



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*)

Citation for published version:

Robledo Sanchez, D, Rubiolo, JA, Cabaleiro, S, Martínez, P & Bouza, C 2017, 'Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*)' *Scientific Reports*, vol. 7, no. 1, 12105. DOI: 10.1038/s41598-017-12459-4

Digital Object Identifier (DOI):

[10.1038/s41598-017-12459-4](https://doi.org/10.1038/s41598-017-12459-4)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Scientific Reports

Publisher Rights Statement:

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



SCIENTIFIC REPORTS



OPEN

Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*)

Diego Robledo^{1,2}, Juan A. Rubiolo¹, Santiago Cabaleiro³, Paulino Martínez¹ & Carmen Bouza¹

Growth is among the most important traits for animal breeding. Understanding the mechanisms underlying growth differences between individuals can contribute to improving growth rates through more efficient breeding schemes. Here, we report a transcriptomic study in muscle and brain of fast- and slow-growing turbot (*Scophthalmus maximus*), a relevant flatfish in European and Asian aquaculture. Gene expression and allelic association between the two groups were explored. Up-regulation of the anaerobic glycolytic pathway in the muscle of fast-growing fish was observed, indicating a higher metabolic rate of white muscle. Brain expression differences were smaller and not associated with major growth-related genes, but with regulation of feeding-related sensory pathways. Further, SNP variants showing frequency differences between fast- and slow-growing fish pointed to genomic regions likely involved in growth regulation, and three of them were individually validated through SNP typing. Although different mechanisms appear to explain growth differences among families, general mechanisms seem also to be involved, and thus, results provide a set of useful candidate genes and markers to be evaluated for more efficient growth breeding programs and to perform comparative genomic studies of growth in fish and vertebrates.

Growth, the main target of aquaculture breeding programs¹, is a complex polygenic trait with high environmental influence. Fast and efficient growth shortens the time required to reach market size, increases food conversion efficiency, and, in sum, leads to higher profitability and food safety. Aquaculture is currently the fastest growing food industry²; however, despite its importance, genetic breeding programs are applied in no more than 10% of cultured species^{1,3}.

Growth is a trait of moderate to high heritability³ that displays a complex development and physiology. It involves both the increase of muscle cell number (hyperplasia) and cell size (hypertrophy), and the balance between both processes depends on the developmental stage, but it is also influenced by the environment. Growth is regulated by the hypothalamic-pituitary axis hormones which also control feeding behaviour⁴, yet local interactions between different muscle cell types and with other nearby tissues also condition muscle growth. Additionally, growth rate is regulated by other physiological and environmental factors, such as food intake, temperature, diet composition, age or sex⁵. Furthermore, identifying key growth-related genes is more complex in teleost than in other vertebrates, since this group of fish has undergone an additional round of genomic duplication⁶ involving neo- or sub-functionalization processes. Yet, although a complex task, understanding the genetic basis of growth variation might aid to achieve more efficient breeding programs. Despite its good response to traditional selection, the sole application of phenotypic and relatedness information to improve growth rate might determine the loss of relevant genetic variation affecting the medium- and long-term performance of breeding programs^{7,8}. In this sense, knowing the molecular mechanisms, genes and pathways involved in growth can have multiple benefits. Beyond increasing our basic knowledge with important implications in different research areas

¹Departamento de Zooloxía, Xenética e Antropoloxía Física, Facultade de Veterinaria, Universidade de Santiago de Compostela, 27002, Lugo, Spain. ²The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG, United Kingdom. ³Cluster de Acuicultura de Galicia (Punta do Couso), Aguiño-Ribeira, 15695, Spain. Correspondence and requests for materials should be addressed to C.B. (email: mcarmen.bouza@usc.es)

	Muscle	Brain
Number of transcripts	30,674	49,034
Number of genes	17,173	23,104
N50	3,734 bp	5,062 bp
N90	1,517 bp	2,122 bp
Total transcripts >500 bp	29,901 (97.5%)	47,517 (96.9%)
Average transcript length	2,859 bp	3,788 bp

Table 1. Muscle and brain transcriptome statistics.

related to growth, identification of allelic variants of moderate to large effect could be useful in marker-assisted selection (MAS), provided a cost-benefit balance. Finally, genome-editing techniques may allow recovering or even introducing variants not present in the population of interest, increasing genetic gain⁹. As such, different studies have been focused on understanding growth differences in different fish species in the last decade^{10–15}.

Turbot is the flatfish with highest aquaculture production, mainly held by Spain and China. Females reach sexual maturity at ~3 years (~50 cm length) and can spawn between 5 and 10 million eggs per season. Reproduction can be achieved along the whole year under a variable controlled photoperiod¹⁶. Thus, production can benefit from the high fecundity of a single selected couple over a relatively long period. Several studies have been carried out in this species to identify quantitative trait loci (QTL) associated with growth^{17–19}, but the information provided, though useful, is still of limited application. Knowing the genes or even the allelic variants behind growth differences is important, not only for a more efficient selection, but also to guide future studies in turbot and other fish species. However, functional studies on differential growth have not been performed in turbot yet.

To gain knowledge on the key genes involved in growth, we carried out a transcriptomic study in fast- and slow-growing turbot to identify differential expression profiles and genetic markers between extreme growth phenotypes. Samples of muscle and brain tissues were taken from family-matched fast- and slow-growth fish selected from eight unrelated full-sib families, and their transcriptomes compared by RNA-Seq. These organs play a key role in the regulation of somatic growth, and muscle also represents the main edible part of fish^{5,20}. In order to evaluate a high number of individuals and to avoid family bias, a pooling strategy was followed. Differential gene expression and marker association between high- and low-growth turbot were obtained. The detected candidate genes and markers were mapped using the recently assembled genome of the species and their positions were compared with previously reported QTLs²¹. The most interesting genes and markers were validated using individual Real-Time PCR (qPCR) and SNP genotyping to disentangle general and family-specific factors involved in turbot growth.

Results

RNA sequencing output. A total of ~183 million paired-end (PE) reads were generated, for an average of ~15 million reads for each of the twelve samples analyzed. After filtering, ~171 million reads were retained (93.44%; >13 million reads per sample). On average, 88.4% of the PE filtered reads aligned to the turbot genome at unique (83.9%) or multiple (4.5%) genomic positions, whereas the remaining 11.6% did not match. The resulting muscle and brain transcriptomes revealed a total of 30,674 and 49,034 transcripts, respectively, belonging to 17,173 and 23,104 genes (Table 1), which included some non-annotated genes in the turbot genome²¹. Annotation against NCBI's non-redundant protein database was successful for 87.2% and 86.7% of the muscle and brain transcriptomes, respectively.

