

CARLA VERMEULEN CARVALHO GRADE

**IDENTIFICAÇÃO DE POSSÍVEIS ELEMENTOS
C/S-REGULATÓRIOS DA MIOSTATINA UTILIZANDO
BIOINFORMÁTICA**

Monografia apresentada ao Departamento de Biologia Celular
como requisito para conclusão do Curso de Ciências
Biológicas do Setor de Ciências Biológicas da Universidade
Federal do Paraná.

Orientadora: Prof. Dra. Lúcia Elvira Alvares

CURITIBA

2005

AGRADECIMENTOS

Agradeço a Deus, a minha família, que sempre esteve ao meu lado, aos meus amigos, que sempre me apoiaram, e, em especial, a minha orientadora Lúcia Elvira Alvares, pela confiança que depositou em mim.

SUMÁRIO

AGRADECIMENTOS	iii
LISTA DE FIGURAS	v
LISTA DE TABELAS	vi
RESUMO	vii
1. INTRODUÇÃO	1
2. OBJETIVOS	15
3. METODOLOGIA	16
4. RESULTADOS	18
5. DISCUSSÃO	56
6. REFERÊNCIAS BIBLIOGRÁFICAS	60

LISTA DE FIGURAS

FIGURA 1 – LOCALIZAÇÃO DO MESODERMA PARAXIAL	1
FIGURA 2 – FORMAÇÃO E MATURAÇÃO DOS PARES DE SOMITOS	2
FIGURA 3 – SEQUÊNCIA DE AMINOÁCIDOS DA MIOSTATINA HUMANA	5
FIGURA 4 – COMPARAÇÃO DE CAMUNDONGOS NORMAIS E MUTANTES PARA O GENE DA MIOSTATINA	8
FIGURA 5 – SÍNTESE DA MIOSTATINA	9
FIGURA 6 – SINALIZAÇÃO DA MIOSTATINA	11
FIGURA 7 – GRÁFICO DE CONSERVAÇÃO DO TIPO <i>PIP</i>	19
FIGURA 8 – EXONS / ELEMENTOS ALTERNATIVOS	20
FIGURA 9 – ELEMENTOS ESTIMULADORES OU <i>ENHANCERS</i>	20
FIGURA 10 – FORMULÁRIO DE LOCALIZAÇÃO	22
FIGURA 11 – OPÇÃO <i>BASE GENOME</i>	22
FIGURA 12 – OPÇÃO <i>BROWSER SETTINGS</i>	24
FIGURA 13 – OPÇÃO <i>ECR</i>	25
FIGURA 14 – OPÇÃO <i>DNA</i>	25
FIGURA 15 – OPÇÃO <i>SYNTENY / ALIGNMENTS</i>	27
FIGURA 16 – OPÇÃO <i>ZOOM</i>	27
FIGURA 17 – CROMOSSOMOS DO GENOMA-BASE	28
FIGURA 18 – OPÇÃO <i>GRAB ECR</i>	29
FIGURA 19 – GRÁFICOS <i>PIP</i> E <i>SMOOTH</i>	30
FIGURA 20 – OPÇÃO <i>GENOME ALIGNMENT</i>	31
FIGURA 21 – SEQUÊNCIA DE QUALIDADE 'RASCUNHO'	33
FIGURA 22 – <i>DOT-PLOT</i>	33
FIGURA 23 – BLOCOS DE HOMOLOGIA	34
FIGURA 24 – ÁRVORE FILOGENÉTICA	35
FIGURA 25 – PERFIS DE CONSERVAÇÃO <i>SMOOTH</i> E <i>PIP</i>	36
FIGURA 26 – BARRA DE FERRAMENTAS	36
FIGURA 27 – ALINHAMENTO CODIFICADO POR CORES	37
FIGURA 28 – ANOTAÇÕES DO MULTIF	38

FIGURA 29 – SEQUÊNCIA REFERÊNCIA	39
FIGURA 30 – PIP (<i>PERCENT IDENTITY PLOT</i>)	40
FIGURA 31 – ORGANIZAÇÃO DO LÓCUS DA MIOSTATINA NO GENOMA HUMANO	45
FIGURA 32 – COMPARAÇÃO DO LÓCUS DA MIOSTATINA DE DIFERENTES VERTEBRADOS	47
FIGURA 33 – ANÁLISE REALIZADA PELO PIPMAKER	48
FIGURA 34 – SÍTIOS DE LIGAÇÃO PARA FATORES DE TRANSCRIÇÃO CONSERVADOS NO LÓCUS DA MIOSTATINA	49
FIGURA 35 – ECRs NA REGIÃO A <i>UPSTREAM</i> DO GENE DA MIOSTATINA	50
FIGURA 36 - SÍTIOS DE LIGAÇÃO PARA FATORES DE TRANSCRIÇÃO CONSERVADOS A <i>UPSTREAM</i> DO LÓCUS DA MIOSTATINA	51
FIGURA 37 - ECRs NA REGIÃO A <i>DOWNSTREAM</i> DO GENE DA MIOSTATINA	52
FIGURA 38 - SÍTIOS DE LIGAÇÃO PARA FATORES DE TRANSCRIÇÃO CONSERVADOS A <i>DOWNSTREAM</i> DO LÓCUS DA MIOSTATINA	52

LISTA DE TABELAS

TABELA 1 – LISTA DOS FATORES DE TRANSCRIÇÃO	53
---	----

RESUMO

A miostatina (GDF-8) pertence à superfamília de Fatores de Crescimento e Diferenciação β (TGF- β) e foi identificada em 1997 por MCPHERRON *et al.*. Ao nocautear o gene da miostatina, MCPHERRON *et al.* notaram que mutantes adultos apresentaram um formato anormal de corpo, com ombros e quadris pronunciados, e quando a pele foi removida de animais mortos, era aparente que os músculos dos mutantes eram muito maiores do que os dos selvagens. Esse aumento muscular revelou ser resultado de hiperplasia e hipertrofia. A compreensão do padrão de expressão da miostatina é importante para a tentativa de se criar estratégias de manipulação desse gene *in vitro* e *in vivo*. No presente trabalho, buscou-se localizar, através do uso de ferramentas da Bioinformática, os elementos *cis*-regulatórios (promotor e *enhancers*) que atuam sobre a atividade de transcrição do gene da miostatina. Através da comparação dos lóci do GDF8 humano, de camundongo e galinha, foram encontradas diferentes regiões evolutivamente conservadas (ECRs) entre essas espécies que, por terem permanecido intactas ao longo da evolução, representam fortes candidatas a elementos regulatórios. No locus da miostatina, foi observado que há uma alta identidade de seqüência entre humano, camundongo e galinha na região dos três exons. Entretanto, com relação aos introns, a similaridade foi elevada quando a comparação foi realizada apenas entre os dois mamíferos; apesar de o nível de similaridade entre humano, camundongo e galinha ter sido baixo nessas regiões, foi identificado um sítio de ligação conservado entre as três espécies para o fator de transcrição XPF1 no primeiro intron. Na região *upstream* ao gene GDF8 foi identificada apenas uma ECR, localizada imediatamente adjacente ao gene. Ao realizar uma busca por sítios de ligação para fatores de transcrição conservados nessa região, foi detectada a presença, dentre outros, de um sítio para o fator TBP (*Tata Binding Protein*), o que reforça a hipótese de que essa região corresponda ao promotor do gene da miostatina. A busca por regiões conservadas entre as três espécies na região *downstream* ao GDF8 revelou a presença de três ECRs. Na ECR1 não foi identificado nenhum sítio de ligação para fatores de transcrição, ao contrário da ECR2. Nesta foram encontrados 15 tipos de sítios, entre eles sítios para fatores relacionados ao desenvolvimento da musculatura esquelética como *MyoD* e miogenina que interagem com o E-box, além de sítios para Pax3. Ainda na ECR2 foram encontrados sítios para fatores relacionados ao desenvolvimento cardíaco, como TEF1 e TBX5. Já na ECR3 foram identificados apenas dois sítios de ligação para fatores de transcrição, dentre eles um sítio para ERR1, que é super-expresso durante diferenciação de adipócitos, indicando que essa ECR possa ser a responsável pela regulação da expressão da miostatina em adipócitos, enquanto que a ECR2 possa regular a transcrição da miostatina em musculatura esquelética e no coração. Os resultados teóricos desse trabalho serão, no futuro, testados com o auxílio de experimentos práticos que possam vir a confirmar os dados encontrados.

1. INTRODUÇÃO

1.1 FORMAÇÃO DA MUSCULATURA ESQUELÉTICA DURANTE A EMBRIOGÊNESE

A capacidade de crescimento de um animal depende em grande parte do número de fibras musculares formadas durante o desenvolvimento embrionário. Em aves e mamíferos, este número mantém-se praticamente inalterado após o nascimento (SWATLAND & KIEFFER 1974; SWATLAND 1984; KONIGSBERG 1986) e, conseqüentemente, o aumento no número de fibras musculares (hiperplasia) não pode concorrer para ganhos de massa muscular. Assim, nos organismos adultos, apenas o aumento do diâmetro das fibras musculares (hipertrofia) pode resultar em ganhos significativos da massa muscular (REHFELDT *et al.* 2000).

As fibras da musculatura esquelética são geradas a partir de blocos segmentares de mesoderma paraxial, conhecidas como somitos (CHRIST & ORDAHL 1995). Durante a gastrulação, o mesoderma paraxial é depositado como uma faixa de células mesenquimais, que se dispõe lateralmente às estruturas axiais do embrião (tubo neural e notocorda) e é delimitado pelo mesoderma intermediário e lateral (STOCKDALE *et al.* 2000) (Figura 1).

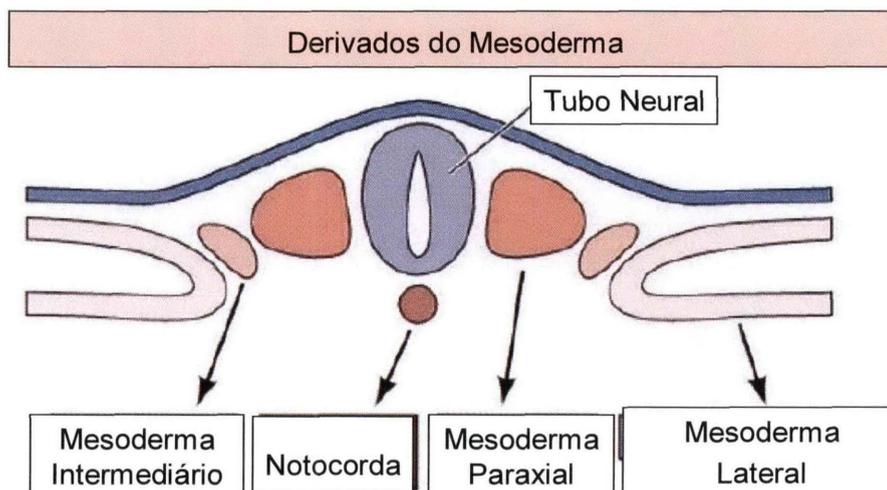


Figura 1. Localização do mesoderma paraxial. As fibras da musculatura esquelética são geradas a partir de blocos segmentares de mesoderme paraxial, conhecidas como somitos. Durante a gastrulação, a mesoderme paraxial é depositada como uma faixa de células mesenquimais, que se dispõe lateralmente às estruturas axiais do embrião (tubo neural e notocorda) e é delimitada pela mesoderme intermediária e lateral.

O mesoderma paraxial pode ser dividido, na direção rostro-caudal, em mesoderma cefálico, somitos, e placa segmentar. O mesoderma cefálico é localizado rostralmente à vesícula ótica e não sofre segmentação. Ele dá origem aos músculos externos do olho, à musculatura da mandíbula, e a alguns ossos do crânio. O mesoderma da placa segmentar surge durante a gastrulação, quando as células do epiblasto invaginam através do nódulo de Hensen e ao longo da linha primitiva (STOCKDALE *et al.* 2000). Conforme o desenvolvimento embrionário progride, a linha primitiva adiciona continuamente material à extremidade posterior do mesoderma paraxial, enquanto células da extremidade anterior condensam-se, formando os somitos (STOCKDALE *et al.* 2000) a uma razão de, aproximadamente, um par a cada uma hora e meia, a partir do estágio 14 de HH (HAMBURGER & HAMILTON 1951) (STOCKDALE *et al.* 2000). A formação e a maturação dos pares de somitos ocorre da região cervical em direção à região caudal do embrião (AOYAMA & ASAMOTO 1988; ORDAHL & LE DOUARIN 1992) (Figura 2). Cada par de somitos sofre um processo de maturação que leva à formação de todo o esqueleto axial, a musculatura do tronco e dos membros (STOCKDALE *et al.* 2000).

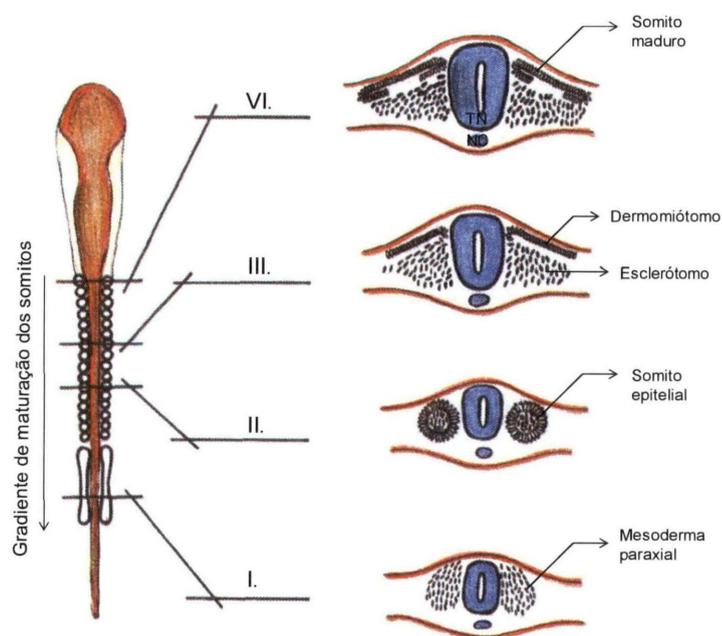


Figura 2. A formação e a maturação dos pares de somitos ocorre da região cervical em direção à região caudal do embrião. Em I., observa-se o mesoderma paraxial como uma faixa de células mesenquimais; em II., já houve a formação dos somitos epiteliais; em III., a porção ventral dos somitos sofreu uma desepitelialização, originando o esclerótomo, enquanto que a porção dorsal, ou dermatomiótomo, permanece como uma camada epitelial; em IV., células do dermatomiótomo desceram e originaram o miótomo (hepaxial e hipaxial).

Dorsalmente, os somitos são cobertos pela ectoderme da superfície e, ventralmente, são opostos à endoderme e à aorta. Lateralmente, o somito está em contato com células do mesoderma intermediário. Uma lâmina basal circunda cada somito recém-formado, conectando-o com as estruturas ao seu redor (STOCKDALE *et al.* 2000).

Inicialmente, o somito é composto de células colunares arranjadas radialmente ao redor de um lúmen. Esse somito epitelial é padronizado rapidamente em compartimentos distintos que, subseqüentemente, dão origem a linhagens celulares distintas. Em resposta a sinais vindos de tecidos adjacentes, a porção ventral dos somitos epiteliais sofre um processo de desepitelialização, para formar o esclerótomo mesenquimal, enquanto que a porção dorsal, ou dermomiótomo, permanece como uma camada epitelial (BRENT & TABIN 2002). O dermomiótomo apresenta subdomínios que abrigam diferentes populações de células precursoras musculares, que irão formar a musculatura esquelética epaxial e hipaxial. Enquanto as células da porção dorso-medial do dermomiótomo originam a musculatura epaxial, representada pelos músculos vertebrais e das costas, as células da porção lateral formam a musculatura hipaxial, que corresponde à musculatura da parede corpórea ventral, dos membros, da língua e, nos mamíferos, a do diafragma. Com exceção dos músculos da cabeça e pescoço, que originam-se do mesoderma paraxial anterior aos somitos e do mesoderma pré-cordal, todos os demais músculos esqueléticos dos organismos vertebrados derivam dos somitos (CHRIST & ORDAHL 1995).

O desenvolvimento da musculatura epaxial tem início quando células do lábio dorso-medial (LDM) dos somitos expandem-se sob o dermomiótomo, formando um folheto contínuo que origina o miótomo epaxial. Esta expansão permite que os mioblastos entrem no miótomo na posição correta para formar a musculatura epaxial. Uma vez atingindo o local apropriado, estas células interrompem o ciclo celular, alongam-se e diferenciam-se, formando as primeiras células musculares do embrião (DENETCLAW *et al.* 1997; KAHANE *et al.* 1998).

A musculatura hipaxial não é uma entidade homogênea e origina-se a partir de dois mecanismos distintos. Na região torácica, o lábio ventro-lateral (LVL) do dermomiótomo permanece arranjado como um epitélio durante o alongamento do dermomiótomo, o que leva à deposição do miótomo hipaxial por um mecanismo semelhante à formação do miótomo epaxial. Neste processo, as porções medial e lateral do miótomo misturam-se, formando uma lâmina contínua de mioblastos. Genes específicos são ativados nestes mioblastos levando à diferenciação muscular. Sob o controle de sinais locais, estas células irão dar origem à musculatura da parede ventral do corpo. Ao nível dos membros e nas regiões occipital e cervical ocorre um mecanismo distinto, que tem início com a desagregação de células do LVL do dermomiótomo, após a separação entre dermomiótomo e esclerótomo. Nas regiões adjacentes aos membros, estas células precursoras miogênicas migram lateralmente em direção a somatopleura para formar musculatura. Occipitalmente, os mioblastos encaminham-se para a base dos arcos branquiais, formando a musculatura da língua. E, nos mamíferos, células situadas acima do botão do membro anterior seguem uma rota posterior e ventral para formar a musculatura do diafragma (DENETCLAW *et al.* 1997; CINNAMON *et al.* 1999).

Apesar dos aspectos particulares, o desenvolvimento das musculaturas epaxial e hipaxial compartilha de mecanismos celulares e moleculares muito semelhantes. Em ambos os casos, a diferenciação das células musculares é acompanhada pela interrupção do ciclo celular

e pela fusão de mioblastos individuais para a produção das fibras musculares multinucleadas, bem como pela ativação de genes músculo-específicos que codificam proteínas estruturais e contráteis. Neste processo, os principais reguladores da miogênese, os fatores miogênicos, se associam com fatores ativadores ou inibidores, em um mecanismo complexo e requintado de controle da expressão gênica durante a ontogenia (LUDOLPH & KONIECZNY, 1995).

1.2 O FATOR GDF-8

Nos últimos anos, a identificação de um novo fator de crescimento e diferenciação denominado miostatina (GDF-8), tem proporcionado grandes avanços para um melhor entendimento das bases moleculares envolvidas na formação e deposição de tecido muscular esquelético em vertebrados.

A miostatina (previamente chamada GDF-8) foi originalmente identificada em uma busca por novos membros da superfamília de fatores de crescimento e diferenciação β (TGF- β) em 1997, por MCPHERRON *et al.* A superfamília de fatores de crescimento TGF- β inclui mais de 30 proteínas de mamíferos estruturalmente relacionadas que têm diversas funções durante o desenvolvimento embrionário e na manutenção da homeostase de tecidos em adultos (LIN *et al.* 2003).

1.3. DESCOBERTA DO GDF-8

Para identificar novos membros da superfamília TGF- β , MCPHERRON *et al.* (1997) construíram oligonucleotídeos degenerados, correspondendo à regiões conservadas dentre os membros conhecidos dessa família, e os usaram como *primers* em PCR (reação em cadeia da polimerase) sobre o DNA genômico de camundongo. Dentre as seqüências obtidas nessa seleção, estava uma designada GDF-8 (*growth and differentiation factor 8*). A seqüência de aminoácidos prevista para GDF-8 continha todas as marcas características dos membros da superfamília TGF- β , incluindo uma seqüência sinal para secreção, um sítio de processamento proteolítico e uma região carboxi-terminal contendo o padrão conservado de resíduos de cisteínas (**Figura 3**). Entretanto, com base numa comparação entre as regiões C-terminais do GDF-8 com as de outros membros da superfamília TGF- β , o GDF-8 parecia não pertencer a nenhuma das subfamílias conhecidas de fatores altamente relacionados, incluindo os TGF- β s, inibinas e BMPs (MCPHERRON *et al.* 1997).

