

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Modelos estatísticos para mapeamento de QTL associados a dados
de contagem**

Karen Pallotta Tunin Kamogawa

**Tese apresentada para obtenção do título de
Doutora em Agronomia. Área de concentração:
Ciência Animal e Pastagens**

**Piracicaba
2009**

Karen Pallotta Tunin Kamogawa
Médico Veterinário

Modelos estatísticos para mapeamento de QTL associados a dados de contagem

Orientador:
Prof. Dr. IRINEU UMBERTO PACKER

**Tese apresentada para obtenção do título de
Doutora em Agronomia. Área de concentração:
Ciência Animal e Pastagens**

**Piracicaba
2009**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Kamogawa, Karen Pallotta Tunin

Modelos estatísticos para mapeamento de QTL associados a dados de contagem /
Karen Pallotta Tunin Kamogawa. - - Piracicaba, 2009.
72 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2009.
Bibliografia.

1. Bovinos 2. Carrapatos 3. Dados de contagem 4. Mapeamento genético 5. Resistência
genética animal I. Título

CDD 636.2089696
K32m

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

*Ao Luiz Fernando, pelo amor e compreensão sempre
presentes em todas as fases deste trabalho;*

*Aos meus pais, por me mostrar o caminho do bem
e ajudar a superar os obstáculos da vida
DEDICO.*

AGRADECIMENTOS

A Deus, pelo dom da vida e a capacidade de amar.

Ao Professor Irineu Umberto Packer, por ser mais do que um orientador, por me ensinar a ter fé, ser confiante, perseverante e humilde. Mais do que a tese, o senhor me ensinou a viver! E por isso lhe serei sempre grata.

Ao Dr. Mário Luiz Martinez, pelo incentivo e apoio sempre demonstrados nos poucos anos em que trabalhamos juntos. Sempre me lembrarei de você.

Ao meu amigo Dr. Luís Fernando Batista Pinto, pelo companheirismo durante esses anos e sua ajuda indispensável para a confecção desta tese.

Ao Dr. Guilherme Rosa (UW-Madison), pelo período em que passamos juntos e que foi essencial para meu processo de doutoramento. Além de um grande mestre, uma grande pessoa.

A Marigel Perez-Cabal, pela amizade inesperada, por me ensinar espanhol e tornar minha estadia longe de casa um pouco menos pesada.

Ao Dr. Fabyano Fonseca e Silva, por sua inestimável ajuda e paciência ao me ajudar.

Ao Dr. Daniel Gianola e sua esposa Graciela pela amizade e ajuda nos momentos difíceis.

A Filó, por me mostrar que uma mulher pode ser mais forte que uma rocha e ainda assim, mais doce que o mel.

A todos meus amigos da ESALQ e do “famoso” 4th floor (UW-Madison), pela amizade e convivência indispensáveis ao nosso dia-a-dia.

As funcionárias, Giovana e Eleonora, sua ajuda, amizade e prestatividade tornaram esta jornada um pouco menos complicada.

A todos os demais funcionários e docentes do Departamento de Zootecnia e Departamento de Zootecnia de não-ruminantes pelos dias de convivência e amizade.

Ao CNPq, pela bolsa de estudos que recebi aqui e para realização do doutorado sanduíche, sem esse apoio este trabalho não teria sido possível.

Aos meus cães, por sempre me mostrar um amor incondicional, mesmo quando não havia tempo para brincadeiras e afagos.

SUMÁRIO

RESUMO.....	6
ABSTRACT.....	7
LISTA DE FIGURAS.....	8
LISTA DE TABELAS.....	10
1 INTRODUÇÃO.....	11
2 DESENVOLVIMENTO.....	13
2.1 Importância do carrapato <i>Boophilus microplus</i> na agropecuária.....	14
2.2 Alternativa ao controle de carrapatos: resistência genética do hospedeiro.....	15
2.3 Introdução ao mapeamento de QTL.....	16
2.4 Métodos utilizados na detecção de QTL em populações não endogâmicas.....	18
2.5 Características quantitativas não-normalmente distribuídas: dados de contagem.....	21
2.6 Material e Métodos.....	22
2.6.1 População experimental.....	22
2.6.2 Coleta dos dados fenotípicos.....	23
2.6.3 Coleta dos dados genotípicos e mapa de ligação.....	24
2.6.4 Estudo de simulação.....	24
2.6.5 Metodologias aplicadas para o mapeamento de QTL.....	25
2.6.6 Comparação dos modelos.....	29
2.6.7 Análise dos dados reais.....	29
2.7 Resultados e Discussão.....	30
2.7.1 Comparação dos modelos.....	30
2.7.2 Análise dos dados reais.....	39
3 CONCLUSÕES.....	54
REFERÊNCIAS.....	55
APÊNDICES.....	60

RESUMO

Modelos estatísticos para mapeamento de QTL associados a dados de contagem

Este estudo teve por objetivo analisar e comparar metodologias estatísticas para fins de mapeamento de QTL associados à resistência a ectoparasitas em bovinos. Os animais, submetidos à infestação artificial, foram periodicamente avaliados por contagens, como número de carrapatos. Estes dados se caracterizam como medidas repetidas e, via de regra, não atendem ou atendem parcialmente as exigências usuais da análise, para mapeamento de QTL, dentre elas a de apresentar distribuição normal e independência dos erros. Ainda não está bem definido qual seria a melhor estratégia para analisar dados com o perfil descrito. Algumas alternativas seriam transformações de dados que permitam o uso dos programas já disponíveis, ou o desenvolvimento de programas que utilizem outras distribuições como Poisson ou Poisson inflada de zeros (ZIP). Esta proposta está inserida na parceria entre EMBRAPA - Gado de Leite e a ESALQ/USP, para desenvolvimento do projeto de mapeamento de QTL em bovinos mestiços (Gir x Holandês), para várias características incluindo a resistência a parasitas. Foram utilizados 263 animais F2, genotipados com 5 marcadores moleculares no cromossomo 23, na tentativa de mapear QTL para característica de resistência a carrapatos. Dados coletados naquela população F2 e dados simulados em diferentes cenários, serão a base para a comparação de estratégias de análise e mapeamento de QTL. Os modelos de mapeamento clássico, assim como a utilização de transformações dos dados originais foram comparados a modelos de regressão Poisson e modelo ZIP. Os modelos Poisson e ZIP apresentaram os melhores resultados quando trabalhamos com dados de contagem inflacionados de zeros, porém em outros cenários a transformação dos dados originais se mostrou igualmente eficiente. Dependendo do propósito do mapeamento (seja ele localizar ou estimar o efeito), cada modelo possui suas vantagens e suas limitações. Assim, sempre é recomendável uma prévia análise descritiva dos dados para que o melhor modelo seja utilizado.

Palavras-chave: QTL; Seleção; Dados de contagem; Modelo ZIP; Bovinos; Carrapatos

ABSTRACT

Statistical models for QTL mapping associated to counting data

This study has as main objective to analyze and compare statistical approaches to QTL mapping for parasites resistance in bovines. The animals, under artificial infestation, were periodically evaluated by counting, as ticks count. These data are characterized as repeated measures and, usually, don't follow or partially follow the usual requirements for the analysis, for QTL mapping, that is to present normal distribution and error independence. It is not clear yet which will be the best strategy to analyze this kind of data. Some alternatives could be data transformation that allows the use of software available on the web, or the development of specific programs that use other types of distribution like Poisson or Zero Inflated Poisson (ZIP). This work is an association between EMBRAPA – Gado de Leite and ESALQ/USP, to the development of the QTL mapping project for crossbred bovines (Gyr x Holstein), for different characteristics including the parasite resistance. Were used 263 animals F2, genotyped for 5 molecular markers on the chromosome 23, aiming to map QTL for characteristics of parasite resistance. Data collected on this F2 population and simulated data in different scenarios will be the base for the strategies of the QTL mapping approaches comparison. The classical mapping models and the use of data transformation of the original data were compared to Poisson regression and ZIP models. The Poisson and ZIP models presented the best results when working with zero inflated count data however in some other scenarios the data transformation showed similar efficiency. Depending on the purpose of the mapping (this meaning locate or estimate the QTL effect) each model has its vantages and its limitations. This way, it is always advisable to make a previous descriptive analysis of the data to better choose the model.

Keywords: QTL; Selection; Count Data; ZIP model; Bovines; Ticks

LISTA DE FIGURAS

Figura 1	- Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano sem uso de transformações.....	39
Figura 2	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano sem transformação dos dados.....	39
Figura 3	- Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano com uso da transformação de Box-Cox.....	40
Figura 4	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação de Box-Cox dos dados.....	40
Figura 5	- Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados.....	41
Figura 6	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados.....	42
Figura 7	- Mapeamento de QTL, para a primeira contagem, através do modelo Poisson....	43
Figura 8	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Poisson.....	43
Figura 9	- Mapeamento de QTL, para a primeira contagem, através do modelo ZIP.....	44
Figura 10	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo ZIP.....	44
Figura 11	- Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano sem uso de transformações.....	46
Figura 12	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano sem transformação dos dados.....	46
Figura 13	- Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano com uso da transformação de Box-Cox.....	47
Figura 14	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação de Box-Cox dos dados.....	47
Figura 15	- Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados.....	48
Figura 16	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados.....	49

Figura 17	- Mapeamento de QTL, para a segunda contagem, através do modelo Poisson.....	50
Figura 18	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Poisson.....	50
Figura 19	- Mapeamento de QTL, para a segunda contagem, através do modelo ZIP.....	51
Figura 20	- Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo ZIP.....	51

LISTA DE TABELAS

Tabela 1	- Mapa consenso dos marcadores utilizados nas análises.....	24
Tabela 2	- Taxa de falsos negativos para cada modelo nos diferentes cenários simulados.....	30
Tabela 3	- Taxa de falsos positivos para cada modelo nas diferentes inflações de zeros simuladas.....	31
Tabela 4	- Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0$).....	33
Tabela 5	- Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0,2$).....	35
Tabela 6	- Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0,5$).....	37
Tabela 7	- Valores das estimativas de posição e efeitos aditivos obtidos para o mapeamento de QTL dos dados reais (primeira contagem) com os diferentes modelos.....	45
Tabela 8	- Valores das estimativas de posição e efeitos obtidos para o mapeamento de QTL dos dados reais (primeira contagem) com os diferentes modelos.....	52

1 INTRODUÇÃO

O melhoramento genético clássico de bovinos, baseado nos esquemas de seleção e acasalamento, tem apresentado bons resultados para as características de moderada e alta herdabilidade. No entanto, outros caracteres de importância econômica (como resistência a doenças, por exemplo) não são contemplados nestes programas por apresentarem baixa herdabilidade. A seleção assistida por marcadores se apresenta como uma possibilidade bastante atraente, pois permitirá identificar precocemente os animais superiores para tais características e selecioná-los para obtenção de um rebanho que as expresse (Khatkar et al, 2004).

As razões pelas quais os produtores podem optar pelo melhoramento genético para resistência a doenças (Gibson e Bishop, 2005) são: (1) aumentar a competitividade através da diminuição dos custos e aumento da produtividade. O percentual de gasto com prevenção de doenças é de aproximadamente 17% do custo total de produção nos países desenvolvidos e entre 35 e 50% nos países em desenvolvimento; (2) reduzir o desenvolvimento de resistência as drogas usadas no controle e tratamento de infestações como antielmínticos, antibióticos, anti-protozoários, acaricidas e vacinas e (3) atender as pressões do mercado consumidor para a diminuição dos resíduos de drogas utilizadas para controle e prevenção de doenças e infestações.

O avanço da biologia molecular e o seqüenciamento do genoma têm permitido estudar, com mais precisão, os efeitos dos genes envolvidos no controle de características quantitativas. A partir desse desenvolvimento, vários estudos demonstraram a possibilidade de mapeamento de QTL (*Quantitative Trait Loci*), uma região do genoma onde podem estar inseridos um ou mais genes associados a características quantitativas, ou de uma maneira menos ambígua, um locus subjacente a uma característica quantitativa (Lynch & Walsh, 1998). Porém um novo problema foi contemplado, uma vez que muitas características de interesse são provenientes de dados de contagem (como contagem de carrapatos e ovos por grama de fezes). Tais dados não possuem distribuição normal, o que é uma pressuposição dos programas de mapeamento de QTL atualmente disponíveis. Assim, é necessário o estudo de métodos

de transformação de dados, o que permitiria a utilização dos programas disponíveis, ou o desenvolvimento de uma abordagem alternativa. Outro problema encontrado é que alguns desses dados de contagem podem apresentar dispersão, quando comparados com outras distribuições conhecidas, como é o caso de dados Poisson com inflação de zeros (ZIP) e outros casos de superdispersão.

No Brasil o projeto de mapeamento de QTL em bovinos é uma parceria entre a EMBRAPA – Gado de Leite, a Universidade de São Paulo e a Universidade Federal de Viçosa. (Martinez et al, 2002) que teve seu início em 1995. Uma população F2 foi formada a partir do cruzamento de fêmeas da raça Gir e machos da raça Holandesa, com o objetivo de identificar QTLs para características de resistência a ecto e endoparasitas e resistência ao calor. Tunin (2005) estudou o cromossomo 23 para peso ao nascimento, peso à desmama e resistência a carrapatos, e encontrou um QTL sugestivo para resistência a carrapatos. Outros autores também estudaram esta população para características de crescimento e resistência a carrapatos em outros cromossomos (5, 7 e 14) (Gasparin et al, 2005; Miyata et al, 2007; Gasparin et al, 2007)

O presente trabalho teve por objetivo estudar modelos estatísticos para mapeamento de QTL para características de contagem inflacionadas de zero. Para isso foram utilizados dados simulados em diferentes cenários para comparação dos diversos modelos. Posteriormente esses modelos foram aplicados em dados reais para demonstrar sua aplicabilidade.

2 DESENVOLVIMENTO

2.1 Importância do carrapato *Boophilus microplus* na agropecuária

O carrapato *Rhipicephalus (Boophilus) microplus* foi descrito pela primeira vez em 1887 por Canestrini e até os dias de hoje figura como um dos ectoparasitas de maior importância na pecuária, principalmente em países de regiões tropicais e subtropicais (Evans et al, 2000). Inicialmente era tido como uma praga com algumas restrições climáticas, porém segundo esse autor, o *Boophilus microplus* pode ser considerado um “carrapato pantropical”. Tal infiltração em condições climáticas aparentemente desfavoráveis, nos leva à conclusão de que este carrapato possui capacidade de adaptação excepcional (Hoogstral, 1973). O ciclo biológico do carrapato *Boophilus microplus* apresenta uma fase parasitária de aproximadamente 21 dias na qual passa pelos estágios de larva, ninfa e adulto, todos em um único hospedeiro.

A fase de vida livre se inicia com a queda das fêmeas ingurgitadas e termina quando as larvas eclodidas encontram um novo hospedeiro. Ainda nessa fase, a fêmea ingurgitada apresenta, inicialmente, um período de pré-postura de 3 dias e morre após a postura. Em temperaturas ao redor de 28°C e umidade relativa de 85%, a postura e a eclosão ocorrem em aproximadamente 18 dias. As larvas recém eclodidas migram para as pontas da vegetação onde localizam o hospedeiro através do odor ou vibrações produzidas pelo mesmo (Olsen, 1974). No hospedeiro, as larvas se fixam em regiões corporais propícias para seu desenvolvimento, tais como posterior da coxa, região perineal e perivulvar. Após sete dias de sua fixação, ocorre a muda de larvas para ninfas e estas se desenvolvem até se tornarem adultas, com marcado dimorfismo sexual em aproximadamente oito dias. A fêmea, após o acasalamento, começa a alimentar-se do sangue do hospedeiro até o ingurgitamento total, que propicia sua queda ao solo, enquanto o macho permanece no bovino à procura de novas fêmeas (Olsen, 1974).