Differential gene expression. A total of 17 and 4 differentially expressed (DE) genes were detected in muscle and brain, respectively, when comparing fast- (FG) and slow- (SG) growing pooled samples using the log₂ FG/SG fold change (FC) (Table 2). FC values were moderate, ranging from 1.18 to -0.86 FC values, positive FC corresponding to higher expression in FG compared to SG (up-regulated onwards), and negative FC to lower expression in FG compared to SG (down-regulated onwards). Since the number of DE genes was low, an extended list of those genes with FDR < 0.3 was also explored (Supplementary Dataset 1). Overall expression patterns were consistent across the three biological pools tested (Fig. 1), especially considering that FG and SG pools belong to eight different families and that the mechanisms driving growth differences are not necessarily the same for all the families. No significant sex-ratio differences between pools and families were found, supporting that sex was not a confounding factor in this study. KEGG enrichment analysis revealed the prevalence of certain pathways among the extended list of DE genes. The pathways consistently enriched were “Glycolysis/Gluconeogenesis” and “Starch and sucrose metabolism” (log₂ fold enrichment of 5.3 and 4.8, respectively) (Supplementary Figure 1), involving a set of 16 up-regulated genes coding for enzymes of the glycolytic pathway.

Among other interesting up-regulated genes in muscle, we detected six myosin heavy chain genes (FDR p-values ranging from 0.016 to 0.295 and FC from 1.00 to 0.48), strongly related to muscle growth; and the elongation factor 1-alpha (p-val = 0.099, FC = 0.70), which suggests enhanced protein synthesis for muscle growth. Interestingly, growth hormone receptor 2 (GHR2) was down-regulated (p-val = 0.063, FC = -1.01), showing an opposite pattern to that expected according to the function of the growth hormone receptor in mammals. Furthermore, the cytosolic creatine kinase (CKs) was up-regulated (p-val = 0.081, FC = 0.59), while its mitochondrial counterpart (CKm) was down-regulated (p-val = 0.170, FC = -0.67), in agreement with the higher production of cytosolic ATP due to the activation of the glycolysis pathway in FG fish. Accordingly, the mitochondrial oxidative phosphorylation showed a lower activity in FG turbot as suggested by down-regulation of ADP-ATP

Gene	FDR p-val	FC	LG	Position
Muscle				
Triosephosphate isomerase	0.001	0.79	LG02	13,361,150
Glucose-6-phosphate isomerase	0.004	0.77	LG04	21,441,582
Beta-enolase	0.008	0.77	LG16	147,95,147
C-binding protein	0.015	1.13	LG10	17,180,345
Glyceraldehyde-3-phosphate dehydrogenase	0.016	0.72	LG02	13,180,325
Phosphoglycerate kinase 1	0.016	0.72	LG08	14,493,171
Fish-egg lectin	0.016	0.99	UN	UN
Myosin heavy fast skeletal muscle	0.016	0.75	LG06	19,857,385
Extensin-2	0.017	1.18	LG09	14,700,577
Mesothelin	0.023	1.10	LG03	14,490,203
Fructose-bisphosphate aldolase a	0.024	0.67	LG21	9,095,011
Flocculation protein flo11	0.027	0.80	LG01	3,075,171
Type ii cytoskeletal 8	0.037	0.95	LG10	18,632,403
Gelsolin	0.037	0.96	LG05	15,564,429
Protein-glutamine gamma-glutamyltransferase e	0.037	0.98	LG05	10,093,262
Thymosin beta-a	0.044	0.87	LG07	7,429,705
Chitin synthase 2	0.044	1.03	LG22	917,070
Brain				
Non annotated transcript	0.000	-0.86	LG03	12,013,135
Histone-lysine n-methyltransferase	0.005	0.60	LG22	1,690,914
Hyaluronan and proteoglycan link protein 1	0.039	0.60	LG12	959,526
Synaptoporin	0.039	0.66	LG11	5,169,652

Table 2. Differentially expressed genes between fast- and slow-growing turbot. FDR p-val: false discovery rate corrected p-value; FC: log2 fold change; LG: linkage group; UN: gene located in a scaffold not assigned to a LG of the turbot genetic map.

translocase 2 (p-val = 0.051, FC = -0.64), which codes for a membrane protein responsible for transporting ATP from the mitochondria to the cytosol.

The most interesting genes detected in the family pools were evaluated in individual samples by qPCR: CKs, CKm, GHR2, lactate dehydrogenase (LDH), muscle glycogen phosphorylase (PYGM) and transforming growth factor β 1 (TGF β 1). Although a notable concordance was observed between individual qPCR and pooled RNA-Seq data when comparing the expression pattern of FG and SG fish, some discrepancies were detected (Fig. 2), suggesting that some mechanisms driving growth might be different among families. Specifically CKm and TGF β 1 showed important variation among families. In fact, when individual qPCR data were pooled following the same distribution of individuals as for RNA-Seq, correlation was positive and significant between both datasets, except for CKm and TGF β 1 (Supplementary Table 1).

A lower number of DE genes were detected in the brain and not so straightforwardly related to growth as in the case of muscle. Among these, we observed down-regulation of phosphorilase kinase alpha 2 (p-val = 0.172, FC = -0.49) encoding for a subunit of the phosphorylase b enzyme, which is expressed in liver and brain. This gene modulates glycogen catabolism and plays an important role in supplying cell energy²². Also, we detected several up-regulated genes associated with sensory control, such as t-box brain protein 1 (p-val = 0.088, FC = 0.61), a transcription factor required for normal brain development and expressed in the olfactory bulb among other brain regions²³; synaptoporin (p-val = 0.039, FC = 0.66), a vesicular transport protein involved in synaptic vesicles and plasticity²⁴; and olfactomedin (p-val = 0.190, FC = 0.46), a secreted glycoprotein involved in regulating chemosensory cilia in olfactory neurons and optic nerve extension²⁵. KEGG pathway and GO enrichment did not reveal any significant term.

Differential SNP allelic frequencies (DAfreq). A total of 146,217 SNPs were called using both muscle and brain reads. The sequences containing these SNPs were annotated according to the turbot genome information²¹. A total of 45,348 SNPs (31%) were found in coding regions, and among these, 16,133 corresponded to non-synonymous mutations, representing an overall Ka/Ks ratio of 0.55. Allelic frequencies were estimated for FG and SG fish and the significance of allelic divergence tested (DAfreq onwards). Significant DAfreq between FG and SG were detected for 75 SNPs in muscle and for 358 in brain (Supplementary Dataset 2). These 435 DAfreq SNPs identified were located in 326 genes: i) 43% in untranslated regions (3' and 5' UTR); ii) 49.2% in coding regions of annotated genes, including mostly synonymous (74%), but also missense variants (24%), and 2% located in splicing regions; iii) 3.4% in introns, and iv) 4.4% in non-annotated genes. The genomic position of these SNPs in the turbot genome²¹ and their co-localization with previously described growth-associated markers^{17,18} was evaluated (Fig. 3). DAfreq SNPs were widespread throughout the genome, although some specific regions at particular linkage groups (LG) of the turbot genetic map showed a higher concentration (i.e. LG02, LG05 or LG13), and some of them co-localized with previously reported growth associated markers. We also

