

**IDENTIFICAÇÃO DE SEQÜÊNCIAS EXPRESSAS (EST)
EM EMBRIÕES DE *Gallus gallus***

ERIKA CRISTINA JORGE

Dissertação apresentada à Escola Superior de
Agricultura Luiz de Queiroz, Universidade de
São Paulo, para a obtenção do título de
Mestre em Agronomia, Área de
concentração: Ciência Animal e Pastagens

PIRACICABA

Estado de São Paulo – Brasil

Outubro - 2002

**IDENTIFICAÇÃO DE SEQUÊNCIAS EXPRESSAS (EST)
EM EMBRIÕES DE *Gallus gallus***

ERIKA CRISTINA JORGE

Bióloga

Orientador: Prof. Dr. **LUIZ LEHMANN COUTINHO**

Dissertação apresentada à Escola Superior de
Agricultura Luiz de Queiroz, Universidade de
São Paulo, para a obtenção do título de
Mestre em Agronomia, Área de
concentração: Ciência Animal e Pastagens

PIRACICABA

Estado de São Paulo – Brasil

Outubro – 2002

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Jorge, Erika Cristina

Identificação de seqüências expressas (EST) em embriões de *Gallus gallus* / Erika Cristina Jorge. - - Piracicaba, 2002.

71 p. : il.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2002.

Bibliografia.

1. Biologia do desenvolvido 2. Biologia molecular 3. Embriogênes animal 4. Galos 5. Genes I. Título

CDD 636.5

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

AGRADECIMENTOS

Ao Prof. Dr. Luiz Lehmann Coutinho, agradeço pela orientação e ensinamentos recebidos e também por todas as valiosas oportunidades e confiança em meu trabalho.

À FAPESP, pelo apoio financeiro ao desenvolvimento deste trabalho.

Aos professores do Departamento de Produção Animal, pelos ensinamentos adquiridos.

À Dra. Cláudia Barros Monteiro Vitorello, agradeço pelos valiosos ensinamentos e especialmente pela amizade e apoio.

À Dra. Lúcia Alvares, pelos ensinamentos e apoio.

Às amigas Helena, Pilar, Clarissa, Raquel e Gisele, agradeço pela ajuda para o desenvolvimento deste trabalho.

À João Paulo Kitajima e Mateus, pela implantação da bioinformática no laboratório.

Aos meus pais e irmãos, pelo carinho, paciência e apoio.

À Miriam, pela amizade.

Ao pessoal do laboratório de Biotecnologia Animal, pela divertida convivência e amizade.

À todos aqueles que contribuíram de alguma forma para a realização deste trabalho.

Especialmente à *Alanis Morissette*, por ter transformado a minha vida para sempre...

SUMÁRIO

	Página
LISTA DE FIGURAS	vii
LISTA DE TABELAS	ix
RESUMO	x
SUMMARY	xii
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	4
2.1 Somitogênese	4
2.2 Formação da Musculatura Esquelética	7
2.3 Formação dos Membros	10
2.3.1 Localização da área do broto de membro	11
2.3.2 Indução do broto de membro	11
2.3.3 Determinação da identidade das asas e pernas	13
2.3.4 Crescimento do broto de membro	14
2.3.5 Diferenciação celular	17
2.4 Expressed Sequence Tags (EST)	18
2.4.1 EST e desenvolvimento embrionário	21
2.4.2 EST para animais domésticos	22
3 MATERIAL E MÉTODOS	24
3.1 Obtenção e coleta dos embriões	24
3.2 Bibliotecas de cDNA	28
3.2.1 Extração do RNA total	28
3.2.2 Seleção da população de RNA mensageiro	28
3.2.3 Síntese do cDNA dupla fita	29

3.2.3.1 Síntese da primeira fita	30
3.2.3.2 Síntese da segunda fita	31
3.2.3.3 Adição do adaptador SalI	32
3.2.3.4 Digestão com NotI	32
3.2.3.5 Coluna de fracionamento	33
3.2.4 Ligação do cDNA ao vetor pSPORT1	33
3.2.5 Transformação	34
3.2.5.1 Bactérias competentes	34
3.2.5.2 Choque térmico	34
3.3 Validação das bibliotecas.	35
3.4 Minipreparação do DNA plasmidial	35
3.5 Sequenciamento	36
3.5.1 Sequenciamento em ABI Prism 377 DNA Sequencer	36
3.5.2 Sequenciamento em Megabace	36
3.6 Análise das EST	37
3.6.1 Phred: análise de qualidade	38
3.6.2 Cap3: clusterização	38
3.6.3 Relatórios das análises	39
3.6.4 BLAST automático	39
3.6.5 Classificação das EST	39
3.6.6 Identificação dos SNPs (Single Nucleotide Polymorphisms).....	41
4 RESULTADOS E DISCUSSÃO.....	42
4.1 Seleção e coleta dos embriões	42
4.2 Bibliotecas de cDNA	43
4.2.1 RNA total e seleção do mRNA	43
4.2.2 Validação das bibliotecas de cDNA	44
4.3 Análise dos dados	46
4.3.1 Clusterização	46
4.3.2 EST biblioteca-específicas	49

4.3.3 Índices de novidade	51
4.3.4 Classificação das EST	52
4.3.5 Categoria dos No hits	54
4.3.6 Genes relacionados ao desenvolvimento embrionário	56
4.3.7 Análise comparativa	61
4.3.8 Frequência de cDNA completos	61
4.3.9 SNPs (Single Nucleotide Polymorphisms)	61
5 CONCLUSÕES	63
REFERÊNCIAS BIBLIOGRÁFICAS	65

LISTA DE FIGURAS

	Página
1 Esquema representativo dos três tecidos embrionários formados a partir dos somitos: esclerótomo, miótomo e dermomiótomo (porções epaxial e hipoaxial).	6
2 Esquema representativo do desenvolvimento dos brotos de membros.	12
3 Esquema representativo do broto de membro em desenvolvimento.	16
4 Representação da obtenção de EST a partir de cDNA.	19
5 Fotografia de um embrião de <i>Gallus gallus</i> no estágio E15 de desenvolvimento onde foi representada, entre as linhas pretas, a região de somitos associada ao tubo neural microdissecada dos embriões.	25
6 Fotografias dos embriões de <i>Gallus gallus</i> nos estádios E21, E24 e E26 selecionados para a micordissecação dos brotos de membros anteriores e posteriores, retiradas do trabalho publicado por Hamburger & Hamilton (1951), com destaque para os diferentes tamanhos de brotos de membros obtidos em cada estágio.	26
7 Fotografias dos embriões de <i>Gallus gallus</i> utilizados para a	

microdissecação dos brotos de membros nos estádios E21, E24 e E26 do desenvolvimento embrionário.	27
8 Esquema representativo da clonagem direcionada adaptado do protocolo do kit SuperScript Plasmid System.	30
9 Fotografia da eletroforese dos <i>pools</i> de RNA total extraídos a partir do embrião inteiro e dos tecidos microdissecados de <i>Gallus gallus</i>	44
10 Fotografia da eletroforese do mRNA selecionado para estimativas de qualidade e quantidade de material.	44
11 Produtos amplificados na PCR de colônia para a verificação da presença e tamanho dos insertos clonados. Fotografia do gel de agarose 1%.	45
12 Análise da clusterização das bibliotecas de cDNA. Distribuição do número de EST em função de sua representatividade nos <i>clusters</i>	49
13 Categorização dos clusters e singletons identificados nas bibliotecas CEMB, CMEM e CSOM. Distribuição das frequências nas doze categorias estabelecidas para a classificação das EST de <i>Gallus gallus</i> deste estudo.	53

LISTA DE TABELAS

	Página
1 Estadiamento dos embriões de <i>Gallus gallus</i>	27
2 Termos utilizados em análise de bioinformática	37
3 Quantidade de RNA total utilizada por biblioteca	43
4 Número de clones seqüenciados e número de clones válidos por biblioteca de cDNA após análise de qualidade (Phred)	46
5 Resultados da clusterização da EST de <i>Gallus gallus</i> via Cap3	47
6 Índices de <i>clusters</i> e <i>singletons</i> biblioteca-específicos, contribuições relativa e específica e novidade das bibliotecas de <i>Gallus gallus</i>	51
7 Nova classificação dos <i>clusters</i> e <i>singletons</i> após análise contra o banco dbEST humano do <i>GenBank</i>	55

IDENTIFICAÇÃO DE SEQUÊNCIAS EXPRESSAS (EST) EM EMBRIÕES DE *Gallus gallus*

Autora: ERIKA CRISTINA JORGE

Orientador: Prof. Dr. LUIZ LEHMANN COUTINHO

RESUMO

O desenvolvimento embrionário requer um rigoroso controle da expressão de inúmeros genes que devem ser ativados no momento e local apropriados para o correto estabelecimento das estruturas e órgãos do organismo. Com a análise das sequências expressas (*Expressed Sequence Tags*, EST) é possível identificar os genes expressos em tecidos específicos em diferentes estádios do desenvolvimento. A fim de identificar genes expressos durante o desenvolvimento embrionário de *Gallus gallus*, EST foram obtidas a partir da extremidade 5' dos clones de três bibliotecas de cDNA construídas a partir de (1) embrião inteiro, (2) membros anteriores e posteriores e (3) somitos e tubo neural. As EST foram analisadas pelos programas Phred, Cap3 e Consed para avaliação de qualidade das bases e clusterização. A análise dos dados resultou em 4998 EST válidas segundo parâmetros estabelecidos para este estudo. Todas as sequências consenso dos *clusters* e *singletons* foram comparadas com as sequências disponíveis no *GenBank* (<http://www.ncbi.nlm.nih.gov>) e classificadas em doze categorias segundo sua função. A categorização revelou que 25% dos genes não apresentaram homólogos no banco utilizado (*Low* e *No hit*). Aproximadamente 15% dos genes não apresentaram função definida (hipotéticos conservados). Os 60% restantes permitiram identificar genes envolvidos com a somitogênese, com a formação da musculatura esquelética e

com a formação dos brotos de membros em vertebrados, além de genes de manutenção e estrutura celular. O conjunto das EST identificadas neste projeto possibilitou a construção de um banco com cerca de 5.000 EST de estádios embrionários de *Gallus gallus*, fonte para a identificação de novos genes, para a elucidação dos processos reguladores do desenvolvimento embrionário e para a identificação de SNPs (*Single Nucleotide Polymorphism*).

STUDY OF EXPRESSED SEQUENCE TAGS (EST) DURING CHICKEN (*Gallus gallus*) EMBRYONIC DEVELOPMENT

Author: ERIKA CRISTINA JORGE

Adviser: Prof. LUIZ LEHMANN COUTINHO

SUMMARY

Embryo development requires rigorous control of expression of various genes, which may become active at an adequate time and site for the correct establishment of organ and structure of the organism. Identification of Expressed Sequence Tags (EST) allows the determination of genes expressed in specific tissues at various stages of development. To identify genes expressed during embryonic development of *Gallus gallus*, EST were obtained from 5' end of clones from three cDNA libraries, derived from (1) whole embryos, (2) limb buds (hindlimb and forelimb), and (3) somites and neural tube. EST sequences were analyzed using the softwares Phred, Cap3 and Consed to evaluate base quality and clustering, resulting in 4998 valid EST, according to the parameters established. All consensus sequences of clusters and singletons obtained were compared with sequences available at GenBank (<http://www.ncbi.nlm.nih.gov>) and classified into twelve categories according to function. Categorization revealed that about 25% of the genes were not described in the database consulted (Low and No hit). About 15% of the genes had no defined function (conserved hypothetical). The remaining sixty percent permitted identification of genes involved with somitogenesis, skeletal muscle and limb bud development in vertebrates and cellular maintenance and

structure. The set of EST identified in this project resulted in a database with around 5000 EST of *Gallus gallus* embryonic stages, which is a new source of gene identification to facilitate investigation of regulatory processes in embryo development and identification of Single Nucleotide Polymorphisms (SNPs).

1 INTRODUÇÃO

A especificação das diferentes estruturas e órgãos do organismo durante a embriogênese requer um rigoroso controle da expressão de inúmeros genes. Genes com função regulatória e estrutural devem atuar no momento e local apropriados, proporcionando as condições para o correto estabelecimento das diferentes linhagens celulares do organismo. A determinação e a subsequente diferenciação destas linhagens precursoras dependem de uma intrincada rede de sinalização que tem início extracelularmente, levando à ação coordenada de múltiplos genes e à ativação de uma cascata de sinalização intracelular (Ludolph & Konieczny, 1995).

A base molecular que controla a embriogênese tem sido objeto de muito interesse em biologia do desenvolvimento. A identificação do conjunto dos genes expressos e das vias sinalizadoras reguladoras envolvidas com a formação de estruturas complexas permitirá gerar uma visão integrada do desenvolvimento embrionário. Um progresso significativo tem ocorrido recentemente com os estudos realizados em dois tecidos embrionários: os somitos e os brotos de membro anteriores e posteriores em desenvolvimento.

Somitos são estruturas homólogas dispostas lado a lado ao longo do tubo neural do embrião (Ludolph & Konieczny, 1995; Nowicki & Burke, 2000). Cada somito é morfológicamente idêntico a outro ao longo do eixo antero-posterior (AP), e eventualmente origina as mesmas estruturas em ambos os lados do embrião: musculatura esquelética, cartilagem e ossos de todo o organismo. Durante o desenvolvimento, estas estruturas iniciam um complexo processo culminando na subdivisão dos blocos em camadas embrionárias (esclerótomo, dermomiótomo e miótomo) precursoras dos derivados somáticos. Portanto, investir na identificação dos genes expressos nesta região

e nesta fase do desenvolvimento embrionário irá contribuir para a elucidação dos mecanismos moleculares iniciais que regulam a formação desses tecidos especializados que colonizam todo o corpo do organismo.

Células precursoras originárias especificamente da região do dermomiótomo migram para colonizar a região dos quatro flancos embrionários que suportam os membros anteriores e posteriores do organismo. Estas células estabelecem-se nestas regiões e iniciam um programa molecular que resulta na formação dos diferentes tecidos especializados que compõem a região dos membros, essencialmente músculos, ossos, cartilagem, vasos sanguíneos e todo o sistema nervoso que predomina nesta área (Capdevila & Belmonte, 2001). Identificar e compreender como este programa se estabelece abrirá novos caminhos para o conhecimento do padrão de desenvolvimento nesta região, além de contribuir com informações genéticas de estádios mais avançados (musculatura esquelética, cartilagem e ossos), informações adicionais àquelas identificadas nos estádios iniciais do desenvolvimento com a análise dos somitos.

Esses dois tecidos embrionários foram selecionados para este estudo com o objetivo principal de identificar a coleção dos genes expressos nos estádios iniciais do desenvolvimento embrionário de *Gallus gallus*. Diferentes metodologias podem ser empregadas para a identificação de genes. Para este estudo optou-se pela análise das seqüências expressas (*Expressed Sequence Tags*, EST), que correspondem a seqüências parciais dos RNA mensageiros expressos em uma célula ou tecido. Com esta metodologia é possível identificar os genes expressos e inferir sobre os seus níveis de expressão, em função da quantidade de EST seqüenciadas de um mesmo mRNA (Hatey et al., 1998). Assim, visando a compreensão dos mecanismos reguladores do desenvolvimento, EST de estádios embrionários diferentes de *Gallus gallus* foram analisadas na tentativa de contribuir com um conjunto de informações genéticas para futuras pesquisas de expressão gênica, para a identificação de novos genes, especialmente aqueles envolvidos com processos específicos do desenvolvimento (somitogênese, formação de musculatura esquelética e formação dos brotos de

membros) e para a identificação de novos marcadores moleculares para programas de melhoramento genético animal.

Especificamente, os objetivos deste estudo foram:

1. Identificar e analisar o conjunto dos genes expressos nos somitos associados ao tubo neural de embriões de *Gallus gallus* no estágio E15 de desenvolvimento (Hamburger & Hamilton, 1951), a partir da construção de uma biblioteca direcionada de DNA complementar (cDNA) e análise de EST da extremidade 5' dos insertos.
2. Identificar e analisar o conjunto dos genes expressos nos brotos de membros anteriores e posteriores de embriões de *Gallus gallus* nos estádios E21, E24 e E26 de desenvolvimento (Hamburger & Hamilton, 1951), a partir da construção de uma biblioteca direcionada de DNA complementar (cDNA) e análise de EST da extremidade 5' dos insertos.

2 REVISÃO DE LITERATURA

A identificação dos mecanismos moleculares que determinam o correto estabelecimento das diferentes linhagens celulares precursoras das estruturas e órgãos do organismo é um dos maiores desafios da biologia do desenvolvimento. O maior interesse é gerar uma visão global do conjunto dos genes expressos e das vias sinalizadoras que se integram no controle do desenvolvimento embrionário. Diferentes tecidos embrionários vêm sendo empregados como modelos nesta área da biologia, entre eles, os somitos e os brotos de membros anteriores e posteriores.

2.1 Somitogênese

Os somitos são blocos de células mesenquimais formados no mesoderme paraxial (MP) dispostos lateralmente às estruturas axiais do embrião, o tubo neural e a notocorda (Ludolph & Konieczny, 1995). A disposição segmentada do MP estabelece unidades repetidas com identidades equivalentes ao longo do embrião, o que é essencial para o correto desenvolvimento dos derivados somíticos: músculos, ossos e cartilagem do organismo.

Durante a fase da gastrulação, que corresponde aos estádios 4 a 15 em galinha (Hamburger & Hamilton, 1951), células mesenquimais são continuamente depositadas à extremidade posterior do MP. A segmentação do MP ocorre continuamente na direção crânio-caudal do embrião. Vários modelos foram desenvolvidos para explicar a periodicidade da segmentação do MP durante o desenvolvimento embrionário. Um dos modelos mais aceitos é a presença de um relógio molecular que traduz a expressão periódica de alguns genes no tempo correspondente à formação dos somitos. Os fatores

de transcrição bHLH *c-hairy1*, *c-hairy2* e *Lunatic Fringe* foram identificados no MP e apresentam ondas cíclicas de expressão a cada 90 minutos, cada onda correspondendo à formação de um somito (Palmeirim et al., 1997). Essas ondas de expressão gênica permitem que a somitogênese ocorra em intervalos sucessivos e regulares, uniformizando o tamanho dos blocos celulares (Stockdale et al., 2000; Pourquié, 2001).

Esses genes com padrão oscilatório no MP são regulados por um mecanismo pós-transcricional que envolve a participação da via de sinalização Notch, um receptor transmembrana capaz de reconhecer dois ligantes: Delta e Serrate. A ligação faz com que Notch sofra uma clivagem proteolítica para que o domínio liberado, associado ao fator de transcrição Su, ative a transcrição de genes com padrão cíclico de expressão. A função do relógio molecular é portanto, regular a ativação de Notch no MP, o que resulta na especificação de células com propriedades de limites posterior e anterior dos somitos (Tsakonas et al. 1999; Delfini et al., 2000; Jiang et al., 2000).

A segmentação do MP em compartimentos celulares definidos ainda é dependente de alterações progressivas na organização celular que determinam a epitelialização do tecido. A base genética que determina essa transição mesênquima-epitélio no MP foi identificada em experimentos com *knockout* do gene que determina o fator de transcrição bHLH *Paraxis* (Burgess et al., 1996). Suposições indiretas também indicam uma participação da via de sinalização *Wnt/β-catenin* na determinação da disposição epitelial do tecido somítico (Stockdale et al., 2000). A especificação genética dos limites dos somitos é dependente da via de sinalização Notch, além da expressão de várias proteínas de superfície celular, como a família de receptores cinase de tirosina Eph e seus ligantes *ephrins* (Bergmann et al., 1995), e as *N-cadherins*, proteínas responsáveis pela formação da lâmina basal e manutenção dos limites dos somitos (Pourquié, 2001).

Após cinco horas da delimitação dos blocos celulares, os somitos iniciam um processo de compartimentalização interna ao longo dos três principais eixos do embrião: dorso-ventral (DV), próximo-distal (PD) e ântero-posterior (AP). Embora esse padrão de compartimentalização dos somitos não seja discernível morfológicamente, padrões específicos de expressão gênica podem ser identificados ao longo desses eixos, e cada

um desses domínios determina o desenvolvimento de um tecido embrionário: esclerótomo, dermomiótomo e o miótomo.

