A pilot study to elucidate effects of artificial selection by size on the zebrafish (*Danio rerio*) fast skeletal muscle transcriptome.

Harish. P1, Mareco. E. A.2, and Garcia de la serrana. D3,4\*.

1School of Biological Sciences, Royal Holloway-University of London, Egham, Surrey (UK).

2Environment and Regional Development Graduate Program, University of Western São Paulo, Presidente Prudente, São Paulo (Brazil).

3Scotish Oceans Institute. School of Biology. University of St Andrews, St Andrews, Scotland (UK).

4Cellular Biology, Physiology and Immunology Department. School of Biology. University of Barcelona, Barcelona (Spain).

\*Corresponding author.

**Abstract**

Fish size is a complex trait determined by the interaction of environmental and genetic factors. While evidence exists that fish body length is determined by muscle fibre composition, the molecular basis behind body size and muscle fibre composition remains yet to be fully understood. Here we use the model zebrafish (*Danio* *rerio*) to explore the molecular mechanisms involved in muscle fibre number, muscle growth and their effect on animal size. In order to do so, we used two lines of zebrafish artificially selected by Amaral and Johnston 2012 for body length to obtain large (L) and small (S) genetic lineages. After selection animals from the L-lineage were significantly bigger in length than individuals from the S-lineage (+6.6%, +6.7% and +5.5% for standard, total and fork lengths) and had larger and more muscle fibres (+15% and +24% on average fibre area and fibre number respectively) with individuals from S-lineage showing a higher fibre density (+17%). RNAseq analysis on fast skeletal muscle revealed differences in gene expression, splice variants (SV) abundance and single nucleotide polymorphism (SNPs) retention between lineages, indicating the complexity of the molecular processes involved in the determination of the trait. Despite the complexity we found that animals from the L-lineage had an enrichment in genes related to the dystroglycan complex (GO:0016011) while genes related to the proteasome (GO:0022624), sarcomere (GO:0030017), muscle homeostasis (GO:0046716) and response to stress (GO:0006950) were enriched on the S-lineage. We also found differences in expression, SV or/and SNPs retention for some key genes involved in the control muscle development such as *mef2ca*, *mef2cb*, *stac3*, *map14a* and components of the circadian system (*per1a*, *per1b*, *per3* and *cry2b*). These preliminary results provide an initial insight of the molecular changes induced by artificial selection for size, that could aid future studies related to determination of fish growth.

**Background**

Somatic growth involves the coordination of several physiological processes including food intake, digestion, absorption and incorporation of the nutrients necessary to sustain growth. Skeletal muscle represents up to 60% of the total body weight in fish, with 90% of it formed by fast skeletal muscle (Johnston et al., 1998), and hence plays a crucial role in determining animal size and mass. Manipulation of traits such as an increased growth rate and final size have a direct commercial impact on the industry, and has been driven by artificial selection in growth (Yáñez et al., 2015), in species such as Atlantic salmon (*Salmo* *salar*, (Gjedrem, 2004), tilapia (*Oreochromis* *niloticus*, (Charo-Karisa et al., 2006)) or gilthead sea bream (*Sparus* *aurata*, (Fernández-Díaz and Yúfera, 1997) (for a complete review on the subject see Gjedrem and Rye, 2018). In recent years, breeding programs are incorporating biotechnology-based approaches to increase efficiency of selection by identifying markers that explain a significant proportion of trait variability and select for them (as an example see Robledo et al., 2017). While there is an increasing number of molecular markers linked to growth, molecular mechanisms behind body size, growth rate or muscle fibre number have not yet been fully described. However, physiological processes such as net protein accumulation and activation of satellite cells are known to play an important role.

Essential to skeletal muscle growth is also a net accumulation of protein by shifting the balance between proteosynthesis and metabolism, a mechanism which is primarily regulated by the Pi3k/Akt/mTor pathway. The phosphorylation and activation of the mTor complex leads to the transcription of several components of the translational systems (such as ribosomal subunits, elongation factors, etc) increasing protein synthesis. Different studies in fish have shown that while growth factors do not independently stimulate mTor phosphorylation (Garcia de la serrana et al., 2013), amino acids alone are can active protein synthesis through mTor pathway, and can act synergistically with growth factors for an increased response and activation of related pathways (Garcia de la serrana et al., 2013; Johnston et al., 2011).

Three main systems control muscle protein degradation that leads to atrophy: the proteasome, lysosomal (autophagy) and calpain/calpastatin (Johnston et al., 2011). Several components of the protein degradation systems are under control of signalling pathways such as Igfs through the Pi3k pathway which, in addition to stimulating protein synthesis, can also inhibit degradation by blocking the Foxo3 transcription factor that controls expression of some key proteasome components (Johnston et al., 2011). Similarly, amino acids (especially essential amino acids) are also able to produce a similar effect and reduce atrophy related components (Garcia de la serrana et al., 2017).

Post-embryonic growth of skeletal muscle is also regulated by the activation of myogenic precursor cells (MPCs) localized in the basal lamina of the muscle fibres (Johnston, 2006; Johnston et al., 2011). The activation of MPCs is a key step during myogenic growth and regeneration, occuring in response to endocrine, immune or nutritional signals. Activation of these MPCs result in an asymmetric division with a population of cells re-entering quiescence to replenish the MPC pool and a set of myoblasts exiting the cell cycle to facilitate either myofibrillogenesis to form new fibres in the fish, or nuclear accretion into pre-existing fibres to maintain an effective nuclear domain and drive myofibrillar growth.

The process of myogenesis, with all its different events, is a complex process that requires the activation and coordination of complex molecular networks and signals including transcription factors (Myod, Myog, Myf5, Myf6, Sox8 or Pax3/7) (Johnston, 2006; Johnston et al., 2011), adaptive proteins (such as Crk, Crkl, Dock1 and Dock5) (e.g. Moore et al., 2007) or growth factors (Igf1, Igf2, Hgf or Gh) (Johnston, 2006; Johnston et al., 2011).

The zebrafish with its high fecundity, low generation time, embryo transparency, regenerative capacity and a high-quality annotated genome, is used as a model in many fields of biology (Dodd et al., 2000). Despite not being used in aquaculture, the zebrafish is considered as an useful model to study fundamental questions of fish physiology such as nutrition, growth or immunology, allowing the researcher to generate strategies which can be translated into the aquaculture industry (Lee-Estevez et al., 2018; Opazo et al., 2017; Ribas and Piferrer, 2014; Ulloa et al., 2014). Previously, Amaral and Johnston used the short generation time of the zebrafish to artificially select two lines of animals by their body size, to simulate the selective pressure that can be present during an industrial breeding program (Amaral and Johnston, 2012a). The authors found that artificial selection for body size affected the levels of maternal mRNA in the oocyte and the transcriptional regulation of several components of the Igf-Akt-mTor system in response to food intake after a short period of starvation (Amaral and Johnston, 2012a). However, their work was restricted to the analysis of a limited number of genes and did not analyse the effect of artificial selection by size on the skeletal muscle fibre histology, the muscle transcriptome, or other genetic variants such as splice variants (SV) and single nucleotide polymorphisms (SNPs).

Next generation sequencing technology has been demonstrated to be a powerful tool to study different aspects of fish skeletal muscle growth, development and physiology (Alexander et al., 2011; Garcia de la serrana Castillo et al., 2012; Garcia de la serrana et al., 2015; Mareco et al., 2015), by allowing the researcher to simultaneously study the transcriptional regulation of multiple genes and under different experimental conditions. The availability of a reference genome, in combination with sequencing technologies can additionally facilitate the study of SV abundance or the identification of SNPs.

In this study, we analysed the fast skeletal muscle of the zebrafish from selected lineages generated by Amaral and Johnston, and investigated muscle fibre composition and used RNAseq to study the effects on fast skeletal muscle global gene expression, SV abundance and different retention of SNPs to better understand the molecular mechanisms behind muscle growth as a product of artificial selection.

