GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Transcript profiling of expressed sequence tags from semimembranosus muscle of commercial and naturalized pig breeds

C.S. Nascimento[1], J.O. Peixoto[2], L.L. Verardo[1], C.F. Campos[1], M.M.C. Weller[1], V.R. Faria[1], M.E. Botelho[1], M.F. Martins[3], M.A. Machado[3], F.F. Silva[4], P.S. Lopes[1] and S.E.F. Guimarães[1]

[1]Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil
[2]Embrapa Suínos e Aves, Concórdia, SC, Brasil
[3]Embrapa Gado de Leite, Juiz de Fora, MG, Brasil
[4]Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: S.E.F. Guimarães
E-mail: sfacioni@ufv.br

Genet. Mol. Res. 11 (3): 3315-3328 (2012)
Received April 27, 2011
Accepted May 27, 2012
Published June 15, 2012
DOI http://dx.doi.org/10.4238/2012.June.15.1

ABSTRACT. In general, genetic differences across different breeds of pig lead to variation in mature body size and slaughter age. The Commercial breeds Duroc and Large White and the local Brazilian breed Piau are ostensibly distinct in terms of growth and muscularity, commercial breeds are much leaner while local breeds grow much slower and are fat type pigs. However, the genetic factors that underlie such distinctions remain unclear. We used expressed sequence tags (ESTs) to characterize and compare transcript profiles in the semimembranosus muscle of these pig breeds. Our aim was to identify differences in breed-related gene expression that might influence growth performance and meat quality. We constructed three non-normalized cDNA libraries from semimembranosus muscle, using two samples from each one, of these three breeds; 6902 high-quality ESTs were obtained. Cluster analysis

was performed and these sequences were clustered into 3670 unique sequences; 24.7% of the sequences were categorized as contigs and 75.3% of the sequences were singletons. Based on homology searches against the Swiss-Prot protein database, we were able to assign a putative protein identity to only 1050 unique sequences. Among these, 58.5% were full-length protein sequences and 17.2% were pig-specific sequences. Muscle structural and cytoskeletal proteins, such as actin, and myosin, were the most abundant transcripts (16.7%) followed by those related to mitochondrial function (12.9%), and ribosomal proteins (12.4%). Furthermore, ESTs generated in this study provide a rich source for identification of novel genes and for the comparative analysis of gene expression patterns in divergent pig breeds.

Key words: Expressed sequence tags; Muscle; cDNA library; Ham; *Sus scrofa*

INTRODUCTION

Selection for efficiency of meat production and meat quality is a major priority in modern pig production. Growth, carcass composition, and meat quality are important traits in pig production and they are known to depend on a variety of underlying factors. One of the main factors determining meat quality is the genetic background determined by breed (Rehfeldt et al., 2008). In addition, the biochemical properties and meat qualities of livestock may vary across individuals, muscles and breeds and the underlying factors remain unclear.

Improvements in growth performance can result in a reduction of meat quality. Commercial pig breeds have been intensively selected over the past decades for rapid, large and efficient accretion of muscle, which is believed to have led to deterioration in meat quality. Also, a muscle fiber type composition is often considered as a crucial inherent contributor to muscle metabolism and postmortem meat quality. Breed probably accounts for most of the genetic factors affecting the muscle fiber composition. Significant differences in meat quality and metabolic profiles between commercial and naturalized pigs have been reported (Lin and Hsu, 2005; Wimmers et al., 2008; Kim et al., 2008). Therefore, the comparison of commercial and naturalized breeds might provide important information about the genetic factors that are responsible for meat quality.

Duroc is an older breed of American domestic pig that forms the basis for many mixed-breed commercial hogs. It has desirable meat quality in terms of marbling, tenderness, and juiciness. Large White, a typical lean-type European breed, is now widely used for commercial production throughout the world (Briggs and Briggs, 1969). Large White is well known for its muscularity and leanness. Piau is a typical naturalized Brazilian pig breed and has been characterized by slow growth rate, high fat deposition and favorable meat quality (Souza et al., 2009; Serão et al., 2011). The Brazilian local pig breed originated from breeds introduced by Portuguese settlers in the XVI century and has also some influence of Dutch and African pigs (Vianna, 1985). These animals were used for breeding in small farms, their main characteristics being rusticity, adaptability to poor conditions of management and feeding and strong resistance to diseases. All the old Brazilian pig breeds are considered as fat type, sup-

plying farmers not only with meat, but also with a large amount of fat. However, in the last decades this fatness has become disadvantageous due to changes in consumer demand and the associated low production efficiency; thus these local pigs are on the way to extinction (Lopes et al., 2002).

Duroc, Large White and Piau are ostensibly distinct breeds with respect to growth and muscularity. The Brazilian naturalized pig breed differs from the commercial lines in terms of its adaptability to poor conditions of management and feeding, resistance to diseases, and fat and muscle deposition (Guimarães and Lopes, 2001). Given the observed differences across these three breeds, they should provide a good basis for the study of molecular mechanisms of breed-specific differences in meat quality.