A primeira subdivisão aparente no somito é a DV. A porção ventral dos somitos sofre uma transição epitélio-mesênquima (de-epitelialização) e origina o esclerótomo, que é responsável pela formação da cartilagem e ossos do corpo. Enquanto que a porção dorsal do somito, chamada dermomiótomo, mantém sua natureza epitelial e é fonte de células que irão originar musculatura esquelética, derme e derivados vasculares (Stockdale et al., 2000). (Figura 1)

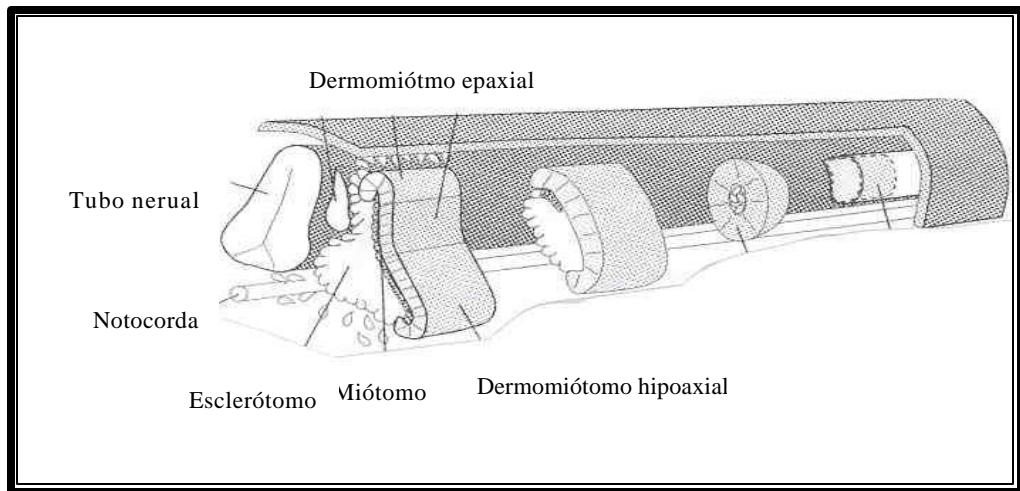


Figura 1 - Esquema representativo dos três tecidos embrionários formados a partir dos somitos: esclerótomo, miótomo e dermomiótomo (porções epaxial e hipoaxial). Figura publicada por Buckingham (2001).

O dermomiótomo tem origem na porção dorsal dos somitos. A porção média do dermomiótomo origina o miótomo, um terceiro tecido embrionário responsável pela formação da musculatura epaxial do corpo (musculatura intercostal e vários músculos da parede do corpo). Uma população de células da porção lateral do dermomiótomo delamina e migra para colonizar outras regiões do corpo, sendo precursores da musculatura hipoaxial (musculatura esquelética dos membros, parede abdominal, tórax e língua) (Stockdale et al., 2000).

As células do miótomo derivam do dermomiótomo em resposta à sinais provenientes de estruturas axiais (tubo neural e notocorda), do mesoderme intermediário e do MP, sob o controle de Wnt, Shh e membros da família de fatores de crescimento *Transforming Growth Factors* (TGF- β), especialmente as *Bone Morphogenic Proteins* (BMP). As células do miótomo caracterizam-se por serem mononucleadas e não proliferativas, já expressando proteínas músculo específicas (Goulding et al., 1994).

A porção lateral do dermomiótomo compromete-se com a formação de duas populações distintas de células precursoras miogênicas: (1) aquela que permanece na região do dermomiótomo e (2) outra que delamina da porção lateral e migra para colonizar outras regiões do embrião. Os precursores não-migratórios associam-se às células do miótomo para formar a musculatura epaxial do corpo. E os precursores migratórios originam a musculatura hipaxial. Embora destinadas a formarem músculos, as células migratórias não expressam fatores miogênicos ou proteínas contráteis enquanto não atingem seus sítios alvos (Birchmeier & Brohmann, 2000).

O fator de transcrição Pax 3 (*Paired type homeodomain*) é necessário para o correto estabelecimento da população migratória de células que delamina da porção lateral do dermomiótomo. Isto foi observado em ratos que apresentam mutações espontâneas nesse gene. Não havendo precursores migratórios, esses ratos apresentam desenvolvimento anormal, com ausência da musculatura dos membros e do diafragma. A expressão de c-Met e Lbx1, outros dois genes importantes no desenvolvimento dos precursores migratórios, também está comprometida nesses ratos mutantes. c-Met é um receptor kinase de tirosina e, associado a *scatter factor/hepatocyte growth factor* (SF/HGF) promove a delaminação da população migratória no dermomiótomo (Dietrich et al., 1999; Scallan et al., 1999). Lbx1 é um fator de transcrição (*homeobox*), também restrito à população migratória (Birchmeier & Brohmann, 2000).

2.2 Formação da musculatura esquelética

O processo de formação da musculatura esquelética é um modelo de desenvolvimento, que envolve a (1) especificação dos precursores nos tecidos embrionários, a (2) ativação da proliferação celular, para a determinação do número de

células musculares do organismo, e finalmente, a (3) diferenciação e (4) maturação dos mioblastos em fibras musculares maduras. Este é um dos modelos mais estudados em desenvolvimento e por isso, vários genes foram identificados dentro dessa rede molecular, determinando que o programa miogênico seja ativado no momento e local apropriados (Arnold & Braun, 2000).

A transcrição de genes de determinação muscular é dependente da ação de uma família de fatores conhecida como reguladores miogênicos (*myogenic regulatory factors*, MRF). Os MRF são fatores de transcrição que apresentam dois domínios funcionais: um domínio hélice-alça-hélice (bHLH) e uma região básica, responsável pela interação ao DNA. Esses fatores ativam a transcrição de genes alvos ligando-se a uma região específica do DNA, o *E-box* (CANNTG), presente nos *enhancers* de genes músculo-específicos. A ligação eficiente ao DNA é obtida por heterodimerização desses fatores com proteínas não miogênicas E2A e membros da família *myocyte enhancer factor 2* (MEF2). Juntos, esses fatores ativam o programa de diferenciação muscular, por indução da transcrição de genes reguladores e estruturais músculo-específicos (Puri & Sartorelli, 2000).

A família de MRFs é composta por quatro genes ativadores de transcrição: MyoD, Myf5, MRF4 e Miogenina. Em geral, esses quatro fatores miogênicos desempenham funções distintas, mas com alguma redundância. MyoD e Myf5 atuam na determinação e proliferação dos mioblastos indiferenciados, enquanto que a miogenina e MRF4 participam da diferenciação dos mioblastos em miotubos (Ludolph & Konieczny, 1995; Dauncey et al., 1996). A existência de várias vias para ativar miogênese durante o desenvolvimento pode ser considerada um mecanismo de proteção contra eventuais falhas em acionar o programa miogênico.

A expressão temporal e espacial dos fatores miogênicos deve ser precisamente regulada por proteínas que atuam tanto positiva quanto negativamente, por associação física ou por competição pela ligação ao DNA. Essas proteínas reguladoras são essenciais, pois comandam o correto estabelecimento da musculatura esquelética no corpo. Diversos fatores foram identificados como reguladores negativos da miogênese. A proteína Id, por exemplo, foi identificada como bloqueadora da heterodimerização dos

fatores miogênicos com as proteínas E2A, impedindo a ligação ao *E-box* no DNA, prevenindo portanto, a ativação do programa de diferenciação dos mioblastos. Da mesma forma atua a proteína Twist no mesoderme e nos somitos (Puri & Sartorelli, 2000).

Proteínas com propriedades oncogênicas também antagonizam o programa miogênico. Proteínas como essas identificadas nos mioblastos pertencem a família Ras e c-Jun, que em associação ao fator de transcrição AP-1, impedem a expressão de MyoD nos mioblastos proliferativos (Bengal et al., 1992). Recentemente Delfini et al. (2000) identificaram a participação do receptor transmembrana Notch na regulação negativa da miogênese, inibindo a expressão de MyoD e a diferenciação muscular em culturas de células miogênicas.

Os conhecidos membros das famílias de fatores de crescimento *Fibroblast Growth Factor* (FGF) e TGF- β exercem uma importante função na miogênese, ativando vias de sinalização implicadas no controle da atividade dos fatores miogênicos. Os FGFs, por exemplo, associados à seus receptores transmembrana (FGFR1 e FGFR4), desempenham um forte efeito inibitório sobre a diferenciação terminal dos mioblastos via fosforilação de MyoD. A inativação de MyoD via fosforilação é essencial para que os mioblastos permaneçam no estado proliferativo (Li et al., 1992).

Os Serum response factors (SRF) são fatores de transcrição MADS-box que atuam positivamente sobre a diferenciação miogênica. Esses MADS-box são motivos de elementos cis regulatórios de proteínas musculares que desempenham papel fundamental na ativação de genes músculo-específicos, como a actina (Croissant et al., 1996). Outros elementos funcionalmente relevantes na ativação de promotores miogênicos são os receptores nucleares para hormônios tireoideanos (TH) e ácido retinóico (RAR). Especificamente, o domínio de ligação ao DNA do receptor de TH também atua como domínio para a ligação de MEF2 (Puri & Sartorelli, 2000). Além de atuarem como fatores de determinação miogênica, MyoD e Myf5 estão potencialmente envolvidos na regulação do ciclo celular e remodelamento da cromatina. Durante a diferenciação miogênica, a atividade das proteínas reguladoras do ciclo celular, as cyclin dependent kinases (cdks), é bloqueada, promovendo a estabilização e a ativação dos fatores

miogênicos. Assim estabelece-se uma forte correlação entre a interrupção do ciclo celular e a ativação de um loop de regulação positiva assegurando que MyoD seja ativado durante a diferenciação muscular (Walsh & Perlman, 1997).

Esses mesmos fatores também podem remodelar a cromatina por domínios protéicos distintos daqueles envolvidos com a ativação transcricional miogênica. MyoD interage com SWI/SNF, subunidades da ATPase remodeladora de cromatina, promovendo a conversão miogênica dos fibroblastos. Também parece interagir com histonas transacetilases PCAF e CBP/p300, que atuam como co-ativadores transcpcionais. Essa acetilação de MyoD aumenta a sua afinidade por promotores músculo-específicos, promovendo a diferenciação muscular (Buckingham, 2001).

2.3 Formação dos membros

A formação dos membros dos vertebrados é um excelente modelo experimental para a determinação dos mecanismos moleculares e celulares envolvidos na regulação do crescimento e na padronização do desenvolvimento embrionário. Nos últimos anos, estudos clássicos em embriologia associados à ensaios de expressão ectópica e *knockout* de genes nos membros, especialmente de ratos e galinhas, têm contribuído para a identificação dessas vias sinalizadoras que determinam a correta formação de todos os tecidos no local e momento apropriados, e o seu crescimento durante o desenvolvimento.

Um mecanismo básico controla o desenvolvimento dos membros com morfologia e função diferenciados dentro de todos os vertebrados. Resumidamente, a formação do membro inicia-se com a expressão de fatores que determinam posicionamento específico do primórdio ao longo do embrião. Em seguida, são estabelecidas interações moleculares entre células do mesênquima e da ectoderme superficial, que acabam por determinar o crescimento coordenado deste primórdio. Os progenitores mesenquimais originam os elementos ósseos, tendões e outros tecidos conectivos do membro maduro e a ectoderme superficial origina a pele e seus apêndices, como escamas, pêlos ou penas (Capdevila & Belmonte, 2001; Johnson & Tabin, 2001; Martin, 2001).

2.3.1 Localização da área do broto do membro.

O primeiro passo no desenvolvimento do membro é a determinação de um grupo de células embrionárias que irão originar o primórdio. Os primórdios dos dois pares de membros originam em quatro áreas específicas nos flancos embrionários, onde grupos de células da mesoderme lateral formam pequenos brotos de células mesenquimais envolvidas por uma ectoderme.

A família de genes Hox codificam para fatores de transcrição *homeodomain* que fornecem a determinação espacial do desenvolvimento de muitas estruturas embrionárias nos vertebrados, incluindo a alocação dos brotos de membro. O primórdio do membro é induzido na região do flanco embrionário em posição específica que contém uma certa combinação de expressão de genes da família Hox. A expressão combinada de Hoxc6, Hoxc8 e Hoxb5, por exemplo, ocorre exatamente na área dos membros anteriores, sugerindo o papel desses genes na especificação dessa estrutura no embrião (Nelson et al., 1996; Nowicki & Burke, 1999).

A sobreposição da expressão dos diferentes integrantes da família de genes Hox é controlada por uma variedade de fatores que incluem reguladores de transcrição, receptores de ácido retinóico (RAR), o gene Krox20, membros da família de cofatores Pbx, além de membros das famílias FGF e TGF- β (Capdevila & Belmonte, 2001). Os mecanismos que controlam a expressão dos genes Hox no embrião não são inteiramente conhecidos, mas provavelmente a interação dos diferentes membros dessa família de genes com uma variedade de reguladores transcricionais devem estabelecer um padrão no eixo embrionário que contribui para alocar o primórdio do membro nos vertebrados.

2.3.2. Indução do broto do membro.

Após a determinação da área de formação dos membros anteriores e posteriores em áreas específicas nos flancos embrionários, células da placa lateral comprometem-se com uma ativa divisão celular, enquanto que células da região entre os membros dividem-se mais lentamente. Essa proliferação diferencial resulta no desenvolvimento de um primórdio ou broto de membro, que consiste de uma massa de células mesenquimais envolvidas por uma ectoderme.

Os tecidos adjacentes (somitos, mesoderme intermediária-MI e mesoderme da placa lateral-MPL) fornecem uma informação posicional que deve ser interpretada pelas células mesodérmicas. Os FGF são os candidatos mais prováveis à indutores de membro nos flancos embrionários. São fatores de crescimento com alta afinidade com receptores tirosina-kinase e que são conhecidos por exercer diferentes funções durante o desenvolvimento embrionário, incluindo a formação do broto do membro (Martin, 2001; Tickle & Münsterberg, 2001). Fgf-8 é o mais aceito como fator indutor de membros. Ele é expresso no MI e mantém as células dos flancos embrionários em estado proliferativo, além de sinalizar para as células do MPL ativarem a expressão de Fgf-10 nessa região. Fgf-10 é portanto o fator mediador do efeito indutivo de Fgf-8 (MI) no MPL (Capdevila & Belmonte, 2001). (Figura 2)

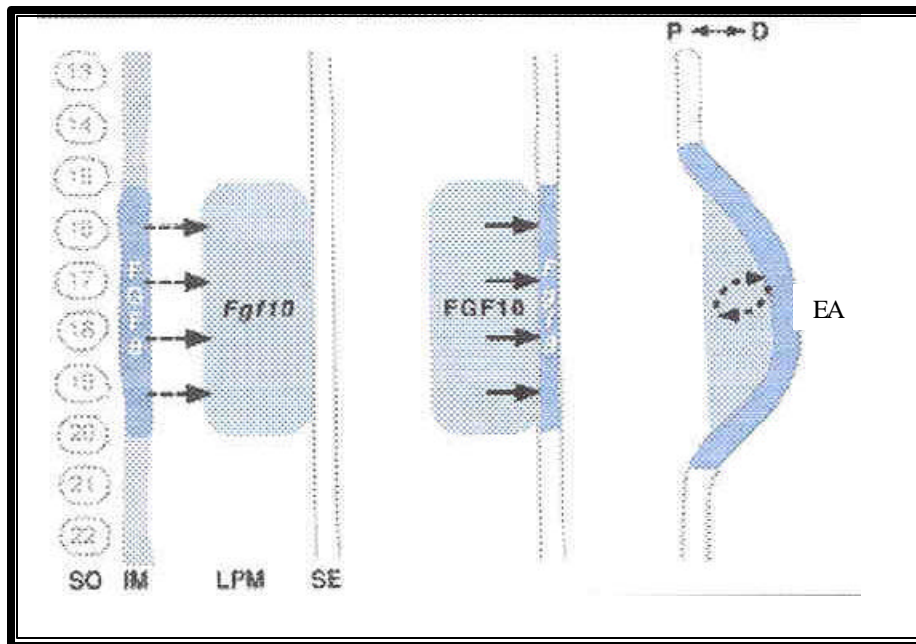


Figura 2 - Esquema representativo do desenvolvimento dos brotos de membros retirado do trabalho publicado por Martin (2001). (SO) correspondem aos somitos; (IM) mesoderme intermediária; (LPM) mesoderme placa lateral; (SE) e (EA) ectoderme superficial.

Outros genes devem atuar como mediadores dos efeitos indutivos de Fgf-8 (MI) e Fgf-10 (MPL). Estudos recentes indicam a participação de uma outra família de fatores de crescimento conhecida como Wnts. A via de sinalização intracelular ativada por Wnts envolve a estabilização de β -*catenin*, um fator de transcrição para genes Wnts. Kawakami et al. (2001) demonstraram a participação de β -*catenin* na regulação da expressão de Fgf-10 durante a indução do membro, mantendo sua expressão restrita à área do MPL dos membros. Dois genes diferentes da família Wnt são indicados como candidatos a indutores: Wnt2B para membros anteriores e Wnt8C para posteriores. Assim, a via Wnts/ β -*catenin* media o *loop* Fgf-8/Fgf-10 que controla a iniciação do membro, e a expressão localizada de Fgf-10 no MPL parece ser fator chave para a indução dos membros (Johnson & Tabin, 1997; Dudley & Tabin, 2000).

2.3.3. Determinação da identidade das asas e pernas

O processo de indução resulta na formação de quatro brotos de membro em flancos específicos do embrião: dois membros anteriores e dois posteriores. Apesar de algumas controvérsias, os membros anteriores e posteriores são considerados estruturas homólogas, o que implica em aceitar que os mecanismos moleculares para a formação dessas duas estruturas sejam basicamente as mesmas (Gilson-Brown, 1998). De fato, a maioria dos genes apresenta o mesmo padrão de expressão nas duas estruturas, o que resulta em padrões similares de formação da estrutura óssea e outras características morfológicas. Porém, algumas diferenças moleculares específicas devem existir a fim de justificar as diferenças morfológicas e funcionais observadas entre esses dois apêndices.

Recentemente, vários genes vêm sendo identificados com expressão exclusiva em membros anteriores ou posteriores, incluindo dois genes da família T-box, Tbx-4, restrito à área do membro anterior e Tbx-5, da posterior (Rodriguez-Esteban et al., 1999; Takeuchi et al., 1999). Outros genes exclusivos incluem membros da família Hox, Hoxc4 e Hoxc5, restritos de membros anteriores, e um outro fator de transcrição, Ptx-1, da classe de proteínas *paired-type homeodomain*, exclusivamente expresso em membros posteriores. Apesar desses avanços recentes, outros genes devem estar envolvidos na

identidade dos membros anteriores e posteriores, e novos estudos são necessários para se determinar o completo mecanismo que controla esse processo.

2.3.4. Crescimento do broto do membro

Após o broto de membro ter sido iniciado nos quatro flancos específicos do embrião, correspondentes aos membros anteriores e posteriores, esse primórdio inicia um processo de multiplicação celular que resulta em seu crescimento nos três eixos: antero-posterior (AP), dorso-ventral (DV) e próximo-distal (PD). Neste momento, algumas regiões específicas são identificadas no broto de membro e as interações moleculares entre elas são responsáveis por seu desenvolvimento. A zona de atividade polarizada (ZAP) padroniza o eixo AP e a ectoderme apical (EA) determina a identidade DV e o crescimento do membro, mantendo as células de mesênquima da ZP (zona de progresso) no estado indiferenciado e proliferativo (Vargesson et al., 1997; Johnson & Tabin, 1997; Dudley & Tabin, 2000; Altabef et al., 1997). Estas áreas podem ser identificadas no esquema representativo da Figura 3.

A EA é uma estrutura epitelial especializada localizada no limite DV do broto de membro e torna-se morfológicamente definida próximo ao estágio 18 em galinhas (Hamburguer & Hamilton, 1951). Além de determinar a polaridade DV no membro, sua integridade é essencial para manter as células da ZP em estado proliferativo após o processo de iniciação do membro, determinando assim, o seu crescimento (Altabef et al., 1997).

Nos últimos anos, vários fatores foram identificados como mediadores da função da EA no crescimento dos membros, incluindo os genes da família de FGFs (Fgf-2, Fgf-4 e Fgf-8), além das vias de sinalização Notch e Wnt (TAVARES et al., 2000). Assim que a expressão é consolidada no MPL durante a indução do membro, Fgf-10 sinaliza para a ectoderme superficial iniciar um programa de expressão gênica, o que inclui a expressão de Fgf-8. Na verdade, há evidências de que Wnt3a interfira nesse *loop* Fgf-10 (MPL) e Fgf-8 (EA). Fgf-10 induz a expressão de Wnt3a na ectoderme e, via β -catenin, ativa a expressão de Fgf-8 (Tickle et al., 2001; Martin, 2001 e Capdevila & Belmonte, 2001). São as interações moleculares entre MPL e EA, reguladas pelo *loop* Fgf-10/Fgf-8

que determinam a contínua proliferação das células da ZP e o conseqüente crescimento do broto do membro.

