A genomic approach to reveal novel genes associated with myotube formation in the model teleost, *Takifugu rubripes*

Jorge M. O. Fernandes, Matthew G. Mackenzie, Greg Elgar, Yuzuru Suzuki, Shugo Watabe, James R. Kinghorn, and Ian A. Johnston

Gatty Marine Laboratory, School of Biology, University of St. Andrews, St. Andrews, Fife, United Kingdom; Medical Research Council Rosalind Franklin Centre for Genomics Research, Genome Campus, Hinxton, Cambridge, United Kingdom; Fisheries Laboratory, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Maisaka, Shizuoka, Japan; Laboratory of Aquatic Molecular Biology and Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan; and Sir Harold Mitchell Building, School of Biology, University of St. Andrews, St. Andrews, Fife, United Kingdom

Submitted 18 April 2005; accepted in final form 27 May 2005

Fernandes, Jorge M. O., Matthew G. Mackenzie, Greg Elgar, Yuzuru Suzuki, Shugo Watabe, James R. Kinghorn, and Ian A. Johnston. A genomic approach to reveal novel genes associated with myotube formation in the model teleost, *Takifugu rubripes*. Physiol Genomics 22: 327–338, 2005. First published May 31, 2005; 10.1152/physiolgenomics.00087.2005.—Little is known about the transcriptional networks that regulate myotube production in vertebrates. In the present study, we have used a genomic approach to discover novel genes associated with myotube formation in fast muscle of the tiger puffer fish, *Takifugu rubripes*. The number of fast muscle fibers per myotome increased until 1.2 kg body mass, and subsequent growth was by fiber hypertrophy alone. Forward and reverse subtracted cDNA libraries were prepared from a 180-g (myotube−) and a 3.4-kg (myotube+) fish, and 1,452 expressed sequence tags (ESTs) were obtained. After these ESTs were grouped into nonredundant clusters and housekeeping and structural genes were eliminated, 57 genes were selected and quantitative PCR was used to investigate their expression levels in different tissues from independent groups of myotube(−) and myotube(+) fish acclimated to the same environmental conditions and diet. Eleven novel genes were found to be consistently differentially expressed, but only four showed appropriate tissue-specific expression. These four genes were upregulated 5–25 times in fast muscle of myotube(−) relative to myotube(+) growth stages, while their expression remained unchanged in the other tissues studied. The novel genes identified, which are also present in other vertebrate genomes, may play a role in inhibiting myotube formation in vertebrate muscle.

different populations of myoblasts (somatic, embryonic, fetal, and adult) appear sequentially during mammalian development (27). The embryonic myoblasts give rise to primary muscle fibers that form a scaffold for the formation of more numerous secondary muscle fibers from the fetal myoblasts (25, 52, 54, 72). The definitive number of muscle fibers in mammals is already established at or shortly after birth (62, 64). Postembryonic growth involves an increase in the length, diameter, and nuclear content of muscle fibers (62, 64). The nuclei required for growth and tissue maintenance are derived from a small population of myogenic progenitors (adult myoblasts) residing between the sarcolemma and basal lamina, called satellite cells (29, 65). Activated satellite cells generate undifferentiated progeny to restore the pool of quiescent satellite cells and proliferating cells that are committed to differentiation (53, 67). During normal growth, myotube formation is inhibited. However, various stimuli including exercise (13), stretch (61), and injury (4) can trigger myotube formation even in adults that have stopped growing. There is strong functional (32, 73) and molecular evidence (14, 66) that satellite cells constitute a heterogeneous population. Although progress has been made in the transcriptional mRNA profiling of satellite cells (10, 11, 66), the genetic network(s) regulating their function in growth and repair remains poorly understood.

Teleost fish represent a good model for studying vertebrate myogenesis, since slow and fast muscle fibers are arranged in anatomically discrete layers and myotube production is not restricted to early developmental stages, reflecting the large difference between embryonic and final adult body size (36). Furthermore, fast and slow muscles have a distinct embryological origin (16) and show different patterns of postembryonic growth (3, 38). For fast muscle, the number of fibers increases until a genetically predetermined body length, and subsequent growth is entirely by fiber hypertrophy (38, 41, 63, 71). For example, in a benthic morph of the arctic char (Salvelinus alpinus), the number of fast muscle fibers per myotomal cross section increased from 20,000 at 6.5-cm fork length (FL) to ~120,000 at 33-cm FL and then remained constant to the maximum size of 55-cm FL (38). In contrast, the number of slow muscle fibers showed a continuous increase with fish length (38).

Embryonic fast muscle fibers arise from the lateral pre-somatic mesoderm (16), and additional fibers are produced from discrete germinal zones in the late embryo, larval, and early juvenile stages by stratified hyperplasia (3, 46). The final and most important stage of myogenesis involves the formation of myotubes throughout the myotomes, called mosaic hyperplasia because of the resulting pattern of fiber diameters (39, 41, 42, 63). The nuclei for fiber recruitment and hypertrophy are derived from a population of muscle progenitor cells (MPCs) that are equivalent to the muscle satellite cells found in mammals (40). MPCs express the MyoD gene family of basic helix-loop-helix transcription factors (myf5, myoD, myogenin, MRF4) that regulate myogenic lineage formation and differentiation (19, 48). In the fast muscle of arctic char, the density
of mononuclear cells expressing MPC markers, such as paired box protein 7 (Pax 7) and forkhead protein K1-α (Fox K1-α), decreased slightly with increasing length but was not related to the cessation of fiber recruitment (38). This is probably because the majority of MPCs provide myonuclei for fiber growth, and relatively few are involved in initiating new myotube formation (38).

