Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia

Grazyella M. Yoshida¹, José M. Yáñez¹,²*

¹ Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile
² Núcleo Milenio INVASAL, Concepción, Chile

*Corresponding author:
E-mail: jmayanez@uchile.cl
Abstract

Background: Body traits are generally controlled by several genes in vertebrates (i.e. polygenes), which in turn make them difficult to identify through association mapping. Increasing the power of association studies by combining approaches such as genotype imputation and multi-trait analysis improves the ability to detect quantitative trait loci associated with polygenic traits, such as body traits.

Results: A multi-trait genome-wide association study (mtGWAS) was performed to identify quantitative trait loci (QTL) and genes associated with body traits in Nile tilapia (*Oreochromos niloticus*) using genotypes imputed to whole-genome sequence (WGS). To increase the statistical power of mtGWAS for the detection of genetic associations, summary statistics from single-trait genome-wide association studies (stGWAS) for eight different body traits recorded in 1,309 animals were used. The mtGWAS increased the statistical power from the original sample size from 13% to 44%, depending on the trait analyzed. The better resolution of the WGS data combined with the increased power of the mtGWAS approach, allowed the detection of significant markers not previously found in the stGWAS. Some lead single nucleotide polymorphisms (SNPs) were found within important functional candidate genes previously associated with growth-related traits. For instance, we identified SNP within the α1,6-fucosyltransferase (*FUT8*), solute carrier family 4 member 2 (*SLC4A2*), A disintegrin and metalloproteinase with thrombospondin motifs 9 (*ADAMTS9*) and heart development protein with EGF like domains 1 (*HEG1*) genes, which have been associated with average daily gain in sheep, osteopetrosis in cattle, chest size in goats, and growth and meat quality in sheep, respectively.
Conclusions: The high-resolution mtGWAS presented, allowed identification of significant SNPs, linked to strong functional candidate genes, associated with body traits in Nile tilapia. These results provide further insights about the genetic variants and genes underlying body trait variation in cichlid fish with high accuracy and strong statistical support.

Keywords: body traits, genome-wide association study, genotype imputation, quantitative trait loci, Oreochromos niloticus, multi-trait
Background

Tilapia is one of the most important fish species cultivated in the world, and is currently farmed in more than 125 countries. Total farmed fish production reached 5.8 tons globally, and Nile tilapia (Oreochromis niloticus) represented more than 70% of this volume [1]. Tilapia is generally sold as whole fish or fillets, making body traits such as body and fillet weight among the most economically important traits for this species, and are the primary breeding objectives in aquaculture breeding programs [2]. The most important body traits in Nile tilapia are typically measured as body weight at a specific age (e.g. body weight at harvest), fillet weight or fillet yield (fillet weight/body weight), both show heritability values ranging from 0.06 to 0.48, when using pedigree-based estimates [3–9]. Previous studies have estimated high values of genetic correlation between harvest weight and fillet weight (>0.96) and moderate to high values between harvest weight and fillet yield (0.21 to 0.74) [7, 9, 10], suggesting that is not possible to improve fillet traits independently of weight [11]. Although, previous reports have also identified negative or null genetic correlation between harvest weight and fillet yield [12], which suggests the importance of assessing these relationships on each particular population. Other body traits which have been proposed as selection criteria, to generate more profitable commercial fish populations, are reduced waste (sum of bones, viscera, head, and fins) and carcass weight, due to their higher heritability values, less correlation to body weight, compared to fillet weight, and null or even negative impact on fillet yield [13, 14].

The availability of a chromosome-level reference genome assembly [15] and high-throughput whole-genome sequencing (WGS) methods [16, 17], has allowed for the assessment of genetic variation of different Nile tilapia populations at a genome-wide level
and the recent development of single nucleotide polymorphism (SNP) panels [18, 19]. The availability of Nile tilapia SNP panels made it possible to use modern molecular breeding approaches; including mapping of quantitative trait loci (QTL) through genome-wide association studies (GWAS), marker-assisted selection (MAS) and genomic selection [20, 21]. The GWAS approach evaluates the association between the genotype and the phenotype of interest, with both sources of information available for a large number of individuals. This method captures the linkage disequilibrium (LD) between markers and causative mutations that tend to be inherited together across generations [22]. GWAS has been applied to provide insights into both the genetic architecture and loci underpinning the genetic variation of growth-related traits in different aquaculture species, including Atlantic salmon and catfish [23–26], using high-density SNPs arrays (ranging from 108 K to 218 K SNPs) and, more recently, Nile tilapia by using a medium-density (50 K) SNP array [20]. These studies revealed the polygenic nature of growth-related traits and identified some genes harboring significant SNPs, which are well-known to be involved in growth and bone development, including meprin A subunit beta-like (MEP1A), fibroblast growth factors (FGF), disintegrin and metalloproteinase domain 12 (ADAM12), myosin light chain kinase (MYLK) and transforming growth-factor beta receptor type 3 (TGFBR3).

The use of ultra-high-density SNPs or WGS can improve the accuracy and power of GWAS to detect quantitative trait loci (QTLs) associated with complex traits [27–30]. Although the cost of WGS is rapidly decreasing, it is still expensive to sequence all available phenotyped individuals in a GWAS design. To solve this, genotype imputation strategies to WGS data can be successfully implemented to detect putative causal loci in a cost-efficient manner. Previous studies using imputed genotypes from WGS for GWAS have been reported in cattle [27, 28], pigs [29, 30] and sheep [31]. In addition, new strategies such as multi-trait
GWAS (mtGWAS) analysis are required to increase the power to detect QTL through GWAS, based on genotype imputation of WGS [32]. mtGWAS improves the power of GWAS through the incorporation of summary information contained in the output of single-trait GWAS (stGWAS). Thus, mtGWAS jointly exploits information from genetically correlated traits to increase statistical power, due to fact that the true SNP effects and their estimated error may be correlated across traits. For instance, multi-trait approaches have been implemented in pertinent software, e.g. MTAG v0.9.0 [33], and successfully applied to boost the discovery of genetic variants associated with important traits in humans [34–36].

To the best of our knowledge, no previous studies have shown the use of genotype imputation to high-density SNPs, combined with mtGWAS in aquaculture species, to uncover putative causative genetic variants associated with body traits. The objective of this study was to use mtGWAS and genotypes high-density SNPs to increase the accuracy and power to identify both QTLs and genes associated with eight body traits in Nile tilapia.

**Results**

**Descriptive statistic, quality control and genetic parameters**

A total of 1,309 animals averaging 370 days-old were phenotyped and genotyped. Average, standard deviation, minimum and maximum phenotypic values for average daily gain (ADG), harvest weight (BWH), waste weight (WW), head weight (HW), hon weight (HON), harvest length (BLH), fillet weight (FW) and fillet yield (FY) are reported in Table 1. The coefficient of variation ranged between 6.86% and 27.47%, with the lowest and the highest values calculated for trait FY and FW, respectively.
For WGS, the call-rate parameter excluded the highest number of SNPs (~12 million), whereas MAF discarded ~7.8 million and ~253K SNPs, for WGS and imputed WGS, respectively. The HWE filtered a low number of markers, ~1.8 million for WGS and 79K for imputed WGS. After quality control to the 50K SNP chip, 5,905, 4,114 and 3,665 SNPs were removed by HWE, MAF and genotyping call-rate, respectively. 29,587 SNPs remained for the subsequent analyses. After applying sample call-rate, all samples in both WGS and 50K SNP chip were retained (Supplementary Table 1).

Heritability estimates based on the SNP-based genomic-relationship matrix (GRM) constructed with about 1 million markers ranged from 0.21 to 0.45 for the body traits analyzed here, with the lowest and the highest value determined for FY and HW, respectively (Table 2). Correlation of SNP effects among all body traits analyzed ranged from 0.20 to 1.00, with small values reported for correlations between FY and the rest of the traits (Fig 1).

**Comparison between single-trait and multi-trait GWAS**

The average gain in statistical power for mtGWAS compared to stGWAS was assessed by the increase in the mean $\chi^2$ statistic and was used to calculate how much larger the stGWAS sample would have to be, to be equivalent to the increase expected in $\chi^2$ statistic. We found that the mtGWAS analysis corresponded to gains equivalent to increase the original samples from 13% to 44%, these values corresponded to an increase in sample size from 1,309 in stGWAS to a value ranging from 1,474 to 1,890 in mtGWAS (Table 2). For instance, the number of SNP surpassing the Bonferroni corrected significance threshold for stGWAS and mtGWAS, respectively, was: 1 and 1,359 for ADG, 1 and 1,209 for BWH, 1 and 1,347 for WW, 0 and 1595 for HW, 1 and 1,138 for HON, 0 and 827 for BLH, 1 and 833 for FW, and 1 and 1,920 for FY. In addition, the maximum -log(p-value) increased from 7.52
to 14.58 for ADG, from 7.63 to 14.39 for BWH, from 7.45 to 14.60 for WW, from 5.71 to 14.39 for HW, from 7.45 to 13.00 for HON, from 5.63 to 17.15 for BLH, from 7.59 to 17.75 for FW, and from 8.50 to 11.62 for FY, when comparing stGWAS against mtGWAS (Table 2).

The stGWAS identified a single significant genomic region on LG16, in position 4,178,535 base pairs (bp), associated with ADG, BWH, WW, HON and FW, and a significant SNP on LG07, in position 16,847,179 bp, for FY (Supplementary Figure 1). When combining the summary statistics of all body traits, using a mtGWAS, we identified several novel genomic regions associated with different traits. The number of SNPs surpassing the genome-wide significance threshold ranged from 827 to 1,920 depending on the trait analyzed, with the lowest and the highest number of significant variants associated with BLH and FW (Table 3). The greatest number of significant variants were located on LG03 and LG12 for all traits, except FW where most variants were located on LG13 (Fig 2). The location of significant variants on different chromosomes, and representation of several loci, suggest these body traits are under polygenic control.