134 **atg**caaaaactgcaactctgtgtttatatttacctgtttatgctg
M Q K L Q L C V Y I Y L F M L
 179 attgttgctgggtccagtgatctaaatgagaacagtgagcaaaaa
 I V A G P V D L N E N S E Q K
 224 gaaaatgtggaaaaagaggggctgtgtaatgcatgtacttggaga
 E N V E K E G L C N A C T W R
 269 caaaacttaaatcttcaagaatagaagccattaagatacaaatc
 Q N T K S S R I E A I K I Q I
 314 ctcagtaaacttcgtctggaacagctcctaactcagcaaatg
 L S K L R L E T A P N I S K D
 359 gttataagacaactttttaccctcctccactccgggaactg
 V I R Q L L P K A P P L R E L
 404 attgatcagtatgatgtccagagggatgacagcagcgatggctct
 I D Q Y D V Q R D D S S D G S
 449 ttggaagatgacgattatcacgctacaacggaacaatcattacc
 L E D D D Y H A T T E T I I T
 494 atgcctacagagtctgattttctaatgcaagtggatggaaaacc
 M P T E S D F L M Q V D G K P
 539 aaatgttgcttctttaaatttagctctaaaatacaatacaataaa
 K C C F F K F S S K I Q Y N K
 584 gtagtaaaggccaactatggatatatttgagaccctcgagact
 V V K A Q L W I Y L R P V E T
 629 cctacaacagtgtttgtgcaaatcctgagactcatcaaacctatg
 P T T V F V Q I L R L I K P M
 674 aaagacggtacaaggtatactggaatccgatctctgaaacttgac
 K D G T R Y T G I R S L K L D
 719 atgaaccagcactggtatttggcagagcattgatgtgaagaca
 M N P G T G I W Q S I D V K T
 764 gtggtgcaaaattggctcaacaacctgaatccaacttaggcatt
 V L Q N W L K Q P E S N L G I
 809 gaaataaaagcttttagatgagaatggatcatgacttggctgtaacc
 E I K A L D E N G H D L A V T
 854 ttcccaggaccaggagaagatgggctgaatccgtttttagaggtc
 F P G P G E D G L N P F L E V
 899 aaggtaacagacacaccaaaaagatccagaagggattttggtcct
 K V T D T P K **R S R R** D F G L
 944 gactgtgatgagcactcaacagaatcacgatgctgctcgttacct
 D **C** D E H S T E S R **C C** R Y P
 989 ctaactgtggattttgaagcttttggatgggattggattatcgct
 L T V D F E A F G W D W I I A
 1034 cctaaaagatataaggccaattactgctctggagagtgatgaattt
 P K R Y K A N Y **C S G E C E F**
 1079 gtatttttacaaaaatatacctcactcatctggtacaccaagca
 V F L Q K Y P H T H L V H Q A
 1124 aaccccagaggttcagcagcccttgctgtactcccacaagatg
 N P R G S A G P **C C** T P T K M
 1169 tctccaattaatagctatattttaatggcaagaacaataata
 S P I N M L Y F N G K E Q I I
 1214 tatgggaaaattccagcagatggttagtagaccgctgtgggtgctca
 Y G K I P A M V V D R **C G C S**
 1259 **tga 1261**
 *

Figura 3 – Sequência de aminoácidos da miostatina humana, prevista pelo programa ORF Finder – “Open Reading Frame Finder”. (<http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>). A proteína prevista apresenta 375 aa e contém um sítio de processamento proteolítico (destacado em vermelho), além de nove resíduos de cisteína conservados na região C-terminal (em azul), típicos dos membros da família TGF- β .

Como a maioria dos membros da família TGF- β , o GDF-8 aparenta ser altamente conservado dentre as espécies. Por análise genômica do tipo *Southern-blot*, seqüências homólogas foram detectadas em todas as espécies de mamíferos analisadas e também em galinhas. Na maioria das espécies, a sonda GDF-8 também detectou um segundo fragmento levemente hibridizado. Outras análises das seqüências homólogas de GDF-8 em DNA de murino e humano revelaram que esse segundo fragmento corresponde a um gene altamente relacionado, o GDF-11.

Para determinar se os sinais de processamento no GDF-8 são funcionais e se o GDF-8 forma dímeros, como os demais membros da superfamília TGF- β , o cDNA do GDF-8 foi expresso estavelmente em células de ovário de hamster chinês (CHO). Análise do tipo *Western* de meio preparado a partir dessas células, usando um anticorpo criado contra o fragmento C-terminal do GDF-8 expresso por bactéria, detectou duas espécies de proteínas com massas moleculares relativas de aproximadamente 52 e 15 kilodaltons, sob condições redutoras (proteína na forma linear), consistente com formas não processadas e processadas do GDF-8, respectivamente.

Sob condições não-redutoras (que permite a formação de dímeros), foram detectadas duas espécies de proteína, com massas moleculares relativas de 101 e 25 kilodaltons, consistentes com as formas diméricas de formas processadas e não processadas do GDF-8, respectivamente. Assim, como outros membros da superfamília TGF- β , o GDF-8 parece ser secretado e proteoliticamente processado, e a região C-terminal aparenta ser capaz de formar dímeros ligados por pontes disulfeto.

Para determinar o padrão de expressão do GDF-8, foi realizada uma hibridização *in situ* em embriões de camundongo, isolados em vários estágios de desenvolvimento. Em todos os estágios examinados, a expressão do mRNA do GDF-8 parecia ser restrita a músculos esqueléticos em desenvolvimento. Em estágios mais iniciais, a expressão do GDF-8 era restrita a somitos em desenvolvimento. Por hibridização *in situ* em embriões inteiros, o mRNA do GDF-8 pôde ser detectado já no dia 9,5 pós-coito em um terço dos somitos. Nesse estágio de desenvolvimento, a hibridização pareceu ser restrita aos somitos rostrais mais maduros. Por volta do dia 10,5 pós-coito, a expressão do GDF-8 era claramente evidente em quase todos os somitos. Com base nas análises da hibridização *in situ* de cortes preparados de embriões do dia 10,5 pós-coito, a expressão do GDF-8 em somitos pareceu ser localizada no compartimento do miótomo. Em estágios mais avançados de desenvolvimento, a expressão foi detectada em uma grande variedade de músculos em desenvolvimento.

Segundo MCPHERRON *et al.* (1997), o GDF-8 continua a ser expresso em animais adultos. Através de análise do tipo *Northern-blot*, a expressão de mRNA do GDF-8 foi vista quase que exclusivamente em musculatura esquelética, dentre os tecidos adultos examinados. Um sinal significativamente menor, porém claramente detectável, foi visto também em tecido adiposo. Com base na análise do tipo *Northern-blot* de RNA preparado de um grande número de diferentes músculos esqueléticos adultos, pode-se verificar que a expressão de GDF-8 pareceu ser generalizada neste tecido, apesar de os níveis variarem dentre músculos individuais.

Para determinar a função biológica do GDF-8, MCPHERRON *et al.* (1997) romperam o gene da miostatina por recombinação homóloga em células-tronco embrionárias de camundongos. Após injeção dos clones recombinantes em blastocistos, as quimeras geraram filhotes heterozigotos quando cruzados com fêmeas da linhagem C57BL/6. Após análises genóticas dos filhotes derivados dos cruzamentos entre F1 heterozigotos, observou-se que a razão de genótipos foi próxima da esperada (1:2:1), mas o número reduzido de homozigotos demonstrou ser estatisticamente significativo.

Muitos homozigotos eram viáveis e férteis quando cruzados com camundongos C57BL/6 e uns com os outros. Entretanto eles eram aproximadamente maiores que seus irmãos heterozigotos e selvagens. A diferença entre mutantes e selvagens pareceu ser relativamente constante, indiferente à idade e sexo em animais adultos. Mutantes adultos apresentaram também um formato anormal de corpo, com ombros e quadris pronunciados. Quando a pele foi removida de animais mortos, era aparente que os músculos dos mutantes eram muito maiores do que os dos selvagens. O aumento na massa muscular esquelética pareceu ser espalhado por todo o corpo. Músculos individuais isolados de animais mutantes homozigotos pesaram de duas a três vezes mais que aqueles isolados de animais selvagens (Figura 4).

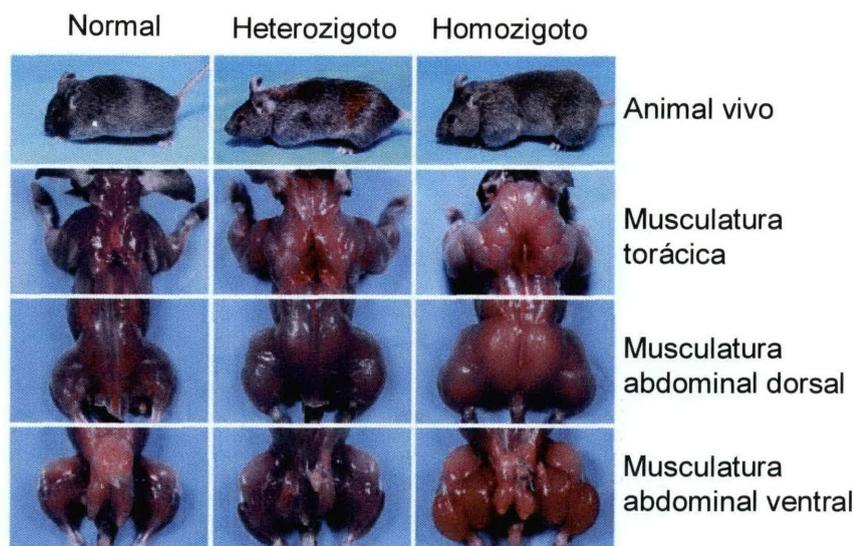


Figura 5. Comparação entre camundongos normais, heterozigotos e homozigotos para mutação no gene da miostatina. Mutantes adultos apresentaram um formato anormal de corpo, com ombros e quadris pronunciados. Quando a pele é removida de animais mortos, os músculos dos mutantes são visivelmente muito maiores do que os dos selvagens, sendo que o aumento na massa muscular esquelética parece ser espalhado por todo o corpo.

Para determinar se o aumento na massa muscular poderia sozinho acarretar na diferença dentre os pesos corporais totais entre mutantes e selvagens, ou se muitos tecidos eram de forma geral maiores nos mutantes, MCPHERRON *et al.* (1997) compararam o peso total corporal com o peso das carcaças. A diferença nos pesos das carcaças entre selvagens e mutantes foi comparável à diferença do peso corporal total.

Para determinar se o aumento na massa muscular esquelética resultou de hiperplasia ou hipertrofia, MCPHERRON *et al.* (1997) realizaram análises histológicas de diferentes grupos musculares. O músculo mutante mostrou-se normal. Nenhum excesso de tecido conjuntivo ou adiposo foi observado e não houve sinais óbvios de degeneração, como grande variação no tamanho das fibras ou núcleos centralizados. Uma quantificação do número de fibras musculares mostrou que na porção mais larga do músculo *tibialis cranialis*, o número total de células foi 86% maior em mutantes. A quantidade de DNA extraída do músculo mutante foi aproximadamente 50% maior que do selvagem. Assim, uma grande parte do aumento da massa muscular esquelética resultou de hiperplasia. Entretanto, a hipertrofia também pareceu contribuir para o aumento da massa muscular. A média do diâmetro das fibras do músculo *tibialis cranialis* e do *gastrocnemius* foi de, respectivamente, sete a 22% maior em mutantes, sugerindo que a área total das fibras musculares foi aumentada em 14 e 49%, respectivamente.

Com base nessas primeiras evidências, estudos posteriores observaram que alterações naturais na seqüência codificadora dessa proteína estariam associadas ao fenótipo

de musculatura dupla em raças bovinas, e em camundongos selecionados para acentuada deposição de massa muscular (MCPHERRON *et al.* 1997; GROBET *et al.* 1997; GROBET *et al.* 1998).

1.4 SÍNTESE

Como outros membros da família TGF- β , a miostatina é sintetizada como uma proteína precursora (Figura 5) que sofre dois eventos de processamento proteolítico para gerar a molécula biologicamente ativa. O primeiro evento remove a seqüência sinal N-terminal de 24 aminoácidos, necessária para enviar a proteína para vias de secreção. O segundo evento gera, através de clivagem, o fragmento C-terminal que possui atividade de ligação com o receptor. Após o processamento proteolítico, o pró-peptídeo e o dímero C-terminal, ligado por pontes disulfeto, permanecem ligados não-covalentemente em um complexo latente (LEE 2004).

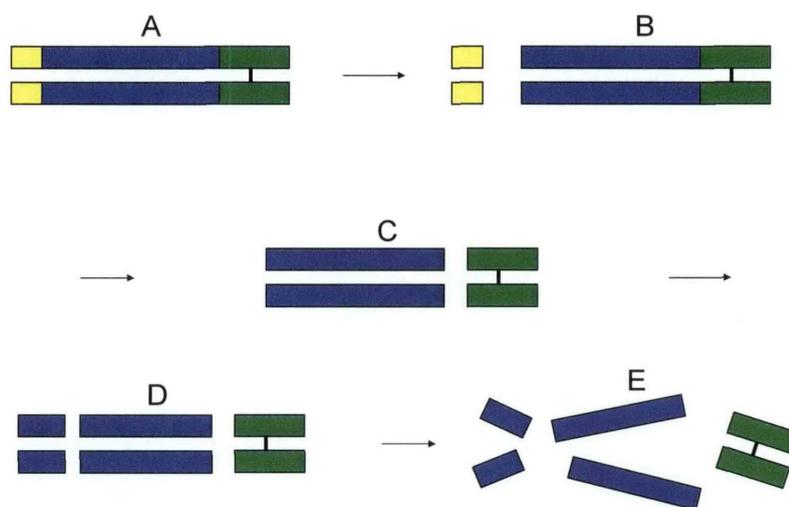


Figura 5. A miostatina é sintetizada como uma proteína precursora (A) que sofre dois eventos de processamento proteolítico para gerar a molécula biologicamente ativa. O primeiro evento (B) remove a seqüência sinal N-terminal de 24 aminoácidos (em amarelo), necessária para enviar a proteína para vias de secreção. O segundo evento (C) gera, através de clivagem, o fragmento C-terminal (em verde) que possui atividade de ligação com o receptor. Após o processamento proteolítico, o pró-peptídeo (em azul) e o dímero C-terminal, ligado por pontes disulfeto, permanecem ligados não-covalentemente em um complexo latente. A ativação desse complexo latente ocorre, provavelmente, após a quebra do pró-peptídeo (D) e, só então, o dímero C-terminal torna-se biologicamente ativo (E).

Com base no que se sabe sobre outros membros da família TGF- β , o pró-peptídeo desempenha um importante papel no dobramento apropriado do domínio C-terminal. O pró-peptídeo também desempenha um papel importante na regulação da atividade do dímero C-terminal. Camundongos transgênicos com super-expressão do pró-peptídeo em músculos,

possuem um fenótipo similar a aqueles nocauteados para a miostatina, em termos de aumento da massa muscular (LEE 2004).

Em adição ao pró-peptídeo, várias outras proteínas também se mostraram capazes de se ligar e inibir a atividade do dímero C-terminal da miostatina. Uma dessas proteínas é a folistatina. Ela é capaz de bloquear a atividade da miostatina em testes com ligação de receptores e genes repórteres e também em camundongos que receberam transplante de células expressando miostatina. Estudos com embriões de galinha mostraram que a folistatina é expressa em células do miótomo e em musculatura em desenvolvimento em estágios em que a miostatina também é expressa e que implantes de *beads* carregadas com folistatina em botões de asa em desenvolvimento podem causar um aumento da expressão da Pax-3 e atraso na diferenciação das células musculares (LEE 2004).

Estudos genéticos em camundongos mostraram que uma super-expressão de folistatina em músculos pode causar aumentos dramáticos no crescimento muscular e que a perda da folistatina resulta em redução da massa muscular já no nascimento (LEE 2004).

Além da folistatina, outras proteínas do plasma também interagem com a miostatina para regulá-la negativamente, incluindo o gene folistatina-relacionado (FLRG), gene específico da interrupção do crescimento (GAS-1) e teletonina (TELLGREN *et al.* 2004).

A super-expressão de proteínas que se ligam a miostatina e formas mutantes da miostatina que não podem ser clivadas, também resultam em camundongos com massa muscular aumentada. Recentemente, camundongos transgênicos foram gerados usando uma mutação do tipo *missense* que atua de maneira dominante negativa. Esses camundongos têm massa muscular esquelética aumentada como resultado de hiperplasia muscular, mas sem hipertrofia. Outros transgênicos da miostatina produzem massa muscular aumentada por hipertrofia e não hiperplasia (TOBIN & CELESTE 2005).

É possível que a miostatina circule como complexos latentes múltiplos e que cada um seja ativado independentemente em resposta a diferentes estímulos. É possível também que o complexo latente possa consistir de uma estrutura de maior ordem, contendo todas essas proteínas (TOBIN & CELESTE 2005).

Entender como a miostatina é ativada desse estado latente será crítico não apenas para saber como a atividade da miostatina é regulada, mas também para identificar novas estratégias para o desenvolvimento de agentes terapêuticos para aplicações clínicas (LEE 2004).

1.5 SINALIZAÇÃO

O dímero C-terminal da miostatina é capaz de se ligar a receptores e ativar uma cascata de sinalização na célula-alvo, como fazem os membros da família TGF- β (LEE 2004). Os membros da superfamília TGF- β utilizam um mecanismo comum de sinalizar ao núcleo. Eles se ligam a receptores na membrana celular, que possuem um domínio intracelular quinase

serina-treonina, levando à reunião de um complexo receptor que fosforila proteínas da família SMAD. As SMADs, então, vão para o núcleo, se ligam ao DNA e recrutam os co-ativadores ou co-repressores transcricionais requeridos para controlar a expressão gênica. Há dois tipos de receptores aos quais se ligam os membros da superfamília TGF- β , chamados receptores do tipo I e tipo II. Geralmente, cada ligante têm seu receptor específico; entretanto, alguns receptores podem ser compartilhados por diferentes ligantes. Quando o ligante se liga ao receptor, ocorre a união de dois receptores do tipo I e dois receptores do tipo II para formar um complexo receptor ativado. Nesse complexo, o receptor de tipo I é fosforilado pelo receptor de tipo II que, por sua vez, fosforila as proteínas SMAD que transmitem o sinal (LIN *et al.* 2003) (Figura 6).

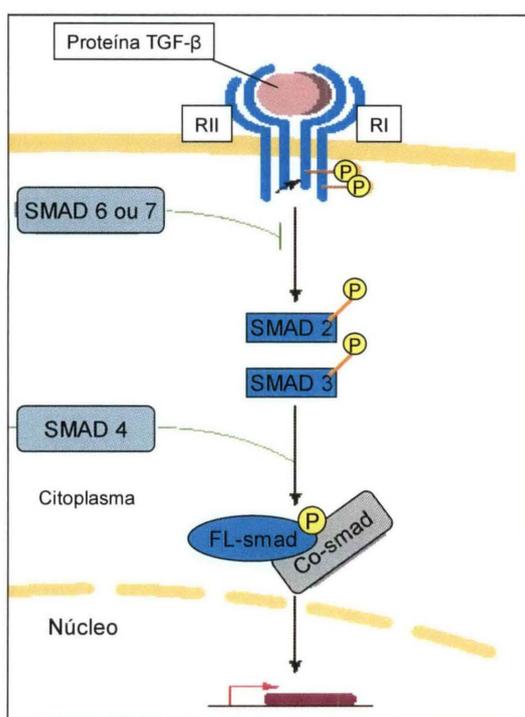


Figura 6. Sinalização da miostatina. Os membros da família TGF- β se ligam a receptores (tipo I / RI ou tipo II / RII) na membrana celular. Ocorre, então, a união de dois receptores do tipo I e dois receptores do tipo II para formar um complexo receptor ativado. Nesse complexo, o receptor de tipo I é fosforilado pelo receptor de tipo II que, por sua vez, fosforila as proteínas SMAD. As SMADs, então, vão para o núcleo, se ligam ao DNA e recrutam os co-ativadores ou co-repressores transcricionais requeridos para controlar a expressão gênica.

1.6 MODO DE AÇÃO

A miostatina é expressa inicialmente no compartimento do miótomo de somitos em desenvolvimento, atuando na regulação do número final de fibras musculares que são formadas (LEE 2004) e continua a ser expressa na linhagem miogênica ao longo do desenvolvimento e em animais adultos (LEE & MCPHERRON 2001) quando é produzida pela musculatura esquelética, circula no sangue e atua para limitar o crescimento das fibras musculares. Em animais adultos, observam-se também claros níveis de RNA da miostatina

presentes em tecido adiposo. Camundongos adultos sem miostatina têm uma redução na gordura corporal total, que é particularmente acentuada em animais mais velhos (LEE 2004).

Apesar de muitos detalhes com relação ao mecanismo de ação da miostatina e os mecanismos pelos quais a atividade da miostatina é regulada *in vivo* permanecem por serem elucidados, é agora firmemente estabelecido que a miostatina é sintetizada pela musculatura esquelética, circula no sangue e atua em uma maneira concentração-dependente como um regulador negativo do crescimento muscular (LEE 2004).

Os efeitos sobre o tamanho das fibras musculares em animais adultos devem resultar de ação da miostatina sobre células satélite. Essas células são mononucleadas, encontradas entre a membrana basal e o sarcolema das fibras musculares, e acredita-se que elas representam a população de células-tronco da musculatura. As células satélite são normalmente quiescentes mas podem entrar no ciclo celular e se fundir com fibras existentes durante o crescimento muscular. Experimentos *in vitro* demonstram que a miostatina inibe a ativação e auto-renovação das células satélite (TOBIN & CELESTE 2005).

Um modelo especulativo é de que níveis circulantes de miostatina mudam sob diferentes condições ambientais, como temperatura e disponibilidade de alimento, ou sob diferentes estados psicológicos, como atividade física, doença, gravidez ou idade, e de que essa alternância de níveis modifique o balanço homeostático metabólico entre estoque de gordura e crescimento muscular. Sob certas condições, pode ser mais vantajoso mudar esse balanço em direção ao estoque de gordura, e sob outras condições, pode ser mais vantajoso mudar esse balanço em direção ao crescimento muscular. Esse tipo de regulação já foi observado em vários estudos (LEE 2004).

As funções biológicas da miostatina levantaram a possibilidade que modificar a via de sinalização mediada pela miostatina, pode ser uma estratégia efetiva para aumentar o crescimento muscular para uma variedade de aplicações clínicas (LEE 2004).

Estabelecer o modo de ação e regulação da miostatina deve trazer efeitos benéficos em termos de manutenção da massa muscular ou promoção do crescimento muscular em casos de degeneração muscular crônica, como a sarcopenia, que é a perda progressiva de musculatura que ocorre na velhice, ou em pacientes com doenças degenerativas da musculatura, como distrofia muscular. Inibição da miostatina pode também ter efeitos benéficos em estados de doenças onde a perda de musculatura é mais aguda. Um exemplo é a perda de musculatura que ocorre na caquexia, que é uma síndrome degenerativa freqüentemente vista em pacientes com doenças crônicas como câncer, AIDS, e sepsis e que contribui para a morbidez e mortalidade. Inibição da atividade miostatina pode também ter efeitos benéficos no tratamento de doenças metabólicas, como a obesidade e a diabetes tipo II (LEE 2004).

