UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

RODRIGO VITORIO ALONSO

**Diagnóstico genético pré-implantacional de embriões bovinos
utilizando plataformas de genotipagem de marcadores moleculares
do tipo SNP**

**São Paulo
2013**

RODRIGO VITORIO ALONSO

Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP

Tese apresentada ao Programa de Pós-Graduação em Reprodução Animal da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Departamento:

Reprodução Animal

Área de Concentração:

Reprodução Animal

Orientador:

Prof. Dr. José Antônio Visintin

São Paulo

2013

Autorizo a reprodução parcial ou total desta obra, para fins acadêmicos, desde que citada a fonte.

DADOS INTERNACIONAIS DE CATALOGAÇÃO-NA-PUBLICAÇÃO

(Biblioteca Virginie Buff D'Ápice da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo)

T.2787
FMVZ

Alonso, Rodrigo Vitorio
Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP / Rodrigo Vitorio Alonso. -- 2013.
95 f. : il.

Tese (Doutorado) - Universidade de São Paulo. Faculdade de Medicina Veterinária e Zootecnia. Departamento de Reprodução Animal, São Paulo, 2013.

Programa de Pós-Graduação: Reprodução Animal.

Área de concentração: Reprodução Animal.

Orientador: Prof. Dr. José Antônio Visintin.

1. Biópsia. 2. Embrião. 3. Bovino. 4. Marcadores moleculares. 5. DNA.
I. Título.

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA VETERINÁRIA E ZOOTECNIA

Comissão de Ética no uso de animais

Of. CEUAVET nº 238/FMVZ/2011
lfrs

São Paulo, 24 de novembro de 2011.

Ao senhor
Prof. Dr. José Antonio Visintin

Assunto: Projeto nº 2465/2011, intitulado "Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP".

A Comissão de Ética no uso de Animais desta Faculdade, reunida no dia 23/11/2011, analisou o protocolo acima mencionado e solicita a Vossa Senhoria encaminhar a carta de doação por parte da empresa Deoxi Biotecnologia Ltda referente ao material biológico utilizado no experimento.

Atenciosamente,

Denise Tabacchi Fantoni
Presidente

FOLHA DE AVALIAÇÃO

Nome: ALONSO, Rodrigo Vitorio

Título: Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP

Tese apresentada ao Programa de Pós-Graduação em Reprodução Animal da Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Aprovado em: ____ / ____ / ____

Banca Examinadora

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

Prof. Dr.: _____

Instituição: _____ Julgamento: _____

DEDICATÓRIA

Dedico essa tese aos meus pais, Gilberto Pontes Alonso e Dulcineia Vitorio Alonso, que nunca mediram esforços para me oferecer o mais alto nível de educação intelectual e emocional, e que sempre me deram a segurança e o suporte necessários para progredir...

À minha esposa, Patrícia Suzuki, que passou a fazer parte da minha vida durante o desenvolvimento desse trabalho e que hoje enfrenta os desafios da vida ao meu lado...

Aos meus mestres, Wilson Machado de Souza, José Fernando Garcia, José Abdo de Andrade Hellú, Sílvia Helena Venturoli Perri, Amilcar Tanuri e José Antônio Visintin, que participaram de forma decisiva em diferentes etapas da minha vida acadêmica e profissional e são exemplos vivos de dedicação e competência.

AGRADECIMENTOS

A conclusão dessa importante etapa da minha vida só foi possível devido à extensa rede de colaboração profissional e de pessoas especiais que participaram e/ou participam da minha vida até hoje. É uma honra ter a oportunidade de registrar meus agradecimentos e minha consideração a cada uma das pessoas e instituições que, de forma direta ou indireta, colaboraram para a realização desse trabalho.

Não posso deixar de iniciar esse agradecimento a Deus, mesmo consciente de que, por muitas vezes, não atingi suas expectativas como filho. Mas Ele, na sua infinita misericórdia, sempre esteve me abençoando, guardando, livrando de todo mal e ouvindo minhas orações e súplicas. Obrigado Senhor.

Agradeço a minha esposa, Patrícia Suzuki, que acompanhou de perto grande parte do desenvolvimento desse trabalho, contribuindo tanto de forma direta (ditando planilhas, acompanhando viagens) como de forma indireta (pelas várias noites que não pude dormir ao seu lado devido às atividades realizadas em outras cidades).

Agradeço também aos meus pais, Gilberto e Dulcineia, que foram os pilares da minha educação e que me apoiaram nas diferentes fases da minha vida. A minha irmã Karine e ao meu cunhado Eduardo, que me deram dois grandes presentes, Rafaela e Daniel, meus queridos sobrinhos.