O posicionamento da EA na região exata da ectoderme superficial ainda é motivo de especulações. Sabe-se que o posicionamento está relacionado com a geração da polaridade dorso-ventral (DV) do membro. As diferenças ao longo desse eixo são evidentes em muitos apêndices de vertebrados, como por exemplo, a palma e as costas das mãos dos humanos, os músculos, tendões e outras estruturas que mostram uma disposição ordenada ao longo desse eixo. A divisão do membro em compartimentos celulares com domínios de expressão diferencial DV, determina que a EA posicione-se na interface exata entre as células da porção dorsal, que expressam *Rfng*, *Wnt7a* e *LMX-1*, e as da porção ventral, domínio da expressão de *EN-1*, fator que previne a expressão de *Rfng* e *Wnt7a* (Capdevila & Belmonte, 2001).

Um grupo de células localizadas no mesênquima posterior do broto do membro é chamado de zona de atividade polarizada (ZAP) e atua como organizador da polaridade AP deste apêndice. Um dos únicos fatores conhecidos capazes de mediar essa atividade polarizadora da ZAP é o *Shh* (Sonic Hedgehog) (Laufer et al., 1994). A atividade deste fator é essencial para a manutenção do crescimento e padronização de estruturas distais e intermediárias do membro. Uma complicada rede de interações genéticas restringe a expressão de *Shh* na ZAP, incluindo genes da família *Hox*, o fator de transcrição *dHAND* (bHLH) e a via de sinalização do ácido retinóico, além de moduladores extracelulares e modificadores pós-traducionais que regulam a disponibilidade extracelular e o limite de ação da proteína *Shh* (Capdevila & Belmonte, 2001).

Apesar disso, a atividade do *Shh* não está restrita à ZAP. Este fator também está envolvido na manutenção do EA, o que ilustra a importância das interações epitélio-mesênquima durante o desenvolvimento do membro. A EA mantém as células do mesênquima da área distal do membro, conhecida como zona de progresso (ZP), em estado proliferativo e indiferenciado. As células da ZP originam a maioria das estruturas mesenquimais do membro. Conforme o membro cresce, as células deixam a ZP, movem-se proximamente e adquirem informação posicional para formar o apêndice maduro, com os padrões anatômicos de ossos, músculos e nervos. *Shh*, provavelmente

associado a Fgf-4 expresso na EA, comandam um *loop* regulador que mantém o crescimento celular e a proliferação das células do mesênquima e mantém a integridade da EA. Portanto, sinais derivados da EA são necessários para manter o crescimento do membro, e sinais derivados do mesênquima são necessários para manter a EA (Laufer et al., 1994).

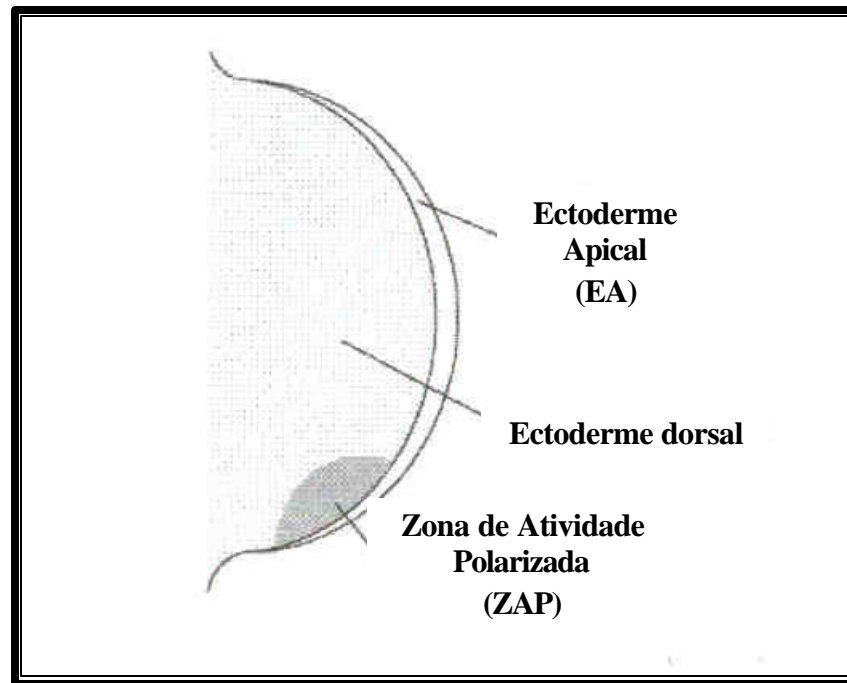


Figura 3 - Esquema representativo de um broto de membro em desenvolvimento retirado do trabalho publicado por Tickle (1999). Nesta representação é possível identificar a região da ectoderme apical (EA) e da zona de atividade polarizada (ZAP) em um broto de membro em desenvolvimento.

Os produtos dos genes Formin e seu antagonista Gremlin são necessários para o estabelecimento do *loop* Shh/Fgf-4. Mutantes para Formin apresentam membros mal formados, EA não organizado e diminuição da expressão de Shh no mesênquima do membro (Zeller et al., 1999). Esse *feedback* Shh/Fgf-4 pode ainda ser regulado por degradação protéica (Sidow et al., 1999). Foi identificado uma proteína denominada Dactylin, membro da família gênica F-box/WD40, que codifica para moléculas

adaptadoras para proteínas que serão destruídas. Mutantes para esse gene também determinam animais com membros mal formados e problemas com a manutenção da EA. Portanto, o mecanismo de manutenção da EA é distinto das atividades de indução e diferenciação do broto do membro. Mas outros estudos ainda são necessários a fim de se determinar a relação entre Formin, BMP e Dactylin no mecanismo de manutenção da EA.

2.3.5. Diferenciação celular

Todos os mecanismos descritos resultam na formação das estruturas reconhecidas nos membros como ossos, músculos, nervos e vasos sanguíneos, além das estruturas epiteliais como cabelos, escamas ou penas. Para alguns desses elementos pouco é conhecido sobre os mecanismos que controlam a diferenciação celular terminal.

Em todos os membros dos vertebrados, as estruturas ósseas proximais (húmero ou fêmur) são as primeiras a serem determinadas e são formadas por células mesenquimais que deixam a ZP. Essas células saem do controle da EA e agregam-se para formar condensações pré-cartilaginosas que originarão os componentes ósseos. Ao mesmo tempo, algumas células localizadas ao redor desses primórdios ósseos comprometem-se com a morte celular programada (MCP), o que contribui com a forma do membro, especialmente nos espaços interdigitais. Os próximos estádios do desenvolvimento ósseo incluem a troca da cartilagem por tecido ósseo por ossificação endocondral, crescimento e alongamento dos ossos por proliferação e diferenciação dos condrócitos, e formação dos ossos por diferenciação dos osteoblastos (Capdevila & Belmonte, 2001).

O mecanismo exato que dirige a diferenciação de todos os tipos celulares que compõem o membro é desconhecido. Sabe-se que membros da família de fatores TGF- β , assim como seus receptores, moduladores extracelulares e tradutores intracelulares, controlam a escolha entre o destino digital e interdigital nos membros (formação óssea ou mcp). BMP parecem exercer função dupla, ora como indutores de MCP, ora como promotores do crescimento cartilaginoso, provavelmente via dois receptores: BmpR-1a (MCP) e Bmp-1b (desenvolvimento cartilaginoso). FGFs também exercem um

importante papel nesse momento do desenvolvimento uma vez que células da ZP devem estar protegidas tanto do estímulo apoptótico quanto do condrogênico, para que o crescimento do membro seja mantido. Outros fatores, além de TGF- β , FGF e BMP e suas respectivas vias sinalizadoras, devem estar envolvidos na diferenciação terminal de todos os tipos celulares presentes nos membros em desenvolvimento.

2.4 Expressed sequence tags (EST).

A determinação completa e precisa da seqüência de nucleotídeos do DNA oferece uma base molecular para auxiliar a compreensão do funcionamento e da estrutura de um organismo. Sendo assim, nos últimos anos a comunidade científica vem concentrando esforços para a realização de projetos na área de genômica, a fim de determinar a seqüência DNA e/ou identificar as regiões codificadoras de genes em genomas em um grande número de organismos. Entre as metodologias disponíveis para a identificação de genes, a análise das seqüências expressas (EST) tem provado ser uma estratégia bastante informativa, uma vez que permite identificar os genes expressos em uma linhagem celular ou tecido em um estágio de desenvolvimento específico.

As EST são seqüências parciais de uma das extremidades da molécula de DNA complementar (cDNA), resultantes do sequenciamento sistemático dos clones de uma biblioteca de cDNA. São portanto seqüências dos RNA mensageiros (mRNA) expressos em uma célula na forma de cDNA, o que permite a caracterização preliminar do transcriptoma de um organismo. Há informação suficiente nas 200-400 bases de nucleotídeos sequenciadas para a identificação desses cDNA e sua localização nos cromossomos (Adams et al., 1991; Hatey et al., 1998). Na Figura 4 está representado um esquema de obtenção de EST a partir de fragmentos de cDNA.

Esta metodologia de caracterização dos transcritos de um tecido (mRNA) facilita a identificação das regiões codificadoras em genomas complexos. As EST são extremamente eficientes para a identificação de genes desconhecidos e de polimorfismos entre indivíduos, fontes importantes para diversas pesquisas biológicas. Permite ainda determinar o perfil de expressão dos genes, de acordo com a abundância de aparecimento das EST referentes a um mesmo gene, permitindo fazer inferências sobre o

nível de atividade dos diferentes genes no tecido e fazer comparações entre diferentes tecidos e diferentes estágios de desenvolvimento. Finalmente, a coleção das EST é fonte essencial para a condução de estudos de expressão gênica, especialmente por *microarrays*, que possibilita avaliar a expressão de inúmeros genes de uma só vez (Adams et al., 1991; Hatey et al., 1998; Banfi et al., 1998).

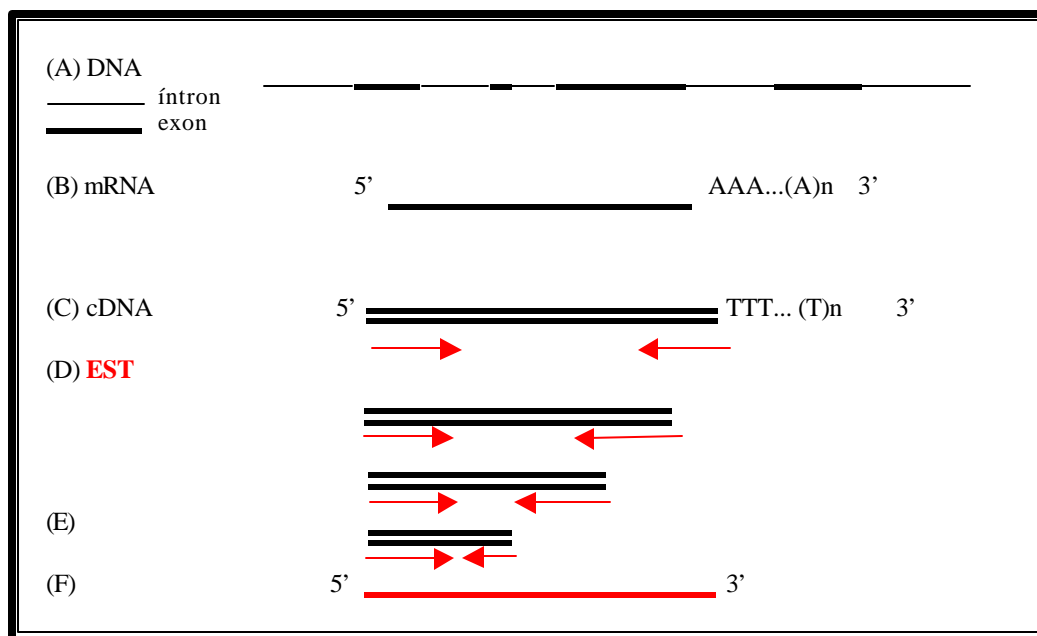


Figura 4 - Representação da obtenção de EST a partir do cDNA. (A) Um gene consiste de duas porções distintas: os *exons* e os *íntrons*. (B) Durante o processo de transcrição, os *íntrons* são eliminados por *splicing* e os *exons* unem-se para a formação da molécula de RNA mensageiro. Esta molécula sofre algumas alterações pós-transcricionais como a adição da cauda poli(A) na extremidade 3'. (C) DNA complementar (cDNA) é sintetizado tendo como molde a molécula de mRNA por ação da enzima transcriptase reversa (RT) e um *primer* oligo d(T). (D) O seqüenciamento das extremidades das moléculas de cDNA produz as EST. (E) cDNA incompletos permitem a identificação de diferentes seqüências de um mesmo gene. Essas seqüências podem sobrepor-se e formar a seqüência consenso de um gene. (F) Seqüência consenso do cDNA completo.

A identificação de EST é uma estratégia que está sendo adotada nos estudos de vários organismos, tais como humanos, roedores, suínos, bovinos, aves e plantas. Um banco de dados onde essas seqüências podem ser depositadas é encontrado no *GenBank* (<http://www.ncbi.nlm.nih.gov>) (dbEST). Existem aproximadamente 12 milhões de EST publicadas neste banco. Dentre essas, mais de quatro milhões foram descritas para humanos (*Homo sapiens*) identificadas por análise de mais de 7200 bibliotecas de cDNA construídas de tecidos diferentes. O número de EST identificadas em camundongos (*Mus musculus* e *domesticus*) ultrapassa dois milhões em mais de 670 bibliotecas de cDNA. Para bovinos (*Bos taurus*) o número de EST publicadas é de 230 mil, suínos (*Sus scrofa*) 110 mil e *Gallus gallus* pouco mais de 60 mil em 47 bibliotecas de cDNA diferentes.

Nos últimos anos, inúmeras bibliotecas de cDNA vêm sendo construídas com o objetivo principal de categorizar as EST identificadas no maior número de tecidos de um organismo e também para auxiliar na localização de genes em genomas complexos, como o de humanos, por exemplo. Sendo assim, Jia et al. (2001) construíram uma biblioteca de medula óssea humana e 4258 EST foram seqüenciadas de ambas as extremidades dos clones. Cerca de 1860 seqüências consenso, foram identificadas quando comparadas contra o banco não redundante do *GenBank*. Entre essas, 60 foram identificadas como genes novos. Entre as EST que encontraram homólogos no banco de dados, foi possível selecionar membros de famílias gênicas conhecidas, além de permitir a localização digital de genes em regiões específicas dos cromossomos a partir da comparação das EST com o genoma humano.

EST também representam uma alternativa para a análise de organismos com organizações cromossômicas complexas, como certas plantas que apresentam altos níveis de poliploidia. Recentemente foi concluído o projeto de sequenciamento em larga escala de EST de cana-de-açúcar (Projeto SUCEST), planta que apresenta uma alta complexidade genômica (Vettore et al., 2001). Neste projeto foram sequenciadas cerca de 300.000 EST a partir de 26 bibliotecas de cDNA de tecidos, tratamentos e estádios de desenvolvimento diferentes. A análise dessas EST permitiu identificar um total de 43.141 seqüências consenso, dos quais 38% não apresentaram homologias com

seqüências existentes no banco de dados público. Do total de clusters, 47% foram biblioteca-específicos, ou seja, formados por EST de uma única biblioteca.

2.4.1 EST e desenvolvimento embrionário.

A metodologia de identificação de EST começa agora a responder inúmeras questões sobre a biologia do desenvolvimento. A coleção dos genes expressos por um organismo apenas será completa quando forem conhecidos os genes expressos em cada estágio do desenvolvimento. Considerando-se que alguns genes são expressos apenas nos estádios embrionários, a importância da utilização desses tecidos não pode ser subestimada. Em humanos, isso é particularmente problemático pela dificuldade de obter-se embriões por razões éticas e técnicas, o que é refletido pelo número de EST publicadas nos bancos públicos de seqüências. A maioria das quatro milhões de EST humanas publicadas é de tecidos adultos (Ko, 2001). Os tecidos embrionários poderiam contribuir para a identificação de um maior número de genes no genoma humano.

A solução para esse problema seria identificar modelos animais que pudessem substituir os embriões humanos, com alto grau de homologia entre as seqüências identificadas. Um modelo muito utilizado é o camundongo (*Ratus Novergicus* e *Mus musculus*), com mais de 500 mil seqüências dos estádios embrionários e fetais depositadas nos bancos, e com alto grau de homologia com as seqüências identificadas em humanos. Apenas recentemente foi publicado por Gou et al. (2001) as análises realizadas em uma biblioteca de cDNA de embriões humanos com quatro semanas de desenvolvimento. Por PCR foram identificados genes específicos relacionados ao desenvolvimento embrionário. Entre 70 clones selecionados aleatoriamente, 53% codificaram para genes conhecidos, 26% com hipotéticos conservados e 17% não apresentaram homólogos no banco de dados e foram designados como EST específicas de embriões humanos. Esta foi a primeira descrição de uma biblioteca de cDNA de embriões humanos, considerado um valioso banco para estudos de expressão gênica e mecanismos reguladores que controlam a embriogênese humana.

2.4.2 EST para animais domésticos.

Embriões de galinha têm sido extensivamente usados como modelos animais para estudos de muitos sistemas biológicos, pela facilidade de manipulação e também por seu rápido desenvolvimento. Mas trata-se de uma espécie que está atrasada em relação a outros organismos em relação à pesquisa genômica, já que poucos resultados de caracterização molecular de estádios embrionários foram publicados até o momento. Isso porque um dos principais objetivos dos projetos EST desenvolvidos para animais domésticos é elaborar mapas genéticos, especialmente um mapa funcional que integre a informação dos mapas de ligação, desenvolvidos por estudos de segregação de marcadores moleculares, aos mapas citogenéticos, que localizam esses marcadores nos cromossomos, para facilitar a localização de genes candidatos para *Loci* de características quantitativas (*quantitative trait loci* - *QTL*).

Um excelente exemplo de genômica para animais domésticos são os resultados dos trabalhos desenvolvidos por Smith et al. (2001b), que construíram uma biblioteca de cDNA normalizada de embriões de galinha da espécie *White Leghorn* com dez dias de desenvolvimento. O objetivo inicial foi identificar SNPs (*Single Nucleotide Polymorphisms*) em 99 EST selecionadas aleatoriamente. O número de SNPs identificados dentro de 37 *amplicons* gerados variou de zero a quatro de um total de 65, com frequência de aparecimento de uma transição a cada 470 bases. Cerca de 35% dos *amplicons* apresentaram apenas um SNP e 19% quatro. Posteriormente, as EST identificadas foram utilizadas para desenvolver *primers* para análise genética em outras espécies de interesse econômico, entre elas pato, peru e galinha d'angola (Smith et al., 2001a). Eles descreveram um total de 26 seqüências genômicas comparativas (CGS) com dez pares de *primers* específicos, o que permitiu estabelecer comparações filogenéticas entre essas espécies. Os índices de similaridade entre as CGS foram superiores entre galinha, pato e peru, sendo galinha d'angola filogeneticamente mais distante. Recentemente, os SNPs identificados nas EST foram analisados em duas linhagens distintas de galinha, postura e corte (Smith et al., 2002). Um total de 12 EST permitiu gerar um fragmento de 6489 pb, onde foi possível identificar uma variação de 0 a 10 SNPs de um total de 44, com frequência de 0,7%. Cerca de 70% desses SNPs foram

compartilhados entre as duas linhagens estudadas. Os SNPs e os pares de *primers* desenvolvidos poderão ser agora utilizados para análise genética em linhagens comerciais de galinhas.

3 MATERIAL E MÉTODOS

3.1 Obtenção e coleta dos embriões

Os ovos de frango de corte da linhagem *Agroceres-Ross* (AG-Ross) utilizados neste trabalho foram obtidos na Granja Ninho Verde, localizada no município de Tietê (SP). Todos os ovos foram incubados à 37 °C em atmosfera úmida para a obtenção de quatro diferentes estádios de desenvolvimento (E15, E21, E24 e E26), determinados segundo Hamburger & Hamilton (1951). Destes estádios foram selecionados embriões inteiros (E26) e dois tecidos diferentes microdissecados com auxílio de material cirúrgico: os somitos associados ao tubo neural (E15) e os membros anteriores e posteriores dos embriões (E21, E24 e E26).

Os embriões totais com seis dias de incubação (E26) foram selecionados para a construção da biblioteca de cDNA denominada CEMB. Dos tecidos microdissecados, embriões no estágio 15 de desenvolvimento (obtidos após 50 horas de incubação) foram utilizados para a extração de todos os somitos visíveis associados ao tubo neural do embrião. As regiões cefálica e caudal, de mesoderme paraxial não segmentado, foram descartadas. A região removida com auxílio de material cirúrgico foi demonstrada na Figura 5. A biblioteca de cDNA construída a partir desse tecido microdissecado foi denominada CSOM.