**Methods**

*Fish size selection*

Animals from the large (L) and small (S) zebrafish were obtained from the lines previously generated by Amaral and Johnston, 2012a. Briefly, wild zebrafish were obtained from Mymensingh, Bangladesh and reared in 25L glass tanks at 27 ± 0.3°C, with a 12:12 photoperiod, in a UV sterilised fresh water recirculation system and fed with bloodworms to satiation once daily and thrice a day from a week prior to breeding. For each lineage, pre-separated males and females were introduced into a breeding tank in a 1 male to 2 female proportion, and eggs were collected the next day. Eggs were immediately cleaned with 10% (v/v) Hank’s solution and maintained in glass tanks (1L) under the same conditions as the recirculation system. After 7 days post fertilisation (dpf) the larvae were fed a combination of ZM-100 (Fish Food Ltd, Hampshire, UK) and microworms. 50% of the water was changed daily until 30dpf, where fish were transferred to the main recirculation system, and fed to satiation with a combination of ZM-200 (Fish Food Ltd, Hampshire, UK) and bloodworms (Ocean Nutrition, Belgium). Four rounds of experimental selection were conducted based on body size at 90 dpf. For the small-lineage, fish with fork length lesser than 75% of the mean fork length were bred with each other, and for the larger lineage, fish with fork length greater than 125% the mean fork length were bred with each other. An unselected line was also produced which was not subject to experimental selection for use as a reference sample set. Three replicates of the unselected (U), small (S) and large (L) lineages were bred with a selection of 20 fish per replicate for propagation. All breeding was conducted at ~120 dpf.

*Fish sampling*

Animals were sampled in 2012 from the 4th generation of selected zebrafish and followed the same protocol as described by Amaral and Johnston (Amaral and Johnston, 2012a). Total length (TL), fork length (FL) and standard length (SL) were measured in 40-50 animals from each line breed (L, U and S).

For muscle morphology analysis, 10 fish from L and 10 from the S lineages were photographed, and a 5-7 mm transverse section was obtained at 0.7 fork length. The muscle steak was photographed and the total cross-sectional area was estimated. The steaks were mounted on cork sheets and frozen in 2-methyl-butane (isopentane) cooled on liquid nitrogen to -159°C. The blocks were wrapped in foil and stored at -80°C, until further analysis. Muscle blocks were equilibrated to -25°C and 7mm sections cut using a cryostat. Sections were transferred onto L-poly-lysine coated slides for later histological analysis. Sections were stained with myosin ATPase to distinguish between fibre types and hematoxylin and eosin for morphometric analysis (Johnston and Temple, 2002). Photographs of tissue sections were taken at a magnification of ×200, and four fields, selected at random, were photographed per block per fish.

*RNA extraction and sequencing*

Fast skeletal muscle was extracted from the frozen histological blocks washed in nuclease free water and homogenised in Lysing Matrix-D (Qbiogene) using 1mL of TRIzure (Sigma) in a FastPrep® machine (Qbiogene) and RNA extracted using chloroform:isopropanol method following manufacturer recommendations. Resulting pellets were washed in 1ml ice-cold 70% ethanol for a total of four times. The samples were air-dried and suspended in 50µl of RNase-free water, at 37°C. The RNA concentration, 260/280 and 260/230 ratios were measured using a NanoDrop® spectrophotometer (Fisher) and had a range between 2-2.15. RNA integrity was confirmed in a 1% (m/v) ethidium bromide agarose gel electrophoresis and further confirmed by the SickKids Hospital Sequencing services (Vancouver, Canada) before library preparation and sequencing.

*Gene expression analysis, splice variants determination and SNP detection*

RNASeq library preparation (1 µg RNA) of 6 individuals from L and 6 from S-lineages, sequencing (paired-end reads on an Illumina HiSEq 2500 sequencer), reads quality filtering, mapping, reads normalization, expression values, splice variants estimation and global statistical analysis for differences in expression were carried out by the bioinformatics department of SickKids Hospital Next Generation Sequencing service. In short, the DNase-treated RNA was used to prepare individually barcoded RNASeq libraries with the TruSeq RNA Sample Prep kit (Illumina). The 12 libraries were randomly distributed in two lanes (6 per line) of HiSeq2500 platform and sequenced following manufacturer’s recommendations (Illumina). Quality filtered paired-end reads were mapped against the zebrafish genome (Zv9) using Cufflinks v0.0.5. Reads abundance was estimated using the RSEM v1.2.15 (Li and Dewey, 2011). Normalized mapped reads for contig size and library depth were used to calculate global differences in gene expression between L and S lineages using the DESEQ v2.13.2 algorithm from the R-Bioconductor package (Love et al., 2014). A false discovery rate (FDR) cut-off of FDR < 0.05 and log2Fold-change threshold of 0.5 was applied. Differences on SV abundance were established by estimating differences in normalised counts at exon level for each gene between the two lineages using the R-build package DEXseq v1.28.0 (Anders et al., 2012).

For SNP detection, low-quality reads (<20 phred) were removed using Trimmomatic program v0.38 (Bolger et al., 2014) and sequence reads from each animals were aligned against the zebrafish genome (Zv9) using BWA-MEM v0.7.12 assembly program (Li and Durbin, 2009). For SNP variant calls Samtools mpileup pipeline v1.6 (Li, 2011) was used. VCFtools v0.1.13 (Danecek et al., 2011), was used to filter the raw variants for strand bias, end distance bias, base quality bias, SNPs around gaps, low coverage and erroneously high coverage. We considered a minimum of 100 reads per animal, with high-confident variation. SNP Annotation was made using snpeff program v4.3t (Cingolani et al., 2012), where consensus for each trait was made and used as a basis for comparison between lineages.

Gene ontology enrichment analysis of the genes showing differences in gene expression, SV abundance and SNPs retention was performed using STRING v10.5 software (Franceschini et al., 2013) with a FDR cut-off of <0.05 (differential gene expression, SV and SNPs detection workflows are shown in Supplementary Figure 1).

*Statistics*

Data on muscle morphology was tested for normality and homoscedasticity, and subsequently analysed using an unpaired student’s t-test (overall significance level equal to 0.05). The distribution of fibre diameters was written using a suite of non-parametric statistical programs written in R (http://www.r-project.org/) developed from those previously described (Johnston et al., 2014). Fast muscle fibres were fitted using a kernel function into smooth probability distribution functions (PDFs) using an in-house program written in R v3.3.0. The average smoothing parameter used (0.284) was similar between groups. Bootstrap techniques (n=1000) were used to distinguish underlying structure in the distributions from random variation. A non-parametric Kolmogorov–Smirnoff two-sample test was used to test the null hypothesis that the PDFs of groups were equal over all diameters (Johnston et al., 2012).

Differences between L and S lineages for fast skeletal muscle average fibre diameter, fibre density, fibre number and animal size were tested using standard R tools built in R-Studio v.1.1.419. Differences in FL, SL and FL were analysed using a two-way ANOVA with *group* (L or S) and *sex* (Male or Female) as factor followed by a Tukey HSD posthoc analysis. Differences were considered significant when p-val<0.05. With the exception of the All graphs, with the exception of the probability distribution functions (PDFs), were produced using the ggplot2 R-build package (Wickham, 2016).

For global DGE between L and S individuals the DESEQ v.2.13.2 a false discovery rate (FDR) cut-off of FDR < 0.05 was applied for significant differences.

**Results and Discussion**

In the present study we found that artificial selection by size increased the average length of the adult zebrafish. Individuals from the large (L) lineage displayed significantly increased standard length (SL), fork length (FL) and total length (TL) than animals from the S-lineage (+6.7%, +6.5% and +5.3% for SL, FL and TL respectively; p-val <0.05 for each measurement) regardless of the sex of the animals (Table 1) as previously observed by Amaral and Johnston (for a comprehensive discussion see Amaral and Johnston, 2012a). This is not surprising, as similar results has been observed for land animals previously (e.g. De Roos et al., 2008; Macarthur, 1949; Rubin et al., 2012), and fish (reviewed in Gjedrem and Rye, 2018) when selected for growth rate or final size.

In this study we report for the first time significant differences in muscle fibre structure and size distribution between the two lineages as a result of size selection, with higher average fibre diameter (+15%; p-val<0.01) and number (+24%; p-val<0.001) (Figure 2A, B) in animals from the L-lineage. However, fibre density was higher in individuals from the S-lineage (+13%; p-val <0.01) (Figure 2C). While it might seem logical that a larger animal requires a higher number of muscle fibres, it cannot be concluded that they must also have a higher fibre density. For instance, coho salmon transgenic for growth hormone (Gh1) had similar average fibre diameter and fibre density than non-transgenic animals of matching size (Johnston et al., 2014). These results also suggest that size selection may have affected the processes of fibre formation, impacting hyperplastic and hypertrophic responses during growth. The effect on energy metabolism resulting from this phenotype, and selection pressures acting on the fish skeletal muscle remains yet to be fully elucidated.

In order to elucidate the molecular basis behind differences in size and muscle structure an RNAseq analysis was performed to investigate differences in gene expression, splice variants (SV) abundance and single nucleotide polymorphism (SNP) frequencies between the two lines.