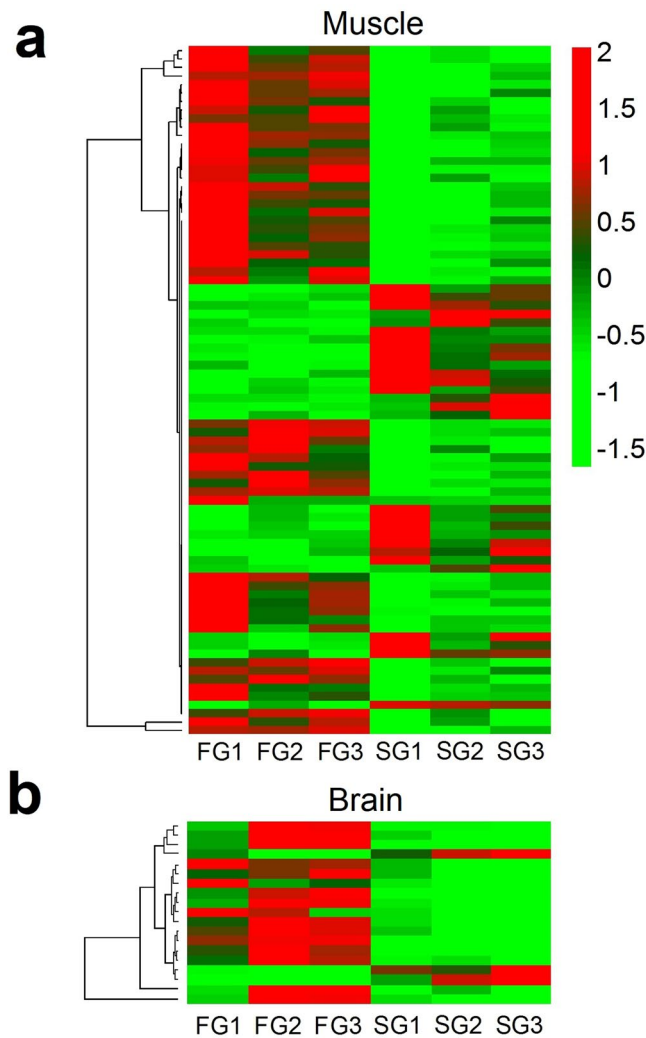


Figure 1. Muscle and brain heatmaps. Heatmaps of genes showing FDR corrected p-values < 0.3 for fast- (FG) and slow- (SG) growing turbot in: (a) muscle and (b) brain. Displayed are DESeq 2 v.3.2 (Love *et al.*⁵⁹) normalized counts for each sample and gene scaled by gene. Genes were hierarchically clustered according to their gene expression using Pearson correlation as a distance measure.

checked for the correspondence between DAFreq SNPs and DE genes in the turbot map (Fig. 3). A selection of the most suggestive SNPs is shown in Table 3. Several DAFreq SNPs co-localized in a narrow region of LG2, with two glycolysis up-regulated genes in muscle, glyceraldehyde-3-phosphate dehydrogenase and triosephosphate isomerase, including a synonymous SNP variant of the gene glycogen synthase kinase 3, which is involved in the negative regulation of skeletal muscle growth (Table 3; Supplementary Dataset 2). Highly significant SNPs showing a large DAFreq (>0.4) were also observed in two expressed genes in brain coding for sensory-related vesicle proteins, synaptotagmin 1 and synaptophysin, mapping at LG11 and LG16, respectively. Further, the muscle up-regulated deoxynucleoside triphosphate SAMHD1 gene showed a missense mutation located at LG10.

Except for the SNP corresponding to fast myosin heavy chain protein, all the SNPs assayed were technically feasible and could be characterized using the SNaPshot methodology (8 out of 9). These SNPs were successfully genotyped in all FG and SG samples, and significant frequency differences for three novel candidate genes were individually validated (Table 3; Supplementary Table 2). Worth noting is the DAFreq observed for SAMHD1, which also presented differential expression between extreme growth phenotypes in this study.

Discussion

Transcriptomes of fast- (FG) and slow- (SG) growing turbot involving eight different families from a breeding population of Atlantic origin were compared in this study. Several genes showing differential gene expression (DE) and/or divergent allelic frequencies at SNP markers (DAFreq) were found when comparing both phenotypic groups. The use of pools of families represents a cost-effective strategy for a preliminary exploration of the most important factors involved in differential growth. This approach enables identifying candidate genes, pathways and genomic regions associated with differential growth rates, as recently reported in other aquaculture species^{26–29}. This strategy was here applied for the first time in the turbot, the most important farmed flatfish worldwide³⁰. We identified relevant candidate genes and genomic regions associated with differential growth in

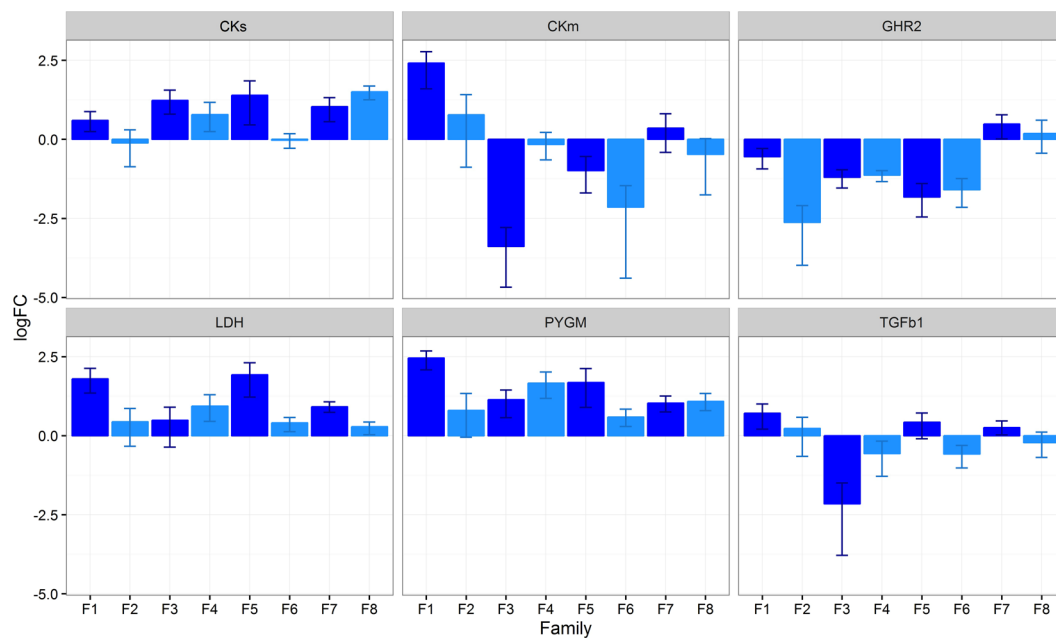


Figure 2. Family expression differences for six candidate genes. Log₂ fold change differences between fast- and slow-growing fish in each family for six candidate genes studied by qPCR. Positive fold changes indicate over expression in fast-growing fish. The genes are lactate dehydrogenase (LDH), muscle glycogen phosphorylase (PYGM), growth hormone receptor 2 (GHR2), cytosolic creatine kinase (CKs), mitochondrial creatine kinase (CKm), and transforming growth factor β 1 (TGfb1).