Agradeço ao Prof. Dr. José Antônio Visintin, que mesmo ocupando o cargo de Diretor da FMVZ/USP, aceitou ser meu orientador, depositando confiança no meu trabalho e na minha pessoa.

Agradeço o Prof. Dr. José Fernando Garcia, que atua na minha vida de forma direta há mais de 14 anos. Foi meu professor durante a graduação em Medicina Veterinária na UNESP-Araçatuba (1999), meu orientador de iniciação científica (2001-2003) e, desde então, nunca mais deixamos de trabalhar juntos. Exemplo de pró-atividade e competência, Fernando foi o grande responsável por direcionar minha vida profissional, com orientações sábias e apoio incondicional. Devo grande parte do que sou a esse homem, que muito mais do que professor, se tornou meu mestre e meu amigo.

Agradeço à minha sócia, Francine Campagnari, por enfrentarmos juntos os desafios diários da nossa empresa (Deoxi Biotecnologia) e por ter a certeza de que posso contar com ela a qualquer momento, para o que der e vier.

Agradeço ao Prof. Dr. Amilcar Tanuri, com quem aprendo diariamente o significado de sabedoria e humildade e por quem tenho grande apreço e admiração.

Agradeço à Prof. Dra. Gisele Zoccal Mingoti pela amizade e colaboração na produção *in vitro* de embriões, na Faculdade de Medicina Veterinária de Araçatuba.

Agradeço a toda equipe da Deoxi Biotecnologia (Valquíria, Karla, Erica, Adriana, Natália, Karol, Neto e Andrea) que de alguma maneira colaboraram em alguma etapa desse trabalho. Agradeço especialmente ao Altair Martins do Carmo (Fiotti), que há mais de 8 anos é meu fiel escudeiro nos trabalhos envolvendo micromanipulação embrionária e sexagem de embriões bovinos. Com sua habilidade e paciência inigualáveis, confesso que seria muito difícil concluir esse trabalho sem a preciosa ajuda do sempre companheiro Fiotti.

Agradeço aos colegas Adriana Santana do Carmo e Yuri Tani Utsunomyia pelas riquíssimas discussões e pelo auxílio nas genotipagens e nas análises estatísticas e de bioinformática.

Agradeço a toda equipe da In Vitro Brasil (Mogi Mirim/SP), pelo excelente trabalho de produção *in vitro* de embriões bovinos e pela confiança depositada no meu trabalho.

Agradeço à Prof. Dra. Caris Maroni Nunes e à toda equipe do Laboratório de Bioquímica e Biologia Molecular Animal (LBBMA), da Faculdade de Medicina Veterinária de Araçatuba (FMVA-UNESP) por cederem parte da estrutura laboratorial necessária para o desenvolvimento desse estudo.

Agradeço a todos os professores que ministraram as disciplinas cursadas durante o doutorado: Mário Binelli, Rubens Paes de Arruda, Anneliese de Souza Traldi, Mayra Elena Ortiz D'Avila Assumpção, Pietro Sampaio Baruselli, Guilherme de Paula Nogueira e Marcelo Vasconcelos Meireles.

Agradeço aos colegas Paulo, Guta, Zé Rodrigo, Everton, Júlia, Henderson, Robertinha, Zé Nélio, Kedson, Rodrigo, Thais, Francisco, Robinson, Manoel e Lilian, pelos ótimos momentos vividos durante as disciplinas cursadas no Campus de Pirassununga e no VRA em São Paulo. Agradeço especialmente ao colega e companheiro de viagem Rafael Cipriano, por gentilmente oferecer a casa de sua família para ficarmos hospedados durante as disciplinas cursadas em São Paulo.

Agradeço ao Dr. José Abdo de Andrade Hellú pelos anos em que trabalhamos juntos e pude aprender muito com seus ensinamentos de vida.

Agradeço à Yeda Watanabe e André Dayan (Vitrogen) que sempre me incentivaram a auxiliaram nos momentos de dificuldade.

Agradeço à Sociedade Brasileira de Tecnologia de Embriões (SBTE), entidade pela qual tenho profunda gratidão, pois proporcionou uma extensa rede de colaboração que impulsionou minha vida profissional.

Agradeço à Harumi, secretária do VRA, por todo apoio e eficiência em auxiliar em todas as questões burocráticas. Da mesma forma, agradeço às bibliotecárias da FMVZ-USP, Elza e Neusa, pela disponibilidade em auxiliar na normatização dessa tese.

Por fim, agradeço à FAPESP pelo auxílio financeiro.