The generation of expressed sequence tags (ESTs) from cDNA clones randomly selected from libraries constitutes an efficient and widely recognized strategy to identify and map genes (Nobis et al., 2003). ESTs that have been identified to date in pigs have been organized in the DFCI Pig Gene Index [http://compbio.dfci.harvard.edu/tgi/]. However, more ESTs are still needed than the current 1,376,756 [Release 14.0 March 11, 2010] sequences of the pig in order to identify, annotate and classify the genes specific to that species. In recent years, many papers have reported pig muscle ESTs (Davoli et al., 1999, 2002; Tang et al., 2007). However, the number of studies comparing different breeds is still limited, particularly in livestock.

The aim of this study was to assess differences in breed-related gene expression across two commercial pig breeds, Duroc and Large White, and the naturalized Brazilian Piau breed, primarily to identify genes that might play a role in skeletal muscle meat quality and growth development of the pig. For this purpose, we used ESTs to characterize and compare gene expression profiles in the semimembranosus (ham) muscle of these breeds. The resulting findings should be useful in improving an understanding of the molecular mechanisms responsible for breed-specific differences in growth performance and meat quality in pigs.

MATERIAL AND METHODS

Animals and sample collection

The semimembranosus muscle sample was collected from two adult male pigs (18 months old) of each breed: Duroc, Large White and naturalized Brazilian Piau breed. The animals were slaughtered by electrocution combined with exsanguinations and the muscle tissue was removed immediately under RNase-free conditions. For each animal about 1 g was collected and stored in RNAlater solution (Ambion, Austin, TX, USA) at -20°C for preservation and protection until RNA extraction procedure.

Construction of the muscle cDNA libraries

Total RNA was extracted using the RNeasy® Maxi kit (Qiagen, Hilden, Germany) according to manufacturer instructions. Contaminating genomic DNA was removed from total RNA solution by DNase Set (Qiagen) during RNA isolation procedure. Poly(A) RNA from muscle tissue was isolated using Oligotex® kit (Qiagen), according to manufacturer instructions.

Superscript™ Plasmid System with Gateway™ technology for cDNA Synthesis and Cloning kit (Invitrogen, Carlsbad, CA, USA) was used for construction of the three cDNA libraries. Double cDNAs were subjected to column chromatography with a cDNA size fractionation column (Invitrogen) and cDNAs with a minimum size of 0.3 kb were used for construction of cDNA libraries. Double-stranded cDNA fragments were cloned directionally into *Sal*I and *Not*I sites of pSPORT1 vector predigested with *Sal*I and *Not*I (Invitrogen), and transformed into *Escherichia coli* DH5α cells by electroporation.

Plasmid DNA purification was performed by modified alkaline lysis procedure and DNA samples were submitted to sequencing (Sambrook and Russell, 2001).

Nucleotide single-pass sequencing

Sequencing reactions were performed with 250 ng recombinant DNA template prepared on 96-well microplates using the DYEnamic ET dye terminator kit (Amersham Pharmacia Biotech, Piscataway, NJ, USA), followed by sequencing reaction clean up to remove residual dye and enzymes. Unidirectional single-pass sequencing was performed on a MegaBACE™ 1000 capillary sequencer (General Electric Healthcare, São Paulo, SP, Brazil).

Sequence processing and assembly

We used Phred (Ewing et al., 1998) to make base calls from sequence traces. The cleaned up sequences were performed on EGAssembler [http://egassembler.hgc.jp/] that provides a freely available web server. Briefly, the low-quality sequences (sequence length <100 bp, the percentage of undetermined bases >3% and low complexity) were discarded using SeqClean [http://compbio.dfci.harvard.edu/tgi/software/] with the option -A to disable the trimming of poly A/T tail.

After pre-processing, all sequences greater than 100 bp were used for contig assembly by using CAP3, a DNA sequence assembly program (Huang and Madan, 1999). The EST datasets were clustered and assembled into contigs with the parameters set greater than 80% similarity over at least a 40-bp fragment. ESTs that did not form contigs were designated as singletons. The resultant contigs and singletons were then referred to as unique sequences.

Transcript functional annotation

For functional annotation we used the Blast2GO v1.3.3 software (Conesa et al., 2005). Unique putative protein products were identified as the best hits obtained with BlastX against the Swiss-Prot protein database. These unique sequences were then annotated according to the following parameters: a Pro-Similarity-Hit-Filter of 5 and a Gene Ontology (GO) weight of 5. In all libraries, the BlastX parameters used an E-value of 10^{-05} and HSP length cutoff 33. The unique sequences were routinely compared with the curated InterPro database (InterProScan, EBI), Goslim, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) to identify functional motifs and gene. Directed acyclic graphs were generated using a sequence filter of 5, an alpha score of 0.6 and a θ node score filter. Frequency of each GO functional category was

then summarized and reported in the level 3 analysis, illustrating general functional categories. The coding region and the ORF from each unique sequence predicted using the results BlastX alignments by the TargetIdentifier software (Min et al., 2005). In order to identify specific genes expressed on muscle, muscle-specific sequences were downloaded from the TiGER website [http://bioinfo.wilmer.jhu.edu/tiger/] and matched against unique sequences using BlastX alignments with an E-value of 10^{-05}.