Infestações por *R. (B.) microplus* possuem sérias implicações médicas e veterinárias uma vez que este parasita é o vetor de muitas doenças (cite algumas doenças aqui) e é o responsável por diversos problemas econômicos para a pecuária

leiteira (Peconick et al, 2008). Com outros artrópodes hematófagos vetores de doenças, os carrapatos também representam um sério problema de saúde pública (Kashino et al, 2005). A Organização das Nações Unidas para Agricultura e Alimentação (FAO) estima que as perdas mundiais devido a infestações por esse carrapato sejam em torno de sete bilhões de dólares por ano.

No Brasil, que possui o maior rebanho comercial de bovinos do mundo constituindo cerca de 200 milhões de animais (IBGE, 2008), a magnitude deste problema pode ser maior do que dois bilhões de dólares por ano (Peconick et al, 2008). Honer e Gomes (1990) estimaram que uma vaca infectada com carrapatos e vermes pode perder em torno de 18 a 47 kg de peso vivo por ano. Na Austrália, Frisch et al. (2000) estimaram que um animal com uma média de 40 carrapatos/dia pode perder peso equivalente a 20 kg/ano. Também foi reportado que animais infestados com mais do que 200 carrapatos por um período de seis semanas pode morrer se não for tratado (Martinez et al, 2006). A infestação causa a supressão do apetite, o que leva a perda de peso vivo e baixas taxas de crescimento. Carrapatos são também vetores dos parasitas *Babesia bovis*, *B. Bigemina* e *Anaplasma marginale* causadores do quadro conhecido no Brasil como “Tristeza parasitária” (Piper et al, 2008).

Os problemas causados pela infestação de carrapatos variam de acordo com a idade do animal, o nível de infestação e se o parasita em questão está atuando como vetor para outras enfermidades. Um bovino, que seja portador de uma alta carga parasitária, pode apresentar uma anemia, variando de leve a moderada, e morte. Os carrapatos acarretam um aumento na taxa de morbidade e mortalidade durante o período de seca, quando as infestações geralmente são maiores, além de retardar os resultados de terminação de animais que se destinarão à venda (Radostits et al, 2000).

O método mais amplamente utilizado para controle do número de carrapatos num rebanho bovino é a aplicação de acaricidas; no entanto é comum observar carrapatos resistentes a um ou mais destes fármacos. São necessários cerca de 10 anos para o aparecimento da resistência a cada classe de fármaco e é muito provável que os novos acaricidas desenvolvidos também encontrem este mesmo problema (Frisch, 1999).

O uso de acaricidas não é um método economicamente viável de prevenção e controle de carrapatos (Pegran et al, 1991). Assim, o desenvolvimento de novas opções de prevenção deste ectoparasita se faz necessária e a procura por bovinos que sejam resistentes é uma alternativa nessa busca.

2.2 Alternativa ao controle de carrapatos: resistência genética do hospedeiro

O uso de químicos agrícolas e seu impacto na segurança alimentar e no meio ambiente em geral nunca estiveram sujeito a tanta atenção pública. A necessidade do aumento da produtividade das empresas agrícolas, não mais para alimentar um mundo faminto e sim para sobreviver aos efeitos da competição, nunca foram tão grandes. A habilidade de redução de custos, aumento da flexibilidade de mercado e fornecer produtos para consumidores cada vez mais exigentes, são componentes essenciais para aumento da produtividade (Frisch et al, 2000).

A redução das preocupações públicas, dos custos com o controle de parasitas e do desenvolvimento de cepas de carrapatos resistentes aos acaricidas, pode ser atingida através do uso de animais geneticamente resistentes aos parasitas (Frisch et al, 2000).

As conseqüências, como doenças, quebras produtivas, altos gastos de produção e até morte, do conjunto parasita-hospedeiro são muito similares entre as espécies. Em bovinos, no entanto, o nível de infestação por carrapatos parece variar de acordo com a raça. Até mesmo depois de repetidas infestações, hospedeiros susceptíveis ainda abrigam um número significativamente maior de parasitas do que raças bovinas resistentes. Portanto, existe no hospedeiro uma variabilidade genética herdável, porém os genes ainda não foram totalmente identificados nem existe conhecimento sobre a imunologia (Carvalho et al, 2008).

A resistência do hospedeiro tem sido amplamente discutida como um método viável de controle de carrapatos. Em bovinos, a resistência genética é bastante desejável, pois é sustentável e ambientalmente mais adequada, assim como reduz os gastos com fármacos e práticas de manejo (Piper et al, 2008). Além de todos os fatores genéticos do hospedeiro e sua resposta imune adaptativa, alguns pesquisadores

acreditam que a resistência pode ser mediada por fatores imunes não adaptativos. Sanidade do rebanho, cor e espessura do couro, área de pele disponível a infestação ou comprimento de pêlo são alguns exemplos de fatores não adaptativos (Kashino et al, 2005). Se a resistência é adquirida então o bovino é susceptível até que tenha experimentado uma infestação, já se a resistência é fisiológica não há a necessidade do contato com o parasita para que a resistência possa manifestar-se (Roberts, 1968).

Já está bastante documentado que raças de *Bos indicus* são naturalmente mais resistentes ao *B. Microplus* do que raças de *Bos taurus*, apesar de existir diferenças consideráveis dentro da subespécie e dentro de raça (Utech et al, 1978). Muitos estudos demonstraram que essa resistência é herdável e que sua presença nos rebanhos *Bos taurus* e *Bos indicus* pode ser melhorada por seleção (Frisch, 1999). Apesar de todo conhecimento a respeito dessa herança, a seleção de rebanhos para essa característica ainda não foi amplamente aplicada à campo, com apenas um estudo apresentado até os dias de hoje (Piper et al, 2008).

Identificar os mecanismos pelo quais a resistência de bovinos previne altas infestações de carrapatos é um passo essencial para o desenvolvimento de marcadores associados à resistência (Piper et al, 2008). Muitos estudos estão sendo desenvolvidos nessa área na tentativa de se encontrar marcadores moleculares que possam ser utilizados em programas de seleção assistida por marcadores (MAS), porém ainda não há marcadores moleculares para uso comercial.

2.3 Introdução ao mapeamento de QTL

Em certas situações, a detecção de regiões do genoma associadas a uma característica de herança quantitativa demanda um número relativamente grande de animais com o fenótipo e genótipo determinados. Por exemplo, dois QTL muito próximos e com efeitos opostos tendem a se eliminar, entretanto, se os efeitos são na mesma direção eles tendem a se apresentar como um único QTL de grande efeito (Barton & Keightley, 2002). Sendo assim, a determinação do número de animais a ser utilizado depende da magnitude do efeito do QTL que se deseja mapear e do

delineamento experimental que se propõe, dentre outros fatores, o que pode levar a perdas de QTL verdadeiros quando se têm amostras pequenas.

Segundo Georges e Andersson (2003), o gado de leite possui vantagens em relação aos outros animais de produção no mapeamento de QTL, devido ao fato de que na maioria das fazendas há intensa anotação de dados referentes a avaliações fenotípicas tais como medidas de qualidade de leite, saúde do animal, assim como grande controle em relação ao pedigree e manejo, principalmente em países de maior desenvolvimento e com maior tecnologia de controle. Estes autores citam que o mapeamento de QTL é beneficiado pelo reduzido tamanho efetivo das populações, que leva à redução na complexidade das características em termos de heterogeneidade de alelos, aliada à maior extensão dos blocos de desequilíbrio de ligação, quando comparado com outras espécies.

A detecção de QTL requer três estágios essenciais: (1) coleta acurada de dados fenotípicos dentro da população experimental; (2) identificação acurada dos genótipos (através de marcadores moleculares, p.ex.) da população; e (3) análise estatística correlacionando os dados fenotípicos com os dados de genótipos. Quanto as análises estatísticas, foram e ainda são abordadas em diferentes metodologias propostas para o mapeamento de QTL. No início, os marcadores eram estudados individualmente (Soller e Genizi, 1978; Weller, 1986). Verificava-se apenas a associação entre cada marcador e a características de interesse, aproximação conhecida como mapeamento por marcas simples. O princípio desta análise é testar a hipótese nula de que não há diferença entre as médias fenotípicas entre diferentes classes genotípicas. O modelo que se pode aplicar a este método é uma regressão simples do fenótipo sobre o genótipo e verificar se o parâmetro da regressão é significativo. O método da máxima verossimilhança pode ser utilizado para estimar os parâmetros e possui o diferencial de considerar a densidade das classes genotípicas estudadas (Weller, 1986).

Métodos mais robustos devem ser utilizados quando o objetivo é identificar e mapear a posição do QTL no grupo de ligação. Assim outras metodologias foram propostas com este fim. O mapeamento por intervalo simples (Lander & Botstein, 1989) introduziu a idéia de avaliar cada posição num intervalo delimitado por dois marcadores.

Este procedimento é mais poderoso que a análise de marcas simples, pois considera informações contidas no mapa de ligação. Porém é limitado pelo fato de não ajustar os efeitos de QTLs presentes fora do intervalo sob análise.

Para a identificação de QTL podem ser utilizados diversos delineamentos experimentais. Por se tratar de animais, são utilizados delineamentos de acasalamentos entre animais não aparentados, visando gerar variabilidade (heterozigosidade), já que a endogamia aumenta a homozigosidade, indesejável do ponto de vista do mapeamento de QTL. Outro problema é que a endogamia pode trazer como consequência o aparecimento de genes deletérios em homozigose, sendo que alguns deles podem ser letais. Entre os delineamentos utilizados em animais podemos citar os delineamentos de meio irmão, de netas e F_2 .

Na presente tese foi utilizado o delineamento experimental F_2 , o qual promove o desequilíbrio de ligação necessário para a identificação de QTL quando são cruzadas duas subespécies, como, por exemplo, *Bos taurus* e *Bos indicus*. Este delineamento é o que fornece melhor poder de detecção quando o número de animais não é expressivo. Por este delineamento também é possível separar os efeitos aditivos e de dominância dos QTL encontrados.

Segundo Soller (1990), QTL só serão detectados quando as duas populações diferirem na frequência dos alelos dos QTL, o que ocorre quando ambas as populações estão fixadas para os alelos de interesse. O delineamento de F_2 é mais usado quando o objetivo é cruzar duas populações divergentes a fim de gerar variabilidade, ou quando o objetivo é introduzir um gene de interesse, presente em uma das populações (como por exemplo, resistência a parasita presente em *Bos taurus*).

2.4 Métodos utilizados na detecção de QTL em populações não endogâmicas

Como comentando anteriormente, a geração de linhagens endogâmicas para construção de populações F_2 está fora de questão em bovinos. Populações não endogâmicas são de grande interesse além das razões meramente experimentais. Por exemplo, QTL detectado através do cruzamento entre linhas endogâmicas usualmente representa diferenças fixas *entre* linhagens e a relevância dos QTL *intra* populações

permanece não esclarecido. Esta distinção de que os cruzamentos entre linhagens endogâmicas detectam QTL responsáveis por diferenças entre - populações, enquanto que as populações não endogâmicas detectam QTL responsáveis por variações intra-populações, torna esses dois métodos complementares e não competitivos. É bom incluir referências. Evite deixar parágrafos sem citações no desenvolvimento.

Populações não endogâmicas possuem várias desvantagens na detecção de QTL em relação a populações endogâmicas (como é o caso de plantas e insetos). Em alguns casos, QTL de grande efeito sob características de interesse em programas de seleção podem apresentar frequências extremas ou até fixas uma vez que a seleção para essas características já são realizadas há algum tempo. Se dois QTL possuem somente dois alelos e se assumirmos cruzamentos ao acaso, a heteroziguidade será no máximo 0,5. Isto é, mesmo na melhor das hipóteses, na média, 50% dos pais não produzem gametas recombinantes neste locus (Knott & Haley, 1992). A heteroziguidade dos marcadores também é menor que um, apesar de que se for utilizado marcadores altamente polimórficos, tais como microssatélites, a heteroziguidade pode ser maior. Na ausência de desequilíbrio de ligação, a fase de ligação entre o QTL e o marcador varia entre as famílias e, desta forma, a informação deve ser obtida com base na *intra*-família (Soller, 1990).

Alguns modelos para mapeamento genético têm sido propostos e podem ser classificados em seis grupos, segundo Hoeschele et al (1997): grupo 1 inclui regressão linear usando marcadores únicos ou múltiplos ligados ao QTL; grupo 2 inclui análises de máxima verossimilhança utilizando marcadores únicos ou múltiplos; grupo 3 inclui regressão do quadrado da diferença fenotípica dos pares de animais aparentados na proporção esperada de idêntico por descendência (IBD) do locus; grupo 4 inclui máxima verossimilhança restrita e modelo linear misto incorporando efeitos alélicos do QTL normalmente distribuídos com a matriz de covariância condicional aos dados observados dos marcadores; grupo 5 inclui análise de ligação exata Bayesiana usando marcadores únicos ou múltiplos ligados aos marcadores e ajustando um QTL bialélico ou de alelos infinitos; e, finalmente, o grupo 6 inclui as análises Bayesiana aproximadas de um QTL suposto bialélico. Estes métodos diferem em relação ao tempo computacional necessário para a realização de cada análise. Métodos que não

demandam um tempo computacional muito grande, como a regressão e a máxima verossimilhança, permitem a permutação dos dados, usada para calcular os valores de limiar de significância no genoma para os testes estatísticos e para incorporar mais facilmente múltiplos QTL e múltiplas características, porém estes métodos podem utilizar somente alguns tipos de parentes (como, por exemplo, irmãos-completos ou meio irmãos) e os métodos de regressão fornecem apenas estimativas somente quanto a localização do QTL (Hoeschele et al, 1997).

Quando se dispões de um grande número de animais, podem-se utilizar métodos baseados em mínimos quadrados para encontrar ligações entre QTL-marcador dentro de pedigrees sem que seja necessário acumular evidências de marcadores individuais ou intervalos ao longo do pedigree. Em muitos casos, no entanto, isto não é possível e muitos métodos para acúmulo de informação ao longo dos pedigrees foram desenvolvidos para análises de marcadores únicos. Entre os métodos para a detecção de ligação entre QTL e marcadores, o proposto por Haseman e Elston (1972) envolve a regressão do quadrado da diferença fenotípica entre duas famílias de irmãos completos, sob o número de alelos de um locus do marcador que eles compartilham e que é idêntico-por-descendência (IBD). Outro método similar foi proposto por Goldgar (1990) que leva em consideração múltiplos marcadores e mais do que duas famílias de irmãos completos.

O desenvolvimento do método da máxima verossimilhança (ML) para a detecção de QTL segregando em populações não endogâmicas é muito mais difícil do que para linhagens endogâmicas. Nas análises de F_2 somente os parâmetros de QTL e a variância residual não são conhecidas, apesar de que em populações segregantes a frequência do alelo na posição do QTL deve ser estimada em adição a esses parâmetros e a fase de ligação entre os marcadores e o QTL é desconhecida e todas as combinações possíveis dos haplótipos devem ser consideradas. Ainda há também variação entre famílias devido a efeitos genéticos não ligados aos marcadores e efeitos ambientais (Knott e Haley, 1992). Em análises de segregação que não abordam a variação entre famílias as estimativas dos parâmetros possuem um viés e genes espúrios podem ser detectados (Knott, Haley and Thompson, 1992). Outros autores

propõem a inclusão do efeito poligênico em adição ao efeito fixo de QTL, como em Fernando e Grossman (1989).

2.5 Características quantitativas não-normalmente distribuídas: dados de contagem

Algumas características importantes para o melhoramento animal são apresentadas na forma de contagem como, por exemplo, número de casos de mastite, tamanho de leitegada, número de carrapatos e número de ovos por grama de fezes (OPG). Diversas metodologias para tratar dados de contagem foram desenvolvidas, como, por exemplo, o modelo de regressão Poisson. Quando há ocorrência de superdispersão, um modelo linear com um resíduo pode ser mais robusto do que um assumindo somente uma amostragem pura de Poisson. Isto ocorre devido ao fato de que o resíduo no modelo linear “absorve” todas as causas de variação não contempladas no modelo (Rodrigues-Motta, 2006). No entanto, alguns problemas de superdispersão pode se dar devido ao número excessivo de contagem de valor zero, relativamente a quantidade já esperada devido a densidade da Poisson, por exemplo.

