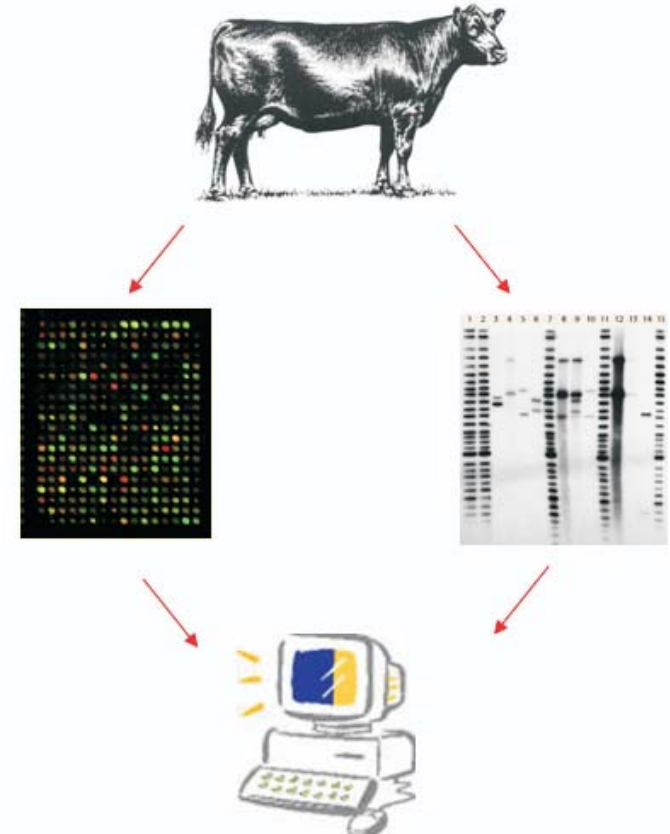


Bioinformática relacionada à descoberta de genes associados com características de interesse econômico em animais pecuários: relatório de pós-doutorado



ISSN 1982-5390

Setembro, 2008

*Empresa Brasileira de Pesquisa Agropecuária
Centro de Pesquisa de Pecuária dos Campos Sulbrasilieiros
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 71

***Bioinformática relacionada à
descoberta de genes associados
com características de interesse
econômico em animais
pecuários: relatório de pós-
doutorado***

Fernando Flores Cardoso

Embrapa Pecuária Sul
Bagé, RS
2008

Exemplares desta publicação podem ser adquiridos na:

Embrapa Pecuária Sul

BR 153, km 603 - Caixa Postal 242

CEP 96401-970 - Bagé, RS

Fone/Fax: (0XX53) 3242-8499

<http://www.cppsul.embrapa.br>

sac@cppsul.embrapa.br

Comitê Local de Publicações da Embrapa Pecuária Sul

Presidente: Alexandre Varella

Secretária-Executiva: Ana Maria Sastre Sacco

Membros: Eduardo Salomoni, Eliara Freire Quincozes, Graciela Olivella Oliveira, Magda Vieira Benavides, Naylor Perez, João Batista Beltrão Marques.

Supervisor editorial: Ana Maria Sastre Sacco

Revisor de texto: Ana Maria Sastre Sacco

Normalização bibliográfica: Graciela Olivella Oliveira

Tratamento de ilustrações: Kellen Pohlmann

Editoração eletrônica: Kellen Pohlmann

Foto da capa: Fernando Flores Cardoso

1ª edição

1ª impressão (2008): tiragem

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados internacionais de Catalogação na Publicação (CIP)

Embrapa Pecuária Sul

Cardoso, Fernando Flores

Bioinformática relacionada à descoberta de genes associados com características de interesse econômico em animais pecuários: relatório de pós-doutorado / Fernando Flores

Cardoso, _ Bagé: Embrapa Pecuária Sul, 2008.

(Embrapa Pecuária Sul. Documentos; 71).

Sistema requerido: Adobe Acrobat Reader.

Modo de acesso:

<<http://www.cppsul.embrapa.br/unidade/publicacoes/list/189>>

Título da página da Web (acesso em 01 jul. 2008).

ISSN 1982-5390

1. Bioinformática. 2. Pecuária. I. Título. II. Série.

CDD 570.285

© Embrapa, 2008

Autor

Fernando Flores Cardoso
Pesquisador A da Embrapa Pecuária Sul

Sumário

Introdução	8
Descrição das atividades desenvolvidas e resultados obtidos	9
a. Desenvolvimento e adaptação de delineamentos experimentais para estudos de genômica genética.....	11
b. Desenvolvimento de ferramentas automatizadas para estudos de genômica genética.....	17
c. Delineamento de um estudo de validação de um micro-arranjo de cDNA para estudos de expressão gênica em suínos.....	18
d. Outras atividades.....	22
Considerações Finais	24
Referências	27
Anexos	29

Lista de Figuras e Tabela

- Fig. 1.** Perfis da varredura genômica, representado pela negativa do logaritmo de base 10 do valor de probabilidade do teste a razão da verossimilhança para diferentes métodos de seleção de amostras para análises de genômica genética..... 15
- Fig. 2.** Quadrado médio do erro (MSE), definido com a variância mais o quadrado do viés das estimativas do efeito do eQTL, para diferentes métodos de seleção de amostras para análises de genômica genética..... 15
- Fig. 3.** Distância média absoluta (MAD) da posição estimada para o eQTL (pico da verossimilhança) em relação a posição efetivamente simulada, para diferentes métodos de seleção de amostras para análises de genômica genética..... 16
- Fig. 4.** Delineamento de referência comum para estudo de validação de micro-arranjo suíno em oito tecidos e quatro animais diferentes..... 20

Fig. 5. Imagens da coleta de tecidos de suínos para extração de RNA nas instalações do Frigorífico do Departamento de Ciência Animal da Michigan State University.....	21
Fig. 6. Representação esquemática da tecnologia de micro-arranjos utilizada para quantificação relativa da expressão gênica de duas amostras, teste e referência.....	21
Fig. 7. Análise de maciez da carne bovina utilizando texturômetro (Texture Analyzer).....	22
Fig. 8. Avaliação da área de olho de lombo com cartelas quadriculadas e a medida da espessura de gordura de cobertura em carcaças bovinas.....	23
Tab. 1. Marcadores genéticos no cromossomo 6 e suas respectivas localizações (cM) no mapa de ligação relativo ao marcador S0099, número de indivíduos com dados perdidos, número de alelos e conteúdo de informação de polimorfismo (PIC).....	12

Bioinformática relacionada à descoberta de genes associados com características de interesse econômico em animais pecuários: relatório de pós-doutorado

Fernando Flores Cardoso

Introdução

Este documento foi elaborado com o objetivo de relatar as atividades desenvolvidas durante treinamento de pós-doutorado realizado no Departamento de Ciência Animal da Universidade do Estado de Michigan (MSU), na área de bioinformática, com foco em estatística genômica, durante o período de 23/06/2006 a 23/12/2006. A MSU está localizada em East Lansing, Michigan, sendo reconhecida globalmente como uma das grandes universidades públicas dos Estados Unidos da América. A MSU foi fundada em 1855 sendo umas das primeiras Universidades criadas naquele país com foco direcionado a ensino, pesquisa e extensão em agricultura, tradição que permanece até hoje. O Departamento de Ciência Animal completou no ano passado 100 anos de excelência em ciência e tecnologia para a produção animal, possuindo tradição, excelente infraestrutura e pessoal altamente capacitado para este fim. Esses fatores, aliados à existência de um projeto na área de interesse em andamento da MSU, no qual haviam recursos disponíveis para custear parcialmente o treinamento, determinaram a escolha do local de treinamento.