EPÍGRAFE

“O temor do Senhor é o princípio da sabedoria, mas os insensatos desprezam o conhecimento e a instrução.”
Proverbios: 1:7

“...adquire pois a sabedoria, emprega tudo o que possuis na aquisição de entendimento.”
Proverbios: 4:7

RESUMO

ALONSO, R. V. **Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP**. [Preimplantation genetic diagnosis of bovine embryos using genotyping platforms of SNP molecular markers]. 2013. 95 f. Tese (Doutorado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2013.

Apesar do grande desenvolvimento das biotecnologias da reprodução animal, como a produção *in vitro* de embriões (PIV), o diagnóstico genético pré-implantacional (PGD) ainda é aplicado com restrição na rotina da transferência de embriões em animais. Os recentes avanços da genômica têm permitido associar características fenotípicas com informações moleculares, possibilitando a realização da seleção assistida por marcadores moleculares. O objetivo do presente estudo foi de realizar o PGD de embriões bovinos em plataforma de alta densidade de SNP (BeadChip/6.909 SNP). A pequena quantidade de DNA genômico (gDNA) obtida na biópsia embrionária é a principal limitação para a análise de grande número de SNP. Dessa forma, a metodologia de amplificação total do genoma (WGA) (Repli-g Mini Kit, Qiagen, Hilden, Alemanha) foi utilizada para aumentar a quantidade de gDNA da biópsia e permitir a análise de milhares de SNP simultaneamente. Foram utilizados 88 embriões bovinos PIV, submetidos à micromanipulação pela técnica de microaspiração, possibilitando a formação de 3 grupos com diferentes números de células biopsiadas: G1) 5-10 (n=28); G2) 10-20 (n=37); G3) >100 - blastocisto eclodido (n=23). Todas as amostras foram submetidas ao mesmo protocolo de WGA e 4µL de cada amostra foram utilizados para a genotipagem em plataforma iScan/Illumina. A qualidade dos genótipos foi avaliada pela análise do Call Rate (CR), 10º percentil do GCScore (GC10), Allele Drop In (ADI) e Allele Drop Out (ADO). O teste de Kruskal-Wallis foi utilizado para investigar diferenças na distribuição das variáveis entre os grupos. O coeficiente de correlação de postos de Spearman revelou correlação significativa entre todas as variáveis. Os resultados apresentaram correlação positiva entre o CR e GC10 (0,99/P<0,001), enquanto as taxas de ADI e ADO apresentaram correlação negativa com o CR e GC10 (ADI/CR: -0,87; ADI/GC10:-0,88; ADO/CR:-0,87; ADO/GC10:-0,86), com P<0,001 para todas as variáveis. O teste de Kruskal-Wallis apontou para diferença significativa em todas as variáveis analisadas (CR, GC10, ADI e ADO) entre os 3 grupos de biópsias (G1,

G2 e G3). O CR médio foi de 59,26%, 78,47% e 95,97%, para G1, G2 e G3, respectivamente. Foi desenvolvido programa computacional (mendelFix) baseado na “Lei da Segregação”, para inferência dos genótipos não determinados baseado no genótipo do progenitores, aumentando o CR dos grupos 1, 2 e 3 para 79,69%, 88,20% e 97,28%, respectivamente. Os resultados do presente estudo permitem concluir que é possível realizar a genotipagem de embriões bovinos em plataforma de alta densidade de SNP com amostras submetidas ao protocolo WGA/MDA, mas o número de células obtidas na biópsia embrionária afeta a qualidade da genotipagem. A associação das biotecnologias descritas no presente trabalho permite a aplicação da seleção assistida por marcadores moleculares em embriões bovinos, contribuindo para acelerar ainda mais o processo de melhoramento genético animal.

Palavras-chave: WGA. Biópsia. Embrião. Bovino. Marcadores moleculares. DNA.

ABSTRACT

ALONSO, R. V. **Preimplantation genetic diagnosis of bovine embryos using genotyping platforms of SNP molecular markers.** [Diagnóstico genético pré-implantacional de embriões bovinos utilizando plataformas de genotipagem de marcadores moleculares do tipo SNP]. 2013. 95 f. Tese (Doutorado em Ciências) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2013.