Como o método da máxima verossimilhança para mapeamento de QTL tem como pressuposição a distribuição normal dos resíduos, trabalhar com dados de contagem com dispersão se torna um problema. Para analisar características quantitativas que desviam da premissa de normalidade, porém mantendo as características desejáveis da distribuição normal, a transformação dos dados é necessária antes de se realizar o mapeamento de QTL.

As transformações matemáticas mais freqüentemente utilizadas para dados de contagem são raiz quadrada e Box-Cox. Quando escolhidas adequadamente, as transformações dos dados são capazes de “normalizar” estes de uma maneira eficiente. Em geral, a transformação pela raiz quadrada é utilizada quando os dados possuem distribuição Poisson (Ascombe, 1948) e a transformação de Box-Cox (Box & Cox, 1964) pode ser aplicada em diferentes tipos de dados por possuir um fator de transformação flexível inserido no modelo (Yang et al, 2006).

Dados de contagem inflacionados por zeros são caracterizados por um número excessivo de zeros. A palavra “inflacionado” é usada para enfatizar que a probabilidade de massa no ponto zero excede o número de zeros esperados sob uma determinada família paramétrica padrão (Tu, 2002). Se não for corretamente modelada, a presença desses zeros extras pode invalidar a distribuição assumida na análise, violando a integridade das inferências científicas (Lambert, 1992).

Atualmente diversos estudos contemplam dados de contagem com inflação de zeros. Extensões do modelo Poisson têm sido propostas com o uso de distribuições geométricas (Mullahy, 1986), modelo binomial negativo (Greene, 1994; Pardoe, 2001), modelo de “barreira” (Mullahy, 1986) e modelo Poisson inflacionado de zeros (ZIP) (Lambert, 1992). Particularmente, o ZIP tem sido usado extensivamente na literatura para acomodar dados de distribuição Poisson quando esta apresenta zeros extras em sua contagem (Rodrigues-Motta, 2006).

A principal razão para se modificar o modelo Poisson padrão é o fato de que os dados de contagem com excesso de zeros é mascarada como super-dispersão, sem ajustar um modelo que explique a sua causa.

No modelo ZIP se assume que zero é o único valor possível com uma probabilidade p e, com uma probabilidade $1-p$ uma variável aleatória Poisson (λ) é observada. Desta forma a probabilidade de um valor igual a zero é dada por uma mistura de dois componentes, uma vez que a distribuição Poisson também permite que valores iguais a zero também ocorram (Rodrigues-Motta, 2006).

2.6 Material e Métodos

2.6.1 População experimental

Os animais experimentais da geração F_2 foram produzidos dos cruzamentos entre animais F_1 mestiços Holandês x Gir, no Campo Experimental Santa Mônica (Valença, RJ), pertencente a EMBRAPA Gado de Leite.

Para isto foram utilizadas 28 fêmeas Gir (submetidas à superovulação) inseminadas com sêmen de quatro touros da raça Holandesa e os embriões transferidos para fêmeas receptoras. Como receptoras foram utilizadas cerca de 300 fêmeas Holandês x Gir, com “grau

de sangue” variando de $\frac{1}{2}$ a $\frac{7}{8}$ entre novilhas e vacas adultas em lactação. Estes acasalamentos foram repetidos com a finalidade de se obter cerca de 150 F₁. Destes 150 indivíduos F₁ (machos e fêmeas), apenas 4 machos foram escolhidos baseados na sua fertilidade, para serem pais da geração F₂. Sessenta e oito fêmeas F₁ foram acasaladas com cada um dos quatro touros, para construir 4 famílias (baseadas nos machos), evitando-se o parentesco entre o reprodutor e as fêmeas a ele designadas. Cada uma das famílias produziu, por meio da superovulação e transferência de embriões, cerca de 100 animais F₂. No total foram obtidos cerca de 400 indivíduos F₂ (200 machos e 200 fêmeas), ou seja, cerca de 100 animais por reprodutor, dos quais 263 foram avaliados genotipicamente para este trabalho, pois a época da realização desta genotipagem da população F₂ não havia sido terminada. Detalhes desta população podem ser obtidos em Gasparin (2007).

2.6.2 Coleta dos dados fenotípicos

A avaliação das características fenotípicas da população foi realizada no Campo Experimental Santa Mônica, da EMBRAPA Gado de Leite.

O número de carrapatos foi obtido através da contagem das fêmeas de carrapatos *Boophilus microplus* que completaram seu ciclo após a infestação artificial com número conhecido de larvas. Dada a importância da precisão na avaliação dessa característica, foram feitas infestações artificiais com 10.000 larvas por animal, e as foram realizadas no dia modal de queda dos carrapatos, que segundo Seifert (1971) ocorre no 21º dia após a infestação.

As larvas foram preparadas em laboratório na EMBRAPA Gado de Leite, incubando-se meio grama de ovos por frasco (equivalente a 5000 larvas). Dois frascos contendo larvas infectantes foram inseridos em colar adaptado e colocado na região cervical do animal, de modo que as larvas pudessem atingir ambos os lados do corpo.

Os animais foram avaliados em grupos de contemporâneos, com idades entre 10 e 14 meses, 20 a 30 animais por grupo de contemporâneo e todos mantidos a pasto desde a infestação até a contagem (21 dias). As infestações foram feitas durante duas épocas do ano, águas e seca, possibilitando que alguns grupos fossem mensurados mais de uma vez.

Foram contadas as fêmeas semi-ingurgitadas, de 4,5 a 8,0 mm de diâmetro, do lado direito do animal. As contagens foram realizadas pela manhã, até aproximadamente 9 horas, quando a maioria dos carrapatos se desprendia dos animais.

2.6.3 Coleta dos dados genotípicos e mapa de ligação

Os animais das gerações parentais, F_1 e F_2 foram genotipados para 5 marcadores do cromossomo 23 (BTA23). Maiores informações sobre os procedimentos utilizados para a genotipagem, marcadores e mapa obtido pode ser encontrado em Karen (2004).

O mapa de ligação utilizado neste trabalho foi o mapa consenso dos bovinos e pode ser obtido pelo <http://www.marc.usda.gov/genome/cattle/cattle.html>. O mapa dos marcadores utilizados estão descritos na tabela 1.

Tabela 1 – Mapa consenso dos marcadores utilizados nas análises

Marcador	Posição (cM)
BM1258	28.307
BB705	43.103
BM1818	58.195
CSSM24	65.715
BM1905	71.648

Fonte: USDA

2.6.4 Estudo de simulação

Para avaliar os diferentes modelos propostos foram realizadas simulações como segue. Foi simulado um único cromossomo com 45 cM de comprimento baseado nos dados reais com cinco marcadores não igualmente espaçados (0; 14,8; 29,9; 37,4; 43,3 cM) para uma população F_2 com tamanho populacional de 263 animais e variando-se o efeito aditivo do QTL em $\alpha = 0,05$; 0,1 e 0,2.

Um suposto QTL afetando a característica de interesse foi simulado a 20cM do primeiro marcador do grupo de ligação. A média populacional simulada foi de $\mu=2$. Estes cenários foram escolhidos a fim de mimetizar os dados reais e a estrutura real do pedigree foi usada para cálculo das probabilidades de origem para a população F_2 com o software QxPak (Perez-Encizo e Misztal, 2004). Os dados fenotípicos foram gerados de uma Poisson e de uma ZIP, com dois diferentes probabilidades de “zeros-verdadeiros” (0,2 e 0,5) com o auxílio do software estatístico SAS[®] 9.1.3.

Em cada cenário de simulação foram realizadas 1000 repetições. Cada um dos cinco diferentes métodos de análise foi testado com cada amostra simulada: os dados na escala original e duas

diferentes transformações para estes dados (raiz-quadrada e Box-Cox) foram analisadas pelo modelo Gaussiano (normal), modelo de regressão Poisson e modelo de regressão ZIP.

Dados sem a presença de QTL também foi simulada para realização das medidas de falsos positivos e falsos negativos.

2.6.5 Metodologias aplicadas para o mapeamento de QTL

Todas as metodologias para mapeamento de QTL apresentadas neste trabalho foram baseadas no método de regressão proposto por Haley et al. (1994), com a diferença de que este método foi entendido para o contexto dos modelos lineares generalizados (McCullagh and Nelder, 1989) para melhor atender as pressuposições do modelo Poisson. Além disso, foi proposto o modelo de Poisson com inflação de zeros para melhor atender a causa comum da superdispersão nos modelos Poisson. O procedimento da regressão para mapeamento de QTL por intervalos será revisto rapidamente; três alternativas foram testadas para lidar com os dados de contagem com diferentes inflações de zeros.

O modelo clássico de regressão para o mapeamento de QTL por intervalos para uma população F_2 não-endogâmica, pode ser descrita por (Haley et al, 1994):

$$y_i = \mu + aq_i + e_i \quad (1)$$

onde y_i é a característica observada (fenótipo) relativo no i -ésimo indivíduo F_2 , μ é a média geral, q_i é o coeficiente da localização para o i -ésimo indivíduo; a é o efeito aditivo do suposto QTL e e_i é o termo residual, assumido distribuição normal com média 0 e variância σ^2 , isto é $e_i \sim N(0, \sigma^2)$. O coeficiente q_i é dado pela combinação dos termos das probabilidades da linha de origem condicionais aos marcadores genotipados (Haley et al. 1994). Uma busca em grade foi realizada como descrito por Baret et al (1998). O princípio da busca em grade é: uma função de máxima verossimilhança foi avaliada em diferentes pontos igualmente espaçados ao longo da amplitude total de variação do parâmetro desejado. Para um delineamento experimental F_2 , esta amplitude varia entre -1 e 1. O número de vezes em que o algoritmo simplex é rodado dentro desta amplitude depende do incremento escolhido. Para cada um dos pontos da grade, uma máxima verossimilhança é estimada usando o método simplex sugerido por Nelder & Mead (1965) em um espaço bi-dimensional (efeito genético e variância residual). Do ponto da grade com o maior valor de verossimilhança, o algoritmo é repetido num espaço tri-dimensional

(com a adição da porcentagem dos machos heterozigotos) a fim de refinar a localização da máxima verossimilhança. Todas as estimativas realizadas foram baseadas nas análises de grade com incremento de 1 cM. O teste estatístico utilizado foi o *t-Student* para o modelo ZIP e teste de χ^2 para todos os outros procedimentos.

Dada a pressuposição de normalidade dos resíduos no modelo (1), podemos concluir que este não é o modelo ideal para análise de dados de contagens. Neste caso, um modelo de análise mais eficiente pode ser obtido por procedimentos de transformação de dados ou pela modelagem direta dos dados em sua escala original com a utilização de modelos mais gerais. Neste trabalho serão comparadas três metodologias, análise clássica utilizando do modelo (1) com transformação dos dados, modelo Poisson e modelo de mistura.

Um método flexível para a transformação de dados não normalmente distribuídos foi proposto por Box e Cox (1964). A família de transformações Box-Cox é dada por:

$$z_i = \begin{cases} \frac{y_i^\lambda - 1}{\lambda}, & \text{for } \lambda \neq 0 \\ \ln y_i, & \text{for } \lambda = 0 \end{cases} \quad (2)$$

onde z_i é a variável transformada e λ é chamado de fator de transformação. O método da máxima verossimilhança é usado para se escolher um valor ótimo de λ que vai gerar a transformação z_i . É importante mencionar que transformações mais usuais como logarítmica e raiz quadrada, são casos especiais da família Box-Cox quando λ é igual a -1, 1 e 0,5, respectivamente.

A transformação por Box-Cox já foi aplicada anteriormente ao mapeamento de QTL com características contínuas de distribuições assimétricas (Yang et al. 2006). Neste trabalho, tal procedimento foi aplicado à dados de contagem, mais especificamente de distribuição Poisson e Poisson inflado de zeros. As análises foram realizadas pelo procedimento TRANSREG do SAS, através da transformação simultânea dos dados de contagem e ajustando o modelo (1) para a grade das posições estipuladas.

Além da transformação Box-Cox que busca o valor de λ que melhor aproxima os dados de uma distribuição normal, também foi utilizada a transformação pela raiz quadrada, dada por:

$$w_i = \sqrt{y_i}, \text{ onde } y_i \geq 0 \quad (3)$$

A transformação pela raiz quadrada é conhecida por ser a mais apropriada quando os dados possuem uma distribuição Poisson (Ascombe, 1948). Portanto, neste trabalho foi considerada como alternativa a metodologia Box-Cox.

A transformação de Box-Cox possui sérias limitações quando a variável resposta contém zeros, uma vez que esta transformação não está definida para $\lambda=0$. Neste caso, uma pequena constante c é adicionada a cada observação para assegurar uma variância aproximadamente mais constante (Yamamura, 1999). Especificamente para dados Poisson, Ascombe (1948) demonstrou que $c=3/8$ é ótima no sentido de que a variância converge mais rapidamente a medida que a média aumenta. Uma vez que este trabalho foi realizado com variáveis Poisson (e variações desta), foi adotado este valor de c para ambas as transformações dadas pelas equações (2) e (3).

Como uma alternativa a transformação dos dados, foi considerada a regressão Poisson como uma extensão do modelo (1). Neste caso, foi assumido que cada observação y_i é desenhada de uma distribuição Poisson com esperança condicional μ_i , a função de densidade desta distribuição pode ser expressa segundo (McCullagh & Nelder, 1989) como:

$$f(y_i; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (4)$$

onde $\lambda = \exp(\mu + aq_i)$. Para a distribuição Poisson a função de ligação usada é a log. A função de ligação relaciona o preditor linear η ao valor esperado de μ de uma variável resposta y . Nos modelos lineares clássicos a média e o preditor linear são os mesmos e a função de ligação identidade é usada em ambos η e μ , podendo estes assumir qualquer valor real. No entanto, quando trabalhamos com contagens e a distribuição é Poisson (ou outra derivada desta) devemos obter valores de μ menores que zero, nesta circunstância a função de ligação identidade não é tão atrativa, em parte porque η pode ser negativo, enquanto μ pode não ser. Modelos para dados de contagens baseados na independência de dados levam naturalmente a efeitos multiplicativos e isto é mais bem expresso pela função de ligação log, onde $\eta = \log \mu$. Desta forma, efeitos aditivos que contribuem em η se tornam efeitos multiplicativos que contribuem para μ e μ é necessariamente positivo (McCullagh and Nelder, 1989).

Considerando independência entre as observações com a função de ligação descrita, a função do logaritmo da verossimilhança pode ser expressa como:

$$LL = \sum_i^n [-\mu_i + y_i \log(\mu_i) - \log(y_i!)] \quad (5)$$

no qual os parâmetros são os mesmo descritos em (4).

O último método utilizado neste trabalho foi o modelo Poisson Inflado de Zeros (ZIP). Este modelo usa o procedimento de modificar a distribuição conjunta discreta das variáveis aleatórias e a função condicional da média das contagens, ao invés de contar com

transformações para obter a normalidade da distribuição condicional de variáveis contínuas latentes (Greene, 1994). Mais especificamente, em uma regressão ZIP um modelo logit com pressuposições binomiais é usado para determinar se uma contagem individual provém de um perfil “sempre zero” ou “nem sempre zero”. A função de densidade de um modelo ZIP tem a seguinte fórmula:

$$f(y_i) = \begin{cases} \varpi_i + (1 - \varpi_i) \exp(-\mu_i) & \text{para } y_i = 0 \\ (1 - \varpi_i) \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} & \text{para } y_i > 0 \end{cases} \quad (6)$$

onde $1 - \varpi_i = P(y_i \sim \text{Poisson}(\mu_i))$.

o logaritmo da verossimilhança é expresso por:

$$LL = \sum_{i=1}^n \left[I(y_i = 0) \log(\varpi_i + (1 - \varpi_i) \exp(-\mu_i)) + I(y_i > 0) (\log(1 - \varpi_i) + y_i \log(\mu_i) - \mu_i - \log(y_i!)) \right] \quad (7)$$

A regressão Poisson foi implementada usando o procedimento GENMOD do SAS com a função de ligação log. O modelo ZIP foi ajustado com o procedimento NLMIXED, seguindo a sintaxe descrita em Liu e Cella (2008).

