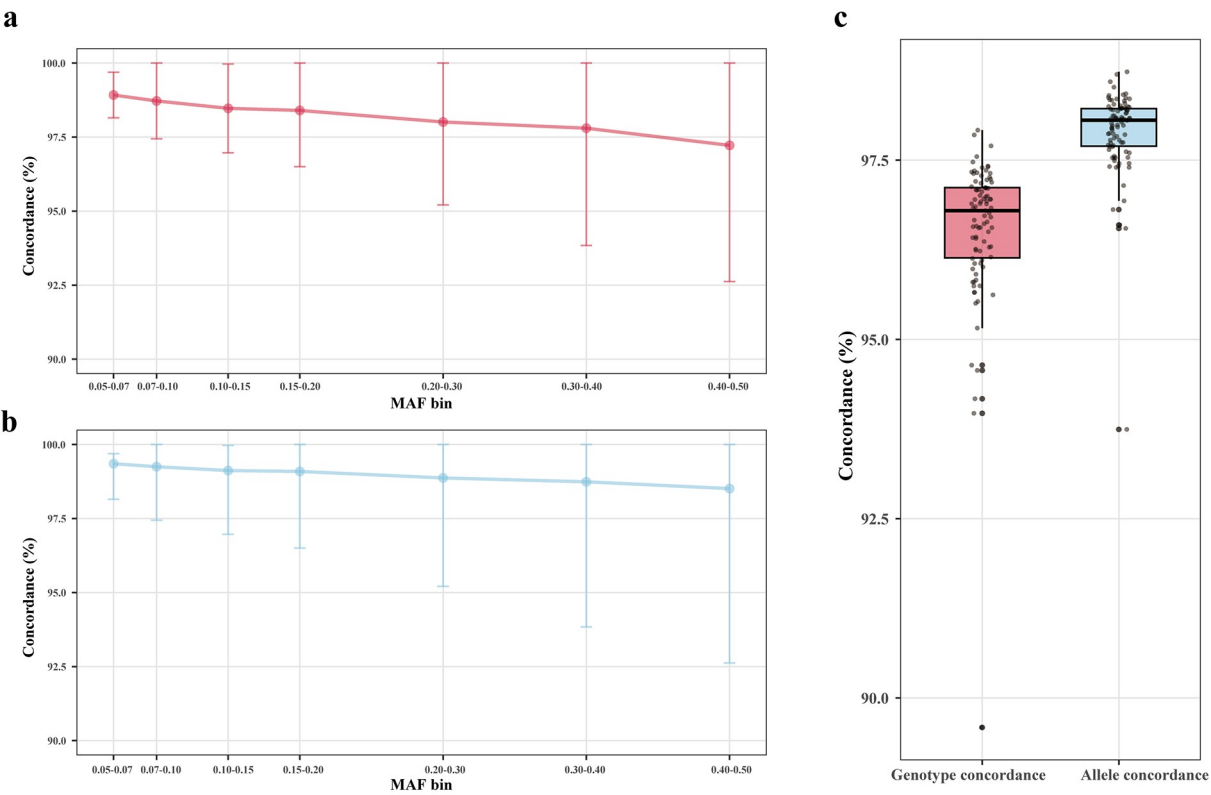
Online Resource 5 Standard deviations of the five principal components of the eye-size traits