The tiger puffer fish, Takifugu rubripes, was the second vertebrate to have its genome sequenced to draft level (2). Although only one-eighth the size of the human genome (~365 Mb), the Takifugu genome has a similar number of predicted gene loci, and 75% of its putative peptides have homologs in the human proteome (2). In addition, the tiger puffer fish genome is compact, generally with short introns and a low proportion of repetitive elements (2), making this organism an ideal tool for comparative genomics (24, 31, 69).

The availability of the Takifugu genome sequence greatly facilitates the identification of novel genes whose expression patterns are correlated with muscle fiber recruitment in fast muscle. Combining normalization and subtraction, suppression subtractive hybridization (SSH) is an excellent method of identifying differentially expressed genes (17). The use of subtracted cDNA libraries to generate single-pass partial cDNA sequences (expressed sequence tags; ESTs) greatly reduces redundancy and, therefore, enables the detection of rare transcripts that would probably not be identified from nonnormalized libraries (22, 49). ESTs are important for the annotation of eukaryotic genomes (47), as they provide valuable information regarding the discovery of novel genes, characterization of transcript structure, population of physical maps, and alternative splicing of transcripts (5, 21). ESTs from teleost fish account for circa 4.7% of the nearly 25 million sequences present in the current release of the EST database, dbEST (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html; January 2005) (6). However, almost 90% of these ESTs have been obtained from zebrafish, rainbow trout (Oncorhynchus mykiss), medaka (Oryzias latipes), and Atlantic salmon (Salmo salar). The two fish species whose genomes have been sequenced are poorly represented in dbEST: a mere 99 ESTs have been identified in Tetraodon nigroviridis, and there are only 25,850 ESTs from T. rubripes (9). The aim of the present study was to generate additional Takifugu EST resources and discover novel genes whose expression patterns are correlated with muscle fiber recruitment in fast muscle. We exploited the different fiber recruitment patterns between slow and fast muscles with growth to distinguish between genes associated with myotube formation and genes whose expression changed simply as a function of body size. The discovery of genetic networks regulating myotube formation is also of commercial interest, since variation in fiber number is known to be an important flesh quality parameter in farmed fish (37, 43).

MATERIALS AND METHODS

Histology and morphometric measurements. Live tiger puffer fish (T. rubripes) were obtained from the fish market at Maisaka City, Japan, during January and February of 2003. Experimental protocols were approved by the Animal Welfare Committee, University of St. Andrews, Scotland. The fish were killed by a blow to the head and transection of the spinal cord. A 0.5-cm steak was cut at 0.7 standard length (SL) (Fig. 1A), and a series of blocks were made to sample the myotomal, dorsal, and anal fin muscles from one side of the trunk. The blocks were frozen in isopentane cooled to its freezing point in liquid nitrogen (~159°C), and 8-μm frozen sections were cut at ~20°C in a cryostat. Sections were air-dried and stained for myosin ATPase and succinic dehydrogenase activities (44) and with the S-58 antibody, which recognizes a subset of slow muscle fibers (e.g., see Ref. 39). The sections were preincubated at pH 4.3 or pH 9.4 before being stained for myosin ATPase to selectively inactivate different fiber types (44). Sections stained with myosin ATPase were used to quantify the number and diameter of the fast muscle fibers. The cross-sectional areas of ~50 fast muscle fibers from each of 20 fields per fish were digitized to obtain a representative sample of the myotomal cross section. Smooth distributions were fitted to the measurements of diameter using a kernel function as described previously (45); the value of the smoothing constant (h) was 0.145. The total fiber number was estimated from the measured fiber cross-sectional areas and the total area of fast muscle per myotomal cross section (45).

Construction of subtracted cDNA libraries. One adult male tiger puffer fish (3.4 kg, 50-cm SL) in which myotube production had ceased [myotube(−) stage] and a juvenile specimen (180 g, 18-cm SL) with active myotube production [myotube(+) stage] were sampled, and their fiber recruitment status was confirmed by histological examination. The large fish was wild-caught and obtained from the Maisaka City fish market, whereas the smaller fish was bred in captivity at the Fisheries Laboratory (Univ. of Tokyo). Both animals were humanely killed by an overdose of phenytoxethanol (Sigma) followed by transection of the spinal cord. Pure samples of fast myotomal muscle were preserved in RNAlater (Ambion) for subsequent RNA extraction. Total RNA was extracted with TRI reagent (Sigma), according to the manufacturer’s instructions. The SMART PCR cDNA synthesis kit (BD Biosciences) was used to synthesize cDNA. After 17 cycles of amplification, the cDNA from both samples was digested with Rsa I (BD Biosciences). The two sets of blunt-ended cDNA fragments were compared by SSH using the PCR-select cDNA subtraction kit (BD Biosciences). Two independent subtractions were performed: a forward subtraction (where the cDNA from the 3.4-kg Takifugu was used as tester, and the cDNA from the 180-g fish was the driver) and a reverse subtraction (using cDNA from the 180-g Takifugu as the tester and cDNA from the 3.4-kg fish as a driver). The subtracted PCR mixtures were ligated onto a T/A PCR4-TOPO vector (Invitrogen) and used to transform chemically competent TOP10 Escherichia coli cells (Invitrogen). Two subtracted cDNA libraries of 1,152 random clones each were generated. Both libraries had insert sizes ranging from 250 to 3,000 bp. Individual clones are available from the Fish Muscle Research Group (FMRG; Gatty Marine Laboratory, Univ. of St. Andrews; http://www.st-andrews.ac.uk/~fmr/).