Para o desenvolvimento do trabalho durante o treinamento tive disponível uma sala ampla e individual, com dois computadores para rodar as análises estatísticas, sendo um deles uma estação de trabalho Dell Precision Xeon com dois processadores "Core Duo" e alta capacidade de

processamento. A estrutura de bibliotecas físicas e virtuais da MSU facilitou o acesso a publicações e referência bibliográficas para o aprendizado durante o treinamento. Além disso, tivemos o benefício de acompanhar atividades dos laboratórios de ciência da carne, biologia molecular e do centro de genômica funcional, todos localizados no departamento.

Além da infra-estrutura foi essencial ao treinamento, em particular à aquisição de experiência prática, a disponibilidade de dados de marcadores moleculares, pedigree e fenótipos da população experimental F2 de suínos desenvolvida pela MSU.

Descrição das Atividades Desenvolvidas e Resultados Obtidos

Durante este treinamento de pós-doutorado em bioinformática foram adquiridos conhecimento e experiência prática em diversas metodologias relacionadas à descoberta de genes associados com características de interesse econômico em animais pecuários. Em particular foram abordados os seguintes temas: a análise de dados de desempenho e de marcadores genéticos para identificar loci de características quantitativas (QTL), análise de experimentos de micro-arranjos de cDNA para avaliar expressão gênica diferencial, delineamento e análise de varredura genômica de expressão gênica global (genômica genética) e mapeamento de desequilíbrio de ligação para identificar QTL.

O contexto do treinamento foi associado à participação no projeto “Integração de perfis transcricionais com dados de marcadores genéticos para identificar genes controlando características de importância econômica em suínos”. Um projeto inovador que integra, pela primeira vez em espécies pecuárias, dados de perfis de expressão gênica e marcadores genéticos de uma população referência F2, para combinar o poder de recombinação com a análise funcional na identificação de genes controlando crescimento do músculo esquelético, deposição de gordura e qualidade da carne.

Apesar de o projeto ser relacionado com genética suína, as ferramentas estatísticas desenvolvidas e utilizadas nesta pesquisa são pertinentes para análise de dados de outras espécies pecuárias, tais como gado de

corte e leite e ovinos. Este projeto envolve 960 suínos F2 gerados a partir de uma população referência Duroc x Pietrain estabelecida na MSU, com uma rica e abrangente base de dados contendo fenótipos de crescimento, mérito de carcaça, qualidade da carne e sensoriais, bem como dados de marcadores genéticos, em suínos F0, F1 e F2. Adicionalmente, estão sendo gerados dados de expressão gênica em micro-arranjos, comparando fenótipos extremos dentro de leitegada para área de músculo de lombo e gordura de cobertura, com base em um delineamento experimental de medidas fenotípicas (fenotipagem) seletivas. Os objetivos primordiais do projeto são estimar efeitos de loci de características quantitativas e expressão gênica em conjunto com fenótipos para fazer inferência sobre potenciais ciclos regulatórios, buscando a descoberta de novos genes candidatos para características de crescimento, mérito de carcaça e qualidade da carne, com importância econômica na produção suína. Os objetivos específicos para uso dessa população são:

1. Conduzir um estudo de perfilhamento transcricional em tecidos do músculo esquelético e adiposo subcutâneo derivados de animais F₂ exibindo fenótipos extremos para área do músculo do lombo e deposição de gordura para identificar genes com expressão diferencial;
2. Realizar uma ampla varredura genômica usando marcadores microssatélites para detectar loci de característica quantitativa para crescimento, mérito de carcaça e qualidade da carne e refinar posições de supostos QTLs no mapa, através do mapeamento de marcadores microssatélites adicionais dentro dessas regiões de QTL.
3. Integrar resultados de perfis de expressão e varreduras genômicas para mapear genes com expressão diferencial no objetivo 1, refinando a varredura genômica pela inclusão de novos marcadores gene-específicos e pela modelagem dos padrões de expressão gênica com características quantitativas;
4. Validar os genes e QTLs derivados experimentalmente em populações comerciais e repassar a informação desenvolvida à indústria.

A seguir são descritas as atividades desenvolvidas, durante o período de treinamento:

a. Desenvolvimento e adaptação de delineamentos experimentais para estudos de genômica genética

A análise genética da expressão gênica global surgiu recentemente como método de preferência para descobrir genes e redes de genes subjacentes à variação em características biológicas complexas. Este novo campo de estudo, também chamado de genômica genética (JANSEN; NAP, 2001), combina análise de ligação com perfis de expressão gênica e caracteres quantitativos, para relacionar a variação na sequência de DNA com a variação fenotípica, na busca do entendimento do papel da regulação da expressão gênica e seus efeitos nas funções biológicas (ROCKMAN; KRUGLYAK, 2006). A abundância do transcrito de milhares de genes avaliada em experimentos de micro-arranjos de DNA complementar (cDNA; ver também seção III.c. deste relatório) é tratado como uma característica quantitativa em uma população estruturada e com informações de parentesco, sendo sua base genética descrita pela associação com posições do mapa de ligação, ou seja por eQTL, e pela sua herdabilidade poligênica. Uma vez que cada transcrito está associado com o gene que o codifica e que provavelmente tem sua posição conhecida no genoma, os eQTL podem ser considerados com tendo regulação local, quando mapeiam próximo a posição genômica do gene que codifica o transcrito, ou como de regulação distantes, quando mapeiam em outras posições do genoma (ROCKMAN; KRUGLYAK, 2006). Combinar informações de eQTLs com QTLs coincidentes para fenótipo de interesse econômico proporciona um maneira poderosa para identificar genes candidatos para as características de interesse econômico.

Entretanto, o delineamento de experimentos em genômica genética tem como desafio maximizar o aproveitamento de micro-arranjos de alto custo individual, o que, em geral, limita bastante o tamanho da amostra possível. Por outro lado, o custo de genotipagem é relativamente menor, especialmente em chips de larga escala, e é de se esperar que na maioria das situações haja um conjunto de indivíduos genotipados disponíveis para o experimento maior do que os recursos disponíveis para o perfilhamento de expressão. Consequentemente, a escolha da amostra a ser usada é um fator crucial para o sucesso e eficiência do experimento e, portanto, diversos métodos para eleger subconjuntos de indivíduos para avaliar expressão gênica foram propostos.

Durante o pós-doutorado fomos solicitados a desenvolver um amplo estudo de simulação para propor e avaliar variações das estratégias que estavam disponíveis na literatura. Especificamente foram propostas modificações nas seguintes estratégias, de modo que as mesmas fossem aplicáveis a populações F2 abertas: dissimilaridade genética de Jin et al. (2004), máxima recombinação (JANNINK, 2005), e o método de Wang e Nettleton (2006). O desempenho desses métodos foi comparado através da precisão da estimação dos efeitos e localização dos eQTL. Duas alternativas adicionais foram testadas: uma sendo a seleção meramente aleatória e a outra baseada no perfilamento dos indivíduos extremos (com máximo e mínimo valor para um fenótipo de interesse) dentro de sub-classe. Além disso, foram comparados diferentes métodos de análise de dados oriundos de fenotipagem seletiva para diferentes níveis de correlação e intensidades de seleção (80/510, 160/510 e 240/510).

O estudo de simulação, para refletir ao máximo uma situação real, foi baseado no mapa de ligação e genótipos dos marcadores observados no cromossomo 6 da população de recurso suína da MSU (Tabela 1).

Tabela 2. Marcadores genéticos no cromossomo 6 e suas respectivas localizações (cM) no mapa de ligação relativo ao marcador S0099, número de indivíduos com dados perdidos, número de alelos e conteúdo de informação de polimorfismo (PIC).