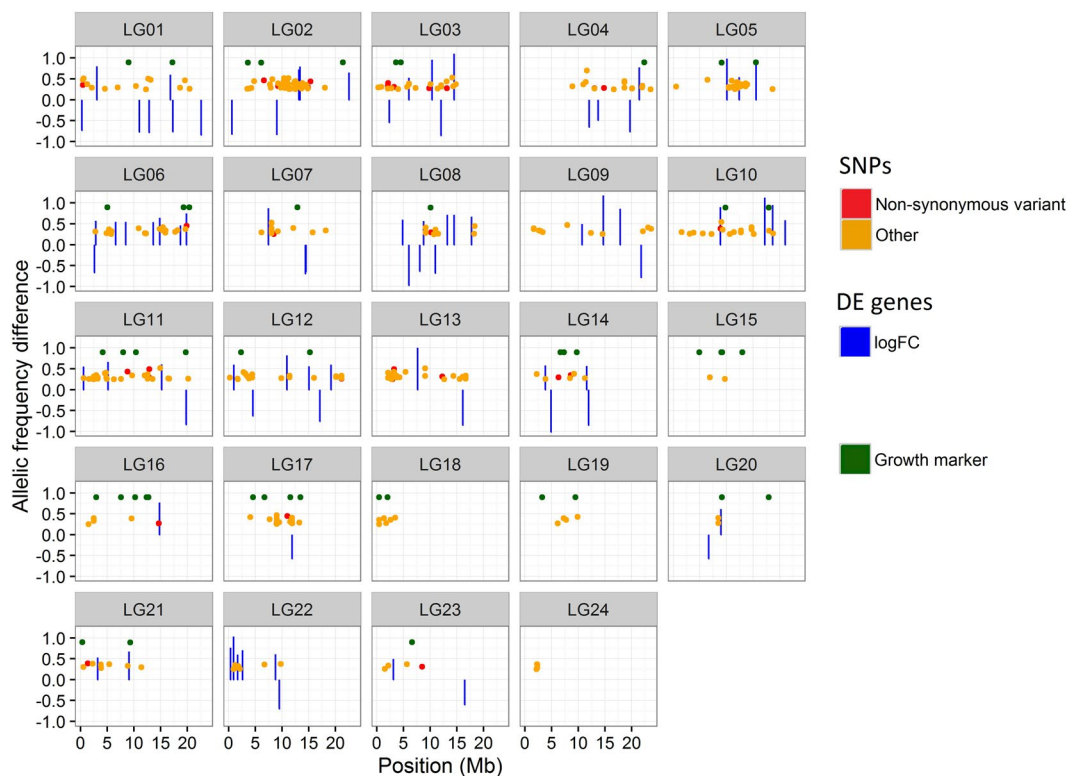


Figure 3. SNPs with differential frequencies along the turbot genome. SNPs showing allelic frequency differences between fast- and slow-growing turbot are shown in red (non-synonymous mutations) or orange (any other type of mutation) in the different chromosomes of the turbot genome. The fold change of those genes with FDR corrected p-values < 0.3 are shown in blue and the position of genetic markers previously found associated with growth traits are shown in green.

Chr	Pos	V1/ V2	DAfreq RNA-seq (−log p-val)	DAfreq SNaPshot	Gene	Annotation
LG01	12,679,184	G/A	0.51 (6.8)	0.11	insulin receptor	Synon
LG02	13,982,806	T/C	0.32 (8.2)	0.15 ^{**/s}	glycogen synthase kinase 3	Synon
LG02	15,305,418	A/T	0.44 (6.9)	0.13	glutamate receptor 2	Missense: Tyr453Phe
LG04	11,572,319	T/A	0.70 (7.8)	0.08	Cklf-like marvel transmembrane domain-containing protein 3	3'UTR
LG06	19,857,348	T/A	0.45 (9.5)	—	myosin heavy chain, fast skeletal muscle	Missense: Thr40Ser
LG10	8,837,360	T/G	0.39 (9.7)	0.30 ^{**/*}	deoxynucleoside triphosphate SAMHD1	Missense: Lys151Gln
LG11	14,899,119	A/G	0.52 (7.2)	0.21 ^{*/s}	synaptotagmin 1	Intron
LG16	9,519,616	C/A	0.39 (7.9)	0.05	synaptophysin protein 1	5'UTR
LG24	2,244,489	A/G	0.37 (7.8)	0.07	glucose-6-phosphatase 3	Synon

Table 3. Selection of SNPs showing divergent allelic frequencies between high and low growth turbot. For each SNP the chromosome (Chr), the position in the chromosome (Pos), the major variant (V1), the minor variant (V2), the allelic frequency difference (DAfreq) and Fisher's test $-\log_{10}$ p-value ($-\log pval$) obtained from the RNA-seq, the DAfreq and significance of genotypic/allelic divergence (G-tests; s: $p < 0.10$, * $p < 0.05$, ** $p < 0.01$) obtained from SNaPshot, the gene the SNP belongs to (Gene) and the SNP annotation regarding the gene (Annotation; Synon: synonymous) are shown.

this species, essential for understanding the mechanisms underlying phenotypic variation of this complex trait and useful for genetic assistance in breeding programs. The transcriptome profiles at 270 dpf here obtained also contribute key information to be further compared with other life-cycle stages with different weight of hyperplasia and hypertrophy into myogenesis⁵. Some candidate genes were individually validated using qPCR in FG and SG fish in turbot families. Aggregated individual qPCR results were mostly in agreement with RNA-Seq pools; however, some genes showed inter-family transcriptomic differences, suggesting that functional strategies associated with growth rate in turbot may be diverse among families, as previously reported for other complex traits in teleosts³¹. Our results are consistent with previous growth-association studies in turbot and other fish, where some genomic regions showed association across families, while others were family specific^{17,18,32}. Furthermore, SNaPshot assays validated the association of some growth-related SNP variants identified through differential allele frequencies between FG and SG pools in the transcriptomic analysis. These gene markers could be useful for studies in other families and/or populations, or even for genome editing approaches for those potentially functional variants.

Up-regulation of all glycolytic enzymes along with lactate dehydrogenase was consistently observed in our study, which supports a higher oxidation rate of glucose by fermentation in muscle to increase growth rate. Unlike other vertebrates, fish muscle is separated in discrete layers of different fibre types. Vertebrates have two main types of striated muscle fibres, red and white, which in fish specialize in low speed swimming and bursts of maximum speed, respectively³³. Red muscle fibres (less than 10% of the myotomal musculature) mainly depend mainly on aerobic, mitochondria driven, energy metabolism³⁴. On the contrary, the metabolism of white muscle fibres (never less than 70% of fish skeletal muscle) relies mainly on anaerobic glycolysis, aided by the intracellular energy shuttle of cytosolic creatine kinase (CKs)³⁵. The higher expression of glycolytic enzymes, lactate dehydrogenase (LDH) and CKs in fast-growing turbot strongly suggests a higher white muscle metabolic activity, which would presumably explain their higher growth. Glycolysis has also been found up-regulated in the muscle of domestic rainbow trout compared to their wild counterparts³⁶, and also in faster-growing fish of the same species; this up-regulation has been associated with an increased muscle energy demand in fast-growing animals²⁹. However, we cannot discriminate cause and consequence from our results. Either high growth rate demands an increasing anaerobic energy metabolism or, conversely, higher activity of glycolytic enzymes, LDH and CKs, results in higher growth rate. Further work is needed to clarify this issue.