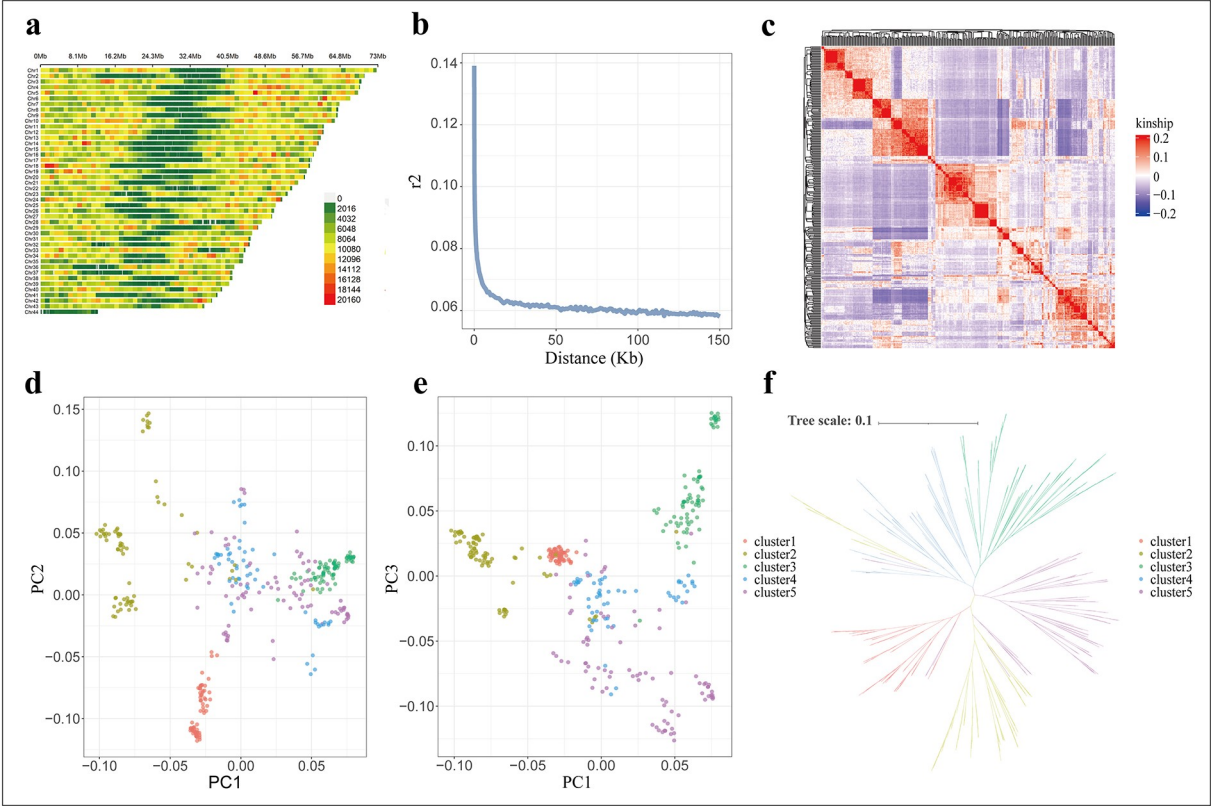
Online Resource 6 Functional annotation and conserved domains of the five positional candidate genes