Marcadores genéticos	Localização do Mapa	Número de genótipos perdidos	Número de alelos	PIC
S0099	0	42	5	0,7231
SW2406	22,4	6	5	0,5157
SW2525	50,6	10	4	0,5535
S0087	81,3	10	4	0,4781
S0220	98,1	13	2	0,2717
SW122	103,9	9	5	0,6357
SW1881	135,7	26	4	0,5809
SW322	164,8	3	7	0,7154
SW2419	181	12	7	0,7533

Foram gerados conjuntos de dados bivariados concebendo-se duas características, uma característica fenotípica tradicional e outra de expressão gênica, com três níveis de correlação entre as características: alto, médio e baixo. Um eQTL com efeito bialélico foi gerado na posição 92 cM, entre os marcadores S0087 e 0220 com efeito pleiotrópico nas duas características, explicando 10% da variação fenotípica de cada uma. Os alelos do eQTL foram considerados fixos nas raças fundadoras da população. O modelo inclui além do efeito aditivo do eQTL, uma média geral, sexo e os efeitos aleatórios de família e poligênico.

Uma vez que os dados completos foram gerados, as cinco estratégias a seguir descritas foram utilizadas para obter sub-amostras da característica de expressão:

- Seleção aleatória (Random) de indivíduos, todos com igual probabilidade.
- Seleção fenotípica (Phenotypic) baseada na escolha dos indivíduos com maior e menor valor para a característica fenotípica dentro de sexo e família. Esta foi a estratégia empregada no projeto da MSU antes da realização deste estudo.
- Dissimilaridade genética (Line Dissim Chr) similar ao proposto por Jin et al. (2004), entretanto, usando a probabilidade de mesma linha de origem em cada posição do genoma como medida de similaridade genética em vez da identidade de estado dos alelos dos marcadores, como originalmente proposto.
- Fenótipos extremos dentro de genótipo do QTL (Nettleton) similar ao proposto por Wang e Nettleton (2006). Os indivíduos foram atribuídos a grupos de genótipo AA, AB e BB baseado nas probabilidades de pertencerem a cada um desses grupos de linhas de origem e os indivíduos de fenótipo mais extremo dentro de cada grupo (caudas da distribuição) foram escolhidos.
- Máxima recombinação (Line Max Recomb) similar ao proposto por Jannink (2005), entretanto considerando linha de origem em vez do genótipo nos marcadores para selecionar os indivíduos que apresentavam o maior número de recombinações.

As análises dos dados simulados foram realizadas usando o software QxPak (PEREZ-ENCISO; MISZTAL, 2004) usando três abordagens: análises bivariadas usando dados completos da característica fenotípica para recuperar parcialmente a informação perdida pela seleção na amostra para expressão; análise de covariância usando o fenótipo como covariável para o modelo da característica de expressão e análise univariada usando somente informações da característica de expressão.

Os métodos de seleção e análise foram comparados pelas seguintes medidas:

- Perfil da varredura genômica, representado pela negativa do logaritmo de base 10 do valor de probabilidade do teste a razão da verossimilhança.
- Quadrado médio do erro (MSE): definido com a variância mais o quadrado do viés das estimativas do efeito do eQTL.
- Distância média absoluta (MAD) da posição estimada para o eQTL (pico da verossimilhança) em relação a posição efetivamente simulada.

As análises bivariadas se mostraram como método mais apropriado para análise de dados com fenotipagem seletiva, pois são capazes de recuperar parcialmente a informação perdida pela pré-seleção, especialmente quando a correlação entre o fenótipo de interesse e característica de expressão gênica tem alta correlação. Por outro lado, as análises de covariância se mostraram inapropriadas por remover parte da variação genética na expressão, diminuindo o poder de detecção dos eQTL. Em linhas gerais, para as análises bivariadas, as diferenças significativas entre os métodos foram observadas na menor proporção de indivíduos selecionados (80/510) e apontam os métodos de dissimilaridade genética e de extremos dentro de genótipo de QTL como apresentando o maior poder de detecção do QTL (ver perfis de varredura genômica na Figura 1) e maior precisão na estimativa dos efeitos do QTL (ver na Figura 2 os menores quadrados médios do erro para esses métodos).

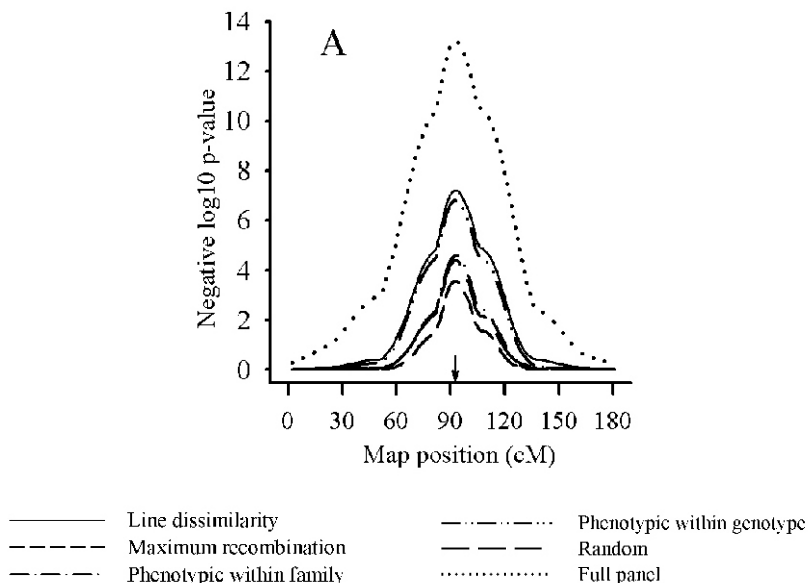


Fig. 1. Perfis da varredura genômica, representado pela negativa do logaritmo de base 10 do valor de probabilidade do teste a razão da verossimilhança para diferentes métodos de seleção de amostras para análises de genômica genética.

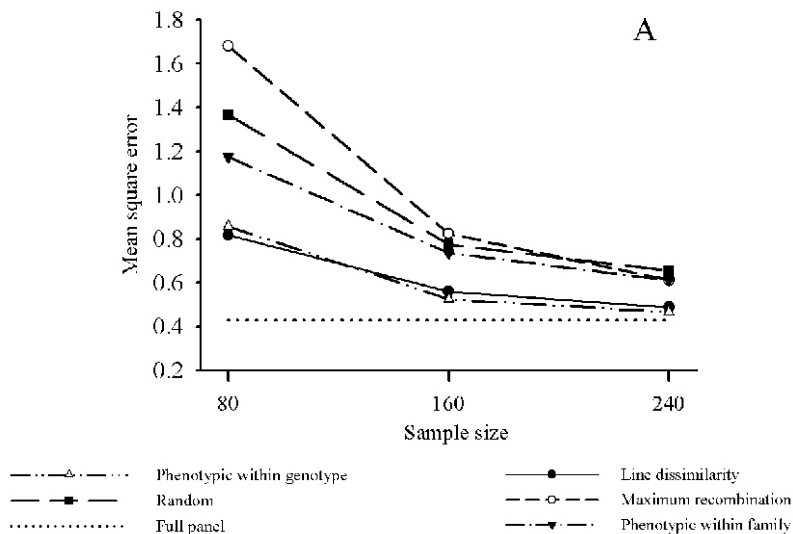


Fig. 2. Quadrado médio do erro (MSE), definido com a variância mais o quadrado do viés das estimativas do efeito do eQTL, para diferentes métodos de seleção de amostras para análises de genômica genética.

Por outro lado, não foram observadas diferenças significativas entre métodos em termos de determinação da posição do QTL, como observado pelas distâncias absolutas médias representadas na Figura 3.

