Int. J. Dev. Biol. 48: 333-337 (2004)

EST analysis of mRNAs expressed during embryogenesis in *Gallus gallus*

ERIKA C. JORGE[1], CLAUDIA B. MONTEIRO-VITORELLO[2], HELENA J. ALVES[1], CLARISSA S. SILVA[1], RACHEL G. BALAN[1], MATEUS PATRÍCIO[1] and LUIZ L. COUTINHO*,[1]

[1]Departamento de Zootecnia and [2]Departamento de Fitopatologia, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, São Paulo, Brasil

ABSTRACT Chicken Expressed Sequence Tags (ESTs) were analyzed to identify genes associated with myogenesis during embryonic development. A total of 6,184 ESTs were generated from three cDNA libraries constructed from whole embryos (Stage 26), somites associated with neural tube (Stage 15), and limb buds (Stages 21, 24 and 26). Clustering and assembly of 4,998 valid ESTs resulted in 2,329 unique sequences with 902 clusters (38.7%) and 1,427 singletons (61.3%). There are more than 400,000 chicken ESTs available at GenBank and we were able to identify 143 novel sequences. From these, 45 sequences found either a human EST homolog or a match with conserved regions among proteins. Most of these sequences were found to be expressed in somites, an important tissue for muscle development and not characterized before. This study revealed the value of micro dissected embryonic libraries for describing gene expression profiles associated with myogenesis and gene discovery.

KEY WORDS: *myogenesis, morphogenesis, chicken embryo, somite, limb bud*

Characterizing Expressed Sequence Tags (ESTs) on a large scale has proven to be an efficient strategy for gene discovery, genome annotation and comparative genomics (Adams *et al.*, 1991; Hatey *et al.*, 1998; Rudd, 2003). Recent progress in genomic research allows numerous questions to be addressed in several areas, including embryonic development. In embryogenesis, the main goal is to understand the molecular mechanisms responsible for correct establishment of diverse structures of an organism. Chicken is an ideal vertebrate model organism for embryology because its development occurs *in ovo* rather than *in utero.* Chicken has also provided an important model for cancer, immune system and gene regulation studies.

Skeletal muscle formation in vertebrates has been largely used as a model for cell fate determination and differentiation. These events are controlled by precise activation of gene expression at the appropriate time and place (Ludolph and Konieczny, 1995). Precursor skeletal muscle cells in vertebrates originate in somite (Pourquié, 2001). During outgrowth, the somites give rise to both sclerotome (ventrally) and dermomyotome (dorsally). The first myogenic precursor cells arise in the medial lip of dermomyotome, that ultimately give rise to the deep muscles of the back (epaxial musculature). Hypoaxial musculature arises in the lateral lip of dermomyotome by two distinct mechanisms: (1) a non-migrating precursor cells, that immediately differentiate in intercostals muscles and muscle of the body wall; and (2) a migratory cell population, precursor of skeletal musculature of the limbs; diaphragm and tongue (Stockdale *et al.*, 2000; Birchmeier and Brohmann, 2000; Buckingham *et al.*, 2003).

Somites and developing limb buds were selected for this work to identify and categorize the profile of genes expressed during skeletal muscle development in *G. gallus,* to identify new genes, besides those involved in embryogenesis, and to perform a comparative study with human sequences.

cDNA libraries and sequencing

A total of 4,998 ESTs (Phred quality > 20 and at least 200 bp) were considered in our analysis (Table 1). The number of cDNA clones that extended from the start-codon (full-length cDNA) was calculated based on BLASTX analysis, and the frequency in each library was 17.4% (41 unique sequences) for EM1; 21% (214) for LB1; and 34.7% (452) for SM1, suggesting that these clones may contain the complete coding sequence. All 4,998 sequences were deposited in the dbEST division of GenBank (access numbers: CD760792 –

Abbreviations used in this paper: EST, expressed sequence tag; LS, library specific; TIGR, The Institute for Genomic Research.