Claramente muito trabalho ainda é necessário para determinar se interferir com a miostatina traria efeitos benéficos para doenças humana. Em caso positivo, um número de

considerações faz da miostatina um alvo particularmente atraente para o desenvolvimento de drogas:

- 1º) Os efeitos da perda da miostatina são altamente específicos, uma vez que efeitos significativos em outros tecidos, que não o muscular e o adiposo, não foram reportados ainda;
- 2º) A miostatina normalmente funciona extracelularmente e é, dessa forma, acessível a uma variedade de agentes farmacológicos em potencial, incluindo proteínas capazes de se ligar a ela;
- 3º) O efeito da miostatina depende da dose, sendo que a inibição parcial da sua atividade resulta em um aumento parcial do crescimento muscular; assim, pode não ser necessário bloquear completamente a sinalização da miostatina para produzir um efeito terapêuticamente benéfico (LEE 2004).

1.7 FATORES DE TRANSCRIÇÃO, PROMOTOR E ENHANCERS

Apesar de necessárias, as RNA polimerases não são suficientes para a regulação transcricional. São requeridas muitas outras proteínas, geralmente chamadas fatores de transcrição gerais, que ajudam a posicionar o complexo da polimerase corretamente. Estes fatores provavelmente também participam de diversas outras funções: regular as interações histonas-DNA nas proximidades da transcrição, afrouxar a dupla-hélice e permitir que o complexo transcricional prossiga a partir do promotor basal. O promotor basal (ou simplesmente promotor) é a seqüência de DNA específica onde a RNA polimerase e os fatores transcricionais gerais se conectam. O promotor está localizado adjacientemente ou próximo ao sítio de iniciação da transcrição (WILT & HAKE 2004).

A maioria dos promotores de genes eucarióticos possui uma pequena seqüência, de aproximadamente 25 nucleotídeos, localizada *upstream* do sítio de iniciação da transcrição, chamada sítio TATA ou TATA Box, pois é composta de adeninas (A) e timinas (T). Quando o fator de transcrição TFIID (ou TBP) interage com o DNA no TATA Box, o DNA altera a sua conformação espacial, fornecendo um "sinal" que pode auxiliar outras proteínas e a RNA polimerase a se ligarem apropriadamente ao promotor (WILT & HAKE 2004).

Proteínas ativadoras ou supressoras geralmente atuam no promotor ou muito próximas a este, impedindo ou incentivando o acesso da polimerase ao promotor. Em eucariotos, a maioria dos genes é regulada por interações proteínas-DNA. Muitas destas interações ocorrem a centenas, às vezes milhares de pares de base de distância, *upstream* ou *downstream* do gene, em regiões do DNA conhecidas como *enhancers* ou estimuladores da transcrição. Exatamente como essas interações remotas entre proteínas reguladoras e DNA regulador afetam a atividade do promotor, não é entendido em detalhes. Há evidências, a partir de diversos estudos, de que a proteína ativadora ou supressora interage fisicamente com a RNA polimerase e, dessa forma, influencia as atividades desta no promotor (WILT & HAKE 2004).

Podem haver diferentes ativadores e supressores atuando sobre um determinado promotor. Diferentes combinações de fatores fornecem uma maneira de regular a atividade específica de um único gene ou um grupo pequeno de genes (WILT & HAKE 2004).

1.8 BIOINFORMÁTICA

A formação de diferentes tipos celulares a partir de um grupo invariável de genes é governada por processos bioquímicos que regulam a atividade gênica. A transcrição, que representa o passo inicial da expressão gênica, é modulada pela interação entre fatores de transcrição que se ligam aos elementos *cis*-regulatórios no DNA, co-fatores adicionais e a influência da estrutura da cromatina. Proteínas trans-ativadoras que controlam o nível de transcrição de cada gene, ligam-se a seqüências *cis*-regulatórias no DNA. Um entendimento completo da interação entre fatores transcripcionais e seqüências *cis*-regulatórias (promotores e *enhancers*) transformaria a pesquisa biológica, fornecendo os meios para interpretar e modelar as respostas das células a diversos estímulos (WASSERMAN & SANDELIN 2004). Contudo, o completo entendimento da rede de regulação gênica ainda permanece distante.

Apesar disto, nos últimos anos surgiram poderosos métodos computacionais para identificar as seqüências regulatórias responsáveis pelo controle da taxa de transcrição de genes de interesse. Para os pesquisadores, estes métodos de bioinformática representam uma ferramenta valiosa para entender a regulação gênica de uma forma focada e acelerar os estudos nesta área. Em geral, os programas disponíveis para a identificação de elementos regulatórios putativos, tais como Mulan (*Multiple-sequence alignment analysis*: OVCHARENKO *et al.* 2005) e Pipmaker (SCHWARTZ *et al.* 2000), baseiam-se no alinhamento de seqüências do genoma de diferentes grupos de organismos filogeneticamente distantes, buscando identificar regiões que tenham sido evolutivamente conservadas. A manutenção destas seqüências conservadas ao longo da evolução sugere que elas tenham um papel funcional, provavelmente regulando a transcrição gênica. Outros programas, tais como MultiTF (OVCHARENKO *et al.* 2005), buscam identificar potenciais sítios de ligação para fatores de transcrição que tenham sido conservados filogeneticamente, reduzindo em até 95% a chance de falso-positivos. Estes métodos são suficientemente eficientes para direcionar investigações laboratoriais (WASSERMAN & SANDELIN 2004), que venham a validar os achados em bioinformática. Estes programas estão disponíveis *on-line* e podem ser utilizados gratuitamente.

2. OBJETIVOS

1. Identificar seqüências que tenham sido conservadas evolutivamente (ECRs) no locus da Miostatina, comparando genomas de diferentes organismos vertebrados.
2. Elaborar um diagrama representando os ECRs, que correspondam a potenciais elementos regulatórios da miostatina (promotor e *enhancers*).
3. Localizar os sítios de ligação para fatores de transcrição, dentro dos ECRs, que sejam conservados filogeneticamente.
4. Classificar os fatores de transcrição identificados, segundo sua categoria e função.
5. Disponibilizar em português as instruções das ferramentas de Bioinformática utilizadas, de forma a facilitar o acesso às mesmas por estudantes e demais interessados.

3. METODOLOGIA

Esse trabalho foi realizado com o auxílio de ferramentas da bioinformática e bancos de dados. Esses programas estão disponíveis *on-line* e podem ser utilizados gratuitamente.

3.1 COMPARAÇÃO DE SEQÜÊNCIAS GENÔMICAS

Regiões evolutivamente conservadas (ECRs - *evolutionary conserved regions*) localizadas entre genes, representam potenciais locais onde possam estar presentes elementos regulatórios. Isso porque esses locais, provavelmente, representam regiões importantes na regulação da atividade gênica, tendo sido submetidos à grande pressão evolutiva, de forma que seria deletéria qualquer alteração em sua estrutura.

O alinhamento e a busca por seqüências que tenham sido conservadas entre diferentes vertebrados durante a evolução filogenética (ECRs) foram realizados pelos programas ECR Browser e Mulan, que podem ser acessados na *homepage* do Centro de Genômica Comparativa, no endereço <http://www.dcode.org>. O Centro de Genômica Comparativa é mantido pelo Laboratório Lawrence Livermore e pela Universidade da Califórnia, que disponibilizam uma coleção de ferramentas de Bioinformática que podem ser acessadas *on-line* em <http://bio.cse.psu.edu/>. Nesse trabalho, foram consideradas como ECRs, regiões que apresentaram no mínimo 100 pares de base e 70% de identidade de seqüência. Esses parâmetros podem ser modificados nos programas utilizados conforme o interesse do pesquisador.

Os organismos selecionados para as comparações genômicas foram o homem, camundongo, galinha, o anuro *Xenopus tropicalis*, paulistinha (zebrafish – *Danio rerio*), *Takifugu rubripes* (baiacu ou Fugu) e *Tetraodon nigroviridis* (baiacu verde pintado). A seqüência do gene da miostatina humana foi utilizada como referência para as comparações de seqüência. As coordenadas do trecho que foi comparado são ch2: 190.614.448 a ch2: 190.871.187, do genoma humano. Essa região abrange, além do gene da miostatina humana, a região intergênica situada *upstream* (até a extremidade 3' do gene que codifica a proteína hipotética MEC13057) e *downstream* (até a extremidade 5' do gene que codifica a proteína hipotética LOC285011) ao gene.

Como ferramenta auxiliar na comparação genômica de diferentes vertebrados foi empregado o programa PipMaker (<http://pipmaker.bx.psu.edu/pipmaker/>).

3.2 IDENTIFICAÇÃO DE POTENCIAIS SÍTIOS DE LIGAÇÃO PARA FATORES DE TRANSCRIÇÃO

Uma vez tendo sido identificadas as ECRs entre os diferentes vertebrados, os potenciais sítios de ligação para fatores de transcrição, que tenham sido conservados ao longo

da evolução, foram localizados com o auxílio do recurso multiTF do Mulan. Em seguida, esses fatores de transcrição, que podem estar envolvidos na regulação da atividade transcricional da miostatina, foram categorizados com relação à classe e função com base nos dados do banco de dados Transfac (<http://www.gene-regulation.com/pub/programs.html>).

4. RESULTADOS

4.1 PROGRAMAS UTILIZADOS

Para melhor compreensão e visando facilitar o acesso por parte de outros estudantes a estas ferramentas de Bioinformática, as instruções para a utilização dos programas ECR Browser, Mulan e PipMaker estão disponíveis abaixo.

4.1.1 ECR BROWSER

4.1.1.1 INTRODUÇÃO

ECR Browser é uma interface gráfica e dinâmica para Regiões Evolutivamente Conservadas (ECRs) em genomas de espécies seqüenciadas incluindo humano, camundongo, rato e Fugu. O perfil de conservação criado ao se alinhar a “seqüência-base” com todas as outras aos pares, é mostrado graficamente para qualquer locus de um genoma. ECRs são identificados como regiões de alta identidade de seqüência contra um *background* de evolução neutro. Ao analisar um alinhamento, o programa detecta e destaca elementos de seqüência de comprimento significativo que são conservados acima de um nível específico de identidade de seqüência entre os dois genomas (de acordo com os parâmetros definidos pelo usuário) e os identifica como ECRs. Visualmente, os ECRs são representados como picos coloridos em um gráfico, cujo eixo X representa posições no genoma-base e o eixo Y representa a porcentagem de identidade de seqüência entres os genomas base e o alinhado, naquela posição específica.

A Figura 7 mostra um gráfico de conservação do tipo *pip* no qual o DNA humano (seqüência-base) é representado no eixo horizontal, enquanto múltiplos alinhamentos do tipo blastz ininterruptos são mostrados no gráfico como pequenas linhas horizontais. O comprimento da linha de alinhamento horizontal corresponde ao comprimento do alinhamento à seqüência-base, enquanto a posição vertical corresponde ao nível de identidade de nucleotídeos desse alinhamento. São utilizados apenas valores significativos de 50 a 100% de identidade de seqüência. Pequenos retângulos vermelhos no topo do gráfico representam os ECRs.

Para demarcar ECRs em relação a aspectos de codificação de proteínas, genes anotados são mostrados como uma linha horizontal azul acima do gráfico, com linhas inclinadas que indicam a orientação da transcrição. Quadrados azuis ao longo da linha correspondem a exons, e quadrados amarelos correspondem a UTRs (*Untranslated regions* – regiões não-traduzidas). Os exons também são mostrados através de picos azuis no gráfico. Picos dentro do perfil de conservação que não correspondem a seqüências transcritas estão destacados em vermelho se forem intergênicos ou em rosa se estiverem em um intron.

Retângulos verdes na base do gráfico mostram a posição de elementos repetitivos no genoma-base e essa anotação é sombreada até o topo em cinza.

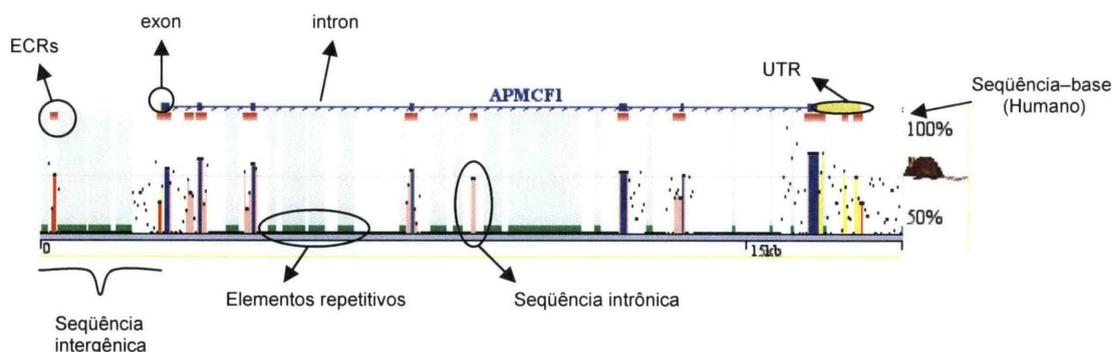


Figura 7. Gráfico de conservação do tipo *pip*, para o gene APMCF1, no qual o DNA humano (seqüência-base) é representado no eixo horizontal, enquanto múltiplos alinhamentos do tipo blastz ininterruptos são mostrados no gráfico como pequenas linhas horizontais.

4.1.1.2 QUAIS ASPECTOS DE UMA SEQÜÊNCIA CORRESPONDEM A ECRS?

4.1.1.2.1 EXONS CODIFICADORES

Uma grande fração de ECRs identificados em qualquer alinhamento de genoma corresponde a exons conservados, codificadores de proteínas (representados por retângulos azuis no gráfico). Devido ao significado funcional das seqüências que codificam proteínas, exons codificadores estão geralmente sob grande pressão seletiva para permanecerem intactos. Por isso, enquanto o *background* neutro diverge e “desaparece” do grupo conservado com o aumento da distância evolutiva entre dois genomas, exons codificadores geralmente permanecem como picos proeminentes de seqüências conservadas. É normal observar uma única e intacta linha horizontal de alinhamento que corresponde a exons codificadores, uma vez que inserções e deleções que mudariam o quadro de tradução não são toleradas.

4.1.1.2.2 GENES NOVOS

Apesar de recentes avanços na descrição do genoma humano, há ainda muitos genes que permanecem desconhecidos. O ECR Browser fornece uma anotação gráfica de genes preditos, acima da marca dos ECRs, para que os níveis de conservação dos exons preditos possam ser examinados atentamente (Figura 8). Já que a maioria dos exons codificadores é conservada nos alinhamentos entre vertebrados, o ECR Browser representa uma ferramenta para encontrar e avaliar novos genes e exons alternativos ainda não descritos. Em alguns casos, o perfil conservado mimetiza parcial ou completamente um transcrito predito de um gene e fornece evidência adicional de que o gene predito é real e funcional. No exemplo abaixo, exons do cromossomo ‘chr5.11.006.a’ previstos pelo programa “Twinscan” dentro de

um modelo gênico sem nenhuma contrapartida de um gene conhecido, correspondem perfeitamente a um agrupamento de ECRs dentro da região analisada, fornecendo evidência extra de que essa previsão corresponde a um gene funcional. ECRs adicionais conservados entre os exons previstos pelo “Twinscan” poderiam representar exons candidatos a compor um transcrito alternativo ou, potencialmente, elementos regulatórios dentro do gene.

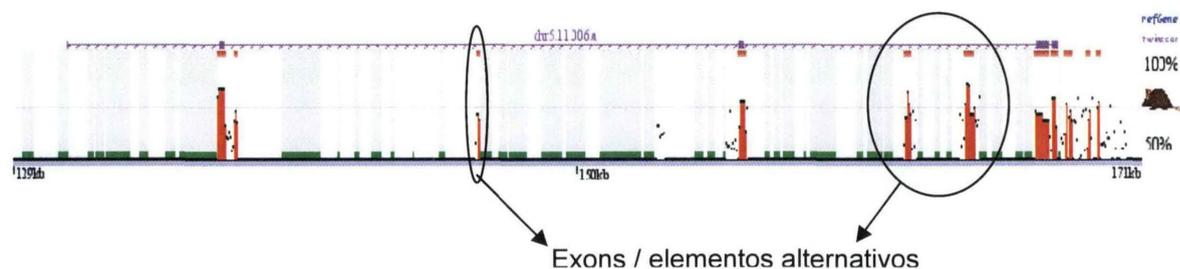


Figura 8. Novos genes e exons alternativos ainda não descritos também podem ser encontrados e avaliados pelo ECR Browser.

4.1.1.2.3 PROMOTORES E ENHANCERS

Para alguns genes, a transcrição é dirigida parcial ou completamente por elementos estimuladores ou *enhancers*, localizados imediatamente *upstream* do promotor (Figura 9). Esses elementos, se permanecerem conservados em duas espécies comparadas, podem ser facilmente identificados como picos vermelhos localizados perto e *upstream* da extremidade 5' do gene.

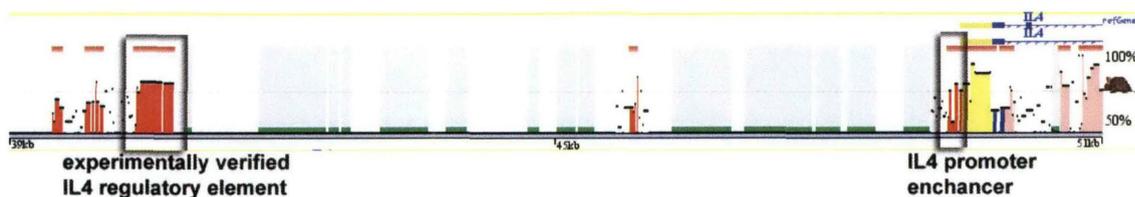


Figura 9. Elementos estimuladores ou *enhancers*, localizados imediatamente *upstream* do promotor.

4.1.1.2.4 ELEMENTOS REGULATÓRIOS DISTAIS

Funções adicionais podem estar escondidas em elementos que se localizam longe dos genes e que regulam os padrões de expressão espacial e temporal da transcrição de genes vizinhos. ECRs regulatórios são freqüentemente conservados através da evolução, e

excelentes candidatos a tais elementos regulatórios distais podem ser identificados no ECR Browser como elementos bem conservados localizados entre genes descritos.

4.1.1.3 PORQUE COMPARAR VERTEBRADOS SUPERIORES COM PEIXES?

O alinhamento das seqüências de humano e camundongo forneceu uma ferramenta valiosa para a descrição de elementos funcionais em ambos os genomas. Entretanto, como diferentes regiões do genoma de vertebrados parecem estar divergindo em razões evolutivas muito diferentes, nenhum tipo de comparação entre apenas duas espécies pode ser utilizada com sucesso para analisar todos os lóci genômicos. Em certas regiões, a conservação humano-camundongo é tão alta que torna difícil o estudo de elementos conservados específicos. Em um estudo recente, NOBREGA *et al.* (2003) descobriram que, em tais regiões, o alinhamento humano-camundongo freqüentemente aponta centenas de ECRs não codificadores por gene. Esses autores descobriram que comparações entre organismos evolutivamente distantes, nesse caso o homem e Fugu, forneceram uma maneira altamente eficiente de examinar todos esses ECRs para encontrar aqueles com a maior probabilidade de serem funcionais. Nove ECRs identificados na comparação entre humano e peixe foram demonstrados como sendo *enhancers* funcionais. Em outras regiões, essas comparações podem levar a nenhum elemento conservado, mesmo perto de genes com funções altamente conservadas. Portanto, não há uma comparação ideal ou regras com relação a distancia evolutiva entre genomas alinhados que permitam que todos os elementos funcionais sejam identificados em uma região de interesse. Por essa razão, ECR Browser fornece acesso a várias comparações genômicas para que o usuário possa escolher a que melhor se encaixe na análise de cada lócus individual.

4.1.1.4 GENOMAS COMPARADOS

A presente versão do ECR Browser contém um alinhamento comparativo de nove genomas diferentes – humano, cachorro, camundongo, rato, Opossum (marsupial), galinha, *Xenopus* e três peixes (Fugu, *Tetraodon* e paulistinha).

4.1.1.5 COMO COMEÇAR A USAR O ECR BROWSER?

A maneira mais rápida de começar a navegar pelo ECR Browser é selecionando o genoma-base e digitando o nome do gene no espaço delimitado para isso na janela de entrada do programa. Outra opção é indicar uma localização no cromossomo, da seguinte forma: chr17:1000-2000, para descrever o cromossomo 17, entre as localizações 1000-2000 pb. Se o usuário digitar um nome parcial de algum gene, todos os genes com nomes parecidos serão

listados, com nomes e localizações completos no genoma-base, e com um *link* direto para a visualização dos perfis no ECR Browser.

O formulário de localização que aparece a seguir (Figura 10), contém a informação do genoma-base que está sendo usada atualmente, bem como o comprimento total do lócus genômico que está sendo visualizado. Através da opção *Jump to*, o usuário tem a possibilidade de efetuar novas análises, mudando a posição cromossômica ou o gene analisado.

The image shows a web interface for the ECR Browser. At the top, it says "ECR Browser on Human (hg16)" and "89980 bps". On the right, there is a URL: "http://ecrbrowser.dcode.org/". In the center, there is a search bar with "GATA3" entered, preceded by "Jump to" and followed by a "Submit" button. Below the search bar, there is a red text instruction: "GENOME ALIGNMENT: Align your sequence to either Human, Mouse, Rat, or Fugu genome".

Figura 10. Formulário de localização.

4.1.1.6 SELECIONANDO O GENOMA BASE

A barra de ferramentas do ECR Browser, localizada na parte superior da tela, contém vários *links* que fornecem acesso a dados importantes e permitem modificações dos parâmetros, estabelecendo o esquema de funcionamento do ECR Browser. A opção mais à esquerda, chamada *Base Genome*, permite a mudança do genoma-base utilizado pelo programa (Figura 11). Quando o genoma-base é alterado, todas as coordenadas exibidas no programa corresponderão ao novo genoma-base escolhido; a lista de cromossomos disponíveis será limitada pela estrutura seqüenciada do genoma, e o *display* gráfico dos genes e outras características corresponderão às descrições daquele genoma. O usuário pode variar de genoma-base para ver estruturas, conteúdos, descrições e outras características gênicas atribuídas a cada espécie.