Despite the great development of animal reproduction biotechnology, such as embryo *in vitro* production (IVP), preimplantation genetic diagnosis (PGD) is still applied with restraint in animals embryo transfer. Recent advances in genomics have associated phenotypic characteristics with molecular information, allowing the development of marker assisted selection. The aim of this study was to perform PGD in bovine embryos using high-throughput SNP platform (BeadChip/6,909 SNP). The small amount of genomic DNA (gDNA) obtained from embryo biopsy is the main limitation for the high-density SNP analysis. Thus, the Whole Genome Amplification (WGA) (Repli-g Mini Kit, Qiagen, Hilden, Germany) was used to increase the amount of gDNA from embryo biopsy and allow the analysis of thousands SNP simultaneously. Eighty-eight IVP bovine embryos were subjected to micromanipulation by microaspiration, allowing the formation of three groups with different numbers of biopsied cells: G1) 5-10 (n=28); G2) 10-20 (n=37); G3) > 100 - hatched blastocyst (n = 23). All samples were subjected to the same WGA protocol, and 4µL of each sample were used for genotyping on iScan/Illumina platform. The genotyping quality was assessed using the Call Rate (CR), 10th percentile GC Score (GC10), Allele Drop In (ADI) and Allele Drop Out (ADO). Kruskal-Wallis test was applied to investigate differences in the distribution of variables among groups. Spearman's rank correlation coefficient revealed a significant correlation between all variables. The results showed a positive correlation between CR and GC10 (0.99/P <0.001), while ADI and ADO rates were negatively correlated with CR and GC10 (ADI/CR: -0.87; ADI/GC10: - 0.88; ADO/CR: -0.87; ADO/GC10: -0.86), P<0.001 for all variables. Kruskal Wallis pointed to significant differences in all variables (CR, GC10, ADO and ADI) among the 3 groups of biopsies (G1, G2 and G3). The CR average was 59.26%, 78.47% and 95.97% for G1, G2 and G3, respectively. Was developed a script (mendelFix) based on the "Law of Segregation", for inference of not determined genotypes based on the parents genotypes, thus increasing the CR

of group 1, 2 and 3 to 79.69%, 88.20% and 97.28%, respectively. The results of this study show that it is possible to perform the genotyping of bovine embryos in high-throughput SNP platform with samples subjected to WGA/MDA protocol, but the number of cells obtained by embryo biopsy affects the quality of genotyping. The association of biotechnologies described in this work allows the application of marker-assisted selection in bovine embryo, contributing to further accelerate the process of animal breeding.

Keywords: WGA. Biopsy. Embryo. Cattle. Molecular markers. DNA

SUMÁRIO

1	INTRODUÇÃO.....	16
2	REVISÃO DE LITERATURA.....	19
2.1	O BRASIL E O AGRONEGÓCIO.....	19
2.2	O BRASIL E A REPRODUÇÃO ANIMAL.....	20
2.3	MELHORAMENTO GENÉTICO: CLÁSSICO E MOLECULAR.....	20
2.4	MARCADORES MOLECULARES.....	21
2.5	MICROMANIPULAÇÃO EMBRIONÁRIA.....	23
2.6	AMPLIFICAÇÃO TOTAL DO GENOMA.....	25
2.6.1	Primer Extension Preamplification (PEP).....	26
2.6.2	Metodologia Omniplex.....	29
2.6.3	Amplificação por Deslocamento Múltiplo (MDA).....	30
2.6.4	WGA com pequenas quantidades de gDNA.....	32
2.6.5	WGA no diagnóstico genético pré-implantacional.....	33
3	HIPÓTESE E OBJETIVOS.....	38
3.1	HIPÓTESE.....	38
3.2	OBJETIVOS.....	38
4	MATERIAL E MÉTODOS.....	40
4.1	PURIFICAÇÃO, QUANTIFICAÇÃO E DILUIÇÃO DE gDNA.....	40
4.2	PRODUÇÃO <i>IN VITRO</i> DE EMBRIÕES BOVINOS.....	40
4.3	MICROMANIPULAÇÃO EMBRIONÁRIA.....	41
4.4	AMPLIFICAÇÃO TOTAL DO GENOMA (WGA).....	44
4.5	PLATAFORMA DE GENOTIPAGEM – ISCAN/ILLUMINA.....	45
5	ANÁLISES ESTATÍSTICAS.....	49
5.1	DEFINIÇÃO DAS VARIÁVEIS DE QUALIDADE DA GENOTIPAGEM.....	49
5.1.1	GenCall Score e Call Rate.....	49
5.1.2	Allele Drop In (ADI) e Allele Drop Out (ADO).....	51
5.2	ESTATÍSTICA DESCRITIVA.....	52
5.3	CORRELAÇÃO ENTRE AS VARIÁVEIS.....	52

5.4	AVALIAÇÃO DA QUALIDADE DA GENOTIPAGEM COM DIFERENTES QUANTIDADES DE CÉLULAS OBTIDAS NA BIÓPSIA EMBRIONÁRIA.....	53
5.5	INFERÊNCIA DE GENÓTIPOS NÃO DETERMINADOS NAS BIÓPSIAS EMBRIONÁRIAS.....	54
6	RESULTADOS	56
6.1	AVALIAÇÃO DA METODOLOGIA WGA/MDA COM QUANTIDADES DECRESCENTES DE gDNA.....	56
6.2	AVALIAÇÃO DA GENOTIPAGEM DE BIÓPSIAS EMBRIONÁRIAS..	59
7	DISCUSSÃO	67
8	CONCLUSÕES	76
	REFERÊNCIAS	77
	APÊNDICES	85