EST sequencing. The subtracted cDNA libraries were sequenced at the MRC Rosalind Franklin Institute for Genomics Research. Colony PCR for insert amplification was performed under limiting conditions of primer and dNTPs using short T3 (5'-ATTACCCCTA- AAG-3') and short T7 (5'-AATACGACTCACTATAG-3') primers. The thermocycling parameters were as follows: initial denaturation at 96°C for 2 min, 35 cycles of amplification (96°C for 20 s, 49°C for 20 s, and 72°C for 45 s), and final extension at 72°C for 5 min. PCR products were directly sequenced using a T3 primer (5'- AATTAAC- CCTACTAAAGG-3') and Big Dye terminator sequencing mix (Applied Biosystems). The sequencing reaction cycle (95°C for 20 s and 60°C for 140 s) was repeated 25 times. DNA sequences were read using an Applied Biosystems 3700 capillary sequencer.

All EST data generated by this project have been submitted to dbEST and may be accessed through GenBank accession numbers CK829215–CK830666. The primer sets used for real-time PCR amplification of the selected genes from the forward and reverse subtracted libraries can be found in Supplemental Tables S1 and S2.

Physiol Genomics • VOL 22 • www.physiolgenomics.org
respectively (available at the Physiological Genomics web site). A complete list of the partial genome databases corresponding to the subtracted libraries generated in this study is available online in the FMRG EST database (http://www.st-andrews.ac.uk/~fmrg/).

These tables show the sequences present in each cluster and the similarity search results against the Takifugu predicted gene set and the nonredundant protein database. The selected genes from the myotube (−) and myotube (+) library that were analyzed by quantitative PCR (qPCR) are listed in Supplemental Tables S3 and S4, respectively. Relative expression levels of the differentially displayed genes were consistently upregulated in multiple tissues of Takifugu are shown in Supplemental Tables S5 and S6.

Bioinformatics. Raw sequence trace data were processed by the EST analysis pipeline developed by the Natural Environment Research Council-Environmental Genomics Thematic Programme Data Centre (NERC-EGTDC; Univ. of Edinburgh). The electropherograms were first analyzed by trace2dbEST (accessible through http://www.nematodes.org/PartiGene), which integrates Phred (20), Cross_match (P. Green, unpublished; available from http://www.phrap.org/phredphrapconsed.html), and basic local alignment search tool (BLAST) (1). Sequences were trimmed by removing vector and E. coli sequence, low-quality base calls, and any sequence after the poly(A) tail. The resulting sequences were then subjected to BLAST similarity searches, using BLASTX against the National Center for Biotechnology Information (NCBI) nonredundant protein sequence database (downloadable from ftp://ftp.ncbi.nlm.nih.gov/blast/db/ nr.tar.gz). Good-quality sequences were submitted to dbEST (6), jointly with their BLAST-based preliminary annotation. PartiGene (26) was then used to cluster the sequences on the basis of their similarity, using CLOBB (55), and the clusters were assembled into contigs with Phrap (P. Green, unpublished; available from http://www.phrap.org/phredphrapconsed.html). The nonredundant partial transcriptome databases thus created were annotated on the basis of the best results from BLAST similarity searches. Clusters were compared with two data sets: the NCBI nonredundant protein sequence database and the Ensembl fugu assembly v.2 cDNA database (ftp://ftpensembl.org/pub/current_fugu/data/fasta/cdna/). Partigene was also used to compare the T. rubripes partial transcriptome databases against the 27.1b.1 release of the T. nigroviridis genome and cDNA data sets (ftp://ftpensembl.org/pub/current_tetraodon/data/fasta/dna/; ftp://ftpensembl.org/pub/current_tetraodon/data/fasta/cdna/). Then the EST translation tool Prot4EST (70) was employed to identify coding regions and predict the corresponding polypeptide translations. The putative polypeptides were classified according to the three ontologies defined by the Gene Ontology (GO) consortium (28), using Annot8er_blast2GO (available from http://www.nematodes.org/PartiGene). Annotation was based on BLAST searches against a GO slim subset of the GO-annotated SwissProt/TrEMBL database (downloadable from ftp://ftp.ebi.ac.uk/pub/ databases/GO/goa/). The candidate genes that were consistently