Despite the relatively low number of animals used (6 per group) and size being a highly complex polygenic trait, a preliminary exploration of the expression data generated by principal components analysis (PCA) show that individuals from S and L lineages segregated in two disperse, but clear, groups along the first component, indicating that artificial selection had some measurable effect on gene transcription (Supplementary File 2). The PCA, however, showed that two individuals from the L group (L06 and L23) were not completely separated (Supplementary File 2). Correlations between transcriptomic data, muscle morphology and further analysis by statistical outlier tests however indicate that the samples were from a single population, and were hence not discarded as outliers.

A total of 91 genes were significantly up-regulated in animals from the L-lineage, as compared to 90 in the S-lineage (Table 2; Supplementary File 3 and 4). Despite the differences in gene expression found, we did not detect ontological enrichment for any GO category on the genes differently expressed. This may suggest that a broad spectrum of molecular networks are involved in size determination making characterisation of enrichment in any specific pathway challenging. However, we identified several components of the circadian clock system, such as *per1a*, *per1b* and *cry2b* (Baggs et al., 2009; Buijs et al., 2013). Despite the majority of studies in relation to the circadian network having focused on their role on the brain (Bhargava et al., 2015; Hermann et al., 2013), there is evidence that clock genes are expressed in others tissues, playing active roles in their regulation (Atwood et al., 2011). For instance in skeletal muscle, components of the clock system regulates the expression of several genes involved in muscle development such as *Myod1*, *Fbox32*, *Pgc-1a/ß* or *Myh1* (Andrews et al., 2010; Aoyama and Shibata, 2017; Chatterjee and Ma, 2016; Riley and Esser, 2017). Similar relationships between clock genes and genes related to muscle development have also been described in fish (Amaral and Johnston, 2012b; Wu et al., 2016). However, while *Myod1* remainsthe primary myogenic regulatory factor regulated by clock genes in mammalian skeletal muscle, studies in fish suggest that *myf5* (Lazado et al., 2014) or *myf6* (Amaral and Johnston, 2012b) may be also targets. It is interesting to notice that, despite not passing the filter of significance adopted in our study (FDR<0.05) we found two MRFs with differences in the number of reads mapped: *myod1* (higher in L-zebrafish with 0.83 fold-change and uncorrected p-val=0.01) and *myf6* (higher in S-zebrafish with 0.72 fold-change and uncorrected p-val=0.01); an effect which can be elucidated with a larger sample size. These results might indicate that circadian genes modulate growth through the regulation of *myod1* and *myf6* expression, but such role will need further experimental confirmation.

We also detected a higher expression of *mapk14a* on animals from the S-lineages (0.97 fold-change, FDR=0.04). The Mapk14a, also known as p38a, plays a critical role in controlling skeletal muscle development, growth, differentiation, metabolism and slow fibre phenotype (Akimoto et al., 2005; Keren et al., 2006; Meissner et al., 2007; Puri et al., 2000; Shi et al., 2008; Wright et al., 2007). p38 signalling is also involved in regulating the exit of myoblasts from the cell-cycle, maintaining them in a proliferative state (Gardner et al., 2015; Perdiguero et al., 2007). In this study we did not quantify the proportion of slow skeletal muscle between the two lineages, therefore it remains unclear if p38a transcription might be related to an increase in slow skeletal muscle fibres.

While RNAseq has proved a useful tool to understand fish skeletal muscle physiology at transcriptional level (Garcia de la serrana et al., 2015; Mareco et al., 2015), analysis of other functional modifications such as single nucleotide polymorphisms (SNPs) and splice variants (SV), are not normally considered due to the lack of reference genome for many fish species. In this study, we took advantage of the zebrafish genome to study the effect of artificial selection on SNPs frequencies and SV abundance. We found a total of 143 genes with differences in SV abundance (Table 3). Around 62% (92) of detected SV had a single exon spliced while the rest had two or more exons involved (Table 3). In 72% of the cases, shorter transcripts variants were found in animals from the S-lineage (Table 3). GO analysis of SV results showed an enrichment of genes with a LUC7 N-terminus domains (PF03194; FDR=0.04) in the L-lineage. The LUC7 is an U1 snRNA associated protein with a role in splice site recognition (Table 4). In the case of SV from the S-lineage animals, we report a significant enrichment of components of the myosin (GO:0005859; FDR=0.001) and sarcomere complex (GO:0030017; FDR=0.005) with myosin-related functional domains (Table 4), which may suggest sarcomere remodelling in response to size selection, a hypothesis whose physiological meaning will need further exploration. Interestingly, we have also found genes of particular interest for muscle development and function such as *stac3, mef2ca* and *per3* (Table 3). Per3 is another component of the clock system, a network that not only regulates the circadian rhythm, but also regulates Myod1 controlling muscle growth and development. The myogenic enhancer factor-2 (Mef2) is a family of transcription factors that tightly regulate muscle development (Black and Olson, 1998; Estrella et al., 2015). The Mef2c transcription factor is crucial for normal muscle growth, maintaining sarcomeric integrity and regulating skeletal muscle glucose homeostasis (Anderson et al., 2015; Potthoff et al., 2007; Taylor and Hughes, 2017). The depletion of *mef2ca* and *mef2cb* causes severe heart malformation in zebrafish (Hinits et al., 2012, 2009). While the SV do not have a major effect on the *mef2ca* functional domains, the splice variant did modify the N-terminal region (FDR=0.02), a region needed for the interaction with other proteins and develop its functions (Ghosh et al., 2009; Perry et al., 2009). Stac3 was initially characterised in fish skeletal muscle (Bower et al., 2012), and is reported to act as a bridge between RyR1 and Dhpr, channel proteins involved in the depolarization-contraction calcium dependent cycle (Campiglio et al., 2018; Horstick et al., 2014; Nelson et al., 2013). It is interesting to note that the splice variant of *stac3* present in L-lineage individuals had one of two SRC homologue 3 domains (SH3) located in the carboxy terminal region of the protein eliminated after splicing (FDR=0.00). SH3 domains are found in proteins that regulates the cytoskeleton and typically interact with adaptor proteins and kinases. How a shorter *stac3* with a missing SH3 domain would affect the RyR-Dhrp system and its repercussions on the Ca2+ influx during depolarization-contraction cycle raise very intriguing questions that need to be investigated.

We also detected numerous SNPs (Supplementary File 5), but focused on those variants that are unique to each lineage (Table 5, Supplementary File 6). Using these criteria, we identified a total of 94 SNPs that were differently retained between lineages distributed in intergenic regions (9), up-stream (15), downs-stream (9), 3’UTR (20), 5’UTR (19) or coding regions (19) (Supplementary File 6). We have to consider that the number of fixed SNPs observed might be an overestimation due to genetic drift. Genetic drift occurs when effective population is small, allowing the random fixation of SNPs. In our case the effective population was relatively small, therefore we cannot discard that a proportion of the fixed SNPs are not due to artificial selection by size and therefore should be treated carefully.

GO analysis of genetic differences on SNPs fixation in the L-lineage animals revealed an enrichment of components in the dytroglycan and sarcoglycan complex (GO:0016011; FDR=0.03, Table 6) that links the actin cytoskeleton with the extracellular matrix, essential to maintain muscle fibre integrity, while genes with SNPs in the S-lineage were enriched in terms related to the proteasome system (GO:0022624; FDR=0.008). Additionally, we found SNPs located in relevant genes for muscle development, such as members of the already mentioned Mef2 family (*mef2aa* and *mef2ca*) but also *map4k5*. In these cases, above mentioned SNPs were not located in coding regions but in UTRs. While SNPs in UTR regions do not modify the amino acid sequence, they can have significant effects on transcriptional regulation by introducing or removing recognition sites for miRNA, as occurs in the example of myostatin in Texel sheep (Clop et al., 2006). Though we have not found any link between SNPs located in UTR regions and differences in gene expression for the present data, we cannot rule out that under a different physiological context such differences could become more prominent.

Though the study was designed as an exploratory experiment, it demonstrates the complexity of the molecular mechanisms involved (differences in gene expression, SV and SNPs) in size determination and skeletal muscle fibre structure. However, the construction of a coherent model incorporating these results is extremely complex and will need the support of further research to fully understand the role of the different modifications observed. Global analysis (combining gene expression, SV and SNPs) resulted in the identification of certain genes and pathways that were affected at multiple levels. A GO enrichment analysis of all genes either differently expressed or showing differences in SV and SNPs variants for each of the lineages was performed for each of the lineages (Table 7). GO analyses for L-lineage appeared to have an enrichment on components from the sarcolemma (GO:0042383, FDR=0.00) and sarcoglycan complex (GO:0016012, FDR=0.004). On the other hand, when all mechanisms found in the S-lineage were analysed together we found an enrichment on genes related to response to stress (GO:0006950, FDR=0.03), mRNA transcription (GO:0061419, FDR=0.03), muscle cell homeostasis (GO:0046716, FDR=0.04), protein binding (GO:00005515, FDR=0.01), metabolic pathways (PathwayID:1100, FDR=0.01) and HIF-1 pathway (PathwayID:1100, FDR=0.01).