RESULTS

cDNA library construction, reads sequencing and sequence assembly

Three non-normalized cDNA libraries were constructed from poly(A) RNA extracted from semimembranosus muscle tissue from adult Duroc, Large White and Piau pig breeds. To identify ESTs, microtiter plates that contained randomly selected positive colonies were prepared for all three libraries. For the commercial breeds, 26 plates that contained colonies from the Duroc library and 29 microtiter plates that contained colonies from the Large White library were sequenced. For the naturalized Piau breed, 41 plates were sequenced. The average bp length for the cloned cDNA in all libraries was approximately 483.82 bp, ranging from 0.3 to 2.3 kb.

The chromatogram files from the sequencing were processed for removal of contaminant sequences and 6902 with high-quality EST sequences were used for computational analysis. A summary with the statistic of ESTs can be seen in Table 1. After processing, high quality ESTs (phred >20) were submitted to the dbEST database, with an average length of 100 bp per sequence. The GenBank® accession Nos. assigned to them were GR552441 to GR556356 and dbEST-Id 66295908 to 66299823.

Table 1. Statistical summary of cDNA library muscle sequences.

	Number of sequences (%)			
Libraries	Duroc	Large White	Piau	Total
Total ESTs sequenced	2126	2649	3607	8382
Trashed	18	15	1447	1480
Hight-quality ESTs	2108	2634	2160	6902
Unique sequences	1416	1568	686	3670
Contigs	418	193	294	905 (24.7)
Singletons	998	1375	392	2765 (75.3)
Mean length (nt)	466.93	577.82	406.70	483.82
GC level (%)	49.43	48.14	52.16	49.91

ESTs = expressed sequence tags; GC = guanine and cytosine.

The CAP3 sequence assembly program was used to group together redundant ESTs, which had overlapping sequences from all the libraries. Clustering to reduce data redundancy resulted in 3670 unique sequences representing different putative transcripts from pig breeds forming 905 (24.7%) contigs and 2765 (75.3%) singletons, defined as sequences that did not assemble into contigs using the defined assembly parameters. The mean unique sequence length ranged from as small as 100 bp to as large as >1400 bp (Figure 1) and the average guanine and cytosine content was 49.91% (Table 1).

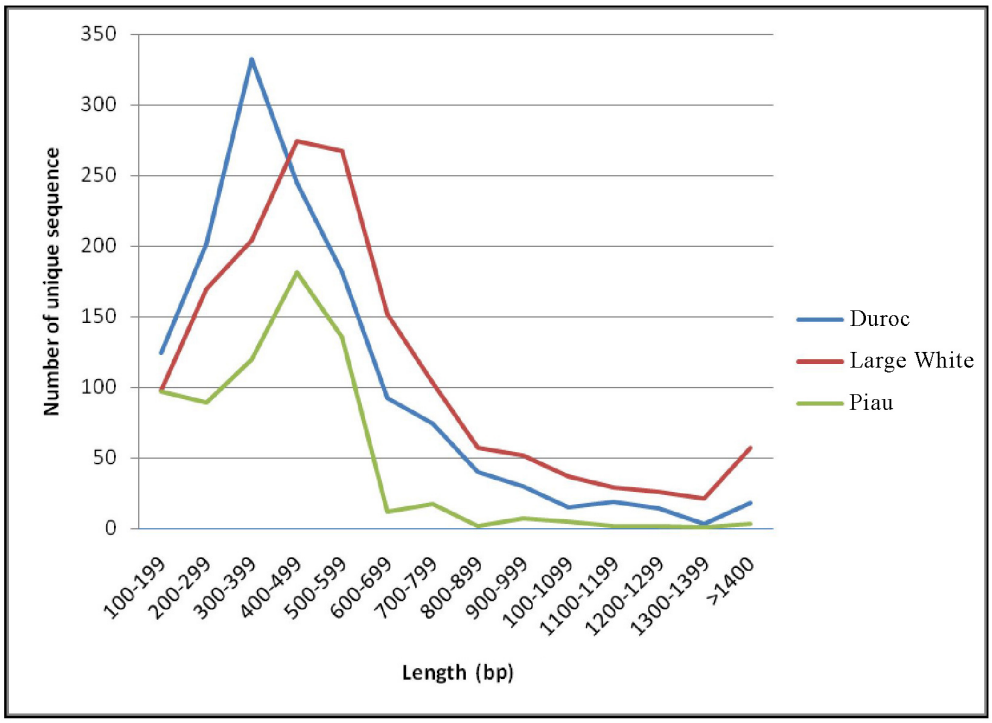

Figure 1. Graphical representation of the unique sequence size distributions by library.

Sequence annotation

For BLAST annotation, contigs and singletons were first searched using the BlastX against the Swiss-Prot protein database using a cutoff E-value of 10^{-5}. The E-value represents the number of alternate alignments, with the same or better total score, that could be expected to occur within the database purely by chance. Table 2 summarizes the statistics of BlastX hits. Based on this comparison, we were able to assign a putative identity to 1050 (28.6%) sequences matching homologous proteins on the three libraries. On the other hand, 2620 (71.4%) had no significant similarity to any proteins in the Swiss-Prot protein database. Classifying the genes based on species names of BlastX homologous sequences revealed 181 (17.2%) pig specific sequences.