Fig. 1. Muscle fiber types in the tiger puffer fish (Takifugu rubripes). A: samples were taken at 0.7 standard length (SL; tip of snout to end of trunk) from a fish of 154 g body wt. B: cross section of the trunk at 0.7 SL illustrating the position of muscle blocks (a–e). Scale bar = 2 cm. C: frozen sections stained to differentiate different muscle fiber types. Scale bar = 100 μm. a: Section stained for myosin ATPase activity at pH 9.4 after 1-min preincubation at pH 4.3. The red fibers in the dorsal fin adductor muscle (da) show a complex pattern of myosin ATPase staining with the interior layers of fibers (left) inactivated by alkaline preincubation, whereas the outer layers (right) contained darkly (arrowheads) and lightly staining fibers. b: Section stained with the S-58 slow muscle myosin antibody. A subset of the red fibers that were unstained for myosin ATPase stained intensely with S-58 (arrowheads). c: Fast myotomal muscle fibers (f) stained darkly for myosin ATPase activity at pH 9.4. d: Section of fast myotomal muscle stained for myosin ATPase activity at pH 9.4 after 2-min preincubation at pH 4.3. The myosin ATPase activity was inactivated in the larger diameter fibers (asterisks) but not the smaller diameter fibers. e: Section of fast myotomal muscle showing one of the internalized strips of red muscle fibers most of which stained with the S-58 slow muscle myosin antibody (arrowheads). Note that the fast muscle fibers (f) were unstained. aa, anal fin adductor muscle.

1 The Supplemental Material for this article (Supplemental Tables S1–S6) is available online at http://physiolgenomics.physiology.org/cgi/content/full/00087.2005/DC1.
differentially expressed were subjected to BLASTN searches against the build 2d of the *T. rubripes* genome database at Ensembl, using the default search configuration.

Preselection of differentially expressed genes. The partial transcriptome databases corresponding to the subtracted libraries were screened for genes with disrupted open reading frames, RNA genes, untranslated regions, metabolic enzymes (aldolase A, creatine kinase, etc.), and abundant structural proteins such as myosin heavy chain and tropomyosin. These noncoding sequences and housekeeping genes were discarded, and two sets of putative genes were selected from each library.

Post hoc screening of selected genes by real-time PCR. To identify genes that were consistently differentially expressed, three myotube(+) (150 ± 10 g, 15.8 ± 0.8-cm SL) and three myotube(−) fish (1.31 ± 0.05 kg, 36.5 ± 1.3-cm SL) were acclimated in the laboratory to the same conditions of temperature (~12°C), photoperiod (12:12-h light-dark cycle), and feeding regime for 3 wk. The animals were humanely killed, and RNA was isolated from fast myotomal muscle. One hundred milligrams of each sample were added to FastRNA Pro Green beads (Qiogene) containing 1 ml of TRI reagent (Sigma). The tissues were homogenized using the FastPrep Instrument (Qiogene) for 40 s at a speed setting of 6.0. Total RNA was isolated from each sample according to the manufacturer’s instructions. Potential contaminating DNA was removed from the RNA preparation using TURBO DNA-free (Ambion). Total RNA was then quantified with the fluorescent nucleic acid stain RiboGreen (Molecular Probes), using the RETROscript kit (Ambion), following the recommended protocol. First-strand cDNA was synthesized from 1 µg of total RNA using the RETROscript kit (Ambion), according to the suggested method. A negative control lacking RT was included. qPCR was performed using an ABI Prism 7000 (Applied Biosystems) real-time thermocycler with SYBR Green chemistry (QuantiTect SYBR Green PCR, Qiagen). The reaction mixtures contained 1X QuantiTect SYBR Green PCR master mix, 1 µl of cDNA, 0.4 µM each primer, and RNase-free water (Qiagen) to a final volume of 25 µl. The primer pairs used to amplify the selected candidate genes from the forward and reverse subtracted libraries are listed in Supplemental Tables S1 and S2, respectively. Amplification was performed in 96-well plates (Applied Biosystems) using the following thermocycling parameters: initial activation at 95°C for 15 min followed by 40 cycles of 15 s at 94°C, 30 s at 56°C, and 30 s at 72°C (extension and acquisition of fluorescence data); 18S rRNA (Ambion) was used as standard, and the fluorescent dye ROX was used as internal reference for normalization of SYBR Green fluorescence. After the final amplification cycle, a dissociation protocol ranging from 60 to 90°C was performed to investigate the specificity of the primers and the presence of primer dimers. PCR efficiencies of each amplicon were estimated by linear regression analysis of the logarithm of SYBR Green fluorescence vs. cycle number, using the computer program LinRegPCR (59). Differences in expression of the candidate genes between the myotube(+) and myotube(−) fish were quantified with the relative expression software tool REST (58). The ratio of expression of a target gene normalized against the standard is computed according to the following mathematical model (57)

\[
\text{Ratio} = \frac{E_{\text{target}}}{E_{\text{standard}}} \frac{\Delta C_{T,\text{target}}(\text{group}1-\text{group}2)}{\Delta C_{T,\text{standard}}(\text{group}1-\text{group}2)}
\]

where E represents the real-time PCR efficiencies and \(\Delta C_T\) is the difference between threshold cycle (CT) values between the two groups. Group differences were tested for significance with REST (58) using a pairwise fixed reallocation randomization test with 2,000 randomizations.

**Gene expression in different tissues.** Fast muscle, slow muscle, heart, liver, brain, and skin were collected from myotube(+) and myotube(−) fish, and total RNA was extracted from 100 mg of each tissue using TRI reagent. After quantification with RiboGreen, first-strand cDNA was synthesized using the Retroscript kit. Primers for the selected candidate genes found to be consistently differentially expressed were used to quantify their expression in these tissues by qPCR as described above. The data obtained were analyzed with the REST software (58).