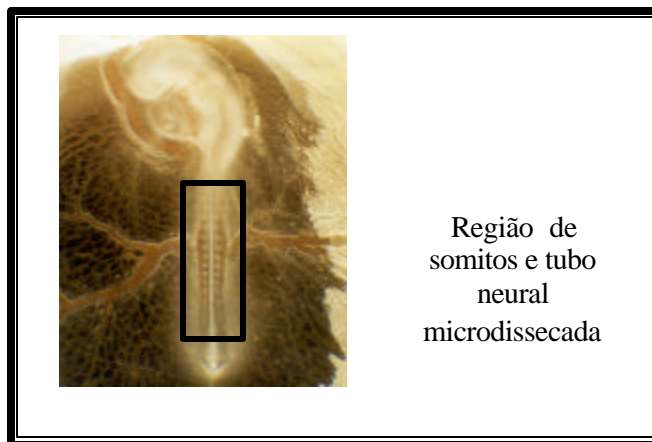


Figura 5 - Fotografia de um embrião de *Gallus gallus* no estágio 15 de desenvolvimento (50 horas de incubação à 37 °C) onde foi representado entre as linhas pretas a região de somitos associada ao tubo neural microdissecada dos embriões.

Os estádios 21 (3 dias e meio), 24 (4 dias e meio) e 26 (6 dias) foram selecionados para representar três momentos diferentes do desenvolvimento dos membros em *Gallus gallus*. Um *pool* de tecidos foi obtido a partir dos quatro membros (anteriores e posteriores) microdissecados dos embriões e a biblioteca de cDNA sintetizada foi denominada CMEM. Na Figura 6 foram apresentadas as fotografias que representam os embriões nos estádios selecionados para a microdissecação dos membros anteriores e posteriores de *Gallus gallus*, retiradas do trabalho publicado por Hamburger & Hamilton (1951), destacando os diferentes tamanhos dos brotos de membro em crescimento nos diferentes estádios. Na Figura 7 foram apresentadas as fotografias dos embriões de *Gallus gallus* que foram utilizados para as microdissecações dos brotos de membros neste estudo.

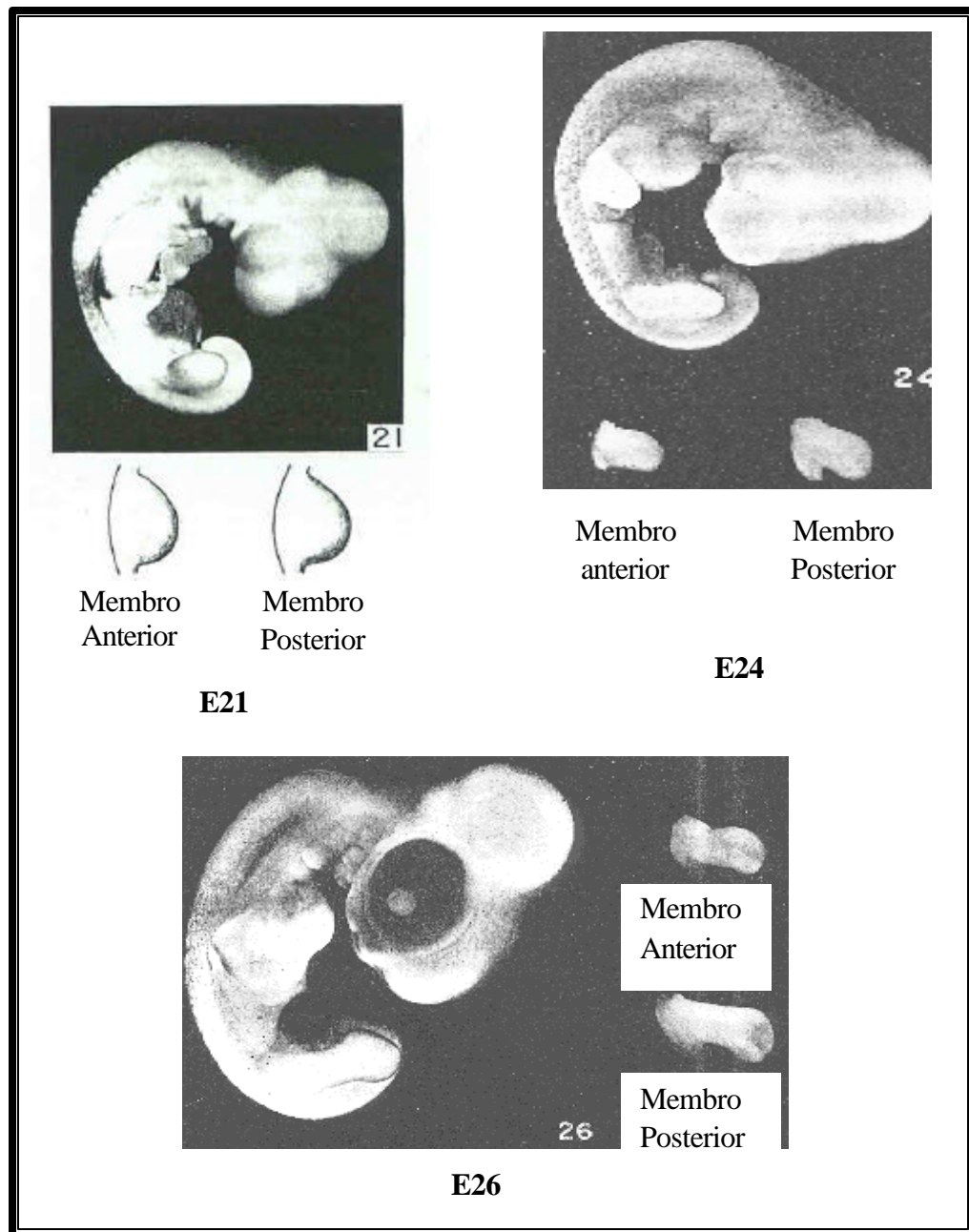


Figura 6 - Fotografias dos embriões de *Gallus gallus* nos estádios E21, E24 e E26 selecionados para a microdissecação dos brotos de membros anteriores e posteriores, retiradas do trabalho publicado por Hamburger & Hamilton (1951), com destaque para os diferentes tamanhos de brotos de membros obtidos em cada estágio.

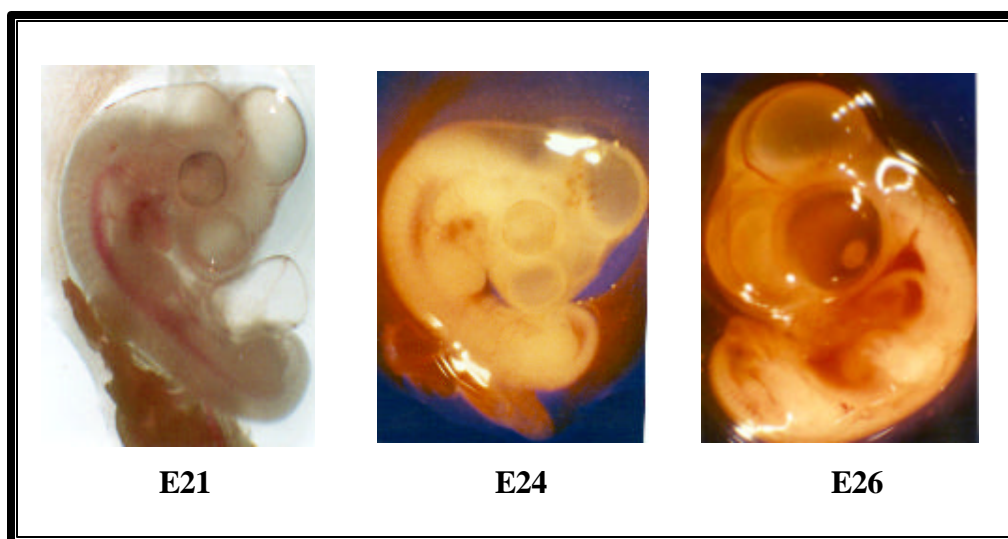


Figura 7 - Fotografias dos embriões de *Gallus gallus* utilizadas para a microdissecação dos brotos de membros nos estádios E21, E24 e E26 do desenvolvimento embrionário.

Na Tabela 1 foram descritos os estádios escolhidos para representar os genes expressos durante o desenvolvimento embrionário de *Gallus gallus*.

Tabela 1. Estadiamento dos embriões de *Gallus gallus*

Biblioteca	Tempo de incubação	Estádio de desenvolvimento
CEMB	6 dias	E26*
CMEM	3 dias e meio	E21*
	4 dias e meio	E24*
	6 dias	E26*
CSOM	50 horas	E15*

* Fonte: Hamburger & Hamilton (1951)

3.2 Bibliotecas de cDNA

3.2.1 Extração de RNA total

O RNA total foi isolado dos tecidos microdissecados utilizando-se o reagente Trizol (Invitrogen), conforme protocolo fornecido pelo fabricante, que corresponde à uma adaptação da metodologia descrita por Chomczynski & Sacchi (1987).

Os tecidos coletados foram imediatamente dissolvidos em Trizol e incubados à temperatura ambiente (TA) por 5 minutos. Adicionaram-se 200 µL de clorofórmio às amostras e, após agitação vigorosa por 15 segundos, as amostras foram centrifugadas à 12.000 x g por 15 minutos a 4 °C. O RNA foi precipitado adicionando-se 500 µL de álcool isopropílico aos sobrenadantes transferidos para tubos limpos. O pelete resultante da nova centrifugação foi lavado com 1 mL de etanol 75% (para RNA: 75% etanol absoluto: 25% água tratada com dietilpirocarbonato - DEPC) e centrifugado a 7.500 x g por 5 minutos a 4 °C. O RNA seco em estufa a 37°C por 10 minutos, foi ressuscitado em 20 µL de água DEPC e quantificado a 260 e 280 nm em espectrofotômetro. As amostras de RNA total foram submetidas a uma eletroforese em gel de agarose 1%, para verificação da integridade do material extraído.

3.2.2 Seleção da população de RNA mensageiro.

O RNA mensageiro (mRNA) foi selecionado da população total utilizando-se o *Kit MessageMaker[®] Reagent Assembly* (Invitrogen). O RNA total foi aquecido por 5 minutos a 65 °C para a desnaturação das estruturas secundárias das moléculas e imediatamente resfriado em gelo. A concentração de sal foi ajustada adicionando-se 0,1 volume de 5 M NaCl (concentração final de 0,5 M NaCl).

A seleção das moléculas de mRNA dentro da população total (RNA ribossomal e transportador) envolveu dois passos semelhantes. Para a primeira seleção, 2 mL de *oligo(dT) cellulose Suspension* foram adicionados à amostra de RNA total. Os *oligos (dT)* contidos nesta resina paream-se à cauda poli(A) das extremidades 3' das moléculas de mRNA. Após aquecimento a 37 °C por 10 minutos, a solução RNA total / *oligo(dT)* foi transferida para uma seringa filtro, onde inúmeras inversões permitiram o

pareamento das moléculas de mRNA aos *oligos (dT)* da resina. O RNA que não conseguiu parear-se à resina foi descartado empurrando-se o êmbolo da seringa. A resina oligo(dT) foi lavada utilizando-se 5 mL de dois diferentes tampões para a remoção dos RNAs ribossomais e transportadores contaminantes. O tampão 1 [0,5 M NaCl e 20 mM Tris-HCl (pH 7,5)] foi adicionado à seringa filtro, misturado por inversão e expelido. O mesmo foi feito com o tampão 2 [0,1 M NaCl e 20 mM tris-HCl (pH 7.5)]. Após as duas lavagens sucessivas, o mRNA foi liberado da resina oligo(dT) adicionando-se 2 mL de água destilada pré-aquecida à 65 °C.

Uma segunda seleção do mRNA já selecionado foi realizada, consistindo basicamente na repetição do mesmo protocolo descrito para a primeira. A mesma seringa filtro foi utilizada e para isso, a resina oligo(dT) foi previamente lavada com 5 mL do tampão 1. A concentração de sal do mRNA foi novamente ajustada à 0,5 M NaCl. A amostra foi transferida para a seringa e incubada à TA por 10 minutos, para permitir uma nova hibridização do mRNA à resina presente. Após expelir o material não anelado, repetiram-se os passos das lavagens com os tampões 1 e 2. Adicionou-se 2 mL de água destilada pré-aquecida a 65 °C à seringa para a liberação da população final de mRNA em um tubo estéril.

O mRNA foi precipitado utilizando-se glicogênio (50 µg/mL), 0,1 volume de acetato de amônio 7,5 M; 2 volumes de etanol absoluto resfriado a -20 °C. Após incubação a -20°C por 12 horas, o material foi centrifugado a 2.800 x g por 30 minutos a 4 °C. O pelete resultante foi lavado com 2 mL de etanol 75 % (para RNA) e seco por aproximadamente 30 minutos a 37 °C. O mRNA final foi ressuscitado em 10 µL de água DEPC e uma fração de 1 µL submetido a uma eletroforese em gel de agarose 1% para verificar a presença e estimar a quantidade de material selecionado.

3.2.3 Síntese do cDNA dupla fita.

Para a síntese do cDNA e clonagem direcionada no vetor pSPORT 1, utilizou-se o *Kit SuperScript Plasmid System* (Invitrogen). A Figura 8 corresponde a um esquema representativo referente à construção de uma biblioteca direcionada de cDNA, que

permite identificar as duas extremidades dos fragmentos de cDNA quando clonados nos vetores.

3.2.3.1 Síntese da primeira fita.

O mRNA selecionado foi utilizado como molde para a síntese da primeira fita de cDNA. Aos 9 μ L de mRNA foram adicionados 2 μ L de *NotI primer adapter*, um iniciador *oligo(dT)* associado à um adaptador contendo um sítio de restrição para a enzima *NotI*. O *primer* atuou como iniciador para a transcriptase reversa (SuperScriptRT II) sintetizar a cópia de cDNA; e o sítio de restrição identificou a posição da extremidade 3' da molécula de cDNA.

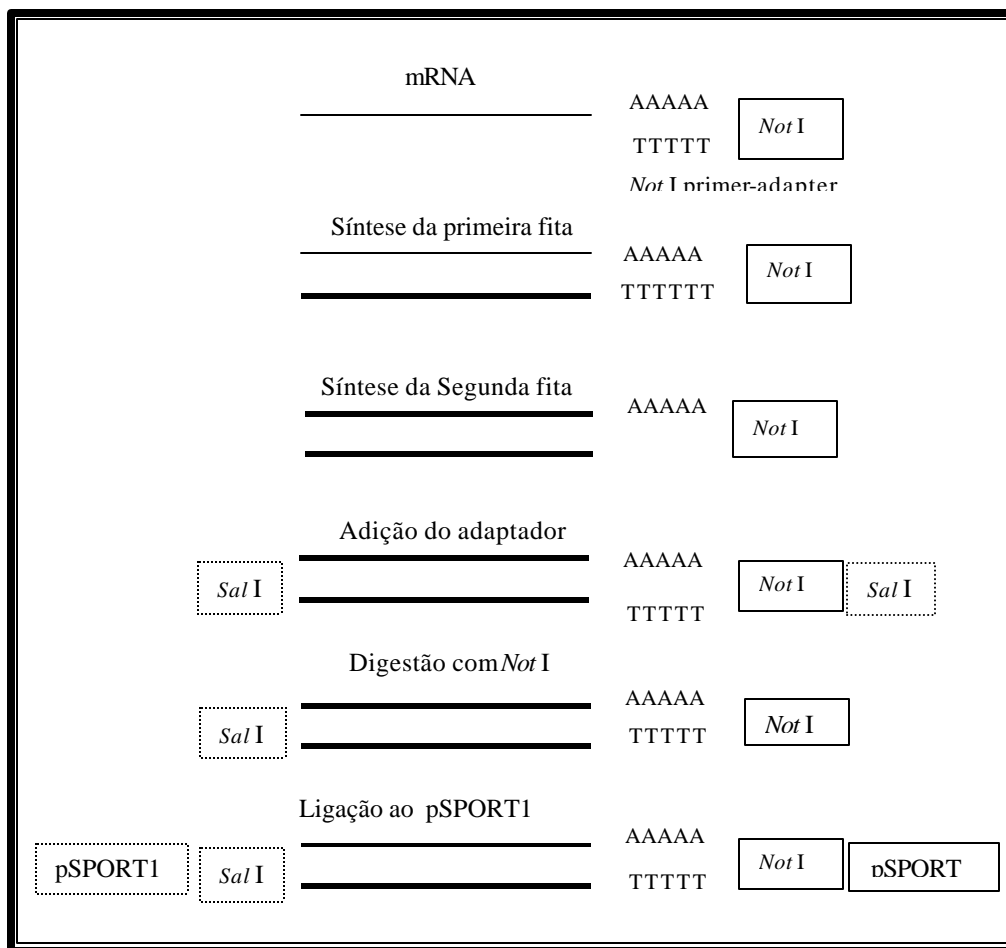


Figura 8 - Esquema representativo da clonagem direcionada adaptado do protocolo do kit *SuperScript Plasmid System* (Invitrogen).

Após aquecimento a 70 °C por 10 minutos, a reação foi resfriada em gelo para permitir a ligação do *primer* adaptador à cauda poli(A) das extremidades 3'. Para a síntese da primeira fita foram adicionados 4 µL de tampão de primeira fita [250 mM Tris-HCl (pH8.3), 375 mM KCl e 15 mM MgCl₂]; 2 µL de 0,1 M DTT e 1 µL de dNTP (10 mM). Após estabilização da temperatura por 2 minutos à 37 °C, 3 µL de SuperScriptRT II (200 U/µL) foram adicionados à reação, que foi posteriormente incubada por 2 horas à 37 °C.

3.2.3.2 Síntese da segunda fita.

O kit *SuperScript Plasmid System* utiliza a substituição via *nick translation* para a síntese da segunda fita de cDNA (OKAYAMA et al., 1982). À reação de primeira fita foram adicionados 93 µL de água DEPC, 30 µL de tampão de segunda fita [100 mM Tris-HCl (pH 6.9); 450 mM KCl; 23 mM MgCl₂, 0.75 mM β-NAD⁺; 50 mM (NH₄)₂SO₄], 3 µL de dNTP (10 mM). A síntese da segunda fita é catalisada pela combinação de 1 µL de *E. coli DNA ligase* (10 U/µL), 4 µL de *E. coli DNA polimerase I* (10 U/µL) e 1 µL de *E. coli RNase H* (2 U/µL). Após incubação da reação a 16 °C por 2 horas, 2 µL de T4 DNA polimerase (5 U/µL) foram adicionados e a reação foi novamente incubada a 16 °C por mais 5 minutos.

Para a precipitação do cDNA dupla fita sintetizado foram adicionados 10 µL de EDTA 0,5 M para terminar a reação de síntese. Em seguida, foram adicionados 150 µL de fenol: clorofórmio: álcool isoamílico (25:24:1). Após centrifugação à TA por 5 minutos à 14.000 x g para a separação das fases, 140 µL da fase superior foi cuidadosamente transferida para um novo tubo. Para a precipitação do cDNA, adicionaram-se 70 µL de acetato de amônio 7,5 M e 500 µL de etanol absoluto resfriado a -20 °C. Após centrifugação a TA por 20 minutos a 14.000 x g, o sobrenadante foi removido. O pelete foi lavado com 500 µL de etanol 70% resfriado a -20 °C. O cDNA foi seco em estufa a 37 °C por 10 minutos, para a completa evaporação do etanol residual.

3.2.3.3 Adição do adaptador *SalI*.

Um novo sítio de restrição foi adicionado, desta vez às duas extremidades do inserto para o reconhecimento da extremidade 5' no vetor pSPORT1 como representado na Figura 8. Para a reação de ligação do adaptador *SalI*, o pelete de cDNA obtido no passo anterior foi ressuspensionado em 25 µL de água DEPC. Em seguida adicionaram-se 10 µL de tampão de *T4 DNA ligase* [250 mM Tris-HCl (pH 7.6); 50 mM MgCl₂; 5 mM ATP; 5 mM DTT, 25% (w/v) PEG 8000], 10 µL do adaptador *SalI* (1 µg/µL) e 5 µL da enzima *T4 DNA ligase* (1 U/µL), em um volume final de 50 µL. A reação foi incubada a 16 °C por 16 horas.

Após as 16 horas de incubação, iniciou-se nova precipitação para a remoção de adaptadores não ligados. Adicionaram-se 50 µL de fenol: clorofórmio: álcool isoamílico (25:24:1) à reação. Após centrifugação a 14.000 x g a TA por 5 minutos, 45 µL da fase superior foi transferida para um tubo limpo. O cDNA foi precipitado com 25 µL acetato de amônio (7.5M) e etanol absoluto resfriado a -20 °C. Após centrifugação nas mesmas condições por 20 minutos, o pelete foi lavado com 200 µL de etanol 70% (para RNA). O pelete foi seco por 10 minutos a 37 °C, para evaporação do etanol residual.