Despite the complexity of the data obtained we have observed that some pathways and components seemed to be more affected. These results suggest that some pathways of the skeletal muscle are under higher pressure during artificial selection for size. Among those networks we found components of the sarcoglycan system repeatedly affected in animals from the L-lineages and elements involved in metabolism and transcription, circadian system and MEF2 family in individuals from the S-lineage. The observed differences in components and molecular networks between the lineages may correlate to the fundamental differences found at the level of the muscle fibre. It might be possible that larger fibres, such as those found in the L-lineage which possesses an increased volume of sarcomere, might require a more robust sarcoglycan system to maintain membrane integrity and avoid derived problems of muscle weakness and dystrophy (Goldstein and McNally, 2010). Similarly, smaller fibres are energetically more demanding than larger fibres, requiring higher energy to maintain ionic homeostasis and membrane potential (Jimenez et al., 2013; Johnston et al., 2012), perhaps suggesting the enrichment of GO terms related to metabolism and transcription. Among the different components affected at different molecular levels by artificial selection, a remarkable number of genes involved in the circadian system (*per1a*, *per1b*, *per3* and *cryo2b*) were affected. The circadian system is a natural oscillator that regulates several biological processes. Studies in zebrafish (Amaral and Johnston, 2012b) and Atlantic cod (Lazado et al., 2014) indicate that the circadian clock system also regulates the expression of key component regulating muscle development such as MyoD and Myf5. We have no indication of how the changes in circadian related genes might affect muscle development. While there is no evidence of changes in the expression of myogenic regulatory factors (MRFs) or differences in SV abundance, we cannot discard their effect at protein level and their role to determining final size. For answering this question, we consider it necessary to carry out further studies.

**Conclusions**

In the present work we analysed the molecular networks involved in final muscle fibre size using a zebrafish model, artificially selected over 4 generations to be either large (L-lineage) or small (S-lineage). After 4 generations of selection we found differences in final size but also on average fibre size and density, indicating that size selection also affected those parameters. RNAseq analysis to study gene expression, SV and SNPs retention showed the complexity of the molecular mechanisms involved in the determination of final size.

**Acknowledgements**

We will like to thank Professor Ian Alistair Johnston from the Scottish Oceans Institute (University of St Andrews) to host the present research in his laboratory facilities and provide the animals for the study. The study and DGS was supported by the Marine Alliance for Science and Technology for Scotland (Scottish Funding Council grant HR09011). Mareco. E. A was supported by the International Cooperation Program CAPES/PGCI BEX: 1450/13-1 and FAPESP 2009/12157-6

**References**

Akimoto, T., Pohnert, S.C., Li, P., Zhang, M., Gumbs, C., Rosenberg, P.B., Williams, R.S., Yan, Z., 2005. Exercise Stimulates Pgc-1alpha Transcription in Skeletal Muscle through Activation of the p38 MAPK Pathway 280, 19587–19593. https://doi.org/10.1074/jbc.M408862200

Alexander, M.S., Kawahara, G., Kho, A.T., Howell, M.H., Pusack, T.J., Myers, J.A., Montanaro, F., Zon, L.I., Guyon, J.R., Kunkel, L.M., 2011. Isolation and transcriptome analysis of adult zebrafish cells enriched for skeletal muscle progenitors. Muscle and Nerve 43, 741–750. https://doi.org/10.1002/mus.21972

Amaral, I.P.G., Johnston, I.A., 2012a. Experimental selection for body size at age modifies early life-history traits and muscle gene expression in adult zebrafish. J. Exp. Biol. 215, 3895–3904. https://doi.org/10.1242/jeb.068908

Amaral, I.P.G., Johnston, I.A., 2012b. Circadian expression of clock and putative clock-controlled genes in skeletal muscle of the zebrafish. AJP Regul. Integr. Comp. Physiol. 302, R193–R206. https://doi.org/10.1152/ajpregu.00367.2011

Anders, S., Reyes, A., Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017. https://doi.org/10.1101/gr.133744.111

Anderson, C.M., Hu, J., Barnes, R.M., Heidt, A.B., Cornelissen, I., Black, B.L., 2015. Myocyte enhancer factor 2C function in skeletal muscle is required for normal growth and glucose metabolism in mice. Skelet. Muscle 5, 1–10. https://doi.org/10.1186/s13395-015-0031-0

Andrews, J.L., Zhang, X., McCarthy, J.J., McDearmon, E.L., Hornberger, T.A., Russell, B., Campbell, K.S., Arbogast, S., Reid, M.B., Walker, J.R., Hogenesch, J.B., Takahashi, J.S., Esser, K.A., 2010. CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. Proc. Natl. Acad. Sci. 107, 19090–19095. https://doi.org/10.1073/pnas.1014523107

Aoyama, S., Shibata, S., 2017. The role of circadian rhythms in muscular and osseous physiology and their regulation by nutrition and exercise. Front. Neurosci. 11, 1–12. https://doi.org/10.3389/fnins.2017.00063

Atwood, A., DeConde, R., Wang, S.S., Mockler, T.C., Sabir, J.S.M., Ideker, T., Kay, S.A., 2011. Cell-autonomous circadian clock of hepatocytes drives rhythms in transcription and polyamine synthesis. Proc. Natl. Acad. Sci. 108, 18560–18565. https://doi.org/10.1073/pnas.1115753108

Baggs, J.E., Price, T.S., Ditacchio, L., Panda, S., Fitzgerald, G.A., Hogenesch, J.B., 2009. Network features of the mammalian circadian clock. PLoS Biol. 7, 0563–0575. https://doi.org/10.1371/journal.pbio.1000052

Bhargava, A., Herzel, H., Ananthasubramaniam, B., 2015. Mining for novel candidate clock genes in the circadian regulatory network. BMC Syst. Biol. 9, 1–14. https://doi.org/10.1186/s12918-015-0227-2

Black, B.L., Olson, E.N., 1998. Transcriptional Control of Muscle Development By Myocyte Enhancer Factor-2 (Mef2) Proteins. Annu. Rev. Cell Dev. Biol. 14, 167–196. https://doi.org/10.1146/annurev.cellbio.14.1.167

Bower, N.I., Garcia de la serrana, D., Cole, N.J., Hollway, G.E., Lee, H.-T., Assinder, S., Johnston, I.A., 2012. Stac3 is required for myotube formation and myogenic differentiation in vertebrate skeletal muscle. J. Biol. Chem. 287, 43936–49. https://doi.org/10.1074/jbc.M112.361311

Buijs, R.M., Escobar, C., Swaab, D.F., 2013. Chapter 15 - The circadian system and the balance of the autonomic nervous system, in: Buijs, R.M., Swaab, D.F. (Eds.), Autonomic Nervous System, Handbook of Clinical Neurology. Elsevier, pp. 173–191. https://doi.org/https://doi.org/10.1016/B978-0-444-53491-0.00015-8

Campiglio, M., Kaplan, M.M., Flucher, B.E., 2018. STAC3 incorporation into skeletal muscle triads occurs independent of the dihydropyridine receptor. J. Cell. Physiol. 9045–9051. https://doi.org/10.1002/jcp.26767

Charo-Karisa, H., Komen, H., Rezk, M.A., Ponzoni, R.W., van Arendonk, J.A.M., Bovenhuis, H., 2006. Heritability estimates and response to selection for growth of Nile tilapia (Oreochromis niloticus) in low-input earthen ponds. Aquaculture 261, 479–486. https://doi.org/10.1016/j.aquaculture.2006.07.007

Chatterjee, S., Ma, K., 2016. Circadian clock regulation of skeletal muscle growth and repair. F1000Research 5, 1549. https://doi.org/10.12688/f1000research.9076.1

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Ruden, D.M., Lu, X., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Landes Biosci. 1–13. https://doi.org/http://dx.doi.org/10.4161/fly.19695

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C., Georges, M., 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat. Genet. 38, 813–818. https://doi.org/10.1038/ng1810

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Roos, A.P.W., Hayes, B.J., Spelman, R.J., Goddard, M.E., 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179, 1503–1512. https://doi.org/10.1534/genetics.107.084301

Estrella, N.L., Desjardins, C.A., Nocco, S.E., Clark, A.L., Maksimenko, Y., Naya, F.J., 2015. MEF2 transcription factors regulate distinct gene programs in mammalian skeletal muscle differentiation. J. Biol. Chem. 290, 1256–1268. https://doi.org/10.1074/jbc.M114.589838

Fernández-Díaz, C., Yúfera, M., 1997. Detecting growth in gilthead seabream, Sparus aurata L., larvae fed microcapsules. Aquaculture 153, 93–102. https://doi.org/10.1016/S0044-8486(97)00017-3

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., Jensen, L.J., 2013. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, 808–815. https://doi.org/10.1093/nar/gks1094

Garcia de la Serrana Castillo, D., Estevez, A., Andree, K., Johnston, I. a, 2012. Fast skeletal muscle transcriptome of the Gilthead sea bream (Sparus aurata) determined by next generation sequencing., BMC genomics. https://doi.org/10.1186/1471-2164-13-181

Garcia de la serrana, D., Johnston, I. A., 2013. Expression of heat shock protein (Hsp90) paralogues is regulated by amino acids in skeletal muscle of Atlantic salmon. PLoS One. 8, e74295.