**RESULTS**

**Histochemical localization of muscle fiber types.** The arrangement of muscle fiber types at 0.7 SL is illustrated in Fig. 1C. The dorsal and ventral fin muscles comprised a red portion adjacent to the neural spines and interhemal bones, respectively. The red muscle layer stained intensely for the mitochondrial marker succinic dehydrogenase (SDHase; not illustrated) and weakly for myosin ATPase after a brief period of preincubation at pH 4.3 and was classified as a “slow muscle” (Fig. 1Ca). A subset of these fibers adjacent to the neural spines stained with the S-58 antibody (Fig. 1Cb) raised against chicken slow muscle myosin (12). The myotomal muscle was primarily composed of fast muscle fibers that stained weakly for SDHase (not illustrated) and intensely for myosin ATPase activity after 1 min of preincubation (Fig. 1Cc). Red myotomal muscle comprised a thin strip of fibers on the lateral surface of the myotome that stained intensely with S-58 (not illustrated). Preincubation at pH 4.3 for longer than 1 min resulted in a progressive inactivation of the myosin ATPase activity of fast fibers. The smallest diameter fibers were the most resistant to inactivation, suggesting they had a different myosin composition than the larger diameter, more mature, fast muscle fibers (Fig. 1Cd). Narrow layers of S-58-positive muscle fibers were also observed in the dorsal and ventral portions of the myotome interspersed with the fast muscle fibers (Fig. 1Ce).

**Postembryonic muscle fiber recruitment.** The smooth distributions of fast muscle fiber diameter are plotted in Fig. 2. The four largest fish examined, ranging from 1.32 to 4.0 kg, had no fibers <5 µm in diameter (solid lines in Fig. 2) and were considered to have stopped recruiting muscle fibers. The relationship between the number of fast muscle fibers per myotomal cross section and fish body mass is shown in Fig. 3. Fast fiber number increased from 94,000 in the smallest fish examined (14-cm SL, 81 g) to an average value of 412,000 in the four largest fish. The threshold body size for the cessation of recruitment was ~1.2 kg and 35-cm SL (arrow in Fig. 3). The maximum fiber diameter increased with body mass, reaching 300 µm at 4 kg.

**Characterization of the subtracted cDNA libraries.** We constructed two subtracted cDNA libraries from fast skeletal muscle of *T. rubripes*, taking care not to sample the internalized strips of slow fibers, as shown in Fig. 1Ce. The forward subtracted library [myotube(−)] comprised genes whose expression is higher in fish that have stopped recruiting muscle fibers, and the reverse subtracted library [myotube(+)] contained genes that are more highly expressed during myotube production. A total of 1,152 clones from each library were single-pass sequenced from the 5′-end, generating 804 and 648 good quality ESTs from the myotube(−) and myotube(+) libraries, respectively. The average read length of the sequences obtained after poly(A) and vector sequence trimming was 510 bp. These EST sequences were then grouped in nonredundant clusters to account for multiple representations of the same genes by several ESTs. The myotube(−) library was composed of 293 putative genes of which 175 were singletons (single EST clusters), while the myotube(+) library
comprised 392 gene clusters, including 262 singletons. The majority of EST clusters (96.5%) could be mapped to the current Ensembl Takifugu genome assembly. The overall redundancy (defined as the ratio between the total number of sequences and the number of clusters) was 2.74 and 1.65 for the myotube(−)/H11002 and myotube(+) library, respectively. This low figure reflects a high efficiency of the subtraction procedure, which is also corroborated by the fact that only 33 clusters were common to both libraries. The common clusters contained several ESTs and corresponded to structural and motor proteins, as well as metabolic enzymes associated with glycolytic metabolism. Fast skeletal muscle myosin heavy chain, tropomyosin 1, tropomyosin 1-2, aldolase A, and glyceraldehyde-3-phosphate dehydrogenase were particularly abundant. The muscle isoform 1 of creatine kinase was also overrepresented by 43 sequences in a single cluster. A comprehensive record of the partial transcriptome databases generated in the present study, including the BLAST annotation results against the nr nonredundant protein database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/nr.tar.gz) and the Ensembl Takifugu predicted gene set (ftp://ftp.ensembl.org/pub/current_fugu/data/fasta/cdna/), can be found on the FMRG EST database website (http://www.st-andrews.ac.uk/~fmrg/).

A small proportion of the EST clusters were found to be mitochondrial or rRNA sequences (Fig. 4). Only 186 clusters from the myotube(−)/H11002 library and 231 from the myotube(+) library had significant hits (expected value below 10^−8) against the NCBI nonredundant protein database (Fig. 4). The remaining 34 and 39% from the myotube(−) and myotube(+) library, respectively, did not match any protein entries and are therefore novel sequences or untranslated regions (UTRs) of known genes (Fig. 4). In addition, these clusters did not correspond to any gene from the Ensembl Takifugu predicted gene set. Approximately 24% of the clusters from both libraries could be translated with ESTscan (34) or DECODER (23) (Fig. 4). Hence, these are likely to represent 168 new putative polypeptides. The remaining sequences could not be translated with the aforementioned software and probably correspond to UTRs or pseudogenes.

The putative translation products of the clusters from both subtracted libraries were grouped into different categories according to the GO slim terms (28), as shown in Figs. 5 and 6. A large proportion of clusters did not have an associated GO slim term and were referred to as “unclassified.” The majority of these clusters represent hypothetical proteins with unknown function and cellular location. In both libraries, >70% of the putative proteins with a GO slim classification were either enzymes or proteins with binding activity (Figs. 5 and 6).