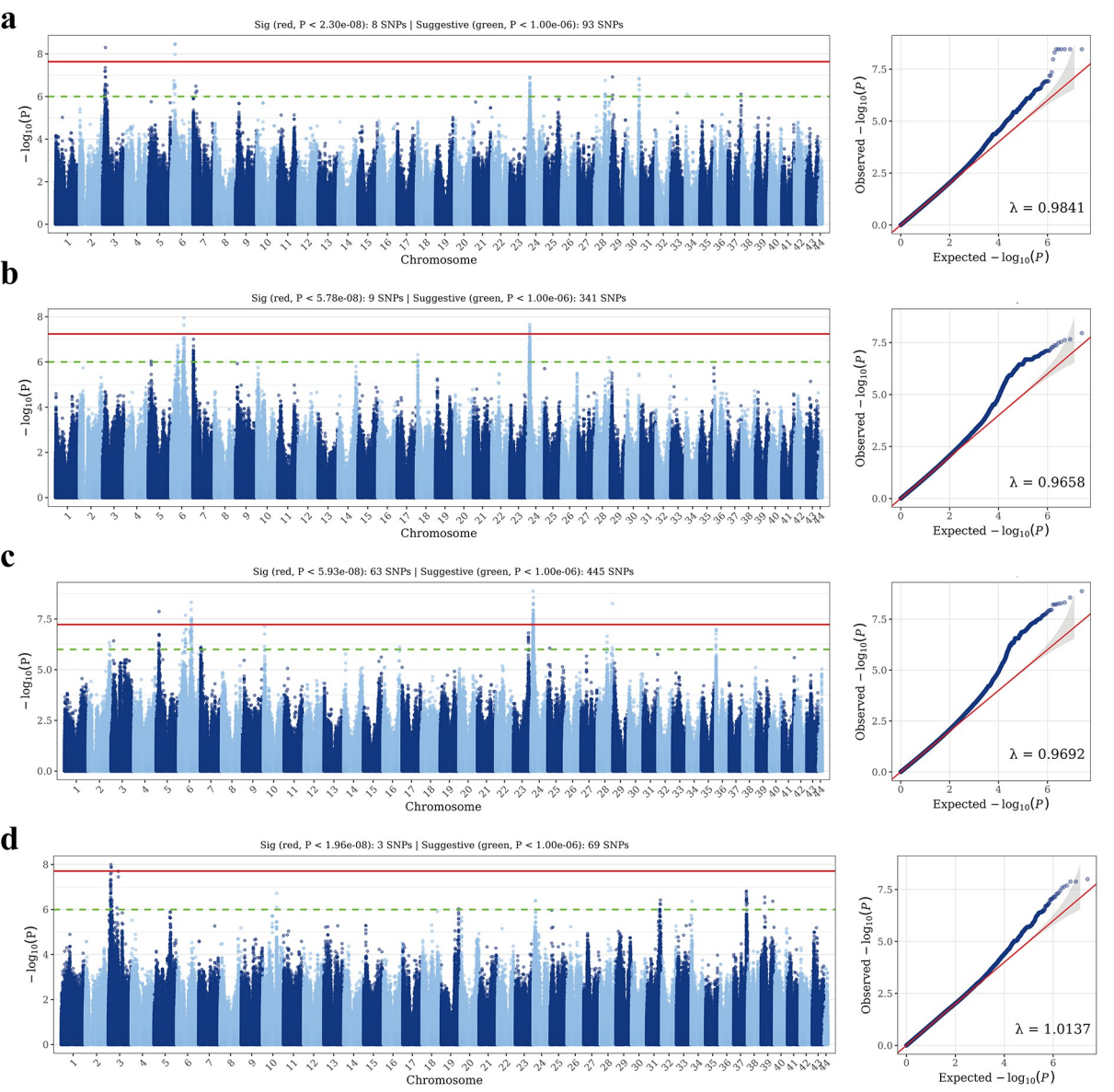
Online Resource 7 Comparison of individual-level and group block bootstrap resampling for the mediation analysis


a
Area
Perim
Feret
MinFeret
b
Histogram of Areas
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
c
Histogram of O-Area
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
d
Histogram of Feret
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
e
Histogram of MinFeret
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
f
Histogram of Perim
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
g
Histogram of BL
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
h
Histogram of BW
Density
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles


a
Histogram of PC1
Density
0
10
20
30
40
50
0.04
0.05
0.06
0.07
Normal Q−Q Plot
Sample Quantiles
0.04
0.05
0.06
0.07
−3
−2
−1
0
1
2
3
Theoretical Quantiles
b
Histogram of PC1 normalized
Density
0.0
0.1
0.2
0.3
−3
−2
−1
0
1
2
3
Normal Q−Q Plot
Sample Quantiles
−3
−2
−1
0
1
2
3
Theoretical Quantiles


a
Concordance (%)
100.0
97.5
95.0
92.5
90.0
0.05-0.07
0.07-0.10
0.10-0.15
0.15-0.20
0.20-0.30
0.30-0.40
0.40-0.50
MAF bin
b
Concordance (%)
100.0
97.5
95.0
92.5
90.0
0.05-0.07
0.07-0.10
0.10-0.15
0.15-0.20
0.20-0.30
0.30-0.40
0.40-0.50
MAF bin
c
Concordance (%)
97.5
95.0
92.5
90.0
Genotype concordance
Allele concordance