3.2.3.4 Digestão com *NotI*.

Esta próxima etapa promoveu a digestão do adaptador *SalI* que havia sido adicionado à extremidade 3' dos insertos, de maneira a permitir a identificação de sítios de restrição distintos em cada extremidade do inserto: *SalI* na extremidade 5' e *NotI* na extremidade 3', como mostrado na Figura 8. Para esta reação, o cDNA precipitado no último passo foi ressuspensionado em 41 µL de água DEPC. Adicionaram-se 5 µL de *React 3* tampão e 4 µL da enzima de restrição *NotI* (15 U/mL). Após um período de incubação a 37 °C de 2 horas, o material foi novamente precipitado com fenol: clorofórmio: álcool isoamílico (25:24:1), seguindo as mesmas quantidades utilizadas no passo anterior, para a eliminação dos adaptadores removidos das extremidades.

3.2.3.5 Coluna de Fracionamento.

O cDNA foi fracionado em uma coluna fornecida pelo *kit SuperScript Plasmid System* (Invitrogen). O cDNA precipitado após digestão foi ressuspenso em 100 μL de tampão TEN [10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 25 mM NaCl]. Foram coletadas 20 frações seguindo-se as instruções descritas pelos fabricantes do kit. No tubo numerado 1 foi coletado todo o efluente dos 100 μL de cDNA adicionados à coluna (cerca de três gotas). No tubo 2, foram coletados outros 100 μL de tampão TEN adicionado à coluna. Nos tubos 3 ao 20 foi coletada uma gota por fração, sempre adicionado 100 μL de tampão TEN a cada 3 gotas coletadas.

Após o fracionamento, os volumes obtidos em cada fração foram medidos e registrados em uma tabela. Por cálculos de volume cumulativo foram selecionadas as frações desejadas; valores acima de 550 μL de volume cumulativo foram descartados (normalmente as frações de 15 a 20) por conterem fragmentos de cDNA menores que 500 pb e adaptadores não ligados, seqüências que seriam preferencialmente clonadas, de acordo com os fabricantes do kit. As frações selecionadas para a ligação foram aquelas imediatamente acima da fração identificada com 550 μL de valor cumulativo (fração 15). A partir da fração 14, três frações foram somadas a fim de garantir quantidades suficientes de cDNA para a ligação do inserto no vetor, e concentradas em *SpeedVac* a um volume de 10 μL . Assim, as frações 14, 13 e 12 foram somadas em uma única fração (fração A), e 11, 10 e 9 em outra (fração B), sendo que a fração B teoricamente apresentava fragmentos maiores do que a primeira (A).

3.2.4 Ligação do cDNA ao vetor pSPORT1.

Para a reação de ligação foram utilizados 1 μL de cDNA concentrado (não quantificado) para 0,5 μL de vetor pSPORT1 *NotI-SalI* cortado (50 ng/ μL). Adicionaram-se 2 μL de tampão de *T4 DNA ligase* [250 mM Tris-HCl (pH7.6), 50 mM MgCl_2 , 5 mM ATP, 5 mM DTT, 25% (w/v) PEG 8000], e 1 μL da enzima *T4 DNA ligase* (50 U/ μL), em volume final de 10 μL . As reações foram incubadas a 4 °C por no mínimo 12 horas.

3.2.5 Transformação.

3.2.5.1 Bactérias Competentes

As bactérias competentes utilizadas na transformação foram preparadas utilizando-se o protocolo descrito por Sambrook et al. (1985). Uma colônia isolada de *E. coli* DH5 α foi inoculada em 3 ml de meio SOB (triptona 2 %; 0,5 % de extrato de levedo; 0,5 % de NaCl; 2,5 mM KCl 1 M; 10 mM MgCl₂ 1 M e 10 mM MgSO₄ 1 M) e incubada sob agitação por 12 horas. Foram transferidos 50 μ L do pré-inóculo para um frasco contendo 50 mL de meio SOB e vigorosamente agitado por aproximadamente 3 horas a 37 °C. Medidas da densidade óptica (O.D.) foram realizadas freqüentemente de forma a monitorar o crescimento das células até atingir O.D. _{600 nm} de 0,5–0,6. Neste momento, adicionou-se 500 μ L de MgCl₂ 1 M às células, que foram imediatamente resfriadas em gelo por 15 minutos. Seguiu-se centrifugação a 2.800 x g por 10 minutos a 4 °C. O sobrenadante foi descartado e o pelete gentilmente ressuspensionado em 10 mL de RF I (KCl 100 mM; MnCl₂.4H₂O 10 mM; acetato de potássio 30 mM; CaCl₂.2H₂O 10 mM; glicerol 15 %). O material em suspensão permaneceu por 15 minutos em gelo e foi centrifugado a 800 x g durante 10 minutos a 4 °C. O sobrenadante foi novamente descartado e o pelete ressuspensionado em 2 mL de RF II (Na-MOPs 10 mM, KCl 10 mM, CaCl₂.2H₂O 75 mM, glicerol 15 %). Estes 2 mL foram divididos em alíquotas de 100 μ L/tubo pré-resfriados em gelo, imediatamente congelados em nitrogênio líquido e estocadas a –80 °C.

3.2.5.2 Choque Térmico.

A transformação das reações de ligação foi realizada em *E. coli* DH5 α competentes, segundo protocolo descrito por Sambrook et al. (1985). O choque térmico ocorreu a 42 °C por 90 segundos. Imediatamente foram adicionados às células 900 μ L de meio SOC (triptona 2 %, extrato de levedura 0,5 %; NaCl 5 M; MgCl₂ 1 M; KCl 1 M e MgSO₄ 1 M). Após agitação por 1 hora a 37 °C, as transformações foram plaqueadas em meio LB ágar 1,2 % com ampicilina (50 μ g/mL) e X-gal (20 mg/mL).

3.3 Validação das bibliotecas.

A fim de verificar o sucesso da síntese do cDNA e o tamanho dos insertos clonados no vetor, algumas colônias isoladas obtidas por transformação foram aleatoriamente selecionadas para um processo chamado validação, via reação em cadeia da polimerase (*Polymerase Chain Reaction - PCR*). Os *primers Universal M13* (5'-GTAAAACGACGGCCAGT-3') e o *Reverso M13* (5'-AACAGCTATGACCATG-3') foram utilizados para flanquear os insertos de cDNA no vetor pSPORT1. Para a amplificação dos insertos de uma colônia inteira foram utilizados 1,5 µL de tampão de PCR (10mM Tris-HCl pH 9, 50 mM KCl e 0,1 % Triton X-100), 0,5 µL de MgCl₂ (25 mM), 0,3 µL de dNTP 10 mM e 0,3 µL da enzima Taq polimerase (5 U/µL), em um volume final de 10 µL. Os cDNAs foram amplificados com uma desnaturação de 3 minutos para o rompimento das bactérias e em 35 ciclos (1 minuto à 96 °C, 1 minuto a 50 °C e 1 minuto a 72 °C). Os produtos da PCR foram visualizados em gel de agarose 1% para verificação da existência e do tamanho dos insertos.

3.4 Minipreparação do DNA plasmidial.

Após a validação das bibliotecas, colônias isoladas foram inoculadas em meio *Circle Grow* com ampicilina (100 mg/mL) e incubadas a 37°C por 22 horas sob agitação constante. As células foram sedimentadas a 3200 x g por 6 minutos à TA e o sobrenadante descartado por inversão. Ao pelete foram adicionados 240 µL da solução GET [glicose 20 %; EDTA 0,5 M (pH 8,0) e Tris-HCl 1 M (pH 7,4)]. Após nova centrifugação, o sobrenadante foi descartado por inversão. O pelete foi ressuscitado em 80 µL de GET e 60 µL foram transferidos para uma microplaca de 250 µL de polipropileno de fundo redondo (tipo Elisa) contendo 2,5 µL de RNase (10 mg/mL). Adicionou-se ao meio uma solução tensoativa de hidróxido de sódio (NaOH 0,2 N) com SDS 1 % (dodecil sulfato de sódio) e, após 10 minutos de incubação a TA, acrescentou-se 60 µL de acetato de potássio 3 M. Após centrifugação a 3.200 x g por 4 minutos a TA, a solução foi filtrada em uma placa *millipore* (MAGV N22). Ao material filtrado acrescentaram-se 110 µL de isopropanol e centrifugou-se a 3.200 x g por 45 minutos. O

sobrenadante foi descartado e o pelete lavado em etanol 70 %. Após secagem a TA por 60 minutos, o pelete foi ressuspenso em 40 μL de água Milli-Q.

3.5 Seqüenciamento.

3.5.1 Seqüenciamento em ABI Prism 377 DNA Sequencer.

O protocolo da reação de seqüenciamento realizado foi modificado a partir do descrito pelos fabricantes do *kit Big Dye Terminator Cycle Sequencing Ready Reaction* da *Perkin-Elmer Applied Biosystems*. O *primer* T7 (3,2 pmoles/ μL - 5'-TAATACGACTCATATAGGG-3') foi utilizado como iniciador para a enzima AmpliTaq DNA polimerase, hibridizando-se próximo à região 5' do inserto no vetor pSPORT1. Para aproximadamente 500 ng de DNA foram utilizados 2 μL de tampão *save money* (200 mM Tris pH9; 5 mM MgCl_2) e 2 μL do *kit Big Dye*, contendo a enzima AmpliTaq DNA polimerase e *Big Dye terminators*, para um volume final de 10 μL . O seqüenciamento foi em 35 ciclos de 45 segundos a 96 °C, 30 segundos a 55 °C e 4 minutos a 60 °C.

As reações de seqüenciamento foram purificadas adicionando-se glicogênio (1 g/L) e acetado de sódio (1,5M) com EDTA (0,25M). Foram adicionados 100 μL de etanol absoluto e após centrifugação a 3.200 x g a TA por 30 minutos, removeu-se o sobrenadante por inversão. Secou-se a placa a TA por 30 minutos e o pelete foi ressuspenso em 3 μL de tampão de corrida (5 formamida : 1 *loading tampão*). As amostras foram desnaturadas a 98 °C por 2 minutos e 1,2 μL foram carregados em gel de poliacrilamida. A eletroforese foi realizada em aparelho de seqüenciamento automático *ABI Prism 377 DNA Sequencer* sob 2400 V durante 7 horas.

3.5.2 Seqüenciamento em Megabace.

Algumas amostras de ambas bibliotecas foram analisadas em seqüenciador automático *Megabace – Amersham Pharmacia*. Para a reação de seqüenciamento de 300 ng de DNA foi utilizado 0,5 μL do *primer* T7 (3,2 pmoles/ μL) e 4 μL de *Dyemamic ET*

terminator sequencing premix , em um volume final de 10 μL . A reação consistiu de 30 ciclos de 20 segundos a 95 °C, 15 segundos a 50 °C e um minuto a 60 °C.

Para a purificação foram adicionados 2 μL de acetado de amônio 7.5 M e 30 μL de etanol absoluto resfriado a -20 °C. Após centrifugação a 3200 x g por 40 minutos, o sobrenadante foi removido. Adicionou-se ao pelete 150 μL de etanol 70% e centrifugou-se a 3200 x g por 10 minutos. Após secagem por 30 minutos a TA, o pelete foi ressuspenso em 10 μL de *loading solution* (Megabace). A eletroforese foi realizada sob 9Kv durante 100 minutos.

3.6 Análise das EST.

Os programas Phred (Ewing et al, 1998), Cap3 (Huang and Madan, 1999) e Consed (Gordon et al., 1998) foram utilizados para análise de qualidade e alinhamento das EST geradas de *Gallus gallus*. Os resultados desta análise geram algumas denominações foram definidas na Tabela 2.

Tabela 2. Termos utilizados em análise de bioinformática.

Termo	Definição
Clusterização	Processo de montagem dos <i>clusters</i> , neste caso realizado pelo programa Cap3 (Huang & Madan, 1999).
<i>Cluster (C)</i>	Agrupamento de EST definido em critérios estabelecidos pelo programa Cap3 (homologia de 30 pb com identidade de 90%).
<i>Singletons (S)</i>	EST que não participam de nenhum <i>cluster</i> , correspondendo a seqüências únicas.

As seqüências consenso dos *clusters* e dos *singletons* foram automaticamente comparadas contra as seqüências depositadas no *GenBank* (Benson et al., 2002) (<http://www.ncbi.nlm.nih.gov>), especialmente os bancos não redundante (banco de

dados de seqüências de nucleotídeos dos EUA, Europa e Japão) e dbEST (Boguski et al., 1993) utilizando-se dos programas BLAST (*Basic Local Alignment Search Tool*) (Altschul et al., 1990), para identificar a presença e o grau de homologia com seqüências descritas para outras espécies. Finalmente, uma categorização foi proposta para a classificação dos *clusters* e das seqüências dos *singletons*, baseada nas categorias utilizadas pelo *The Institute for Genomic Research* (TIGR) (<http://www.tigr.org>).

3.6.1 Phred: Análise de qualidade.

Os eletroferogramas gerados nos seqüenciamentos dos clones de cDNA foram inicialmente analisados pelo programa Phred (Ewing et al., 1998), que avaliou a qualidade dos picos correspondentes à cada base seqüenciada, conferindo um valor de probabilidade de erro a cada uma. Para este estudo foram consideradas apenas aquelas EST que apresentaram no mínimo 200 pb com qualidade Phred superior a 20, índice de probabilidade que corresponde a um erro a cada 1000 bases. As EST que não atingiram esses parâmetros foram removidas do banco. O programa *Cross-Match* (Ewing et al., 1998; <http://www.geospiza.com> ou <http://www.phrap.org>) foi utilizado para identificar seqüências de vetor, adaptadores adicionados às extremidades dos insertos de cDNA e cauda poli(A) presente na extremidade 3' dos cDNA seqüenciados. Nenhuma contaminação com DNA de *E. coli* foi identificada nas bibliotecas.

3.6.2 Cap3: Clusterização.

As EST aprovadas pelos parâmetros do programa Phred foram selecionadas para o processo de clusterização via o programa Cap3 (Huang & Madan, 1999). Os critérios utilizados foram um mínimo de 30 pb de homologia entre as seqüências com 90% de identidade (-o 30 e -p 90) para promover a sobreposição das EST, resultando na formação dos *clusters*. Aquelas seqüências que não entraram na formação dos *clusters* foram identificadas como *singletons*.

3.6.3 Relatórios das análises.

Após as seqüências terem sido avaliadas por estes programas de bioinformática, um relatório geral foi gerado contendo as seguintes informações:

- Número de seqüências avaliadas por biblioteca analisada.
- Número de seqüências que entram na formação de *clusters*.
- Número total de *clusters* formados.
- Número total de *singletons*.
- Índice de novidade, calculados automaticamente a partir da soma do número de *clusters* gerados ao número de seqüências únicas (*singletons*) e dividindo pelo número total de seqüências válidas.
- Número de seqüências com mais de 400 bases com qualidade Phred superior a 20.

3.6.4 BLAST automático.

As seqüências dos *clusters* e dos *singletons* de *Gallus gallus* foram comparadas com aquelas depositadas no *GenBank* (Benson et al., 2002) (<http://www.ncbi.nlm.nih.gov>) utilizando-se os programas BLASTX e BLASTN (Altschul et al., 1990), para identificar a presença e o grau de homologia com as seqüências descritas para outras espécies. O programa BLASTX foi utilizado para comparar as seqüências de nucleotídeos das EST, traduzidas em todos os quadros de leitura, contra o banco de seqüências não redundante (nr) do *GenBank*. O programa BLASTN foi utilizado para comparar as seqüências que não apresentaram homologia com o banco nr contra o dbEST humano (Boguski et al., 1993). Todos os resultados foram automaticamente e disponibilizados na página do laboratório de Biotecnologia Animal da ESALQ-USP (<http://biotecnologia.esalq.usp.br>).

3.6.5 Classificação das EST.

As seqüências dos *clusters* e dos *singletons* foram classificadas baseando-se nas categorias propostas pelo *The Institute for Genomic Research* (TIGR) conhecidas como *Expressed Gene Anatomy Database* (EGAD). A classificação de uma seqüência em uma

categoria foi determinada de acordo com a função exercida por essa proteína no organismo. A função foi verificada no próprio resultado do BLAST (<http://www.ncbi.nlm.nih.gov>), ou então, por busca no banco de proteínas conhecido como SwissProt (<http://www.expasy.ca>). As categorias consideradas para essa classificação foram:

- **Expressão Gênica.** Esta categoria incluiu proteínas envolvidas com a transcrição e tradução dos genes, especialmente as ribossomais, os fatores de transcrição, além das chaperonas e as heat shock proteins.
- **Metabolismo.** Proteínas envolvidas nas principais vias metabólicas do organismo como Glicólise e Ciclo de Krebs.
- **Estrutura Celular.** Nesta categoria foram incluídas proteínas de membrana, proteínas de citoesqueleto, além das proteínas musculares como miosina e actina.
- **Divisão celular.** Proteínas reguladoras do ciclo celular.
- **Matriz Extracelular.** Proteínas com propriedades adesivas expressas nos espaços extracelulares para manter a estrutura do tecido.
- **Proteínas Nucleares.** Fizeram parte dessa categoria as Histonas e as *DNA binding proteins*, como as *Zinc Finger proteins*, por exemplo.
- **Proteínas Transportadoras.** Essa categoria foi criada para incluir especialmente hemoglobinas.
- **Desenvolvimento.** Nesta categoria foram incluídos os genes identificados com função específica no desenvolvimento do organismo, especialmente envolvidas no processo de somitogênese, miogênese e na formação dos membros anteriores e posteriores.
- **Hipotéticas conservadas.** Proteínas identificadas sem função definida.
- **Low hit.** Sequências que apresentaram probabilidades de similaridades baixas (*E-value* superior a 10^{-5}).
- **No hit.** Sequências sem homologias com o *GenBank*.
- **Não classificadas.** Proteínas que não foram classificadas em nenhuma das categorias propostas.

3.6.6 Identificação dos SNPs (*Single Nucleotide Polymorphisms*)

Os polimorfismos de um nucleotídeo conhecidos como *Single Nucleotide Polymorphisms* (SNPs) foram identificados nos *clusters* das EST seqüenciadas de *Gallus gallus*, como indicativo da variabilidade presente entre os indivíduos estudados. Esses SNPs foram identificados analisando-se as discrepâncias em bases com altos índices de qualidade dentro de um mesmo *cluster* (*high quality discrepancy*). Os índices de qualidade das bases consideradas foram determinados segundo o programa Phred, sendo superiores a 40 (probabilidade de um erro a cada 10.000 bases).

4 RESULTADOS E DISCUSSÃO

4.1 Seleção e coleta dos embriões.

Para a identificação dos genes expressos durante o desenvolvimento embrionário de *Gallus gallus* foram escolhidos quatro diferentes estádios de desenvolvimento (E15, E21, E24 e E26), seguindo o estadiamento proposto por Hamburguer & Hamilton (1951). As fotografias dos embriões utilizados neste estudo, referentes aos estádios de desenvolvimento escolhidos, foram apresentadas nas Figuras 5, 6 e 7 (Material e Métodos).

O estádio E15 foi selecionado para a microdissecação dos somitos associados ao tubo neural. Neste estádio foi possível visualizar a presença de 24 a 27 somitos dispostos ao longo do eixo antero-posterior do embrião. Este estádio foi selecionado por apresentar níveis consideráveis de expressão de fatores relacionados à formação dos miótomos. Optou-se pela associação do tubo neural ao somitos microdissecados exatamente pela influência molecular dessa estrutura axial ao desenvolvimento dos somitos, especialmente na formação das três camadas embrionárias precursoras de musculatura esquelética, cartilagem e ossos do organismo (esclerótomo, dermomiótomo e miótomo), com a expressão dos fatores de crescimento como FGF e TGF- β (Goulding et al., 1994).

Os estádios E21, E24 e E26 foram selecionados para este estudo por corresponderem aos estádios iniciais de desenvolvimento dos brotos de membros, estádios que apresentaram os brotos visíveis nos flancos embrionários. Estes estádios corresponderam às fases de estabelecimento das diferentes linhagens celulares e início do processo de diferenciação e crescimento dos brotos de membro.

4.2 Bibliotecas de cDNA

4.2.1 RNA total e seleção do mRNA.

A partir do embrião inteiro no estágio E26 e dos tecidos específicos (somitos e brotos de membros) microdissecados dos embriões, três *pools* de RNA total foram obtidos para a construção das bibliotecas de cDNA (CEMB, CMEM e CSOM). Na Tabela 3 foram relacionadas as quantidades de embriões utilizadas para cada extração de RNA total de cada biblioteca de cDNA construída. Para a biblioteca de embrião inteiro de *Gallus gallus* (CEMB), dez embriões no estágio 26 de desenvolvimento foram suficientes para a obtenção de 1,2 mg de RNA total. Para a biblioteca de membros (CMEM), dos três diferentes estágios de desenvolvimento selecionados (E21, E24 e E26) foi obtido um *pool* de um total de 60 embriões, que rendeu 2,0 mg RNA total. Para a biblioteca de somitos associados ao tubo neural (CSOM), 200 embriões no estágio E15 de desenvolvimento foram necessários para a obtenção de 1,5 mg de RNA total.