A comparison between the T. rubripes partial transcriptomes and the T. nigroviridis genome assembly revealed that ~74% of clusters from the myotube(−) library and 69% from the myotube(+) library had matches to T. nigroviridis genes. Of the putative genes that did not match the T. nigroviridis predicted gene set, only four clusters from each library were not present in the current T. nigroviridis genome assembly.
Selection of candidate genes and expression in different tissues. Both subtracted libraries were screened for pseudogenes, UTRs, short coding sequences, housekeeping genes, abundant structural proteins, and common clusters. After discarding these ESTs, we selected a set of 24 putative genes from the myotube(−) library (Supplemental Table S3) and 33 gene clusters from the myotube(+) library (Supplemental Table S4). Using qPCR, we have determined the relative expression of the selected genes in the original myotube(−) and myotube(+) cDNAs used for SSH. A single dissociation peak was obtained for each primer set, thus confirming the presence of only one PCR product. The transcript levels of 18S rRNA did not change significantly in fast muscle from myotube(−) and myotube(+) fish (data not shown). Hence, this validated standard was used to normalize the qPCR data. Three quarters of the genes from both libraries were found to be significantly upregulated, and most of the remaining ones showed an unaltered expression level (data not shown), revealing that the subtraction was efficient; only two genes from the myotube(−) library were significantly downregulated.

We then performed a post hoc gene expression analysis of the selected, potentially differentially expressed genes by qPCR using independent fast muscle samples from three myotube(−) and three myotube(+) growth stage fish acclimated for 3 wk to the same diet and environmental conditions. It was found that nine and two genes from the myotube(−) and myotube(+) libraries, respectively, were consistently and significantly differentially expressed between the two sample groups (Table 1). We have also analyzed the relative expression levels of the nine myotube(−) and two myotube(+) selected genes in slow muscle, heart, liver, skin, and brain of myotube(−) and myotube(+) fish, respectively. Genes whose transcript levels were higher in fast muscle of myotube(−) fish were classified as 1) strong candidates, 2) intermediate candidates, and 3) unlikely prospects, according to the following criteria: 1) high myotube(−)/myotube(+) expression ratios in fast muscle but not in other tissues, and expression limited to fast muscle in myotube(−) fish; 2) high myotube(−)/myotube(+) expression ratios in fast muscle and weak expression in a limited range of other tissues in myotube(−) fish; and 3) high myotube(−)/myotube(+) expression ratios in multiple tissues and wide expression in myotube(−) fish. Expression of the four strong candidate genes, FRC258, FRC363, FRC386, and FRC405, was preferentially higher in fast muscle of myotube(−) fish by an average factor of 25, 5, 16, and 21, respectively (Fig. 7, A, C, E, and G). In myotube(−) fish, the expression of the strong candidates was restricted to fast muscle, with expression levels at least fourfold higher in fast muscle than in any of the other tissues studied (Fig. 7, B, D, F, and H). Their expression patterns indicated that these genes might be involved in the inhibition of myotube formation, since their higher expression levels are fast muscle specific rather than just being related to the size of the fish. FRC258, FRC363, FRC386, and FRC405 were found to be present in scaffolds 278 (76,165–77,989 bp), 237 (230,648–231,231 bp), 62 (193,206–196,264 bp), and 572 (129,878–131,877 bp), respectively, of the current Takifugu genome assembly. All four strong candidates had unambiguous best reciprocal BLAST hits that corresponded to putative orthologs in Tetraodon and human.

FRC167, FRC214, and FRC272 were considered to be intermediate candidate genes. Expression of these genes was higher in fast muscle of myotube(−) fish by an average 5-fold, 15-fold, and 13-fold, respectively, but in most cases their transcript levels also varied with size in other tissues (Supplemental Tables S5 and S6).

The remaining two genes from the myotube(−) library (FRC40 and FRC263) and the genes FRC177 and FRC300 from the myotube(+) library that showed consistent differential expression were unlikely prospects (Supplemental Tables S5 and S6).

**DISCUSSION**

We have identified four novel genes (FRC258, FRC363, FRC386, and FRC405) that were expressed at significantly and consistently higher levels in fast muscle of *T. rubripes* that had stopped recruiting muscle fibers. The tissue and temporal expression patterns of these genes indicate that they are strong candidates for playing a role in the inhibition of myotube formation. Transcript levels of the gene FRC258 were 25-fold higher in myotube(−) fish. Conceptual translation of its nine exons predicted on Ensembl revealed that it is an ortholog of the human cardiomyopathy-associated 5 protein (22 and 65% identity with human and *T. nigroviridis*, respectively). This
403-residue protein, not previously identified in teleosts, has several conserved domains including B302, fibronectin, and SPRY receptor domains. The presence of such domains suggests that FRC258 might be involved in regulation of transcription and/or in cell migration, differentiation, and adhesion (18).