Para um melhor entendimento do modelo de regressão ZIP com amostras independentes, sejam $\mathbf{y} = (y_1, \dots, y_n)'$ variáveis aleatórias independentes tais que y_i é zero com probabilidade p_i , ou $y_i \sim \text{Poisson}(\lambda_i)$ com probabilidade $1 - p_i$. Então, $\Pr(y_i = 0) = p_i + (1 - p_i)e^{-\lambda_i}$, onde $e^{-\lambda_i}$ é a probabilidade de se observar zero sob a distribuição Poisson e $y_i = k$ com probabilidade $(1 - p_i)e^{-\lambda_i} \lambda_i^k / k!$, $k = 1, 2, \dots, \infty$. Além disso, seja os parâmetros $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ e $\mathbf{p} = (p_1, \dots, p_n)'$ que satisfaz os modelos

$$\begin{aligned} \log(\boldsymbol{\lambda}) &= \mathbf{X}_\lambda \boldsymbol{\beta}_\lambda \\ \text{logit}(\mathbf{p}) &= \left\{ \log\left(\frac{p_i}{1 - p_i}\right) \right\} = \mathbf{X}_p \boldsymbol{\beta}_p, \end{aligned} \quad (8)$$

onde \mathbf{X}_λ e \mathbf{X}_p são matrizes de incidência das covariáveis relacionadas aos parâmetros desconhecidos $\boldsymbol{\beta}_\lambda$ e $\boldsymbol{\beta}_p$ para os vetores $\log(\boldsymbol{\lambda}) = \{\log \lambda_i\}$ e $\left\{ \log\left(\frac{p_i}{1 - p_i}\right) \right\}$, respectivamente.

Sob a teoria de modelos lineares generalizados (McCullagh e Nelder, 1989), log e logit são as funções de ligação “naturais” ou transformações da média linearizada da Poisson e probabilidades de sucessos de Bernoulli, respectivamente.

2.6.6 Comparação dos modelos

Os modelos foram comparados com base em medidas de desempenho e confiabilidade nas detecções, como descrito a seguir:

1. Variância (VAR): foi calculada a variância dos valores do efeito aditivo (α) obtido com as 1000 simulações;
2. Viés: foi obtido somando-se a diferença entre o valor do efeito aditivo (α) obtido e o valor simulado dividido pelo número de simulações (1000).

Com os valores de variância e de viés, podemos calcular o erro quadrático médio (EQM);

3. MDA (média das distâncias absolutas): calculada através do valor absoluto da diferença entre a posição onde foi calculado o maior valor da máxima verossimilhança dada pela simulação e a posição simulada para o QTL;
4. Falso positivo (FP): porcentagem de vezes em que foi encontrado um QTL significativo ($p\text{-valor} \leq 0,05$), em qualquer posição, quando nenhum QTL foi incluído na simulação;
5. Falso negativo (FN): porcentagem de vezes em que não foi encontrado um QTL significativo na posição 20cM, quando este foi simulado.

2.6.7 Análise dos dados reais

Os dados reais foram analisados com todos os modelos descritos anteriormente para que se tivesse uma idéia do comportamento dos mesmos em condições reais. Os dados foram então divididos em dois arquivos de análises segundo a idade em que os animais foram submetidos às contagens de carrapatos (primeira e segunda contagem).

Para a determinação do limiar em que um QTL seria considerado significativo, foi utilizado testes de permutação seguindo o modelo sugerido por Churchill e Doerge (1994), onde a estrutura do pedigree e dos marcadores é mantida, quebrando-se apenas a estrutura dos dados fenotípicos. Foram realizadas 1000 permutações, dentro de cada modelo de análise utilizado, para determinação do limiar de significância de 5%.

2.7 Resultados e Discussão

2.7.1 Comparação dos modelos

Na tabela 2 e 3 estão apresentados os resultados da comparação dos modelos segundo as taxas de falsos negativos e positivos para cada cenário simulado.

Tabela 2- Taxa de falsos negativos para cada modelo nos diferentes cenários simulados

Modelo	$\alpha=0,05$	$\alpha=0,1$	$\alpha=0,2$
P(0) = 0			
Normal	74,3	25,1	0
Box-Cox	70,2	25,0	1
Raiz-Quadrada	74,4	25,6	1
Poisson	74,6	25,2	0
ZIP	77,2	25,9	0
P(0) = 0,2			
Normal	38,8	70,2	18,5
Box-Cox	32,3	75,4	30,2
Raiz-Quadrada	44,7	81,3	43,0
Poisson	69,4	44,4	4,4
ZIP	78,0	37,1	2,0
P(0) = 0,5			
Normal	91,1	87,0	61,6
Box-Cox	93,7	94,6	92,0
Raiz-Quadrada	92,4	90,7	80,3
Poisson	60,0	48,9	21,5
ZIP	86,4	57,1	4,3

Onde α representa o efeito do QTL simulado, P(0) representa a probabilidade de zeros na contagem simulada.

Tabela 3 – Taxa de falsos positivos para cada modelo nas diferentes inflações de zeros simuladas

	P(0) = 0	P(0) = 0,2	P(0) = 0,5
Normal	18,7	15,9	17,9
Box-Cox	17,5	15,1	16,3
Raiz-Quadrada	18,6	15,9	16,2
Poisson	17,6	54,7	79,6
ZIP	17,5	17,3	17,6

P(0) representa a probabilidade de zeros na contagem simulada.

No caso onde a probabilidade de zeros foi igual a zero, a distribuição Poisson apresentou taxa de falsos negativos (FN) em torno de 74 e 25 para efeito do QTL iguais a 0,05 e 0,1, respectivamente; e quase nulo quando o efeito foi de 0,2 (Tabela 2). A transformação Box-Cox apresentou melhor taxa de FN que os outros procedimentos. Interessante notar que neste caso o modelo ZIP apresentou maiores taxas de FN. Assim, parece que a inclusão de um parâmetro extra relativo ao “zero perfeito” leva a uma pequena penalidade se os dados possuem uma distribuição Poisson sem inflação, especialmente em situações com um pequeno efeito do QTL.

Com uma inflação moderada de zeros e considerando um pequeno efeito do QTL o modelo ZIP apresenta uma taxa de FN um pouco maior do que o modelo Poisson e quase duas vezes maior do que o modelo Gaussiano com ou sem a transformação dos dados. No entanto, com QTL de maior efeito (0,1 e 0,2) o modelo ZIP atua de maneira muito superior aos outros métodos, uma vez que este modelo permite que haja uma separação entre os zeros verdadeiros (provenientes de uma distribuição de Bernoulli) e os zeros falsos (provenientes de uma distribuição Poisson).

Finalmente, com uma inflação extrema de zeros ($P(0)=0,5$) o modelo de regressão Poisson atuou melhor em situações onde o efeito do QTL é pequeno a moderado e o modelo ZIP apresentou o segundo melhor resultado, nas mesmas condições. Apesar de tudo, sob um efeito de QTL grande, o modelo ZIP ultrapassa todos os outros métodos.

Em termos da taxa de falsos positivos (FP) todos os modelos apresentaram aproximadamente a mesma taxa de FP quando não há inflação de zeros. No entanto, quando há inflação moderada de zeros, o modelo de regressão Poisson apresentou quase três vezes a taxa de FP do que os demais métodos. Este resultado pode explicar o porquê do modelo Poisson apresentar uma menor taxa de FP do que o modelo ZIP em alguns cenários com inflação de zero, como quando o efeito do QTL foi pequeno ou moderado, parece que o modelo Poisson é menos conservativo. Ainda interessante, o modelo Gaussiano, utilizando-se ou não a transformações dos dados, apresentou resultados muito similares ao modelo ZIP em termos da taxa de FP.

Em geral, os modelos Gaussianos possuem taxa de FN maior e empatam em relação à taxa de FP e o modelo Poisson, senão há inflação de zeros, tem a mesma taxa de FP que o apresentado pelo modelo ZIP e ganha, por pouco, do modelo ZIP em termos da taxa de FN se o efeito do QTL é pequeno. Entretanto, se o efeito for maior, o modelo ZIP possui uma taxa de FP e FN menor que todos os outros métodos testados.

Nas tabelas 4, 5 e 6 estão demonstrados os resultados obtidos para as medidas de desempenho de cada modelo em cada cenário de simulação.

Tabela 4 – Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0$)

	Medidas de Desempenho		
	VAR	VIÉS	MDA
<u>$a = 0,05$</u>			
Normal	0,120	0,347	12,764
Box-Cox	0,034	0,148	12,687
Raiz Quadrada	0,004	0,023	12,746
Poisson	0,002	0,004	12,779
ZIP	0,003	-0,003	12,894
<u>$a = 0,1$</u>			
Normal	0,083	0,687	8,165
Box-Cox	0,029	0,273	9,097
Raiz Quadrada	0,003	0,055	8,253
Poisson	0,002	0,030	8,194
ZIP	0,002	0,030	8,180
<u>$a = 0,2$</u>			
Normal	0,073	1,311	3,634
Box-Cox	0,047	0,452	6,234
Raiz Quadrada	0,002	0,074	3,672
Poisson	0,001	0,002	3,671
ZIP	0,001	0,002	3,686

Onde a representa o efeito do QTL simulado.

Na tabela 4, onde não há inflação de zeros ($P(0)=0$, distribuição Poisson), observamos de uma maneira geral que o modelo Gaussiano, sem uso de transformações, foi o que obteve os piores valores de variância e viés comparados com todos os outros métodos em qualquer efeito de QTL considerado, o que já era esperado uma vez que este modelo assume normalidade de resíduos, o que não ocorre na distribuição Poisson. De maneira geral também podemos observar que os modelos Poisson e ZIP foram os que melhor se saíram nestas mesmas categorias o que também já era esperado uma vez que estamos utilizando dois modelos que assumem a distribuição Poisson dos dados.

Quando consideramos um efeito pequeno do QTL, as transformações mostraram ajustar bem os dados com valores de variância e viés muito menores do que os obtidos com o modelo sem a transformação (tabela 4), sendo ainda que a transformação pela Raiz Quadrada resultou em variância muito próxima da obtida com os modelos Poisson e ZIP. No entanto, em relação às MDA, todos os modelos testados apresentaram valores muito semelhantes. Isso mostra que apesar dos modelos Poisson e ZIP apresentarem um menor erro na determinação do efeito, quando se trata da acurácia na localização do QTL, todos os modelos mostram o mesmo desvio. O mesmo pode ser observado quando consideramos valor moderado a grande do efeito do QTL, com uma única observação de que, em relação à MDA, o modelo Gaussiano com a transformação Box-Cox apresentou o pior valor quando todos os outros métodos apresentaram valores semelhantes para essa medida.

Tabela 5 – Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0,2$)

	Medidas de Desempenho		
	VAR	VIÉS	MDA
<u>$\alpha = 0,05$</u>			
Normal	0,262	0,303	14,153
Box-Cox	0,131	0,161	14,396
Raiz Quadrada	0,018	0,014	14,495
Poisson	0,008	0,002	14,387
ZIP	0,003	0,008	13,349
<u>$\alpha = 0,1$</u>			
Normal	0,214	0,550	12,152
Box-Cox	0,114	0,276	13,222
Raiz Quadrada	0,015	0,022	13,510
Poisson	0,006	0,010	12,184
ZIP	0,002	0,005	9,620
<u>$\alpha = 0,2$</u>			
Normal	0,156	1,040	7,061
Box-Cox	0,085	0,446	10,329
Raiz Quadrada	0,011	0,032	9,546
Poisson	0,004	0,008	7,107
ZIP	0,002	0,002	3,948

Onde α representa o efeito do QTL simulado.

Na tabela 5, onde já é considerada uma inflação de zeros de 20% ($P(0)=0,2$), mais uma vez o modelo Gaussiano sem transformações foi o que obteve pior determinação do efeito do QTL (seja o efeito simulado pequeno ou grande). Já os modelos Poisson e ZIP foram os que melhor obtiveram melhores efeitos do QTL, sendo que já é possível observar para efeitos moderado e grande, uma superioridade do modelo ZIP na determinação do efeito do QTL simulado.

Em relação às MDA, quando o efeito do QTL é pequeno, todos os modelos se comportam de maneira semelhante, com exceção do modelo ZIP que apresenta um valor menor, mas ainda bastante desviado da posição simulada (tabela 5). Quando consideramos QTL com efeito moderado, alguns padrões anteriormente observados se mantêm, a transformação pela raiz dos dados se mostra eficiente em relação a determinação do efeito e o modelo Poisson fica muito próximo ao modelo ZIP, porém a superioridade do modelo ZIP já pode ser notada ao observamos que o erro na determinação do efeito é muito baixo e a localização do QTL é mais precisa do que os outros modelos considerados. Isso é comprovado quando observamos a simulação de um efeito grande de QTL.

Tabela 6 – Resultados das medidas de desempenhos dos diferentes modelos testados para o mapeamento de QTL com probabilidade de zeros igual a zero ($P(0)=0,5$)

	Medidas de Desempenho		
	VAR	VIÉS	MDA
<u>$a = 0,05$</u>			
Normal	0,358	0,175	14,599
Box-Cox	0,044	-0,018	14,985
Raiz Quadrada	0,026	-0,008	14,756
Poisson	0,027	0,011	14,615
ZIP	0,005	0,006	14,000
<u>$a = 0,1$</u>			
Normal	0,296	0,319	14,125
Box-Cox	0,041	-0,048	15,179
Raiz Quadrada	0,024	-0,024	14,690
Poisson	0,022	0,014	14,151
ZIP	0,003	0,007	11,102
<u>$a = 0,2$</u>			
Normal	0,228	0,613	11,359
Box-Cox	0,039	-0,102	14,757
Raiz Quadrada	0,019	-0,047	13,319
Poisson	0,016	0,017	11,380
ZIP	0,003	0,005	6,130

Onde a representa o efeito do QTL simulado.

Na tabela 6, onde estão representados as medidas de desempenho num cenário extremo de inflação de zeros ($P(0)=0,2$), a superioridade do modelo ZIP quando há inflação de zeros pode ser notada em todos os cenários simulados, com exceção das MDA quando considerado um efeito pequeno do QTL onde todos os métodos se comportaram de maneira semelhante. Em relação à determinação do efeito do QTL, pode ser notado que tanto o modelo Gaussiano com a transformação Raiz Quadrada quanto o modelo Poisson apresentaram resultados muito próximos, mostrando que neste quesito a transformação pode ser empregada para obter uma estimativa aproximada do efeito, porém não para a determinação da localização do QTL.

De maneira geral, é possível observar que quando os dados apresentam uma inflação de zeros moderada ou extrema, a melhor opção é adotarmos o modelo ZIP, uma vez que este modelo leva em consideração tanto a distribuição dos zeros “perfeitos” (que seguem uma distribuição de Bernoulli) quanto os zeros provenientes da distribuição Poisson. Porém, na impossibilidade de se utilizar um modelo específico, como quando utilizamos programas que assumem a normalidade dos resíduos, a transformação através da Raiz Quadrada mostrou-se uma boa alternativa para estimarmos o efeito do QTL, porém sem a localização do mesmo.

2.7.2 Análise dos dados reais

Os dados reais foram analisados por todos os métodos testados com as simulações descritas na seção anterior. Isso nos permite ter uma idéia das possíveis interpretações errôneas que podem ser cometidas, quando o método mais adequado para cada tipo de dados não é utilizado.

Os dados da primeira contagem foram analisados e os limiares de 5% de significância para cada modelo foram estimados através de 1000 permutações. Nas figuras de 1 a 19 estão os resultados dessas análises.