1 INTRODUÇÃO

Atualmente o Brasil ocupa posição de destaque internacional na indústria da carne, pois é detentor do maior rebanho bovino comercial do mundo com aproximadamente 213 milhões de cabeças. A bovinocultura de corte representa a maior fatia do agronegócio brasileiro, gerando faturamento de mais de R\$50 bilhões/ano e oferecendo cerca de 7,5 milhões de empregos, colocando o país como um dos maiores exportadores de carne bovina nos últimos anos (ABIEC, 2011).

O avanço da pecuária nacional é fruto do trabalho de renomados grupos de pesquisa nas áreas do melhoramento genético e das biotecnologias da reprodução animal, com destaque para a produção *in vitro* de embriões bovinos. Em 2011, o Brasil foi responsável pela produção de 86% dos embriões bovinos produzidos *in vitro* no mundo, o equivalente a 318.116 embriões (STROUD, 2012).

Com a finalização e publicação do genoma bovino em 2009, novas tecnologias pós-genômicas tem sido desenvolvidas. Atualmente existem plataformas de genotipagem capazes de avaliar de centenas a milhares de marcadores moleculares simultaneamente. Essas tecnologias permitem a associação de informações moleculares com características produtivas, possibilitando a interação do melhoramento genético clássico com o melhoramento genético molecular, pelas estratégias de seleção assistida por marcadores moleculares e/ou seleção genômica.

Uma das principais vantagens da seleção genômica é a obtenção de informações zootécnicas logo após o nascimento dos animais, proporcionando relevante redução de custo e de tempo no processo de melhoramento genético. Dessa forma, a associação de técnicas reprodutivas com as recentes metodologias genômicas abrem perspectivas para acelerar ainda mais o processo de seleção animal, pela realização das análises genéticas ainda em fase embrionária.

A maior limitação para a realização do diagnóstico genético pré-implantacional utilizando plataformas de genotipagem de alta densidade de SNP é a pequena quantidade de DNA genômico (gDNA) obtida na biópsia embrionária. Entretanto, diferentes metodologias de amplificação total do genoma têm sido

desenvolvidas com o objetivo de aumentar a quantidade de gDNA a partir de poucas células obtidas na biópsia embrionária, viabilizando a realização de análises genéticas múltiplas por diferentes técnicas de biologia molecular.

O objetivo do presente estudo foi avaliar as metodologias de amplificação por deslocamento múltiplo (MDA) e de genotipagem em plataforma de alta densidade de SNP, utilizando embriões bovinos produzidos *in vitro* e submetidos à biópsia pela técnica de microaspiração de diferentes quantidades de células.

Apesar do grande desenvolvimento e aplicação das biotecnologias da reprodução animal, como a produção de embriões *in vitro*, as técnicas de diagnóstico genético pré-implantacional ainda são pouco desenvolvidas e aplicadas com restrição aos programas de transferência de embriões. Paralelamente, os rápidos avanços na área da genômica estão cada vez mais próximos da cadeia produtiva, permitindo a identificação e seleção dos animais geneticamente superiores cada vez mais cedo e com maior acurácia. Dessa forma, os resultados aqui apresentados poderão contribuir de forma significativa para o progresso científico e tecnológico, associando as biotecnologias da reprodução animal com as recentes metodologias genômicas, com reais perspectivas de sua aplicação a campo.

2 REVISÃO DE LITERATURA

2.1 O BRASIL E O AGRONEGÓCIO

De acordo com o relatório da projeção do agronegócio brasileiro de 2010/2011 a 2020/2021, desenvolvido pelo Ministério da Agricultura, Pecuária e Abastecimento (BRASIL, 2011), o mundo vive um período de grande apreensão, em particular após a eclosão da crise financeira e recessão da economia mundial ocorrida em 2008. Segundo os autores, durante a preparação do relatório intensificaram-se as discussões sobre os preços agrícolas, não somente como um dos itens pressionando os índices de inflação interna, como também uma preocupação internacional com a elevação dos preços dos alimentos nos últimos anos.