Most of the lead SNPs were on LG01, LG03 and LG12 for ADG, BWH, WW, HW, HON and BLH. Some variants were common between body traits, such as two SNPs at positions 24,557,870 and 24,557,984 on LG12, that were the most significant SNPs (p-value < 9.893E-14) common in ADG, BWH, WW, HW, and HON. The lead SNPs for FW and FY were found on LG04 and LG13, and none of those were identified in other body traits (Table 3).
Candidate genes

The full list of genes located within 100 kb upstream and downstream of the lead SNP is available in additional file (Supplementary Table 2). Some lead SNPs for ADG, BWH, WW, HON, BLH are close to potential candidate genes, including *collagen type IV alpha 1 chain* (*COL4A1*) and *growth differentiation factor 6* (*GDF6*) on LG16 and LG22, respectively, and *ankyrin repeat and SOCS box containing 2* (*ASB2*) associated with BWH and HON, located on LG19. The genes intercepted by lead SNPs, located in exonic or intronic regions are shown in Table 4. Some of these genes have been associated with body traits in previous studies. For trait FW, *A disintegrin and metalloproteinase with thrombospondin motifs 9* (*ADAMTS9*), located in LG05, was intercepted by a SNP in an exon region at 29,062,243 bp. Two lead SNPs for WW, located on LG09, at positions 14,670,077 and 14,674,835 bp, were intercepting introns of the gene *solute carrier family 4 member 2* (*SLC4A2*). Intronic regions of *α1,6-fucosyltransferase* (*FUT8*) and the *heart development protein with EGF like domains 1* (*HEG1*), located on LG15 and LG16, were intercepted by lead SNPs associated with ADG and FY, respectively. Two SNPs within *nucleoporin 107* (*NUP107*), located on LG17, were associated with both BWH and HON, on positions 19,609,147 and 19,612,729 bp, respectively, with the first SNP hitting an intronic region and the second one located in an exon region. Others genes such as *Coiled-Coil Domain Containing 102A* (*CCDC102A*), *SLIT-ROBO Rho GTPase Activating Protein 1* (*SRGAP1*), *MutS Homolog 6* (*MSH6*) *Myosin VI* (*MYO6*), *Myosin XVI* (*MYO16*), and *Kinectin 1* (*KTN1*) were intercepted by one or more lead SNPs, but no clear evidence of a close association with body and growth traits has been reported.
Discussion

We found moderate to high heritability values for ADG, BWH, WW, HW, HON, BLH, FW and FY, which is consistent with previous estimates for Nile tilapia calculated using pedigree and genomic methods [8, 9, 20, 21]. The additive genetic variance and heritability estimated for BWH using genotypes imputed to high-density genotypes increased about 15% in comparison to the value previously estimated for the same population using a 50K SNP panel [20].

The use of genomic information can help in the identification of QTLs controlling complex traits which are economically important for aquaculture purposes, such as growth-related traits. Previous studies have identified loci and candidate genes associated with growth-related traits in aquaculture species [20, 23, 24, 26, 37, 38]. However, similar to what we found when using stGWAS (Supplementary Figure 1), few or no markers surpassed the genome-wide significance threshold or represented a small proportion of genetic variance in GWAS. No studies have found evidence of major QTLs for growth-related traits, and GWAS signals were moderate even when a relatively large sample size (>4,600 animals) and more than 100K markers were used, as in the case of GWAS for body weight in Atlantic salmon [23].

To increase the statistical power in order to detect genetic association between SNPs and traits of interest, recent studies have used mtGWAS, which can leverage multiple input summary statistics of the same trait with different measures or different traits with a high genetic correlation [33, 39, 40]. We combined the use of genotypes imputed to high-density genotype and the mtGWAS approach implemented in MTAG software to increase the statistical power and accuracy of QTL detection [33]. The imputation proceeded from a
medium-density (50K) SNP panel to high-density, where the markers from the reference dataset were previously selected based on quality control, and an expected accuracy of imputation higher than 0.80. The mtGWAS increases statistical power by using information from different traits that are genetically correlated with each other [33]. Here, the correlation of the overall SNP effects ranged from 0.86 to 1.00, except for the correlation between FY and all of the other traits, which ranged from 0.20 to 0.47 (Fig 1), and the samples were overlapped for all traits. The better resolution of the genotypes imputed to high-density genotype, combined with the power of the mtGWAS approach, lead to the detection of several novel significant markers not previously found when using stGWAS.

A difference in the number of significant SNPs between stGWAS and mtGWAS is expected given the substantial increase in statistical power which has been documented by the mtGWAS approach. However, it has also been shown that original associations detected by single-trait GWAS can disappear when running multi-trait GWAS. For instance, in the paper describing the application of mtGWAS [32], the increase of significant lead SNPs was from two up to four times higher when comparing mtGWAS against stGWAS. Nevertheless, there were also SNPs associated in the stGWAS analyses which were not found to be associated when running a multi-trait GWAS. If the SNP association is not confirmed by the mtGWAS, we may assume that the previous association identified by the stGWAS is spurious and interpretations on these unconfirmed associations have to be taken with caution.

We found numerous significant markers associated with body traits, dispersed in almost all linkage groups (LG; Fig 2), probably due to the polygenic architecture of these traits. However, a major common association peak on LG12 was found for all traits analyzed, except for FW where the major peak was found on LG13; suggesting that part of the genetic variation that affects body traits might be presented on these linkage groups. Unfortunately,
no one gene was intercepted by the two most significant lead SNPs in this region, but a nearby gene on LG12, *hydroyxysteroid 17-Beta Dehydrogenase 4 (HSD17B4)*, a possible regulator of muscle development in Berkshire pigs, has been reported to play an important role during the early stages of myogenesis when expression of its mRNA is significantly high [41].

Lead SNPs, identified in this study, were located close or intercepted several strong functional candidate genes associated with body and growth traits in previous studies. Some genes were found in windows within 100 kb downstream and upstream from the lead SNP, such as *COL4AI*, located in LG16, associated with different body traits, including ADG, BWH, WW, HON and BLH. In catfish *COL4AI* was identified within QTLs associated with body length and body length of the fish without the head. Collagen is an important component of the extracellular matrix of cartilage and bone, playing a key role in skeletal development [42]. We also found *GDF6*, located in LG22, was associated with different traits, including ADG, BWH, WW, HON, BLH. *GDF6* in zebrafish is related with reduced eye size and different skeletal defects [43]. In a study to compare the orthologous sequences from 14 species (including human, mice, livestock, fugu, and zebrafish), the *GDF6* gene was found to control developmental patterning of skeletal joints [44]. Inactivation of the *GDF6* gene can cause defects in the joints, ligaments, and cartilage formation in mouse [45]. *ASB2*, located in chromosome 19, was associated with BWH and HON. In Atlantic salmon, the *ASB2* gene is not involved in muscle differentiation but may play an important role in growth inhibition. The high expression of *ASB2* observed skeletal muscle of fasting fish, is strongly downregulated in response to feeding [46].

We also found strong candidate genes intercepted by lead SNPs that may contribute to a better understanding of the biological mechanisms controlling body traits in Nile tilapia. Growth is considered a continuous function during the life of an animal and ADG is an
important trait that plays an essential role in rapid growth. ADG was previously reported as a selection criteria in a breeding program for Nile tilapia, which has been applying selection for at least five generations [47]. We found a lead SNP associated with ADG on chromosome LG15, located in an intronic region of the FUT8 gene, which has been associated with ADG (from birth to six months-age) in a sheep population from Irian [48]. In mice, the disruption of the FUT8 gene induces severe growth retardation and early mortality during postnatal development [49–52]. The insulin-like growth factor binding protein-3 (IGFBP-3) has growth inhibitory effects, and the alteration in the function of low-density lipoprotein receptor-related protein-1 (LRP-1) is a result of the loss of core fucosylation that might cause an elevated serum concentration of IGFBP-3 in FUT8-null mice [52]. The loss of function of FUT8 has also been reported to be related to down-regulation of transforming growth factor-beta 1 (TGF-β1) receptor and epidermal growth factor (EGF) receptor, proteinase-activated receptor and integrin activity, which contributes to emphysema-like changes in the lung, and growth retardation in FUT8-null mice [51].

Two lead SNPs associated with BWH and HON were found on LG17, in an intronic and exonic region of the NUP107 gene that plays an important role in the development of vertebrate embryos. The zygotic deficiency of NUP107 in zebrafish embryos can result in loss of pharyngeal skeletons, degeneration of intestinal and retinal epithelia, and implications in cartilage and bone formation [53]. In fibroblasts of senescent humans, and organs of aged mice, a decreased level of suggested the attenuation of hyporesponsiveness to growth [54].

The waste weight is the sum of the weight of the head, viscera, bones, and fins, and has been suggested as a phenotypic record to improve fillet yield through the application of various index (e.g. fillet to waste ratio). However, based on simulated data of ten generations of selection using real genetic parameters of five real farmed fish populations, direct selection
on fillet yield was generally the best approach to improve fillet yield [13]. The potential limitation for selecting against waste weight is the potential for decreasing the volume of essential organs in the visceral cavity. A negative genetic correlation (< -0.52) between fillet yield and head, and bone development has been reported in rainbow trout [55]. We found a lead SNP that intercepts the SLC4A2 gene, a strong biological candidate for waste weight in Nile tilapia. The loss of this gene causes emaciation and achlorhydric [56], generating growth retardation, reduced osteoclast numbers and/or a reduction in osteoclast activity, resulting in osteopetrosis in mice [57]. Osteopetrosis is a skeletal disorder that can affect humans and animals, characterized by the formation of overly dense bones [57, 58]. In Red Angus cattle, a deletion mutation in SLC4A2 is associated with an osteopetrosis phenotype [58].