TargetIdentifier program was used to investigate whether the first methionine within the amino acid translation represented the true N-terminal methionine. From the alignment of the Swiss-Prot protein database, 1050 hits matched sequences, a total of 614 (58.5%) revealed the full-length protein sequence (Table 2). Two hundred and four sequences were derived from unique sequences with partial 5' sequences and 62 were of 3' sequences. Other 170 sequences have a 5'-stop codon but do not have a start codon and were classified as ambiguous sequences.

Table 2. BlastX comparisons of all unique sequences to Swiss-Prot protein database sequences.

Libraries	Duroc	Large White	Piau	Total (%)
Hit unique sequence	218	430	402	1050 (28.6)
Contigs with match	83	131	200	414 (39.4)
Singlets with match	135	299	202	636 (60.3)
Size (nt)	353.95	584.43	594.34	510.91
Full-length	153	270	191	614
Partial length 5'	24	64	116	204
Partial length 3'	24	16	22	62
Ambiguous	17	80	73	170
Pig hits	26	80	75	181 (17.2)
No hit unique sequence	1198	1138	284	2620 (71.4)
Contigs with no match	335	62	94	491 (18.7)
Singlets with no match	863	1076	190	2129 (81.3)
Size (nt)	448.41	573.56	328.9	450.29

This table provides a working annotation of unique sequences described in this study, based on BlastX comparisons to Swiss-Prot protein database sequences.

Identification of the most abundant genes in each library

BlastX results were examined to identify libraries containing the greatest number of abundant genes. In this study, muscle structural and cytoskeletal proteins like actin, myosin and others (16.7%) were the most abundant transcripts followed by mitochondrial function (12.9%), ribosomal proteins (12.4%) and others (58.1%). The most representative class of proteins is represented in Table 3.

Table 3. Summary of the most abundant class of proteins (%).

Proteins class	Duroc	Large White	Piau	Frequency
Structural muscle/cytoskeletal	21 (9.6)	76 (17.7)	78 (19.4)	175 (16.7)
Mitochondrial	18 (8.3)	52 (12.1)	65 (16.2)	135 (12.9)
Ribosomal	17 (7.8)	70 (16.3)	43 (10.7)	130 (12.4)
Others	162 (74.3)	232 (54.0)	216 (53.7)	610 (58.1)
Total	218	430	402	1050

Gene ontology and KEGG classifications

We used Blast2GO to obtain global insights into the transcripts on our EST datasets by determining the significant molecular processes involved. The automatic procedure by Blast2GO-assigned GO terms were categorized in terms of their GO terms at the 3rd level annotation (Figure 2).

The KEGG were queried for sequence encoding enzymes. A total of 56 (from 1050 BlastX hits) sequences were mapped to 12 pathways, with 27 sequences representing metabolic enzymes characterized by unique Enzyme Commission numbers. More than half of the sequences (61.8%) were included in basic metabolic processes; the deduced gene products were associated with molecules involved in "Oxidative phosphorylation"-like cytochrome-c oxidase (N = 18) (Map: 01100) and NADH dehydrogenase, ubiquinone (N = 15) (Map: 01100), and H^+-transporting two-sector ATPase (N = 7) (Map: 00190). These products had the highest representation amongst the sequences mapped to KEGG pathways. Other unique sequences

were assigned into pathways involving environmental information processing (17.6%), genetic information processing (7.4%), human diseases (7.4%), and cellular processes (5.9%).