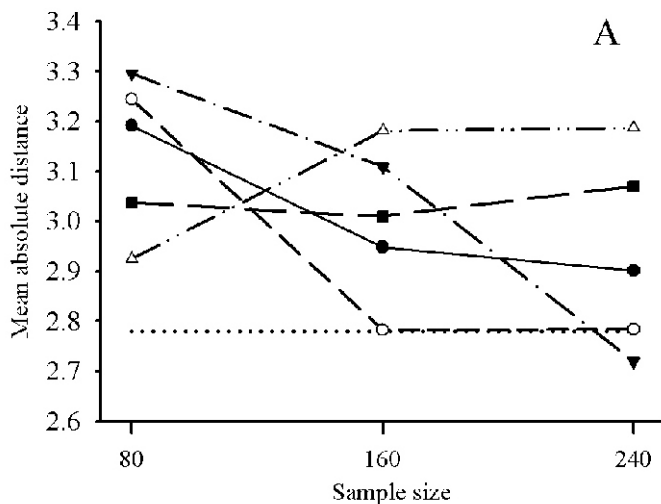


Fig. 3. Distância média absoluta (MAD) da posição estimada para o eQTL (pico da verossimilhança) em relação a posição efetivamente simulada, para diferentes métodos de seleção de amostras para análises de genômica genética.

Dentre as estratégias avaliadas, o uso da dissimilaridade genética e de extremos fenotípicos dentro genótipo de QTL foram os mais eficientes na recuperação da informação e, portanto, no uso dos recursos disponíveis, sendo recomendados para o delineamento experimental de estudos de expressão gênica global. Cabe notar que o método de dissimilaridade genética não requer a detecção prévia de QTL para um fenótipo de interesse nem informação fenotípica alguma, sendo, portanto, mais robusto para ser empregado em situações gerais.

b. Desenvolvimento de ferramentas automatizadas para estudos de genômica genética

Conforme mencionando anteriormente nesse relatório, as análises de genômica genética envolvem o uso da abundância do transcrito de RNA (expressão) em milhares de genes considerando-a como uma característica quantitativa (fenótipo) para varredura genômica visando identificar loci associados com a variação em expressão gênica (eQTL). Desta forma, para viabilizar tais análises, é imprescindível ter ferramentas e procedimentos automatizados de implementação.

Foi nosso papel durante o treinamento desenvolver um procedimento que integrando dois pacotes estatísticos, o SAS e o Qxpak (PEREZ-ENCISO; MISZTAL, 2004), permitisse a automatização no processo de análise estatística em genômica genética. Nesse procedimento uma programação em macro do SAS foi desenvolvida para fazer a leitura inicial e formatação dos dados e montar os arquivos necessários para rodar o Qxpak, inclusive o arquivo de parâmetros que define o modelo a ser utilizado. Uma vez concluída a geração dos arquivos necessários a própria macro invoca o programa Qxpak que faz a varredura genômica para detecção dos eQTL e gera um arquivo de saída contendo todos os resultados da análise (ANEXO I). Cada análise é gravada em um diretório específico e uma segunda macro, também em linguagem do SAS, faz a leitura em série dos arquivos de saída com resultados de todas as características analisadas, extrai as informações relevantes e sumariza por itens (componentes de variância, teste de significância, soluções para efeitos do modelo, etc.) (ANEXO II).

Esses procedimentos serão utilizados para identificar genes candidatos com efeito direto e/ou regulador na expressão gênica e a associação entre expressão gênica e os fenótipos para produção e qualidade da carne.

c. Delineamento de um estudo de validação de um micro-arranjo de cDNA para estudos de expressão gênica em suínos

O delineamento experimental e execução de um estudo de validação para um novo micro-arranjo de oligonucleotídeos, recentemente desenvolvido pelo consórcio do genoma suíno nos EUA, tiveram parte durante o treinamento. O delineamento previu a coleta e análise de oito tecidos suínos (fígado, músculo esquelético, gordura subcutânea, rim, cérebro, amígdala e linfonodo mesentérico) de relevância para as pesquisas dos membros do comitê. Todos os tecidos foram coletados em quatro indivíduos de mesmo sexo, idade, condições de saúde e manejo, no mesmo dia de abate, permitindo o controle de variáveis experimentais e legitimando as comparações nos padrões de expressão gênica nos diferentes tecidos. Foi proposto e adotado um delineamento de referência comum onde todos os tecidos foram testados com respeito a um pool desses mesmos oito tecidos (Figura 4). Na Figura 4 também são representadas hibridizações do tecido com si próprio, as quais foram realizadas em uma amostra de cada tecido de um dos animais e duas amostras da referência comum, com o objetivo de validação da técnica e de obter a distribuição da hipótese de nulidade, uma vez que neste caso não há expressão diferencial (mesma amostra).

O delineamento principal teve 32 micro-arranjos (oito tecidos x quatro indivíduos) e teve os corantes invertidos (dye swap) de duas formas: entre animais dentro de tecido e dentro de animais entre tecidos. Isto teve a finalidade de controlar qualquer possível interação (preferência) de corante por tecido.

Um modelo misto foi proposto para a análise dos dados visando identificar expressão diferencial de genes entre tecidos. O modelo neste tipo de delineamento proposto inclui os efeitos fixos de corante, tecido e interação corante por tecido e os efeitos aleatórios de arranjo, animal e interação animal por tecido.

As amostras foram coletadas em oito fêmeas (somente quatro foram usadas nas análises) em idade de abate oriundas da fazenda experimental da MSU e abatidas no frigorífico do Departamento de Ciência Animal.

Imediatamente após a morte do animal os tecidos foram extraídos, picados e congelados em nitrogênio líquido (Figura 5). O RNA foi extraído dessas amostras para ser usado no experimento, o qual foi conduzido no Centro de Genômica Animal do mesmo departamento. Esse RNA foi reversamente transcrito em DNA complementar (cDNA) o qual foi hibridizado nas placas de micro-arranjos. Em cada placa foram colocados o cDNA oriundo de um tecido mais a referência, cada um marcado com um corante diferente, verde ou vermelho. Os dados de abundância da transcrição do RNA (expressão gênica) foram obtidos através de scanner a laser que excita os corantes fluorescentes que previamente são adicionados aos tecidos. Isto é possível pois quanto maior a quantidade do transcrito maior será a ligação da amostra à sonda do gene impressa no micro-arranjo. Se a amostra corada em verde (vermelha) for a mais abundante a imagem será verde (vermelha), caso as duas amostras sejam expressas igualmente a imagem será amarela e finalmente se não houver expressão em nenhuma das amostras, será preta. A imagem obtida foi analisada usando o software Genepix 6.0 que detecta o local de cada gene ("spot"), corrige para a intensidade de luz de fundo ao redor do "spot" e obtém a intensidade de luz nos dois canais, verde e vermelho. Na Figura 6 é apresentada uma representação esquemática dessa técnica.

Os dados deste experimento ainda estavam sendo coletados ao final do treinamento, entretanto continuarei integrado no projeto, atuando a distância no aperfeiçoamento dos modelos e procedimentos e dando suporte às análises de dados.