Garcia de la serrana, D., Devlin, R.H., Johnston, I.A., 2015. RNAseq analysis of fast skeletal muscle in restriction-fed transgenic coho salmon (Oncorhynchus kisutch): An experimental model uncoupling the growth hormone and nutritional signals regulating growth. BMC Genomics 16, 1–17. https://doi.org/10.1186/s12864-015-1782-z

Gardner, S., Gross, S.M., David, L.L., Klimek, J.E., Rotwein, P., 2015. Separating myoblast differentiation from muscle cell fusion using IGF-I and the p38 MAP kinase inhibitor SB202190. Am. J. Physiol. - Cell Physiol. 309, C491–C500. https://doi.org/10.1152/ajpcell.00184.2015

Garcia de la serrana, D., Fuentes, E. N., Martin, S. A. M., Johnston, I. A., Macqueen, D. J., 2017. Divergent regulation of insulin-like growth factor binding protein genes in cultured Atlantic salmon myotubes under different models of catabolism and anabolism. Gen Comp Endocrinol. 247, 53-65.

Ghosh, T.K., Song, F.F., Packham, E.A., Buxton, S., Robinson, T.E., Ronksley, J., Self, T., Bonser, A.J., Brook, J.D., 2009. Physical Interaction between TBX5 and MEF2C Is Required for Early Heart Development. Mol. Cell. Biol. 29, 2205–2218. https://doi.org/10.1128/MCB.01923-08

Gjedrem, T., 2004. Genetic improvement of cold-water fish species. Aquac. Res. 31, 25–33. https://doi.org/10.1046/j.1365-2109.2000.00389.x

Gjedrem, T., Rye, M., 2018. Selection response in fish and shellfish: a review. Rev. Aquac. 10, 168–179. https://doi.org/10.1111/raq.12154

Goldstein, J.A., McNally, E.M., 2010. Mechanisms of muscle weakness in muscular dystrophy. J. Gen. Physiol. 136, 29–34. https://doi.org/10.1085/jgp.201010436

Hermann, C., Saccon, R., Senthilan, P.R., Domnik, L., Dircksen, H., Yoshii, T., Helfrich-F??rster, C., 2013. The circadian clock network in the brain of different Drosophila species. J. Comp. Neurol. 521, 367–388. https://doi.org/10.1002/cne.23178

Hinits, Y., Osborn, D.P.S., Hughes, S.M., 2009. Differential requirements for myogenic regulatory factors distinguish medial and lateral somitic, cranial and fin muscle fibre populations. Development 136, 403–414. https://doi.org/10.1242/dev.028019

Hinits, Y., Pan, L., Walker, C., Dowd, J., Moens, C.B., Hughes, S.M., 2012. Zebrafish Mef2ca and Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. Dev. Biol. 369, 199–210. https://doi.org/10.1016/j.ydbio.2012.06.019

Horstick, E.J., Linsley, J.W., Dowling, J.J., Hauser, M.A., Kristin, K., Ashley-koch, A., Saint-amant, L., Satish, A., Cui, W.W., Zhou, W., Sprague, S.M., Stamm, D.S., Powell, C.M., Marcy, C., 2014. Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy. Nat. Commun. https://doi.org/10.1038/ncomms2952.Stac3

Jimenez, A.G., Dillaman, R.M., Kinsey, S.T., 2013. Large fiber size in skeletal muscle is metabolically advantageous. Nat. Commun. 2150. https://doi.org/10.1038/ncomms3150.Large

Johnston, I. a, Garcia de la Serrana, D., Devlin, R.H., 2014. Muscle fibre size optimisation provides flexibility to energy budgeting in calorie-restricted Coho salmon transgenic for growth hormone. J. Exp. Biol. 3392–3395. https://doi.org/10.1242/jeb.107664

Johnston, I., Temple, G., 2002. Thermal plasticity of skeletal muscle phenotype in ectothermic vertebrates and its significance for locomotory behaviour. J. Exp. Biol. 205, 2305–2322. https://doi.org/temperature activite physiologie comportement muscle nage adaptation evolution plasticite caracteristiques morphologiques

Johnston, I.A., 2006. Environment and plasticity of myogenesis in teleost fish. J. Exp. Biol. 209, 2249–2264. https://doi.org/10.1242/jeb.02153

Johnston, I.A., Bower, N.I., Macqueen, D.J., 2011. Growth and the regulation of myotomal muscle mass in teleost fish. J. Exp. Biol. 214, 1617–1628. https://doi.org/10.1242/jeb.038620

Johnston, I.A., Calvo, J., Guderley, H., Fernandez, D., Palmer, L., 1998. Latitudinal variation in the abundance and oxidative capacities of muscle mitochondria in perciform fishes. J. Exp. Biol. 201, 1–12.

Johnston, I.A., Kristjansson, B.K., Paxton, C.G.P., Vieira, V.L.A., Macqueen, D.J., Bell, M.A., 2012. Universal scaling rules predict evolutionary patterns of myogenesis in species with indeterminate growth. Proc. R. Soc. B Biol. Sci. 279, 2255–2261. https://doi.org/10.1098/rspb.2011.2536

Keren, A., Tamir, Y., Bengal, E., 2006. The p38 MAPK signaling pathway: A major regulator of skeletal muscle development. Mol. Cell. Endocrinol. 252, 224–230. https://doi.org/10.1016/j.mce.2006.03.017

Lazado, C.C., Kumaratunga, H.P.S., Nagasawa, K., Babiak, I., Giannetto, A., Fernandes, J.M.O., 2014. Daily rhythmicity of clock gene transcripts in Atlantic cod fast skeletal muscle. PLoS One 9, 1–12. https://doi.org/10.1371/journal.pone.0099172

Lee-Estevez, M., Figueroa, E., Cosson, J., Short, S.E., Valdebenito, I., Ulloa-Rodríguez, P., Farías, J.G., 2018. Zebrafish as a useful model for immunological research with potential applications in aquaculture. Rev. Aquac. 10, 213–223. https://doi.org/10.1111/raq.12156

Li, B., Dewey, C.N., 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12. https://doi.org/10.1186/1471-2105-12-323

Li, H., 2011. A statistical framework for SNP calling , mutation discovery , association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8

Macarthur, J.W., 1949. Selection for Small and Large Body Size in the House Mouse. Genetics 34, 194–209.

Mareco, E.A., Garcia de la Serrana, D., Johnston, I.A., Dal-Pai-Silva, M., 2015. Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (Piaractus mesopotamicus). BMC Genomics 16, 1–13. https://doi.org/10.1186/s12864-015-1423-6

Meissner, J.D., Chang, K., Kubis, H., Nebreda, A.R., Gros, G., Scheibe, R.J., 2007. The p38 ␣ / ␤ Mitogen-activated Protein Kinases Mediate Recruitment of CREB-binding Protein to Preserve Fast Myosin Heavy Chain IId / x Gene Activity in Myotubes \* 282, 7265–7275. https://doi.org/10.1074/jbc.M609076200

Moore, C. A., Parkin C. A., Bidet, Y., Ingham, P. W., 2007. A role for the Myoblast city homologues Dock1 and Dock5 and the adaptor proteins Crk and Crk-like in zebrafish myoblast fusion. Development. 134, 3145-53.