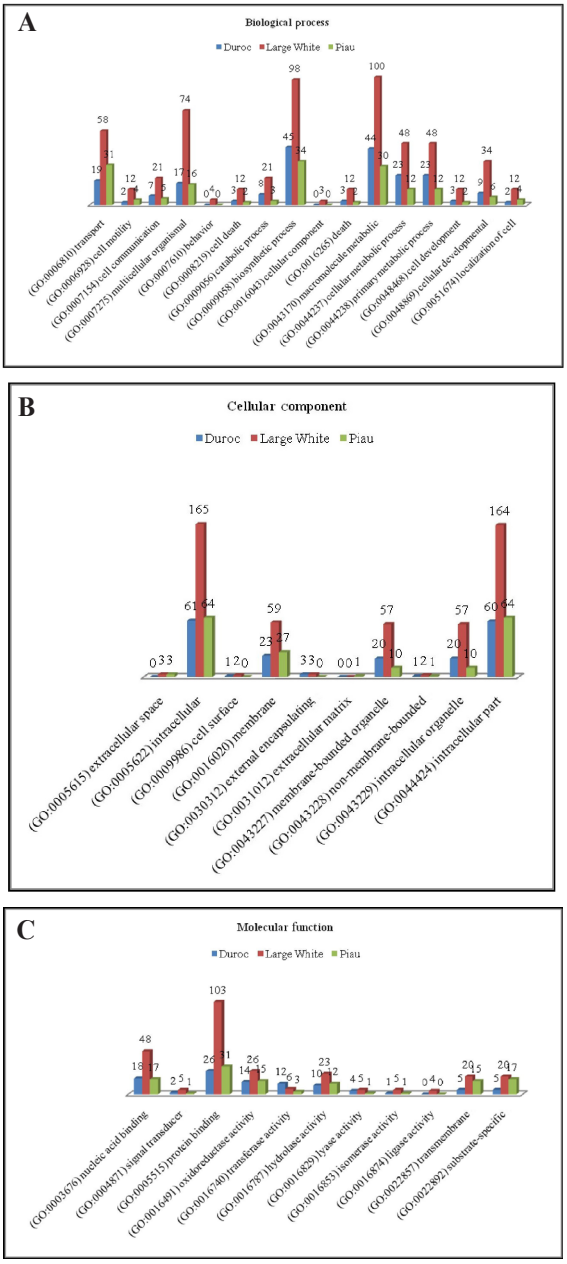

Figure 2. Gene ontology (GO) level distribution bar chart for muscle cDNA libraries. GO terms allowed assigning 210 annotations to Duroc (blue) library, 503 annotations to Large White (red) library, and 201 annotations to Piau (green) library. **A.** Biological process. **B.** Cellular component. **C.** Molecular function.

Comparisons of gene expression among commercial and Piau skeletal muscle

The Gene Venn plot was used to show the comparisons across the three pig DNA libraries to unique proteins. These unique sequences were classified according to their homolog in the BlastX searches on the Swiss-Prot protein database (Figure 3). The Venn diagram indicates that several genes were partially shared by two or more libraries, the other being exclusive for one of them. Ninety-two hits have homologous protein sequences exclusively from Duroc dataset, i.e., alpha-crystallin B chain (E-value = $1E^{-55}$); from Large White, 109 proteins, i.e., SH3 cysteine-rich domain-containing protein 3 (E-value = 1 E^{-52}), and another 249 for Piau dataset, i.e., inhibitor of growth protein 1 (E-value = 2 E^{-52}). Ninety unique sequences were found only in both Duroc and Large White, and include several muscle homologous sequences, i.e., tropomyosin, enolase beta, EEF1A-2, and catalase. Seventy-one were found in both Large White and Piau, i.e., nebulin, desmin, myozenin 1, and emerin. Ten homologous proteins had matched only in Piau and Duroc, i.e., Triadin. Thirty-eighty of these hits were homologous in the three cDNA libraries, i.e., aldolase A.

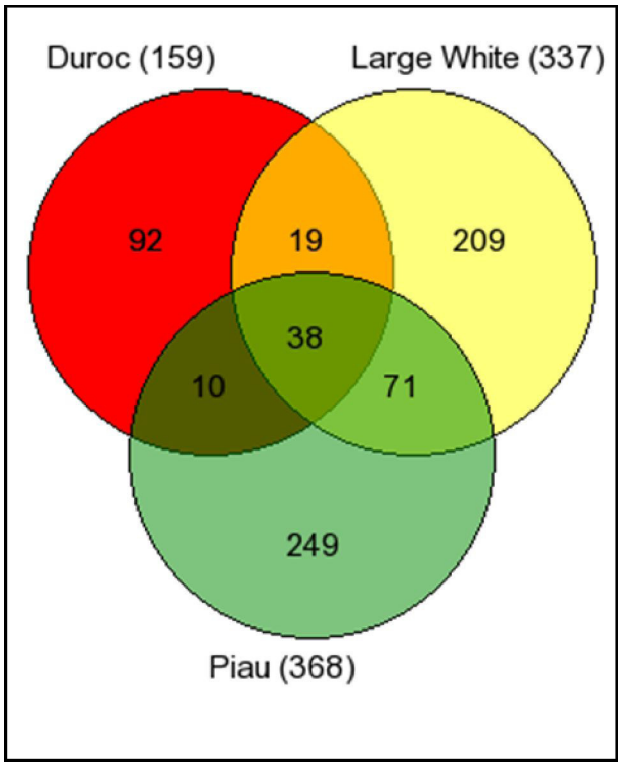

Figure 3. Venn diagram of the 864 hits with BlastX homology search (E-value <1 E^{-5}) into Duroc, Large White and Piau datasets. Red is representative of blast hits in Duroc, yellow indicates blast hits in Large White and green indicates blast hits in Piau datasets, orange indicates proteins expressed in Duroc and Large White only, light orange expressed in Duroc, Large White and Piau. The number of hits in each class is shown, including the number of total matches from each cDNA library against Swiss-Prot protein database.

Muscle-specific transcripts

The functional annotation of pig transcripts resulted in the identification of 149 muscle-specific proteins including all unique sequences (Figure 4). The group of unique sequences related to cytoskeleton components and molecules involved in contractility was composed of 86 different sequences. The most abundantly (>5) transcripts were characterized in order of frequency in the cDNA libraries that are listed in Table 4. Some of the identified molecules were structural components of the cytoskeleton or contribute to its organization. They are represented, among others, by microfilaments such as the beta polypeptide of actin alpha 1 (N = 18), tropomyosin 2 (beta) isoform 1 (N = 17), enolase 3 (N = 14), muscle creatine kinase (N = 12), and myosin, heavy chain 1, skeletal muscle (N = 11).