O relatório do Departamento de Agricultura dos Estados Unidos (USDA) relata que os principais fatores apontados para a elevação dos preços dos produtos agrícolas são a demanda mundial crescente para grãos, oleaginosas e produtos da pecuária; a desvalorização do dólar americano; a manutenção dos preços de energia em alta; e o crescimento da produção de biocombustíveis, ressaltando atenção especial para a demanda de óleos vegetais e de carnes (USDA, 2008). De acordo com a previsão da Organização das Nações Unidas para Alimentação e Agricultura (FAO, 2011), será necessário duplicar a produção de carne até o ano 2050 para suprir o aumento na demanda de alimentos provocada pelo grande crescimento populacional dos próximos 40 anos.

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o rebanho bovino brasileiro chegou a 212,8 milhões de cabeças, em 2011, representando um aumento de 1,6% em relação a 2010 (PRODUÇÃO DA PECUÁRIA MUNICIPAL, 2011). Conforme os dados publicados pela Associação Brasileira das Indústrias Exportadoras de Carne (ABIEC, 2011), a bovinocultura de corte representa a maior fatia do agronegócio brasileiro, gerando faturamento de mais de R\$ 50 bilhões/ano e oferecendo cerca de 7,5 milhões de empregos.

2.2 O BRASIL E A REPRODUÇÃO ANIMAL

Nos últimos anos, o Brasil tem se destacado como líder mundial na produção *in vitro* (PIV) de embriões bovinos, sendo responsável por 86% da produção mundial, o equivalente a 318.116 embriões PIV em 2011 (STROUD, 2012). Esse destaque não se restringe à PIV, mas engloba diversas tecnologias da reprodução animal, como a criopreservação e sexagem de espermatozoides (CARVALHO et al., 2010), inseminação artificial em tempo fixo (NASSER et al., 2011), técnicas de micromanipulação de gametas e embriões, que possibilitam o desenvolvimento de biotecnologias como a clonagem (ASSIS NETO et al., 2011; SMITH et al., 2011), transgenia e o diagnóstico genético pré-implantacional, como a sexagem de embriões (ALONSO et al., 2008), entre outras. O progressivo e constante desenvolvimento dessas tecnologias tem provocado um grande impacto positivo na pecuária nacional com relação às biotecnologias relacionadas ao melhoramento genético animal.

2.3 MELHORAMENTO GENÉTICO: CLÁSSICO E MOLECULAR

O melhoramento animal, apesar de ter a fundamentação teórica desenvolvida há muitos anos, tem recebido grandes contribuições tecnológicas, que são as principais responsáveis pelos progressos genéticos que têm sido observados nas mais diferentes espécies de animais domésticos exploradas comercialmente (KEPLER, 1999). Impulsionada pela globalização e pela eficiência de produção em todos os setores, as pesquisas em genética animal no Brasil e no mundo tem avançado exponencialmente com novas tecnologias e metodologias (LÔBO et al., 2011).

Atualmente, o Brasil conta com diferentes programas de melhoramento genético de bovinos, que há mais de duas décadas desenvolvem trabalho fundamental para o aumento da produtividade da pecuária nacional e são os grandes responsáveis pelo sucesso obtido nesta área.

Mais recentemente, para assistir a seleção animal, tanto dentro como fora da esfera reprodutiva, informações de marcadores de DNA têm sido propostas, desenvolvidas e aplicadas na produção de bovinos, graças à associação de dados genético-moleculares com fenótipos de interesse medidos pela seleção clássica (VIGNAL et al., 2002; VAN EENENNAAM, 2005). Esta abordagem é conhecida como seleção assistida por marcadores moleculares (do inglês: Marker Assisted Selection – MAS), consistindo-se na genotipagem de marcadores sabidamente em desequilíbrio de ligação (do inglês: Linkage Disequilibrium – LD) com genes relacionados a características de interesse, aliando esta informação com a seleção tradicional (HAYES, 2007).

Conforme Dekkers (2004), nos últimos anos, os avanços na área da genética molecular proporcionaram a identificação de múltiplos genes ou marcadores genéticos associados a diversas características de interesse nos animais de produção, tais como precocidade sexual, ganho de peso, qualidade de carne, produção e qualidade de leite, resistência a parasitas, entre outras. Essas informações genéticas têm sido utilizadas em programas de seleção assistida por marcadores.

2.4 MARCADORES MOLECULARES

Historicamente, os marcadores de DNA empregados na MAS foram os microssatélites, pequenas sequências de 1 a 7 nucleotídeos que se repetem em tandem e limitados na ordem de até 100 repetições (CURI; LOPES, 2001), tendo sua genotipagem baseada na Reação em Cadeia pela Polimerase (do inglês: Polymerase Chain Reaction – PCR) (OKADA; TAMATE, 2000).