Fillet traits are key economic traits for aquaculture species and new insights regarding the underlying genetic variants controlling them can help to enhance yield. We found two lead SNPs associated with FW and FY, intercepting an exonic and intronic region of genes ADAMTS9 and HEG1, respectively. The ADAMTS9 gene is highly expressed during embryo development and continues to be expressed in adult tissues of mice [59, 60]. A significant expression of ADAMTS9 during skeletal development of mouse was suggested by Jungers et al. (2005), including mandible, ossification centers, initial condensation of mesenchyme to form the cartilage centers, the perichondrium around formed cartilage, in the proliferative zones of cartilage and long bones. Skeletal development may be correlated with organic growth [61]. ADAMTS9 is responsible for the regulation of the epidermal growth factor receptor (EGFR) and TGF-β1 [62, 63]. Tang et al. (2019) identified a 22-bp indel in ADAMTS9 associated with chest width, chest width index, and chest circumference index, and 14-bp indel was associated with height across the hip in cashmere goats, which suggests that ADAMTS9 could be a molecular marker to improve goat growth traits [62].
A lead SNP associated with FY intercepted the *HEG1* gene, located in LG16. The *HEG1* gene was initially reported to be responsible for regulating zebrafish heart growth, and is fundamental for the development of heart and blood vessels [65]. Recently the *HEG1* gene was identified as one of several novel genes associated with human skeletal muscle growth, exhibiting a significant correlation with the percentage of change in lean mass [66, 67]. In a comparative transcriptomic analysis to identify differentially expressed genes related to product performance and meat quality from the longissimus dorsi in sheep, Cheng et al. (2020) identified six different expressed genes, including *HEG1*, commonly expressed in the three biological processes in sheep that were related to growth, development, and meat quality [66].

Some genes such as *MSH6*, *SRGAP1*, *MYO6*, *MYO16*, *KTN1* and other genes presented in Table 4, intercepted by lead SNPs, are thought to be involved in different biological functions, such as some types of cancer [69, 70], hearing loss [71] and schizophrenia [72]. The mutation of *MSH6*, for example, may increase the risk of developing colorectal carcinomas [73, 74], and *MYO16* appears to have an important role in neural development and the function of the nervous system [75]. The functionality of these genes with growth traits in Nile tilapia is unclear and suggests that some of the significant SNPs found, and the function of the identified genes must be better characterized.

**Conclusions**

We used fine-mapping association analysis for body traits in Nile tilapia and found that mtGWAS provided substantial improvements in the number of significant SNPs identified when compared to stGWAS. These results confirm the increase of statistical power.
to identify trait-specific genetic associations in multi-trait analysis. Interestingly, we found many lead SNPs within or nearby genes related to cartilage, bone, skeletal growth and development in humans, mice, livestock, and aquaculture species. These results can provide further knowledge and a better understanding of genetic variants and genes underlying complex body traits in Nile tilapia.

**Material and Methods**

**Animals and phenotypes**

We used a total of 1,309 phenotyped animals from 72 families (mean = 18, minimum = 7, and maximum = 25 animals per family) belonging to a breeding nucleus owned by Aquacorporación International group (ACI), Costa Rica. More details about the breeding program, the origin of the Nile tilapia population and production conditions are described in detail in previous studies [18, 20, 76]. Briefly, a mating design of two dams per sire was used to produce the 72 full-sib families. The eggs of each full-sib family were incubated and reared in separate hapas until individual tagging by using PIT (passive Integrated Transponder)-tags at an average weight and age of 13 g (SD = 8 g) and 104 days (SD = 18 days), respectively. After tagging, the fish were grown in excavated ponds for about 370 days until harvest. All animals were slaughtered by hypothermia in ice slurry at commercial plant, and different body traits were measured at harvest time and in the processing plant: body weight at harvest (BWH in g), fillet weight (FW in g), waste weight (WW in g = BWH − FW), head weight (HW in g), hon weight (HON in g = BWH − (HW + Viscera)), body length at harvest (BLH in cm), average daily gain (ADG in g = (BWH - bodyweight at tagging)/(age at harvest - age at tagging)), and fillet yield (FY in % = FW/BHW*100).
Genotypes and imputation to whole-genome sequences

Genomic DNA was extracted and purified from 1,309 fin clip samples using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer’s protocol. The samples were genotyped using a 50K SNP Illumina BeadChip [18] and filtered using Hardy-Weinberg Disequilibrium (HWE, p-value $10^{-6}$), minor allele frequency (MAF < 0.01), and a genotyping call-rate for SNPs and samples of <0.90. After quality control 29,587 SNPs and 1,309 samples were retained.

Initially 26.6 million non-redundant variants were identified through Illumina HiSeq 2500 re-sequencing from 143 animals, were obtained from breeding nucleus owned by Aquacorporación International group (ACI), Costa Rica, for imputation of medium-density genotypes to WGS [18]. Quality control was performed using the following thresholds: HWE (p-value $10^{-8}$), MAF < 0.01 and call-rate for SNPs < 0.80. A total of 5,011,051 SNPs were retained after applying the quality control. In order to estimate the overall accuracy of imputation and remove the variants with low imputation accuracy we used a five-fold cross validation scheme. Briefly, the 143 animals with data from the WGS-derived genotypes were randomly divided into five exclusive reference sets (80% of animals genotyped with ~5 million SNP genotypes) and the remaining animals were used as the validation set (20% of animals with medium-density genotypes). The accuracy of imputation was estimated as the correlation between true and imputed genotypes ($R^2$ value). A total of 1,324,420 SNPs with $R^2$ value higher than 0.80 were used as the final ultra-dense SNP panel for imputation. The 143 animals and 1,324,420 SNPs were used as a reference dataset to impute the 1,309 animals with medium-density SNP genotypes using the software FImpute v. 3.0 [77]. A post-
imputation quality control excluded SNPs with MAF < 0.05 and HWE p-value < 10^{-8}, resulting in a total of 992,494 SNPs available for downstream analyses.

**Single-trait genome-wide association**

The single-trait genome wide association analyses (stGWAS) were performed using the *mlma* option of the software GCTA v. 1.24 [78], which was used to apply the following linear mixed model:

\[
y_{ij} = \mu + b_1 \cdot \text{age}_j + b_2 \cdot \text{SNP}_i + a_{ij} + e_{ij}
\]  

where \(y_{ij}\) is the phenotypic value of the \(j\)-th animal, \(\mu\) is the fixed effect of the overall mean, \(b_1\) and \(b_2\) are the regression coefficients for age and the allele substitution effect for SNP, respectively, \(\text{age}_j\) is the age covariate of animal \(j\)-th and \(\text{SNP}_i\) is the \(i\)-th SNP genotype of animal \(j\), coded as 0, 1 and 2 for genotype \(A_1A_1\), \(A_1A_2\) and \(A_2A_2\), respectively, \(a_{ij}\) is the vector of random polygenic effect of the \(j\)-th animal \(\sim N(0, G \sigma_a^2)\), with \(G\) the genomic relationship matrix (GRM) calculated using the imputed genotypes [78, 79] and \(e_{ij}\) is the vector of random residual effect \(\sim N(0, I \sigma_e^2)\), with \(I\) an identity matrix and \(\sigma_e^2\) the residual variance. The GRM is calculated based on the relationship from a genome-wide sample of SNPs, combined using a common-sense weighting scheme [78]. The GRM restricted maximum likelihood (GREML) [78] implemented in GCTA was used to estimate the genetic and residual variances. Heritability \((h^2)\) was calculated as \(h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)\). For each SNP, the allele substitution effect and its p-value were also estimated using GCTA.
Multi-trait genome-wide association

The summary statistics from stGWAS was used as input for the multi-trait analysis of GWAS (mtGWAS) using the software MTAG v0.9.0 [33]. In MTAG, the SNP effect estimated for each trait can be improved when different traits that are correlated are included in the analysis. This multi-trait approach can increase the power to detect loci in any one of the traits assessed. The first step of MTAG is to filter variants based on discarding non common SNPs, duplicated SNPs, or SNPs with strand ambiguity. In our study, out of the 992,494 SNPs available after imputation and initial quality control, a total of 183,401 SNPs with strand ambiguity were filtered out. The remaining 809,093 SNPs were used for mtGWAS analyses. A bivariate linkage disequilibrium (LD) score regression is used so summary statistics do not need to come from independent samples [33]. The MTAG output consists of a file per trait with updated results of SNP effects and p-values from a mtGWAS, which can be interpreted in the same way as stGWAS. Significance thresholds were determined for both single-trait and mtGWAS using Bonferroni correction (0.05/ number of SNPs).

To calculate how much larger the stGWAS sample size would have to be to give the same mean $\chi^2$ statistics than mtGWAS, the following equation was used [33]:

$$N_{\text{GWAS equivalent}} = N_{\text{GWAS}} \frac{\chi^2_{\text{mtGWAS}}}{\chi^2_{\text{stGWAS}}}$$

(2)

where, $\chi^2_{\text{mtGWAS}}$ and $\chi^2_{\text{stGWAS}}$ are the mean $\chi^2$ statistic for mtGWAS and stGWAS results, respectively, and $N_{\text{GWAS}}$ is the number of actual sample size in stGWAS (1,309 animals).
Identification of QTL and candidate genes

The most significant SNPs per chromosome per each trait detected, using mtGWAS, was selected as the lead SNP, and further used to search for candidate genes based on proximity to the variant. Genes located within 100 kb upstream and downstream of the lead SNP were considered putative candidate genes associated with the trait. The gene search was performed using BLAST (Basic Local Alignment Search Tool) against the latest version of the *Oreochromis niloticus* reference genome (*O_niloticus_UMD_NMBU [15]*) , which is publicly available at NCBI (GenBank assembly accession GCA_001858045.3).