Expression of the FRC363 gene was fivefold higher in myotube(−) fish. The Ensembl-predicted transcript structure for this gene comprises five exons that translate into a 483-residue protein with several conserved domains. The presence of POZ/BTB and BTB/Kelch domains suggests that FRC363 might be involved in protein-protein interactions (8). The Takifugu FRC363 protein shared 34 and 76% identity with the corresponding human and T. nigroviridis proteins, respectively. Transcript abundance of the gene FRC386 was found to be 16-fold higher in myotube(−) fish. This cluster matches a novel Ensembl prediction composed of 3 exons that encode a polypeptide of 85 residues with 93% similarity to its ortholog in T. nigroviridis. Despite not having GO classification, the FRC167 protein contains a conserved domain typical of the LYR family, which includes proteins from the NADH-ubiquinone oxidoreductase complex. Transcript levels of the gene FRC214 were 15 times higher in myotube(−) fish compared with myotube(+) fish. The predicted transcript structure for this gene comprises 4 exons that translate into a 430-residue protein with POZ/BTB and Kelch domains and that share 35 and 65% identity with the corresponding human and T. nigroviridis protein, respectively. The FRC272 transcript, which was 13 times more highly expressed in fast muscle of myotube(−) fish, is composed of 6 exons that correspond to a 144-residue protein, with orthologs in human (58% identity) and T. nigroviridis (71% identity). This hitherto uncharacterized protein has a bipartite nuclear localization signal and an A1pp domain, indicating that it might be a cyclic phosphodiesterase.

The remaining four genes that were consistently differentially expressed in fast muscle of myotube(−) fish (FRC40 and FRC263) and myotube(+) fish (FRC177 and FRC300) showed ubiquitous tissue expression patterns and were classified as unlikely prospects. Numerous cellular processes vary with body size, e.g., teleost skeletal muscle aerobic enzyme activities show negative allometry, while glycolytic enzyme activities scale positively with body mass (68). We conclude that FRC40 and FRC263 from the myotube(−) library as well as FRC177 and FRC300 from the myotube(+) library are likely...
to be involved in biological processes related to body size but unconnected to the control of myotube formation.

The candidate genes potentially involved in myotube formation were identified among 1,452 ESTs generated from two subtracted cDNA libraries. Even though the programs for gene discovery are becoming increasingly accurate in predicting transcriptional units, it is often the case that these methods are unable to identify correctly the exon/intron boundaries and alternatively spliced transcripts (60). Clearly, this poses problems for genome annotation that can only be solved by combining gene predictions with experimental data, particularly ESTs (47). Aside from reducing the redundancy of the libraries, the SSH approach makes it possible to identify rare transcripts. Indeed, we have found a total of 278 EST clusters

Table 1. List of genes that are consistently and significantly differentially expressed in fast muscle of Takifugu rubripes at 2 distinct growth stages

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>Library</th>
<th>GenBank Accession No.</th>
<th>BLASTX Hit (GenBank Accession No.)</th>
<th>Conserved Domains (CDD Accession No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRC00040</td>
<td>Myotube (-)</td>
<td>CK829265</td>
<td>No significant hit found</td>
<td>No putative domains detected</td>
</tr>
<tr>
<td>FRC00167</td>
<td>Myotube (-)</td>
<td>CK829482</td>
<td>Mus musculus 4930469P12Rik protein (AAH21522)</td>
<td>L YR family (pfam05347)</td>
</tr>
<tr>
<td>FRC00214</td>
<td>Myotube (-)</td>
<td>CK829703</td>
<td>Pan troglodytes hypothetical gene (AK044523)</td>
<td>Kelch (smart00612)</td>
</tr>
<tr>
<td>FRC00258</td>
<td>Myotube (-)</td>
<td>CK829660</td>
<td>Gallus gallus hypothetical protein (XP_424765)</td>
<td>SPLa and RY anodine receptor (smart00449)</td>
</tr>
<tr>
<td>FRC00263</td>
<td>Myotube (-)</td>
<td>CK829990</td>
<td>Tetraodon nigroviridis unnamed protein product (CAF96211)</td>
<td>No putative domains detected</td>
</tr>
<tr>
<td>FRC00272</td>
<td>Myotube (-)</td>
<td>CK829695</td>
<td>Mus musculus expressed sequence A314976 (NP_997102)</td>
<td>Appr-1*-p processing enzyme (smart00506)</td>
</tr>
<tr>
<td>FRC00363</td>
<td>Myotube (+)</td>
<td>CK829877</td>
<td>No significant hit found</td>
<td>BTB/Kelch (IP000210)</td>
</tr>
<tr>
<td>FRC00386</td>
<td>Myotube (+)</td>
<td>CK829928</td>
<td>Gallus gallus CGI-20 protein (XP_414758)</td>
<td>No putative domains detected</td>
</tr>
<tr>
<td>FRC00405</td>
<td>Myotube (+)</td>
<td>CK829961</td>
<td>Danio rerio kelch-like protein Klhl (AAS84610)</td>
<td>BTB/Kelch (IP000210)</td>
</tr>
<tr>
<td>FRC00177</td>
<td>Myotube (+)</td>
<td>CK830424</td>
<td>Homo sapiens protein kinase C and casein kinase substrate in neurons 3 (NP_057307)</td>
<td>Src homology 3 (CD00174)</td>
</tr>
<tr>
<td>FRC00300</td>
<td>Myotube (+)</td>
<td>CK830205</td>
<td>Mus musculus 1700018O18Rik protein (AAH46793)</td>
<td>Na+/melibiose symporter (COG2211)</td>
</tr>
</tbody>
</table>
Fig. 7. Expression patterns of the strong candidate genes. A, C, E, and G: relative expression of the strong candidates in fast muscle (FM), slow muscle (SM), heart (HT), liver (L), skin (SK), and brain (B) of *T. rubripes*. Data are represented as myotube(−)/myotube(+) average expression ratios normalized against 18S rRNA levels. B, D, F, and H: relative spatial expression of the selected genes in myotube(−) fish. Mean transcript levels in different tissues normalized to the levels in fast skeletal muscle with 18S rRNA as standard. Gene clusters FRC258, FRC363, FRC386, and FRC405 are strong candidates, since their expression is upregulated mainly in fast muscle of myotube(−) fish relative to myotube(+) fish, and in myotube(−) fish their expression is restricted to fast muscle. Error bars indicate the standard error (*n* = 4). Where error bars are absent, data represent the average of 2 values, which did not differ by >30%.
found in the assembly (35) revealed that a total of 198 clusters could not be annotated of the Takifugu genome is stressed by the fact that most of these clusters were not predicted in Ensembl. These ESTs might correspond to UTRs, nonconserved exons, pseudogenes, or novel genes. Of these, 168 clusters could be translated with ESTscan or DECODER, and, therefore, they might represent previously unidentified genes.