***Address correspondence to:** Dr. Luiz Lehmann Coutinho. Departamento de Zootecnia, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, São Paulo, Brasil. Fax: +55-19-3429-4285. e-mail: llcoutin@esalq.usp.br

Note: All 4,998 valid EST sequences studied in this paper were deposited in the dbEST division of GenBank (accession numbers: CD760792 – CD765430).

0214-6282/2004/$25.00
© UBC Press
Printed in Spain
www.ijdb.ehu.es

CD765430).

Clustering and assembly of ESTs

Clustering analyses were performed intra-library and inter-library (EM1/LB1/SM1) (Table 1). Most of the clusters from all three libraries were assembled with the alignment of up to four ESTs: 81 (92%) in EM1; 388 (80%) in LB1; and 361 (88%) in SM1 (Fig. 1). The number of four ESTs in a cluster has been used as a threshold to calculate redundancy index in other works (Tirunagaru *et al.*, 2000; Jia *et al.*, 2001; Boardmann *et al.*, 2002). Comparable novelty index (56.4%) was observed for a similar number of ESTs analyzed from other cDNA libraries (Tirunagaru *et al.*, 2000; Jia *et al.*, 2001). The somites library (SM1) had a novelty index of 54.6% and provided a more diversified mRNA population as compared to limb buds (P <.0001, using a chi-square test) and thus, a superior gene discovery rate.

Library-specific ESTs

EST sequencing from different libraries allows identification of genes putatively transcribed within a certain tissue or during a particular developmental phase. Our analysis revealed 2,090 (89%) library-specific unique sequences (1,427 singletons and 663 clusters) (Table 2). Joint analysis assembled 902 clusters rather than 984 (sum of clusters for each library). The same result was observed among the singletons (1,427 instead of 1,569) (Table 2). The level of reorganization among clusters and singletons in the joint analysis is also indicative of low redundancy among libraries.

From the 902 clusters, 321 (ca. 26.5%) were assembled by ESTs expressed in at least two of the libraries. These common clusters represent ESTs that are concurrently expressed in the different tissues studied. Most of them represent housekeeping genes and

TABLE 1

ANALYSIS OF EACH cDNA LIBRARY (EM1, LB1, SM1) AND ALL DATA GATHERED (EM1 LB1 SM1) REGARDING NUMBER OF ESTs AND CLUSTERIZATION

	EM1	LB1	SM1	EM1 LB1 SM1
Total number of valid EST	383	2,328	2,287	**4,998**
Total number of clusters	88	485	411	902
Number of EST in clusters	232	1,801	1,396	3,571
Total number of singletons	151	527	891	1,427
Total number of unique sequences	239	1,012	1,302	**2,329**
Novelty index*	62.4 %	43.5%	54.6%	45.7%

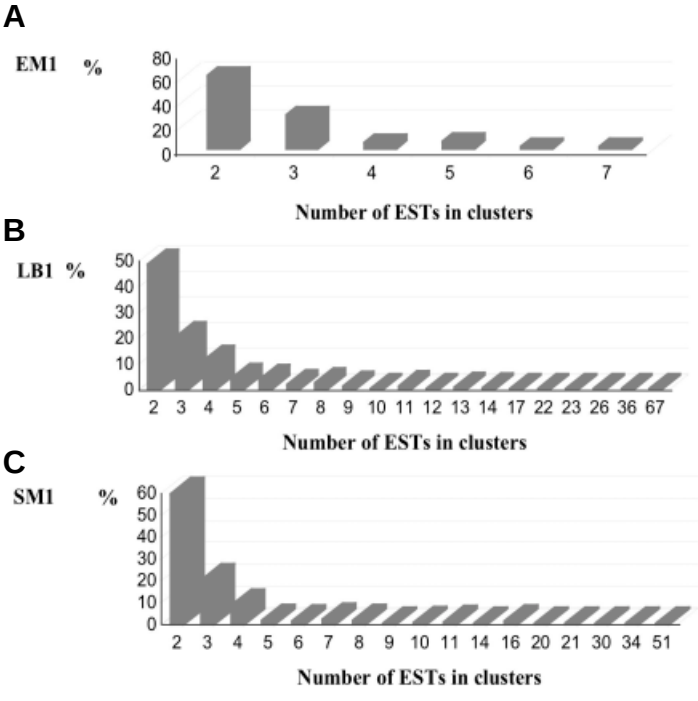
* Novelty indexes were assessed for each library by calculating the fraction of each library composed of unique sequences (tentative consensus plus singletons) within each library.