Abbreviations

**ADAM12**: disintegrin and metalloproteinase domain 12; **ADAMTS9**: A disintegrin and metalloproteinase with thrombospondin motifs 9; **ADG**: average daily gain; **ASB2**: ankyrin repeat and SOCS box containing 2; **BLAST**: Basic Local Alignment Search Tool; **BLH**: harvest length; **bp**: base pairs; **BWH**: harvest weight; **CCDC102A**: Coiled-Coil Domain Containing 102A; **COL4A1**: collagen type IV alpha 1 chain; **EGF**: receptor and epidermal growth factor; **EGFR**: epidermal growth factor receptor; **FGF**: fibroblast growth factors; **FUT8**: α1,6-fucosyltransferase; **FW**: fillet weight; **FY**: fillet yield; **GDF6**: growth differentiation factor 6; **GREML**: genomic-relationship matrix restricted maximum likelihood; **GRM**: genomic-relationship matrix; **GWAS**: genome-wide association studies; **HEG1**: heart development protein with EGF like domains 1; **HON**: hon weight; **HSD17B4**: hydroxysteroid 17-Beta Dehydrogenase 4; **HW**: head weight; **HWE**: Hardy-Weinberg Disequilibrium; **IGFBP-3**: insulin-like growth factor binding protein-3; **KTN1**: Kinectin 1; **LD**: linkage disequilibrium; **LG**: linkage group; **LRP-1**: low-density lipoprotein receptor-
related protein-1; MAF: Minor Allele Frequency; MAS: marker-assisted selection; MEP1A: meprin A subunit beta-like; MSH6: MutS Homolog 6; mtGWAS: multi-trait genome-wide association study; MYLK: myosin light chain kinase; MYO16: Myosin XVI; MYO6: Myosin VI; NUP107: nucleoporin 107; QTL: quantitative trait loci; SLC4A2: solute carrier family 4 member 2; SNP: single nucleotide polymorphisms; SRGAP1: SLIT-ROBO Rho GTPase Activating Protein 1; stGWAS: single-trait genome-wide association studies; TGFBR3: transforming growth-factor beta receptor type 3; TGF-β1: transforming growth factor-beta 1; WGS: whole-genome sequence; WW: waste weight.

**Declarations**

**Acknowledgements**

We are grateful to Acuacorporación Internacional for providing the Nile tilapia dataset.

**Authors’ contributions**

GMY participated in the study design, performed the analyses and drafted the manuscript. JMY conceived and designed the study; contributed to the analysis, discussion and writing. Both authors have reviewed and approved the manuscript.

**Funding**

This research was financially supported by Production Development Corporation (CORFO project number 14EIAT-28667) a Chilean governmental organization. GMY is supported by Fondecyt/Conicyt Postdoctoral Grant n. 3190553, and JMY is grant supported by Núcleo Milenio INVASAL funded by Chile's government program, Iniciativa Cientifica Milenio
from Ministerio de Economía, Fomento y Turismo. The funders had no role in the design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Availability of data and materials**

The datasets generated and/or analysed during the current study are available in the Figshare repository, [https://figshare.com/s/9b265a22b7e138c5a839](https://figshare.com/s/9b265a22b7e138c5a839) and [https://figshare.com/s/1fa22386fd5bae0366e0](https://figshare.com/s/1fa22386fd5bae0366e0). The *Oreochromis niloticus* reference genome is publicly available at NCBI (GenBank assembly accession GCA_001858045.3, [https://www.ncbi.nlm.nih.gov/genome/?term=GCA_001858045.3](https://www.ncbi.nlm.nih.gov/genome/?term=GCA_001858045.3)).

**Ethics approval and consent to participate**

Nile tilapia sampling procedures were approved by the Comité de Bioética Animal from the Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile (Certificate N° 18179-VET-UCH).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.
References

1. FAO. Fisheries and aquaculture software. FishStatJ - Software for Fishery and Aquaculture Statistical Time Series. 2019. http://www.fao.org/fishery/.

2. Gjedrem T. Genetic improvement for the development of efficient global aquaculture: A personal opinion review. Aquaculture. 2012;344–349:12–22. doi:10.1016/j.aquaculture.2012.03.003.

3. Bentsen HB, Gjerde B, Nguyen NH, Rye M, Ponzoni RW, Palada de Vera MS, et al. Genetic improvement of farmed tilapias: Genetic parameters for body weight at harvest in Nile tilapia (Oreochromis niloticus) during five generations of testing in multiple environments. Aquaculture. 2012;338–341:56–65. doi:10.1016/j.aquaculture.2012.01.027.

4. Khaw HL, Bovenhuis H, Ponzoni RW, Rezk MA, Charo-Karisa H, Komen H. Genetic analysis of Nile tilapia (Oreochromis niloticus) selection line reared in two input environments. Aquaculture. 2009;294:37–42. doi:10.1016/j.aquaculture.2009.05.025.

5. Ponzoni RW, Hamzah A, Tan S, Kamaruzzaman N. Genetic parameters and response to selection for live weight in the GIFT strain of Nile Tilapia (Oreochromis niloticus). Aquaculture. 2005;247:203–10. doi:10.1016/j.aquaculture.2005.02.020.

6. Rezk MA, Ponzoni RW, Khaw HL, Kamel E, Dawood T, John G. Selective breeding for increased body weight in a synthetic breed of Egyptian Nile tilapia, Oreochromis niloticus: Response to selection and genetic parameters. Aquaculture. 2009;293:187–94. doi:10.1016/j.aquaculture.2009.03.019.

7. Rutten MJM, Bovenhuis H, Komen H. Genetic parameters for fillet traits and body
measurements in Nile tilapia (Oreochromis niloticus L.). Aquaculture. 2005;246:125–32. doi:10.1016/J.AQUACULTURE.2005.01.006.

8. Thodesen (Da-Yong Ma) J, Rye M, Wang Y-X, Yang K-S, Bentsen HB, Gjedrem T. Genetic improvement of tilapias in China: Genetic parameters and selection responses in growth of Nile tilapia (Oreochromis niloticus) after six generations of multi-trait selection for growth and fillet yield. Aquaculture. 2011;322–323:51–64. doi:10.1016/J.AQUACULTURE.2011.10.010.

9. Nguyen NH, Ponzoni RW, Abu-Bakar KR, Hamzah A, Khaw HL, Yee HY. Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), Oreochromis niloticus. Aquaculture. 2010;305:1–5. doi:10.1016/J.AQUACULTURE.2010.04.007.

10. Gjerde B, Mengistu SB, Ødegård J, Johansen H, Altamirano DS. Quantitative genetics of body weight, fillet weight and fillet yield in Nile tilapia (Oreochromis niloticus). Aquaculture. 2012;342–343:117–24.

11. Powell J, White I, Guy D, Brotherstone S. Genetic parameters of production traits in Atlantic salmon (Salmo salar). Aquaculture. 2008;274:225–31.

12. Yáñez JM, Joshi R, Yoshida GM. Genomics to accelerate genetic improvement in tilapia. Anim Genet. 2020::age.12989. doi:10.1111/age.12989.

13. Fraslin C, Dupont-Nivet M, Haffray P, Bestin A, Vandeputte M. How to genetically increase fillet yield in fish: New insights from simulations based on field data. Aquaculture. 2018;486:175–83.

14. Bosworth B, Waldbieser G, Garcia A, Tsuruta S, Lourenco D. Heritability and response to selection for carcass weight and growth in the Delta Select strain of channel catfish, Ictalurus punctatus. Aquaculture. 2020;515:734507.
15. Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerdinger WJ, Roberts RB, et al. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. Gigascience. 2019;8:giz030. doi:10.1093/gigascience/giz030.

16. Lu S, Zhu J, Du X, Sun S, Meng L, Liu S, et al. Genomic selection for resistance to Streptococcus agalactiae in GIFT strain of Oreochromis niloticus by GBLUP, wGBLUP, and BayesCπ. Aquaculture. 2020;523:735212.

17. Cáceres G, López ME, Cádiz MI, Yoshida GM, Jedlicki AM, Palma-Véjares R, et al. Fine Mapping Using Whole-Genome Sequencing Confirms Anti-Müllerian Hormone as a Major Gene for Sex Determination in Farmed Nile Tilapia (Oreochromis niloticus L.). G3 (Bethesda). 2019::g3.400297.2019. doi:10.1534/g3.119.400297.

18. Yáñez JM, Yoshida G, Barria A, Palma-Véjares R, Travisany D, Díaz D, et al. High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Validation Through Whole-Genome Resequencing in Nile Tilapia (Oreochromis niloticus). Mar Biotechnol. 2020.

19. Joshi R, Árnyasi M, Lien S, Gjøen HM, Alvarez AT, Kent M. Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile Tilapia (O. niloticus). Front Genet. 2018;9:472. doi:10.3389/fgene.2018.00472.

20. Yoshida G, Llorente J, Correa K, Soto J, Salas D, Yáñez J. Genome-Wide Association Study and Cost-Efficient Genomic Predictions for Growth and Fillet Yield in Nile Tilapia (Oreochromis niloticus ). G3 Genes|Genomes|Genetics. 2019;9:g3.400116.2019. doi:10.1534/g3.119.400116.

21. Joshi R, Skaarud A, de Vera M, Alvarez AT, Ødegård J. Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (Oreochromis niloticus). Aquaculture. 2020;516:734641.
22. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol. 2012;8.

23. Yoshida GM, Lhorente JP, Carvalheiro R, Yáñez JM. Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (Salmo salar L.). Anim Genet. 2017;48:698–703.

24. Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. BMC Genomics. 2015;16:1–9. doi:10.1186/s12864-015-2117-9.

25. Geng X, Liu S, Yao J, Bao L, Zhang J, Li C, et al. A genome-wide association study identifies multiple regions associated with head size in catfish. G3 Genes, Genomes, Genet. 2016;6:3389–98.

26. Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S, et al. Identification of novel genes significantly affecting growth in catfish through GWAS analysis. Mol Genet Genomics. 2018;293:587–99.

27. Sanchez MP, Govignon-Gion A, Croiseau P, Fritz S, Hozé C, Miranda G, et al. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genet Sel Evol. 2017;49:1–16.

28. Van Binsbergen R, Bink MCAM, Calus MPL, Van Eeuwijk FA, Hayes BJ, Hulsegge I, et al. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. Genet Sel Evol. 2014;46:1–13.

29. Wu P, Wang K, Zhou J, Chen D, Yang Q, Yang X, et al. GWAS on Imputed Whole-Genome Resequencing From Genotyping-by-Sequencing Data for Farrowing Interval of Different Parities in Pigs. Front Genet. 2019;10:1012. doi:10.3389/fgene.2019.01012.

30. Van Den Berg S, Vandenplas J, Van Eeuwijk FA, Bouwman AC, Lopes MS, Veerkamp
RF. Imputation to whole-genome sequence using multiple pig populations and its use in genome-wide association studies. Genet Sel Evol. 2019;51:2. doi:10.1186/s12711-019-0445-y.

31. Al Kalaldeh M, Gibson J, Duijvesteijn N, Daetwyler HD, MacLeod I, Moghaddar N, et al. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. Genet Sel Evol. 2019;51:1–13.

32. Crispim AC, Kelly MJ, Guimarães SEF, E Silva FF, Fortes MRS, Wenceslau RR, et al. Multi-trait GWAS and new candidate genes annotation for growth curve parameters in brahman cattle. PLoS One. 2015;10.

33. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet. 2018;50:229–37.

34. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50:1112–21.

35. Hill WD, Marioni RE, Maghzian O, Ritchie SJ, Hagenaars SP, McIntosh AM, et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Mol Psychiatry. 2019;24:169–81.

36. Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC, et al. Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. Cell Rep. 2017;21:2597–613.

37. Reis Neto RV, Yoshida GM, Lhorente JP, Yáñez JM. Genome-wide association analysis for body weight identifies candidate genes related to development and metabolism in rainbow trout (Oncorhynchus mykiss). Mol Genet Genomics. 2019;;1–9. doi:10.1007/s00438-018-
1518-2.

38. Gutierrez AP, Yáñez JM, Fukui S, Swift B, Davidson WS. Genome-Wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic Salmon (Salmo salar). PLoS One. 2015;10:e0119730. doi:10.1371/journal.pone.0119730.

39. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.

40. Han X, Gharahkhani P, Mitchell P, Liew G, Hewitt AW, MacGregor S. Genome-wide meta-analysis identifies novel loci associated with age-related macular degeneration. J Hum Genet. 2020;1–9.

41. Jo JL, Hwang JH, Kwon SG, Park DH, Kim TW, Kang DG, et al. Association between a non-synonymous HSD17B4 single nucleotide polymorphism and meat-quality traits in Berkshire pigs. Genet Mol Res. 2016;15. doi:10.4238/gmr15048970.

42. Olsen BR, Reginato AM, Wang W. Bone Development. Annu Rev Cell Dev Biol. 2000;16:191–220. doi:10.1146/annurev.cellbio.16.1.191.

43. den Hollander AI, Biyanwila J, Kovach P, Bardakjian T, Traboulsi EI, Ragge NK, et al. Genetic defects of GDF6 in the zebrafish out of sight mutant and in human eye developmental anomalies. BMC Genet. 2010;11:102. doi:10.1186/1471-2156-11-102.

44. Portnoy ME, McDermott KJ, Antonellis A, Margulies EH, Prasad AB, Kingsley DM, et al. Detection of potential GDF6 regulatory elements by multispecies sequence comparisons and identification of a skeletal joint enhancer. Genomics. 2005;86:295–305.

45. Settle SH, Rountree RB, Sinha A, Thacker A, Higgins K, Kingsley DM. Multiple joint and skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5 genes. Dev Biol. 2003;254:116–30.

46. Bower NI, Johnston IA. Discovery and characterization of nutritionally regulated genes
associated with muscle growth in Atlantic salmon. Physiol Genomics. 2010;42 A:114–30.

47. de Oliveira CAL, Ribeiro RP, Yoshida GM, Kunita NM, Rizzato GS, de Oliveira SN, et al. Correlated changes in body shape after five generations of selection to improve growth rate in a breeding program for Nile tilapia Oreochromis niloticus in Brazil. J Appl Genet. 2016;57.

48. Pasandideh M, Rahimi-Mianji G, Gholizadeh M. A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. J Genet. 2018;97:493–503.

49. Wang X, Inoue S, Gu J, Miyoshi E, Noda K, Li W, et al. Dysregulation of TGF-β1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. Proc Natl Acad Sci U S A. 2005;102:15791–6.

50. Wang X, Gu J, Miyoshi E, Honke K, Taniguchi N. Phenotype Changes of Fut8 Knockout Mouse: Core Fucosylation Is Crucial for the Function of Growth Factor Receptor(s). Methods in Enzymology. 2006;417:11–22.

51. Wang X, Fukuda T, Li W, Gao C-X, Kondo A, Matsumoto A, et al. Requirement of Fut8 for the expression of vascular endothelial growth factor receptor-2: a new mechanism for the emphysema-like changes observed in Fut8-deficient mice. J Biochem. 2009;145:643–51.

52. Lee SH, Takahashi M, Honke K, Miyoshi E, Osumi D, Sakiyama H, et al. Loss of core fucosylation of low-density lipoprotein receptor-related protein-1 impairs its function, leading to the upregulation of serum levels of insulin-like growth factor-binding protein 3 in Fut8-/- mice. J Biochem. 2006;139:391–8.

53. Zheng X, Yang S, Han Y, Zhao X, Zhao L, Tian T, et al. Loss of zygotic NUP107 protein causes missing of pharyngeal skeleton and other tissue defects with impaired nuclear pore function in zebrafish embryos. J Biol Chem. 2012;287:38254–64.
54. Kim SY, Kang HT, Choi HR, Park SC. Reduction of Nup107 attenuates the growth factor signaling in the senescent cells. Biochem Biophys Res Commun. 2010;401:131–6.

55. Haffray P, Bugeon JÔ, Pincent C, Chapuis H, Mazeiraud E, Rossignol MN, et al. Negative genetic correlations between production traits and head or bony tissues in large all-female rainbow trout (Oncorhynchus mykiss). Aquaculture. 2012;368–369:145–52.

56. Gawenis LR, Ledoussal C, Judd LM, Prasad V, Alper SL, Stuart-Tilley A, et al. Mice with a targeted disruption of the AE2 Cl-/HCO 3- exchanger are achlorhydric. J Biol Chem. 2004;279:30531–9.

57. Wu J, Glimcher LH, Aliprantis AO. HCO3-/Cl- anion exchanger SLC4A2 is required for proper osteoclast differentiation and function. Proc Natl Acad Sci U S A. 2008;105:16934–9.

58. Meyers SN, McDaneld TG, Swist SL, Marron BM, Steffen DJ, O’Toole D, et al. A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. BMC Genomics. 2010;11:337. doi:10.1186/1471-2164-11-337.

59. Clark ME, Kelner GS, Turbeville LA, Boyer A, Arden KC, Maki RA. ADAMTS9, a novel member of the ADAM-TS/metallopondin gene family. Genomics. 2000;67:343–50.

60. Jungers KA, Le Goff C, Somerville RPT, Apte SS. Adamts9 is widely expressed during mouse embryo development. Gene Expr Patterns. 2005;5:609–17.

61. González-Cerón F, Rekaya R, Aggrey SE. Genetic analysis of bone quality traits and growth in a random mating broiler population - ScienceDirect. Poult Sci. 2015;94:883–9.

62. Shao B, Feng Y, Zhang H, Yu F, Li Q, Tan C, et al. The 3p14.2 tumour suppressor ADAMTS9 is inactivated by promoter CpG methylation and inhibits tumour cell growth in breast cancer. J Cell Mol Med. 2018;22:1257–71.

63. Wang Q, Hao R, Zhao X, Huang R, Zheng Z, Deng Y, et al. Identification of EGFR in
pearl oyster (Pinctada fucata martensii) and correlation analysis of its expression and growth traits. Biosci Biotechnol Biochem. 2018;82:1073–80.

64. Tang Q, Zhang X, Wang X, Wang K, Yan H, Zhu H, et al. Detection of two insertion/deletions (indels) within the ADAMTS9 gene and their associations with growth traits in goat. Small Rumin Res. 2019;180:9–14.

65. Kleaveland B, Zheng X, Liu JJ, Blum Y, Tung JJ, Zou Z, et al. Regulation of cardiovascular development and integrity by the heart of glass-cerebral cavernous malformation protein pathway. Nat Med. 2009;15:169–76.

66. Seaborne RA, Strauss J, Cocks M, Shepherd S, O’Brien TD, Van Someren KA, et al. Human Skeletal Muscle Possesses an Epigenetic Memory of Hypertrophy. Sci Rep. 2018;8:1–17.

67. Seaborne RA, Strauss J, Cocks M, Shepherd S, O’Brien TD, van Someren KA, et al. Methylome of human skeletal muscle after acute & chronic resistance exercise training, detraining & retraining. Sci Data. 2018;5:1–9.

68. Cheng S, Wang X, Zhang Q, He Y, Zhang X, Yang L, et al. Comparative Transcriptome Analysis Identifying the Different Molecular Genetic Markers Related to Production Performance and Meat Quality in Longissimus Dorsi Tissues of MG × STH and STH Sheep. Genes (Basel). 2020;11:183. doi:10.3390/genes11020183.

69. Lei C, Du F, Sun L, Li T, Li T, Min Y, et al. MIR-143 & MIR-145 inhibit gastric cancer cell migration & metastasis by suppressing MYO6. Cell Death Dis. 2017;8:e3101–e3101. doi:10.1038/cddis.2017.493.