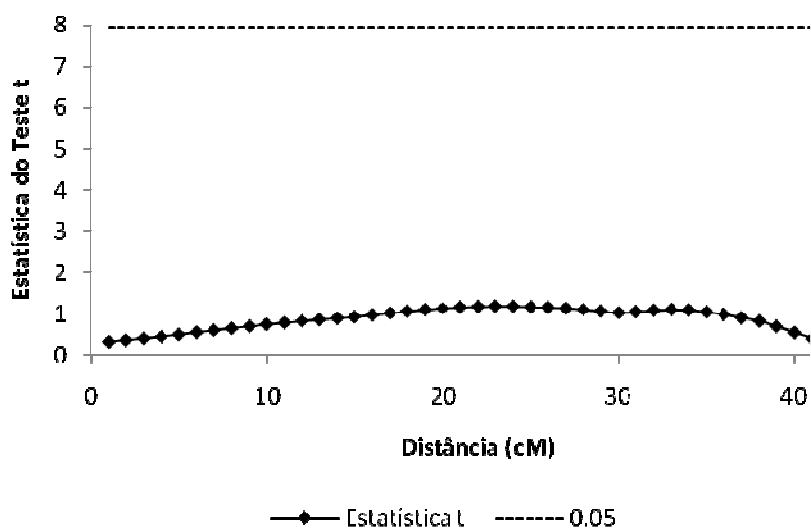


Figura 1 – Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano sem uso de transformações

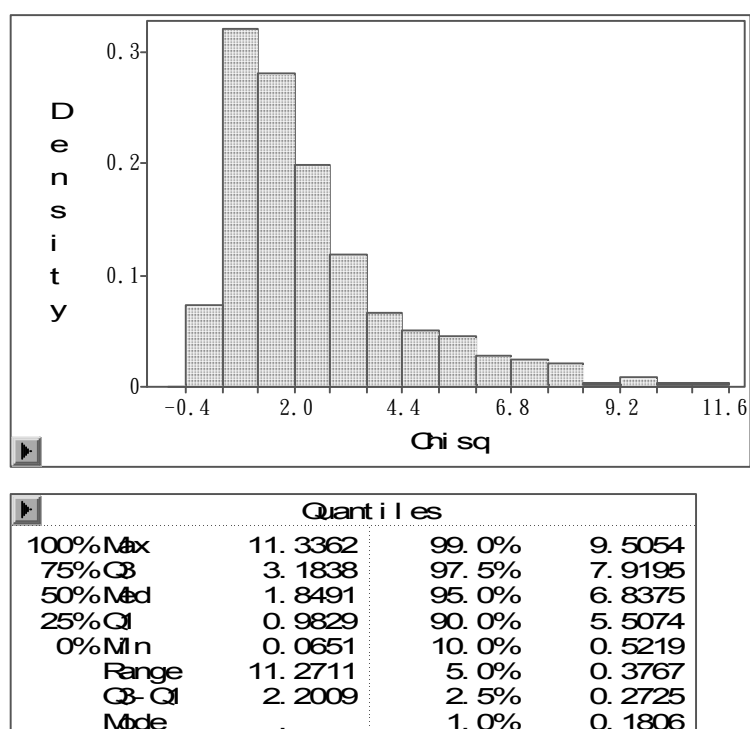


Figura 2 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano sem transformação dos dados

Nas figuras 1 e 2 estão os resultados obtidos para o mapeamento de QTL para contagem de carrapatos, na primeira contagem, sob o cromossomo 23, usando o

modelo gaussiano sem transformações de dados. Pela figura 1 pode se concluir que não há nenhum indício de QTL, porém como já foi discutido anteriormente, este não seria o melhor modelo para análise desses dados.

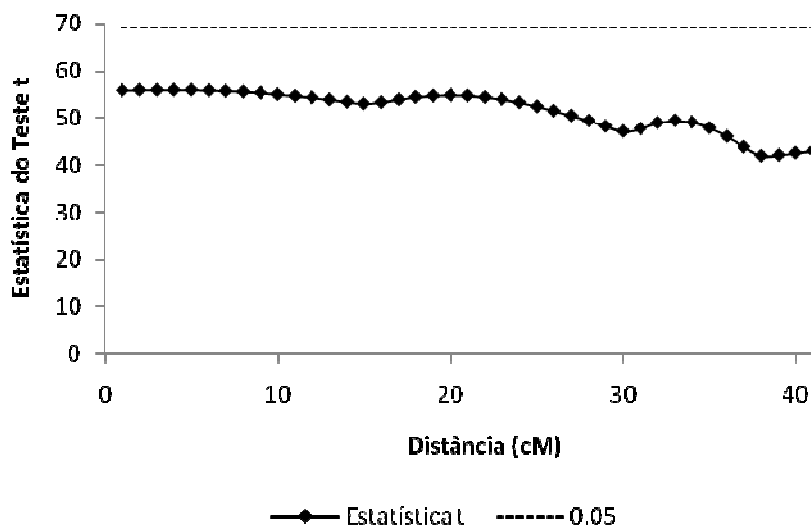


Figura 3 - Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano com uso da transformação de Box-Cox

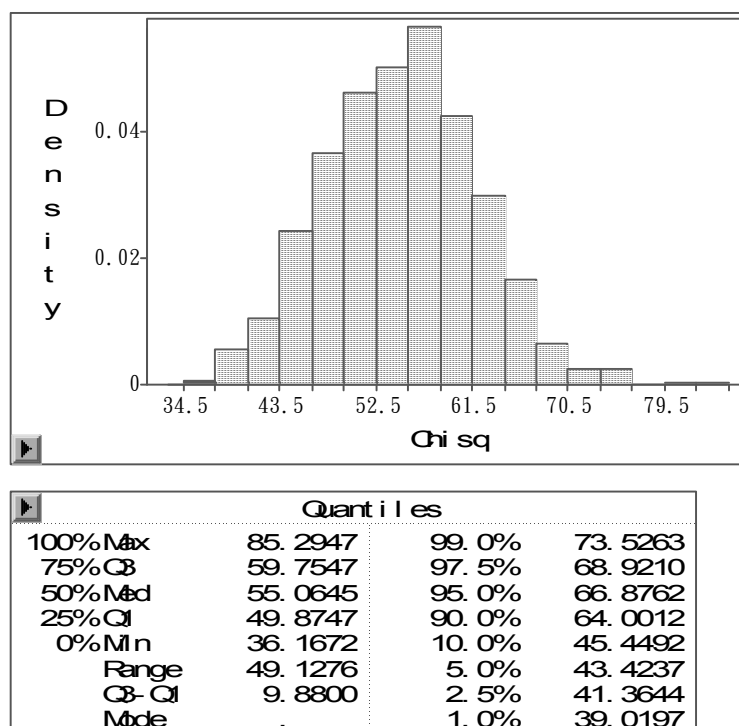


Figura 4 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação de Box-Cox dos dados

Nas figuras 3 e 4 estão os resultados do mapeamento de QTL para os dados reais pelo modelo Gaussiano com a transformação dos dados por Box-Cox. Como foi discutido anteriormente, este modelo também não seria o melhor para analisar este perfil de dados e sua utilização, neste caso, nos indica a ausência de QTL.

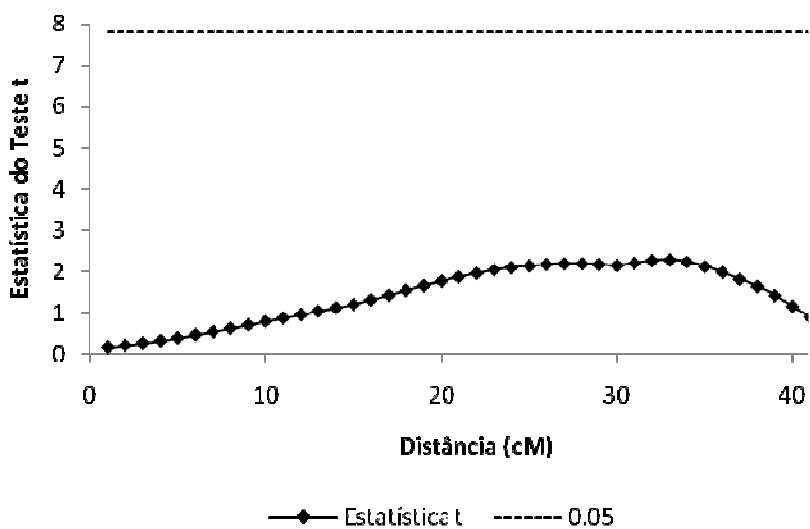


Figura 5 - Mapeamento de QTL, para a primeira contagem, através do modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados

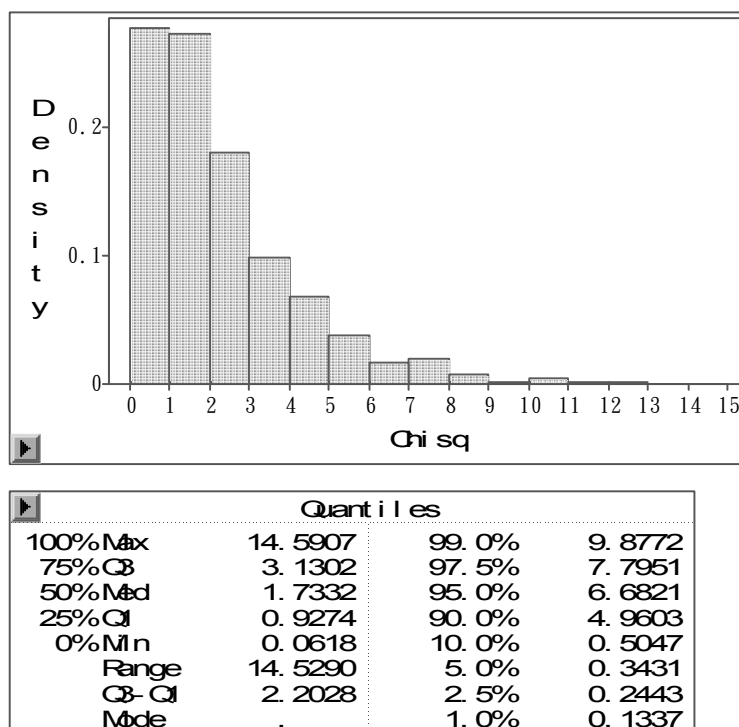


Figura 6 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados

Nas figuras 5 e 6 estão os resultados do mapeamento de QTL pelo modelo Gaussiano com os dados transformados através da Raiz Quadrada. Apesar de ser um modelo um pouco melhor para determinação do efeito do QTL, não é o mais preciso na determinação da posição, como já descrito anteriormente.

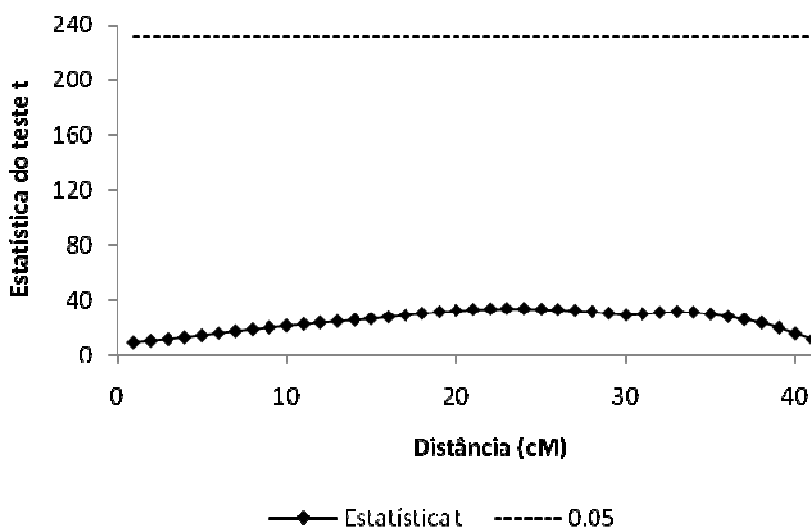


Figura 7 - Mapeamento de QTL, para a primeira contagem, através do modelo Poisson

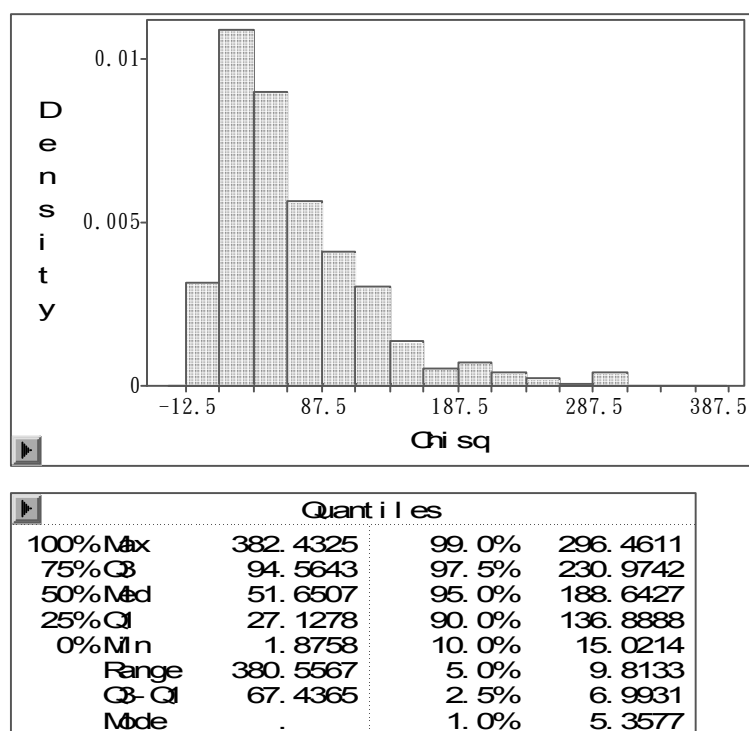


Figura 8 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Poisson

Nas figuras 7 e 8 estão os resultados de mapeamento de QTL, para a primeira contagem, pelo modelo de Poisson. Este modelo já seria mais confiável para a determinação da posição, porém para este perfil de dados não encontra nenhum QTL.

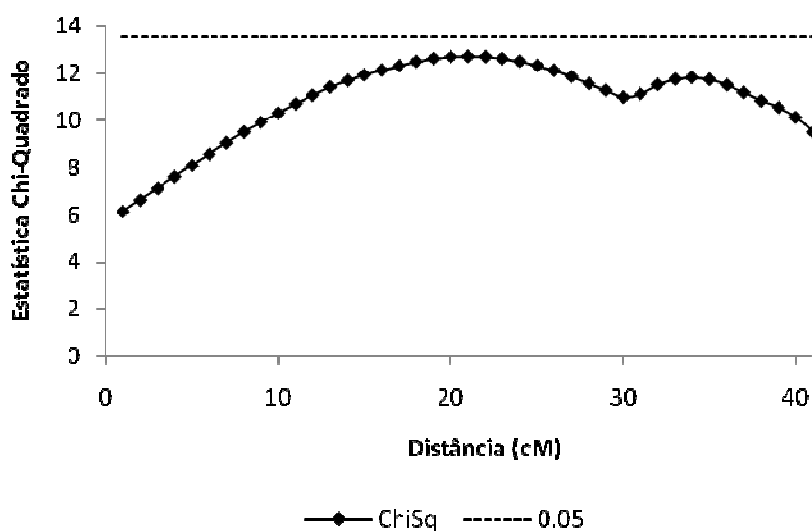


Figura 9 - Mapeamento de QTL, para a primeira contagem, através do modelo ZIP

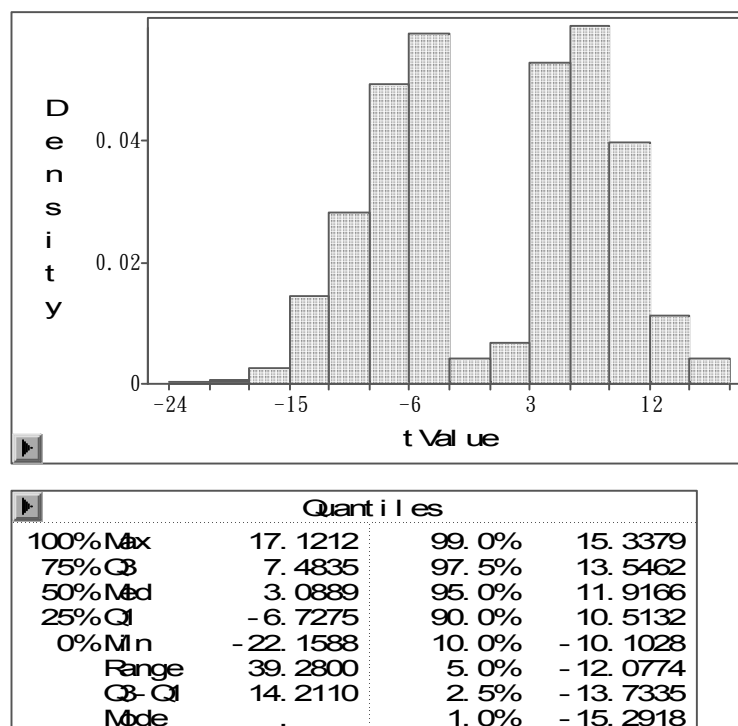


Figura 10 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo ZIP

Nas figuras 9 e 10 estão os resultados do mapeamento de QTL com o uso do modelo ZIP. Por este modelo, como já discutido anteriormente, pode-se ter uma maior precisão na determinação da posição do QTL. Por este resultado pode se observar um QTL sugestivo próximo a posição 20 cM.