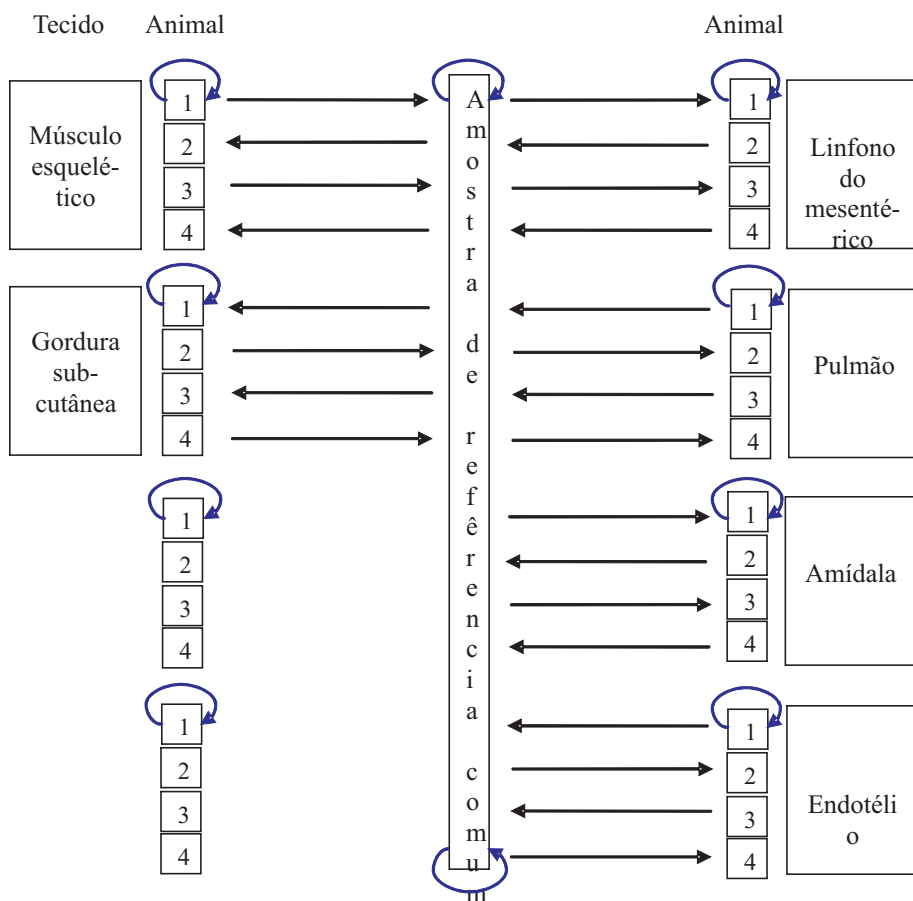


Fig. 4. Delineamento de referência comum para estudo de validação de micro-arranjo suíno em oito tecidos e quatro animais diferentes. Cada seta indica um slide e sentido indica o corante. Setas circulares indicam a prova de uma amostra com ela mesma (self-self).



Fig. 5. Imagens da coleta de tecidos de suínos para extração de RNA nas instalações do Frigorífico do Departamento de Ciência Animal da Michigan State University.

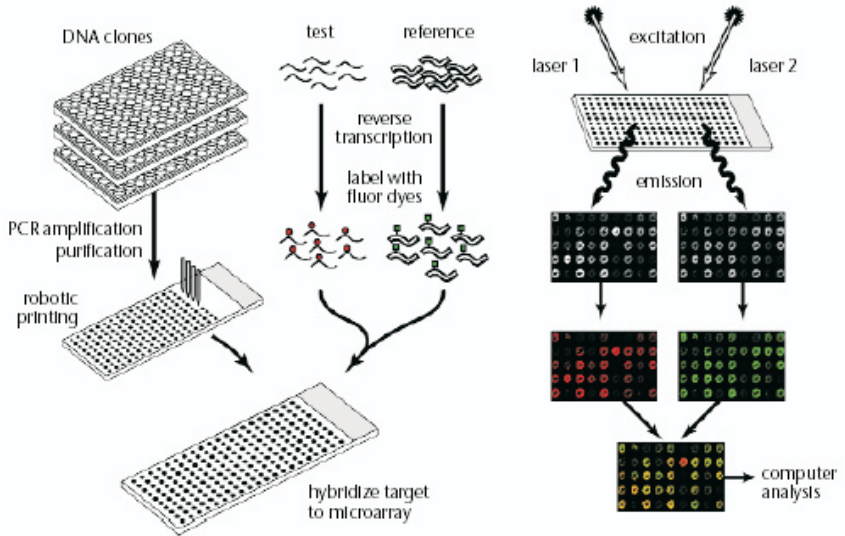


Fig. 6. Representação esquemática da tecnologia de micro-arranjos utilizada para quantificação relativa da expressão gênica de duas amostras, teste e referência.

d. Outras atividades

Embora o foco de treinamento no projeto tenha sido em estatística genômica, tive a oportunidade de acompanhar outras atividades do projeto e do laboratório de ciência da carne, entre elas a extração de DNA de células brancas para genotipagem de marcadores adicionais em regiões promissoras visando aumentar a resolução de mapeamento de QTLs, previamente detectados na população.

No laboratório de ciência da carne, realizamos todo o processo de avaliação da maciez da carne em 30 amostras de novilhos de uma população experimental da MSU, formada pelo cruzamento das raças Angus e Limousine para identificação de genes de importância econômica. O processo iniciou com a preparação dos bifes de uma polegada, passando pela cocção destes em assadeira apropriada, preparação das amostras em cilindros paralelos a fibra muscular, e finalizando com calibração e avaliação no texturômetro (Texture Analyzer) (Figura 7). Foi praticada também a avaliação da área de olho de lombo com cartelas quadriculadas e a medida da espessura de gordura de cobertura em carcaças bovinas (Ver Figura 8.).



Fig. 7. Análise de maciez da carne bovina utilizando texturômetro (Texture Analyzer).



Fig. 8. Avaliação da área de olho de lombo com cartelas quadriculadas e a medida da espessura de gordura de cobertura em carcaças bovinas.

Além das atividades relacionadas ao projeto de genômica, durante a estada nos EUA, foi concluída a elaboração de um artigo científico, referente a avaliação genética de populações multirraciais (CARDOSO et al., 2007), o qual já foi submetido e aceito pelo periódico indexado Journal of Animal Science. Esse trabalho também foi apresentado em pôster na seção solene de comemoração dos 100 do Departamento de Ciência Animal da MSU.

Tive também a oportunidade de participar, junto com o Dr. Alain C. E. Moreau, de uma reunião do grupo de trabalho WG3 da ISO (International Standard Organization), que trata de padrões de identificação eletrônica de animais e sua aplicação nos sistemas de rastreabilidade e certificação, e que congrega especialistas nesta área de diversas partes do mundo. A reunião ocorreu em setembro de 2006 na Universidade do Estado de Kansas, em Manhattan/KS. Nesta ocasião, além de participar das discussões técnicas do grupo, visitamos um local de leilão de bovinos, o maior do estado, na cidade de Salinas, que tinha recentemente adotado um sistema automático para controle dos animais comercializados, através de brincos e/ou bolus de identificação eletrônica e leitoras com antenas em forma de painel localizadas na saída da mangueira de comercialização, as quais estão conectadas ao sistema de informação da companhia leiloeira. Ainda neste tema, visitamos a empresa Digital Angel que fabrica esses dispositivos com a marca Destron® em Saint Paul/Minnesota. Além de conhecer o programa de pesquisa da empresa, acertamos a doação de identificadores eletrônicos, leitoras e antenas para serem usados em animais de um projeto de cruzamentos da Embrapa Pecuária Sul.

Também em setembro, apresentei um seminário na Universidade de Wisconsin/Madison sobre avaliação genética de populações cruzadas na presença de incerteza na paternidade, utilizando modelos hierárquicos de Bayes com variâncias estruturadas. Esta apresentação no Departamento de Ciência do Leite foi a convite do Professor Dr. Guilherme J. M. Rosa.

Considerações Finais

Considerando que bioinformática é uma área emergente no cenário internacional da pesquisa genética, de grande importância na pesquisa animal e priorizada pela Embrapa, o treinamento foi extremamente produtivo

para a Empresa e para seu pesquisador.

Embora tenha sido na mesma instituição na qual obtive o doutoramento, a capacitação foi em área complementar a minha formação anterior, focada em métodos quantitativos aplicados à avaliação genética de bovinos de corte, a partir de medidas fenotípicas e informações de pedigree. Foram adquiridos conhecimentos e experiência prática em técnicas modernas que possibilitam obter mapas genômicos e analisar simultaneamente a expressão de milhares de genes, através do desenvolvimento de metodologia para integrar a informação de genômica genética, tais como genótipos de marcadores moleculares, dados fenotípicos e perfis de expressão gênica em micro-arranjos, nos esquemas usados para melhorar geneticamente populações pecuárias. Soma-se a isso a oportunidade ímpar de tomar parte num projeto pioneiro que integra, pela primeira vez em espécies pecuárias, dados de perfis de expressão gênica e marcadores genéticos de uma população referência F2, para combinar o poder de recombinação com a análise funcional na identificação de genes controlando crescimento do músculo esquelético, deposição de gordura e qualidade da carne. Cabe ressaltar novamente que apesar de o projeto ser relacionado com genética suína, as ferramentas estatísticas utilizadas e/ou desenvolvidas nesta pesquisa (p. ex. análise de loci de característica quantitativa – QTL, mapeamento de desequilíbrio de ligação, análise de micro-arranjos, etc.) são pertinentes para análise de dados das outras espécies pecuárias.