Nelson, B.R., Wu, F., Liu, Y., Anderson, D.M., McAnally, J., Lin, W., Cannon, S.C., Bassel-Duby, R., Olson, E.N., 2013. Skeletal muscle-specific T-tubule protein STAC3 mediates voltage-induced Ca2+ release and contractility. Proc. Natl. Acad. Sci. 110, 11881–11886. https://doi.org/10.1073/pnas.1310571110

Opazo, R., Valladares, L., Romero, J., 2017. Comparison of gene expression patterns of key growth genes between different rate growths in zebrafish (Danio rerio) siblings. Lat. Am. J. Aquat. Res. 45, 766–775. https://doi.org/10.3856/vol45-issue4-fulltext-12

Perdiguero, E., Ruiz-, V., Gresh, L., Hui, L., Ballestar, E., Sousa-victor, P., Baeza-raja, B., Esteller, M., Serrano, A.L., Erwin, F., 2007. Genetic analysis of p38 MAP kinases in myogenesis : fundamental role of p38 a in abrogating myoblast proliferation 1245–1256. https://doi.org/10.1038/sj.emboj.7601587

Perry, R.L.S., Yang, C., Soora, N., Salma, J., Marback, M., Naghibi, L., Ilyas, H., Chan, J., Gordon, J.W., McDermott, J.C., 2009. Direct Interaction between Myocyte Enhancer Factor 2 (MEF2) and Protein Phosphatase 1  Represses MEF2-Dependent Gene Expression. Mol. Cell. Biol. 29, 3355–3366. https://doi.org/10.1128/MCB.00227-08

Potthoff, M.J., Arnold, M.A., McAnally, J., Richardson, J.A., Bassel-Duby, R., Olson, E.N., 2007. Regulation of Skeletal Muscle Sarcomere Integrity and Postnatal Muscle Function by Mef2c. Mol. Cell. Biol. 27, 8143–8151. https://doi.org/10.1128/MCB.01187-07

Puri, P.L., Wu, Z., Zhang, P., Wood, L.D., Bhakta, K.S., Han, J., Feramisco, J.R., Karin, M., Wang, J.Y.J., 2000. Induction of terminal differentiation by constitutive activation of p38 MAP kinase in human rhabdomyosarcoma cells 574–584.

Ribas, L., Piferrer, F., 2014. The zebrafish (Danio rerio) as a model organism, with emphasis on applications for finfish aquaculture research. Rev. Aquac. 6, 209–240. https://doi.org/10.1111/raq.12041

Riley, L.A., Esser, K.A., 2017. The Role of the Molecular Clock in Skeletal Muscle and What It Is Teaching Us About Muscle-Bone Crosstalk. Curr. Osteoporos. Rep. 15, 222–230. https://doi.org/10.1007/s11914-017-0363-2

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2017. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquac. 1–13. https://doi.org/10.1111/raq.12193

Rubin, C.-J., Megens, H.-J., Barrio, A.M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A.M., Andersson, L., 2012. Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. 109, 19529–19536. https://doi.org/10.1073/pnas.1217149109

Shi, H., Scheffler, J.M., Pleitner, J.M., Zeng, C., Park, S., Hannon, K.M., Grant, A.L., Gerrard, D.E., 2008. Modulation of skeletal muscle fiber type by mitogen-activated protein kinase signaling. FASEB J. 22, 2990–3000. https://doi.org/10.1096/fj.07-097600

Taylor, M. V., Hughes, S.M., 2017. Mef2 and the skeletal muscle differentiation program. Semin. Cell Dev. Biol. 72, 33–44. https://doi.org/10.1016/j.semcdb.2017.11.020

Ulloa, P.E., Medrano, J.F., Feijo, C.G., 2014. Zebrafish as animal model for aquaculture nutrition research. Front. Genet. 5, 1–6. https://doi.org/10.3389/fgene.2014.00313

Vélez, E.J., Lutfi, E., Azizi, S., Perelló, M., Salmerón, C., Riera-Codina, M., Ibarz, A., Fernández-Borràs, J., Blasco, J., Capilla, E., Navarro, I., Gutiérrez, J., 2017. Understanding fish muscle growth regulation to optimize aquaculture production. Aquaculture 467, 28–40. https://doi.org/10.1016/j.aquaculture.2016.07.004

Wright, D.C., Han, D.H., Garcia-Roves, P.M., Geiger, P.C., Jones, T.E., Holloszy, J.O., 2007. Exercise-induced mitochondrial biogenesis begins before the increase in muscle PGC-1α expression. J. Biol. Chem. 282, 194–199. https://doi.org/10.1074/jbc.M606116200

Wu, P., Li, Y.L., Cheng, J., Chen, L., Zhu, X., Feng, Z.G., Zhang, J.S., Chu, W.Y., 2016. Daily rhythmicity of clock gene transcript levels in fast and slow muscle fibers from Chinese perch (Siniperca chuatsi). BMC Genomics 17, 1–14. https://doi.org/10.1186/s12864-016-3373-z

Yáñez, J.M., Newman, S., Houston, R.D., 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress, Frontiers in Genetics. https://doi.org/10.3389/fgene.2015.00128

Table 1. Male and female zebrafish body parameters after 4 generation of artificial selection for length

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Lineage* | *L* | *U* | *S* |
| *Male* | SL (mm) | 32.5±2.1b | 30.0±3.1a# | 30.2±2.1a |
|  | FL (mm) | 38.1±2.5b | 34.9±3.6a# | 35.3±2.5a# |
|  | TL (mm) | 39.7±2.7b | 36.3±3.7a# | 37.3±2.6a |
| *Female* | SL (mm) | 34.5±3.0b | 35.4±4.0a# | 32.3±2.7a |
|  | FL (mm) | 40.3±3.4b | 41.0±4.5b# | 37.9±3.1a# |
|  | TL (mm) | 41.7±3.4 | 42.6±4.6# | 39.8±3.1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Lineage* | *L* | *U* | *S* |
| *Male* | SL | 32.5±2.1b | 30.0±3.1a# | 30.2±2.1a |
|  | FL | 38.1±2.5b | 34.9±3.6a# | 35.3±2.5a# |
|  | TL | 39.7±2.7b | 36.3±3.7a# | 37.3±2.6a |
| *Female* | SL | 34.5±3.0b | 35.4±4.0a# | 32.3±2.7a |
|  | FL | 40.3±3.4b | 41.0±4.5b# | 37.9±3.1a# |
|  | TL | 41.7±3.4 | 42.6±4.6# | 39.8±3.1 |

Values are represented as mean±SD (N=50). Differences between lineages within each sex are indicated with different letters (p-val<0.05). Differences between sexes for each lineage are indicated with a # (p-val<0.05). L=large, U=unselected, S=small, SL=standard length, FL= fork length and TL= total length.

Table 2. Top 10 genes with significant changes in their expression between L and S lineages after DESEQ analysis of the transcriptome ranked by Log2 fold-change.

|  |  |  |  |
| --- | --- | --- | --- |
| Lineage | Gene Name | Log2FC | FDR |
| L | *irbp* | 5.4 | 0.03 |
|  | *tir4bb* | 5.4 | 2.58E-06 |
|  | *nr1d1* | 4.8 | 2.84E-07 |
|  | *ccdc147* | 4.4 | 0.04 |
|  | *per1a* | 3.9 | 0.03 |
|  | *helt* | 3.4 | 0.01 |
|  | *camk1ga* | 2.7 | 2.14E-04 |
|  | *per1b* | 2.6 | 5.29E-06 |
|  | *hspb6* | 2.6 | 0.01 |
|  | *znf395b* | 2.4 | 2.63E-04 |
|  |  |  |  |
| S | *prph* | 7.2 | 5.35E-12 |
|  | *stk33* | 6.0 | 1.70E-04 |
|  | *vwc2* | 5.7 | 1.67E-03 |
|  | *ccdc103* | 5.1 | 3.54E-03 |
|  | *dnase1l4.1* | 4.6 | 4.89E-03 |
|  | *hspa13* | 4.5 | 1.85E-16 |
|  | *pdha1b* | 3.7 | 0.02 |
|  | *il6* | 3.5 | 3.02E-03 |
|  | *plin2* | 3.1 | 3.29E-05 |
|  | *asb15b* | 3.1 | 0.03 |

Log2FC= Fold-change in 2 logarithmic scale. L=Large, S=Small, FDR=False Discovery rate.