70. He H, Bronisz A, Liyanarachchi S, Nagy R, Li W, Huang Y, et al. SRGAP1 Is a Candidate Gene for Papillary Thyroid Carcinoma Susceptibility. J Clin Endocrinol Metab. 2013;98:E973–80. doi:10.1210/jc.2012-3823.
71. Melchionda S, Ahituv N, Bisceglia L, Sobe T, Glaser F, Rabionet R, et al. MYO6, the human homologue of the gene responsible for deafness in Snell’s waltzer mice, is mutated in autosomal dominant nonsyndromic hearing loss. Am J Hum Genet. 2001;69:635–40.

72. Rodriguez-Murillo L, Xu B, Roos JL, Abecasis GR, Gogos JA, Karayiorgou M. Fine mapping on chromosome 13q32-34 and brain expression analysis implicates MYO16 in schizophrenia. Neuropsychopharmacology. 2014;39:934–43. doi:10.1038/npp.2013.293.

73. Shia J, Zhang L, Shike M, Guo M, Stadler Z, Xiong X, et al. Secondary mutation in a coding mononucleotide tract in MSH6 causes loss of immunoexpression of MSH6 in colorectal carcinomas with MLH1/PMS2 deficiency. Mod Pathol. 2013;26:131–8. doi:10.1038/modpathol.2012.138.

74. Hendriks YMC, Wagner A, Morreau H, Menko F, Stormorken A, Quehenberger F, et al. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: Impact on counseling and surveillance. Gastroenterology. 2004;127:17–25.

75. Bugyi B, Kengyel A. Myosin XVI. In: Advances in Experimental Medicine and Biology. Springer; 2020. p. 405–19. doi:10.1007/978-3-030-38062-5_18.

76. Yoshida GM, Barria A, Cáceres G, Correa K, Jedlicki A, Cadiz MI, et al. Genome-wide patterns of population structure and linkage disequilibrium in farmed Nile tilapia (Oreochromis niloticus). Front Genet. 2019;10:745. doi:10.3389/FGENE.2019.00745.

77. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478. doi:10.1186/1471-2164-15-478.

78. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

79. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common
SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
Supplementary Information

Additional file 1: Supplementary Figure 1. Manhattan plot for single-trait GWAS (stGWAS) for body traits in Nile tilapia. Manhattan plots of SNPs associated with: (A) Average daily gain. (B) Body weight at harvest. (C) Waste weight. (D) Head weight. (E) Hon weight. (F) Body length at harvest. (G) Fillet weight. (H) Fillet yield. The x-axis presents genomic coordinates along chromosomes 1–23 in Nile tilapia. On the y-axis the negative logarithm of the SNPs associated p-value is displayed. All SNPs above the dashed black line marks the genome-wide significant threshold associated with the trait after Bonferroni correction (−log10 (p-value > 7.30e-8)).

Additional file 2: Supplementary Table 1. Summary results from genotype quality control of whole-genome sequence (WGS), imputed WGS, and 50K single nucleotide polymorphism (SNP) chips for Nile tilapia.

| Parameters                      | Genotypes data sets |
|---------------------------------|---------------------|
|                                 | WGS\textsuperscript{1} | Imputed WGS\textsuperscript{2} | 50K\textsuperscript{3} |
| N samples                       | 143                 | 1,309                      | 1,309                   |
| Initial SNPs                   | 26,650,009          | 1,324,420                  | 43,271                  |
| Minor allele frequency         | 7,788,652           | 253,076                    | 4,114                   |
| Call-rate                      | 12,061,149          | -                          | 3,665                   |
| Hardy-Weinberg equilibrium     | 1,789,157           | 78,850                     | 5,905                   |
| Final SNPs                     | 5,011,051           | 992,494                    | 29,587                  |
Minor allele frequency (MAF) < 0.01, call-rate < 0.80 and Hardy-Weinberg equilibrium (HWE) < 1e-8; MAF < 0.05 and HWE < 1e-8; MAF < 0.01, call-rate <0.80 and HWE < 1e-6.

Additional file 3: Supplementary Table 2. Genomic regions and candidate genes for all lead SNP associated with body traits based on multi-trait GWAS for Nile tilapia.

| Marker | LG¹ | Position² | Allele | MAF³ | p-value  | Genes⁴ |
|--------|-----|-----------|--------|------|----------|--------|
| 1:39153024 | 1 | 39153024 | [G/A] | 0.052 | 2.82E-11 | HDGFL3, CCDC102A |
| 1:39193509 | 1 | 39193509 | [A/G] | 0.052 | 2.82E-11 | HDGFL3, CCDC102A, NADSYN1 |
| 1:3958113 | 1 | 39558113 | [G/A] | 0.052 | 2.82E-11 | CCNE1, ZNF536 |
| 1:39628599 | 1 | 39628599 | [A/G] | 0.052 | 2.82E-11 | ZNF536 |
| 3:47327036 | 3 | 47327036 | [A/G] | 0.113 | 6.77E-11 | Uncharacterized |
| 4:24859012 | 4 | 24859012 | [C/A] | 0.059 | 3.38E-11 | USP31, HS3ST4 |
| 5:6037462 | 5 | 6037462 | [T/G] | 0.055 | 6.15E-09 | DDIT3, MARS, ARHGAP9, GLI1, R3HDM2 |
| 6:26052504 | 6 | 26052504 | [A/C] | 0.060 | 1.93E-08 | Uncharacterized |
| 7:54783508 | 7 | 54783508 | [A/G] | 0.051 | 8.36E-09 | GTF3C6, FAM107B |
| 8:396284 | 8 | 396284 | [T/C] | 0.108 | 3.92E-09 | Uncharacterized |
| 9:16267509 | 9 | 16267509 | [C/T] | 0.077 | 1.68E-10 | PLXDC2, MALRD1, WDR60, ESYT2 |
| 9:16435917 | 9 | 16435917 | [C/T] | 0.075 | 4.76E-09 | WDR60, ESYT2, NCAPG2, PTPRN2 |
| 12:24557870 | 12 | 24557870 | [A/G] | 0.069 | 2.63E-15 | HSD17B4, DMXL1, SEMA6A |
| 12:24557984 | 12 | 24557984 | [T/C] | 0.069 | 2.63E-15 | HSD17B4, DMXL1, SEMA6A |
| 13:21626153 | 13 | 21626153 | [A/G] | 0.101 | 3.80E-08 | GTH-RII, GTF2A1L, STON1, PPP1R21, FOXN2, FBXO11, MSH6, KCNK12, GOLGA4, ITGA9 |
| 13:21626426 | 13 | 21626426 | [T/C] | 0.101 | 3.80E-08 | GTH-RII, GTF2A1L, STON1, PPP1R21, FOXN2, FBXO11, MSH6, KCNK12, GOLGA4, ITGA9 |
| 15:14410587 | 15 | 14410587 | [A/G] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14413544 | 15 | 14413544 | [A/C] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14413799 | 15 | 14413799 | [T/C] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14413803 | 15 | 14413803 | [T/G] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14414140 | 15 | 14414140 | [C/T] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14414368 | 15 | 14414368 | [T/G] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14419133 | 15 | 14419133 | [C/T] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14420897 | 15 | 14420897 | [C/T] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14421396 | 15 | 14421396 | [G/A] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14421648 | 15 | 14421648 | [C/A] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14421941 | 15 | 14421941 | [G/A] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14423479 | 15 | 14423479 | [A/G] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14457958 | 15 | 14457958 | [C/T] | 0.088 | 4.86E-08 | GPHN, FUT8 |
| 15:14651060 | 15 | 14651060 | [G/A] | 0.090 | 1.15E-09 | TMEM121, SLC24A4 |
| 15:14651117 | 15 | 14651117 | [C/A] | 0.090 | 1.15E-09 | TMEM121, SLC24A4 |
| Genes                                      | GSEA Score | FDR  | Body weight at harvest | GSEA Score | FDR  |
|--------------------------------------------|------------|------|------------------------|------------|------|
| TMEM121, SLC24A4                           | 15:14662118| 14662118 [A/C] 0.090 1.15E-09 |          |
| ABHD13, TNFSF13B, MYO16, IRS2, COL4A1      | 16:20105934| 20105934 [G/A] 0.053 1.07E-09 |          |
| ABHD13, TNFSF13B, MYO16, IRS2, COL4A1      | 16:20116545| 20116545 [T/C] 0.053 1.07E-09 |          |
| Uncharacterized                            | 17:20586593| 20586593 [G/T] 0.101 3.49E-09 |          |
| Uncharacterized                            | 17:20601790| 20601790 [T/C] 0.101 3.49E-09 |          |
| Uncharacterized                            | 18:27368147| 27368147 [A/G] 0.098 1.60E-10 |          |
| TMEM260, KTN1, MSH4, VCPKMT, ASPG, KIF26A | 19:11094375| 11094375 [T/C] 0.054 7.95E-09 |          |
| MYCT1, SYNE1, CCDC170, RMND1, TRNAL-UAA, SHPRH | 19:20435177| 20435177 [T/C] 0.065 2.84E-08 |          |
| PITHD1, ELOA, RPL11, EF-1A                 | 22:11635404| 11635404 [C/A] 0.083 1.49E-08 |          |
| CFAP69, KDM1B, DPY19L4, INTS8, GDF6, PKHD1L1, NUDCD1 | 22:11998439| 11998439 [G/A] 0.059 2.80E-12 |          |