subunits of ribosomal proteins. The 663 contigs assembled by ESTs from a single library are divided as: 51 (58%) specific to EM1; 335 (69%) to LB1; and 277 (67%) to SM1 (Table 2). All singletons identified in joint cluster analysis (1,427) were considered to be library-specific. The level of library-specific unique sequences most likely reflects differential gene expression among tissues analyzed. The specific contribution index indicated that SM1 is the library that contributed to most of the new sequences (Table 2).

Functional annotation: gene expression profile of skeletal muscle precursor cells

All unique sequences were compared with the protein database of GenBank using BLASTX and were functionally annotated into twelve categories as proposed by TIGR (Fig. 2). A website supplement was created and all detailed information are available at https://maltes.esalq.usp.br/Chicken-Embryo/ (Table A).

From a total of 2,329 unique sequences, 1,474 (63.3%) were



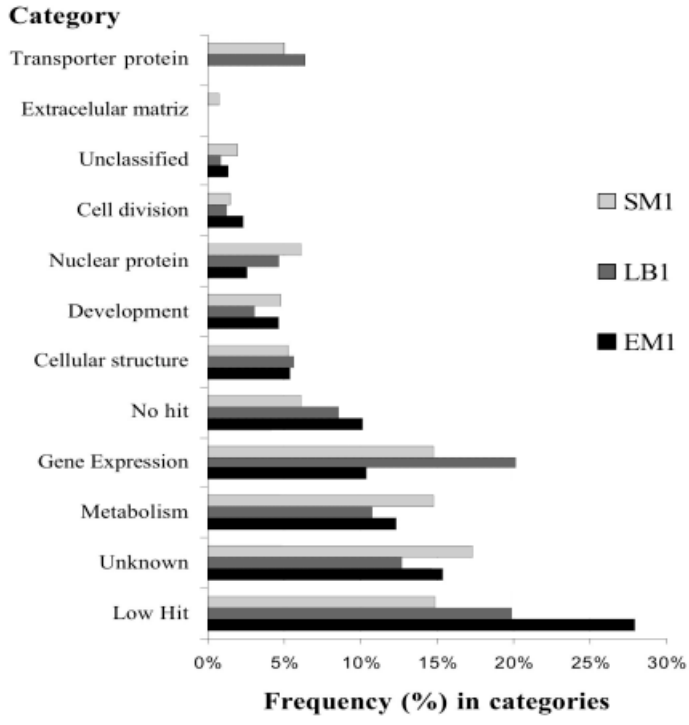


Fig. 1 (Left). Intra-library clustering distribution. *Analysis of clusters assembled by Cap3 in* **(A)** *EM1 library,* **(B)** *LB1 library and* **(C)** *SM1 library.*

Fig. 2 (Right). Functional annotation of the unique chicken sequences into twelve categories as proposed by TIGR. *See text for details.*

TABLE 2

IDENTIFICATION OF SEQUENCES SPECIFIC TO A SINGLE cDNA LIBRARY (EM1, LB1, SM1)

Library	# of clusters	LS* clusters	# of singletons	LS* singletons	# of LS* unique sequences	Specific Contribution**
EM1	88	51 (58%)	151	107 (70.8%)	158	3.0%
LB1	485	335 (69%)	527	471 (89.4%)	806	16.0%
SM1	411	277 (67%)	891	849 (95.3%)	1,126	22.5%
S	-	663	-	1,427	2,090	-