a
0Mb
8.1Mb
16.2Mb
24.3Mb
32.4Mb
40.5Mb
48.6Mb
56.7Mb
64.8Mb
73Mb
0
2016
4032
6048
8064
10080
12096
14112
16128
18144
20160
b
r2
0.14
0.12
0.10
0.08
0.06
0
50
100
150
Distance (Kb)
c
kinship
0.2
0.1
0
−0.1
−0.2
d
PC2
PC1
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00
0.05
e
PC3
PC1
0.10
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00
0.05
cluster1
cluster2
cluster3
cluster4
cluster5
f
Tree scale: 0.1
cluster1
cluster2
cluster3
cluster4
cluster5


a
Sig (red, P < 2.30e-08): 8 SNPs | Suggestive (green, P < 1.00e-06): 93 SNPs
− log10(P)
Chromosome
Observed − log10(P)
Expected − log10(P)
λ = 0.9841
b
Sig (red, P < 5.78e-08): 9 SNPs | Suggestive (green, P < 1.00e-06): 341 SNPs
− log10(P)
Chromosome
Observed − log10(P)
Expected − log10(P)
λ = 0.9658
c
Sig (red, P < 5.93e-08): 63 SNPs | Suggestive (green, P < 1.00e-06): 445 SNPs
− log10(P)
Chromosome
Observed − log10(P)
Expected − log10(P)
λ = 0.9692
d
Sig (red, P < 1.96e-08): 3 SNPs | Suggestive (green, P < 1.00e-06): 69 SNPs
− log10(P)
Chromosome
Observed − log10(P)
Expected − log10(P)
λ = 1.0137