Tabela 3. Quantidade de RNA total utilizada por biblioteca.

Biblioteca	Estádio de desenvolvimento	No. de embriões	Quantidade de RNA total (mg)	Total (mg)
CEMB	E26*	10	1,2	1,2
CMEM	E21*	30	0,85	
	E24*	20	0,80	
	E26*	10	0,35	2,0
CSOM	E15*	200	1,5	1,5

* Fonte: Hamburger & Hamilton (1951)

Além da quantidade, a qualidade do RNA total extraído também é fator fundamental para a qualidade das bibliotecas construídas. Na Figura 9 foi apresentada a fotografia do gel de agarose 1% para a verificação da integridade do RNA total extraído. Os três *pools* apresentaram íntegras as bandas referentes aos RNA ribossomais (28S e 18S), indicativo de qualidade do material extraído.

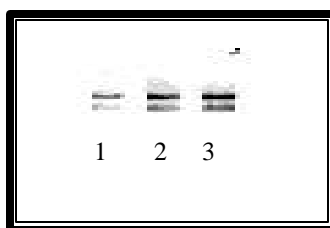


Figura 9 - Fotografia da eletroforese dos *pools* de RNA total extraídos a partir do embrião inteiro e dos tecidos microdissecados de *Gallus gallus*.

Foram necessárias quantidades superiores à 1 mg de RNA total para uma seleção eficiente de mRNA, quantidades exigidas pelos fabricantes do *kit* utilizado (Invitrogen). A qualidade e a quantidade do mRNA selecionado foram estimadas em eletroforese em gel de agarose 1%. A partir da fotografia do gel também foi possível verificar contaminações de RNA ribossomais, que comprometem a qualidade das bibliotecas. Para todos os mRNA selecionados neste projeto, as qualidades foram semelhantes ao representado no gel da Figura 10.

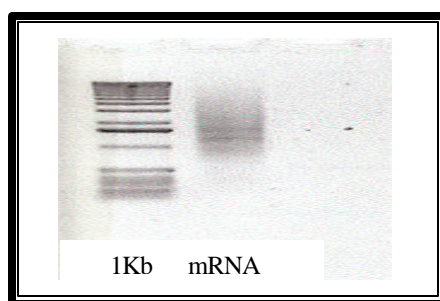


Figura 10 - Fotografia da eletroforese do mRNA selecionado para estimativas de qualidade e quantidade de material.

4.2.2 Validação das Bibliotecas de cDNA.

Todo o processo de síntese do cDNA dupla fita foi realizado sem qualquer marcação radioativa, diferente do proposto pelos fabricantes do *kit SuperScript Plasmid*

System (Invitrogen). Dessa forma, não houve como acompanhar a eficiência da síntese das duas fitas e da adição dos adaptadores às extremidades dos fragmentos de cDNA sintetizados, processo só confirmado após clonagem no vetor pSPORT1. Assim, para a verificação do sucesso do processo de síntese e clonagem dos fragmentos de cDNA e também para a verificação do tamanho relativo dos insertos clonados, alguns clones obtidos após transformação foram utilizados para um processo chamado validação das bibliotecas. Este processo foi importante pois permitiu verificar a qualidade das bibliotecas construídas antes do sequenciamento dos clones.

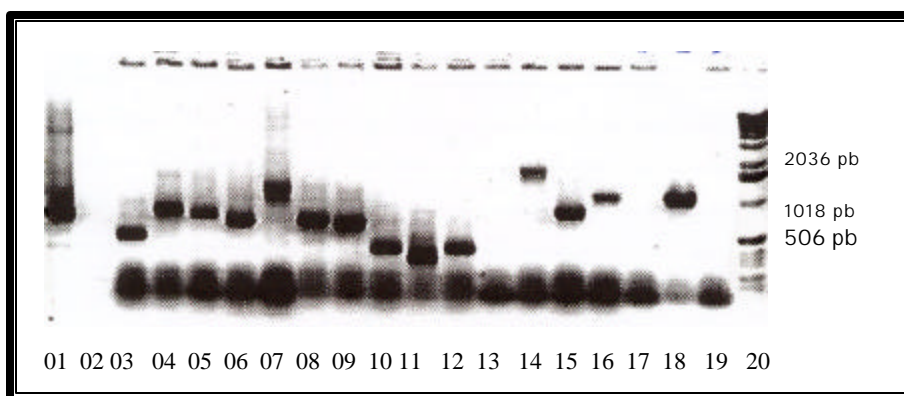


Figura 11 - Produtos amplificados na PCR de colônia para a verificação da presença e tamanho dos insertos clonados. Fotografia do gel de agarose 1%. (01) controle positivo; (02-19) colônias selecionadas aleatoriamente; (20) 1Kb DNA ladder.

Para o processo de validação, algumas colônias foram utilizadas para uma PCR de colônia. Na Figura 11 foi revelada a fotografia do gel de agarose 1% para a visualização dos produtos amplificados. Os tamanhos variaram entre 500 e 1000 pb para todas as bibliotecas, com médias de 600 pb para CEMB, 700 pb para CSOM e 600 pb para CMEM. As médias relativamente baixas de tamanhos de insertos resultaram da baixa eficiência de fracionamento da coluna utilizada neste estudo, mas isto não comprometeu a representatividade das bibliotecas. Posteriormente foi verificado que esses insertos pequenos foram obtidos pela ação incompleta da enzima SuperScriptRT II durante a síntese da primeira fita de cDNA. Isto significa que as bibliotecas de cDNA

construídas representaram mRNA de tamanhos variados, porém com cDNA pequenos, com a extremidade 5' provavelmente na porção central do mRNA, região mais conservada dessa molécula.

4.3 Análise dos dados

Um total de 6184 insertos da extremidade 5' foi seqüenciado dos clones obtidos das três bibliotecas de cDNA construídas neste estudo: 424 para CEMB; 2880 para CMEM e 2880 para CSOM (Tabela 4). O programa Phred (EWING et al, 1998) realizou a análise de qualidade das seqüências geradas e aquelas EST que não atingiram os padrões de qualidade estabelecidos (200 pares de bases com qualidade Phred superior a 20) foram removidas do banco. Também foram identificadas pelo programa *Cross-Match* as seqüências de vetor, adaptadores e cauda poli(A) dos fragmentos de cDNA clonados. Esta primeira análise resultou em 4998 EST válidas, o que representou um índice de sucesso de 80,8%. Para cada biblioteca foram selecionados 383 EST de CEMB, 2328 para CMEM e 2287 para CSOM, como descrito na Tabela 4.

Tabela 4. Número de clones seqüenciados e número de clones válidos por biblioteca de cDNA após análise de qualidade (Phred).

Biblioteca	No. de clones Seqüenciados	No. de clones válidos
CEMB	424	383
CMEM	2880	2328
CSOM	2880	2287
Total	6184	4998

4.3.1 Clusterização.

Na Tabela 5 foram descritos os resultados da clusterização das 4998 EST válidas. Foram realizadas três análises de clusterização isoladas (para cada biblioteca) e uma conjunta (CEMB/CMEM/CSOM). A clusterização das EST da biblioteca CEMB

revelou a formação de 151 *singletons* e 88 *clusters* formados com 232 EST; para a biblioteca CMEM, 527 *singletons* e 485 *clusters* com 1801 EST; e finalmente para a biblioteca CSOM, 891 seqüências únicas e 411 *clusters* com 1396 EST. A análise conjunta identificou 1427 *singletons* e 902 *clusters* formados com 3571 EST, revelando a identificação de 2329 seqüências consenso totais nas três bibliotecas.

As bibliotecas CMEM e CSOM apresentam aproximadamente o mesmo número de clones válidos e por isto podem ser comparadas. O número de *clusters* entre as bibliotecas CMEM e CSOM é semelhante, porém os *clusters* da biblioteca CMEM contêm um maior número de EST em relação à biblioteca CSOM. Consequentemente a biblioteca CSOM apresenta um maior número de *singletons*. Assim, um maior número de seqüências consenso (*clusters* mais *singletons*) foi identificado para a biblioteca CSOM. Isoladamente, 239 *clusters* mais *singletons* foram formados para biblioteca CEMB; 1012 para CMEM e 1302 para a biblioteca CSOM. Estas observações sugerem que o tecido utilizado para a biblioteca CSOM é menos especializado que o da biblioteca de membros, pois a população de RNAm seqüenciada é mais diversificada.

Tabela 5. Resultados da clusterização das EST de *Gallus gallus* via Cap3.

<i>Biblioteca</i>	No. <i>cluster</i>	No. de EST em <i>clusters</i>	No. de <i>singletons</i>	No. de <i>clusters e singletons</i>
CEMB	88	232	151	239
CMEM	485	1801	527	1012
CSOM	411	1396	891	1302
CEMB CMEM CSOM	902	3571	1427	2329

Nos gráficos da Figura 12 foram distribuídos os números de EST alinhadas em *clusters* em relação à sua representatividade nas bibliotecas (% nos *clusters*). Para a biblioteca de embrião (CEMB), 60,5% das EST seqüenciadas formaram *clusters*; para CMEM 77% e 61% para CSOM. Para todas as bibliotecas, a maioria dos *clusters* foi

formada com o alinhamento de até quatro EST. Para CEMB, *clusters* formados com até 4 EST representaram 92%; para CMEM, 80% e CSOM, 87,8%.

Estes índices de frequência de EST nos *clusters* formados estão relacionados com a redundância das bibliotecas construídas. *Clusters* formados com cinco ou mais EST (cerca de 10 a 20%) foram considerados como altamente expressos nesse estudo. Esse número foi determinado como limite para considerar um *cluster* representando um gene altamente expresso seguindo os dados publicados por JIA et al. (2002), que construíram uma biblioteca de cDNA de medula óssea de humanos e analisaram 4258 EST, número de EST próximo ao analisado de *Gallus gallus* neste estudo. Para esses autores, a biblioteca de medula óssea construída apresentou um baixo índice de redundância, com cerca de 6% das EST ocorrendo mais de cinco vezes.

Para a biblioteca CMEM, os genes mais expressos codificaram para diferentes subunidades de proteínas ribossomais, formando *clusters* com 14 (0,6%, subunidade L10), 17 (0,7%, subunidade L4) e 26 (1,1%, subunidade L7A) EST. Hemoglobinas, globinas e citocromo C oxidase foram outros genes que apresentaram altos índices de expressão, formando *clusters* com 22 (hemoglobina cadeia alfa, 0,9%), 67 (globinas alfa, 2,8%) e 23 EST (Citocromo C oxidase, 0,9%). O Fator de alongação tipo 1 também foi identificado entre os mais expressos, formando um *cluster* com 36 EST (1,5%). O perfil das categorias mais expressas em CSOM correspondeu ao mesmo encontrado em CMEM. O *cluster* formado com 20 EST (0,8%, subunidade S2) codificou para uma proteína ribossomal; 24 (1,0%) EST para hemoglobina cadeia alfa e 51 (2,2%) para alfa-globinas e dois diferentes polipeptídeos de citocromo C Oxidase formaram *clusters* com 30 (polipeptídeo I, 1,3%) e 34 (polipeptídeo III, 1,4%) EST. Para a biblioteca CEMB o perfil dos genes mais expressos foi diferente. O *cluster* formado com o maior número de EST (7; 1,8%) foi identificado para uma proteína hipotética conservada, com 98% de identidade com *Mus musculus*. O *cluster* formado com seis EST codificou para Tubulina cadeia α .

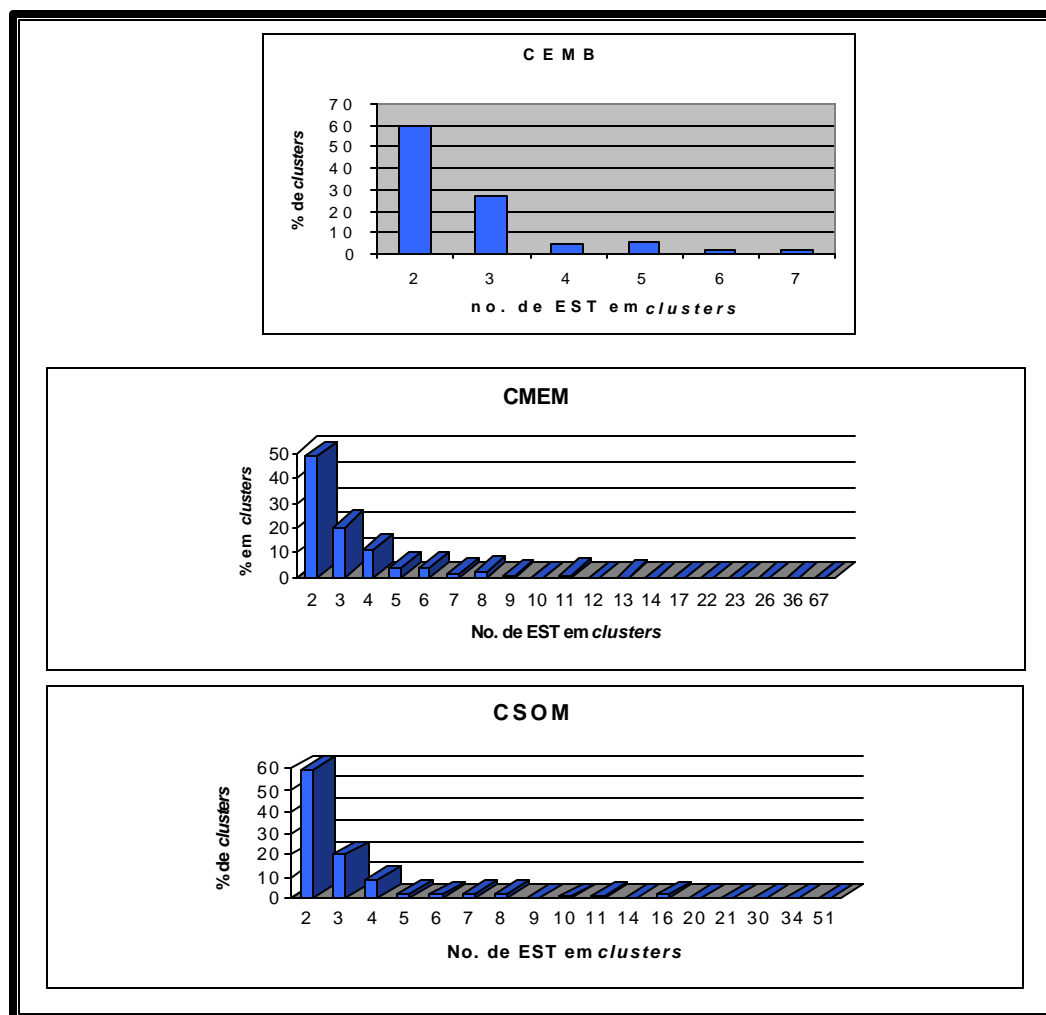


Figura 12 - Análise da clusterização das bibliotecas de cDNA. Distribuição do número de EST em função de sua representatividade nos *clusters*.

4.3.2 EST biblioteca-específicas.

A análise de clusterização conjunta para as três bibliotecas de cDNA construídas (CEMB/CMEM/CSOM) revelou o número de *clusters* e *singletons* biblioteca-específicos, ou seja, o número de EST que foram identificadas especificamente em cada uma das bibliotecas. É importante ressaltar que esses *clusters* e *singletons* biblioteca-específicos foram identificados nas 4998 EST analisadas, o que não implica em dizer que uma EST identificada em apenas uma biblioteca (biblioteca-específica) não seria

encontrada nas outras caso um maior número de EST tivesse sido seqüenciado. Portanto, esses dados de especificidade foram determinados na amostragem obtida de *Gallus gallus* neste estudo. Todos os dados obtidos dos *clusters* e *singletons* biblioteca-específicos foram indicados na Tabela 6.

Cerca de 26,5% dos *clusters* não foram biblioteca-específicos, ou seja, foram formados por EST expressas em pelo menos duas bibliotecas. Isto sugere que estas EST estão sendo coordenadamente expressas nos diferentes tecidos estudados ou que são expressas em resposta às necessidades comuns ao desenvolvimento embrionário. Por outro lado, cerca de 73,5% dos *clusters* foram formados por EST expressas em uma única biblioteca. Desses, 51 *clusters* (58%) foram biblioteca-específicos para CEMB, 335 (69%) para CMEM e 277 (67%) para CSOM. Estes altos índices de *clusters* biblioteca-específicos nas bibliotecas analisadas indicam uma expressão diferencial das EST entre os tecidos analisados provavelmente em função da expressão gênica tecido-específica e também pelo nível de abundância diferencial dos mRNA entre os tecidos. Todos os *singletons* identificados na análise de clusterização conjunta foram considerados como biblioteca-específicos. Dos 1427 *singletons* gerados nesta análise, 107 foram encontrados em CEMB, 471 em CMEM e 849 em CSOM.

As contribuições específicas de cada biblioteca foram calculadas dividindo-se o número de *clusters* e *singletons* biblioteca-específicos pelo número de total de EST analisadas nas bibliotecas (4998), índice que revelou a contribuição dos *clusters* e os *singletons* biblioteca-específicos para o projeto. Os dados foram apertados na Tabela 6. Para a biblioteca CEMB este índice foi calculado em 3%; para CMEM 16% e 22.5% para CSOM. Os 3% obtidos na biblioteca CEMB foram resultado do baixo número de clones seqüenciados nesta biblioteca.

Os índices de contribuição específica obtidos nas bibliotecas de *Gallus gallus* foram relevantes quando comparados aos índices obtidos nas bibliotecas de cDNA do projeto SUCEST (*Sugarcane* EST). Para as 26 bibliotecas de cDNA construídas no projeto de Cana-de-açúcar que resultaram em 237,954 EST válidas, cerca de 53,2% dos *clusters* foram formados com EST de pelo menos duas bibliotecas e os outros 46,8% foram identificados como *clusters* únicos. Os *singletons* não foram considerados na

análise. Os índices de contribuição específica foram calculados em função do número total de *clusters* (43.141) e variaram de 0,55 a 4,6% entre as bibliotecas (Vettore et al., 2001). Mais uma vez, justifica-se essas diferenças em função do número de EST seqüenciadas nos dois projetos e nas diferenças dos cálculos de contribuição específica entre as bibliotecas.

Tabela 6. Números de *clusters* e *singletons* biblioteca-específicos e índices de contribuições relativa, específica e novidade das bibliotecas de *Gallus gallus*.

Biblioteca	No. de <i>clusters</i>	<i>clusters</i> biblioteca específicos	No. de <i>singletons</i>	<i>singletons</i> biblioteca específicos	Contribuições		Índice de Novidade
					Relativa	Específica	
CEMB	88	51	151	107	41,2%	3,0%	62,4%
CMEM	485	335	527	471	36,6%	16,0%	43,5%
CSOM	411	277	891	849	49,2%	22,5%	54,6%

4.3.3 Índices de novidade.

Os índices de novidade apresentados na Tabela 6 revelaram a contribuição dos *clusters* e *singletons* totais de cada biblioteca para o projeto. Para a biblioteca CEMB o índice de novidade foi calculado em 62,4%; para CMEM em 43,5% e 54,6% para CSOM. A análise conjunta revelou um índice total de 45,7% de novidade. Estes índices de novidade, associados às contribuições específicas dos *clusters* e *singletons* biblioteca-específicos, permitiram inferir sobre as eficiências das bibliotecas construídas, especialmente em gerar novas informações para o banco de dados de *Gallus gallus* deste estudo.

Assim, após o seqüenciamento de 4998 clones válidos, a cada dez seqüências geradas nestas bibliotecas, cerca de quatro eram novas para o banco. Para a biblioteca CEMB este índice foi elevado (seis novas informações a cada dez) exatamente pelo baixo número de seqüências obtidas dessa biblioteca, o que significa que estes clones

poderiam continuar sendo seqüenciados a fim de gerar novas informações para o banco. Entre as bibliotecas CMEM e CSOM, CSOM apresentou um índice de novidade mais elevado em relação a CMEM. Somado aos índices de contribuições relativas (número de *clusters* biblioteca-específicos somado ao número de *singletons* biblioteca-específicos pelo total de EST) foi possível inferir que o seqüenciamento da biblioteca CSOM resultaria em um maior número de seqüências novas para o banco de dados do que as outras bibliotecas. Estas observações indicam que a estratégia de fazer bibliotecas de diferentes regiões do embrião é mais eficiente do que seqüenciar uma única biblioteca de embrião completo.