Este treinamento, portanto, enriqueceu a amplitude de minha formação e me permitirá aumentar a abrangência e complexidade de minhas linhas e projetos de pesquisa em genética animal, desenvolvidos na Embrapa Pecuária Sul. Objetivamente, os conhecimentos adquiridos serão úteis em dois projetos do MP1 relacionados com genética animal na Embrapa: um já em andamento vinculado a rede Bife de Qualidade, tratando do melhoramento genético da qualidade da carne bovina na raça Nelore e em sistemas de cruzamentos, sob coordenação da Embrapa Pecuária Sudeste; e outra mais abrangente, envolvendo diversas espécies pecuárias, que está em fase de estruturação e apresentação da proposta final para formação da rede de genômica animal da Embrapa. Esta sob a coordenação da Embrapa Biotecnologia e Recursos Genéticos. Neste último projeto teremos a responsabilidade de coordenar o projeto componente de ferramentas e métodos quantitativos para análises de genômica animal, isto muito em função da capacitação ao qual se refere este relatório.

rio.

Essa capacitação gerou também a oportunidade de aumentar nossa cooperação internacional e já temos uma proposta de pesquisa em negociação com a própria Michigan State University e a Universidade de Barcelona.

Esperamos pelo menos três publicações decorrentes diretamente do trabalho de pós-doutorado e, portanto, a Embrapa fará parte desse desenvolvimento inovador em genética avançada, terá seu nome registrado nas publicações do projeto, e teve um de seus colaboradores capacitado para desenvolver e aplicar metodologia de estatística genômica para identificar genes relacionados a características de importância econômica em animais pecuários, o que certamente contribuirá para a consolidação da posição de vanguarda da Empresa em tecnologias agropecuárias.

Referências

CARDOSO, F. F.; ROSA, G. J. M.; TEMPELMAN, R. J. Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle. **Journal of Animal Science**, Savoy, v. 85, n. 4, p. 909-918, Apr. 2007.

JANNINK, J. L. Selective phenotyping to accurately map quantitative trait loci. **Crop Science**, Ames, v. 45, n. 3, p. 901-908, May/June 2005.

JANSEN, R. C.; NAP, J. P. Genetical genomics: the added value from segregation. **Trends in Genetics**, London, v. 17, n. 7, p. 388–391, July 2001.

JIN, C.; LAN, H.; ATTIE, A. D.; CHURCHILL, G. A.; BULUTUGLO, D.; E YANDELL, B. S. Selective phenotyping for increased efficiency in genetic mapping studies. **Genetics**, Baltimore, v.168, n. 4, p. 2285–2293, 2004.

PEREZ-ENCISO, M.; E MISZTAL, I. Qxpak: a versatile mixed model application for genetical genomics and QTL analyses. **Bioinformatics**, Oxford, v. 20, n. 16. p. 2792- 2798, Nov. 2004.

ROCKMAN, M. V.; E KRUGLYAK, L. Genetics of global gene expression. **Nature Reviews Genetics**, London, v. 7, n. 11, p. 862-872, 2006.

WANG, D.; E NETTLETON, D. Identifying Genes Associated with a Quantitative Trait or Quantitative Trait Locus via Selective Transcriptional Profiling. **Biometrics**, Washington, v. 62, n. 2, p. 504–514, June 2006.

Anexos

I. Procedimento automatizado para análises de varredura genômica em larga escala usando SAS e Qxpak.

```
* Procedure for automated QTL Scan using Qxpak
  by Fernando F. Cardoso 2006;
* Routine to run chromosome scans with qxpak;
* F. F. Cardoso 2006;
options nodate ps = 66 ls = 80 NOXWAIT XSYNC;
* Establish root directory to run the analysis (where qxpak files should
be);
* Growth traits;
%macro growth(trait, group, varnum, model,u0);
%do i = 1 %to 19;
%let root =
c:\Docume ~ 1\fernando\MyDocu ~ 1\MSUPRP\500PRPF2\Qxpak\Univar ~
2\run;
%let chr = SSC&i;
x cd &root;

data test;
/* Qxpak Parameter File */
%let file = msuprp.par; * parameter file name;
  group = "&group";
  file "&root\&file" lrecl = 100;
```

```

put '# MSU Pig Resourse Population F2 510 pigs analysis - Mixed
Model';
    put 'ML_OPTION';
    put 'y';
    *put 'MULTITRAIT_OPTION';
    *put 'y 3';
    put 'PRINT_RESIDUALS';
    put 'y';
    put 'DATAFILE';
    put 'msuprp_growth.dat';
    put 'OUTFILE';
    put 'msuprp.out';
    put 'MARKERFILE';
    put 'msuprp.mkr';
    put 'PEDIGREEFILE';
    put 'msuprp.ped';
    put 'MARKER_POSITIONS';
    put 'SSC1 0 21.1 49.7 77.5 108.9 179.7 235.1';
    put 'SSC2 0 41 53.7 64.8 93.5 103.2 110.1 142';
    put 'SSC3 0 22.7 68.7 86.4 101.3 124.3 132.4 141.2 159.6';
    put 'SSC4 0 29.2 54.1 61.8 73.8 88 131.3';
    put 'SSC5 0 11.5 53.5 81.5 108.9 127.4 147.6 159.7';
    put 'SSC6 0 22.4 50.6 81.3 98.1 103.9 135.7 164.8 181';
    put 'SSC7 0 28.1 48 92.5 135.8 151.6 164 186.7';
    put 'SSC8 0 22.9 54.9 95.1 110.7 124.3 165.4';
    put 'SSC9 0 13.2 43.1 63.7 72.5 97.1 127.5';
    put 'SSC10 0 20.2 44.3 56.8 79.4';
    put 'SSC11 0 53 65.2 119';
    put 'SSC12 0 31.2 47.2 61 94';
    put 'SSC13 0 13.7 41 54.3 65.2 85.9 103.3 122.7';
    put 'SSC14 0 26.6 45.9 64.7 85.2 95.7 117.7';
    put 'SSC15 0 24.7 54.8 64.5 82.2 96';
    put 'SSC16 0 30.9 64.8 99.5';
    put 'SSC17 0 22.7 43 94.9';
    put 'SSC18 0 43.1 54.8';
    put 'SSC19 0 14.8 37.3 47.7 56.1 89.2 146.6 169.6';
    put 'NUMBER_OF_MCMC_ITERATIONS';
    put '1000';
    put 'QTL';
    put 'qtl1 ' "&model &chr " ' all';