Table 3. Differently expressed Splice Variants (SV) found between L and S lineages.

|  |  |  |  |
| --- | --- | --- | --- |
| *Gene location (Chromosome)* | *Gene name* | *Exons* *affected*  | *FDR* |
| Ch1 | *asah1b* | E3 (S) | 0.02 |
|  | *cd99* | E13(S)-E14(S) | 0.017-0.00 |
|  | *commd1* | E2(L) | 0.02 |
|  | *mrpl16* | E6(S) | 0.00 |
|  | *mrpl39* | E11(L) | 0.00 |
|  | *odz3* | E1(L) | 0.03 |
|  | *tmem39a* | E10(S) | 0.02 |
|  | *mynnl\** | E1(S) | 0.00 |
|  | *luc7l2* | E3(L) | 0.00 |
|  | *cd22l\** | E3(L)-E5(L)-E9(S) | 0.02-0.03-0.00 |
|  | *znf501l\** | E3(S) | 0.00 |
| Ch2 | *cyp7a1a* | E1(S) | 0.02 |
|  | *fam102ba* | E1(S) | 0.00 |
|  | *Si\_ch211\_14a17.11* | E3(S) | 0.00 |
|  | *u2surp* | E17(L) | 0.03 |
|  | *tp53\** | E8(S) | 0.04 |
| Ch3 | *lgals2a* | E1(S)-E3(L)-E4(L) | 0.00-0.00-0.02 |
|  | *uqcrc2a* | E6(L) | 0.00 |
|  | *luc7l3\** | E9(L) | 0.04 |
|  | *znf45l\** | E5(L) | 0.00-0.00 |
| Ch4 | *mapre3b* | E1(S) | 0.03 |
|  | *ms4a17a.6* | E6(S) | 0.00 |
|  | *ms4a17a.9* | E7(S) | 0.05 |
|  | *vegfab* | E3(S) | 0.03 |
|  | *nlrp12\** | E9(L) | 0.00 |
| Ch5 | *alad* | E13(S) | 0.00 |
|  | *aldh2.2* | E3(S)-E9(S)-E10(S)-E12(L)-E13(L)-E18(L) | 0.02-0.02-0.03-0.00-0.00-0.00 |
|  | *anxa1b* | E1(L) | 0.02 |
|  | *iqgap2* | E15(S) | 0.02 |
|  | *mthfd2* | E17(S)-E18(S) | 0.00-0.00 |
|  | *myhz1.1* | E8(S) | 0.00 |
|  | *myhz1.2* | E8(S) | 0.00 |
|  | *myhz2* | E23(L)-E24(L)-E25(L)-E26(L)-E33(S)-E34(S) | 0.02-0.00-0.00-0.00-0.00-0.00 |
|  | *nccrp1* | E6(S) | 0.04 |
|  | *prdx4* | E6(S) | 0.04 |
|  | *zgc\_63910* | E1(S) | 0.03 |
| Ch6 | *alg6* | E8(L) | 0.00 |
|  | *ercc3* | E15(S) | 0.00 |
|  | *lnpb* | E1(S) | 0.02 |
|  | *lrrc40* | E6(L) | 0.03 |
|  | *sfrs3a* | E1(S) | 0.04 |
|  | *cenps\** | E5(S) | 0.02 |
| Ch7 | *aktip* | E10(S) | 0.03 |
|  | *aprt* | E2(S) | 0.00 |
|  | *cd82a* | E10(S)-E11(S) | 0.00-0.00 |
| Ch7 | *nae1* | E17(S) | 0.05 |
|  | *nqo1* | E5(S) | 0.01 |
|  | *poln* | E25(S) | 0.03 |
|  | *tnnt3b* | E15(S) | 0.00 |
|  | *cda\** | E5(S) |  |
|  | *zgc\_112102* | E2(L) |  |
|  | *leap2\** | E2(L) | 0.01 |
| Ch8 | *cbwd* | E16(S)-E17(S) | 0.04-0.04 |
|  | *mkrn4* | E8(S) | 0.05 |
|  | *rps8b* | E1(S) | 0.04 |
|  | *odf2b* | E18(S) | 0.00 |
| Ch9 | *crfb5* | E3(S) | 0.00 |
|  | *egfl6* | E5(L) | 0.03 |
|  | *scrn3* | E6(S)-E7(S) | 0.00-0.00 |
|  | *stac3* | E12(L)-E13(L) | 0.00-0.00 |
|  | *tfpia* | E7(L) | 0.00 |
|  | *tmem182* | E1(L) | 0.02 |
|  | *tuba8l3* | E4(L) | 0.05 |
| Ch10 | *akap2* | E5(S) | 0.03 |
|  | *mef2ca* | E1(S) | 0.02 |
|  | *rps25* | E1(L) | 0.00 |
|  | *pcolcea\** | E1(S) | 0.04 |
| Ch11 | *agpat2* | E1(S) | 0.03 |
|  | *per3* | E1(L)-E2(L)-E4(L)-E5(L)-E21(S) | 0.02-0.00-0.03-0.00-0.01 |
|  | *sos1* | E16(S) | 0.00 |
|  | *hnrnpa1a\** | E1(S)-E2(S) | 0.00-0.03 |
| Ch12 | *pde6g* | E5(L) | 0.00 |
|  | *tefa* | E1(S) | 0.00 |
|  | *uqcrc2b* | E7(L) | 0.00 |
| Ch13 | *adss* | E1(S) | 0.03 |
|  | *epcam* | E1(L)-E2(L)-E6(S)-E10(S) | 0.00-0.03-0.00-0.00 |
|  | *h2afy2* | E5(L)-E6(L) | 0.00-0.00 |
|  | *lycat* | E10(S) | 0.00 |
|  | *mpc1* | E1(S) | 0.04 |
|  | *ndufaf1* | E2(L) | 0.00 |
|  | *oip5-as1\** | E1(L)-E2(S)-E3(S) | 0.00-0.00-0.00 |
|  | *soul2* | E4(S) | 0.00 |
|  | *srsf5a* | E8(L) | 0.00 |
|  | *pbld2\** | E9(L)-E10(L)-E14(L)-E15(L) | 0.00-0.00-0.00-0.00 |
|  | *dusp13a\** | E5(S) | 0.01 |
| Ch15 | *sgcg* | E1(L) | 0.00 |
|  | *ndufc2\** | E1(L)-E2(L) | 0.00-0.00 |
|  | *phykpl\** | E12(S) | 0.00 |
| Ch16 | *esrp1* | E1(L) | 0.02 |
|  | *vegfaa* | E7(S)-E8(S) | 0.03-0.00 |
|  | *leng8\** | E1(S) | 0.03 |
| Ch17 | *fcf1* | E1(S) | 0.00 |
|  | *fosl2* | E4(L)-E5(S) | 0.01-0.00 |
|  | *tmem30aa* | E1(S) | 0.03 |
| Ch18 | *cd276* | E1(S) | 0.00 |
|  | *fan1* | E13(L) | 0.00 |
|  | *hdc* | E13(L) | 0.05 |
|  | *ssr3* | E1(L) | 0.00 |
|  | *st14a* | E20(L) | 0.01 |
| Ch19 | *btr18* | E20(S) | 0.04 |
|  | *tbk1* | E21(S) | 0.03 |
| Ch20 | *dll4* | E10(S) | 0.03 |
|  | *lyrm2* | E10(S) | 0.03 |
|  | *mep1a.1* | E14(L) | 0.03 |
|  | *pcmt* | E1(S)-E2(S) | 0.03-0.00 |
|  | *sell* | E9(S) | 0.00 |
|  | *ccl38\** | E4(L) | 0.03 |
|  | *Wu\_fa18f11* | E1(S) | 0.01 |
| Ch21 | *bbs1* | E17(S) | 0.00 |
|  | *cct6a* | E3(S) | 0.02 |
|  | *cdk9* | E9(S) | 0.02 |
|  | *dpysl3* | E1(S) | 0.00 |
|  | *fstl4* | E4(L)-E5(L)-E13(S) | 0.00-0.01-0.00 |
|  | *fundc2* | E7(S) | 0.02 |
|  | *klhl8* | E2(S) | 0.00 |
|  | *mob1bb* | E1(S) | 0.01 |
|  | *mzt2b* | E1(S) | 0.05 |
|  | *ssh2b* | E17(S) | 0.02 |
|  | *timd4* | E1(S) | 0.03 |
|  | *txnl1* | E8(S)-E9(S)-E10(S) | 0.00-0.00-0.01 |
|  | *ube2b* | E7(L) | 0.00 |
|  | *zcchc10* | E5(S) | 0.00 |
| Ch22 | *capzb* | E1(S) | 0.03 |
|  | *dram2b* | E1(S) | 0.05 |
|  | *mccc1* | E21(S) | 0.00 |
|  | *plekhb2* | E1(S) | 0.00 |
|  | *gls2b\** | E2(S) | 0.01 |
|  | *tnfaip8* | E1(S) | 0.00 |
|  | *trim35* | E6(L) | 0.03 |
|  | *dnase1l4.2\** | E7(S)-E8(S) | 0.00-0.00 |
| Ch23 | *ctc1* | E22(S) | 0.00 |
|  | *ebp* | E1(S) | 0.00 |
|  | *myl9a* | E5(S)-E6(S) | 0.05-0.00 |
|  | *rca2.1* | E1(S)-E8(L) | 0.02-0.02 |
| Ch24 | *ctdspla* | E7(L) | 0.03 |
|  | *grnb* | E16(S) | 0.00 |
|  | *mybpc2b* | E3(L) | 0.00 |
|  | *smyhc2* | E2(S)-E3(S)-E6(S) | 0.02-0.00-0.03 |
|  | *smyhc3* | E22(S) | 0.02 |
| Ch25 | *sgcb* | E1(L)-E2(S) | 0.00-0.00 |
|  | *tnni2a.1* | E6(S) | 0.02 |
|  | *ube2q2* | E2(S) | 0.00 |
|  | *tpm1\** | E2(S) | 0.03 |

The exon column indicates the gene exon/s affected by the splice variant. The lineage where the SV with the missing exon is more abundantly found are indicated in parenthesis as L (Large) and S (Small). \*Indicates genes whose annotation has been updated using the latest version of the zebrafish genome (GRCz11).