| 1:39153024 [G/A] 0.052 2.44E-10 | HDGFL3, CCDC102A |
| 1:39193509 [A/G] 0.052 2.44E-10 | HDGFL3, CCDC102A, NADSYN1 |
| 1:39558113 [G/A] 0.052 2.44E-10 | CCNE1, ZNF536 |
| 1:39628599 [A/G] 0.052 2.44E-10 | ZNF536 |
| 3:47050238 [G/A] 0.111 2.55E-10 | Uncharacterized |
| 3:47051526 [C/T] 0.111 2.55E-10 | Uncharacterized |
| 3:47051658 [C/T] 0.111 2.55E-10 | Uncharacterized |
| 3:47064569 [A/G] 0.111 2.55E-10 | Uncharacterized |
| 3:47128152 [T/C] 0.111 2.55E-10 | TLR2 |
| 3:47130072 [A/G] 0.111 2.55E-10 | TLR2 |
| 3:47130123 [G/T] 0.111 2.55E-10 | TLR2 |
| 3:47130132 [G/T] 0.111 2.55E-10 | TLR2 |
| 3:47132237 [T/C] 0.111 2.55E-10 | TLR2 |
| 3:47132276 [T/C] 0.111 2.55E-10 | TLR2 |
| 3:47132284 [C/T] 0.111 2.55E-10 | TLR2 |
| 3:47132357 [C/A] 0.111 2.55E-10 | TLR2 |
| 3:47132400 [G/A] 0.111 2.55E-10 | TLR2 |
| 3:47132429 [A/G] 0.111 2.55E-10 | TLR2 |
| 3:47132474 [A/G] 0.111 2.55E-10 | TLR2 |
| 3:47138842 [T/C] 0.111 2.55E-10 | TLR2 |
| 3:47142840 [A/G] 0.111 2.55E-10 | TLR2 |
| 4:35484394 [T/C] 0.097 5.60E-09 | CDC27 |
| 5:6037462 [T/G] 0.055 1.10E-08 | DDIT3, MARS, ARHGAP9, GLI1, R3HDM2 |
| 6:4119361 [T/C] 0.127 3.13E-08 | ERO1B, TBCE, RBM34 |
| 7:54783508 [A/G] 0.051 3.58E-08 | GTF3C6, FAM107B |
| 8:542536 [A/G] 0.109 8.55E-10 | Uncharacterized |
| 8:542852 [A/G] 0.109 8.55E-10 | Uncharacterized |
| 8:543350 [G/A] 0.109 8.55E-10 | Uncharacterized |
| 8:543673 [A/G] 0.109 8.55E-10 | Uncharacterized |
| 8:544553 [C/A] 0.109 8.55E-10 | Uncharacterized |
| 8:544584 [T/C] 0.109 8.55E-10 | Uncharacterized |
### Table 1: Gene Expression Data

| Gene Symbol | Log2 Fold Change | P-value | Description |
|-------------|-----------------|---------|-------------|
| ABHD13, TNFSF13B, MYO16, IRS2, COL4A1 | 3.04E-19 | 0.059 | Uncharacterized |
| OTUB2, CCDC197, ASB2, FAM181A, MARS, ARHGAP9, GLI1, R3HDM2 | 2.45E-16 | 0.083 | Uncharacterized |
| PITHD1, ELOA, RPL11, EF-1A | 6.01E-09 | 0.098 | Uncharacterized |

### Table 2: Head weight data

| Gene Symbol | Log2 Fold Change | P-value | Description |
|-------------|-----------------|---------|-------------|
| HDGFL3, CCDC102A | 3.47E-10 | 0.052 | Uncharacterized |
| HDGFL3, CCDC102A, NADSYN1 | 3.47E-10 | 0.052 | Uncharacterized |
| CCNE1, ZNF536 | 3.47E-10 | 0.052 | Uncharacterized |
| ZNF536 | 3.47E-10 | 0.052 | Uncharacterized |
| MARS, ARHGAP9, GLI1, R3HDM2 | 3.04E-10 | 0.059 | Uncharacterized |
| MARS, ARHGAP9, GLI1, R3HDM2 | 3.04E-10 | 0.059 | Uncharacterized |
| MARS, ARHGAP9, GLI1, R3HDM2 | 3.04E-10 | 0.059 | Uncharacterized |
Body length at harvest

| Chromosome | Position | Gene(s)                                                                 |
|------------|----------|-------------------------------------------------------------------------|
| 1:39153024 | 1        | HDGFL3, CCDC102A                                                        |
| 1:39193059 | 1        | HDGFL3, CCDC102A, NADSYN1                                               |
| 1:39558113 | 1        | CCNE1, ZNF536                                                           |
| 1:39628599 | 1        | ZNF536                                                                  |
| 3:47137003 | 3        | TLR2                                                                    |
| 4:17982574 | 4        | SREBF1, TOM1L2, DRC3                                                    |
| 5:6044340  | 5        | MARS, ARHGAP9, GLI1, R3HDM2                                             |
| 6:3269011  | 6        | ASPSCR1, NOTUM, TMC6, TMC8, TK1                                         |
| 7:54783508 | 7        | GTF3C6, FAM107B                                                         |

URI1, RXYLT1, SRGAP1, CCDC113, PARP12, TRNAG-UC

UR21, RXYLT1, SRGAP1, CCDC113, PARP12, TRNAG-UC

UR11, RXYLT1, SRGAP1, CCDC113, PARP12, TRNAG-UC

URI1, RXYLT1, SRGAP1, CCDC113, PARP12, TRNAG-UC

URI1, RXYLT1, SRGAP1, CCDC113, PARP12, TRNAG-UC
| Gene Name | Gene Symbol | Chromosome | Band | SNP Position | Allele | Minor Allele Frequency | Minor Allele Inversion | Weight |
|-----------|-------------|------------|------|--------------|--------|-----------------------|------------------------|--------|
| EXD1      |             | 8:545833   | T/C  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
| RASGRP1   |             | 8:546104   | C/A  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
| SYNDIG1L  |             | 8:546281   | G/A  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
| VSX2      |             | 8:546414   | G/A  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
| LIN52     |             | 8:546879   | A/G  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:546888   | G/T  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:549599   | T/C  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:560038   | C/T  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:561805   | C/T  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:567773   | C/T  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 8:567891   | C/T  | 0.109        |        | 4.10E-08              | Uncharacterized        |        |
|          |             | 9:27601979 | A/G  | 0.054        |        | 4.28E-08              | Uncharacterized        |        |
| TBRG4     |             | 11:19476671| T/G  | 0.063        |        | 1.60E-09              | TBRG4, ATP4A, TRNAR-ACG|        |
|          |             | 12:27146675| C/T  | 0.079        |        | 1.96E-13              | GPX8, MCIDAS, ISCA1, PSAT1|        |
|          |             | 13:23915471| A/G  | 0.050        |        | 3.07E-08              | EGLN1, TSNAX, DISC1, SIPAIL2|        |
|          |             | 14:18442958| A/G  | 0.059        |        | 3.08E-08              | FCHSD2, YIF1B, BLOC1S3|        |
|          |             | 15:13929129| T/C  | 0.082        |        | 5.00E-08              | KCNK3, SLC35F6, HLX   |        |
|          |             | 16:20105934| G/A  | 0.053        |        | 3.34E-11              | ABHD13, TNFSF13B, MYO16, IRS2, COL4A1|        |
|          |             | 17:25198694| A/G  | 0.098        |        | 1.28E-09              | POMGNT1                |        |
|          |             | 18:18194228| A/G  | 0.051        |        | 1.23E-09              | EFCAB14, ZNF830        |        |
|          |             | 19:9747813 | T/C  | 0.079        |        | 6.81E-09              | EXD1, RASGRP1, SYNDIG1L, VSX2, LIN52|        |
|          |             | 20:32775965| A/G  | 0.108        |        | 5.07E-08              | GATA5, RBBP8NL, HCK, TM9SF4, PLAGL2|        |
|          |             | 22:11998439| G/A  | 0.059        |        | 7.13E-18              | CFAP69, KDM1B, DPY19L4, INTS8, GDF6, PKHD1L1, NUDCD1|        |
|          |             | 22:11998439| G/A  | 0.059        |        | 7.13E-18              | CFAP69, KDM1B, DPY19L4, INTS8, GDF6, PKHD1L1, NUDCD1|        |

**Hon weight**

| Gene Name | Gene Symbol | Chromosome | Band | SNP Position | Allele | Minor Allele Frequency | Minor Allele Inversion | Weight |
|-----------|-------------|------------|------|--------------|--------|-----------------------|------------------------|--------|
| HDGFL3    |             | 1:39153024 | G/A  | 0.052        |        | 1.28E-10              | HDGFL3, CCDC102A       |        |
|          |             | 1:39193509 | A/G  | 0.052        |        | 1.28E-10              | HDGFL3, CCDC102A, NADSYN1|        |
|          |             | 1:39558113 | G/A  | 0.052        |        | 1.28E-10              | CCNE1, ZNF536          |        |
|          |             | 1:39628599 | A/G  | 0.052        |        | 1.28E-10              | ZNF536                 |        |
|          |             | 3:47137003 | A/C  | 0.110        |        | 1.17E-10              | TLR2                   |        |
|          |             | 4:35484394 | T/C  | 0.097        |        | 1.32E-08              | CDC27                  |        |
|          |             | 5:6040474  | T/C  | 0.059        |        | 3.85E-08              | MARS, ARHGAP9, GLI1, R3HDM2|        |
|          |             | 5:6040483  | C/T  | 0.059        |        | 3.85E-08              | MARS, ARHGAP9, GLI1, R3HDM2|        |
|          |             | 5:6042034  | A/G  | 0.059        |        | 3.85E-08              | MARS, ARHGAP9, GLI1, R3HDM2|        |
|          |             | 8:390896   | T/C  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:391578   | G/A  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:393162   | G/T  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:393184   | A/G  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:393250   | C/A  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:397651   | T/C  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:398534   | G/A  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
|          |             | 8:399361   | T/G  | 0.102        |        | 5.45E-08              | Uncharacterized        |        |
| Chromosome | Start | End | Gene(s) |
|------------|------|-----|---------|
| 8:399394   | 8    | 399394 | [C/A] 0.102 5.45E-08 Uncharacterized |
| 8:400216   | 8    | 400216 | [C/A] 0.102 5.45E-08 Uncharacterized |
| 9:16328834 | 9    | 16328834 | [A/G] 0.077 5.32E-08 PLXDC2, MALRD1, WDR60, ESYT2, NCAPG2, PTPRN2 |
| 12:24557870| 12   | 24557870 | [A/G] 0.068 9.89E-14 HSD1B4, DMXL1, SEMA6A |
| 12:24557894| 12   | 24557894 | [T/C] 0.097 8.99E-14 HSD1B4, DMXL1, SEMA6A |
| 15:23976527| 15   | 23976527 | [T/C] 0.090 5.18E-08 FILIP1, SENP6, MYO6 |
| 16:20105934| 16   | 20105934 | [G/A] 0.053 1.91E-10 ABHD13, TNFSF13B, MYO16, IRS2, COL4A1 |
| 17:19609147| 17   | 19609147 | [C/T] 0.101 3.82E-08 LG17H12ORF66, TMEM19, RAB3IP, PROSER2, UPF2, SLC35E3, NUP107, MDM2 |
| 17:19612729| 17   | 19612729 | [C/T] 0.101 3.82E-08 LG17H12ORF66, TMEM19, RAB3IP, PROSER2, UPF2, SLC35E3, NUP107, MDM2 |
| 18:27357619| 18   | 27357619 | [T/C] 0.115 3.56E-08 Uncharacterized |
| 19:11468492| 19   | 11468492 | [C/T] 0.053 4.89E-08 Uncharacterized |
| 22:11998439| 22   | 11998439 | [G/A] 0.059 5.59E-10 Uncharacterized |