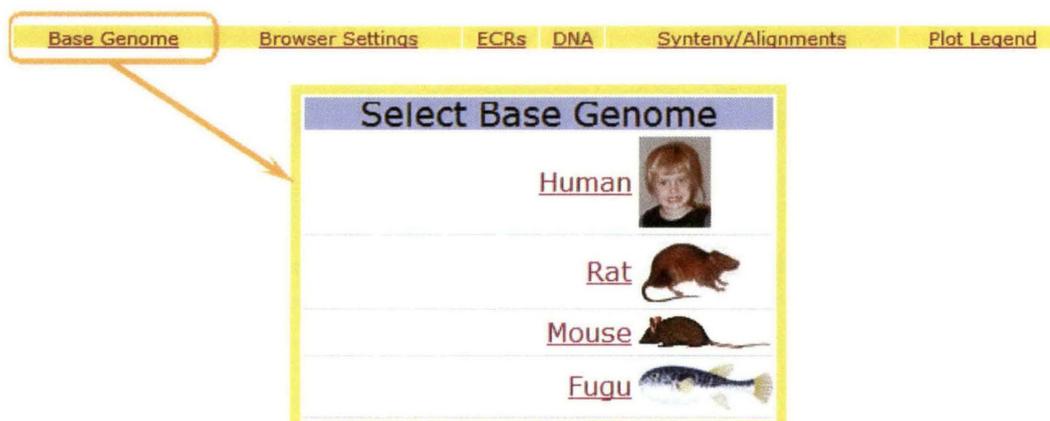


Figura 11. A opção *Base Genome* permite a mudança do genoma-base utilizado pelo programa.

4.1.1.7 BROWSER SETTINGS

A segunda opção a partir da esquerda, chamada *Browser Settings* (Figura 12), fornece a flexibilidade de mudar dinamicamente os parâmetros utilizados para gerar perfis de conservação. O usuário tem uma opção de visualizar o nível de conservação usando qualquer seleção de genomas disponíveis que tenham sido alinhados com o genoma-base. Por exemplo, é possível visualizar a conservação da seqüência humana com apenas a dos roedores, apenas a dos peixes, de um roedor e um peixe, ou todas as espécies. Essa opção também permite ao usuário mudar o estilo do gráfico de conservação, para visualizar *displays* com picos ou barras. Vários trechos de genes preditos estão disponíveis e podem ser selecionados em adição à principal descrição gênica (RefSeq – seqüência referência); a disponibilidade desses trechos depende da disponibilidade de dados correspondentes no programa *UCSC Genome Browser*.

Para fornecer um efeito eficiente de *zoom* que irá possibilitar a visualização de um lócus genômico longo em apenas uma janela, o ECR Browser permite que o perfil de conservação seja dividido em várias camadas. Cada camada representa uma parte do lócus genômico visualizado, onde o comprimento de cada parte e suas posições dentro do lócus são marcados por números abaixo de cada traço. O número total de camadas é definido pelo usuário (*Number of layers*), bem como a altura de cada camada (*Layer height*). Esse valor, multiplicado pelo número de camadas definirá a altura total da imagem.

O ECR Browser detecta regiões evolutivamente conservadas (ECRs) de uma maneira dinâmica. O usuário pode alterar os parâmetros para a identificação de ECRs, conforme a porcentagem de identidade de seqüência desejada. Dessa maneira, a detecção de apenas ECRs muito longos ou de alta conservação, por exemplo, podem ser selecionados. É importante usar parâmetros de detecção de ERCs adequados a cada situação. Somente dessa maneira, o usuário poderá analisar de maneira apropriada os alinhamentos entre seqüências de alta similaridade ou muito divergentes, como alinhamentos rato-camundongo ou humano-peixe, ou em regiões que foram sujeitas a pressões evolutivas muito diferentes.

Base Genome **Browser Settings** ECRs DNA Synteny/Alignments Plot Legend

ECR Browser on Human (hg16) Settings

Display alignments with

- Mouse (mm3)
- Rat (rn3)
- Fugu (fu3)
- Tetraodon (to6)
- Zebrafish (zf2)

Graph type Smooth Pip-plot

Display gene features Genscan RefSeq Twinscan

Number of layers

Layer height pixels

Detect ECRs
(Evolutionary Conserved Regions)

min length bps
min identity %

submit & return

Figura 12. A opção *Browser Settings* fornece a flexibilidade de mudar dinamicamente os parâmetros utilizados para gerar perfis de conservação.

4.1.1.8 LISTA DE ECRS NO LÓCUS

A posição genômica absoluta de cada ECR detectado no lócus visualizado, assim como detalhes sobre cada um deles, está disponível através do *link* ECRs, localizado na barra superior do ECR Browser. Os ECRs são organizados de acordo com as espécies selecionadas pelo usuário, para que sejam utilizadas no diagrama de conservação. Caso múltiplos lóci de uma das espécies sejam detectados como contendo significativa homologia com uma posição no genoma-base, será apresentada uma lista de ECRs correspondendo a todos os lóci homólogos (Figura 13).

A posição genômica listada para cada ECR corresponde à posição da homologia no genoma-base e há um *link* direto para acessar essa localização.

Base Genome Browser Settings **ECRs** DNA Synteny/Alignments Plot Legend

List of ECRs from the chr17:48543517-48547000 region in the Human (hg16) genome

Position	Genomic position (link to the sequence)	Length	Percent identity	mm3 chr11(mm3) alignment
2-446	chr17:48543518-48543962	445bp	86.5%	
964-1331	chr17:48544480-48544847	368bp	76.4%	
1578-1758	chr17:48545094-48545274	181bp	68.0%	
1808-2119	chr17:48545324-48545635	312bp	78.2%	
2234-2440	chr17:48545750-48545956	207bp	70.0%	
2994-3484	chr17:48546510-48547000	491bp	83.5%	

Position	Genomic position (link to the sequence)	Length	Percent identity	fu3 scaffold 965(fu3) alignment
190-368	chr17:48543706-48543884	179bp	68.7%	
1837-2097	chr17:48545353-48545613	261bp	75.5%	
3297-3456	chr17:48546813-48546972	160bp	68.1%	

Position	Genomic position (link to the sequence)	Length	Percent identity	scaffold 1247(fu3) alignment
206-383	chr17:48543722-48543899	178bp	68.5%	
1859-2087	chr17:48545375-48545603	229bp	70.3%	

Figura 13. Lista de ECRs correspondendo a todos os lóci homólogos do alinhamento.

4.1.1.9 SEQÜÊNCIA DE DNA BASE

O *link* chamado DNA, na barra superior do ECR Browser, fornece um acesso à seqüência genômica do genoma-base que deu origem ao gráfico de conservação. A seqüência obtida ao se clicar nesse *link* estará em letras maiúsculas, exceto pelas regiões que correspondem a elementos repetitivos, que estarão em letras minúsculas. (Figura 14)

Base Genome Browser Settings ECRs **DNA** Synteny/Alignments Plot Legend

```
> Human (hg16) chr17:48543517-48547000
GCTCAGTACACAGCCCCAGGGTTGGGCGGGGCCCCGGGAGAGGCATGGTG
CAGGGTGGCTGGCTGAGGCGGCTGCTGCTGTAAGTGGGGTCCACTCAGGT
TCTGTGCGTGATACCAGGAGTTGGTGGGGTCGTCAGGTAGCTGGGGGAG
GCACTGTATGGGAGCGGGGGGCGAGCTGACTGCGGGCAGGGGCGGAGT
GGAGTGGGAAGAGGTGTCCAGAGGGCGGGTGATGGTGGTGAGTTGCAGG
CCAIGGAATCACTGTTATTGGGACTGTGCTCCAGCGGCACCTCCCCGTTT
TTGTAGAGTTTCTTGAACCTTGAACGGCGGTTCTGGAACCAAGATTTCAC
CTGGGCCAGAGAAGAAAGGGGTAGCTAGTTAGCCTCTCAGAATGCTAAAA
CAACCTTGGGAAGAACCTAGCCAAGAAAAGAAAGGCCAGAACCCAATTT
CACATAAGAATCAGGCAAAAAACAGGCCTCCTGACTCCCAGTTTCTTGA
TCTGGGatagaattgtggctaggaacggtcagctgcctgggtccaaattc
tgctctgtcacttactaagcgtgacctgggcaagtgtttaacctctct
tgttcttgacttctctgtctgcaaaatggggatgagaacagcactatt
acctgcctcatagggtcattctgaggctcaataaggtacgacatataaa
aaacttaaagcagcctggatgtagtaagcactcaataaaaAGGCCGTTGA
GGTGGGAGATGGACGGTTTGTAGCATTTCATTACATTAGACACCCATAA
```

Figura 14. A opção DNA fornece um acesso à seqüência genômica do genoma-base que deu origem ao gráfico de conservação.

4.1.1.10 NAVEGANDO POR *LINKS* DE SINTENIA, ACESSANDO ALINHAMENTOS, *DOT-PLOTS* (GRÁFICOS DE PONTOS) E DESCRIÇÕES DE SÍTIOS DE LIGAÇÃO PARA FATORES DE TRANSCRIÇÃO EVOLUTIVAMENTE CONSERVADOS.

As relações de sintenia que dão origem a gráficos de conservação no ECR Browser, podem ser visualizadas através do *link Synteny/Alignments*, localizado na barra superior do programa. No exemplo (Figura 15), a região de 35Kb do cromossomo 17 do genoma humano (hg 16) é analisada na busca de possíveis combinações com outros genomas que estão listados um abaixo do outro. Cada região homóloga de um genoma combinante contém um perfil de homologia azul-cinza, onde a cor azul corresponde à região de sintenia no genoma base. Observa-se que apenas 2/3 desse locus humano reconhece uma clara região homóloga em genomas de roedores. As relações de sintenia com peixes são ainda mais localizadas, limitadas a um pequeno segmento central no locus humano.

A coluna *Position* fornece as coordenadas precisas da região combinada no genoma alinhado; essa coluna é um *link* para uma versão do ECR Browser que usa a seqüência das espécies selecionadas como genoma-base. Seguindo o *link Position* é possível visualizar a mesma região funcional em várias espécies. Por exemplo, ao selecionar a visualização do gene GATA3, usando o humano como genoma-base, pode-se visualizar o mesmo gene em outra espécie (mesmo que o gene homólogo ainda não tenha sido descrito naquela espécie).

A coluna *Length* mostra o tamanho da região homóloga, e pode ser usada com precisão para detectar expansão ou encolhimento de regiões específicas em diferentes linhagens evolutivas.

O *link Alignments/Graphs* envia um alinhamento homólogo para a ferramenta zPicture (<http://zpicture.dcode.org/>), que fornece acesso à várias opções adicionais para manipular a seqüência e visualizar alinhamentos. Uma das opções do zPicture cria um *dot-plot* de similaridade do alinhamento, de forma a visualizar diferentes ordenamentos e localizações de regiões dentro das 2 seqüências sendo comparadas.

O *link* chamado *Binding Sites* conecta o ECR Browser com a ferramenta rVista (<http://rvista.dcode.org/>), que serve para identificar e descrever sítios para fatores de transcrição evolutivamente conservados (TFBS) nos alinhamentos. O rVista irá excluir predições de até 95% de falsos positivos TFBSs, mantendo uma alta sensibilidade na pesquisa.

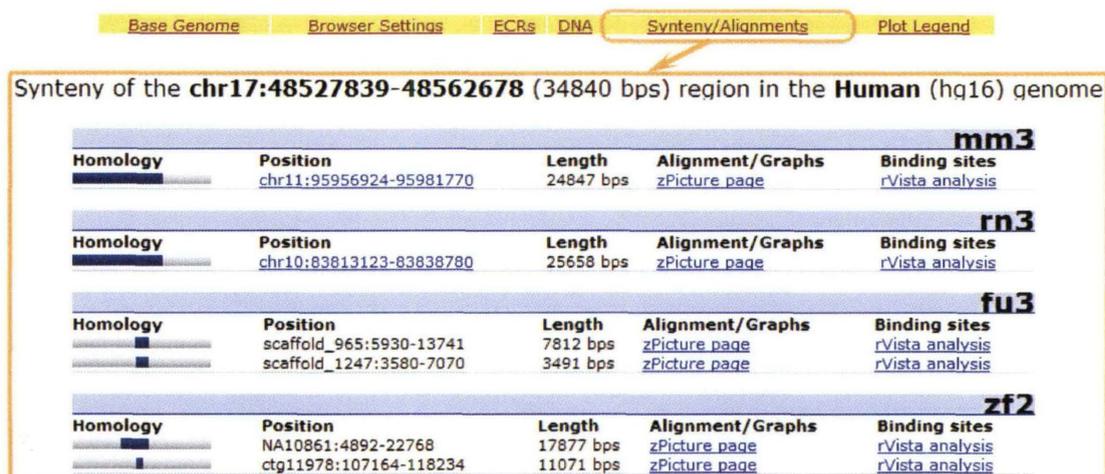


Figura 15. Através da opção *Synteny/Alignments*, pode-se visualizar as relações de sintenia que dão origem a gráficos de conservação no ECR Browser.

4.1.1.11 RECENTRALIZANDO EM UMA DADA LOCALIZAÇÃO

Um clique do mouse em qualquer posição do gráfico do ECR Browser resultará em uma recentralização do gráfico de conservação em uma dada localização. (Essa é uma característica similar ao mapquest.com e outras ferramentas de manipulação de mapas). Por exemplo, se você está interessado em uma visualização detalhada das regiões promotoras do GATA3, você pode primeiro visualizar o gene GATA3, clicar na extremidade 5' do gene (isso irá recentralizar no local de início de transcrição do gene) e então fazer um *zoom* de x vezes.

4.1.1.12 USANDO O ZOOM E MOVENDO

Vários botões na parte de baixo do gráfico do ECR Browser (Figura 16) são responsáveis por funções de *zoom* e movimentação. Botões azuis com os símbolos < e > irão mover o gráfico para a esquerda e direita, respectivamente, preservando 1/3 do intervalo original e expandindo-o apropriadamente enquanto mantém intacto o comprimento total do lócus visualizado. Botões amarelos fornecem funções de *zoom* positivo, enquanto os verdes são responsáveis por *zoom* negativo. É possível fazer *zoom* de 1,5, 3 e 10 vezes usando esse programa. Nota: *zoom* negativo de intervalos muito longos pode resultar em atrasos significativos no tempo de resposta, pois o programa precisará analisar muitos alinhamentos diferentes para criar um perfil de conservação.



Figura 16. Funções de *zoom* e movimentação do gráfico de conservação.

4.1.1.13 MOVENDO DENTRO DO CROMOSSOMO E DE UM CROMOSSOMO PARA OUTRO

No topo do *display* gráfico do ECR Browser, todos os cromossomos da espécie representada pelo genoma-base escolhido, estão listados como números (Figura 17). O cromossomo que está sendo estudado é destacado por um fundo de cor diferente. É possível pular de um cromossomo para outro, clicando no símbolo do cromossomo correspondente. À esquerda do *display* gráfico do ECR Browser, é mostrada uma imagem de cariótipo com *links* ativos do cromossomo selecionado. A posição do lócus genômico, dentro do cromossomo que está sendo exibido, é mostrada como uma barra vermelha na imagem do cromossomo. Regiões heterocromáticas não seqüenciadas (correspondentes a centrômeros e telômeros) são mostradas como regiões mais estreitas na imagem do cromossomo. A escala do cromossomo, em Mb, é mostrada imediatamente à direita da imagem do cromossomo. Um clique do mouse na imagem do cromossomo ativo/escolhido resultará na mudança do *display* do ECR para o local correspondente ao clique do mouse.

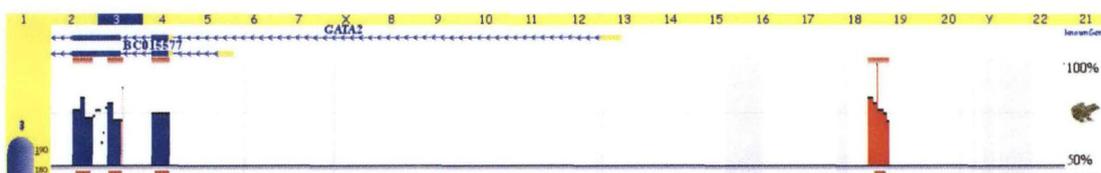


Figura 17. No topo do *display* gráfico do ECR Browser, todos os cromossomos da espécie representada pelo genoma-base escolhido, estão listados como números.

4.1.1.14 FUNÇÃO *GRAB ECR*

Detalhes de seqüência e alinhamento para cada ECR em destaque no gráfico de conservação do ECR Browser, podem ser obtidos usando a função *Grab ECR* do programa. Um clique do mouse no botão *Grab ECR* (que muda de cor depois do recarregamento do *browser*), seguido de um segundo clique do mouse em qualquer pico colorido (ECR) do gráfico, resulta no aparecimento de uma nova página da Internet, descrevendo o ECR correspondente àquele pico (Figura 18). São fornecidos a localização cromossômica, o comprimento, a porcentagem de identidade do alinhamento e o conteúdo GC do ECR. Em adição, são mostrados o alinhamento completo e seqüências correspondendo àquele ECR em ambas as espécies, base e alinhada. Seqüências e alinhamentos de outras espécies podem ser obtidos usando a função *Grab ECR* para recuperar um pico dos alinhamentos descritos no gráfico de conservação com o genoma daquela espécie. Um *link* adicional pode ser usado para mandar o alinhamento de ECR para o *rVista*, para a detecção de sítios de ligação para fatores de transcrição evolutivamente conservados naquele ECR. Em adição a essas funções, *links* para a ferramenta *oligo/primer design* são fornecidos para as seqüências base e alinhada.

Favor notar que os *popups* devem ser permitidos no seu *browser* para que a função *Grab ECR* funcione adequadamente. Caso contrário, a nova janela com a descrição detalhada do ECR não irá aparecer depois que o usuário clicar num elemento conservado.

The diagram illustrates the process of clicking a 'Grab ECR' button, which leads to a detailed popup window for an Evolutionary Conserved Region (ECR). The popup window displays the following information:

ECR :: Evolutionary Conserved Region		Human _[hg16] - Mouse _[mm3] ECR				
Type	408 bps, 75.0% identity					
Location	chr17:48536190-48536597 _[hg16]					
Transcription factor binding sites	rVista					
Oligo/primers design	Human Mouse					
GC count	Human .. GC: 59.56%, AT: 40.44%, OTHER: 0.00% Mouse .. GC: 64.00%, AT: 36.00%, OTHER: 0.00%					
Alignment						
	10	20	30	40	50	60
Human	CaaTtCTGCTCCCTCtGGCCTtCtgACTtCTtAGtTGCCCaccCtCCCetGgCTGAaGg					
Mouse	CetTcCTGCTC					
	5860					
	70					
Human	CaCatCAtCCT					
Mouse	CgCccCAcCCT					
	5800					

Sequences

```

> Human
CAATTCTGCTCCCTCTGGCCTTCTGACTTCTTTAGTTGCCACCCTCCCCTGGCTGAAGG
CACATCATCCTCCCCACTCACCATCAGTGGTTCTGAGCTCAGGGCAGGCCTAACTGAC
CTAGGGAGTGGGTGTGGGACTGTGCCACACACATTCATGACGTGGCCAGGCACCCCTCAG
TGTGCAGGGAAGCACACGCTGCTGCCAGGTGCGAGGTGAACTGCCTGGGGTGGGGGATAC
CCCTCCCCTGCTCCAGCTAATCCTCCAGGGTGTGGAGTCTCTGTTCCTGCTGACTCAGC
AGACACTTCCAATCCTGCTCTGATTTGCCCCAGGAGACAACCCAGGAATTTGGGCCTTGA
GGCCCCACCCACTGTTCCCATCCCCAGGTTTGTCTCCTTCCAAGGTGC
> Mouse
CCTTCCTGCTCCCTCGGGCCTCCGCACTCCTCCAGCTGCCCTGCCCTGGACTGAGGA
CGCCCCACCCTTCCCCGACTCCCATTTGGTGATTCTGAGCTCAGGGTACAAGTACTGCTG
GCCCCAGGGGTGGGTGTGGGACTGTGCCACACGCGTCGCATGACGTGGCTGGGTGCCCTC
GGTGTGCGCGTCAGGGAAGCTTGTGCTGAACGACATTGCCAGGTGCGAGGTGAACTGCCT
CTCCACCCACGGGGTGGGGACGCCCTCCCTCATGCAGCTGATCCTCCAGGGTGTGGA
GCCTCCAGCTTACTCAGCAAACACTTCCAGCCCTGAACATAATTTGCTCTGGGAGACAACC
CAGGAATTTGGGCCACAGGCCCTGCCTGTTCTGTTCTCCTTCTCAGGCTGCCCCCTCCCC
GAATG
  
```

Figura 18. Um clique do mouse no botão *Grab ECR*, seguido de um segundo clique do mouse em qualquer pico colorido (ECR) do gráfico, resulta no aparecimento de uma nova página da Internet, descrevendo o ECR correspondente àquele pico

4.1.1.15 LINK DINÂMICO PARA O UCSC BROWSER

O botão <UCSC Genome Browser>, localizado da barra superior do ECR Browser, fornece acesso dinâmico para descrições detalhadas do locus genômico que é fornecido pelo UCSC Genome Browser. A seleção do genoma base, sua fixação e dados exatos de localização serão enviados para o UCSC Genome Browser para mostrar exatamente a mesma região que está sendo vista no ECR Browser. Ao usar essa função, é possível estudar muitas camadas de descrições funcionais adicionais para qualquer locus que esteja disponível no UCSC Genome Browser. Isso inclui anotações sobre mRNA, EST, SNPs, descrições gênicas detalhadas, etc. Favor notar que esse *link* não é funcional para genomas base que não foram obtidos do UCSC (no momento, isso inclui genomas de peixes).