```

```
put 'EFFECT';
if group='gr0-6wk' then do;
    put 'sex cross 3';
    put 'litter cross 4 diagonal';
    put 'parity cross 5';
    put 'nba cov 6';
    put 'u0 cross 1 add_animal msuprp.ped';
    put 'TRAIT';
    put "&trait &varnum" ' sex litter parity nba' " &u0" '
qtl1';
end;
else do;
    put 'rep cross 2 diagonal';
    put 'sex cross 3';
    put 'pen cross 7 diagonal';
    put 'finisher cross 8';
    if group='fin10wk' then
        put 'age cov 11';
    else if group='fin13wk' then
        put 'age cov 12';
    else if group='fin16wk' then
        put 'age cov 13';
    else if group='fin19wk' then
        put 'age cov 14';
    else if group='fin22wk' then
        put 'age cov 15';
    put 'u0 cross 1 add_animal msuprp.ped';
    put 'TRAIT';
    put "&trait &varnum" ' rep sex pen finisher age' " &u0" '
' qtl1'; end;
    put 'TEST';
    put 'qtl1';
/* Creates a run file */
%let runfile = run; * parameter file name;
file "&root\&runfile" lrecl = 100;
put "&file";
run;
/* Run Qxpak */
data _null_;
```

```

command = "qxpak <run> out";
  call system(command);
run;
/* Create folder & copy output files */
x cd ..;
x md &chr; x cd &chr;
x md &trait; x cd &trait;
%if &u0 = u0 %then %let folder = fix_ad_u0;
%else %let folder = fix_ad;
x md &folder; x cd &folder;
x copy &root\msuprp*. *;
x copy &root\run;
x copy &root\out;
%end;
%mend growth;

%growth (LMA10WK ,fin10wk,21 ,fix_ad , );

```

II. Procedimento automatizado para análises do arquivo de saída de resultados de varredura genômica em larga escala usando Qxpak.

```

* Rotine to read Qxpak output file
  By Fernando F. Cardoso 2006;
options nodate ps = 66 ls = 80;
data varf;
run;
data solf;
run;
data resf;
run;
data qtlf;
run;
data qtlpeakf;
run;
data pfpf;
run;
%let dir = C:\Documents and Settings\fernando\My
Documents\MSUPRP\500PRPF2\Qxpak\Univar ~ 1\;
%macro read(subdir,analysis,trait,outfile,order,nran,ntrait);

```



```
%let file = &subdir.\&analysis.\&outfile;
/* Variance components */
data var;
  infile "&dir&file" trunccover;
  group = "&subdir  ";
  analysis = "&analysis  ";
  trait = "&trait  ";
  order = &order;
  nran = &nran;
  ntrait = &ntrait;
  input str $ 1-256;
  if str = 'RESIDUAL variances' then
    if ntrait = 1 then
      do;
        count + 1;
        veffect = 'resid  ';
        model = count-2*order + 1;
        input t1 $ / t1 $ var;
        if model = 0 | model = 1 then output;
        if nran > 0 then
          do i = 1 to nran;
            input / t1 $ t1 $ veffect $/ t1 $ / t1 $ var;
            if model = 0 | model = 1 then output;
          end;
        end;
      end;
    else if ntrait = 2 then
      do;
        count + 1;
        veffect = 'resid_cwt ';
        model = count-2*order + 1;
        input t1 $ / t1 $ var @;
        if model = 0 | model = 1 then output;
        veffect = 'resid_cov ';
        input var;
        if model = 0 | model = 1 then output;
        veffect = 'resid  ';
      end;
    input t1 $ t1 var;
    if model = 0 | model = 1 then output;
    if nran > 0 then
```

```

do i = 1 to nran;
  input / t1 $ t1 $ eff $ / t1 $ / t1 $ var @;
  veffect = Compress(ef| | '_cwt', '');
  if model = 0 | model = 1 then output;
  veffect = Compress(ef| | '_cov', '');
  input var;
  if model = 0 | model = 1 then output;
  veffect = eff;
  input t1 t1 $ var;
  if model = 0 | model = 1 then output;
end;
end;
keep group analysis trait model veffect var;
run;
data varf;
  set varf var;
  if var^ = .;
run;
/* QTL SCAN */
data qtl;
  infile "&dir&file" trunccover;
  group = "&subdir ";
  analysis = "&analysis ";
  trait = "&trait ";
  order = &order;
  input str $ 1-256;
  if str = 'Positions df 2logLRatio Pvalue Residual_Variances'
then
  do;
    count + 1;
    model = count - order + 1;
  if model = 1 then
  do i = 1 to 10000 until(loglratio = .);
    input pos1 pos2 pos3 df logLRatio Pvalue ResidVar;
    log10p = -log10(pvalue);
    raw_p = pvalue;
    if (loglratio > 40 & pvalue = .) then raw_p = 0; *overflow of
PROBCHI function;
    if pos2 <= 235.1 then do; ssc = 1; posssc = pos2;
end;

```

```
else if pos2 <= 377.1 then do;ssc = 2; posssc = pos2- 235.1; end;
else if pos2 <= 536.7 then do;ssc = 3; posssc = pos2- 377.1; end;
else if pos2 <= 668.0 then do;ssc = 4; posssc = pos2- 536.7; end;
else if pos2 <= 827.7 then do;ssc = 5; posssc = pos2- 668.0; end;
else if pos2 <= 1008.7 then do;ssc = 6; posssc = pos2- 827.7; end;
else if pos2 <= 1195.4 then do;ssc = 7; posssc = pos2-1008.7; end;
else if pos2 <= 1360.8 then do;ssc = 8; posssc = pos2-1195.4; end;
else if pos2 <= 1488.3 then do;ssc = 9; posssc = pos2-1360.8; end;
else if pos2 <= 1567.7 then do;ssc = 10; posssc = pos2-1488.3; end;
else if pos2 <= 1686.7 then do;ssc = 11; posssc = pos2-1567.7; end;
else if pos2 <= 1780.7 then do;ssc = 12; posssc = pos2-1686.7; end;
else if pos2 <= 1903.4 then do;ssc = 13; posssc = pos2-1780.7; end;
else if pos2 <= 2021.1 then do;ssc = 14; posssc = pos2-1903.4; end;
else if pos2 <= 2117.1 then do;ssc = 15; posssc = pos2-2021.1; end;
else if pos2 <= 2216.6 then do;ssc = 16; posssc = pos2-2117.1; end;
else if pos2 <= 2311.5 then do;ssc = 17; posssc = pos2-2216.6; end;
else if pos2 <= 2366.3 then do;ssc = 18; posssc = pos2-2311.5; end;
else if pos2 <= 2535.9 then delete; *ssc = 19; *posssc = pos2-2366.3;
*end;
    output;
    end;
end;
keep group analysis trait model pos1 pos2 pos3 df logLRatio Pvalue
ResidVar ssc posssc log1Op raw_p;
run;
```

* Proportion of False Positives (PFP), FDR, q values to determine QTL Significance over the n replicates;

```
proc iml;
    use qtl;
    read all var {raw_p} into pvalue;
    ntest = NROW(pvalue);
    nbin = max(round(ntest/25), 10); * Number of bins to split the pvalues
distribution;
    bwidth = 1/nbin;
    bcount = J(nbin, 1, 0);
    nbar = J(nbin, 1, 0);
    p = J(ntest, 6, 0); * p cols: 1-order, 2-pvalue, 3-pfp, 4-fdr, 5-q_pfp, 6-q_fdr;
    p[, 2] = pvalue;
    nfull = 0;
```

```

ntnull = 0;
nreject = 0;
alpha = 0.05;
do i = 1 to ntest;
  if (pvalue[i] <= alpha) then nreject = nreject + 1;
  x = 0;
  do j = 1 to nbin while(x = 0);
    if (pvalue[i] <= j/nbin) then
      do;
        bcount[j] = bcount[j] + 1;
        x = 1;
      end;
  end;
end;
end;
end;
x = 0;
do j = 1 to nbin while(x = 0);
  nbar[j] = SUM(bcount[j:nbin])/(nbin-j + 1);
  if (bcount[j] <= nbar[j]) then
    do;
      ntnull = nbin * nbar[j];
      nfnnull = ntest - ntnull;
      ptnull = ntnull / ntest;
      x = 1;
    end;
  end;
end;
pfp_alpha = alpha * ntnull / nreject;
pthr_alpha = alpha * nreject / ntnull;
thr_alpha = CINV(1 - pthr_alpha, 1);
pthr_01 = .01 * nreject / ntnull;
thr_01 = CINV(1 - pthr_01, 1);
fdr_alpha = ntest * alpha / nreject;
power_alpha = (nreject - alpha * ntnull) / nfnnull;
do i = 1 to ntest;
  p[i, 1] = i;
  p[i, 3] = min(pvalue[i] * ntnull / nreject, 1);
  p[i, 4] = min(pvalue[i] * ntest / nreject, 1);
end;
CALL SORT(p, {2});
p[ntest, 5] = p[ntest, 2] * ntnull / ntest;
p[ntest, 6] = p[ntest, 2];