| Waste weight |
|--------------|
| 1:39153024  | 1    | 39153024 | [G/A] 0.052 8.43E-11 HDGFL3, CCDC102A |
| 1:39193509  | 1    | 39193509 | [A/G] 0.052 8.43E-11 HDGFL3, CCDC102A, NADSyn1 |
| 1:39558113  | 1    | 39558113 | [G/A] 0.052 8.43E-11 CCNE1, ZNF536 |
| 1:39628599  | 1    | 39628599 | [A/G] 0.052 8.43E-11 ZNF536 |
| 3:47327036  | 3    | 47327036 | [A/G] 0.113 1.63E-10 Uncharacterized |
| 4:34954060  | 4    | 34954060 | [T/C] 0.084 5.49E-10 PSMD3, SAMD14 |
| 5:6040474   | 5    | 6040474 | [T/C] 0.059 9.93E-09 MARS, ARHGAP9, GLI1, R3HDM2 |
| 5:6040483   | 5    | 6040483 | [C/T] 0.059 9.93E-09 MARS, ARHGAP9, GLI1, R3HDM2 |
| 5:6042034   | 5    | 6042034 | [A/G] 0.059 9.93E-09 MARS, ARHGAP9, GLI1, R3HDM2 |
| 6:4119361   | 6    | 4119361 | [T/C] 0.127 1.63E-08 ERO1B, TBCE, RBM34 |
| 7:3766803   | 7    | 3766803 | [A/G] 0.056 3.58E-08 Uncharacterized |
| 7:49374205  | 7    | 49374205 | [G/T] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374211  | 7    | 49374211 | [G/A] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374217  | 7    | 49374217 | [A/G] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374223  | 7    | 49374223 | [T/G] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374230  | 7    | 49374230 | [A/G] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374233  | 7    | 49374233 | [G/T] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374263  | 7    | 49374263 | [T/C] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374277  | 7    | 49374277 | [A/G] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 7:49374283  | 7    | 49374283 | [G/A] 0.056 3.58E-08 RELN, TP53I11, CD82 |
| 8:3796164   | 8    | 3796164 | [G/A] 0.113 5.08E-08 KCNJ2 |
| 9:14670077  | 9    | 14670077 | [C/T] 0.059 5.22E-08 SLC4A2, ABCF2, CHPF2, SMARC3D |
| 9:14674835  | 9    | 14674835 | [A/G] 0.059 5.22E-08 SLC4A2, ABCF2, CHPF2, SMARC3D |
| 10:32346497 | 10   | 32346497 | [C/T] 0.056 3.58E-08 Uncharacterized |
| 10:32346502 | 10   | 32346502 | [A/G] 0.056 3.58E-08 Uncharacterized |
| 11:19628608 | 11   | 19628608 | [T/G] 0.069 3.87E-09 INO80C, GALNT1 |
| Chromosome:Position | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 |
|--------------------|--------|--------|--------|--------|--------|
| 12:24557870 12 | 24557870 [A/G] 0.069 | 2.51E-15 | HSD17B4, DMXL1, SEMA6A |
| 12:24557984 12 | 24557984 [T/C] 0.069 | 2.51E-15 | HSD17B4, DMXL1, SEMA6A |
| 13:29981822 13 | 29981822 [A/G] 0.056 | 3.58E-08 | Uncharacterized |
| 14:64504 14 | 64504 [T/C] 0.056 | 3.58E-08 | USF1 |
| 14:64506 14 | 64506 [T/C] 0.056 | 3.58E-08 | USF1 |
| 14:64730 14 | 64730 [G/A] 0.056 | 3.58E-08 | USF1 |
| 14:24825295 14 | 24825295 [C/T] 0.056 | 3.58E-08 | Uncharacterized |
| 15:14651060 15 | 14651060 [G/A] 0.090 | 9.42E-10 | TMEM121, SLC24A4 |
| 15:14651117 15 | 14651117 [C/A] 0.090 | 9.42E-10 | TMEM121, SLC24A4 |
| 15:14662118 15 | 14662118 [A/C] 0.090 | 9.42E-10 | TMEM121, SLC24A4 |
| 16:20105934 16 | 20105934 [G/A] 0.053 | 5.48E-11 | ABHD13, TNFSF13B, MYO16, IRS2, COL4A1 |
| 16:20116545 16 | 20116545 [T/C] 0.053 | 5.48E-11 | ABHD13, TNFSF13B, MYO16, IRS2, COL4A1 |
| 17:31075239 17 | 31075239 [A/G] 0.067 | 1.76E-08 | ZMAT3, STXB3, HENMT1, FAM102B, TM4SF4, |
| 18:17100458 18 | 17100458 [T/C] 0.056 | 1.67E-08 | ZNF521, SS18, PSMA8, KCTD1 |
| 22:11998439 22 | 11998439 [G/A] 0.059 | 2.13E-12 | CFAP69, KDM11B, DPY19L4, INTS8, GDF6, PKHD1L1, NUDCD1 |

### Fillet waste

| Chromosome:Position | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 |
|--------------------|--------|--------|--------|--------|--------|
| 3:15396071 3 | 15396071 [A/G] 0.092 | 1.17E-08 | TRNAR-UCU, TRNAA-CGC, MSANTD1, HGFAC, DOK7 |
| 4:34954382 4 | 34954382 [T/C] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34954397 4 | 34954397 [A/G] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34958811 4 | 34958811 [A/G] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34958990 4 | 34958990 [G/A] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34959371 4 | 34959371 [C/A] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34960461 4 | 34960461 [T/C] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 4:34963090 4 | 34963090 [T/C] 0.107 | 1.19E-14 | PSMD3, SAMD14 |
| 5:29062243 5 | 29062243 [G/A] 0.058 | 3.45E-11 | PRICKLE2, ADAMTS9, MAGI1 |
| 5:29062243 5 | 29062243 [G/A] 0.058 | 3.45E-11 | PRICKLE2, ADAMTS9, MAGI1 |
| 6:6596302 6 | 6596302 [A/G] 0.065 | 3.97E-09 | Uncharacterized |
| 8:396284 8 | 396284 [T/C] 0.108 | 8.77E-11 | Uncharacterized |
| 12:24525556 12 | 24525556 [A/G] 0.062 | 2.38E-11 | PRR16, HSD17B4, DMXL1, SEMA6A |
| 13:30002073 13 | 30002073 [A/G] 0.174 | 1.78E-18 | Uncharacterized |
| 16:1683743 16 | 1683743 [G/A] 0.052 | 1.90E-08 | KDEL1C1, TEX30, NEPRO, RALB |
| 18:27368147 18 | 27368147 [A/G] 0.098 | 2.11E-09 | Uncharacterized |
| 22:11635404 22 | 11635404 [C/A] 0.083 | 1.15E-08 | PITHD1, ELOA, RPL11, EF-1A |

### Fillet yield

| Chromosome:Position | Gene 1 | Gene 2 | Gene 3 | Gene 4 | Gene 5 |
|--------------------|--------|--------|--------|--------|--------|
| 6:33824877 6 | 33824877 [T/G] 0.055 | 4.50E-10 | XYL1T, RPS15A, COQ7 |
| 12:26984411 12 | 26984411 [G/A] 0.066 | 2.39E-12 | Uncharacterized |
| 13:17730096 13 | 17730096 [C/A] 0.113 | 2.01E-08 | SLC25A16, ALOX5, ZFAND4, MARVELD1, AVP1I, MARCH8 |
| 13:17730605 13 | 17730605 [C/A] 0.113 | 2.01E-08 | SLC25A16, ALOX5, ZFAND4, MARVELD1, AVP1I, MARCH8 |
| 14:30148797 14 | 30148797 [G/T] 0.056 | 3.01E-09 | Uncharacterized |
| 16:12574352 16 | 12574352 [G/A] 0.051 | 3.84E-08 | ZNF148, SLC12A8, HEG1, KANSL1L |
1Linkage group. 2In base pairs. 3Minor allele frequency. 4Genes found within a window of 100 kb downstream and upstream from the lead SNP using Oreochromis niloticus as the genome reference (O_niloticus_UMD_NMBU).