4.1.1.16 GRÁFICOS PIP X GRÁFICOS SMOOTH

Qualquer perfil do ECR Browser pode ser visualizado como um gráfico do tipo *pip* ou como um gráfico *smooth* (Figura 19). A principal diferença entre esses dois tipos de visualização é o método de construção do gráfico de conservação. No caso do gráfico *pip*, cada alinhamento sem intervalos é visualizado como uma linha horizontal separada. O comprimento da linha corresponde ao comprimento do alinhamento, enquanto sua altura corresponde à porcentagem de identidade de seqüência do alinhamento. Gráficos *smooth* são construídos a partir de uma janela de 100pbs que desliza através do alinhamento. Essa janela, centralizada em cada nucleotídeo da seqüência-base, é usada para calcular o número de combinações dentro dela. Esse número fornece uma porcentagem de identidade em uma janela deslizante centralizada em uma posição determinada. A contagem final das porcentagens de identidade em cada janela é utilizada para calcular a altura do gráfico de conservação do tipo *smooth* em cada ponto. Basicamente, o gráfico *smooth* é uma média “suave” (do inglês, *smooth*) do gráfico *pip*. Os gráficos *smooth* apresentam uma visualização simplificada e mais clara no perfil de conservação, mas perdem informações com relação à distribuição de espaços no alinhamento.

A mudança do tipo de visualização do gráfico de conservação é feita através da opção *Browser/Settings*, na barra superior do ECR Browser.

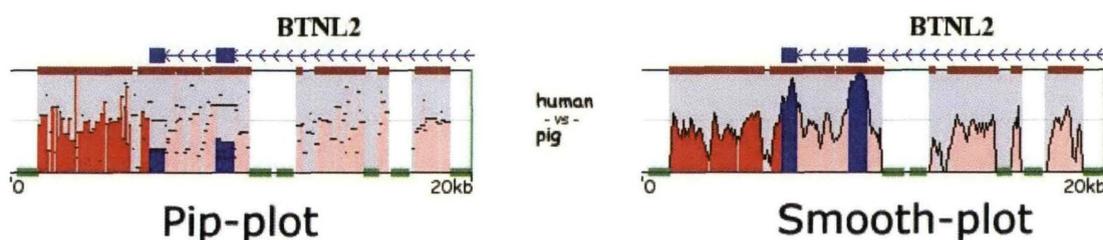


Figura 19. Gráficos dos tipos *pip* e *smooth*.

4.1.1.17 MUDANDO A LARGURA DA IMAGEM DO ECR BROWSER

Para mudar a largura da imagem do ECR Browser, apenas redimensione a janela do seu *browser* e o ECR Browser irá imitar a mudança na largura da janela do *browser* para cobrir todo o espaço da janela do *browser*.

4.1.1.18 ALINHAMENTO GENÔMICO

A característica *Genome Alignment* do ECR Browser tem o objetivo de fornecer a capacidade de mapear uma seqüência submetida pelo usuário com base em um dos genomas-base (humano, camundongo, rato ou Fugu) e de alinhar a seqüência submetida com a região homóloga do genoma escolhido (Figura 20). A seqüência internalizada pode ser submetida na forma de um arquivo FASTA ou pode ser automaticamente carregada do *GenBank* por um número de acesso conhecido.

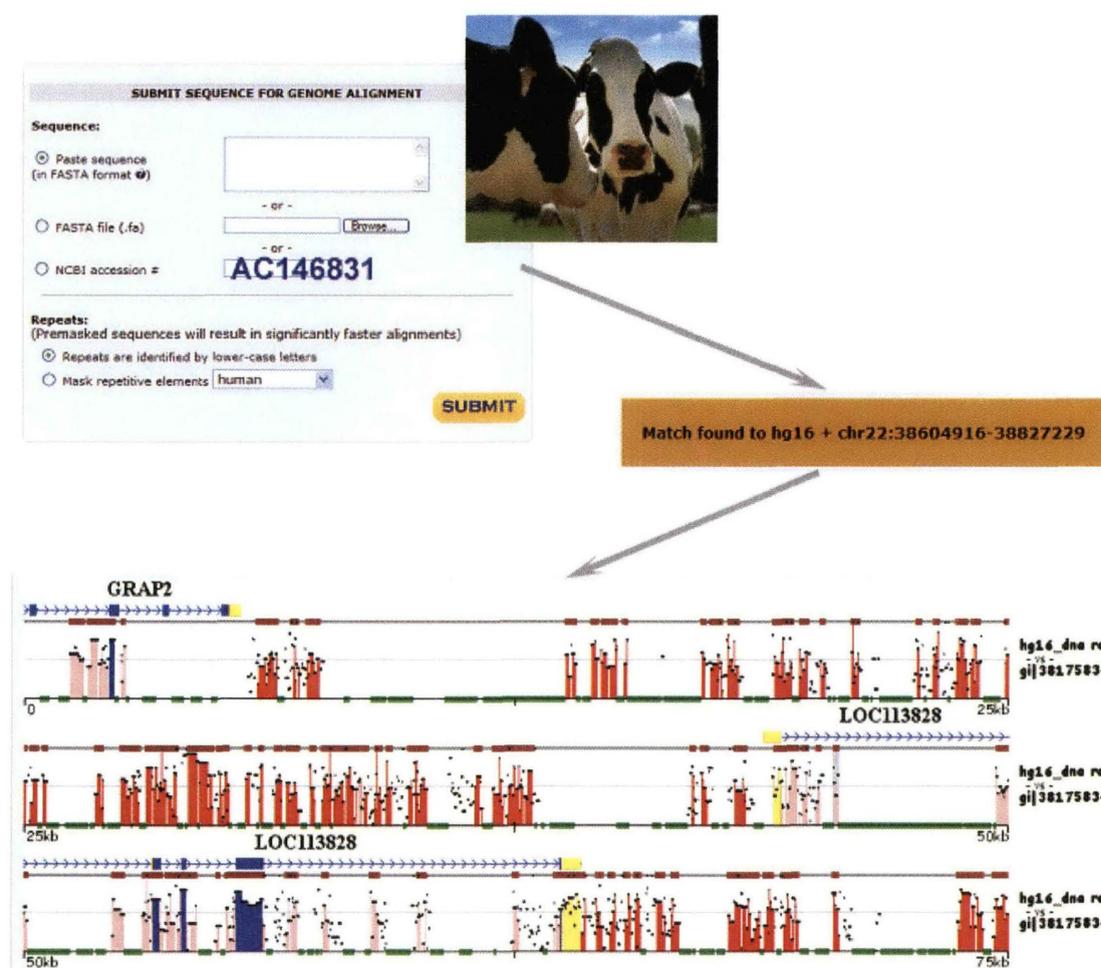


Figura 20. A opção *Genome Alignment* do ECR Browser tem o objetivo de fornecer a capacidade de mapear uma seqüência submetida pelo usuário com base em um dos genomas-base e de alinhar a seqüência submetida com a região homóloga do genoma escolhido.

4.1.2 MULAN – MULTIPLE SEQUENCE LOCAL ALIGNMENT AND CONSERVATION VISUALIZATION TOOL

4.1.2.1 INTRODUÇÃO

O programa Mulan executa múltiplos (dois ou mais) alinhamentos de seqüência com uma estratégia de alinhamento local completo, rápida e eficiente, que garante uma recapitulação de rearranjos de seqüências evolutivas em qualquer espécie. Esse programa combina as ferramentas *refine* e *tba* para alinhar seqüências de qualidade “rascunho” ou “terminada”. O Mulan fornece uma interface gráfica dinâmica para alinhar e visualizar perfis de conservação entre espécies pouco ou muito relacionadas evolutivamente.

Os formatos de entrada, carregamento automático de dados do *UCSC Genome Browser*, descrição de genes, descrição de elementos repetitivos, e relatórios de progresso foram previamente descritos nas instruções *zPicture* e o usuário deve recorrer a esse material para maiores detalhes. Essa introdução é centrada principalmente em novas características exclusivas do Mulan.

Também, para maiores detalhes, as principais características da ferramenta Mulan foram descritas em I. Ovcharenko *et al*, Mulan: Multiple-Sequence Local Alignment and Visualization for Studying Function and Evolution, *Genome Research*, 15, 184-194, (2005).

4.1.2.1.1 ALINHAMENTO DE SEQÜÊNCIAS DE QUALIDADE “RASCUNHO” (Figura 21)

É possível alinhar várias seqüências representadas como *contigs* múltiplos (um exemplo: arquivo FASTA de múltiplos *contigs*) a uma única seqüência *contig* referência usando a ferramenta Mulan. Para tanto, o usuário precisa selecionar o número de espécies (incluindo a seqüência referência) e clicar no botão *Select* à direita na página principal do Mulan, em <http://mulan.dcode.org>. Essa opção fornece um fácil acesso para uma eficiente ordenação e orientação (O&O) de seqüências secundárias baseado na homologia guiada pela seqüência referência. Duas opções de visualização são implementadas para ajudar o usuário com o O&O: descrição posicional de *contigs* nos perfis de conservação e um *dot-plot* (gráfico de pontos) de homologia, codificado por cores (Figura 22).

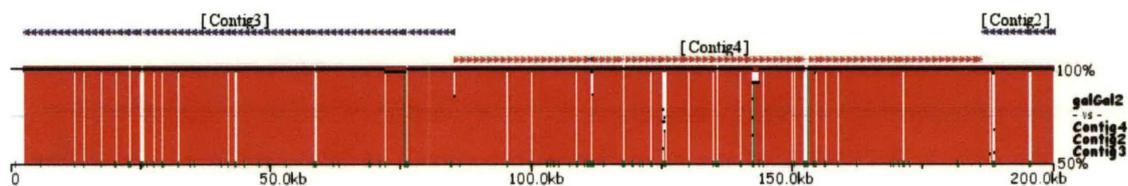


Figura 21. Uma seqüência de galinha de qualidade “rascunho” que consiste de três *scaffolds* (*Contig2*, *Contig3* e *Contig4*) foi alinhada acima de um segmento homólogo do genoma de galinha. A camada de cima indica a ordem e localização dos *contigs* impostos pela seqüência referência. A cor roxa corresponde a alinhamentos da fita reversa, e a cor vermelha a alinhamentos da fita senso.

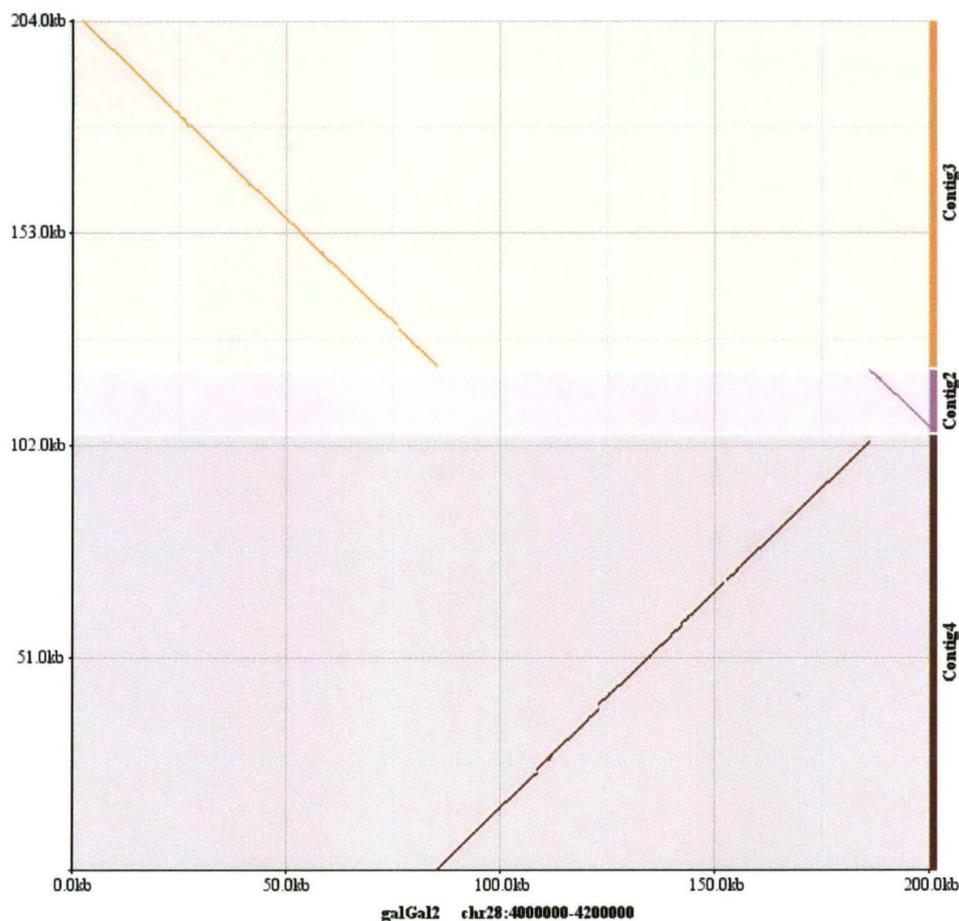


Figura 22. Um *dot-plot* de seqüências de galinha com três *contigs* alinhados com a seqüência homóloga contígua. Cores diferentes correspondem a diferentes *contigs*, como mostra o eixo vertical.

4.1.2.1.2 ALINHAMENTO DE SEQÜÊNCIAS DE QUALIDADE “TERMINADA”

A maioria das opções de análise de seqüências do Mulan corresponde a alinhamentos de seqüências de qualidade “terminada” (seqüências que são representadas por um único *contig*). Como está esquematizado na Figura 23, múltiplas seqüências serão automaticamente separadas em blocos de homologia, e cada um desses blocos será representado como um alinhamento de múltiplas seqüências, independentemente da ordem ou orientação da subsequência responsável pela construção do bloco de uma espécie em particular. Em resumo, a natureza de alinhamento local do Mulan garante uma recapitulação dos rearranjos evolutivos de seqüências em todas as espécies.

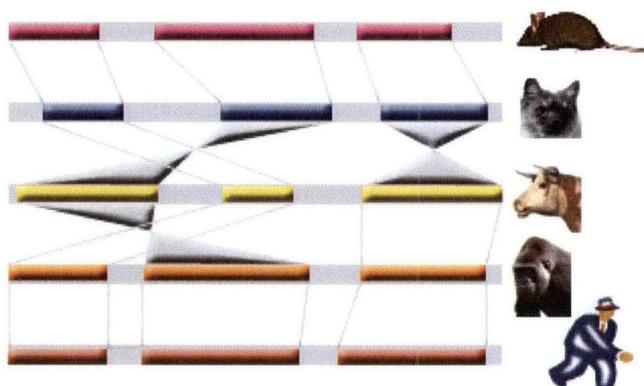


Figura 23. Blocos de homologia.

Especificações corretas de relações filogenéticas entre as seqüências de entrada, representam dados guias importantes para os alinhamentos tba utilizados pela ferramenta Mulan. O Mulan não requer que o usuário especifique manualmente as relações filogenéticas, mas, ao invés, o programa as prevê no primeiro passo do processo de alinhamento, mostrando-as como uma árvore filogenética (Figura 24) e pede por uma confirmação do usuário. Nesse passo intermediário, os usuários podem modificar as previsões automáticas das relações filogenéticas entre as seqüências de entrada para refinar os laços evolutivos entre as espécies.

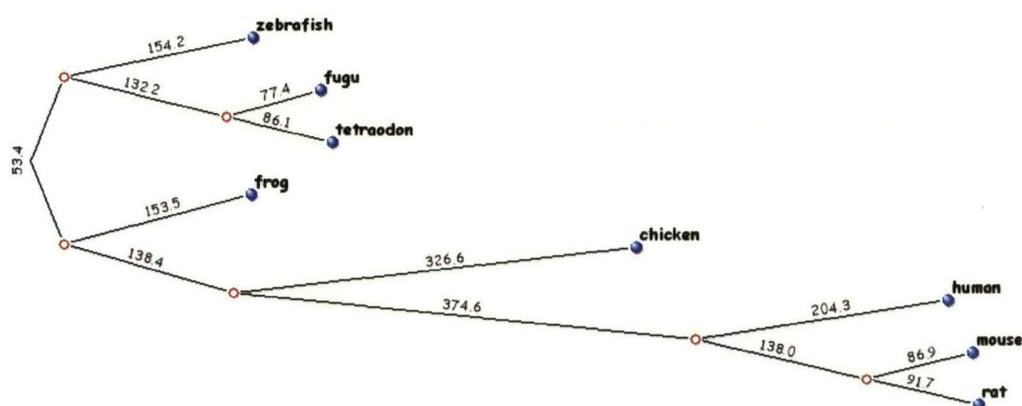


Figura 24. Árvore filogenética automaticamente criada pelo Mulan para descrever relações evolutivas entre as seqüências do humano, roedores, galinha, sapo e peixe para o lócus GATA3. Cada ramo da árvore estima um número de substituições por 1kb da seqüência.

4.1.2.2 RESULTADOS

4.1.2.2.1 PERFIS DE CONSERVAÇÃO DINÂMICOS

O Mulan mostra alinhamentos de seqüências múltiplas como perfis de conservação, gráficos e compactos, como gráficos do tipo *smooth* (com uma janela deslizante que conta o número de combinações errôneas para cada região de 100pb) ou gráficos do tipo *pip* (que mostram localização de porcentagem de identidade de blocos de alinhamento sem falhas) (Figura 25)

Os perfis de conservação do Mulan podem ser replotados dinamicamente baseado na seleção de parâmetros de exibição de dados, como selecionado pelo usuário através da barra superior na página de exibição do perfil de conservação (Figura 26). Dentre outros parâmetros, o usuário pode especificar parâmetros arbitrários para a detecção de ECRs, selecionar uma de quatro possíveis opções de visualização gráfica (incluindo análise filogenética de seqüências altamente relacionadas), e selecionar dinamicamente o organismo de referência.

ECRs são mostrados como blocos vermelho-escuro no topo da camada de conservação de cada espécie e eles estão ligados com os alinhamentos subjacentes. Clicando em um ECR, o usuário é automaticamente redirecionado para a página da *web* que fornece o alinhamento de seqüência múltipla correspondente. Diferentes características funcionais estão codificadas por cores no alinhamento para uma fácil separação visual um do outro (Figura 27).

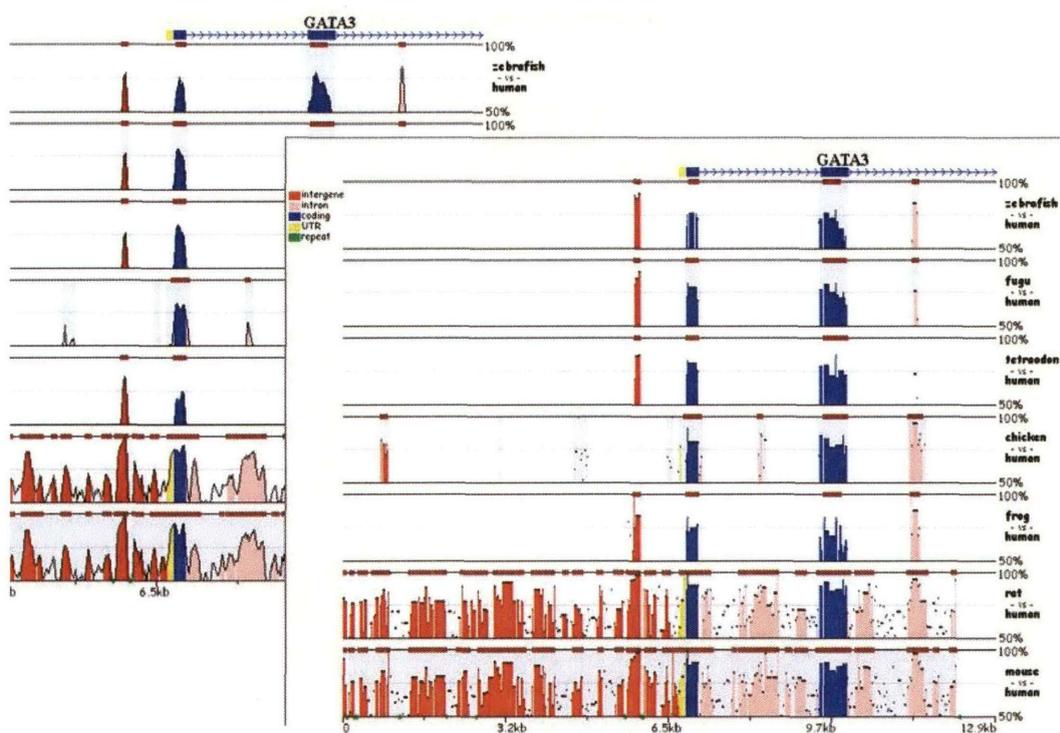


Figura 25. Perfis de conservação do tipo *smooth* e *pip* do lócus GATA3 em genoma de humano, roedores, galinha, sapo e peixe como gerados pelo Mulan.

Request ID: [example](#) <http://mulan.dcode.org/>

Picture settings	Smooth graph	Width	ECR length	ECR similarity	Bottom cut-off	Graph height	Remove legend
<input type="checkbox"/>	<input type="checkbox"/>	15000 bases	at least 100 bases	at least 70 %	50 %	80 pixels	<input type="checkbox"/>
From	To	ECRs present in		Visualization type	Base specie		
1	-1	1 species		standard stacked-pairwise	human		
		1		standard stacked-pairwise	human		
		2		color density by interspecies conservation	mouse		
		3		phylogenetic shadowing	rat		
		4		summary conservation	zebrafish		
		5			fugu		
		6			tetraodon		
		7			chicken		
					frog		

Figura 26. Barra de ferramentas de seleção de parâmetros englobando perfis de conservação gráficos.