Na tabela 7 estão os resultados numéricos obtidos com cada modelo segundo a posição e o efeito estimados. Por esta tabela podemos concluir que se realmente há presença de um QTL no cromossomo 23, se trata de um QTL de efeito moderado (explicando cerca de 20% da variância fenotípica) e, de acordo com o obtido através das simulações, podemos estimar sua posição ao redor de 20 cM.

Tabela 7 – Valores das estimativas de posição e efeitos aditivos obtidos para o mapeamento de QTL dos dados reais (primeira contagem) com os diferentes modelos

Modelo	Estimativas	
	Posição (cM)	Efeito do QTL
Gaussiano	23	2,59
Box-Cox	20	1,37
Raiz Quadrada	33	0,30
Poisson	23	0,15
ZIP	21	0,36

Os termos Box-Cox e Raiz Quadrada referem-se, respectivamente, ao modelo Gaussiano com cada uma das transformações.

A seguir são apresentados os resultados obtidos para a segunda contagem de carrapatos e as respectivas análises de permutação para determinação do limiar a 5% de significância. Como o perfil de dados desta contagem é um pouco diferente dos anteriormente descritos (maior número de zeros), podemos visualizar o comportamento dos diferentes modelos com inflação real de zeros.

Nas figuras 11 e 12 estão os resultados de mapeamento de QTL e seu respectivo teste de permutação para o modelo Gaussiano. Nesta análise poderíamos concluir pela sugestão de um QTL próximo a posição 20cM, porém como mostrado através das simulações, este modelo de análise não retorna resultados confiáveis para este perfil de dados.

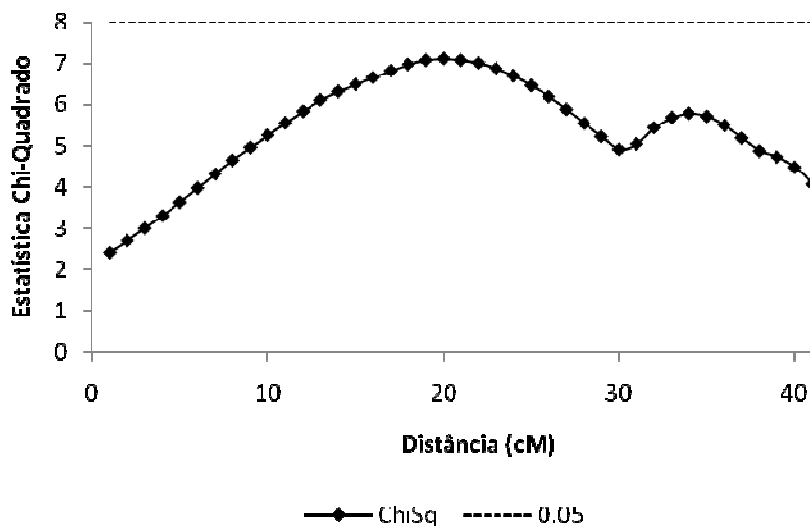


Figura 11 - Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano sem uso de transformações

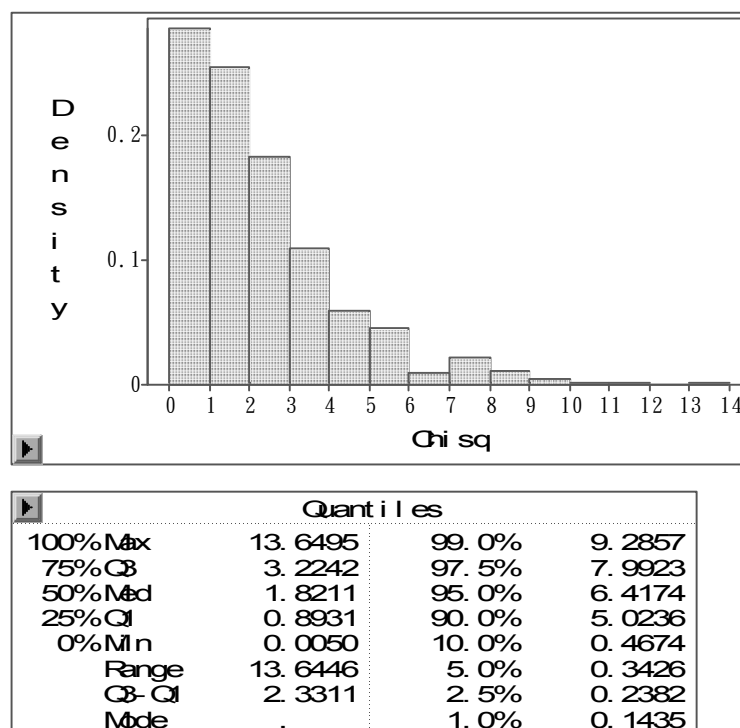


Figura 12 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano sem transformação dos dados

Nas figuras 13 e 14 estão os resultados para o mapeamento de QTL pelo modelo Gaussiano após a transformação dos dados pelo método de Box-Cox. Os resultados

são muito parecidos com os obtidos com a primeira contagem, apenas com uma pequena variação do valor obtido para o limiar.

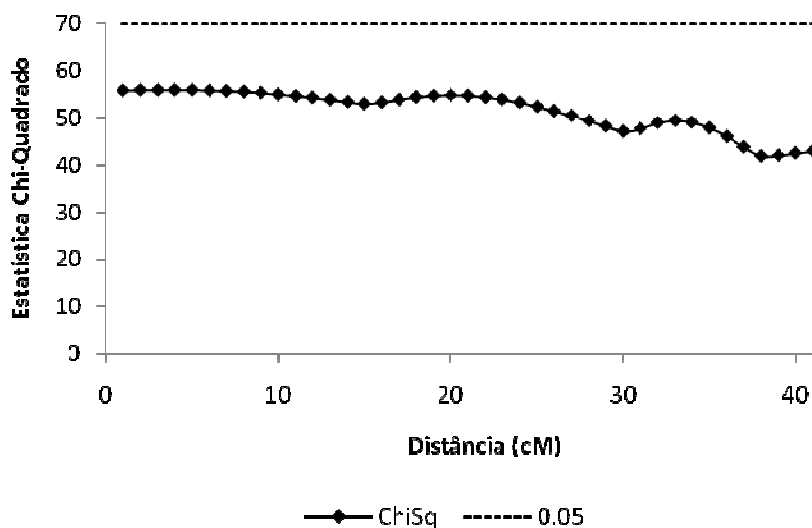


Figura 13 - Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano com uso da transformação de Box-Cox

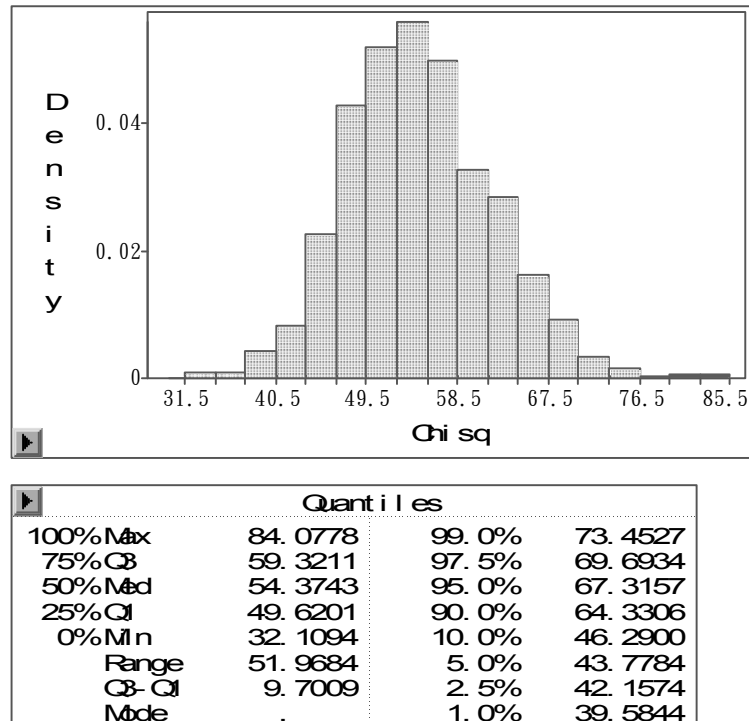


Figura 14 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação de Box-Cox dos dados

Nas figuras 15 e 16 estão os resultados obtidos pra o mapeamento de QTL com o modelo gaussiano após a transformação dos dados pela Raiz Quadrada. Podemos verificar que o formato da curva obtida está mais próximo ao que se esperaria de um mapeamento onde há indicativo de QTL, porém tal conclusão não pode ser feita apenas com a visualização deste gráfico, uma vez que este modelo não possui boa acurácia na determinação da posição do QTL.

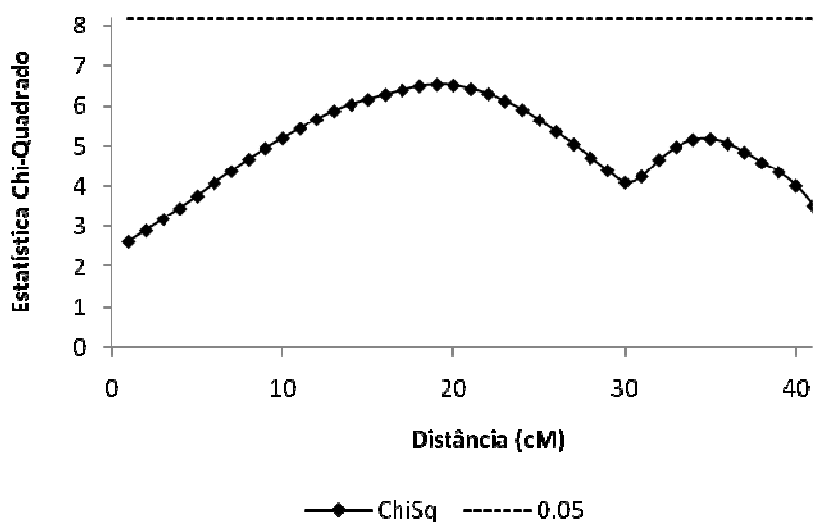


Figura 15 - Mapeamento de QTL, para a segunda contagem, através do modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados

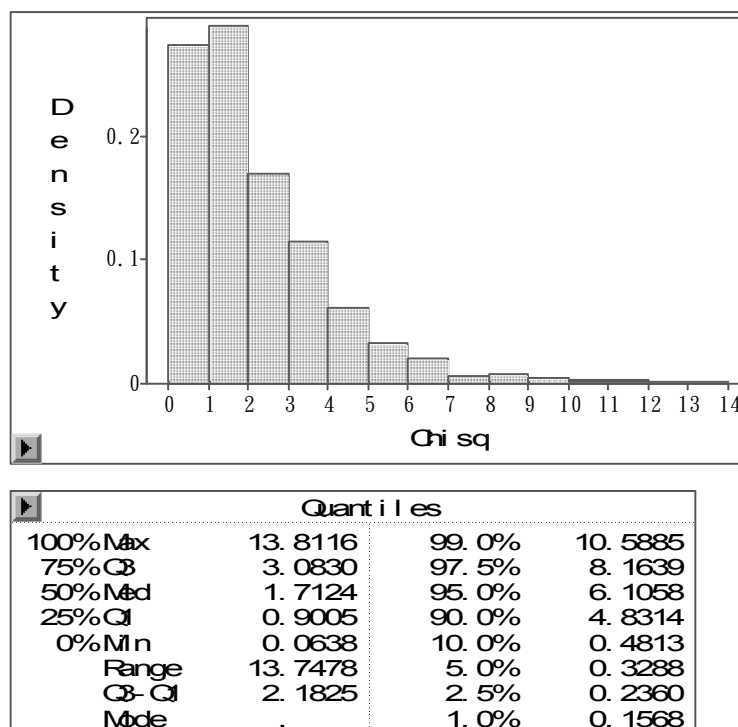


Figura 16 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Gaussiano com uso da transformação pela Raiz Quadrada dos dados

Nas figuras 17 e 18 estão os resultados do mapeamento de QTL com o modelo Poisson. Segundo o obtido através das simulações, este modelo pode mostrar a posição do QTL com uma maior confiabilidade, pois é o modelo que mais se aproxima da representação da distribuição dos dados reais.

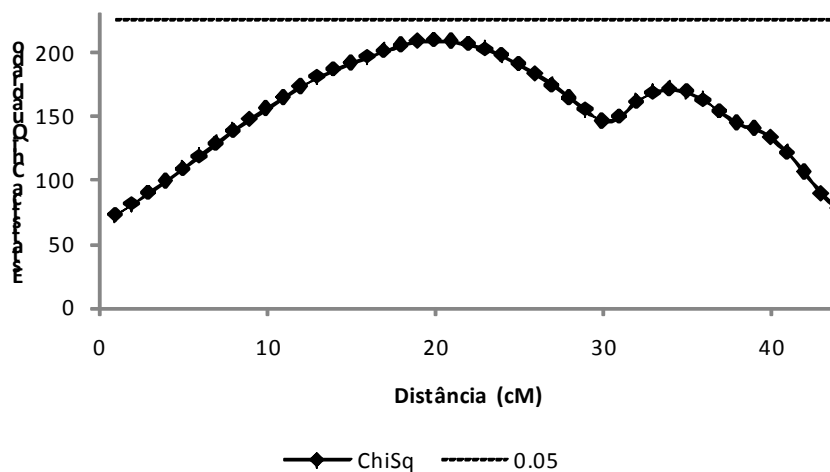


Figura 17 - Mapeamento de QTL, para a segunda contagem, através do modelo Poisson

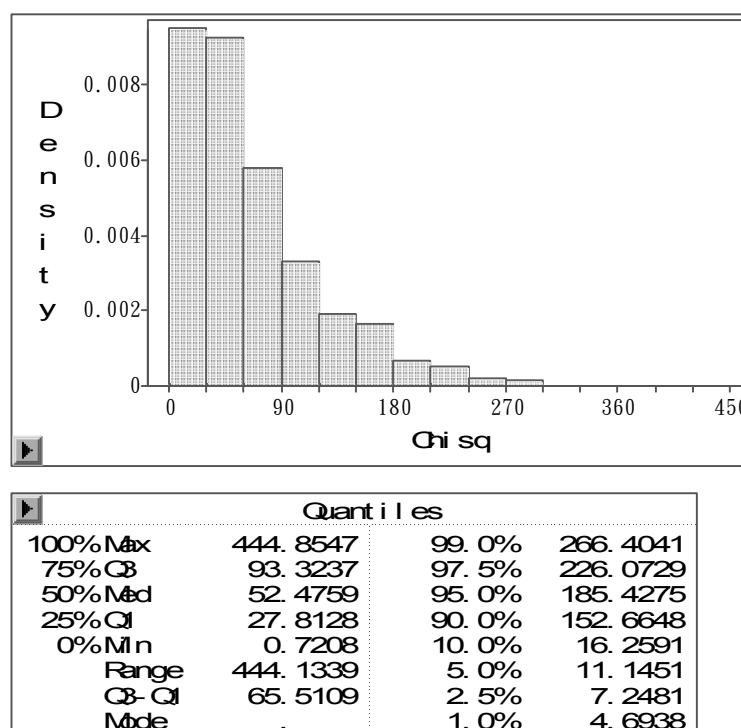


Figura 18 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo Poisson

Nas figuras 19 e 20 pode se observar que o ZIP é o melhor modelo de análise dos dados de contagem com inflação de zeros. Através do gráfico podemos estimar a posição onde há um indicativo de QTL em 20 cM.