* LS, Library specific; ** Specific contribution was calculated by dividing the number of library-specific unique sequences by the total number of ESTs analyzed in the libraries (4,998).

similar to other sequences and 855 (36.7%) gave either low hit (BLASTX *E*-value >10^{-5}) or no hit against that database. Among the most abundant transcripts identified were hemoglobin alpha chain (164 ESTs; 3.2%); cytochrome c oxidase (87 ESTs; 1.7%); different ribosomal protein subunits (77 ESTs; 1.5%); and elongation factor 1 (36 ESTs; 0.7%). Most of them grouped into the categories of 'gene expression' (~12%) and 'metabolism' (~15%) as expected, since these categories include gene products of cellular maintenance.

The 'unknown' category grouped 769 ESTs (average of 15%). Some of them (28) were clusters composed of more than four ESTs, meaning that they represent highly expressed genes, and were identified as library-specific (data available as Table B of our website). All sequences classified as 'unknown', 'no-hit' and 'low-hit' represent a group of expressed sequences identified in embryonic tissues with no described biological function, and future studies are necessary to determine their function.

Unique sequences that identified homologues with known genes are of special interest due to their association with embryogenesis. The 'development genes' category was composed of 130 unique sequences (~5%) of the three-cDNA libraries (Fig. 2; Table C of the website). Notation data revealed that the profile of genes related to development was different for each of the tissues analyzed, although the expression profile of housekeeping genes was similar. Regarding skeletal muscle formation, myogenic factors (MyoD, Myf5, MRF4 and myogenin) responsible for muscle cell determination and differentiation (Ludolph and Konieczny, 1995), were not identified in the libraries analyzed. The EST analysis of non-normalized cDNA libraries, which is our case, preferentially, selects for abundant messenger RNAs. Genes often expressed at comparatively low levels, such as myogenic regulators, are only found after exhaustive sequencing. Prior work had determined MyoD to be 100 times less abundant than β–actin in stages E21-E26 chicken embryos (Alvares *et al.*, 2003). In the present study, eleven ESTs (0.47%) coding for β-actin were identified in the LB1 library. This observation indicates that by sequencing less than 3,000 clones, only the more abundant genes are identified. In the SM1 library, no β-actin EST was identified, while one EST coding for GAPDH was found. These calculations indicate that to find the myogenic factors in our libraries, it would require sequencing about ten times as many EST.

Proteins of the 'Extracellular matrix' were specifically identified in the SM1 library (14; 0.6%), where they might have a function of adhesion in the spaces between consecutive somites (Bergmann *et al.*, 1995). Proteins that regulate cell cycle represented 2% (44) of the categorized sequences. Regulation of myoblast cell cycle is an essential pre-requisite for proliferation and correct differentiation of muscle fibers (Walsh and Perlman, 1997; Polesskaya and Rudnicki, 2002).

Identification of putative new chicken genes

One of the major goals of high throughput EST sequencing is gene discovery. Among the 2,329 unique sequences, 143 (~6%) were identified as having no hit at protein databases (Table D of the website). This analysis was also performed considering the recent submission of 320,000 chicken ESTs in the GenBank (Boardman *et al.*, 2002). It is known that embryo tissues present a high rate of gene discovery, with similar index of normalized cDNA libraries (Boardman *et al.*, 2002).

Of the 143 putative novel *G. gallus* sequences identified, 41 (29%) were similar to cDNA clones found at the human dbEST (BLASTN, *E*-value <10^{-5}, 60% coverage of the unique sequence). Analysis of reading frame and search for conserved regions performed with Blocks (http://blocks.fhcrc.org/blocks) allowed us to identify 22 different motifs within these sequences (Table 3).