Two suggestive candidates for follow-up experiments are the glycolytic enzymes triosephosphate isomerase 1 (TPI) and glyceraldehyde-3-phosphate dehydrogenase (G3PDH). Both genes were up-regulated in turbot muscle and they are closely linked in LG2 (181 kb; ~ 0.3 cM²¹). Further, these two genes lie within the confidence interval of a growth-related QTL in turbot^{17,19}, in the vicinity (< 500 kb; ~ 1 cM¹⁸) of several DAfreq SNPs. TPI markers have been found associated with length variation in other fish³⁷, and its expression has also been correlated with body size, muscle fat content and growth potential in bulls³⁸, as also observed for G3PDH in fast-growing rainbow trout²⁹. Other clusters of a few DE genes in short genomic stretches (< 1 Mb) were observed in our study, suggesting positional clustering of co-expressed genes of related pathways as reported for Arctic char³¹. The integration of structural and functional data provides clues for understanding the genetic architecture underlying growth in turbot and fish in general. The co-localization of several DE genes with DAfreq SNPs and growth-related QTL suggests that multiple genes could be underneath these growth-related genomic regions in turbot, as observed for complex traits in other fish³¹.

A higher activity of the glycolytic pathway was also supported by the up- and down-regulation of cytosolic (CKs) and mitochondrial (CKm) creatine kinase, respectively, in the muscle of turbot. Creatine is used as an energy reservoir in its phosphorylated form, and CK is responsible for creatine phosphorylation, capturing

available cellular energy; but it can also catalyse the reverse reaction, using phosphocreatine to regenerate ATP, acting as a metabolic regulator³⁹. While CKs activates glycolysis by removing the ATP generated, CKm is coupled to the mitochondrial oxidative phosphorylation capturing the ATP generated^{40,41}. In this context, the up-regulation of CKs observed in FG turbot is likely consistent with the higher anaerobic metabolism in the white muscle. Both CKs and CKm genes have been reported to be highly regulated in muscle between large and small rainbow trout, associated with reactive oxygen stress imbalance²⁹.

Growth hormone receptor 2 (GHR2), a fish-specific paralogous gene, was found down-regulated in the muscle of six out of eight turbot families, a similar expression pattern to that observed in fast-growing rainbow trout^{29,41}. Growth hormone pathways are crucial for vertebrate growth, and specifically the variation at GHR2 has been also associated with growth in different tilapia species⁴². GHR2 is involved in fish myogenesis and compensatory growth⁴³, although further functional analyses are required to understand its signal transduction mechanisms. Interestingly, GHR2 is located within the confidence intervals of a QTL explaining up to 18% of the phenotypic variance for body weight and length at LG14 in turbot¹⁷.

Unlike muscle, only a few DE genes were detected in brain when comparing FG and SG turbot. Detection of DE genes could be hindered by sampling the whole brain instead of dissecting specifically the hypothalamus or hypophysis, a very difficult task before the onset of sex growth dimorphism in turbot. Even so, to our knowledge, this is the first transcriptomic study showing brain differences between FG and SG in turbot. Interestingly, some up-regulated genes had been previously related to sensory control, including t-box brain protein 1, synaptotagmin and olfactomedin, the latter two associated with domestication behaviour and feeding preference in other fish, and related to higher growth rate^{44,45}. DAFreq SNPs were also found in two relevant genes coding for vesicle proteins, synaptophysin, associated with behavioural variation in domesticated zebrafish⁴⁴, and synaptotagmin 1, related to memory formation and retention, and feeding response in fish⁴⁶. Moreover, these SNPs mapped within the intervals of growth QTL explaining more than 10% of observed variance¹⁷, where major candidate genes are also located (lumican at LG16 and myogenin at LG11, respectively)¹⁹. Synaptotagmin 1 SNP was validated using SNaPshot, showing significant and suggestive differences between FG and SG fish at genotypic and allelic level, respectively. All these data suggest that sensory-related genes and pathways regulating learning and memory in brain might explain differences in feeding behaviour and, thus, be involved in turbot growth rate.

Numerous DAFreq SNPs were detected in muscle and brain transcriptomes distributed across all chromosomes of the turbot genome, even in linkage groups such as LG9 and LG22, where growth QTLs have not been previously reported²¹. Although many of them may reflect linkage disequilibrium with causal polymorphisms or be false positives, others may explain growth differences in turbot, especially if they co-localize with previously reported QTL, or better, if they are within candidate genes showing differential expression. Thus, this list represents a catalog of candidate markers affecting gene expression and/or protein function with small to moderate effects associated with growth. Some differential SNPs concentrated in genomic regions co-localizing with previous growth markers reported in turbot²¹, particularly focused on those found across different families, which may represent the QTL variation in the wild Atlantic population of origin^{17,18}. For example, a cluster of differential variants were detected in LG12, close to 3/9CA15 explaining up to 9% of length variance; or in LG16, close to Sma-USC-E11, explaining up to 25% of Fulton's factor variance¹⁷, and apparently conserved in different fish species²¹. The largest accumulation of DAFreq SNPs is observed in LG2 where, as previously mentioned, there are also two glycolytic enzymes showing differential expression. One of the SNPs in these regions is located in glycogen synthase kinase 3 (GSK3). This SNP has been validated individually and shows significant genotypic frequency differences between FG and SG fish. GSK3 is involved in the negative regulation of muscle growth and differentiation⁴⁷. A decrease in the activity of this protein is positively related to muscle growth. Further molecular and biochemical analysis are needed to determine if the identified SNP has an impact on GSK3 and consequently on the differential growth rates.

DAFreq SNPs within DE genes represent the most robust evidence for causality. In this study, DAFreq SNPs were often found in the vicinity of DE genes, but rarely mapped within DE genes. Two noteworthy exceptions were missense variants in the gene SAM and HD domain containing deoxynucleoside triphosphate (SAMHD1) located in linkage group LG10 and the myosin heavy fast skeletal gene located at turbot LG6. SAMHD1 was found to be up-regulated and its non-synonymous SNP individually validated, being significant at allelic ($p = 0.034$) and genotypic ($p = 0.008$) level. This gene encodes a major regulator of dNTP reservoir in the cell, playing an essential role in cell-cycle progression and cell proliferation, with impact on viral replication and cellular DNA polymerization^{48,49}. SAMHD1 dNTP reduction has been associated with a reduction of the number of mitochondria per cell by impeding mtDNA replication⁴⁹. This is consistent with down-regulation of CKm in fast-growing turbot. Furthermore, this gene mapped within the confidence interval of a suggestive QTL for growth in turbot¹⁸. On the other side, the myosin heavy fast skeletal gene was in the vicinity (<500 kb; ~1 cM) of a growth QTL marker¹⁸ and was up-regulated together with other myofibrillar component genes, as also reported in fast-growing rainbow trout²⁹. These observations support their interest as growth candidates for future studies.