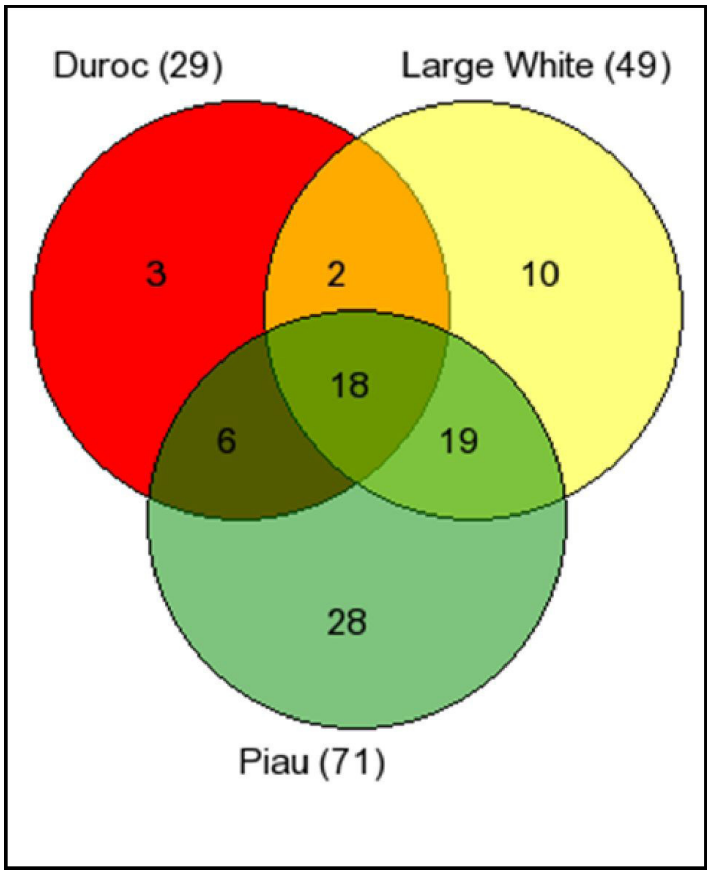

Figure 4. Venn diagram of protein muscle specific on cDNA libraries identified by BlastX homology search (E-value <1 E^{-5}). Red represents proteins expressed in Duroc; yellow indicates proteins expressed in Large White; green indicates genes expressed in Piau datasets; orange indicates proteins expressed in Duroc and Large White only; light orange is expressed in Duroc, Large White and Piau. The number within the parentheses indicates muscle proteins found in each library.

Table 4. The most abundant transcripts (>5) in *Sus scrofa* semimembranosus muscle cDNA libraries.

Protein name	Protein ID	Description	Duroc	Large White	Piau	Total
ACTA1	NP_001091.1	Actin, alpha 1, skeletal muscle	4	9	5	18
TPM2	NP_003280.2	Tropomyosin 2 (beta) isoform 1	5	7	5	17
ENO3	NP_443739.2	Enolase 3	4	8	2	14
CKM	NP_001815.2	Muscle creatine kinase	2	7	3	12
MYHC1	NP_005954.3	Myosin, heavy chain 1, skeletal muscle,	2	5	4	11
MYBPC2	NP_004524.3	Myosin binding protein C, fast type	2	4	4	10
MYL1	NP_524144.1	Fast skeletal myosin alkali light	1	4	4	9
CRYAB	NP_001876.1	Crystallin, alpha B	1	4	3	8
PYGM	NP_005600.1	Muscle glycogen phosphorylase isoform 1	1	4	3	8
PGAM2	NP_000281.2	Phosphoglycerate mutase 2	1	4	2	7
TNNC2	NP_003270.1	Fast skeletal muscle troponin C	1	5	1	7
TNNI2	NP_003273.1	Fast-twitch skeletal muscle troponin I	1	4	2	7
ATP2A2	NP_001672.1	ATPase, Ca^{2+} transporting, slow twitch 2	1	3	2	6
CA3	NP_005172.1	Carbonic anhydrase III	2	3	1	6
MYLPF	NP_037424.2	Myosin light chain, phosphorylatable,	1	4	1	6
MYLC2	NP_000423.2	Slow cardiac myosin regulatory light	2	2	1	5
PDLIM5	NP_001011515.1	PDZ and LIM domain 5 isoform d	2	1	2	5
TNNT1	NP_003274.3	Troponin T1, skeletal, slow isoform a	1	2	2	5
MYOZ1	NP_067068.1	Myozenin 1	0	4	1	5

DISCUSSION

We conducted an initial expression analysis, providing novel insight into the structure of ESTs from non-normalized libraries of pig semimembranosus muscle. A global analysis of 6902 high-quality ESTs revealed 3670 unique sequences from commercial and naturalized breeds and the sequence assembly resulted in approximately 24.7% contigs and 75.3% singletons. In addition, 686 (9.9%) unique sequences were only represented by transcript from the Piau naturalized pig breed. At the present time, there are no sequences from this breed on public sequence databases. On the other hand, data from Meishan and other breeds have already been described in more than one study (Tang et al., 2007; Ferraz et al., 2008). All ESTs were deposited in the dbEST database and will not only be a valuable addition to the current archived sequences from pig, but will also provide a major resource for the comparative analysis of gene expression profiles between commercial and naturalized breeds.