Possible protein variants were identified considering that the

TABLE 3

NOVEL CHICKEN SEQUENCES THAT FOUND EITHER A HUMAN EST HOMOLOG OR A MATCH WITH CONSERVED REGIONS AMONG PROTEINS USING BLOCKS (HTTP://BLOCKS.FHCRC.ORG/BLOCKS)

Clone	*GenBank Id*	Motifs	E-value
GGEZEM1004B02.g	**CD763275**	NCRI*	
GGEZEM1005A07.g	**CD763356**	Tyrosine kinase catalitic domain;	2.8E-05
GGEZLB1002H11.g	**CD763557**	NCRI*	
GGEZLB1002E07.g	**CD763520**	CUB domain;	5.6E-12
GGEZLB1005G12.g	**CD763793**	KOG3655: Drebins and related actin binding protein;	5E-53
GGEZLB1010F04.g	**CD764174**	Myristolated alanine rich C kinase; 30 amino acids	1E-10
GGEZLB1013B10.g	**CD764381**	KOG2410: Gamma-glutamyltransferase;	1E-27
GGEZLB1013D10.g	**CD764405**	NCRI*	
GGEZLB1017F09.g	**CD764592**	NCRI*	
GGEZLB1018G07.g	**CD764673**	NCRI*	
GGEZLB1019A04.g	**CD764693**	Immunoglobulin and major histocompatibility complex 2;	4.7E-05
GGEZLB0119H09.g	**CD764762**	CXYORF1 protein;	2E-34
GGEZLB1020H02.g	**CD764832**	GG14977-PA	2E-36
GGEZLB1021H03.g	**CD764914**	NCRI*	
GGEZLB1029C04.g	**CD765383**	NCRI*	
GGEZSM1002G12.g	**CD760951**	RGS-GAIP interacting protein; 68 amino acids	8E-22
GGEZSM1005C11.g	**CD761065**	NCRI*	
GGEZSM1007B01.g	**CD761195**	PolyC binding protein;	8E-62
GGEZSM1007F01.g	**CD761240**	Hypothetical ptn RIKEN cDNA;	2E-06
GGEZSM1007H02.g	**CD761261**	Alpha-2 macroglobulin family;	2E-36
GGEZSM1010A10.g	**CD761421**	NCRI*	
GGEZSM1010A12.g	**CD761423**	Hypothetical protein;	4E-51
GGEZSM1010F05.g	**CD761447**	Digeorge syndrome critical region gene 8; 70 amino acids	7E-13
GGEZSM1011H01.g	**CD761539**	Uncharacterized conserved protein;	6E-10
GGEZSM1014F02.g	**CD761769**	Hypothetical protein;	1E-26
GGEZSM1014G05.g	**CD761783**	Unnamed protein product;	1E-75
GGEZSM1018G03.g	**CD761947**	Na+ iodide/Myo_inositol;	1E-19
GGEZSM1019G11.g	**CD762036**	Zinc finger transcription factor;	1.6E-31
GGEZSM1020C12.g	**CD762079**	GABA receptor associated protein;	5E-62
GGEZSM1023C01.g	**CD762320**	Abhydrolase 2 phospholesterase/carboxylesterase	7E-51
GGEZSM1024E02.g	**CD762432**	Seryl tRNA synthetase;	1E-24
GGEZSM1024H04.g	**CD762468**	Hypothetical protein;	1E-05
GGEZSM1025H05.g	**CD762558**	WW domain containing protein 2 - Protein Kinase C;	5E-83
GGEZSM1026A10.g	**CD762575**	Na+/K+ ATPase alpha subuint;	2E-98
GGEZSM1026H11.g	**CD762656**	Nuclear protein export factor;	2E-16
GGEZSM1028E02.g	**CD762786**	Chromosome 11 ORF; Uncharacterized conserved protein;	5E-57
GGEZSM1029A07.g	**CD762829**	Oligophrenin 1 and related Rho GTPase;	1E-08
GGEZSM1030F10.g	**CD762976**	NCRI*	
GGEZSM1030G09.g	**CD762986**	NCRI*	
GGEZSM1030H10.g	**CD762997**	KOG2301: Voltage gated Ca2+ channels;	5E-35
GGEZSM1006G05.g	**CD761167**	Dolichyl P Mannose	1E-79
GGEZSM1011F08.g	**CD761525**	Synaptonemal complex protein	1E-43
GGEZSM1021G01.g	**CD762198**	Corticosteroid desydrogenase	6E-44
GGEZSM1022B01.g	**CD762228**	Almstrom syndrome protein	2E-07