Os microssatélites foram intensamente explorados na prospecção de *loci* de características quantitativas (do inglês: Quantitative Trait *Locus* – QTL) em gado de corte, o que resultou na identificação de vários QTLs ao longo dos diferentes cromossomos bovinos (WU et al., 2005; MIZOGUCHI et al., 2006; IMAI et al., 2007; TAKASUGA et al., 2007; ABE et al., 2008; GUTIÉRREZ-GIL et al., 2008; JIANG et al., 2009). Entretanto, apesar da especulação comercial e do desenvolvimento de testes diagnósticos, a MAS não atingiu a significância aspirada pela comunidade

científica por apresentar baixo ganho genético e dificuldades técnicas inerentes à natureza bioquímica dos microssatélites.

Com o início da intitulada era Pós-Genômica, marcada pela conclusão do projeto genoma bovino (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009), esforços para identificar polimorfismos de sítio único (do inglês: Single Nucleotide Polimorphism – SNP) ganharam cena. Marcos neste cenário foram os trabalhos de Van Tassell et al. (2008) e do consórcio HapMap (THE BOVINE HAPMAP CONSORTIUM, 2009), os quais contribuíram com a identificação de aproximadamente 23 mil e 34 mil SNP, respectivamente.

Os SNP são variações bialélicas em locais específicos de um genoma, definidas por apenas um nucleotídeo (BROOKES, 1999). Estes marcadores não estão uniformemente distribuídos ao longo do genoma, e ocorrem com menor frequência em regiões codificantes do que em regiões não codificantes (LI; SADLER, 1991).

Com o progresso da identificação de SNP, novas abordagens para genotipagem alcançaram popularidade. A última geração de tecnologias faz uso de microarranjos baseados em oligonucleotídeos fixados em lâminas de vidro, chamados “SNP Chips”. Cada “chip” é composto por dezenas (ou até centenas) de milhares de pontos microscópicos contendo sondas de DNA específicas. A hibridização de uma sequência alvo com a sonda é detectada e quantificada por fluorescência ou quimioluminescência. Estas tecnologias permitem a análise de centenas de milhares de SNP em um único ensaio, reduzindo os custos de geração de dados (CAETANO, 2009).

Nessa linha, inúmeras empresas transnacionais iniciaram no mercado oferecendo serviços dentro do campo da geração de dados SNP com metodologias de baixo custo, de execução rápida e de alta automatização. A empresa Illumina Inc. desenvolveu, em colaboração com diversos centros de pesquisa internacionais, microarranjos contendo de 3 mil a ~800 mil SNP distribuídos ao longo do genoma bovino (ILLUMINA, 2013). Com uma cobertura genômica de alta densidade, foi possível por em prática uma nova abordagem: o estudo de associação por todo o genoma (do inglês: Genome-wide Association Study - GWAS).

De acordo com Lôbo (2011) o progresso genético pode ser maximizado se os melhores animais forem identificados precocemente e tiverem sua genética propagada massivamente. Uma das principais vantagens do uso de marcadores

moleculares em programas de seleção é relacionada à precocidade de avaliação dos animais para características específicas, uma vez que esta tecnologia emprega amostras de DNA, permitindo análises desde momentos imediatamente após o nascimento, ou até mesmo durante a fase embrionária pré-implantação, podendo ser inclusive incorporada a programas de produção *in vitro* e transferência de embriões, acelerando os sistemas de seleção e melhoramento genético animal (GARCIA, 2001; GARCIA; PORTO-NETO, 2006).

2.5 MICROMANIPULAÇÃO EMBRIONÁRIA

O diagnóstico genético pré-implantacional (do inglês: Preimplantation Genetic Diagnosis - PGD) é uma biotecnologia utilizada para determinar características genéticas de embriões em fases precoces de desenvolvimento (CENARIU et al., 2012). Em humanos, essa tecnologia permite a identificação e prevenção da transmissão de defeitos genéticos, bem como o aumento nas taxas de concepção de pacientes com problemas reprodutivos (ZHENG et al., 2011). Em animais, as tecnologias de PGD são utilizadas tanto para a identificação e prevenção de doenças genéticas como para a seleção de embriões com características desejáveis, sendo que a identificação do sexo de embriões foi sua primeira aplicação nos animais domésticos (GARCIA, 2001).

A primeira e fundamental etapa para realização do PGD é a realização de uma técnica de micromanipulação embrionária, que permita a obtenção de células embrionárias sem prejudicar a viabilidade do embrião após a biópsia (PONSART et al., 2012). Em humanos, a maioria dos trabalhos descreve a realização da biópsia no estágio de 8-16 blastômeros, com a retirada de um único blastômero para realização das análises genéticas (ALMODIN et al., 1999). Entretanto, a maior parte dos trabalhos com embriões bovinos descreve a realização da biópsia em estádios mais avançados do desenvolvimento, compreendendo desde mórula até blastocisto expandido.