Conclusion

The transcriptome comparison of fast- and slow-growing turbot from eight full-sib unrelated families enabled the identification of genes showing differential expression and SNP association with growth across the turbot genome. Sound candidate genes and genomic regions emerged as potential biomarkers of differential growth in turbot. Up-regulation of the glycolytic pathway in fast-growing turbot was consistent with a lower mitochondrial activity presumably related to higher proportion of white muscle fibres. A handful of sensory-related genes showing differential brain expression or divergent allelic frequencies suggest differential behaviour related to feeding between growth phenotypes. Nevertheless, family transcriptome variability of particular genes suggests that heterogeneous mechanisms might also be involved in differential growth rate, in accordance with the polygenic nature of this complex trait. Future studies should aim to expand these findings in other developmental stages

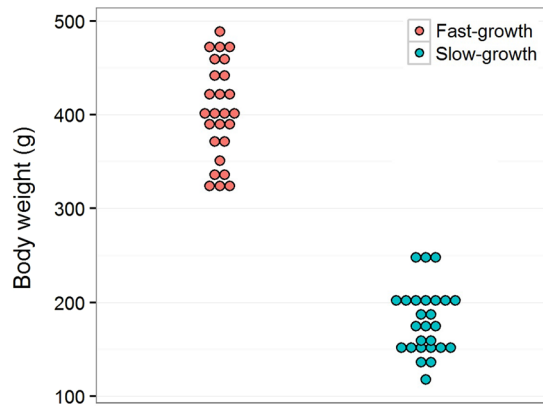


Figure 4. Sampling design. Body-weight distribution of fast- (FG) and slow- (SG) growth sampled groups.

to better comprehend growth differences in turbot. Nonetheless, all genomic resources gathered here have been useful for validating differential genetic variation between growth phenotypes and will aid in further comparative studies of growth in fish, providing a set of markers with an effect in turbot growth to be evaluated in future studies.

Methods

Families and sampling. Turbot from eight unrelated full-sib families coming from a pedigreed breeding strain of Atlantic origin were reared in tanks at 18 °C in the facilities of CETGA (Aquaculture cluster of Galicia; Ribeira, Spain). Fish were weighed at 270 days post fertilization (dpf), before significant differences in growth rate between males and females occurred⁵⁰. The 3–4 largest and smallest fishes from each family were sampled, rendering a total of 26 individuals for each phenotype: Fast- (FG; average weight = 402 g, SD = 51.38) and slow (SG; average weight = 180 g, SD = 35.65) growing turbot (Fig. 4; Supplementary Dataset 3). We genotyped all animals using a sex-linked molecular tool^{51,52} to identify putative sex-associated growth differences using contingency chi-square tests ($p < 0.05$). Muscle and brain tissues were collected and individually conserved in RNAlater (Qiagen, Valencia, CA). Animals were treated according to the Directive 2010/63/UE of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for experimentation and other scientific purposes. Experimental protocols were approved by the Institutional Animal Care and Use Committee of the University of Santiago de Compostela (Spain).

RNA sequencing. RNA extraction was performed using the RNeasy mini kit (Qiagen) with DNase treatment following manufacturer's instructions. RNA quality and quantity were evaluated in a Bioanalyzer (Agilent Technologies, USA) and in a NanoDrop® ND-1000 spectrophotometer (Nanodrop Technologies, USA), respectively. Three libraries were constructed for each growth group (FG and SG) and tissue (muscle and brain), totaling 12 libraries. Eight to nine individuals from 2–3 different families were pooled in each library: 3 FG and 3 SG libraries for each tissue (Supplementary Dataset 3), every family being present in both FG and SG groups. Pooled samples were barcoded and prepared for sequencing at the Wellcome Trust Centre for Human Genetics, Oxford, where 100 bp paired-end (PE) reads were obtained on an Illumina HiSeq. 2000. The quality of the sequencing output was assessed using FastQC v.0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Quality filtering and removal of residual adaptor sequences was conducted on read pairs using Trimmomatic v.0.32⁵³. Specifically, Illumina adaptors were clipped from the reads, leading and trailing bases with a Phred score < 15 were removed and the read trimmed if a sliding window average Phred score over four bases was less than 20. Only reads where both PE reads encompassed a length greater than 32 bp post-filtering were retained. Filtered reads were mapped to the recently assembled turbot genome¹⁸ using Tophat2 v.2.0.12⁵⁴ that leverages the short read aligner Bowtie2 v.2.2.3⁵⁵ with default parameters. Cufflinks v.2.2.1⁵⁶ was used to build gene transfer format (GTF) files and cufflinks gffread utility was employed to obtain final fasta files for both muscle and brain transcriptomes. Both transcriptomes were annotated against the NCBI non-redundant protein database (nr) using BLAST v.2.2.28+⁵⁷.

RNA-Seq differential expression. Uniquely mapped PE-reads were counted and assigned to genes using FeatureCounts⁵⁸ included in the SourceForge Subread package v.1.5.0. Count data were used to estimate differential gene expression between FG and SG samples using the Bioconductor package DESeq. 2 v.3.2⁵⁹ in R v.3.2.2⁶⁰. Genes with Benjamini-Hochberg false discovery rate (FDR) corrected p -values < 0.05 were considered differentially expressed (DE) genes. Due to the low amount of significant DE genes, those showing FDR corrected p -values < 0.3 (standard p -values < 0.005 and 0.0005 for muscle and brain, respectively) were also explored to identify a wider list of candidate genes for investigating enriched functions and pathways. Log₂ library-size normalized counts were used for generating heatmaps using “aheatmap” function in R/NMF⁶¹. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was assessed using KOBAS 2.0⁶² and Gene Ontology (GO) terms using Blast2GO⁶³, the DE genes were compared to the whole muscle or brain transcriptomes and those terms or pathways showing FDR corrected P -values < 0.05 were considered enriched.

Differential allelic frequencies. Pileup files describing the base-pair information at each genomic position were generated from the alignment files using the mpileup function of Samtools⁶⁴ discarding those aligned reads with a mapping quality < 20 and those bases with a Phred score < 20. SNP calling required a minimum read depth of 50 and a minimum of 5 reads for the least frequent allele. Differential allele frequency (DAfreq) between FG and SG samples were obtained using PoPoolation2⁶⁵ and the significance of DAfreq between the two groups was tested by Fisher's exact test. Those SNPs showing Bonferroni corrected p-values < 0.05 and further DAfreq > 0.25 were considered significant. All the SNPs were functionally annotated using SnpEff v.4.2⁶⁶ and the turbot genome protein annotation²¹. SNP physical location was compared with that of previously reported growth-associated QTLs in turbot^{17,18}, based on other farmed families sharing an Atlantic origin with the CETGA families here studied. All these families come from broodstocks derived from the turbot Atlantic area that shows very low genetic differentiation¹⁶.