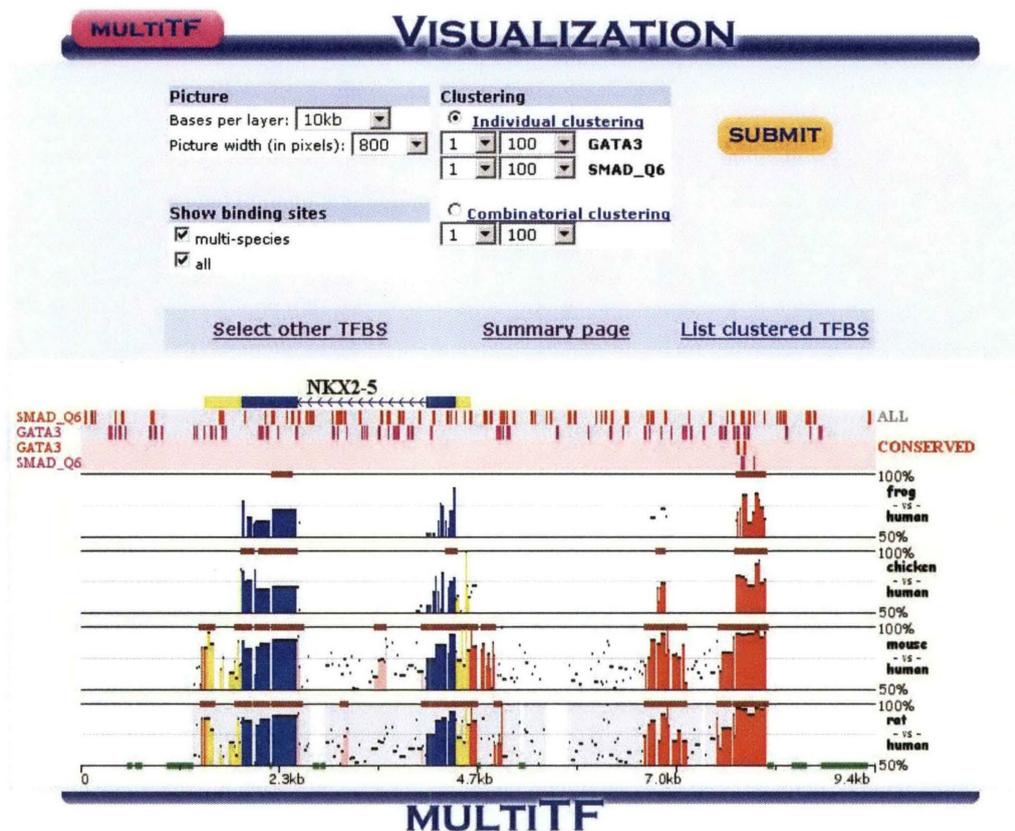


Figura 28. Anotações do MultiTF de TFBS no locus Nkx2.5. Fatores de transcrição de Smad (em vermelho) e Gata3 (em roxo) estão descritos como linhas acima dos gráficos de conservação. Cinco sítios Gata3 e 3 Smad estão altamente conservados e mapeiam em um elemento *enhancer* experimentalmente definido.

4.1.2.2.3 ALINHAMENTOS TEXTUAIS DE MÚLTIPLAS SEQÜÊNCIAS

Alinhamentos locais de seqüências múltiplas gerados pelo Mulan através do uso do programa de alinhamento tba podem ser representados como uma soma de pequenos blocos de alinhamentos contíguos, com cada um deles incluindo um subconjunto de espécies de entrada. Por exemplo, um pedaço de seqüência que é específico apenas de peixes será alinhado dentre as seqüências, englobando, exclusivamente, as linhagens de peixe, enquanto os elementos específicos de roedores serão alinhados por blocos de alinhamento que consistem de seqüências de roedores, apenas. Não importa a ordem ou a orientação de seqüências que constroem um bloco particular, uma vez que o alinhamento completo é feito em um padrão que maximiza o *score* total através de todos os possíveis blocos de alinhamento. Essa natureza local dos alinhamentos de seqüências múltiplas do Mulan não permite uma geração de um arquivo de alinhamento uniforme do tipo ClustalW. Ao invés disso, o usuário pode fazer o *download* de alinhamentos tba originais no formato FASTA *maf*, ou obter projeções do alinhamento de seqüências múltiplas completo para uma das seqüências de

espécies como uma seqüência referência. A seleção da seqüência referência é dinâmica, permitindo o estudo de projeções de alinhamentos de cada espécie de entrada (Figura 29).

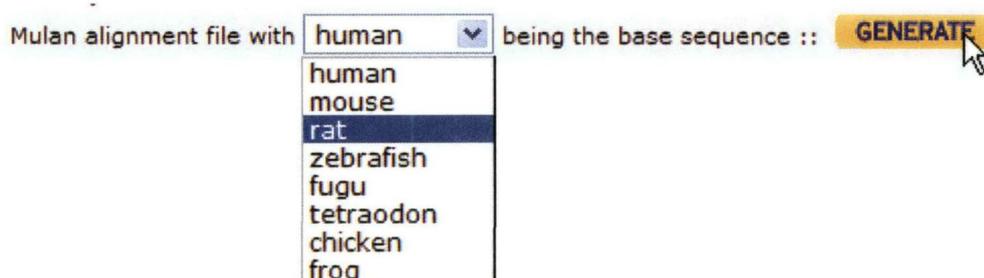


Figura 29. Seleção de uma seqüência referência para a geração de um alinhamento textual do Mulan.

4.1.2.2.4 EXTRAÇÃO DINÂMICA DE ECRS

Cada alinhamento de pares de bases envolvendo uma seqüência referência pode ser investigado para a presença de ECRs. Subseqüentemente, a listagem de posições de ECRs está sendo gerado, listando também seus comprimentos e porcentagem de identidade de seqüência. O usuário pode mudar dinamicamente os critérios para a detecção de ECRs, especificando o comprimento mínimo e a porcentagem de identidade de seqüência mínima. Por exemplo, elementos ultraconservados em alinhamentos humano-camundongo podem ser facilmente extraídos usando critérios de identidade de 100% e 200pb.

4.1.2.2.5 OUTRAS CARACTERÍSTICAS

Há muitas outras características disponíveis da página da *web* de resultados do Mulan. Eles incluem perfis de conservação aos pares, gráficos de pares de bases, atualização dinâmica dos títulos das seqüências, atualização de descrições de genes, tba, refinamento, alinhamentos blast e blastz, e um acesso a todos os arquivos de dados intermediários e de entrada.

4.1.3 PIPMAKER

4.1.3.1 VISÃO GERAL

O PipMaker computa alinhamentos de regiões similares em duas seqüências de DNA. Os alinhamentos resultantes são resumidos em um gráfico de porcentagem de identidade, ou *pip* (*percent identity plot*) (Figura 30).

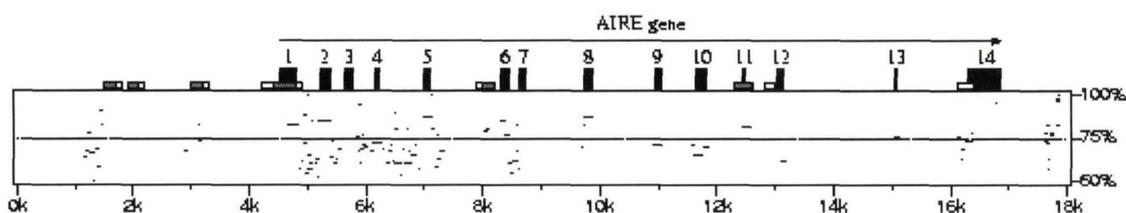


Figura 30. *PIP* (*percent identity plot*). Ícones que podem aparecer ao longo da parte superior da caixa, têm os seguintes significados:

Caixas pontudas brancas são repetições L1.

Triângulos cinza-claro são SINEs, ao invés de MIR.

Triângulos pretos são MIRs.

Caixas pontudas pretas são LINE2s.

Triângulos e caixas pontudas cinza-escuro são outros tipos de repetições intercaladas, como elementos LTR e DNA transposons.

Caixas pequenas cinza-escuro são ilhas CpG onde a razão CpG/GpC excede 0.75.

Caixas pequenas brancas são ilhas CpG onde a razão CpG/GpC fica entre 0.6 e 0.75.

Para gerar um *pip*, como o gráfico acima, o PipMaker requer quatro arquivos fornecidos pelo usuário. O arquivo da primeira seqüência (*first sequence*) é descrito ao longo do eixo horizontal. Repetições intercaladas na primeira seqüência são indicadas por vários tipos de triângulos, cujas localizações são fornecidas por um arquivo “máscara” da primeira seqüência (*mask file of the first sequence*). O usuário gera esse arquivo usando o programa RepeatMasker, disponível na *web* na Universidade de Washington. Um arquivo fornecendo as posições do gene e seus exons (*file of gene and exon positions*) permite que o PipMaker desenhe as localizações dos exons e indique a orientação do gene mostrado como caixas pretas e longas flechas, respectivamente. Finalmente, o usuário fornece um arquivo da segunda seqüência (*second sequence*). Ilhas CpG na primeira seqüência são independentemente determinadas pelo PipMaker e são mostradas como caixas baixas.

O PipMaker compara a primeira e a segunda seqüência. Os alinhamentos são representados de acordo com a posição no arquivo da primeira seqüência. A linha horizontal que atravessa a região central do gráfico indica 75% de identidade de seqüência. Essa versão

do PipMaker compara a primeira seqüência com a segunda seqüência e também com o seu complemento reverso; então regiões combinantes não precisam acontecer na mesma orientação e posição relativa nas duas seqüências. O PipMaker avançado pode, opcionalmente, impor uma condição de que regiões combinantes devem aparecer na mesma ordem relativa e/ou na mesma orientação. Para testar o PipMaker, copie os quatro arquivos fornecidos para o seu computador e os submeta ao PipMaker para gerar o exemplo de *pip* acima.

4.1.3.2 ENTRADA DE DADOS NO PIPMAKER

O PipMaker processa os conteúdos dos quatro arquivos seguintes. Para cada um deles, você pode colar os dados na área para texto com várias linhas, ou, se seu *browser* suportar, dar o nome do arquivo no campo de uma linha, subsequente.

Arquivo de dados de seqüência (*First sequence data file*): é um arquivo no formato FASTA contendo a primeira seqüência, com nucleotídeos em letra maiúscula. O comprimento máximo é de dois milhões de nucleotídeos. Por exemplo:

```
>exatamente uma linha de cabeçalho, o resto contém ACGT SOMENTE
ACGTACGTACGT
CGTACGTACGTA
GTACGTACGTAC
TACGTACGTACG
```

Arquivo de seqüência máscara (*First sequence mask file*): é a documentação produzida pelo RepeatMasker para a primeira seqüência. (NÃO INCLUA a seqüência máscara gerada pelo RepeatMasker.) Assim, o arquivo da seqüência máscara contém entradas como as seguintes:

```
413 5.6 0.0 0.0 HUMAN 1 54 (92195) C Alu SINE/Alu (238) 62 9 SINE/Alu
(238) 62 9
```

Em certas circunstâncias, o RepeatMasker produz alguns resumos estatísticos que precedem linhas do formato acima – essas linhas adicionais são ignoradas pelo PipMaker. A inclusão desse arquivo é opcional, mas é fortemente recomendado para seqüências de mamíferos, pois, caso contrário, combinações biologicamente insignificantes devido a repetições, serão computadas. Se esse arquivo for omitido, então o *pip* não irá incluir ícones mostrando as posições de elementos repetidos intercalados.

Arquivo de exons para a primeira seqüência (*Exons file for the first sequence*). Um arquivo de texto opcional fornecendo as posições de unidades transcricionais da primeira seqüência. A primeira linha é mostrada no lado de cada página do *pip*. Linhas subseqüentes são usadas pelo PipMaker para desenhar as localizações de genes. A orientação de um gene (< ou >), suas posições de início e término, e nome devem estar em uma linha, seguido de linhas separadas, especificando as posições de início e término de cada exon. É também permitido iniciar a linha com '['; em todo caso a linha indicando a posição do gene na primeira seqüência é desenhada sem '>'. Uma linha opcional começando com um sinal "+" pode indicar os primeiros e últimos nucleotídeos da região traduzida (incluindo o códon de iniciação, Met, e o códon de término). Linhas em branco são ignoradas. Os exons devem ser especificados em ordem crescente de localização mesmo se o gene estiver na fita reversa (<). Assim, o arquivo dos exons deve começar como segue:

```

Direção início término nome
> 100 800 Gene 1

100 200 -
300 400 - início e término de diferentes exons
600 800 -

< 1000 2000 Segundo Gene
+ 1100 1900 - primeiro e último nucleotídeo da região traduzida
1000 1200
1800 2000
...

```

Se os números dos exons fornecidos pelo PipMaker não forem adequados, um número (ou um nome) pode ser especificado no final das linhas de posições de exons.

Arquivo de dados da segunda seqüência (*Second sequence data file*): é um arquivo no formato FASTA contendo a segunda seqüência. O comprimento máximo permitido é de 2Mb.

4.1.3.3 SAÍDA DE DADOS DO PIPMAKER

Os três arquivos seguintes são retornados a você como anexos de uma mensagem de e-mail.

Gráfico de porcentagem de identidade (PIP). Esse *pip* é um documento PDF. Se você não possui um leitor de PDF para mostrar o diagrama, você pode conseguir um nos *links get Ghostscript here* ou *Acroread*, do *Adobe*. No momento, o *Acroread* tem melhor suporte para *hiperlinks*, que são usados pelo *PipMaker*. *Ghostview* também pode mostrar *PostScript*. *PipMaker* avançado pode, opcionalmente, gerar *PostScript* ao invés de PDF. *Pips* são descritos com mais detalhes abaixo.

Forma resumida dos alinhamentos. Essa informação é útil para a identificação precisa de posições de seqüência, que correspondem a regiões conservadas indicadas no *pip*. Há uma correspondência de um para um entre linhas nesse arquivo e os pequenos segmentos na linha horizontal do *pip* (onde cada um deles corresponde a uma região entre espaçamentos sucessivos em um alinhamento local). Por exemplo, suponha que após estudar o *pip* do topo dessa página, você se interesse pelos três espaços em torno da posição 3k. No arquivo resumido do alinhamento, você vai encontrar algo como:

```
...
2870-2926 <--> 2281-2337 68% (57 nt)
3117-3128 <--> 2500-2511 83% (12 nt)
3129-3179 <--> 2513-2563 73% (51 nt)
...
```

A primeira linha afirma que as posições 2870-2926 da primeira seqüência (57 pares de base), alinham-se com as posições 2281-2337 da segunda seqüência, sem interrupções e com 68% de identidade de seqüência.

Forma textual tradicional dos alinhamentos. Por exemplo, o seguinte alinhamento corresponde à primeira linha da forma resumida mostrada acima:

```

0      .      :      .      :      .      :      .      :      .      :
2870  GCCCAGGCCTGGGCAGCGAGAGGGCCCTGCTCCCCGCTCAAGGCTCCCAG
      |||||:| ||::||::||:|||| |||||:|:|:| || :|:::|
2281  GCCCAGACATGAACAATGGGAGGCCCTGCTCTCCACTCACAACTTTAG

50      .
2920  GACATTC
      :|||
2331  AACATTC
```

4.1.3.4 INTERPRETANDO UM *PIP*

O *pip* consiste em fileiras que mostram conservação de seqüência e outras características ao longo de segmentos da primeira seqüência. Cada pequena linha horizontal dentro da caixa grande corresponde a uma seção de um alinhamento unido por intervalos sucessivos (ou um término da seqüência). Por exemplo, suponha que um dos alinhamentos computados para suas duas seqüências comece como segue:

```

0      .      :      .      :      .      :      .      :      .      :
19163  GCGGCTCCATGTACCTGCGGGCAAGGGGCTGGTGTGGAAAGCCCCACGG
      ||:|||| || ::|||||||:|::: ||:-||| :::|||| ||||||
6465   GCAGCTACAGACCACCTGCAGGTGTGGA CTGTCACGGAAACCCCCACGT

50      .      :      .      :      .      :      .      :      .      :
19213  CATGGTGGAAAGTCCGAAATTCTACAGGGCCTCTTTGTTAAACCTC
      -||:|||||||:|:|||||||:|:|:|:|:|:| | |:|---
6514   G TGATGGAAAGTCCAAAATTCTACAGGAGTCTTTCTGTTGATCTCCAGT
...

```

Nessa porção do alinhamento, há três pedaços sem intervalos. O primeiro cobre as posições 19163-19190 da primeira seqüência, com 64% de identidade de nucleotídeos; o segundo cobre as posições 19192-19231 com 68% de identidade; e o terceiro cobre as posições 19215-19259, com 78% de identidade. Isso seria descrito no *pip* como três segmentos de linha horizontais que indicam as posições na primeira seqüência e a porcentagem de identidade de seqüência.

4.1.3.5 MENSAGENS DE ERRO

O usuário deve receber precisamente uma mensagem de e-mail do PipMaker – seus resultados ou uma mensagem de erro. Mensagens de erro começam com linhas indicando até que ponto a computação procedeu, e devem terminar com uma linha explicando a causa do fracasso. Normalmente, a mensagem de e-mail deverá ser enviada dentro de poucos minutos desde sua submissão. Entretanto, é possível que seu programa de e-mail se negue a aceitar a mensagem do PipMaker em casos onde uma grande quantidade de dados é produzida.

4.2 LOCALIZAÇÃO DO GENE DA MIOSTATINA NO GENOMA HUMANO

A seqüência referência usada no decorrer deste trabalho foi a do lócus do gene da miostatina humana, que fica localizado na posição 2q32.2 do cromossomo 2 (Figura 31A - obtida através do site Ensembl, em <http://www.ensembl.org>). Em termos estruturais, o gene da miostatina é composto por três exons e dois introns, que codificam um peptídeo de 376 aminoácidos. Este gene fica disposto entre as seqüências que codificam as proteínas hipotéticas MEC13057 (*upstream*) e LOC285011 (*downstream*), orientado no sentido 3'-5' (Figura 31). Para fins de análise dos possíveis elementos regulatórios da transcrição da miostatina, foi analisado o trecho compreendido entre os nucleotídeos 190614448 a 190871187, que demarcam os limites do lócus da miostatina em relação às seqüências MEC13057 e LOC285011 (Figura 31).

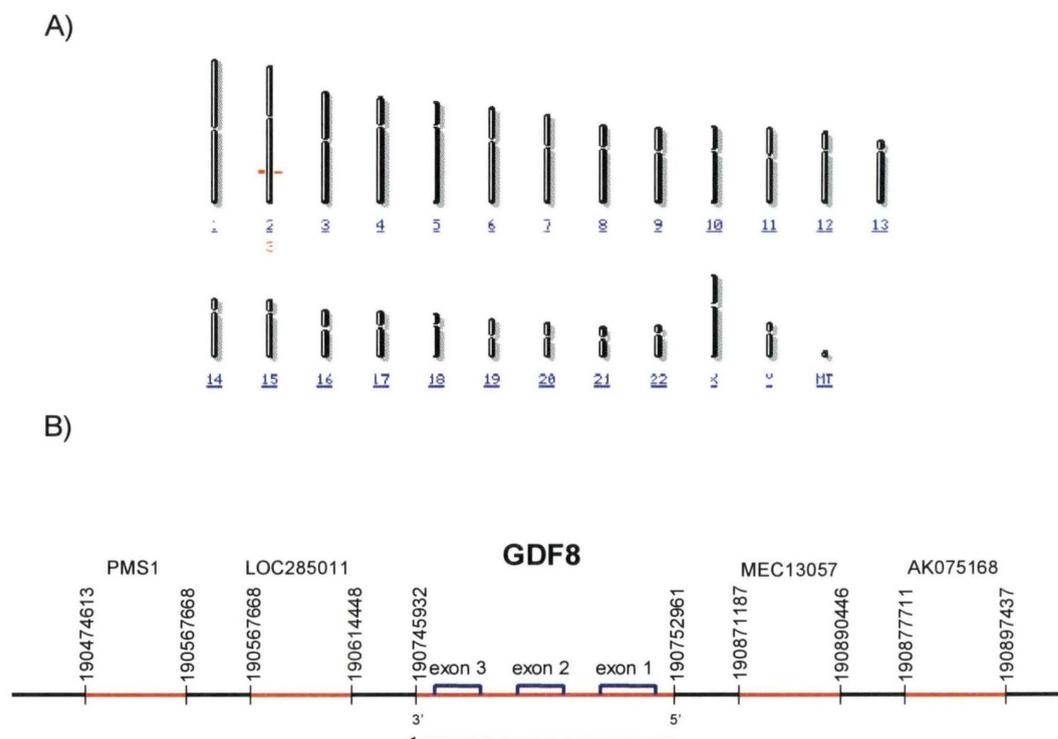


Figura 31. Organização do lócus da miostatina no genoma humano.

A) A região codificadora da miostatina fica localizada no cromossomo 2, região 2q32.2.

B) O gene da miostatina é composto por três exons e dois introns, e fica compreendido entre os nucleotídeos 190752961 e 190745932. A orientação do gene é indicada pela seta. Flanqueando o gene da miostatina ficam as proteínas hipotéticas MEC13057 e LOC285011.

4.3 COMPARAÇÃO DO LÓCUS DA MIOSTATINA DE DIFERENTES VERTEBRADOS

Com auxílio das ferramentas de bioinformática ECR Browser (<http://ecrbrowser.dcode.org/>) e Mulan (<http://mulan.dcode.org/>), foi feita uma comparação dos lóci do gene da miostatina, bem como das regiões intergênicas adjacentes, de diferentes organismos vertebrados (homem, camundongo, galinha, *Xenopus*, *Fugu*, *Tetraodon* e Paulistinha) (Figura 32).

Através dessa comparação, pode-se observar que os genes da miostatina humana e de camundongo são muito semelhantes (Figura 32), havendo elevado nível de identidade de seqüência tanto nos exons quanto em introns e regiões não traduzidas (*untranslated regions* - UTRs). Já quando se compara os lóci de galinha e homem, observa-se elevada identidade de seqüência nas porções referentes aos três exons. Entretanto, as seqüências intrônicas demonstram ser pouco similares entre esses dois organismos. As regiões das UTRs apresentam identidade de seqüência relativamente elevada, especialmente na UTR localizada na região 3', onde se pode observar pelo menos duas ECRs evidentes (Figura 32). A comparação do lócus da miostatina humana com a seqüência correspondente de peixes e anfíbios revelou não haver identidade de seqüência significativa entre as mesmas. Por esse motivo, o gene da miostatina de peixes e anfíbios foi retirado das análises e estudado à parte.