Lineages were the exon are lost after alternative splicing are indicated in parenthesis as E=exon, FDR= False discovery rate.

Table 4. GO enrichment analysis from genes with significant differences in SV abundance between L and S lineages.

|  |  |  |  |
| --- | --- | --- | --- |
| *Lineage* | *Pathway ID/GO ID* | *Description* | *FDR* |
| L | PF03194 | LUC7 N\_terminus | 0.04 |
|  |  |  |  |
| S | GO:0005859 | muscle myosin complex | 0.001 |
|  | GO:0030017 | sarcomere | 0.005 |
|  | GO:0005737 | cytoplasm | 0.026 |
|  | GO:0032982 | myosin filament | 0.030 |
|  | GO:0030016 | myofibril | 0.046 |

GO, Pathway ID and description were obtained from STRING online tool. GO= Gene Ontology, L=Large, S=small, ID=identity and FDR= False Discovery Rate

Table 5. Exons located SNPs specific for L or S lineages identified in the transcriptome.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Coordinates* | *Lineage* | *Gene* | *Gene location* | *Reference* | *Alternative* |
| chr1:47,641,836-47,641,876 | L | *mhc1zca* | exon | G | T |
| chr12:19,437,129-19,437,169 | L | *atpaf2* | exon | C | T |
| chr12:20,472,005-20,472,040 | L | *gspt1* | exon | C | T |
| chr13:31,865,289-31,865,327 | L | *six4a* | exon | G | A |
| chr1:33,037,934-33,037,974 | S | *cd99* | exon | T | C |
| chr1:57,638,976-57,639,016 | S | *metrnl* | exon | C | G |
| chr10:8,287,458-8,287,498 | S | *tln1* | exon | C | T |
| chr12:19,551,929-19,551,969 | S | *neurl2* | exon | C | T |
| chr12:20,212,618-20,212,658 | S | *aco2* | exon | C | T |
| chr12:27,982,366-27,982,406 | S | *synpo2lb* | exon | C | A |
| chr12:27,982,383-27,982,423 | S | *synpo2lb* | exon | C | G |
| chr12:28,778,477-28,778,517 | S | *psme3* | exon | T | C |
| chr13:42,568,565-42,568,605 | S | *cisd1* | exon | A | G |
| chr20:16,239,376-16,239,416 | S | *zyg11* | exon | C | T |
| chr20:16,239,379-16,239,419 | S | *zyg11* | exon | C | A |
| chr3:53,252,390-53,252,430 | S | *trim35* | exon | C | T |
| chr6:28,939,846-28,939,886 | S | *atp1b1a* | exon | G | C |
| chr6:31,702,554-31,702,594 | S | *pgm1* | exon | A | G |
| chr6:32,089,738-32,089,778 | S | *dock7* | exon | C | T |

Coordinates indicated the location of the SNPs based on the zebrafish genome assembly Zv9 version. Reference nucleotide is based on the Zv9 version of the zebrafish genome.

S=Small, L=Large.

Table 6. GO enrichment analysis from genes showing SNPs differently fixed between L and S lineages.

|  |  |  |  |
| --- | --- | --- | --- |
| Lineage | *Pathway ID/GO ID* | Description | FDR |
| L | GO:0016011 | dystroglycan complex | 0.028 |
|  | GO:0016012 | sarcoglycan complex | 0.028 |
|  |  |  |  |
| S | GO:0022624 | proteasome accessory complex | 0.006 |
|  | GO:0044444 | cytoplasmic part | 0.006 |
|  | GO:0005622 | intracellular | 0.021 |
|  | 3050 | Proteasome | 0.016 |

GO, Pathway ID and description were obtained from STRING online tool. GO= Gene Ontology, L=Large, S=small, ID=identity and FDR= False Discovery Rate

Table 7. Global GO enrichment analysis of genes differently expressed genes, genes affected by SV variations and genes containing SNPs variant.

|  |  |  |  |
| --- | --- | --- | --- |
| Lineage | *Pathway ID/GO ID* | Description | FDR |
|  | GO:0042383 | sarcolemma | 0.000 |
| L | GO:0016011 | dystroglycan complex | 0.004 |
|  | GO:0016012 | sarcoglycan complex | 0.004 |
|  |  |  |  |
| S | GO:0033554 | cellular response to stress | 0.003 |
|  | GO:0006950 | response to stress | 0.036 |
|  | GO:004471 | single-organism biosynthetic process | 0.036 |
|  | GO:0061419 | positive regulation of transcription from RNA polymerase II promoter in response to hypoxia | 0.036 |
|  | GO:0046716 | muscle cell cellular homeostasis | 0.046 |
|  | GO:0005515 | protein binding | 0.017 |
|  | GO:0016829 | lyase activity |  |
|  | 1100 | Metabolic pathways | 0.017 |
|  | 4066 | HIF-1 signaling pathway | 0.017 |
|  |  |  |  |

GO, Pathway ID and description were obtained from STRING online tool. GO= Gene Ontology, L=Large, S=small, ID=identity and FDR= False Discovery Rate

**Figures**

Figure 1. Probability density functions of muscle fibre diameter from individuals of the L and S lineages.



Probability density functions (PDFs) of muscle fibre diameter. Zebrafish were selected based on size over four generations, and histological analyses were performed on 10 individuals per lineage. The dashed lines represent the average PDFs of groups and the solid line the PDF of the combined groups. The shaded area represents the 1000 bootstraps of the combined group. Green and red dashed lines represent S and L lineages respectively.

Figure 2. Fast skeletal muscle fibre area, number and density from individuals of the L and S lineages.



Boxplot showing differences in average fibre area (A), number (B) and density (C) between L (red) and S (green) lineages. Zebrafish were selected based on size over four generations, and histological analyses were performed on 10 individuals per lineage. Statistical differences are indicated as \* (p-val<0.05), \*\* (p-val<0.01) or \*\*\* (p-val<0.001).

Supplementary Materials

Supplementary File 1. DGE, SV and SNPs workflow.

Workflow diagram showing the main steps (green boxes) taken for the estimation of differentially expressed genes, splice variants abundance and SNPs variants identification from the RNAseq data. Programs used and their verions used in each step are indicated. Steps with key decisions (yellow rombus) and discarded data (red boxes) are also indicated.

Supplementary File 2. Principal Components analysis

Principal Components Analysis of the differential gene expression analysis from all genes analysed in the RNAseq experiment. A) Individuals from the L and S lineates are shown as dark or light blue circles respectively B) Individuals from the L and S lineages are represented by dots with their own colour code.

Supplementary File 3. Differential Gene Expression (DGE)

Genes identified during the DGE analysis between individuals from the S and L lineages after DESEQ analysis of the transcriptome as a whole.

Id= gene identity; baseMean= average normalized reads from the two lineages for an individual gene; baseMeanL= average normalized reads from individuals of the L-lineage for a given gene; baseMeanS= average normalized reads from individuals of the S-lineage for a given gene; FDR= False discovery rate.

Supplementary File 4. DGE individual counts

DESEQ normalized reads (counts) from each individual zebrafish sequenced for each gene identified.

Supplementary File 5. SNPs identified in the RNAseq data.

List of SNPs identified on S and L-lineages. Gene name, Ensembl identifier and reference sequence are based on the Zv9 version of the zebrafish genome.

Supplementary File 6. Differently retained single nucleotide polymorphisms (SNPs).

List of SNPs identified only in one of the lineages or the other. Coordinates are based on the zebrafish genome assembly Zv9. S=Small, L=Large.

UTR= Unstralated Region, S=small-lineage, L=Large lineage