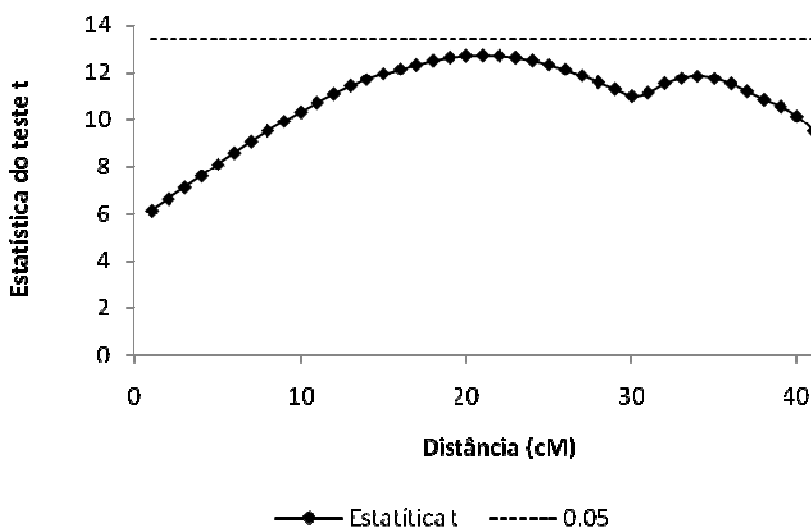


Figura 19 - Mapeamento de QTL, para a segunda contagem, através do modelo ZIP

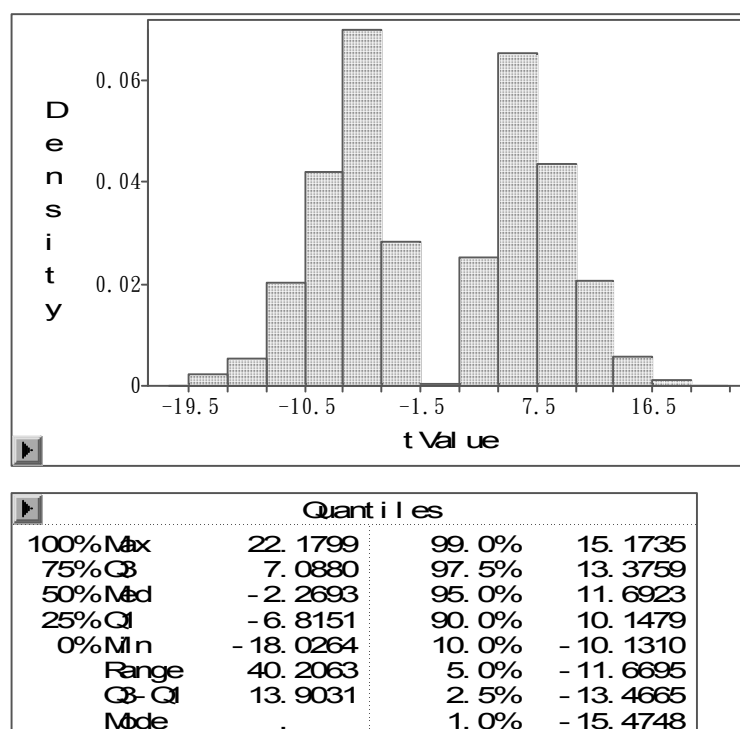


Figura 20 – Saída do programa SAS mostrando o resultado do teste de permutação para obtenção do valor de limiar para o modelo ZIP

Na tabela 8 estão os valores obtidos com cada modelo em relação ao efeito e a posição. Através dela, percebe-se que o QTL possui um efeito moderado (explicando

cerca de 20% da variação fenotípica) assim como o obtido com os dados da primeira contagem, indicando que provavelmente se trata do mesmo QTL.

Tabela 8 – Valores das estimativas de posição e efeitos obtidos para o mapeamento de QTL dos dados reais (primeira contagem) com os diferentes modelos

Modelo	Estimativa	
	Posição (cM)	Efeito do QTL
Gaussiano	20	6,76
Box-Cox	20	1,37
Raiz Quadrada	19	0,64
Poisson	20	0,41
ZIP	21	0,36

Os termos Box-Cox e Raiz Quadrada referem-se, respectivamente, ao modelo Gaussiano com cada uma das transformações.

3 CONCLUSÕES

Quando trabalhamos com dados que fogem da distribuição normal, como os dados provenientes de contagens, uma análise descritiva dos dados é sempre indicada. Se houver inflação de zeros, recomenda-se o uso do modelo ZIP para a análise desses dados. Se há presença de zeros na contagem que não possa ser considerado inflação, pode ser usada tanto o modelo Poisson ou um método de transformação dos dados (o presente trabalho recomenda Raiz Quadrada). A vantagem do uso da transformação é que há possibilidade do uso de qualquer programa disponível atualmente na internet, porém a desvantagem é que quando se usa os dados transformados, a interpretação dos dados não é possível diretamente, sendo necessária uma transformação reversa desses resultados.

Foram confirmados os já conhecidos resultados que a transformação pela Raiz Quadrada pode ser aplicada à dados que possuem distribuição Poisson. Porém se não é possível confiar que os dados em questão possuem distribuição Poisson, uma família mais geral de transformações, como é o caso da Box-Cox, é aconselhada.

Foi encontrado um indicativo de QTL no cromossomo 23 na posição 21 cM para os dados de contagem de carrapatos em bovinos mestiços, porém outros estudos devem ser realizados com o aumento do número de animais e com o adensamento do mapa para este cromossomo.

REFERÊNCIAS

- ASCOMBE, F.J. The transformation of Poisson, binomial and negative-binomial data. **Biometrika**, Oxford, v.35, p.246-254, Dec.1948.
- BARET, P.V.; KNOTT, S.A.; VISSCHER, P.M. On the use of linear regression and maximum likelihood for QTL mapping in half-sib designs. **Genetical Research**, Cambridge, v.72, p. 149-158, Jun. 1998.
- BARTON, N.H.; KEIGHTLEY, P.D. Understanding quantitative genetic variation. **Nature Reviews Genetics**, London, v.3, p.11-21, Jan.2002.
- BOX, G.E.P.; COX, D.R. An analysis of transformations. **Journal of the Royal Statistical Society**, London, n.26, v.B, p.211-252, May.1964.
- CARVALHO, W.A.; BECHARA, G.H.; MORE´, D.D.; FERREIRA, B.R.; SILVA, J.S.; SANTOS, I.K.F. *Rhipicephalus* (Boophilus) *microplus*: Distinct acute phase proteins vary during infestations according to the genetic composition of the bovine hosts, *Bos Taurus* and *Bos indicus*. **Experimental Parasitology**, Amsterdam, v.118, p.587-591, Oct. 2008.
- CHURCHILL, G.A.; DOERGE, R.W. Empirical threshold values for quantitative trait mapping. **Genetics**, London, v.138, p.963-971, Nov. 1994
- EVANS, D.E.; MARTINS, J.R.; GUGLIELMONE, A.A. A review of the ticks (Acari, Ixodida) of Brazil, their hosts and geographic distribution – 1. The state of Rio Grande do Sul, Southern Brazil. **Memórias do Instituto Oswaldo Cruz**, Rio de Janeiro, v.95, n.4, p.453-470, Aug.. 2000.
- FERNANDO, R.L.; GROSSMAN, M. marker assisted selection using Best linear unbiased prediction. **Genetics Selection Evolution**, Les Ulis, v.21, p.467-477, Aug.1968.
- FRISCH, J.E. Towards a permanent solution for controlling cattle ticks. **International Journal for Parasitology**, Melbourne, v.29, p.57-71, Dec. 1999.
- FRISCH, J.E.; O'NEILL, C.J.; KELLY, M.J. Using genetics to control cattle parasites – the Rockhampton experience. **International Journal for Parasitology**, Melbourne, v.30, p.253-264, Dec. 2000.
- GASPARIN, G. **Mapeamento de QTL para resistência a parasitas e características de crescimento nos cromossomos cinco e sete de uma população experimental F2 de bovinos Gir x Holandês**. 2007. 83p. Tese (Doutorado em Genética e Evolução) – Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, São Carlos, 2007.
- GASPARIN, G.; MIYATA, M. COUTINHO, L.L.; MARTINEZ, M.L.; SILVA, M.V.G.B.; MACHADO, M.A.; CAMPOS, A.L.; REGITANO, L.C.A. Quantitative trait locus affecting birth weight on bovine chromosome 5 in F2 Gyr x Holstein population. **Genetics and Molecular Biology**, Ribeirão Preto, v.28, p. 670-676, Dec. 2005.

GASPARIN, G.; MIYATA, M.; COUTINHO, L.L.; MARTINEZ, M.L.; TEODORO, R.L.; MACHADO, M.A.; SILVA, M.V.G.B.; SONSTERGARD, T.S.; REGITANO, L.C.A. Mapping of quantitative trait loci controlling (*Rhipicephalus* (*Boophilus*) *microplus*) resistance on bovine chromosomes 5, 7 and 14. **Animal Genetics**, Oxford, v.38, p.453-459, Oct. 2007.

GEORGES, M.; ANDERSSON, L. Positional identification of structural and regulatory quantitative trait nucleotides in domestic animal species. In: COLD SPRING HARBOR SYMPOSIA ON QUANTITATIVE BIOLOGY, 2003. Springfield. **Proceedings...** Springfield: Cold Spring Harbor Laboratory Press, 2003. p.68.

GIBSON, J.P.; BISHOP, S.C. Use of molecular markers to enhance resistance of livestock to disease: a global approach. **Revue scientifique et technique / Office International des Epizooties**, Paris, v.24, n.1, p.343-353, Jan. 2005.

GOLDGAR, D.E. Multipoint analysis of human quantitative genetic variation. **American Journal of Human Genetics**, Boston, v.4, n.6, p.957-967, Dec. 1990.

GREENE, W.H. **Accounting for excess zeros and sample selection in poisson and negative binomial regression models**. Department of Economics, University of New York, EUA. Mar.1994.p.1-36 (Technical Report,)

GUERRERO, F.D.; BENDELE, K.G.; CHEN, A.C.; LI, A.Y.; MILLER, R.J.; PLEASANCET, E.; VARHOL, R.; ROUSSEAU, M.E.; NENE, V.M. Serial analysis of gene expression in the southern cattle tick following acaricide treatment of larvae from organophosphate resistant and susceptible strains. **Insect Molecular Biology**, St Albans, v.16, n.1, p.49-60, Jul. 2007.

HALEY, C.H.; KNOTT, S.A.; ELSEN, J-M. Mapping quantitative trait loci in crosses between outbred lines using least squares. **Genetics**, Baltimore, v.136, p.1195-1207, Mar. 1994.

HASEMAN, J.K.; ELSTON, R.C. The investigation of linkage between a quantitative trait and a marker locus. **Behavior Genetics**, Netherlands, v.2, n.1, p.3-19, Mar.1972.

HOESCHELE, I.; UIMARI, P.; GRIGNOLA, F.E.; ZHANG, Q.; CAGE, K.M. Advances in statistical methods to map quantitative trait loci in outbred populations. **Genetics**, Baltimore, v.147, p.1445-1457, Nov. 1997.

HOOGSTRAL, H. Acarina (ticks), In: GIBBS, A.J. **Viruses and invertebrates**. Holland: North-Holland Publishing, 1973. cap. 5. p. 89-103.

HONER, M.R.; GOMES, A. **O manejo integrado de mosca dos chifres, berne e carrapatos em gado de corte**. Cidade:EMBRAPA-CNPGC.,1990.p. 1-14. (Circular Técnica, 22)

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=1269&id_pagina=1>. Acesso em :18 fev. 2009.

KASHINO, S.S.; RESENDE, J.; SACCO, A.M.S.; ROCHA, C.; PROENÇA, L.; CARVALHO, W.A.; FIRMINO, A.A.; QUEIROZ, R.; BENAVIDES, M.; GERSHWIN, L.J.; SANTOS, I.K.F.M. *Boophilus microplus*: The pattern of bovine immunoglobulin isotype responses to high and low tick infestations. **Experimental Parasitology**, Amsterdam, n.110, p.12-21, Jan. 2005.

KHATAR, M.S.; THOMSON, P.C.; TAMMEN, I.; RAADSMA, H.W. Quantitative trait loci in dairy cattle: review and meta-analysis. **Genetic Selection Evolution**, Les Ulis, v.36, n.2, p.163-190, Apr.2004.

KNOTT, S.A.; HALEY, C.S. Maximum likelihood mapping of quantitative trait loci using full-sib families. **Genetics**, Baltimore, v. 132, p.1211-1222, Dec. 1992.

KNOTT, S.A.; HALEY, C.S.; THOMPSON, R. Methods of segregation analysis for animal breeding data: parameter estimates. **Heredity**, London, v.68, p.313-320, Apr.1992.

LAMBERT, D. Zero-Inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, Alexandria, v.34, n.1, p.1-14, Feb.1992.

LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Baltimore, v.121, p.185-199, Jan. 1989.

LIU, W.; CELA, J. **Count data models in SAS®**. **SAS Global Forum**: Statistics and data analysis, Jun. 2008.p. 1-12. (paper 371)

LYNCH, M; WALSH, B. **Genetics and analysis of quantitative traits**. Massachusetts: Sinauer Associates, 1998. 980p.

MARTINEZ, M.L.; MACHADO, M.A.; NASCIMENTO, C.S.; SILVA, M.V.G.B.; TEODORO, R.L.; FURLONG, J.; PRATA, M.C.A.; CAMPOS, A.L.; GUIMARÃES, M.F.M.; AZEVEDO, A.L.S.; PIRES, M.F.A.; VERNEQUE, R.S. Association of BoLA-DRB3.2 alleles with tick (*Boophilus microplus*) resistance in cattle. **Genetics and Molecular Research**, Natal, v.5, n.3, p.513-524, Aug. 2006.

McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2nd ed. Londres: Chapman and Hall, 1989. 511p.

MULLAHY, J. Specification and testing of some modified count data models. **Journal of Econometrics**, North-Holland, v.33, p.341-365, Dec. 1986.

MIYATA, M.; GASPARIN, G.; COUTINHO, L.L.; MARTINEZ, M.L.; MACHADO, M.A.; SILVA, M.V.G.B.; CAMPOS, A.L.; SONSTERGARD, T.S.; ROSÁRIO, M.F.; REGITANO, L.C.A. QTL mapping for growth traits on bovine chromosome 14. **Genetics and Molecular Biology**. Riberão Preto, v.30, p. 364, Jan. 2007.

NELDER, J.A.; MEAD, R. A simplex method for function minimization. **The Computer Journal**, Oxford, v.7, n.4, p.308-313, Jul. 1965.

OLSEN, O.W. **Animal parasites**: their life cycles and ecology. 3rd .ed. Maryland: University Park Press, 1974. 562p.

PARDOE, I. A Bayesian sampling approach to regression model checking. **Journal of Computational and Graphical Statistics**, Alexandria, v.10, p.617-627, May, 2001.

PECONICK, A.P.; SOSSAI, S.; GIRÃO, F.A.; RODRIGUES, M.Q.R.B.; SOUZA E SILVA, C.H.; GUZMAN, F.; PATARROYO, A.M.; VARGAS, M.I.; PATARROYO, J.H. Synthetic vaccine (SBm 7462) against the cattle tick *Rhipicephalus (Boophilus) microplus*: Preservation of immunogenic determinants in different strains from South America. **Experimental Parasitology**, Amsterdam, v.119, p.37-43, Dec. 2008.