We have also found several clusters with significant matches to uncharacterized putative proteins in other model systems, from Caenorhabditis elegans (50) to mouse (51). These putative proteins have been conserved through hundreds of millions of years (56) and are therefore likely to exert important, yet unknown, biological functions. Comparison between the Takifugu partial transcriptome and the Tetraodon genome assembly (35) revealed that a total of 198 clusters could not be found in the Tetraodon genome. Of these, 109 had matches to Takifugu predicted genes, and only 8 could not be found in the Tetraodon genomic sequence. These results might be due to gene prediction faults in the Tetraodon genome assembly, or they might represent genuine differences in gene repertoire. Even though T. rubripes diverged from T. nigroviridis only 18–30 million years ago (30), there are many significant physiological differences between these two species. For example, T. rubripes is a temperate marine species that can reach up to 70 cm in total length, whereas T. nigroviridis is a tropical brackish water fish that attains a maximum size of ~15 cm. Our current knowledge of the genetic networks that regulate myotube production and control fiber number in teleosts is extremely limited. There are only a few reports regarding factors involved in embryonic myotube formation in mammals. Dickson et al. (18) found that the alternative splicing of the neural cell adhesion molecule was regulated during myogenesis and that the constitutive expression of the short isoform induced myoblast fusion. Dedieu et al. (15) have recently shown that proteases from the calpain family are involved in induced myoblast fusion. Dedieu et al. (15) have recently shown that proteases from the calpain family are involved in induced myoblast fusion. These proteases may be involved in myotube formation in the adult stages of other vertebrates. Our current knowledge of the genetic networks that regulate myotube production and control fiber number in teleosts is extremely limited. There are only a few reports regarding factors involved in embryonic myotube formation in mammals. Dickson et al. (18) found that the alternative splicing of the neural cell adhesion molecule was regulated during myogenesis and that the constitutive expression of the short isoform induced myoblast fusion. Dedieu et al. (15) have recently shown that proteases from the calpain family are involved in induced myoblast fusion. Dedieu et al. (15) have recently shown that proteases from the calpain family are involved in induced myoblast fusion.

The nature of the MPCs responsible for the processes of myotube formation and hypertrophic growth is unknown. However, environmental factors such as temperature (41) and photoperiod (42) alter the number of myotubes formed while producing proportional changes in the myonuclei content of muscle fibers. These results are most consistent with a single population of MPCs responsible for both processes; however, the alternative hypothesis of separate populations of MPCs fated to produce myotubes or participate in hypertrophic growth cannot be ruled out. It is of interest that transcript levels of the four strong candidate genes identified in the present study were significantly higher in the fast muscle of fish that had stopped producing myotubes, suggesting they may be part of a genetic network inhibiting myotube production in adult stages. Future studies can use powerful experimental inference to test this hypothesis by determining whether orthologs of these genes are upregulated at the body length at which myotube production ceases in two other model species (T. nigroviridis and Danio rerio). Functional analysis of these genes using an in vitro myotube formation assay is also of interest. It is noteworthy that orthologs of all four strong candidate genes were also present in the human and murine genomes, suggesting that they may play a role in inhibiting myotube formation in the adult stages of other vertebrates.

ACKNOWLEDGMENTS

We thank Dr. Phil Snell (Rosalind Franklin Centre for Genomics Research, UK) for assistance with EST sequencing. We are grateful to Dr. Ralf Schmidt, Dr. Mark Blaxter, and Alasdair Anthony (Univ. of Edinburgh, UK) for helping us implement the EST pipeline. We are thankful to Mizuno Naoki (Univ. of Tokyo, Japan) for obtaining and keeping the fish and to Dr. Daisuke Ikeda (Univ. of Tokyo, Japan) for kindly providing tissue samples from T. rubripes. We thank the RIKEN Genome Exploration Research Group (Yokohama, Japan) for granting us permission to use DECODER.

GRANTS

This work was financed by an Environmental Genomics Thematic Grant (NER/S/2001/00250) from the Natural Environment Research Council of the United Kingdom.

REFERENCES