In the analysis of BlastX searches, a high percentage (71.4%) showed no significant similarity to entries on Swiss-Prot protein database and a low percentage of unique sequences (17.2%) were similar to *Sus scrofa* protein. Our results were similar to those found in other reported studies by Davoli et al. (1999, 2002) and Wang et al. (2006) for pig sequences. Furthermore, Bai et al. (2003) developed a porcine skeletal muscle microarray containing 5500 ESTs, but only 10% of clones have been identified. Our cDNA library experiment also suggested that most of the novel tags had come from unknown transcripts.

Among our abundantly represented genes, structural muscle and cytoskeletal, mitochondrial and ribosomal proteins were found. As expected, the most abundant transcripts expressed in semimembranosus muscle are related to structural muscle proteins, the primary component of the muscle (16.7%). Our results are in agreement with other studies on these genes (Davoli et al., 1999). For instance, Davoli et al. (1999) found 16.2% of mitochondrial genes and 9% of actin α skeletal muscle gene. It is known that genes involved in energy metabolism and in contraction are expected to be highly expressed in the muscle tissue due to the energy production and protein synthesis requirements (Mégy et al., 2002). Moreover, our results confirmed that pig tissue is characterized by a high level of expression of the mitochon-

drial genes that have been identified in human and porcine skeletal muscle.

In addition, analysis of biological function suggested that the unique transcripts differed relatively between the two genetic groups in certain functional categories of genes. Structural muscle and cytoskeletal, mitochondrial and ribosomal proteins were abundant in all pig breeds and genes in these functional categories exhibited different expression patterns among pig breeds. Functional annotation analysis showed that ENO3 (see Table 4) is relative highly expressed in Large White and Duroc breeds than in Piau. Enolase is a glycolytic enzyme (2-phospho-D-glycerate hydrolyase) that catalyses the interconversion of 2-phosphoglycerate to phosphoenolpyruvate in the glycolytic pathway. Wu et al. (2008) reported the first evidence that in skeletal muscle the expression of ENO3 is different between Yorkshire and Meishan pig breeds and propose that the ENO3 might play an important and different role in different breeds in terms of transcript level during muscle development. In addition, these authors detected a single-nucleotide polymorphism and carried out association studies, which concluded that the polymorphism may be associated with variation in several carcass traits and meat quality traits of interest for pig breeding.

Several proteins involved in energy metabolism and oxidative phosphorylation encoded by mitochondrial DNA showed differential abundance at the RNA level of the pig muscle datasets. In this study, we report increases of the abundance to NADH dehydrogenase (NDUAB, NDUAD, NDUB2, NDUB4, NDUB6, NDUB9, NDUBA, NDUS6), NADH-ubiquinone oxidoreductase (NU1M, NU2M, NU3M, NU4M, NU4LM, NU5M, and NU6M) and cytochrome c oxidase (CYB, QCR7, QCR9, and QCR10). Proteins encoded by these genes are subunits of the multi-subunit NADH: ubiquinone oxidoreductase (complex I) and this protein contains NADH dehydrogenase and oxidoreductase activities. It plays an important role in transferring electrons from NADH to the respiratory chain (Brandt, 2006). In the present study, the 55% of enzymes that encoded proteins to complex I were detected in the Piau dataset, 37% on Large White and only 8% on Duroc.

Finally, the expression of energy metabolism enzymes (NADH dehydrogenase, cytocrome c oxidase and cytocrome b reductase) was elevated in the Piau skeletal muscle compared to the commercial breeds, suggesting differences in oxidative and glycolytic metabolism across pig breeds. This indicates that the Piau breed possesses a higher oxidative capacity than the Large White and Duroc breeds. Kim et al. (2008) reported elevated expression of the oxidation-related metabolism gene NADH dehydrogenase in Korean naturalized pig muscle associated to meat quality indicated by a higher content of oxidative fibers and muscle fat, as well as redder meat color. Similar results were found by Hocquette et al. (1998) for skeletal muscles with a higher fat content that was also more oxidative. Furthermore, these studies demonstrated relationships between metabolic properties, especially oxidative metabolism and fat content in skeletal muscle. It must be in mind that despite all the original data accessed by the present study, any major inference still needs to be validated not only by a larger sample per breed but also by a larger number of sampled breeds.