Real-Time PCR. The same RNA samples used for RNA-Seq were analysed individually by qPCR. Primers for selected genes (LDH, PYGM, GHR2, CKs, CKm, and TGFβ1) were designed using the Primer 3 software⁶⁷ (Supplementary Table 3). To determine DE between conditions the $\Delta\Delta CT$ method was used with the slow-growth (SG) condition set as normalizer, and the ribosomal protein S4 (RPS4) and ubiquitin (UBQ) as housekeeping genes. These two genes were previously validated for gene expression determination by qPCR in turbot⁶⁸. Reactions were performed using a qPCR Master Mix Plus for SYBR Green I No ROX (Eurogentec) following the manufacturer instructions, and qPCR was carried out on a MX3005 P (Agilent Technologies). Analyses were performed using the MxPro software. A t-test was used to determine significant differences between conditions. Two technical replicates of each sample were included.

SNP validation. PCR and SNaPshot primers for SNPs included in selected candidate genes were designed using the NCBI primer designing tool and blast checked against the nr database⁶⁹. Primer length was modified to allow multiplexing for nine SNPs showing DAfreq (Supplementary Table 4). Total DNA was purified after protein precipitation (5 M NaCl) with freezing cold absolute ethanol (1 mL). Motility rate for each potential single-base extension product was calculated using SNaPshot Primer Focus kit (Thermo Fisher Scientific) and samples were divided into two multiplex reactions and one singleplex (Supplementary Table 4). A sample of DNA from each individual was amplified and the SNaPshot minisequencing reaction was carried out using ABI Prism SNaPshot Multiplex kit (Applied Biosystems) in an ABI Prism 3730 DNA sequencer according to manufacturer's instructions. The genotyping analysis was performed using GeneMapper ID software (Applied Biosystems). Exact G-tests were used to determine the significance of genotypic and genic differentiation between FG and SG population groups using Genepop 4.2⁷⁰.

References

- Chavanne, H. *et al.* A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. *Aquacult. Int.* **24**, 1287–1307 (2016).
- FAO: Food and Agriculture Organization of the United Nations. The state of world fisheries and aquaculture 2014, <http://www.fao.org/3/a-i3720e.pdf> (2014).
- Gjedrem, T., Robinson, N. & Rye, M. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* **350–353**, 117–129 (2012).
- Kawauchi, H. & Sower, S. A. The dawn and evolution of hormones in the adeno-hypothesis. *Gen. Comp. Endocrinol.* **148**, 3–14 (2006).
- Johnston, I. A., Bower, N. I. & Macqueen, D. J. Growth and the regulation of myotomal muscle mass in teleost fish. *J. Exp. Biol.* **214**, 1617–1628 (2011).
- Jaillon *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
- Daetwyler, H. D., Villanueva, B., Bijma, P. & Woolliams, J. A. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* **124**, 369–76 (2007).
- Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. L. Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genet. Sel. Evol.* **48**, 33 (2016).
- Jenko, J. *et al.* Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Gen. Sel. Evol.* **47**, 55 (2015).
- Hemmer-Hansen, J., Nielsen, E. E., Meldrup, D. & Mittelholzer, C. Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod. *Gadus morhua*. *Mol. Ecol. Resour.* **1**, 71–80 (2011).
- Salem, M. *et al.* RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS One* **7**, e36264 (2012).
- Xia, J. H. *et al.* Whole genome scanning and association mapping identified a significant association between growth and a SNP in the IFABP-a gene of the Asian seabass. *BMC Genomics* **14**, 295 (2013).
- Tian, C. *et al.* Single nucleotide polymorphisms in growth hormone gene and their association with growth traits in *Siniperca chuatsi* (Basilewsky). *Int. J. Mol. Sci.* **15**, 7029–7036 (2014).
- Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B. & Davidson, W. D. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One* **10**, e0119730 (2015).
- Tsai, H. Y. *et al.* Verification of SNPs associated with growth traits in two populations of farmed Atlantic salmon. *Int. J. Mol. Sci.* **17**, 5 (2016).
- Bouza, C. *et al.* Turbot. *AquaTrace Species Leaflet* <https://aquatrace.eu/leaflets/turbot> (2014).
- Sánchez-Molano, E. *et al.* Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* **12**, 473 (2011).
- Rodríguez-Ramilo, S. T. *et al.* Identification of quantitative trait loci associated with resistance to viral haemorrhagic septicaemia (VHS) in turbot (*Scophthalmus maximus*): A comparison between bacterium, parasite and virus diseases. *Mar. Biotechnol.* **16**, 265–276 (2014).
- Robledo, D. *et al.* Integrative transcriptome, genome and quantitative trait loci resources identify single nucleotide polymorphisms in candidate genes for growth traits in turbot. *Int. J. Mol. Sci.* **17**, 243 (2016).
- Won, E. T. & Borski, R. J. Endocrine regulation of compensatory growth in fish. *Front. Endocrinol.* **4**, 74 (2013).
- Figueroa, A. *et al.* Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA Res.* **23**, 181–192 (2016).
- Hendrickx, J. *et al.* Mutations in the phosphorylase kinase gene PHKA2 are responsible for X-linked liver glycogen storage disease. *Hum. Mol. Genet.* **4**, 77–83 (1995).