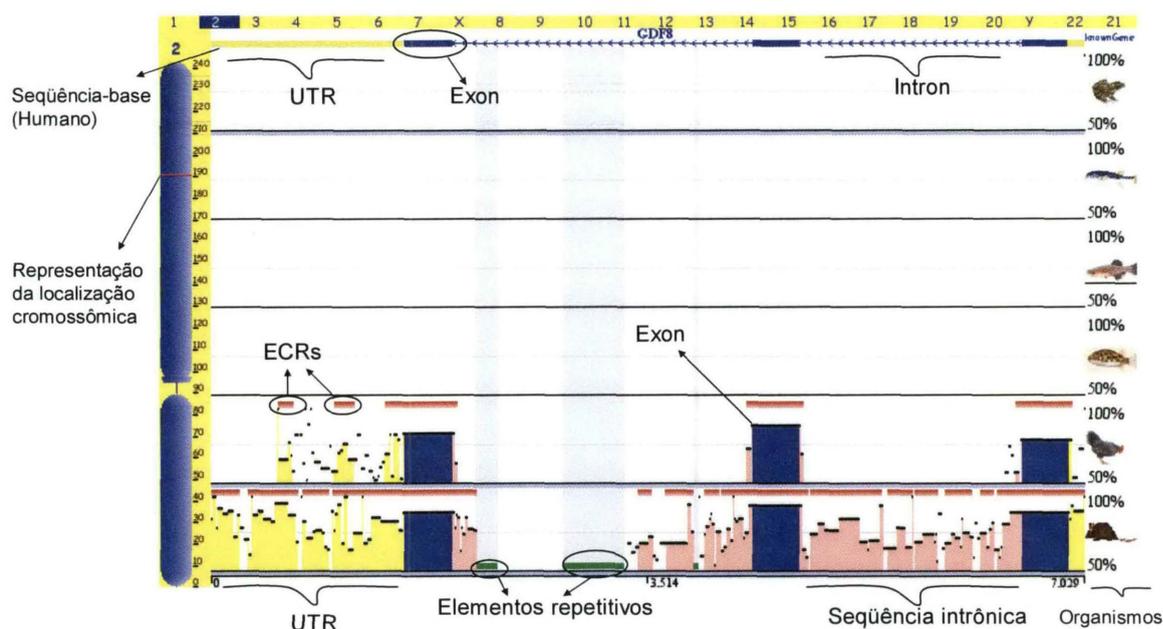


Figura 32. Comparação do lócus da miostatina de diferentes vertebrados. Através da figura, pode-se observar que em humano, camundongo e galinha, o gene da miostatina é composto por três exons conservados (em azul) e dois introns; entre humano e camundongo, há muita conservação de seqüência, mesmo em introns e UTRs; já entre humano e galinha, há apenas ECRs na região das UTRs, além dos três exons. O gene da miostatina de peixes e anfíbios não apresenta similaridade com o do humano, segundo a figura.

A análise da região genômica que contém o gene da miostatina de peixes foi realizada pelo programa PipMaker (<http://pipmaker.bx.psu.edu/pipmaker/>), onde foi feita a comparação entre as seqüências do gene GDF8 de Fugu (que foi usado como seqüência referência), *Tetraodon*, Paulistinha, humano e do gene GDF11 humano (Figura 33). As seqüências comparadas abrangiram a região correspondente ao gene, além de 600 pb das seqüências flangeadoras das extremidades 5' e 3' e foram obtidas no site Ensembl (<http://www.ensembl.org>).

A representação da seqüência do Fugu mostra que nos peixes, o gene da miostatina, assim como em mamíferos e aves, é composto por três exons e dois introns (Figura 33). A identidade de seqüência é elevada entre os três peixes, especialmente entre Fugu e *Tetraodon*. A comparação do lócus GDF8 do Fugu com o humano revelou uma baixa identidade de seqüência, que se resume ao terceiro exon. Entretanto, o GDF8 do Fugu apresentou um nível de similaridade relativamente elevado com o GDF11 humano, demonstrando que o gene da miostatina de peixes é mais próximo ao GDF11 de mamíferos do que ao próprio GDF8.

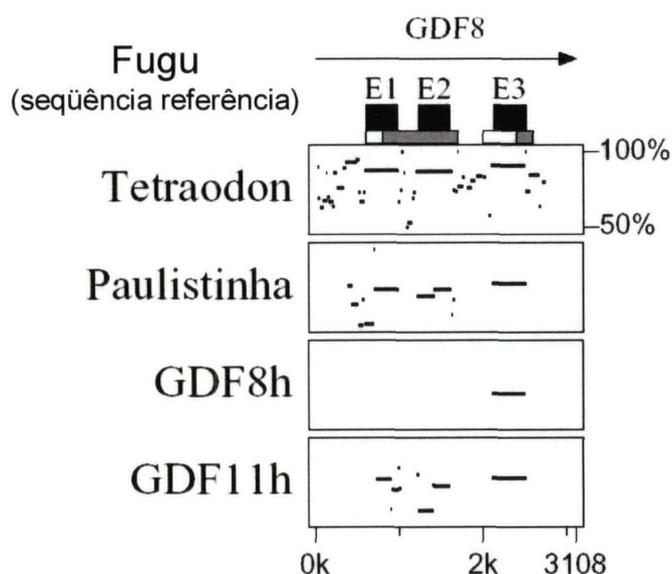


Figura 33. Análise realizada pelo programa PipMaker, das seqüências do gene GDF8 de Fugu (seqüência referência), Tetraodon, Paulistinha e humano, além do GDF11 humano. A seqüência do Fugu mostra a presença de três exons (E1, E2 e E3) e dois introns, tal como nos demais vertebrados. Os traços pretos em cada um dos boxes indica a porcentagem de identidade de seqüência (de 50 a 100%) entre o organismo em questão e a seqüência referência. Através da figura pode-se observar que entre os três peixes há um alto nível de identidade de seqüência, principalmente entre Fugu e Tetraodon. Contudo, quando se compara a seqüência referência (Fugu) com o GDF8 humano, há pouca similaridade. Entretanto, o elevado nível de identidade de seqüência entre GDF8 do Fugu e o GDF11 de humanos, demonstra que o gene da miostatina de peixes é mais próximo ao GDF11 de mamíferos que ao próprio GDF8.

4.4 IDENTIFICAÇÃO DE POSSÍVEIS ELEMENTOS REGULATÓRIOS DA TRANSCRIÇÃO DA MIOSTATINA DE MAMÍFEROS E AVES

A partir dos dados obtidos na Figura 33, foi decidido que as demais análises para identificação de possíveis elementos regulatórios da transcrição do GDF8, seriam feitas com base nos genomas humano, de camundongo e de galinha.

Para facilitar as comparações efetuadas, inicialmente apenas a região correspondente ao gene GDF8 mais UTRs foi analisada. Na Figura 34 são mostrados os sítios de ligação para fatores de transcrição conservados entre os três organismos, de acordo com os resultados da análise realizada pelo programa Mulan. Este programa faz uma busca por potenciais sítios de ligação para fatores de transcrição que sejam conservados entre as seqüências comparadas, reduzindo em até 95% a chance de falsos positivos. Esta análise é feita com base nas informações fornecidas pelo banco de dados Transfac (*The Transcription Factor Database* - <http://www.gene-regulation.com>).

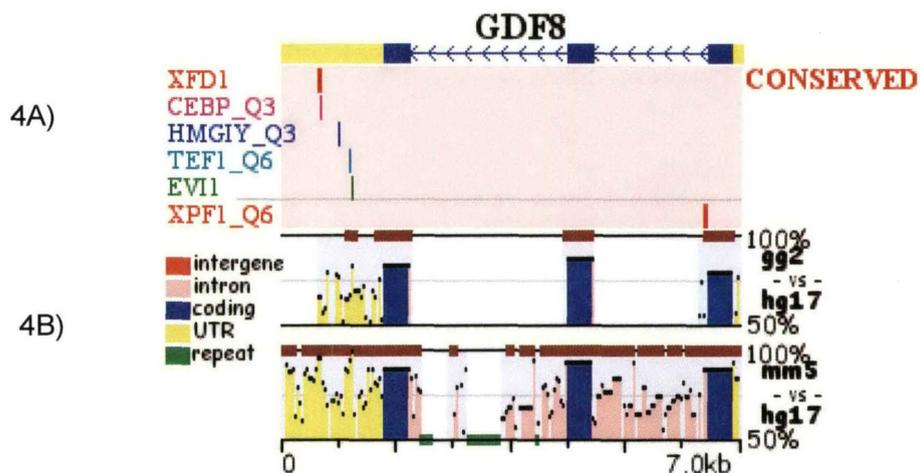


Figura 34A. Sítios de ligação para fatores de transcrição conservados entre humano, galinha e camundongo, no locus da miostatina.

Figura 34B. Localização dos sítios de ligação para fatores de transcrição no locus da miostatina, numa comparação entre humano (hg17), galinha (gg2) e camundongo (mm5).

Em uma segunda etapa, foi realizada a análise de similaridade genômica entre esses três organismos para as regiões intergênicas a 3' (*downstream*) e a 5' (*upstream*) do locus da miostatina.

Na região *upstream* ao gene GDF8, foi encontrada apenas uma ECR entre os três organismos analisados (Figura 35). Devido ao posicionamento dessa ECR com relação ao gene e à sua orientação, presume-se que se trate do promotor do gene da miostatina. A análise realizada para identificação de potenciais sítios de ligação para fatores de transcrição indicou a presença de sítios conservados, entre eles TBP (*TATA Binding Protein*), o que reforça a possibilidade de essa região se tratar do promotor (Figura 36).

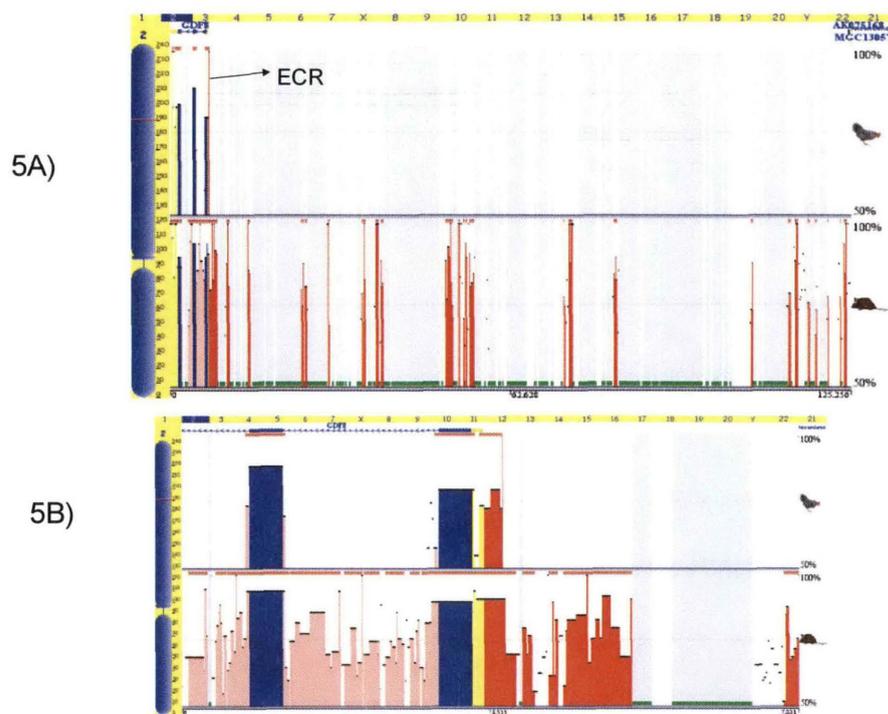


Figura 35. ECRs na região intergênica localizada na extremidade 5' da miostatina.
 A) Há apenas uma ECR que se mantém conservada entre humano, camundongo e galinha, que provavelmente representa o promotor do gene.
 B) Região do promotor putativo em detalhe.

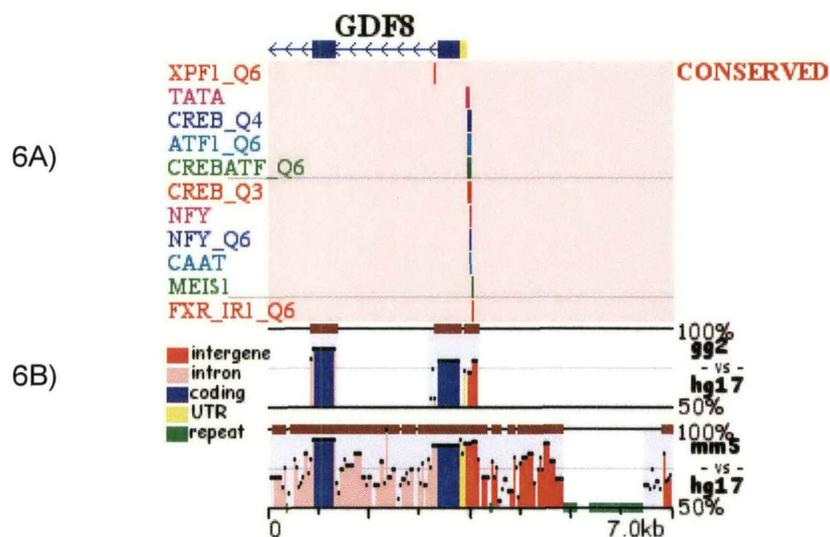


Figura 36A. Sítios de ligação para fatores de transcrição conservados entre humano, galinha e camundongo, na região do promotor putativo.

Figura 36B. Localização dos sítios de ligação para fatores de transcrição na região do promotor putativo, numa comparação entre humano, galinha e camundongo.

A análise de similaridade para a região *downstream* ao gene da miostatina, mostrou a presença de três ECRs, como mostra a Figura 37. Nesta região os sítios de ligação para fatores de transcrição conservados entre os três organismos ficam localizados apenas nas ECRs 2 e 3a (Figura 38).

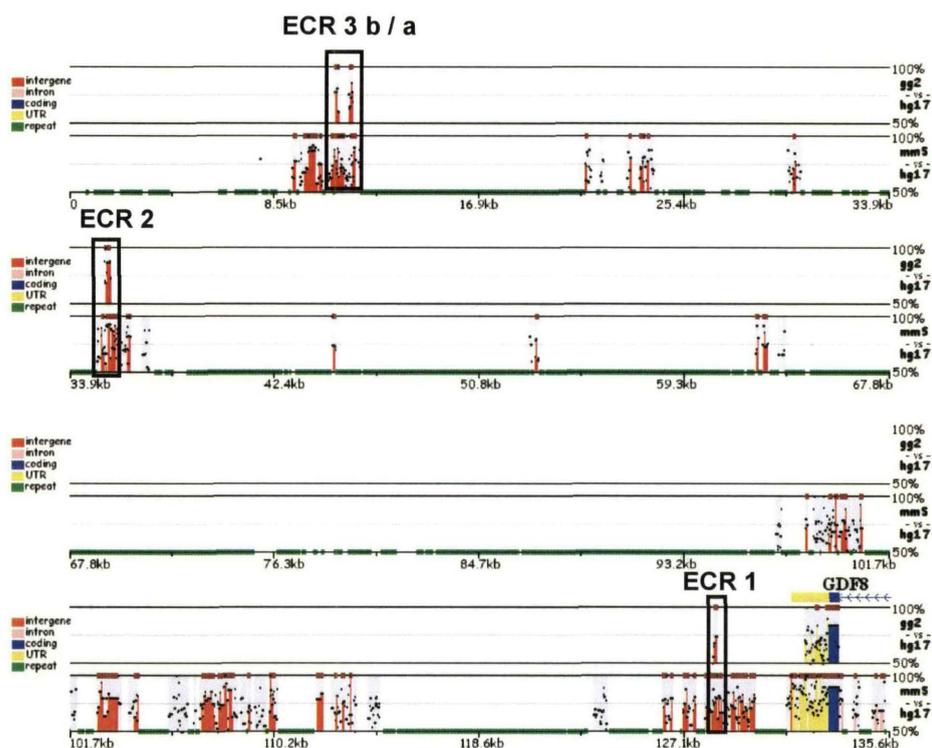


Figura 37. ECRs na região intergênica localizada na extremidade 3' do gene miostatina. Há três regiões conservadas (ECRs 1, 2, 3 a/b) entre humano, camundongo e galinha.

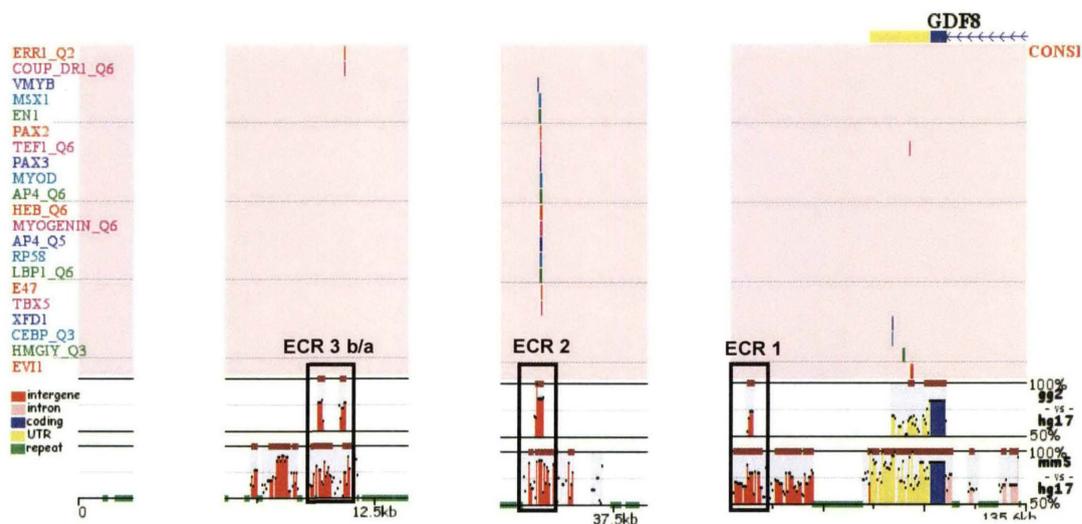


Figura 38A. Sítios de ligação para fatores de transcrição conservados entre humano, galinha e camundongo, na região intergênica à 3' do locus da miostatina.

Figura 38B. Localização dos sítios de ligação para fatores de transcrição na região intergênica à 3' do locus da miostatina, numa comparação entre humano, galinha e camundongo.

Um sumário apresentando o nome, a localização e a função dos fatores de transcrição conservados encontrados no locus da miostatina e nas regiões *upstream* e *downstream* ao gene, estão descritos na Tabela 1.

Tabela 1. Nomes, localizações e funções dos fatores de transcrição para os quais foram encontrados sítios de ligação conservados no locus da miostatina e nas regiões *upstream* e *downstream* ao gene.

Fatores de Transcrição	Posição	Função
AP4_Q5 AP4_Q6	ECR2	Interação funcional com o fator AP-1.
CEBP_Q3	UTR3'	Criticamente envolvido na homeostase energética e pode ter um papel na regulação no balanço entre crescimento e diferenciação celular; camundongos nocauteados para este fator são incapazes de armazenar glicogênio hepático e morrem de hipolicemia; sujeito à auto-regulação positiva; a insulina regula C/EBPalpha e media sua desfosforilação em adipócitos completamente diferenciados.
COUP_ DR1_Q6	ECR3	COUP – membro da super-família do receptor de hormônio esteróide.
E47	ECR2	Parte do complexo MEF-1; induz hiperfosforilação de MyoD;
EN1	ECR2	Necessário para o desenvolvimento normal do cérebro, mutações causam deleções letais; Também atua no desenvolvimento da musculatura esquelética
ERR1_Q2	ECR3	Super-expresso durante a diferenciação de adipócitos; ao se ligar ao gene MCAD, o elemento regulatório parece baixar a transativação induzida por ácido retinóico mediada pelos heterodímeros RAR-alpha/RXR-alpha.
EVI1	ECR4; UTR3'	Envolvido em transformação mielóide
HEB_Q6	ECR2	Ativador transcricional, liga-se com elementos E-box.
HMGIIY_Q3	UTR3'	Fator auxiliar para outros fatores transcricionais como NF-kappaB ou ATF-2 para formar um complexo transcricional.
LBP1_Q6	ECR2	Repressor; a ligação com LBP-1 impede o TFIID de se ligar ao TATA box.

MSX1	ECR2	Pode desempenhar um papel importante na formação dos membros e na diferenciação precoce do epitélio ocular o gene é induzido no broto do membro por um sinal ectodérmico apical;
MYOD MYOD_Q6	ECR2	Ativador transcricional músculo específico, colaborando como outros fatores para a ativação gênica.
MYOGENIN_Q6	ECR2	Ativador transcricional músculo-específico; pode precisar cooperar com outros fatores de transcrição NF-1, MEF-2, COMP1; no broto do membro em desenvolvimento é co-expresso com <i>MyoD</i> ; a miogenina constitui um círculo autoregulatório positivo com MEF-2.
PAX2	ECR2	Ativo durante a transição mesênquima-epitélio no desenvolvimento precoce dos rins; iniciação tumoral através da inibição da transcrição de p53.
PAX3	ECR2	Ativador transcricional; importante para o desenvolvimento de musculatura dos membros; necessário para a migração, mas não para diferenciação de precursores de musculatura dos membros; mutações podem causar fenótipo <i>Spotch</i> (anormalidades na pigmentação da íris e cabelo, deslocamento de olhos, exencefalia, <i>spina bifida</i> , falta de musculatura nos membros); expressão é detectada entre os dias embrionários 8,5 e 14,5; ausência de Sonic hedgehog (Shh) bloqueia a expressão de Pax-3/7; inibição da diferenciação miogênica em células do mioblasto cultivadas (células C2C12).
RP58	ECR2	Repressor (media a repressão transcricional em seqüência com múltiplas cópias do consenso CAGATGTT); associa-se ao núcleo celular com cromatina condensada (provavelmente heterocromatina).
TBX5	ECR2	Crítico para o desenvolvimento de membros e do coração, especialmente na septação cardíaca e na especificação de câmaras; mutações causam a Síndrome de Holt-Oram (HOS), que resulta em anomalias esqueléticas e cardíacas.
TEF1_Q6	ECR2; UTR3'	Ativador; é requerido para o desenvolvimento normal do coração; interferência na interação TEF-1/TFIID por um fator específico de um célula B.