CONCLUSIONS

In the present study, 181 unique sequences could be assigned to protein homologous in porcine and other 603 sequences in the non-redundant clustered dataset showed to be new expression tags or did not have homologous proteins in pigs. We identified several genes cod-

ing some of the main myofibrillar proteins, such as myosin heavy chain and myosin light chain, tropomyosin, troponin, tropomodulin 4, and actin in skeletal muscle and many several isoforms or splice variants were found. These genes, including those encoding the complex of myofibrillar proteins, displayed significant abundance of transcripts across breeds. In this regard, the use of the Brazilian naturalized Piau breed demonstrated the usefulness of the analysis of locally adapted pig breeds to better understand the physiology underlying growth in livestock production by comparing breeds with different growth and metabolic profiles. In addition, a significant achievement of this study was to increase the sequence information for these breeds through 869 putative novel sequences. Furthermore, ESTs generated in this study provided not only a rich source for identification of novel genes in pigs but also it was shown that divergent breeds may differently express important genes in myogenesis. Moreover, the comparative analysis of gene expression patterns in divergent breeds was done and provided strong evidence that different phenotypes are generated by divergent gene expression.

ACKNOWLEDGMENTS

Research supported by the Brazilian agencies FAPEMIG, CAPES, FINEP, and CNPq.

REFERENCES

Bai Q, McGillivray C, da CN, Dornan S, et al. (2003). Development of a porcine skeletal muscle cDNA microarray: analysis of differential transcript expression in phenotypically distinct muscles. *BMC Genomics* 4: 8.

Brandt U (2006). Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* 75: 69-92.

Briggs HM and Briggs DM (1969). In Modern Breeds of Livestock. The MacMillan Co., New York.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Davoli R, Zambonelli P, Bigi D, Fontanesi L, et al. (1999). Analysis of expressed sequence tags of porcine skeletal muscle. *Gene* 233: 181-188.

Davoli R, Fontanesi L, Zambonelli P, Bigi D, et al. (2002). Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig. *Anim. Genet.* 33: 3-18.

Ewing B, Hillier L, Wendl MC and Green P (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175-185.

Ferraz AL, Ojeda A, Lopez-Bejar M, Fernandes LT, et al. (2008). Transcriptome architecture across tissues in the pig. *BMC Genomics* 9: 173.

Guimarães SEF and Lopes OS (2001). Uso de recursos genéticos nativos no mapeamento genético de suínos. *Ação Ambiental* 15: 27-28.

Hocquette JF, Ortigues-Marty I and Pethick D (1998). Nutritional and hormonal regulation of energy metabolism in skeletal muscles of meat-producing animals. *Livest. Prod. Sci.* 56: 115-143.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877.

Kim NK, Lim JH and Song MJ (2008). Comparisons of *longissimus* muscle metabolic enzymes and muscle fibre types in Korean and western pig breeds. *Meat Sci.* 78: 455-460.

Lin CS and Hsu CW (2005). Differentially transcribed genes in skeletal muscle of Duroc and Taoyuan pigs. *J. Anim. Sci.* 83: 2075-2086.

Lopes PS, Guimarães SEF, Pires AV, Soares MAM, et al. (2002). Performance, Carcass Yield and Meat Quality Traits of F_2 Crosses Between Brazilian Native and Commercial Pigs for QTL Mapping. Proceedings in World Congresss on Genetics Applied to Livestock Production: Montpellier, 155-158.

Mégy K, Audic S and Claverie JM (2002). Heart-specific genes revealed by expressed sequence tag (EST) sampling. *Genome Biol.* 3: RESEARCH0074.

Min XJ, Butler G, Storms R and Tsang A (2005). TargetIdentifier: a webserver for identifying full-length cDNAs from EST sequences. *Nucleic Acids Res.* 33: W669-W672.

Nobis W, Ren X, Suchyta SP, Suchyta TR, et al. (2003). Development of a porcine brain cDNA library, EST database, and

microarray resource. *Physiol. Genomics* 16: 153-159.
Rehfeldt C, Henning M and Fiedler I (2008). Consequences of pig domestication for skeletal muscle growth and cellularity. *Livest. Sci.* 116: 30-41.
Sambrook J and Russell DW (2001). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor.
Serão NV, Veroneze R, Ribeiro AM, Verardo LL, et al. (2011). Candidate gene expression and intramuscular fat content in pigs. *J. Anim. Breed. Genet.* 128: 28-34.
Souza CA, Paiva SR, Pereira RW, Guimaraes SE, et al. (2009). Iberian origin of Brazilian local pig breeds based on Cytochrome b (MT-CYB) sequence. *Anim. Genet.* 40: 759-762.
Tang Z, Li Y, Wan P, Li X, et al. (2007). LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. *Genome Biol.* 8: R115.
Vianna AT (1985). Os Suínos. 14ª ed. Editora Nobel, São Paulo.
Wang XL, Wu KL, Li N, Li CL, et al. (2006). Analysis of expressed sequence tags from skeletal muscle-specific cDNA library of Chinese native Xiang pig. *Yi Chuan Xue Bao* 33: 984-991.
Wimmers K, Ngu NT, Jennen DG, Tesfaye D, et al. (2008). Relationship between myosin heavy chain isoform expression and muscling in several diverse pig breeds. *J. Anim. Sci.* 86: 795-803.
Wu J, Zhou D, Deng C, Wu X, et al. (2008). Characterization of porcine ENO3: genomic and cDNA structure, polymorphism and expression. *Genet. Sel. Evol.* 40: 563-579.