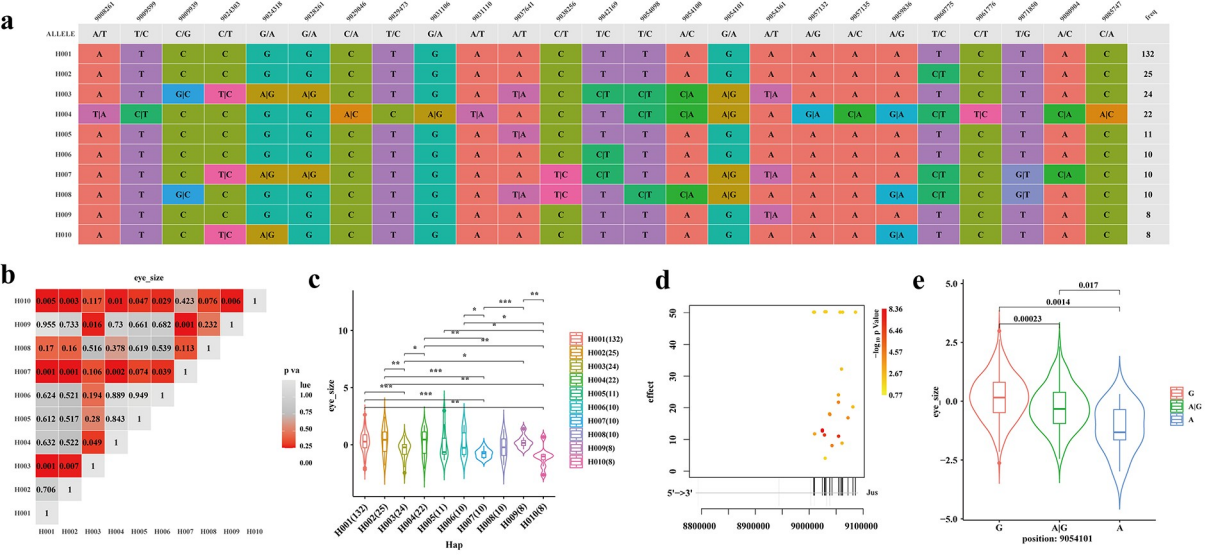

a
ALLELE
A/T T/C C/G C/T G/A G/A C/A T/C G/A A/T A/T C/T T/C T/C A/C G/A A/T A/G A/C A/G T/C C/T T/G A/C C/A
H001 A T C C G G C T G A A C T T A G A A A A T C T A C 132
H002 A T C C G G C T G A A C T T A G A A A A C|T C T A C 25
H003 A T G|C T|C A|G A|G C T G A T|A C C|T C|T C|A A|G T|A A A A T C T A C 24
H004 T|A C|T C C G G A|C C A|G T|A A C T C|T C|A A|G A G|A C|A G|A C|T T|C T C|A A|C 22
H005 A T C C G G C T G A T|A C T T A G A A A A T C T A C 11
H006 A T C C G G C T G A A C C|T T A G A A A A T C T A C 10
H007 A T C T|C A|G A|G C T G A A T|C C|T T A A|G T|A A A A C|T C G|T C|A C 10
H008 A T G|C C G G C T G A T|A T|C T C|T C|A A|G A A A G|A C|T C G|T A C 10
H009 A T C C G G C T G A A C T T A G T|A A A A T C T A C 8
H010 A T C T|C A|G G C T G A A C T T A G A A A G|A T C T A C 8
b
eye_size
H010 0.005 0.003 0.117 0.01 0.047 0.029 0.423 0.076 0.006 1
H009 0.955 0.733 0.016 0.73 0.661 0.682 0.001 0.232 1
H008 0.17 0.16 0.516 0.378 0.619 0.539 0.113 1
H007 0.001 0.001 0.106 0.002 0.074 0.039 1
H006 0.624 0.521 0.194 0.889 0.949 1
H005 0.612 0.517 0.28 0.843 1
H004 0.632 0.522 0.049 1
H003 0.001 0.007 1
H002 0.706 1
H001 1
H001 H002 H003 H004 H005 H006 H007 H008 H009 H010
p va
lue
1.00
0.75
0.50
0.25
0.00
c
eye_size
Hap
H001(132)
H002(25)
H003(24)
H004(22)
H005(11)
H006(10)
H007(10)
H008(10)
H009(8)
H010(8)
d
effect
5'-->3'
Jus
8800000 8900000 9000000 9100000
−log10 p Value
8.36
6.46
4.57
2.67
0.77
e
eye_size
0.017
0.0014
0.00023
G
A|G
A
position: 9054101

Table 1 Summary statistics for the phenotypic traits measured in the P. vannamei GWAS panel (n = 350)

Description	Number of Individuals	Mean	Max	Min	SD	CV
BL (mm)	350	197.58	262.26	130.19	22.86	0.12
BW (g)	350	49.19	91.50	13.60	14.83	0.30
Area (mm^2)	350	42.92	85.36	22.66	11.04	0.26
Perim (mm)	350	26.35	36.37	18.60	3.49	0.13
Feret (mm)	350	9.16	12.67	6.24	1.19	0.13
MinFeret (mm)	350	6.97	9.88	4.84	0.90	0.13
O-Area (mm^2)	350	50.87	97.79	23.71	12.89	0.25

Table 2 Heritability (diagonal), phenotypic (lower triangle) and genetic (upper triangle) correlations among the eye-size and body-size traits. Values are estimates (standard error). P values are from Wald tests

	Area	**O-Area**	**Feret**	**MinFeret**	**Perim**	**BL**	**BW**
Area	0.457(0.134) p<0.001	0.895(0.046) p<0.001	0.862(0.089) p<0.001	0.989(0.027) p<0.001	0.903(0.047) p<0.001	0.680(0.127) p<0.001	0.685(0.152) p<0.001
O-Area	0.898(0.106) p<0.001	0.419(0.138) p<0.001	0.857(0.060) p<0.001	0.903(0.069) p<0.001	0.873(0.059) p<0.001	0.622(0.141) p<0.001	0.877(0.108) p<0.001
Feret	0.855(0.127) p<0.001	0.874(0.136) p<0.001	0.493(0.127) p<0.001	0.989(0.040) p<0.001	0.999(0.002) p<0.001	0.863(0.067) p<0.001	0.788(0.121) p<0.001
MinFeret	0.804(0.072) p<0.001	0.761(0.193) p<0.001	0.786(0.114) p<0.001	0.312(0.113) p=0.006	0.972(0.021) p<0.001	0.701(0.171) p<0.001	0.668(0.201) p<0.001
Perim	0.794(0.137) p<0.001	0.852(0.134) p<0.001	0.967(0.002) p<0.001	0.947(0.062) p<0.001	0.465(0.133) p<0.001	0.766(0.097) p<0.001	0.708(0.142) p<0.001
BL	0.645(0.161) p<0.001	0.663(0.157) p<0.001	0.719(0.151) p<0.001	0.636(0.148) p<0.001	0.706(0.156) p<0.001	0.484(0.135) p<0.001	0.863(0.067) p<0.001
BW	0.495(0.162) p<0.001	0.538(0.168) p<0.001	0.577(0.162) p<0.001	0.510(0.157) p<0.001	0.584(0.167) p<0.001	0.756(0.143) p<0.001	0.412(0.141) p=0.003