* NCRI, no conserved region identified

reading frames found in long EST showed high degree of amino acid identity with small regions of large proteins. CD764174 is similar to a myristolated alanine rich C kinase (100% identity, 30 amino acids) and CD761447 is similar to DiGeorge syndrome critical region gene 8 (52%, 70 amino acids).

Among the 102 (71%) remaining sequences identified as 'no hit' against the human EST database, we were able to identify conserved motifs in four of them, which are dolichyl p mannose (CD761167); synaptonemal complex protein (CD761525); corticosteroid desydrogenase (CD762198); and almstrom syndrome protein (CD762228) (Table 3). All these 45 sequences (41 that found a match in the human dbEST plus these four sequences showing conserved motifs) probably represent new genes for *G. gallus* and deserve a special attention in future studies. Further experimental analysis have to be performed to consider all the others as new genes as they may represent splice variants which have not been sequenced by other chicken EST projects, artifacts or chimeric EST. Therefore, the strategy to analyze ESTs from specific embryonic tissues or micro dissected-tissues (somites and limb buds) effectively contributed to identification of novel genes.

Comparative analysis

Comparative genomics allows identification of a subset of highly conserved genes that likely have common functional roles across vertebrate species. Similarity searches using tBLASTx was used to determine sequence conservation between *G. gallus* and *Homo sapiens*. Approximately 69% of our collection has a significant match to the human dbEST, considering a tBLASTx *E*-value < 10^{-5} and at least 60% protein sequence coverage. The average identity between *G. gallus* and *H. sapiens* amino acid sequences was 80.3%. No human dbEST match was found for 722 (31%) unique sequences.

Comparative analysis with human dbEST was also conducted among the various functional categories to establish the most evolutionarily conserved groups. Proteins classified in 'Gene expression' and 'Cell structure' categories presented the greatest tBLASTx average identity (86%), followed by 'Extracellular matrix', 'Cell division' and 'Nuclear proteins' (83%). 'Developmental' (81%) exhibited a larger degree of conservation when compared to 'Metabolism' (77%) and 'Unknown' (76.3%).

Among the unique sequences classified in 'Development' category, 93% (121 unique sequences) found a match with a human orthologue. Fibroblast growth factor receptors (CD763195, CD763444), Bone morphogenic protein 1 (CD763520), as well as other genes with putative function in embryonic development were identified with average identity greater than 85% (Pourquié, 2001; Buckingham *et al.*, 2003).

Chicken is an important model organism for embryology and development studies and our results reveal the potential to use the sequences expressed in *G. gallus* to study morphogenesis in other vertebrate species. This is particularly critical for humans due to the ethical and technical difficulties in obtaining embryos or embryonic tissues, such as somites and neural tube. Most of the 5 million human ESTs published to date are from adult tissues (Ko, 2001).

In summary, we have analyzed ESTs obtained from chicken skeletal muscle precursor tissues to determine gene expression profile during myogenesis, to identify novel genes and chicken homologues of known genes. This approach has led to the identification of several genes related to morphogenesis and skeletal muscle development. Although much is known about muscle formation in vertebrates, a collection of genes expressed during the crucial times of muscle cell determination, migration and differentiation was not available for chicken embryos. More importantly, this study revealed the value of micro dissected embryonic libraries for gene discovery.