4.3.4 Classificação das EST.

Os *clusters* e *singletons* de *Gallus gallus* identificados neste estudo foram classificados de acordo com sua função biológica em doze categorias propostas pelo *The Institute for Genomic Research* (TIGR) conhecidas como *Expressed Gene Anatomy Database* (EGAD): (1) metabolismo, (2) expressão gênica, (3) estrutura celular, (4) proteína nuclear, (5) proteína transportadora, (6) matriz extracelular, (7) divisão celular, (8) desenvolvimento, (9) hipotética conservada, (10) *Low hit*, (11) *No hit* e (12) não classificadas. As freqüências para cada categoria foram calculadas em função do número total de EST expressas nas bibliotecas (somatório dos *singletons* ao número de EST dentro de cada *cluster*). Estes dados foram apresentados no gráfico da Figura 13.

As freqüências nas categorias *Low hit* e *No hit* foram calculadas separadamente, mas ambas podem ser consideradas seqüências sem homologia com o banco de dados utilizado inicialmente (banco não redundante do *GenBank*). Analisadas em conjunto, essas duas categorias apresentaram as maiores freqüências para as três bibliotecas construídas: para a CEMB totalizaram 38%; 28,7% para CMEM e 22% para CSOM. A categoria hipotética conservada reúne outros importantes indícios para a identificação de novos genes nos tecidos estudados, uma vez que classificam EST identificadas em outras espécies, mas sem função biológica definida. As freqüências para essa categoria foram de 15,4% para a biblioteca CEMB, 12,7% para CMEM e 17,3% para CSOM. As categorias *Low hit*, *No hit* e hipotéticas conservadas analisadas em conjunto totalizam

mais de 40% das EST classificadas neste estudo, formando um importante catálogo de seqüências expressas para futuras pesquisas moleculares.

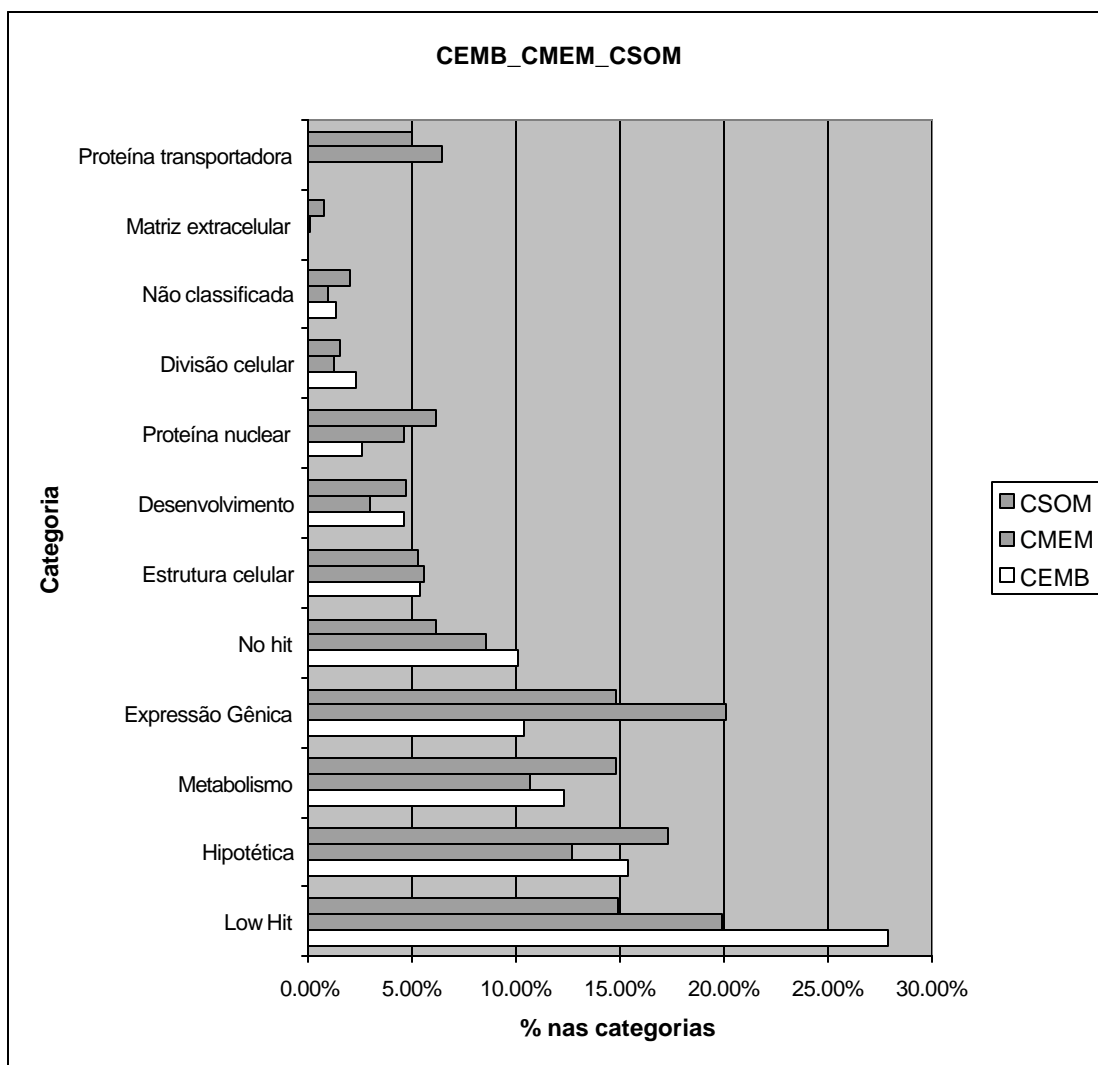


Figura 13 - Categorização dos *clusters* e *singletons* identificados nas bibliotecas CEMB, CMEM e CSOM. Distribuição das freqüências nas doze categorias estabelecidas para a classificação das EST de *Gallus gallus* deste estudo.

Metabolismo e expressão gênica foram as outras duas categorias que apresentaram altas freqüências de EST classificadas neste estudo. Isto pode ser

justificado por representarem categorias que incluem genes de manutenção de uma célula (*housekeeping genes*). Para CEMB, 12,3% dos *clusters* e *singletons* identificados foram classificados na categoria metabolismo; 10,7% em CMEM e 14,8% em CSOM. Em expressão gênica os índices ficaram em 10,4% para CEMB, 20,1% para CMEM e 14,8% para CSOM.

As categorias estrutura celular, proteína nuclear, proteína transportadora e matriz extracelular apresentaram individualmente frequências inferiores a 6%. Proteínas transportadoras (hemoglobinas) e de matriz extracelular não foram encontradas na biblioteca CEMB, mas foram identificadas nas bibliotecas CMEM e CSOM. As proteínas de matriz extracelular foram identificadas especialmente na biblioteca CSOM (0,7%), onde exercem importante função de adesão nos espaços entre somitos consecutivos (Bergmann et al., 1995). A categoria divisão celular apresentou índices inferiores à 2% em todas as bibliotecas analisadas. Essas seqüências reguladoras do ciclo celular são de particular interesse para este estudo por estarem relacionadas ao processo de diferenciação da musculatura esquelética, onde a parada no ciclo celular é pré-requisito essencial. Cerca de 2% das EST não puderam ser classificadas nas categorias estabelecidas.

4.3.5 Categoria dos *No hits*.

Os *clusters* e *singletons* classificados inicialmente na categoria *No hit* foram comparados contra o dbEST humano utilizando-se o programa BLASTN, a fim de confirmar a existência dessas seqüências nas bibliotecas construídas em outros bancos do *Genbank*. Dos 26 *clusters* e *singletons* classificados como *No hit* na biblioteca CEMB, 7 encontraram homólogos no dbEST humano e foram re-classificadas como hipotéticas conservadas; os outros 19 enquadraram-se na categoria *Low hits*. Para a biblioteca CMEM, 129 *clusters* e *singletons* não apresentaram homólogos com seqüências do banco não redundante (*GenBank*). Desses, apenas 30 apresentaram homólogos no dbEST humano; os 98 restantes apresentaram baixo grau de identidade (*Low hit*). Uma única seqüência não foi identificada neste banco e portanto permaneceu na categoria *No hit*. Para a biblioteca CSOM, dos 118 *clusters* e *singletons* classificados

na categoria *No hit*, 10 foram encontrados no mesmo banco e os 108 restantes apresentaram baixo grau de identidade (*Low hit*). Todas as seqüências que encontraram homólogos no dbEST humano e foram re-classificadas como hipotéticas conservadas, apresentaram grau de identidade em torno de 90% com cDNAs humanos. Este alto índice de identidade entre as seqüências nucleotídeos de *Gallus gallus* e humanos revela a importância de *Gallus gallus* como modelo animal para identificação de genes relacionados ao desenvolvimento embrionário em outras espécies, especialmente em humanos.

Tabela 7. Nova classificação dos *clusters* e *singletons* após análise contra o banco dbEST humano do *GenBank*.

Biblioteca	Categoria	Total	%
CEMB	Hipotética conservada	66	17,2%
	<i>Low hit</i>	126	32,9%
	<i>No hit</i>	0	-
CMEM	Hipotética conservada	327	14,0%
	<i>Low hit</i>	562	24,2%
	<i>No hit</i>	1	-
CSOM	Hipotética conservada	423	17,7%
	<i>Low hit</i>	463	19,4%
	<i>No hit</i>	0	-

Com esta nova análise, uma nova classificação foi proposta, onde os *clusters* e *singletons* inicialmente classificados na categoria *No hit* foram redistribuídos para a categoria hipotética conservada e *Low hit*. O total de seqüências hipotéticas conservadas identificadas na biblioteca CEMB passou de 15,4% para 17,2%; e de *Low hit* de 27,9% para 32,9%, enquanto nenhuma seqüência *No hit* foi identificada quando analisadas nessas novas condições. Os índices também foram alterados para a biblioteca CMEM. Após a nova classificação, os índices ficaram em 14,0% de hipotéticas conservadas e

24,2% de *Low hit*. E para CSOM, 17,7% de hipotéticas conservadas e 19,4% de *Low hit*. Estes dados da re-classificação dos *clusters* e *singletons* foram descritos na Tabela 7.

4.3.6 Genes relacionados ao desenvolvimento embrionário.

As EST geradas a partir dos somitos e dos membros anteriores e posteriores, permitiram identificar alguns genes determinantes ou reguladores do desenvolvimento de *Gallus gallus*. Dos *clusters* e *singletons* identificados nas bibliotecas, cerca de 5% foram categorizados com função específica em desenvolvimento embrionário (4,6% em CEMB, 3,0% em CMEM e 4,7% em CSOM), sem considerar outras categorias, como divisão celular e proteínas de matriz extracelular, que também agrupam genes de importância significativa para o correto desenvolvimento das estruturas estudadas (somitos, musculatura esquelética e brotos de membros anteriores e posteriores). Entre os genes envolvidos com o desenvolvimento embrionário identificados, cerca de 67% não foram descritos para *Gallus gallus* até o momento.

É importante ressaltar que a técnica de obtenção e análise de EST permite identificar os genes mais expressos em um determinado tecido do organismo em um estágio específico do desenvolvimento embrionário. Sendo assim, seria interessante tentar identificar os eventos relevantes para cada tecido analisado, a fim de se estabelecer uma visão geral dos principais mecanismos e processos de desenvolvimento identificados nas bibliotecas construídas.

Nos somitos, blocos de células mesenquimais formados no mesoderme paraxial, foi observado a expressão de genes de matriz extracelular. Essas proteínas normalmente exercem função de adesão e são extremamente importantes para a delimitação dos somitos (Pourquié, 2001). Foram identificados receptores transmembrana kinase de tirosina EPH, importantes na especificação genética dos limites entre somitos consecutivos e as *N-cadherins*, proteínas responsáveis pela formação da lâmina basal e manutenção desses limites (Bergmann et al., 1995). Outras proteínas identificadas com a mesma função foram *Nidogen 1* e *Fibulin 2* (proteínas de matriz extracelular), *Vinculin* e ZO-1 MDCK (proteínas envolvidas com a adesão celular) e *alpha-catenin*, proteínas que se associam as *cadherins*.

As estruturas axiais do embrião (notocorda e tubo neural) coordenam a compartimentalização dos somitos em esclerótomo, dermomiótomo e miótomo, tecidos embrionários precursores de cartilagem, ossos e musculatura esquelética (hipoaxial e epaxial) do corpo, respectivamente (Stockdale et al., 2000). Entre as principais vias sinalizadoras das estruturas axiais para controlar esse processo estão a família de fatores de crescimento TGF- β , a via Wnt/***b-catenin*** e Shh (Goulding et al., 1994). Membros da família TGF- β foram encontrados especificamente na biblioteca de somitos associados ao tubo neural, entre eles *Serine/threonine protein kinase*, proteína BAMBI e Serine/threonine kinase 1. Além deles também foram identificados ***b-catenin***, mediador da via Wnt e Smoothelin *L1*, modulador da via Shh.

Ainda seguindo cronologicamente a seqüência de eventos de desenvolvimento a partir dos somitos, uma população de células mesenquimais delamina da porção lateral do dermomiótomo e migra para colonizar diferentes regiões do corpo como precursores de células musculares hipoaxiais (Birchmeier & Brohmann, 2000). Alguns genes foram identificados com expressão restrita à população migratória e foram encontrados tanto na biblioteca de membros em desenvolvimento quanto na de somitos. Entre eles, *Scatter factor/Hepatocyte Growth Factor* (SF/HGF) e *Plexin*, homólogo à c-Met, ambos receptores de SF/HGF (Dietrich et al, 1999; Scaal et al., 1999); *fibronectin* e *Lamin-b*, responsáveis pelo agrupamento, migração e organização de células mesenquimais durante o desenvolvimento embrionário.

Com isso, podemos observar que no estágio de desenvolvimento selecionado para análise de genes expressos nos somitos associados ao tubo neural (estádio 15, segundo Hamburger & Hamilton, 1951) foi possível identificar a expressão de fatores responsáveis pela formação dos tecidos embrionários (esclerótomo, dermomiótomo e miótomo), assim como a expressão de fatores que regulam o início do estabelecimento dos precursores (condrogênicos e miogênicos) em locais específicos do embrião, representados pelo processo de delaminação e migração dos precursores miogênicos a partir do dermomiótomo.

Em relação à formação da musculatura esquelética, os fatores miogênicos (MyoD, Myf5, MRF4 e miogenina), responsáveis pela transcrição de genes músculo-

específicos (Ludolph & Konieczny, 1995), não foram identificados nas bibliotecas analisadas, provavelmente em função dos baixos níveis de expressão desses fatores nos estádios analisados. Mas já foi comprovado que a ativação do programa miogênico inicia-se desde a formação dos somitos (Stockdale et al, 2000; Puri & Sartorelli, 2000). Em trabalhos anteriores, os fatores miogênicos apresentaram índices de expressão em torno de 10^{-4} fmoles, quantificados em embriões inteiros de *Gallus gallus* (E21-E26). Genes constitutivos, como a β -actina por exemplo, foram identificados com índices de 10^{-2} fmoles (Alvares, 2001) nos mesmos embriões. Na biblioteca CMEM foram identificadas 11 EST (0,47%) que codificaram para β -actina, indicando que nesta biblioteca os genes representados apresentaram índices de expressão iguais ou superiores à 10^{-2} fmoles. Na biblioteca CSOM, nenhuma EST para β -actina foi identificada e apenas uma EST para GAPDH (outro gene constitutivo) foi encontrada. Estes índices indicam que para encontrar os fatores miogênicos nas bibliotecas analisadas seria necessário seqüenciar cerca de 100 vezes mais EST.

Alguns oncogenes antagonistas do programa miogênico foram identificados nas bibliotecas. Dentre eles, destacam-se as proteínas c-Jun e aquelas da família Ras, além do fator de transcrição AP-1, todos atuando na inibição da expressão de MyoD em mioblastos proliferativos (Bengal et al., 1992). A proteína transmembrana Notch (Tsakonas et al., 1999), FGF e seu receptor FGFR1, que também atuam inibindo a diferenciação terminal dos mioblastos, foram identificados nas bibliotecas analisadas (Li et al., 1992).

Após o período proliferativo, onde se determina o número de precursores miogênicos do organismo, a parada irreversível no ciclo celular dos mioblastos é pré-requisito essencial para o início da diferenciação dos mioblastos em miotubos. Foi significativa a presença de fatores reguladores do ciclo celular e de fatores que estabilizam a expressão de MyoD entre as EST analisadas. Entre outros, Cdks (*cyclin dependent kinases*), o produto do gene PBR (*peripheral benzodiazepine receptor*), Plx1, MN1, BTG1, fizzy1, *Prohibitin*, *cell division control preotein 2*, septin 7 e MCT1 (Buckingham et al., 2001).

Alguns indícios de início do processo de diferenciação dos mioblastos foram identificados, como por exemplo, a expressão dos *serum response factors* Ras-Rab 5A, Ras-Rab 12 e uma proteína homeobox conhecida como BARH-like 1B, que atuam na ativação de genes músculo-específicos (Croissant et al., 1996). Outros elementos relevantes na ativação do programa miogênico encontrados nas bibliotecas foram os receptores nucleares para hormônios tireóideanos (TH), como *Thyroid receptor interacting protein 12* e *Thyroid hormone receptor-associated protein*; e os membros da via do ácido retinóico, como *leucine zipper nuclear factor*, *P450 retinoid metabolizing protein* e *Retinoic acid binding protein I*. Além dessas proteínas, um outro indício de diferenciação muscular foi a identificação das subunidades da *Chromatin remodeling ATPase*, conhecidas como SWI/SNF, e das histonas transacetilases PCAF. Tanto as subunidades SWI/SNF quanto as histonas PCAF também são importantes para a ativação de promotores músculo-específicos (Puri & Sartorelli, 2000; Buckingham et al., 2001).

Entre as EST de membros analisadas foi freqüente a presença de genes característicos de tecido indiferenciado. Foram identificados por exemplo, algumas proteínas com propriedades de adesão celular, entre elas *Angio-associated migratory cell protein*, *Thrombospondin 4*, *β-cadherin*, *protocadherin*, e *Semaphorin 3F*, proteínas fundamentais para a manutenção do tecido mesenquimal que compõe o broto de membro em desenvolvimento. Outra característica que indicou a presença de tecido indiferenciado nos estádios analisados foi a presença de um fator de transcrição tipo *homeobox* que atua na manutenção da linhagem de células mesenquimais, conhecido como *paired mesoderm homeobox protein*. *Vimentin*, filamento intermediário de células mesenquimais e Rack1-P1, proteína sem função descrita mas com altos índices de expressão nesse tecido, também foram encontrados.

Algumas proteínas já descritas na literatura com função conhecida no desenvolvimento dos membros foram identificadas. Da família de genes Hox, por exemplo, foram identificados Hox D12, D3, B2, B5 e A7. Esses genes controlam a alocação de diversas estruturas no embrião, inclusive a posição dos brotos de membro nos flancos embrionários (Nelson et al., 1996; Nowicki & Burke, 1999). O primórdio de

membro, por exemplo, é induzido em posição específica no embrião que contém uma certa combinação de expressão desses genes, mas a combinação correta ainda não foi determinada (revisado por Capdevila & Belmonte, 2001).

Após determinar sua posição, as células mesenquimais posicionadas na região dos flancos embrionários iniciam um processo de proliferação celular, o que resulta na formação do primórdio de membro. Essa proliferação celular é induzida por um feedback estabelecido entre Fgf8 no mesoderme intermediário e Fgf10 no mesoderme lateral (Tickle et al., 2001; Martin, 2001), mediado pela via de sinalização Wnt/*b-catenin* (Kawakami et al., 2001). Desse sistema indutor foram identificados apenas os genes da via de sinalização Wnt, entre eles *b-catenin* e seu modulador positivo *Ser/Thr protein kinase* PAR-1A.

O crescimento é determinado por interações moleculares entre duas regiões específicas do membro, a EA (ectoderme apical) e o ZAP (mesênquima posterior) (Dudley & Tabin, 2000; Vargesson et al., 1997). Essa sinalização é estabelecida basicamente entre o fator Shh (*Sonic hedgehog*), expresso na ZAP, e Fgf4 expresso na EA (Laufer et al., 1994; Zeller et al., 1994). Nenhum desses dois fatores foram identificados entre as EST de membros. Este *loop* Shh/FGF é regulado por uma complexa rede de interações genéticas que envolvem a expressão de Formin e seu antagonista Gremlin, BMP e Noggin (antagonista), e dactylin, proteína da família F-box/WD40, que codifica para moléculas adaptadoras para proteínas que serão destruídas. Dessa fase que compreende o crescimento do membro apenas as F-box/WD40 foram identificadas (Sidow et al., 1999).