Table 3 Leading SNPs associated with the composite eye-size phenotype. Alleles, allele frequencies, effect sizes (β), phenotypic variance explained (PVE), significance and genomic annotation are given for the primary analysis and for the analysis with body length as a covariate

Chromosome	Position	Allele	Allele frequency	Primary			BL-adjusted			Genomic annotation
				β	PVE	$-\log_{10} P$	β	PVE	$-\log_{10} P$	
3	8284439	G/A	0.154	-0.525	0.03594	7.193	-0.308	0.01237	7.998	Intergenic (SLC22)
3	9054098	C/T	0.13	-0.54	0.03301	6.918	-0.325	0.01196	7.881	Intergenic (Jus)
3	9054101	A/G	0.169	-0.529	0.03933	8.301	-0.28	0.01102	7.642	Intergenic (Jus)
3	30808740	T/C	0.446	0.239	0.01412	2.138	0.243	0.0146	7.699	Intergenic (BBIP1, AHR)
6	12640972	G/A	0.054	0.705	0.02541	8.458	0.092	0.00043	0.457	Intergenic (G2B)
6	12808861	C/G	0.08	0.607	0.02714	7.974	0.043	0.00014	0.243	Intergenic (G2B)

Table 4 Causal mediation analysis with body length as the mediator. The average causal mediation effect (ACME), average direct effect (ADE), total effect and proportion mediated are shown with 95% confidence intervals obtained by block bootstrap resampling of the five genetic groups (2,000 resamples)

Chromosome	Position (bp)	Effect pattern	ACME		ADE		Total effect (95% CI)	Proportion mediated	
			Value (95% CI)	P	Value (95% CI)	P		Value (95% CI)	P
3	8284439	Partially mediated	-0.431 (-0.738, -0.028)	0.023	-0.288 (-0.391, -0.162)	<0.001	-0.720 (-1.037, -0.332)	0.600 (0.083, 0.774)	0.023
3	9054098	Partially mediated	-0.393 (-0.727, -0.038)	<0.001	-0.290 (-0.390, -0.198)	<0.001	-0.683 (-1.088, -0.299)	0.576 (0.128, 0.727)	<0.001
3	9054101	Partially mediated	-0.403 (-0.708, -0.075)	<0.001	-0.238 (-0.308, -0.172)	<0.001	-0.641 (-0.948, -0.336)	0.629 (0.219, 0.775)	<0.001
3	30808740	Direct	0.082 (-0.046, 0.257)	0.178	0.247 (0.176, 0.344)	<0.001	0.329 (0.224, 0.499)	0.249 (-0.173, 0.554)	0.178
6	12640972	Fully mediated	0.808 (0.380, 0.895)	0.007	0.056 (-0.094, 0.167)	0.533	0.865 (0.497, 0.925)	0.935 (0.748, 1.128)	<0.001
6	12808861	Fully mediated	0.573 (0.069, 0.924)	0.008	0.028 (-0.067, 0.168)	0.653	0.600 (0.054, 0.983)	0.954 (0.730, 1.340)	<0.001

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.tif
- ESM2.tif
- ESM3.tif
- ESM4.docx
- ESM5.docx
- ESM6.docx
- ESM7.docx