Experimental Procedures

Embryo collection of G. gallus *and RNA extraction*

Eggs from a broiler-chicken line from *Agroceres-Ross* (AG-Ross, Tietê, São Paulo, Brazil) were incubated at 37ºC in a moist atmosphere to obtain embryo stages as defined by Hamburger and Hamilton (1951). Three cDNA libraries were constructed from: (1) whole embryos at stage E26 (6 days; EM1); (2) somites associated with the neural tube at stage E15 (50 hours incubation; SM1); and (3) limb buds in three developmental stages E21, E24 and E26 (3.5 days, 4.5 days and 6 days, respectively; LB1). For the SM1 library, only well-defined somites identified in the stage E15 and the length of neural tube/notochord associated with them were microdissected.

Pools of embryos (10) and microdissected tissues (60 embryos for LB1 and 200 for SM1) were homogenized in Trizol Reagent (Invitrogen Co, Carlsbad, CA, USA) to isolate total RNA. Poly(A)$^+$ RNA was purified using the MessageMaker® Reagent Assembly Kit (Invitrogen), following manufacturer's directions.

cDNA libraries

cDNA libraries were constructed using 1-2 μg poly(A)$^+$ RNA and the SuperScript Plasmid System kit (Invitrogen) according to manufacturer's protocol. Fractions containing cDNA larger than 500 pb were ligated into pSPORT1 (Invitrogen) at the *Sal* I-*Not* I site.

Plasmid preparation and sequencing

Plasmid DNA was purified using a modified alkaline lysis method (Sambrook *et al.*, 1989). Sequencing reactions were conducted using the DYEnamic ET Dye Terminator sequencing kit (Amersham Biosciences, Piscataway, NJ, USA). Clones were sequenced from the 5' end with T7 primer. Sequencing reactions were analyzed on MegaBACE 1000 automated DNA sequencer (Amersham Biosciences).

Data processing

Base calling, quality control and clustering of the *G. gallus* ESTs were performed using softwares Phred (Ewing *et al.*, 1998), Cap3 (Huang and Madan, 1999) and Consed (Gordon *et al.*, 1998). All EST sequences with Phred quality greater than 20 and at least 200 bp were considered for clustering. Prior to clustering, vector sequences, adaptors and poly(A)$^+$ were identified and screened out (Ewing *et al.*, 1998). Chicken rRNAs, mitochondrial DNA and bacterial contamination were removed from the database.

Sequence similarity searches of the unique sequences were conducted using BLAST programs (Altschul *et al.*, 1990). The GenBank databases used were non-redundant (nr) and dbEST (*G. gallus* and *Homo sapiens*). All searches against dbEST were carried out locally. Twelve categories were established for classification of unique sequences according to their putative biological functions based on the ones proposed by The Institute for Genomic Research (TIGR, http://www.tigr.org): gene expression; metabolism; cell division; extracellular matrix; nuclear proteins; transporter proteins; development; conserved of unknown function; low hit (assuming a BLASTX *E*-value < 10^{-5}); no hit; and not classified. All data gathered in this project was organized with the database management system MySQL version 3.23.54 (www.mysql.com). Supplementary information is provided in our website (https://maltes.esalq.usp.br/Chicken-Embryo).

Acknowledgements

This work was financed by Fundação de Amparo a Pesquisa do Estado de São Paulo, FAPESP. L.L. Coutinho is a recipient of a research productivity scholarship from Conselho Nacional de Pesquisa, CNPq.

References

ADAMS, M.D., KELLY, J.M., GOCAYNE, J.D., DUBNICK, M., POLYMEROPOULOS, M.H., XIAO H., MERRIL C.R., WU, A., OLDE, B., MORENO, R.F., KERLAVAGE, A., MCCOMBIE, W.R. and VENTER, J.C. (1991). Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. *Science* 252: 1651-1656.

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. and LIPMAN, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215(3): 403-410.

ALVARES, L.E., MANTOANI, A., GABRIEL, J.E., and COUTINHO, L.L. (2003). Standard-curve competitive RT-PCR quantification of myogenic regulatory factors in chicken embryos. *Braz. J. Med. Biol. Res.* 36:1629-1641.