Na diferenciação celular terminal do membro, um dos eventos que se destacaram foi a morte celular programada (MCP) (revisado por Capdevila & Belmonte, 2001). Uma série de fatores envolvidos com a ativação ou regulação desse processo foi identificada, entre eles, caspase 9, *death associated protein*, *BH3 death agonist*, *deffender against cell death*, *growth arrest specific 2* entre outros.

4.3.7 Análise comparativa.

A análise das EST também permitiu determinar a média do grau de identidade entre as seqüências de aminoácidos de *Gallus gallus* e seus ortólogos. As EST de *Gallus gallus* apresentaram grau de identidade superior à 70% com *Homo sapiens*, *Rattus norvegicus* e *Mus musculus* (roedores). Para a biblioteca CEMB por exemplo, a média de identidade entre as EST de *Gallus gallus* deste projeto e seqüências descritas para *Homo sapiens* foi de 78,7%. A biblioteca CMEM e a CSOM apresentaram média de 75% de identidade com esta mesma espécie. Entre *Gallus gallus* e roedores (*Rattus norvegicus* e *Mus musculus*) a média de identidade foi de 86,9% (CEMB), 76,4% (CMEM) e 75,3% (CSOM). Estes altos índices de identidade entre as EST deste projeto e outros organismos, somados ao fato de *Gallus gallus* ser considerado um excelente modelo animal para estudos de desenvolvimento embrionário, permitem empregar as seqüências geradas de tecidos de *Gallus gallus* para identificação de novos genes em outras espécies, especialmente relacionados ao desenvolvimento embrionário.

4.3.8 Frequência de cDNA completos.

Uma análise importante realizada nas bibliotecas analisadas foi a identificação de cDNA com seqüência completa nos resultados do BLASTX. Para a biblioteca CEMB, 17,4% dos cDNA identificados eram completos, sendo 22 *singletons* e 19 *clusters*. Para a biblioteca CMEM, 214 (21%) também foram identificados como completos, 74 *singletons* e 140 *clusters*; e 452 (34,7%) em CSOM, sendo 249 *singletons* e 203 *clusters*. Estes índices de 20 a 30% identificados nas bibliotecas analisadas estão de acordo com os índices de aproximadamente 30% de insertos completos nas bibliotecas do projeto SUCEST (Sugarcane EST) (Vettore et al., 2001) e também com os índices do Projeto de EST de Ratos (Marra et al., 1999).

4.3.9 SNPs (*Single Nucleotide Polymorphism*)

O número de polimorfismos de um nucleotídeo denominado SNPs (*Single Nucleotide Polymorphism*) foi calculado para os 902 *clusters* gerados de todas as EST de *Gallus gallus* deste projeto, analisando-se as discrepâncias em bases com altos índices

de qualidade (phred superior a 40, uma base a cada 10.000) dentro de um mesmo *cluster* (*high quality discrepancy*). Um total de 394 SNPs foi identificado nos 902 *clusters* gerados. O número de SNPs variou de 1 a 20 no mesmo *cluster*; e a frequência de aparecimento foi calculada em 1 SNP a cada 1200 pares de bases. Smith et al. (2001) identificaram uma frequência de uma transição a cada 470 bases em embriões da linhagem *White Leghorn*, diferenças provavelmente justificadas pelos altos índices de qualidade bases utilizados para identificar SNPs neste trabalho.

Existem mais de 4 milhões de SNPs publicados para diversos organismos (ratos, humanos, chipanzé e galinhas, entre outros) no dbSNP do *GenBank* (<http://www.ncbi.nlm.nih.gov>). Desses, mais de 2 milhões foram publicados para o genoma humano. Estima-se que ocorra um SNP a cada 1000 bases no genoma (dado retirado no mesmo banco). A identificação desses polimorfismos é essencial para estudos de evolução e mecanismos de manutenção de variabilidade genética entre organismos, além de permitir detectar associações genéticas com doenças complexas em humanos, como diabetes, obesidade e hipertensão, ou a associações genéticas com características produtivas em animais domésticos.

5 CONCLUSÕES

As conclusões do presente estudo foram:

- O conjunto das EST identificadas neste estudo possibilitou a construção de um banco com 4998 EST dos estádios embrionários de *Gallus gallus*, fonte para futuras pesquisas no campo da biologia molecular e do desenvolvimento.
- As EST obtidas neste estudo permitiram identificar genes relacionados ao desenvolvimento que podem contribuir para elucidar os processos de formação de musculatura esquelética e membros anteriores e posteriores dos vertebrados através de estudos de expressão temporal e espacial dos genes relacionados. Cerca de 67% dos genes de desenvolvimento identificados não foram descritos para *Gallus gallus*.
- Este banco de EST dos estádios embrionários representa uma significativa fonte para a descoberta de novos genes em *Gallus gallus*. Cerca de 25% dos genes identificados não encontraram homólogos no banco de dados utilizado e 15% não apresentaram função definida (hipotéticos conservados).
- Este estudo revelou que *Gallus gallus* pode ser utilizado como modelo animal para estudos comparativos entre espécies, em função do alto grau de homologia identificado entre as seqüências desta espécie, humanos e

roedores. Isto é extremamente interessante considerando-se as dificuldades em obter-se informações dos estádios embrionários e fetais, especialmente em humanos.

- As EST obtidas neste estudo contribuíram para a identificação de SNPs (*Single Nucleotide polymorphism*), forma mais freqüente de variações no DNA e considerada a nova geração de marcadores moleculares que podem ser utilizados para a identificação de variações em importantes processos biológicos, genéticos, farmacológicos e médicos, além de representar importantes ferramentas para programas de melhoramento genético animal.

REFERÊNCIAS BIBLIOGRÁFICAS

- 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.; VENTER, J.C. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. **Science**, v. 252, p.1651-1656, 1991.
- ALTABEF, M.; CLARKE, J.D.W.; TICKLE, C. Dorso-ventral ectodermal compartments and origin of apical ectodermal ridge in developing chick limb. **Development**, v.124, p. 4547-4556, 1997.
- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, 1990.
- ALVARES, L.E. Quantificação da expressão dos fatores miogênicos e de Pax3 durante o desenvolvimento embrionário de aves. Rio Claro, 2001. 132 p., Tese (Doutorado), Instituto de Biociências, Universidade Estadual Paulista “Júlio Mesquita filho”.
- ARNOLD, H.H.; BRAUN, T. Genetics of Muscle Determination and Development. **Current Topics in Developmental Biology**, v. 48, p. 129-164, 2000.
- BANFI, S.; GUFFANTI, A.; BORSANI, G. How to get the best of dbEST. **Trends in Genetic**, v.14, n. 2, p. 80-81, 1998.
- BENGAL, E.; RANSONE, L.; ASCARFMANN, R.; DWARKI, V.L.J.; TAPSCOTT, J.; WEINTRAUB, H.; VERMA, I.M. Funcional antagonism between c-Jun and MyoD proteins: a direct physical association. **Cell**, v. 68, p. 507-519, 1992.
- BENSON, D.A.; KARSCH-MIZRACHI, I.; LIPMAN, D.J.; OSTELL, J.; RAPP, B.A.; WHEELER, D.L. GenBank. **Nucleic Acids Research**, v. 30, n. 1, p. 17-20, 2002.

- BERGMANN, A.D.; CHENG, H.J.; BRAMBILLA, R.; KLEIN, R.; FLANAGAN, J.G. 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. **Molecular Cell Biology**, v.15, p. 4921-4929, 1995.
- BIRCHMEIER, C.; BROHMANN, H. Genes that control the development of migrating muscle precursor cells. **Current Opinion in Cell Biology**, v.12, p. 725-730, 2000.
- BOGUSKI, M.S.; LOWE, T.M.; TOLSTOSHEV, C.M. bdEST: Database for Expressed Sequence Tags. **Nature Genetics**, v. 4, n. 4, p. 332-333, 1993.
- BUCKINGHAM, M. Skeletal muscle formation in vertebrates. **Current Opinion in Genetics & Development**, v. 11, p. 440-448, 2001.
- BUMSTEAD, N.; YOUNG, J.R.; TREGASKES, C.; PALYGA, J.; DUNN, P.P. Linkage mapping and partial sequencing of 10 cDNA loci in the chicken. **Animal Genetics**, v. 25, n. 5, p. 337-41, 1994.
- BURGUESS, R.; CSERJESI, P.; LIGON, K.L.; OLSON, E.N. Paraxis: a bHLH protein expressed in paraxial mesoderm and developing somites. **Developmental Biology**, v. 168, p. 296-306, 1995.
- CAPDEVILA, J.; BELMONTE, J.C.I. Patterning mechanisms controlling vertebrate limb development. **Annual Review of Cell Development Biology**, v.17, p. 87-132, 2001.
- CHOMCZYNSKI, P.; SACCHI, N. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry**, v. 162, p. 156-159, 1987.
- COSSU, G.; TAJBAKHSI, S.; BUCKINGHAM, M. How is myogenesis initiated in the embryo? **Trends in Genetic**, v. 12, n. 6, p. 218-223, 1996.
- COUTINHO, L.L.; MORRIS, J.; MARKS, H.L.; BUHR, R.J.; IVARIE, R. Delayed somite formation in a quail line exhibiting myofiber hyperplasia is accompanied by delayed expression of myogenic regulatory factors and myosin heavy chain. **Development**, v. 117, p. 563-569, 1993.
- CROISSANT, J.D.; KIM, J.; EICHELE, G.; GOERING, L.; LOUGH, J.; PRYWES, R.; SCHWARTZ, R.J. Avian serum response factor expression restricted primarily to

- muscle cell lineages in required for α -actin gene transcription. **Developmental Biology**, v. 177, p. 250-264, 1996.
- CROSS, M.K.; MORAN-RIVARD, L.; VELASQUEZ, T.; NAKATSU, M.N.; JAGLA, K.; GOULDING, M. Lbx1 is required for muscle precursor migration along a lateral pathway into limb. **Development**, v.127, n. 2, p. 413-424, 2000.
- DAUNCEY, MJ.; GILMOUR, RS. Regulatory factors in the control of muscle development. **Proceedings of the Nutrition Society**, v. 55, p. 543-559, 1996.
- DELFINI, M.C.; HIRSINGER, E.; POURQUIÉ, O.; DUPREZ, D. Delta 1- activated Notch inhibits muscle differentiation without affecting Myf5 and Pax3 expression in chick limb myogenesis. **Development**, v. 127, p. 5213-5224, 2000.
- DIETRICH, S.; ABOU-REBYEH, F.; BROHMANN, H.; BLADT, F.; SONNENBERG-RIETHMACHER, E.; YAMAAI, T.; LUMSDEN, A; BRAND-SABERI, B.; BIRCHMEIER, C. The role of SF/HGF and eMET in the development of skeletal muscle. **Development**, v. 126, p. 1621-1629, 1999.
- DUDLEY, AT.; TABIN, C.J. Constructive antagonism in limb development. **Current Opinon in Genetics & Development**, v. 10, p. 387-392, 2000.
- EWING, B.; HILLIER, L.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. **Genome Research**, v. 8, n.3, p. 175-185, 1998.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. **Genome Research**, v. 8, n. 3, p. 186-194, 1998.
- GILSON-BROWN, J.J.; AGULNIK, S.I.; SILVER, L.M.; NISWANDER L.; PAPAIOANNOU, V.E. Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development. **Development**, v. 125, p. 2499-2509, 1998.
- GOU, D.M.; CHOW, L.M.; CHEN, N.Q.; JIANG, D.H.; LI, W.X. Construction and characterization of a cDNA library from 4-week-old human embryo. **Gene**, v. 278, n. 1/2, p. 141-147, 2001.
- GOULDING, M.; LUMSDEN, A.; PAQUETTE, AJ. Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. **Development**, v.120, p.957-971, 1994.

- HAMBURGER, V.; HAMILTON, H.L. A series of normal stages in the development of the chick embryo. **Journal of Morphology**, v. 88, p. 49-92, 1951.
- HATEY, F.; TOSSER-KLOPP, G.; MARTINATO, C.C.; MULSANT, P.; GASSER, F. Expressed Sequence Tags for Genes: a review. **Genetic Selection and Evolution**, v. 30, p. 521-554, 1998.
- HUANG, X.; MADAN, A. Cap3: A DNA sequence assembly program. **Genome Research**, v. 9, n. 9, p. 868-877, 1999.
- JIA, L.; YOUNG, M.F.; POWELL, J.; YANG, L.; HO, N.C.; HOTCHKISS, R.; ROBEY, P.G.; FRANCOMANO, C.A. Gene expression profile of human bone marrow stromal cells: high-throughput expressed sequence tag sequencing analysis. **Genomics**, v. 79, n. 1, p. 7-17, 2002.
- JIANG, Y.J.; EANE, B.L.; SMITHERS, L.; HADDON, C.; ISH-HOROWICZ, D.; LEWIS, J. Notch signaling and the synchronization of the somite segmentation clock. **Nature**, v. 408, p. 475-479, 2000.
- JOHNSON, R.L.; TABIN, C. Molecular Models for vertebrate limb development. **Cell**, v. 90, p. 979-990, 1997.
- JUDSON, R.; SALISBURY, B.; SHNEIDER, J.; WINDEMUTH, A.; STEPHENS, J.C. How many SNPs does a genome-wide haplotype map require? **Pharmacogenomics**, v. 3, n. 3, p. 379, 2002.
- KABLAR, B.; KRASTEL, K.; YING, C.; TAPSCOTT, S.J.; GOLDHAMER, D.J.; RUDNICKI, M.A. Myogenic determination occurs independently in somites and limb buds. **Developmental Biology**, v. 206, p. 219-231, 1999.
- KAWAKAMI, Y.; CAPDEVILA, J.; BÜSCHER, D.; ITOH, T.; RODRIGUEZ ESTEBAN, C.; BELMONTE, J.C.I. Wnt signals control FGF-dependent limb initiation and EA induction in the chick embryo. **Cell**, v. 104, p. 891-900, 2001.
- KO, M. Embryogenomics: developmental biology meets genomics. **Trends in Biotechnology**, v. 19, p. 511-518, 2001.
- LAUFER, E.; NELSON, C.E.; JOHNSON, R.L.; MORGAN, B.A.; TABIN, C. Shh and FGF act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. **Cell**, v. 79, p. 993-1003, 1994.

- LI, L.; ZHOU, J.; JAMES, G.; HELLER-HARISON, R.; CZECH, M.; OLSON, E. FGF inactivates myogenic helix-loop-helix proteins through phosphorylation of a conserved protein kinase C site in their DNA binding domains. **Cell**, v. 71, p. 1181-1194, 1992.
- LUDOLPH, D.C.; KONIECZNY, S.F. Transcription factor families: muscling in on the myogenic program. **FASEB**, v. 9, n. 12, p. 1595-1604, 1995.
- MARRA, M.; HILLIER, L.; KUCABA, T.; ALLEN, M.; BARSTEAD, R.; BECK, C.; BLISTAIN, A; BONALDO, M.; BOWERS, Y.; BOWLES, L.; CARDENAS, M.; CHAMBERLAIN, A; CHAPPELL, J.; CLIFTON, S.; FAVELLO, A; GEISEL, S.; GIBBONS, M.; HARVEY, N.; HILL, F.; JACKSON, Y.; KOHN, S.; LENNON, G.; MARDIS, E.; MARTIN, J.; WATERSTON, R. An encyclopedia of mouse genes. **Nature Genetics**, v. 21, n. 2, p. 191-194, 1999.
- MARTIN, G. Making a vertebrate limb: new players enter from the wings. **BioEssays**, v. 23, p. 865-868, 2001.
- NELSON, C.E.; MORGAN, B.A.; BURKE, A.C.; LAUFER, E.; DIMAMBRO, E. Analysis of Hox gene expression in the chick limb bud. **Development**, v. 122, p. 1449-1466, 1996.
- NOWICKI, J.L.; BURKE, A.C. Hox genes and morphological identity: axial versus lateral patterning in the vertebrate mesoderm. **Development**, v. 127, p. 4265-4275, 2000.
- OLSON, E.N. Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. **Developmental Biology**, v. 172, p. 2-14, 1995.
- PALMEIRIM, I.; HENRIQUE, D.; ISH-HOROWICZ, D.; POURQUIÉ, O. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. **Cell**, v. 91, p. 639-648, 1997.
- PARR, B.A.; McMAHON, A.P. Dorsalizing signal Wnt-7A required for normal polarity of D-V and A-P axes of mouse limb. **Nature**, v. 374, p. 350-353, 1995.
- POURQUIÉ, OLIVEIR. Vertebrate Somitogenesis. **Annual Review of Cell Developmental Biology**, v. 17, p. 311-350, 2001.

- PURI, P.L.; SARTORELLI, V. Regulation of muscle regulatory factors by DNA-binding, interacting proteins, and post-transcriptional modifications. **Journal of Cellular Physiology**, v. 185, n. 2, p. 155-73, 2000.
- RODRIGUEZ-ESTEBAN, C.; TSUKUI, T.; YONEI, S.; MAGALLON, J.; TAMURA, K.; BELMONTE, J.C.I. The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. **Nature**, v. 398, p. 814-818, 1995.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. 2. ed. Plainview: Cold Spring Harbor Laboratory Press, 1989. 3 v.
- SCAAL, M.; BONAFEDE, A.; DATHE, V.; SACHS, M.; CANN, G.; CHRIST, B.; BRAND-SABERI, B. SF/HGF is a mediator between limb patterning and muscle development. **Development**, v. 126, p. 4885-4893, 1999.
- SMITH, E.J.; SHI, L.; SMITH, G. Expressed Sequence Tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library. 3. DNA sequence analysis of genetic variation in commercial chicken populations. **Genome**, v. 45, n. 2, p. 261-267, 2002.
- SMITH, E.J.; SHI, L.; PREVOST, L.; DRUMMOND, P.; RAMLAL, S.; SMITH, G.; PIERCE, K.; FOSTER, J. Expressed Sequence Tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library: 2. Comparative DNA sequence analysis of guinea fowl, quail, and turkey genomes. **Poultry Science**, v. 80, n. 9, p. 1263-1272, 2001a.
- SMITH, E.J.; SHI, L.; DRUMMOND, P.; RODRIGUEZ, L.; HAMILTON, R.; RAMLAL, S.; SMITH, G.; PIERCE, K.; FOSTER, J. Expressed Sequence Tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library: 1. DNA sequence characterization and linkage analysis. **Journal of Heredity**, v. 92, n. 1, p. 1-8, 2001b.
- STOCKDALE, F.E.; NIKOVITS, W.J.; CHRIST, B. Molecular and cellular biology of avian somite development. **Developmental Dynamics**, v. 219, p.304-321, 2000.
- TAVARES, A.T.; TSUKUI, T.; BELMONTE, J.C.I. Evidence that members of the Cut/Cux/CDP family may be involved in EA positioning and polarizing activity during chick limb development. **Development**, v. 127, p. 5133-5144, 2000.

- TAKEUCHI, J.K.; KOSHIBA-TAKEUCHI, K.; MATSUMOTO, K.; VOGEL-HÖPKER, A.; NAITOH-MASUO, M.; OGURA, K.; TAKAHASHI, N.; YASUDA, K.; OGURA, T. Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds. **Nature**, v. 398, p. 810-814, 1999.
- TICKLE, C. Morphogen gradients in vertebrate limb development. **Cell & Developmental Biology**, v. 10, p. 345-351, 1999.
- TICKLE, C.; MÜNSTERBERG, A. Vertebrate limb development – the early stages in chick and mouse. **Current Opinion in Genetics & Development**, v. 11, p. 476-481, 2001.
- TIRUNAGARU, V.G.; SOFER, L.; CUI, J.; BURNSIDE, J. An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones. **Genomics**, v. 66, n. 2, p. 144-151, 2001.
- TSAKONAS, S.A.; RAND, M.D.; LAKE, R.J. Notch signaling: cell fate control and signal integration in development. **Science**, v. 284, p. 770-776, 1999.
- VARGESSON, N.; CLARKE, J.D.W.; VICENT, K.; COLES, C.; WOLPERT, L.; TICKLE, C. Cell fate in the chick limb bud and relationship to gene expression. **Development**, v. 124, p. 1909-1918, 1997.
- VETTORE, AL.; SILVA, F.R.; KEMPER, E.L.; ARRUDA, P. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, p.1-4, 2001.
- VOGT, T.F.; DUBOULE, D. Antagonists go out on a limb. **Cell**, v. 99, p. 563-566, 1999.
- WALSH, K.; PERLMAN, H. Cell cycle exit upon myogenic differentiation. **Current Topics in Genetics & Development**, v. 7, p. 597-602, 1997.
- ZELLER, R.; HARAMIS, A.G.; ZUNINGA, A.; McGUIGAN, C.; DONO, R. Formin defines a large family of morphoregulatory genes and functions in establishment of the polarising region. **Cell Tissue Research**, v. 296, p. 85-93, 1999.