```

```
do i = ntest-1 to 1 by -1;
  p[i,5] = min(p[i,2]*ntnull/i,p[i+1,5]);
  p[i,6] = min(p[i,2]*ntest/i,p[i+1,6]);
end;
CALL SORT(p,{1});
q_pfp = p[,5];
create qtlp from q_pfp [colname = 'q_pfp'];
append from q_pfp;
* print alpha nfnnull ntnull ntest nreject pfp_alpha fdr_alpha power_alpha
thr_alpha thr_01 pthr_alpha pthr_01;
create pfp var {alpha ptnull ntnull ntest nreject pfp_alpha fdr_alpha
power_alpha pthr_alpha pthr_01};
append;
quit;
data qtl;
merge qtl qtlp;
run;
data qtlf;
set qtlf qtl;
if logratio^ = .;
run;
data pfp;
set pfp;
group = "&subdir ";
analysis = "&analysis ";
trait = "&trait ";
run;
data pfpf;
set pfpf pfp;
if alpha^ = .;
run;

* QTL peaks, range and CI peaks, dominant peak;
%do iter = 1 %to 18;
data qtl;
set qtl;
if ssc = &iter;
run;
data qtlpeak;
run;
```

* QTL Local maxima;

```

proc iml;
  use qtl;
  read all var {logLRatio}into loglr;
  read all var {posssc}into pos;
  read all var {q_pfp}into qvalue;
  read all var {df}into df;
  npos = NROW(pos);
  peak = J(npos,1,0);
  domin = J(npos,1,0);
  lowrange = J(npos,1,0);
  upprange = J(npos,1,0);
  lowci = J(npos,1,0);
  uppci = J(npos,1,0);
  ssc = &iter;
  alpha = 0.10;
  *Check every position for local maximum;
  *First position;
  if (loglr[1] > loglr[2]) & (qvalue[1] < alpha) then
    do;
      peak[1] = 1;
      lowrange[1] = pos[1];
      t = loc(qvalue > alpha) [> <];
      if t > 0 then
        upprange[1] = pos[t];
    else
      upprange[1] = pos[npos];
      lowci[1] = pos[1];
      t = loc(loglr < (loglr[1] - CINV(.95,df[1])));
      if NCOL(t) > 0 then
        uppci[1] = pos[t [> <]];
      else
        uppci[1] = pos[npos];
      j = loc(pos = min(uppci[1],upprange[1]));
      t = max(loglr[1:j]);
      if loglr[1] >= t then
        domin[1] = 1;
    end;
  *print (pos[i]) (loglr[i]) t j k;
  do i = 2 to npos-1;

```

```
if (loglr[i] > loglr[i-1]) & (loglr[i] > loglr[i + 1]) & (qvalue[i] < alpha) then
do;
  peak[i] = 1;
  t = loc(qvalue[1:i] > alpha);
*print pos qvalue i t (pos[t]);
  if NCOL(t) > 0 then
    lowrange[i] = pos[1 + t[ < > ]];
  else
    lowrange[i] = pos[1];
    t = loc(qvalue[i:npos] > alpha);
*print pos qvalue i t (pos[t]);
  if NCOL(t) > 0 then
    upprange[i] = pos[i-2 + t[ > < ]];
  else
    upprange[i] = pos[npos];
    t = loc(loglr[1:i] < (loglr[i] - CINV(.95,df[i])));
*print 'lowci' pos loglr i t (pos[t]) (CINV(.95,df[i]));
  if NCOL(t) > 0 then
    lowci[i] = pos[1 + t[ < > ]];
  else
    lowci[i] = pos[1];
    t = loc(loglr[i:npos] < (loglr[i] - CINV(.95,df[i])));
  if NCOL(t) > 0 then
uppci[i] = pos[i-2 + t[ > < ]];
  else
    uppci[i] = pos[npos];
    j = loc(pos = max(lowci[i],lowrange[i]));
    k = loc(pos = min(uppci[i],upprange[i]));
    t = max(loglr[j:k]);
*print (pos[i]) (loglr[i]) t j k;
  if loglr[i] >= t then
    domin[i] = 1;
  end;
end;
*Last position;
if (loglr[npos] > loglr[npos-1]) & (qvalue[npos] < alpha) then
do;
  peak[npos] = 1;
  upprange[npos] = pos[npos];
  t = loc(qvalue > alpha);
```

```

if NCOL(t) > 0 then
  lowrange[npos] = pos[t[ < > ]];
  else
    lowrange[npos] = pos[1];
  uppci[npos] = pos[npos];
t = loc(loglr < (loglr[npos] - CINV(.95,df[npos])));
if NCOL(t) > 0 then
  lowci[npos] = pos[t[ < > ]];
  else
    lowci[npos] = pos[1];
  j = loc(pos = max(lowci[npos],lowrange[npos]));
t = max(loglr[j:npos]);
if loglr[npos] > = t then
  domin[npos] = 1;
  end;

* print analysis iter pos loglr qvalue lowrange upprange lowci uppci
domin;
nmax = NCOL(loc(peak = 1));
if nmax > 0 then
  do;
    ssc = J(nmax,1,ssc);
    pos = pos[loc(peak = 1)];
    loglr = loglr[loc(peak = 1)];
    qvalue = qvalue[loc(peak = 1)];
    lowrange = lowrange[loc(peak = 1)];
    upprange = upprange[loc(peak = 1)];
    lowci = lowci[loc(peak = 1)];
    uppci = uppci[loc(peak = 1)];
    domin = domin[loc(peak = 1)];
  create qtlpeak var {ssc pos loglr qvalue lowrange upprange lowci
uppci domin};
  append;
  end;

* print analysis iter pos loglr qvalue lowrange upprange lowci uppci
domin;
quit;
data qtlpeak;
set qtlpeak;

```



```
group = "&subdir  ";
analysis = "&analysis  ";
trait = "&trait  ";
run;

data qtlpeakf;
  set qtlpeakf qtlpeak;
  if loglr^ = .;
run;
%end;

*proc printto;
*run;

/* Solutions */
data sol;
  infile "&dir&file" trunccover;
  group = "&subdir  ";
  analysis = "&analysis  ";
  trait = "&trait  ";
  order = &order;
  input str $ 1-256;
  if str = 'trait/effect level  solution  std. error' then
    do;
      count + 1;
model = count-2*order + 1;
      if model = 1 then
        do i = 1 to 10000 until(solution = .);
          input t1 effect level solution stderr;
            output;
        end;
      end;
      keep group analysis trait model effect level solution stderr;
run;
data solf;
  set solf sol;
  if solution^ = .;
run;
/* Residuals */
data res;
```

```

infile "&dir&file" trunccover;
group = "&subdir ";
analysis = "&analysis ";
trait = "&trait ";
order = &order;
input str $ 1-256;
if str = 'RECORD Y RESIDUE G-HAT Y-Xb' then
do;
    count + 1;
    model = count-2*order + 1;
input t1;
if model = 1 then
do i = 1 to 10000 until(resid = .);
    input record y resid ghat y_xb;
    output;
end;
end;
keep group analysis trait model record y resid ghat y_xb;
run;
data resf;
set resf res;
if resid^ = .;
run;
%mend read;
%read (fin22wk,fix_ad_u0,BF1022wk,msuprp.out,2,2,1);

```

Embrapa

Pecuária Sul

**MINISTÉRIO DA AGRICULTURA,
PECUÁRIA E ABASTECIMENTO**