VMYB	ECR2	Ativador transcricional; oncogene, transforma células mielóides e eritróides (fracamente);
XFD1	UTR3'	Ativado no estágio de blástula, desligado após o final da neurulação; é ativado pela activina A.
XPF1_Q6	Intron 1; Promotor putativo	Envolvido na ativação transcricional na expressão gênica pâncreas- específica.
ATF1	Promotor putativo	Ativador; pode inibir a atividade de alguns fatores não-relacionados; media a resposta a concentrações aumentadas de cálcio ou cAMP, mas até um determinado ponto; não auxilia na ativação gênica por E1A, em contraste com CRE- BP1/ATF-2.
CAAT	Promotor putativo	Proteína nuclear ubíqua, envolvida na regulação de genes globina.
CREB; CREBATF	Promotor putativo	Ativador; media resposta cAMP após fosforilação em Ser-133 por PKA dentro do domínio quinase-induzível (KID); KID coopera com o domínio Q2 na estimulação transcricional; Q2 interage com o componente TAF-110 do TFIID; desfosforilação pode ser através de fosfatase da proteína nuclear 2A (PP2A).
FXR-alpha FXR_IR1_Q6	Promotor putativo	Ativador como um heterodímero com RXR-alfa e, dessa forma, fornece ativação do gene dependente de ácidos biliares.
MEIS1	Promotor putativo	Pode ser ativado por inserção retroviral durante leucogênese, como é o caso em camundongos.
NF-Y	Promotor putativo	Ativador; Importante regulador da transcrição do gene MHC classe II;
TBP- TATA-binding protein	Promotor putativo	Requerido para transcrição pelas polimerases I, II e III; A transcrição basal pela pol II é mediada pelo TBP, Sua atividade é regulada por ligação competitiva entre fatores que atuam positiva (ex. TFIIA) e negativamente (ex. NC2, NC1);

5. DISCUSSÃO

5.1 EVOLUÇÃO DO GENE DA MIOSTATINA

5.1.1 PEIXES E ANFÍBIOS

Seqüências relacionadas à miostatina estão presentes em um número de diferentes espécies de peixes, algumas das quais possuem dois genes da miostatina distintos, denominados MSTN-1 e MSTN-2 (KERR *et al.* 2005). A presença de dois genes miostatina no genoma destes organismos é explicada pelo evento de duplicação genômica que ocorreu durante a evolução dos peixes teleósteos e que, aparentemente, precedeu a separação das superordens Acanthopterygii (Fugu) e Ostariophysii (paulistinha) (KERR *et al.* 2005).

As análises para comparação do locus da miostatina de diferentes vertebrados realizadas neste trabalho confirmaram que a estrutura do gene GDF8 de peixes (MSTN-1) e anfíbios difere da estrutura do GDF8 dos demais vertebrados. Os dados obtidos revelaram que há similaridade entre o GDF8 do Fugu e o humano apenas na região do terceiro exon. Essa região codifica o fragmento C-terminal, que é a parte biologicamente ativa da proteína miostatina e que, por isto, vem sendo submetida a uma maior pressão seletiva. Além disto, o gene GDF8 do Fugu demonstrou maior identidade de seqüência com o GDF11 humano do que com o próprio GDF8 humano, demonstrando a proximidade evolutiva entre estes dois genes e a existência de uma provável seqüência ancestral comum a ambos. A divergência de seqüência entre a miostatina de peixes em relação à dos demais vertebrados pode ter levado a diferenças na própria função e regulação da miostatina (LEE 2004).

A existência de diferenças funcionais para a miostatina dos peixes e demais vertebrados é reforçada pelo padrão de expressão deste gene durante a embriogênese e vida pós-natal. Em peixes, o mRNA da miostatina é detectado em tecidos e órgãos como cérebro, olhos, intestino, filamentos das brânquias, gônadas e rins, enquanto em aves e mamíferos a sua expressão é restrita à musculatura esquelética (MACCATROZZO *et al.* 2001). Esse padrão de expressão em peixes sugere que as ações biológicas da miostatina podem não estar restritas ao músculo esquelético, mas podem, adicionalmente, influenciar outros tecidos dos peixes. Estudos recentes tentaram reproduzir o fenótipo de musculatura dupla observado em mamíferos no Paulistinha, mas sem sucesso (XU *et al.* 2003; AMALI *et al.* 2004). Isso sugere que, ou as funções biológicas da miostatina não são bem conservadas em todos os vertebrados, ou há um tipo adicional de miostatina (outro gene) ainda não descrito, que supriria a ausência da primeira (KERR 2005).

No futuro, deverão ser comparados os loci MSTN-1 e MSTN-2 de diferentes peixes bem como de anfíbios, com o intuito de ampliar as análises sobre a regulação da miostatina realizadas neste trabalho.

5.1.2 AVES E MAMÍFEROS

A comparação do locus da miostatina realizada entre humano, camundongo e galinha, demonstrou que as regiões correspondentes aos exons foram muito conservadas ao longo da evolução, confirmando os resultados de *Southern blot* previamente obtidos por LEE & MCPHERRON (2001). Sabe-se atualmente que as seqüências da miostatina de humano, rato, murino, porcino, peru e galinha são idênticas na porção C-terminal biologicamente ativa da molécula (LEE & MCPHERRON 2001).

A função da miostatina também parece ser conservada entre estas espécies. Assim como no camundongo, mutações no gene da miostatina normalmente são associadas com hiperplasia e hipertrofia muscular em outros organismos. Por exemplo, nos bovinos diversas mutações no gene da miostatina vem sendo associadas com o fenótipo de dupla musculatura. Análises do gene da miostatina em um grande número de raças de musculatura dupla identificaram pelo menos sete diferentes mutações, incluindo mudanças no quadro de leitura e mutações de ponto em resíduos de aminoácidos altamente conservados (KAMBADUR *et al.* 1997; MCPHERRON & LEE 1997; GROBET *et al.* 1997, 1998; CAPPUCIO *et al.* 1998; MARCHIELLI *et al.* 2003). Coletivamente, essas mutações são responsáveis pela grande maioria das raças de musculatura dupla caracterizadas, sugerindo que o gene da miostatina é o predominante, senão o único locus para a formação de musculatura dupla em gado (LEE 2004). Recentemente, foi reportado o primeiro caso de mutação da miostatina em uma criança humana. Ao nascer, essa criança apresentava ser extremamente musculosa. Análises genéticas revelaram que ela tinha uma mutação de ponto no seu gene da miostatina. A mutação de ponto resultou em um mRNA com *splicing* errôneo e níveis indetectáveis da proteína da miostatina. Com quatro anos de idade, a criança continua saudável, tem massa muscular aumentada e é extremamente forte. Esses resultados sugerem fortemente que a miostatina exerce um papel similar na regulação da massa muscular esquelética em humanos, como o faz em outras espécies (TOBIN & CELESTE 2005).

Esses achados abrem a possibilidade de que agentes farmacológicos capazes de bloquear a atividade da miostatina possam ter aplicações na promoção do desenvolvimento da musculatura em doenças humanas, bem como em animais de criação (LEE & MCPHERRON 2001).

5.2 POTENCIAIS REGIÕES REGULADORAS DA TRANSCRIÇÃO DA MIOSTATINA

Até o momento, pouco se sabe com relação à regulação da atividade transcricional do gene da miostatina. Neste trabalho, observou-se que a região do gene da miostatina, compreendida entre os nucleotídeos 190745932 e 190752961 do cromossomo 2 de humano, apresentou os exons altamente conservados entre as três espécies analisadas, como já era esperado. Na região correspondente aos dois introns há uma elevada identidade de seqüência

entre o gene humano e de camundongo, sugerindo que possa haver um ou mais elementos regulatórios exclusivos dos mamíferos localizados nos introns. Embora haja pouca similaridade quando esta comparação inclui a galinha, na região do primeiro intron foi encontrado um sítio de ligação para o fator de transcrição XPF1, que está envolvido na ativação transcricional da expressão gênica pâncreas-específica. Embora não pareça haver nenhuma correlação direta entre este fator e o desenvolvimento muscular, o pâncreas através da produção da insulina é um importante regulador do metabolismo de carboidratos dos vertebrados, que de maneira geral atua de maneira contrária ao metabolismo de proteínas.

Com relação às UTRs que flanqueiam o gene GDF8, observou-se alta identidade de seqüência nessas regiões entre humano e camundongo, e a similaridade é menor quando se compara também a galinha. A UTR 5' não apresentou nenhum sítio de ligação para fatores de transcrição conservado entre os três organismos. Já na UTR 3', foi observada a presença de cinco potenciais sítios de ligação para fatores de transcrição. Entre eles destaca-se o sítio para o fator CEBP, criticamente envolvido na homeostase energética e que pode ter um papel na regulação no balanço entre crescimento e diferenciação celular (MAGENTA *et al.* 2003).

No presente trabalho, as análises de comparação do locus da miostatina realizadas entre humano, camundongo e galinha revelaram a presença de apenas uma região conservada entre estes três grupos na região a *upstream* do gene GDF8 (entre os nucleotídeos 190752961 e 190871187 do cromossomo dois de humano). Essa ECR é constituída de 260 pares de base e, devido à sua localização em relação ao gene, representa um promotor em potencial. A busca por potenciais sítios de ligação para fatores de transcrição apontou a presença do sítio de ligação para a proteína de ligação ao TATA Box (TBP), que é requerida para a transcrição pelas polimerases I, II e III. Esse é um grande indício de que essa ECR realmente corresponda ao promotor do gene da miostatina. Estudos práticos serão necessários para a confirmação desses resultados. Além do sítio para a TBP, foram encontrados ainda outros dez tipos de sítios de ligação para fatores de transcrição gerais, porém nenhum sítio para fatores relacionados à formação de musculatura.

Em contraste, na região *downstream* ao gene GDF8 (entre os nucleotídeos 190614448 e 190745932 do cromossomo 2 de humano) foram encontradas três regiões conservadas entre humano, camundongo e galinha. A ECR1 possui 156 pb e não foi identificado nenhum sítio de ligação para fatores de transcrição nessa região que tenha sido conservado evolutivamente.

A ECR2 possui 222 pb e nela foram identificados 15 sítios de ligação para fatores de transcrição, os de fatores que interagem com o E-box. Este achado é particularmente importante pois o E-box (CANNTG) é uma das seqüências reconhecidas como sendo um componente regulatório crítico na expressão de genes relacionados à musculatura. E-boxes são sítios de ligação para os fatores de transcrição básicos hélice-volta-hélice, coletivamente chamados de fatores regulatórios miogênicos (MRFs). Os MRFs incluem os fatores de transcrição *MyoD*, *Myf5*, miogenina e MRF4 (SPILLER *et al.* 2002), que são ativadores transcricionais de genes da musculatura esquelética, capazes de ativar o programa de

diferenciação muscular mesmo quando expressos em células não-musculares (BRENT & TABIN 2002). *MyoD* e *Myf5* são expressos em mioblastos e miotúbulos durante a embriogênese e demonstrou-se que eles são requeridos para a formação de musculatura esquelética. Em comparação a *MyoD*, o mRNA da miogenina aumenta significativamente quando os mioblastos se diferenciam e é, dessa forma, requerido para a formação dos miotúbulos. Entre os sítios de ligação para fatores de transcrição capazes de ligar-se ao E-box que foram encontrados neste trabalho estão um sítio para o fator HEB, sítios para os fatores *MyoD* e miogenina e para o fator E47, que é capaz de formar heterodímeros com *MyoD*. A importância dos E-boxes na regulação da miostatina foi demonstrada em bovinos onde foram identificados cerca de 10 E-boxes distribuídos em três agrupamentos, além de um sítio de ligação para o fator MEF2, também envolvido na diferenciação muscular. Ensaio adicionais revelaram que *MyoD* parece ser um importante regulador da atividade do promotor da miostatina (SPILLER *et al.* 2002).

Além dos E-boxes, que caracteristicamente interagem com proteínas relacionadas ao desenvolvimento da musculatura esquelética, foi encontrado um sítio de ligação para o fator transcricional Pax3, também localizado na ECR2. Este fator é expresso no dermomiótomo em desenvolvimento durante a compartimentalização dos somitos e é, então, regulado positivamente nos lábios dorsomedial e ventrolateral do dermomiótomo. Pax3 é importante para o desenvolvimento muscular (BRENT & TABIN 2002), assim como o fator EN1 (*Engrailed 1*), que também atua no desenvolvimento da musculatura esquelética.

Ainda na ECR2, foram observados sítios de ligação para o fator de transcrição TEF1 (também observado na UTR3'), requerido para o desenvolvimento normal do coração, e para o TBX5. Esse último é crítico para o desenvolvimento de membros e coração, especialmente na septação cardíaca e especificação de câmaras e, segundo AMTHOR *et al.* 2002, estudos recentes detectaram a presença da miostatina em outros tecidos, que não o músculo esquelético, incluindo as fibras de Purkinje do coração.

Os demais sítios de ligação para fatores de transcrição encontrados na ECR 2 não têm relação com o desenvolvimento muscular.

Já na ECR3a, foram observados apenas dois tipos de sítios de ligação para fatores de transcrição. Entre eles, está o sítio para o fator de transcrição ERR1, que é super-expresso durante diferenciação de adipócitos. Isso dá suporte ao achado de MCPHERRON *et al.* (1997), que afirmaram ter observado um sinal claramente detectável de mRNA do GDF8 em tecido adiposo, através de análise do tipo *Northern-blot*. Desta forma, o ECR3a poderia ser responsável pela regulação da expressão da miostatina em adipócitos. Na ECR3b não foi encontrado nenhum sítio de ligação para fatores de transcrição.

Os resultados teóricos desse trabalho serão, no futuro, testados com o auxílio de experimentos práticos que possam vir a confirmar os dados encontrados.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AMALI AA, LIN CJ, CHEN YH, WANG WL, GONG HY, LEE CY, KO YL, LU JK, HER GM, CHEN TT, WU JL: **Up-regulation of muscle-specific transcription factors during embryonic somitogenesis of zebrafish (*Danio rerio*) by knock-down of myostatin-1.** *Dev. Dyn.* 2004, **229**:847-856.

AMTHOR H: **The regulation and action of myostatin as a negative regulator of muscle development during avian embryogenesis.** *Developmental Biology* 2002, **251**:241-257.

AOYAMA H & ASAMOTO K: **Determination of somite cells: independence of cell differentiation and morphogenesis.** *Development* 1988, **104**:15-28.

BRENT AE, TABIN CJ: **Developmental regulation of somite derivatives: muscle, cartilage and tendon.** *Current Opinion in Genetics & Development* 2002, **12**:548-557.

CAPPUCIO I, MARCHITELLI C, SERRACCHIOLI A, NARDONE A, FILIPPINI F, AJMONE-MARSAN P, VALENTINI A: **A G-T transversion introduces a stop codon at the mh locus in hypertrophic marchigiana beef subjects.** *Anim. Genet.* 1998, **29**:51.

CHAUVIGNE F, GABILLARD JC, WEIL C, RESCAN PY: **Effect of refeeding on IGFI, IGFI, IGF receptors, FGF2, FGF6, and myostatin mRNA expression in rainbow trout myotomal muscle.** *General and Comparative Endocrinology* 2003, **132**:209-215.

CHRIST B, ORDAHL CP: **Early stages of chick somite development.** *Anat. Embryol.* 1995, **191**:381-396.

CINNAMON Y, KAHANE N, KALCHEIM C: **Characterization of the early development of specific hypaxial muscles from the ventrolateral myotome.** *Development* 1999, **126**:4305-4315.

DENETCLAW WF, CHRIST B, ORDAHL CP: **Location and growth of epaxial myotome precursor cells.** *Development* 1997, **124**:1601-1610.

GROBET L, MARTIN LJR, PONCELET D, PIROTTIN D, BROUWERS B, RIQUET J, SCHOEBERLEIN A, DUNNER S, MENISSIER F, MASSABANDA J, FRIES R, HANSET R, GEORGES M: **A deletion in the bovine myostatin gene causes the doubled-muscling phenotype in cattle.** *Nature Genet.* 1997, **17**:71-74.

GROBET L, PONCELET D, ROYO LJ, BROUWERS B, RIQUET J, PIROTTIN D, MICHAUX C, MENISSIER F, ZANOTTI M, DUNNER S, GEORGES M: **Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle.** *Genome* 1998, **9**:210-213.

Gu Z, ZHANG Y, SHI P, ZHANG YP, ZHU D, LI H: **Comparison of avian myostatin genes.** *Animal Genetics* 2004, **35**:462-504.

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

KAHANE N, CINNAMON Y, KALCHEIM C: **The origin and fate of pioneer myotomal cells in the avian embryo.** *Mech. Dev.* 1998, **74**:59-73.

KALCHEIM C, CINNAMON Y, KAHANE N: **Myotome formation: a multistage process.** *Cell Tissue Res* 1999, **296**:161-173.

KAMBADUR R, SHARMA M, SMITH TPL & BASS JJ: **Mutations in myostatin (GDF8) in double-muscled belgian blue and piedmontese cattle.** *Genome Res.* 1997, **7**:910-915.

KERR T, ROALSON EH, RODGERS BD: **Phylogenetic analysis of the myostatin gene sub-family and the differential expression of a novel member in zebrafish.** *Evolution & Development* 2005, **7**:5:390-400.

KLEINJAN DA, SEAWRIGHT A, CHILDS AJ, VAN HEYNINGEN V: **Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription.** *Developmental Biology* 2004, **265**:462-477.

KONIGSBERG IR: **The embryonic origin of muscle.** *Myology* 1986, **1**:39-71.

LEE S-J, MCPHERRON AC: **Regulation of myostatin activity and muscle growth.** *PNAS* 2001, **98**:9306-9311.

LEE S-J: **Regulation of muscle mass by myostatin.** *Annu. Rev. Cell. Dev. Biol.* 2004, **20**:61-86.

LIN S-Y, MORRISON JR, PHILLIPS DJ, KRETZER DM: **Regulation of ovarian function by the TGF- β superfamily and follistatin.** *Reproduction* 2003, **126**:133-148.

LUDOLPH DC & KONIECZNY SF: **Transcription factor families: muscling in on the myogenic program.** *FASEB* 1995, **9**:1595-1604.

MACCATROZZO L, BARGELLONI L, CARDAZZO B, RIZZO G, PATARNELLO T: **A novel second myostatin gene is present in teleost fish.** *FEBS Letters* 2001, **509**:36-40.

MCPHERRON AC, LAWIER AM, LEE S-J: **Regulation of skeletal muscle mass by a new TGF- β superfamily member.** *Nature* 1997, **387**:83-90.

MAGENTA A, CENCIARELLI C, SANTA F, FUSCHI P, MARTELLI F, CARUSO M, FELSANI A: **MyoD stimulates RB promoter activity via the CREB/p300 nuclear transduction pathway.** *Mol. Cell. Biol.* 2003, **23**: 2893- 2906.

ORDAHL CP, LE DOUARIN NM: **Two myogenic lineages within developing somite.** *Development* 1992, **114**:339-353.

NOBREGA M, OVCHARENKO I, AFZAL V, RUBIN E: **Scanning human gene deserts for long-range enhancers.** *Science* 2003, **302**: 413.

OVCHARENKO I, LOOTS GG, GIARDINE BM, HOU M, MA J, HARDISON RC, STUBBS L, MILLER W: **Mulan: multiple-sequence local alignment and visualization for studying function and evolution.** *Genome Research* 2005, **15**:184-194.

REHFELDT C, FIEDLER I, DIETL G, ENDER K: **Myogenesis and postnatal skeletal muscle cell growth as influenced by selection.** *Livest. Prod. Sci.* 2000, **66**:177-188.

RESCAN PY: **Muscle growth patterns and regulation during fish ontology.** *General and Comparative Endocrinology* 2005, **142**:111-116.

SCHWARTZ S, ZHANG Z, FRAZER KA, SMIT A, RIEMER C, BOUCK J, GIBBS R, HARDISON R, MILLER W: **Pipmaker: a web server for aligning two genomic DNA sequences.** *Genome Res.* 2000, **10**:577-586.

SPILLER MP, KAMBADUR R, JEANPLONG F, THOMAS M, MARTYN JK, BASS JJ, SHARMA M: **The myostatin gene is a downstream target gene of basic helix-loop-helix transcription factor myod.** *Mol. Cell. Biol.* 2002, **22**:7066-7082.

STOCKDALE FE, NIKOVITS W, CHRIST B: **Molecular and cellular biology of avian somite development.** *Developmental Dynamics* 2000, **219**:304-321.

SWATLAND HJ, KIEFFER NM: **Fetal development of the double-musled condition in cattle.** *J. Anim. Sci.* 1974, **38**:1090-1102.

SWATLAND HJ: **Structure and development of meat animals.** Prentice-Hall, Inc., Englewood Cliffs, NJ. 1984.

TELLGREN A, BERGLUND AC, SAVOLAINEN P, JANIS CM, LIBERLES DA: **Myostatin rapid sequence evolution in ruminants predates domestication.** *Molecular Phylogenetics and Evolution* 2004, **33**:782-790.

TOBIN JF, CELESTE AJ: **Myostatin, a negative regulator of muscle mass: implications for muscle degenerative diseases.** *Current Opinion in Pharmacology* 2005, **5**:328-332.

WASSERMAN WW & SANDELIN A: **Applied bioinformatics for the identification of regulatory elements.** *Nat. Rev. Genet.* 2004, **5**:276-287.

WEBER TE, BOSWORTH BG: **Effects of 28 day exposure to cold temperature or feed restriction on growth, body composition, and expression of genes related to muscle growth and metabolism in channel catfish.** *Aquaculture* 2005, **246**:483-492.

WILT FH & HAKE SC: **Principles of developmental biology.** W. W. Norton & Company, Inc., New York, 2004.

XU C, WU G, ZOHAR Y & DU SJ: **Analysis of myostatin gene structure, expression and function in zebrafish.** *J. Exp. Biol.* 2003, **206**:4067-4079.

ZIMMERS TA, DAVIES MV, KONIARIS LG, HAYNES P, ESQUELA AF, TOMKINSON KN, MCPHERRON AC, WOLFMAN NM, LEE SJ: **Induction of cachexia in mice by systemically administered myostatin.** *Science* 2002, **296**:1486-1488.