23. Hevner, R. F. *et al.* Tbr1 regulates differentiation of the preplate and layer 6. *Neuron* **29**, 353–366 (2001).
24. Knaus, P., Marqu e-Pouey, B., Scherer, H. & Betz, H. Synaptorin, a novel putative channel protein of synaptic vesicles. *Neuron* **5**, 453–462 (1990).
25. Anholt, R. R. H. Olfactomedin proteins: central players in development and disease. *Front. Cell Dev. Biol.* **2**, 6 (2014).
26. Kocmarek, A. L., Ferguson, M. M. & Danzmann, R. G. Differential gene expression in small and large rainbow trout derived from two seasonal spawning groups. *BMC Genomics* **15**, 57 (2014).
27. Choi, M., Kim, G., Kim, J. & Lim, H. K. Differentially-expressed genes associated with faster growth of the Pacific abalone. *Haliotis discus hannai. Int. J. Mol. Sci.* **16**, 27520–27534 (2015).
28. Ulloa, P. E. *et al.* RNA sequencing to study gene expression and SNP variations associated with growth in zebrafish fed a plant protein-based diet. *Mar. Biotechnol. (NY)* **17**, 353–363 (2015).
29. Danzmann, R. G. *et al.* Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics* **17**, 60 (2016).
30. Robledo, D. *et al.* Integrating genomic resources of flatfish (Pleuronectiformes) to boost aquaculture production. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **21**, 41–55 (2017).
31. Norman, J. D., Ferguson, M. M. & Danzmann, R. G. Transcriptomics of salinity tolerance capacity in Arctic charr (*Salvelinus alpinus*): a comparison of gene expression profiles between divergent QTL genotypes. *Physiol. Genomics* **46**, 123–137 (2014).
32. Lv, W. *et al.* QTL variations for growth-related traits in eight distinct families of common carp (*Cyprinus carpio*). *BMC Genet.* **17**, 65 (2016).
33. Bone, Q. Locomotor muscle in *Fish Physiology* (eds Hoar, W. & Randall, D.) 361–424 (Academic Press, 1978).
34. S nger, A. & Stoiber, W. Muscle fibre diversity and plasticity in *Muscle Development and Growth* (ed. Johnston, I.) 187–250 (Academic Press, 2001).
35. Kiessling, A. *et al.* Spawning induces a shift in farmed rainbow trout whitemuscle energy metabolism from a glucose to a lipid dependence. *Fish Physiol. Biochem.* **14**, 439–448 (1995).
36. Tymchuk, W., Sakhrani, D. & Devlin, R. Domestication causes large-scale effects on gene expression in rainbow trout: Analysis of muscle, liver and brain transcriptomes. *Gen. Comp. Endocrinol.* **164**, 175–183 (2009).
37. Xiao, S. *et al.* Gene map of large yellow croaker (*Larimichthys crocea*) provides insights into teleost genome evolution and conserved regions associated with growth. *Sci. Rep.* **5**, 18661 (2015).
38. Bernard, C., Cassar-Malek, I., Renand, G. & Hocquette, J. F. Changes in muscle gene expression related to metabolism according to growth potential in young bulls. *Meat Sci.* **82**, 205–212 (2009).
39. Wallimann, T. *et al.* The phospho-creatine circuit: molecular and cellular physiology of creatine kinases, sensitivity to free radicals and enhancement by creatine supplementation in *Molecular Systems Bioenergetics: Energy for Life* (ed. Saks, V.) 195–264 (Wiley, 2007).
40. Kraft, T., Hornemann, T., Stolz, M., Nier, V. & Wallimann, T. Coupling of creatine kinase to glycolytic enzymes at the sarcomeric I-band of skeletal muscle: A biochemical study *in situ*. *J. Muscle Res. Cell. Motil.* **21**, 691–703 (2000).
41. Kondo, H. *et al.* Effects of feed restriction on the expression profiles of the glucose and fatty acid metabolism-related genes in rainbow trout *Oncorhynchus mykiss* muscle. *Fish. Sci.* **78**, 1205–1211 (2012).
42. Liu, F. *et al.* A genome scan revealed significant associations of growth traits with a major QTL and GHR2 in tilapia. *Sci. Rep.* **4**, 7256 (2014).
43. Fuentes, E. N., Vald s, J. A., Molina, A. & Bj rnsson, B. T. Regulation of skeletal muscle growth in fish by the growth hormone–insulin-like growth factor system. *Gen. Comp. Endocrinol.* **192**, 136–148 (2013).
44. Drew, R. E. *et al.* Brain transcriptome variation among behaviourally distinct strains of zebrafish (*Danio rerio*). *BMC Genomics* **13**, 323 (2012).
45. Balasubramanian, M. N. *et al.* Molecular pathways associated with the nutritional programming of plant-based diet acceptance in rainbow trout following an early feeding exposure. *BMC Genomics* **17**, 449 (2016).
46. He, S. *et al.* Insights into food preference in hybrid F1 of *Siniperca chuatsi* (♀) × *Siniperca scherzeri* (♂) mandarin fish through transcriptome analysis. *BMC Genomics* **14**, 601 (2013).
47. van der Velden, J. L. *et al.* Glycogen synthase kinase 3 suppresses myogenic differentiation through negative regulation of NFATc3. *J. Biol. Chem.* **283**, 358–366.
48. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379–382 (2011).
49. Franzolin, E., Salata, C., Bianchi, V. & Rampazzo, C. The deoxynucleoside triphosphate triphosphohydrolase activity of SAMHD1 protein contributes to the mitochondrial DNA depletion associated with genetic deficiency of deoxyguanosine kinase. *J. Biol. Chem.* **290**, 25986–25996 (2015).
50. Imsland, A. K., Folkvord, A., Grung, G. L., Steansson, S. O. & Taranger, G. L. Sexual dimorphism in growth and maturation of turbot, *Scophthalmus maximus* (Rafinesque, 1810). *Aquacult. Res.* **28**, 101–114 (1997).
51. Mart nez, P. *et al.* Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). *Genetics* **183**, 1443–1452 (2009).
52. Robledo, D. *et al.* Gene expression analysis at the onset of sex differentiation in turbot (*Scophthalmus maximus*). *BMC Genomics* **16**, 973 (2015).
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114–2120 (2014).
54. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
57. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
58. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
59. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol.* **15**, 550 (2014).
60. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
61. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
62. Wu, J., Mao, X., Cai, T., Luo, J. & Wei, L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* **34**, W720–W724 (2006).
63. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
64. Li, H. *et al.* The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Kofler, R., Pandey, R. V. & Schl tterer, C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436 (2011).

66. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strainw1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
67. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
68. Robledo, D. *et al.* Analysis of qPCR reference gene stability determination methods and a practical approach for efficiency calculation on turbot (*Scophthalmus maximus*) gonad dataset. *BMC Genomics* **15**, 548 (2014).
69. Vera, M. *et al.* Validation of single nucleotide polymorphism (SNP) markers from an immune Expressed Sequence Tag (EST) turbot database. *Aquaculture* **313**, 31–41 (2011).
70. Rousset, F. Genepop'007: a complete reimplement of the Genepop software for Windows and Linux. *Mol. Ecol. Resources* **8**, 103–106 (2008).

Acknowledgements

This work was funded by Spanish Ministry of Economy and Competitiveness and European Regional Development Funds (AGL2012-35904), Ministry of Science and Innovation (Consolider Ingenio, Aquagenomics, CSD2007-00002), and Local Government. Xunta de Galicia (GRC2014/010). DR was supported by a FPU fellowship funded by Spanish Ministry of Education, Culture and Sport (AP2012-0254) and a postdoctoral contract funded by the Biotechnology and Biological Science Research Council (BBSRC) grant BB/N024044/1. Thanks to Manuel Vera for his support on SNP validation and to Lucía Ínsua for technical assistance. We acknowledge the support of High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for generating the sequencing data and Centro de Supercomputación de Galicia (CESGA) for computational analyses. We thank to the Cluster de la Acuicultura de Galicia (CETGA) for providing the samples, which are part of an experiment within the EU project FISHBOOST (grant agreement no. 613611). We thank Antti Kausse and Anna Sonesson (FISHBOOST work package leader and coordinator, respectively) for their permission to use these samples.

Author Contributions

C.B and P.M. conceived and designed the study. S.C. was responsible for fish rearing and sampling. D.R. and J.R. performed RNA-Seq and qPCR analyses, respectively, inferring the biological significance of DE genes. J.R. and C.B. carried out the SNaPshot experiments and analyses. C.B., D.R. and J.R. analysed the data and wrote the manuscript. All authors have read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-12459-4>.

Competing Interests: The authors declare that they have no competing interests.

Accession codes: Raw sequence reads have been deposited in NCBI's Short Read Archive (SRA) under BioProject accession PRJNA322741.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017