BERGMANN, A.D., CHENG, H.J., BRAMBILLA, R., KLEIN, R. and FLANAGAN, J.G. (1995). ELF-2, a new member of the Eph ligand family, is segmentally expressed in mouse embryos in the region of the hindbrain and newly forming somites. *Mol. Cell Biol.* 15: 4921-4929.

BIRCHMEIER, C. and BROHMANN, H. (2000). Genes that control the development of migrating muscle precursor cells. *Curr. Opin. Cell Biol.* 12: 725-730.

BOARDMAN, P.E, SANZ-EZQUERRO, J., OVERTON, I.M., BURT, D.W., BOSCH, E. FONG, W.T., TICKLE, C., BROWN, W.R., WILSON, S.A., and HUBBARD, S.J. (2002). A comprehensive collection of chicken cDNAs. *Curr. Biol.* 12(22): 1965-1969.

BUCKINGHAM, M., BAJARD, L., CHANG, T., DAUBAS, P., HADCHOUEL, J., MEILHAC, S., MONTARRAS, D., ROCANCOURT, D. and RELAIX, F. (2003). The formation of skeletal muscle: from somite to limb. *J. Anat.* 202: 59-68.

EWING, B., HILLIER, L., WENDL, M.C. and GREEN, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. *Genome Res.* 8(3): 175-185.

GORDON, D. ABAJIAN, C., and GREEN, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res.* 8(3): 195-202.

HAMBURGER, V. and HAMILTON, H.L. (1951). A series of normal stages in the development of the chick embryo. *J. Morphol.* 88: 49-92.

HATEY, F., TOSSER-KLOPP, G., MARTINATO, C.C., MULSANT, P. and GASSER, F. (1998). Expressed Sequence Tags for Genes: a review. *Genet. Sel. Evol.* 30: 521-554.

HUANG, X. and MADAN, A. (1999). Cap3: A DNA sequence assembly program. *Genome Res.* 9(9): 868-877.

JIA, L., YOUNG, M.F., POWELL, J., YANG, L., HO, N.C., HOTCHKISS, R., ROBEY, P.G. and FRANCOMANO, C.A. (2001). Gene expression profile of human bone marrow stromal cells: high-throughput expressed sequence tag sequencing analysis. *Genomics* 79: 7-17.

KO, M. (2001). Embryogenomics: developmental biology meets genomics. *Trends Biotechnol.* 19: 511-518.

LUDOLPH, D.C. and KONIECZNY, S.F. (1995). Transcription factor families: muscling in on the myogenic program. *FASEB J.* 9(12): 1595-1604.

POLESSKAYA, A. and RUDNICKI, M.A. (2002). A MyoD-dependent differentiation checkpoint: ensuring genome integrity. *Dev. Cell* 3: 1-2.

POURQUIÉ, O. (2001). Vertebrate Somitogenesis. *Annu. Rev. Cell Dev. Biol.* 17: 311-350.

RUDD, S. (2003). Expressed Sequence Tags: alternative or complement to whole genome sequences. *Trends Plant Sci* 8(7): 321-329.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. (1989). Molecular cloning: a laboratory manual. (ed. Plainview: Cold Spring Harbor Laboratory Press).

STOCKDALE, F.E., NIKOVITS, W.J. and CHRIST, B. (2000). Molecular and cellular biology of avian somite development. *Dev. Dyn.* 219: 304-321.

TIRUNAGARU, V.G., SOFER, L., CUI, J. and BURNSIDE, J. (2000). An expressed sequence tag database of T-cell enriched activated chicken splenocytes: sequence analysis of 5251 clones. *Genomics* 66: 144-151.

WALSH, K. and PERLMAN, H. (1997). Cell cycle exit upon myogenic differentiation. *Curr. Opin. Genet. Dev.* 7: 597-602.

Received: November 2003
Reviewed by Referees: December 2003
Modified by Authors and Accepted for Publication: March 2004