PEGRAM, R.G.; JAMES, A.D.; OOSTERMIJK, G.P.M.; KILLORN, K.J.; LEMCHE, J.; GHIROTTI, M.; TEKLE, Z.; CHIZYUKA, H.G.B.; MWASE, E.T.; CHIZHUKA, F. Studies on the economics of ticks in Zambia. **Experimental and Applied Acarology**, Netherlands, v.12, p.9-26, Sep.1991.

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

PIPER, E.K.; JACKSON, L.A.; BAGNALL, N.H.; KONGSUWAN, K.K.; LEW, A.E.; JONSSON, N.N. Gene expression in the skin of *Bos Taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus (Boophilus) microplus*. **Veterinary Immunology and Immunopathology**, doi:10.1016/j.vetimm.2008.06.11, 2008 (In Press).

RADOSTIS, O.M.; GAY, C.G.; BLOOD, D.C.; HINCHCLIFF, K.W. **Veterinary medicine**: A textbook of the diseases of cattle, sheep, pigs, goats and horses. 9th .ed. Londres: Saunders, 2000. 1877p.

ROBERTS, J.A. Acquisition by the host of resistance to the cattle tick, *Boophilus microplus* (Canestrini). **The Journal of Parasitology**, Melbourne, v.54, n.4, p.657-662, 1968.

RODRIGUES-MOTTA, M. **Zero-inflated poisson models for quantitative genetic analysis of count data with applications to mastitis in dairy cows**. 2006. 275p. Tese (PhD in Animal Sciences) – Department of Animal Sciences, University of Wisconsin, Madison, 2006.

SAS INSTITUTE. **The SAS system for windows**. SAS institute North Caroline, 2002 1 CD ROM.

SEIFERT, G.W. Variations between and within breeds of cattle in resistance to field infestations of the cattle tick (*Boophilus microplus*). **Australian Journal of Agriculture Research**, Collingwood, v.22, p.159-168, Jan.1971.

SOLLER, M. Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. **Journal of Dairy Science**, Savoy, v.73, p.2628-2846, Mar.1990.

SOLLER, M.; GENIZI, A. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. **Biometrics**, Oxford, v.34, p.47-55, Mar. 1978.

TU, W. Zero-inflated data. In: _____. **Encyclopedia of Envirometrics**. Chichester: John Wiley, 2002. 2800p.

TUNIN, K.P. **Mapeamento de QTL para características de resistência a carrapato, peso ao nascimento e peso a desmama no cromossomo 23 de bovinos da geração F2, cruzamento Holandês X Gir**. 2004. 48p. Dissertação (Mestrado em Ciência Animal e Pastagens) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2004.

WELLER, J.I. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. **Biometrics**, Oxford, v.42, p.627-640, Sep.1986.

YAMAMURA, K. Transformation using $(x + 0.5)$ to stabilize the variance of populations. **Research on Population Ecology**, Japan, v.41, p.229-234, Dec.1999.

YANG, R.; YI, N.; XU, S. Box-Cox transformations for QTL mapping. **Genetica**, Oxford, v.128, p.133-143, Sep.2006.

UTECH, K.B.W.; SEIFERT, G.W.; WHARTON, R.H. Breeding Australian illawarra shorthorn cattle for resistance to *Boophilus microplus*. I. Factors affecting resistance. **Australian Journal of Agriculture Research**, Collingwood, v.29, p.411-422, Sep.1978

APÊNDICES

APÊNDICE A – Código do SAS® utilizado para mapeamento de QTL pelo modelo Gaussiano.

Montar um arquivo de entrada contendo a identificação dos animais, as IBD para cada posição (cM) e a característica de interesse (y), seguindo o exemplo:

Animal	position	qtl	y
1	1	0.4889	2
2	1	-0.0323	1
3	1	-0.0003	0
4	1	0.9985	4

Neste exemplo, cada valor das probabilidades de IBD para os casos de heterozigose foram somados de acordo com o proposto por Haley & Knott (1992), dando origem ao valor chamado de “qtl” para cada animal a cada 1 cM.

Com este arquivo, utilizar o código abaixo:

```

PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=all2
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

ods listing close;
Proc GENMOD data=ALL2; by position;
model y = qtl / d=normal ;
  ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq Probchisq;
If parameter ne 'qtl' then delete;
run;

PROC EXPORT DATA= WORK.alfa1
              OUTFILE= "direção de onde se deseja criar o arquivo de
                        saída"
              DBMS=EXCEL2000 REPLACE;
              SHEET="Normal";
RUN;

```

Para o cálculo dos valores de limiar, utilizar o código abaixo e com o arquivo resultante proceder a uma análise da distribuição dos valores das permutações (y) através da ferramenta “Interactive Data Analysis”, obtendo-se o histograma e os quantis:

```

options nosource nonotes;
PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=threshold
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data resul;
input position Estimate Chisq ProbChiSq;
cards;
;
run;

data dados (keep = animal y); set threshold;
if position=1;
run;

%macro permutation;

%do k=1 %to 500;

* Permutation;

data permuta;
do index=1 to 263;
u=uniform(0); output;
end;
run;

data permuta; merge dados permuta;
run;

proc sort data=permuta;
by u;
run;

data permuta; merge dados (drop=y) permuta (keep=y);
run;

*****;

data ibd (keep = animal position qtl); set threshold;
run;

proc sort data=permuta;
by animal;
run;

```

```

proc sort data=ibd;
by animal;
run;

data final;
merge ibd permuta;
by animal;
run;

proc sort data=final;
by position animal;
run;

ods listing close;

Proc GENMOD data=final; by position;
model y = qtl / d=normal ;
ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq ProbChiSq;
If parameter ne 'qtl' then delete;
run;

proc sort data=alfa1;
by descending Chisq;
run;

data alfa1; set alfa1 (obs=1);
run;

data resul;
set resul alfa1;
run;

%end;

%mend;

%permutation;

```

APÊNDICE B – Código do SAS® utilizado para mapeamento de QTL pelo modelo Gaussiano com dados transformados pela raiz dos dados originais

Utilizando o mesmo arquivo de entrada do apêndice A, utilizar o código abaixo:

```

PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=all2
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

```

```

data all3; set all2;
yT= SQRT(y+0.375);
run;

ods listing close;
Proc GENMOD data=ALL3; by position;
model yT = qtl / d=normal ;
  ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq;
If parameter ne 'qtl' then delete;
run;

PROC EXPORT DATA= WORK.alfa1
              OUTFILE="direção de onde se deseja criar o arquivo de
                        saída"
              DBMS=EXCEL2000 REPLACE;
              SHEET="SQRT";
RUN;

```

Para o cálculo dos valores de limiar, utilizar o código abaixo e com o arquivo resultante proceder a uma análise da distribuição dos valores das permutações (y) através da ferramenta “Interactive Data Analysis”, obtendo-se o histograma e os quantis:

```

options nosource nonotes;
PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=threshold
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data resul;
input position Estimate Chisq ProbChiSq;
cards;
;
run;

data dados (keep = animal y); set threshold;
if position=1;
run;
%macro permutation;

%do k=1 %to 500;

* Permutation;

data permuta;
do index=1 to 263;
u=uniform(0); output;
end;

```



```

run;

data permuta; merge dados permuta;
run;

proc sort data=permuta;
by u;
run;

data permuta; merge dados (drop=y) permuta (keep=y);
run;

*****;

data ibd (keep = animal position qtl); set threshold;
run;

proc sort data=permuta;
by animal;
run;
proc sort data=ibd;
by animal;
run;

data final;
merge ibd permuta;
by animal;
run;

proc sort data=final;
by position animal;
run;

data all3; set all2;
yT= SQRT(y+0.375);
run;

ods listing close;

Proc GENMOD data=ALL3; by position;
model yT = qtl / d=normal ;
ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq;
If parameter ne 'qtl' then delete;
run;

proc sort data=alfa1;
by descending Chisq;
run;

data alfa1; set alfa1 (obs=1);
run;

data resul;

```

```

set resul alfa1;
run;

%end;

%mend;

%permutation;

```

APÊNDICE C – Código do SAS® utilizado para mapeamento de QTL pelo modelo Gaussiano com dados transformados por Box-Cox dos dados originais

Utilizando o mesmo arquivo de entrada do apêndice A, utilizar o código abaixo:

```

PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=all2
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data all3; set all2;
yT= y+0.375;
run;

proc sort data=all3; by position; run;
proc transreg data=all3 noprint; by position;
model BoxCox(yT / lambda=-2 to 2 by .05) = identity(qtl);
output out=saida;
run;
data saida; set saida; keep position TyT qtl;
run;

ods listing close;
Proc GENMOD data=saida; by position;
model TyT = qtl / d=normal ;
ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq;
If parameter ne 'qtl' then delete;
run;

PROC EXPORT DATA= WORK.alfa1
OUTFILE= "direção de onde se deseja criar o arquivo de
saída"
DBMS=EXCEL2000 REPLACE;
SHEET="Box-Cox";
RUN;

```

Para o cálculo dos valores de limiar, utilizar o código abaixo e com o arquivo resultante proceder a uma análise da distribuição dos valores das permutações (y) através da ferramenta “Interactive Data Analysis”, obtendo-se o histograma e os quantis:

```

options nosource nonotes;
PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=threshold
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data resul;
input position Estimate Chisq ProbChiSq;
cards;
;
run;

data dados (keep = animal y); set threshold;
if position=1;
run;

%macro permutation;

%do k=1 %to 500;

* Permutation;

data permuta;
do index=1 to 263;
u=uniform(0); output;
end;
run;

data permuta; merge dados permuta;
run;

proc sort data=permuta;
by u;
run;

data permuta; merge dados (drop=y) permuta (keep=y);
run;

*****;

data ibd (keep = animal position qtl); set threshold;
run;

proc sort data=permuta;
by animal;
run;
proc sort data=ibd;

```

```

by animal;
run;

data final;
merge ibd permuta;
by animal;
run;

proc sort data=final;
by position animal;
run;

data all3; set all2;
yT= y+0.375;
run;

proc sort data=all3; by position; run;
proc transreg data=all3 noprint; by position;
model BoxCox(yT / lambda=-2 to 2 by .05) = identity(qt1);
output out=saida;
run;
data saida; set saida; keep position TyT qt1;
run;

ods listing close;
Proc GENMOD data=saida; by position;
model TyT = qt1 / d=normal ;
ods output ParameterEstimates=alfa;
run;

proc sort data=alfa;
by descending Chisq;
run;

data alfa; set alfa (obs=1);
run;

data resul;
set resul alfa;
run;

%end;

%mend;

%permutation;

```

APÊNDICE D – Código do SAS[®] utilizado para mapeamento de QTL pelo modelo Poisson.

Utilizando o mesmo arquivo de entrada do apêndice A, utilizar o código abaixo:

```

PROC IMPORT
DATAFILE='direção do arquivo de entrada'

```

```

OUT=all2
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

ods listing close;
Proc GENMOD data=all2; by position;
model y = qtl / d=poisson ;
  ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq;
If parameter ne 'qtl' then delete;
run;

PROC EXPORT DATA= WORK.alfa1
              OUTFILE= "direção de onde se deseja criar o arquivo de
                        saída"
              DBMS=EXCEL2000 REPLACE;
              SHEET="Poisson";
RUN;

```

Para o cálculo dos valores de limiar, utilizar o código abaixo e com o arquivo resultante proceder a uma análise da distribuição dos valores das permutações (y) através da ferramenta “Interactive Data Analysis”, obtendo-se o histograma e os quantis:

```

options nosource nonotes;
PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=threshold
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data resul;
input position Estimate Chisq ProbChisq;
cards;
;
run;

data dados (keep = animal y); set threshold;
if position=1;
run;
%macro permutation;

%do k=1 %to 500;
* Permutation;
data permuta;
do index=1 to 263;
u=uniform(0); output;
end;
run;

```

```

data permuta; merge dados permuta;
run;

proc sort data=permuta;
by u;
run;

data permuta; merge dados (drop=y) permuta (keep=y);
run;
*****;

data ibd (keep = animal position qtl); set threshold;
run;

proc sort data=permuta;
by animal;
run;
proc sort data=ibd;
by animal;
run;

data final;
merge ibd permuta;
by animal;
run;

proc sort data=final;
by position animal;
run;

ods listing close;
Proc GENMOD data=all2; by position;
model y = qtl / d=poisson ;
ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position parameter estimate Chisq;
If parameter ne 'qtl' then delete;
run;

proc sort data=alfa1;
by descending Chisq;
run;

data alfa1; set alfa1 (obs=1);
run;

data resul;
set resul alfa1;
run;

%end;
%mend;
%permutation

```

APÊNDICE E – Código do SAS® utilizado para mapeamento de QTL pelo modelo ZIP.

Utilizando o mesmo arquivo de entrada do apêndice A, utilizar o código abaixo:

```

PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=all2
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

ods listing close;
Proc NLMIXED data= all2; by position;
parms pzero=0 m=0 alfa=0;

* Zeros verdadeiros;
eta0=pzero;
exp_eta0=exp(eta0);
p0=exp_eta0/(1+exp_eta0);

*Poisson;
etap=m+alfa*qtl;
exp_etap=exp(etap);

* Construindo a função de máxima verossimilhança do modelo ZIP;
if y=0 then ll= log(p0+(1-p0)*exp(-exp_etap));
    else ll= log(1-p0)+y*etap-exp_etap-lgamma(y+1);

model y ~ general (ll);
ods output ParameterEstimates=alfa;
run;

data alfa1; set alfa;
Keep position Estimate tvalue Probt tabs;
If Parameter ne "alfa" then delete;
run;

PROC EXPORT DATA= WORK.alfa1
    OUTFILE= "direção de onde se deseja criar o arquivo de
    saída"
    DBMS=EXCEL2000 REPLACE;
    SHEET="ZIP";

RUN;

```

Para o cálculo dos valores de limiar, utilizar o código abaixo e com o arquivo resultante proceder a uma análise da distribuição dos valores das permutações (y)

através da ferramenta “Interactive Data Analysis”, obtendo-se o histograma e os quantis:

```

options nosource nonotes;
PROC IMPORT
DATAFILE='direção do arquivo de entrada'
OUT=threshold
DBMS=TAB REPLACE;
GETNAMES=YES;
RUN;

data resul;
input position Estimate tValue Probt;
cards;
;
run;

data dados (keep = animal y); set threshold;
if position=1;
run;

%macro permutation;

%do k=1 %to 1000;

* Permutation;

data permuta;
do index=1 to 263;
u=uniform(0); output;
end;
run;

data permuta; merge dados permuta;
run;

proc sort data=permuta;
by u;
run;

data permuta; merge dados (drop=y) permuta (keep=y);
run;

*****;

data ibd (keep = animal position qtl); set threshold;
run;

proc sort data=permuta;
by animal;
run;
proc sort data=ibd;
by animal;
run;

```



```

data final;
merge ibd permuta;
by animal;
run;

proc sort data=final;
by position animal;
run;

ods listing close;

Proc NLMIXED data= final; by position;
parms pzero=0 m=0 alfa=0;

* True zeros;
eta0=pzero;
exp_eta0=exp(eta0);
p0=exp_eta0/(1+exp_eta0);

*Poisson;
etap=m+alfa*qt1;
exp_etap=exp(etap);

* Build the ZIP log likelihood;
if y=0 then ll= log(p0+(1-p0)*exp(-exp_etap));
    else ll= log(1-p0)+y*etap-exp_etap-lgamma(y+1);

model y ~ general (ll);
ods output ParameterEstimates=alfa ;
run;

data alfa1; set alfa;
Keep position Estimate tvalue Probt tabs;
If Parameter = "alfa";
tabs=abs(tvalue);
run;

proc sort data=alfa1;
by descending tabs;
run;

data alfa1 (drop=tabs); set alfa1 (obs=1);
run;

data resul;
set resul alfa1;
run;

%end;

%mend;

%permutation;

```