Diferentes técnicas de micromanipulação podem ser utilizadas para a remoção de células embrionárias, visando à obtenção de DNA para realização de análises genéticas. Ponsart et al. (2012) descreveram as duas principais técnicas

utilizadas na micromanipulação de embriões bovinos: 1) Microsecção: realizada pelo corte da zona pelúcida e separação das células com o auxílio de uma microlâmina; 2) Microaspiração: realizada pela perfuração da zona pelúcida e aspiração das células com o auxílio de 2 micropipetas (Holding e Biópsia).

Cenariu et al. (2012) compararam a taxa de concepção de embriões bovinos produzidos *in vivo* e biopsiados com as diferentes técnicas, com resultado de 43% com a técnica de microaspiração e de 31% com a microsecção.

Alonso (2008) avaliou os fatores que afetam a viabilidade de embriões bovinos produzidos *in vitro*, submetidos à biópsia pela técnica de microaspiração. Foram analisados 4.650 embriões PIV, sendo que os principais fatores envolvidos com a taxa de mortalidade embrionária foram a qualidade do embrião antes da realização da biópsia e a quantidade de material aspirado na biópsia. Os embriões de qualidade 1, 2 e 3, apresentaram taxa de mortalidade de 7,9%, 21,7% e 48,9%, respectivamente. O autor definiu como “biópsia padrão” a aspiração de 5-15% da massa celular embrionária viável e “biópsia fora de padrão” a aspiração de mais do que 15%, sendo que as taxas de mortalidade para os diferentes tipos de biópsia foram de 9,9% e 27,4%, respectivamente.

Lauri et al. (2013) avaliaram a viabilidade dos embriões submetidos à diferentes tamanhos de biópsias por microsecção, sendo que os embriões submetidos à biópsia de 10 células apresentaram taxa de mortalidade de 11%, enquanto que nos de 30 células a taxa aumentou para 38%.

Embora alguns autores descrevam o número de células biopsiadas de embriões bovinos no estágio de mórula e blastocisto, a contagem exata de células obtidas pela técnica de microaspiração de embriões bovinos PIV não pode ser realizada precisamente sem a utilização de técnicas de coloração. Entretanto, a quantidade de células pode ser estimada com base na relação do número médio de células com o estágio embrionário e pela proporção da massa embrionária viável obtida na biópsia.

De acordo com Soom et al. (1997), o número médio de células de embriões bovinos produzidos *in vivo* aumenta de acordo com o desenvolvimento embrionário (Estágio/nº de células): Mórula/46,4 $\pm$ 19,1; Blastocisto inicial/105,0 $\pm$ 23,4; Blastocisto/115,0 $\pm$ 5,6; Blastocisto expandido/159,0 $\pm$ 14,1; Blastocisto eclodido/172,4 $\pm$ 28,8.

Xu et al. (1992) descreveram o número de células de embriões bovinos produzidos *in vitro* de acordo com o desenvolvimento embrionário (Estádio/nº de células): Mórula/ $45,4 \pm 1,17$; Blastocisto/ $101,5 \pm 3,05$; Blastocisto expandido/ $174,7 \pm 5,87$; Blastocisto eclodido/ $195,9 \pm 7,72$. Os dados apresentados por Mori et al. (2002) com relação ao número de células de embriões bovinos PIV são de $84,6 \pm 30,5$ células para blastocistos e de $119,5 \pm 41,4$ para blastocistos expandidos.

2.6 AMPLIFICAÇÃO TOTAL DO GENOMA

Apesar da possibilidade apresentada por Garcia (2001), a principal limitação para implementação das tecnologias de MAS em programas de transferência de embriões é a pequena quantidade de DNA genômico obtida no procedimento de biópsia embrionária, devido ao reduzido número de células que se pode retirar do embrião sem prejudicar o seu desenvolvimento (POLISSENI et al., 2010), restringindo a avaliação de múltiplos testes no diagnóstico genético pré-implantacional (SPITS et al., 2006).

Neste contexto, diversos trabalhos científicos têm apontado para as técnicas de amplificação total do genoma (do inglês: "Whole Genome Amplification" – WGA) como um método revolucionário no PGD tanto na área animal como humana (BARKER et al., 2004; HANDYSIDE et al., 2004; HOLBROOK et al., 2005; SPITS et al., 2006; ALONSO et al., 2007; COSKUN; ALSMAD, 2007; ALONSO et al., 2008; HUMBLLOT et al., 2010; LE BOURTHIS, et al., 2010, 2011; POLISSENI et al., 2010; CHEN et al., 2011; TREFF et al., 2011; PONSART et al., 2012; ZHENG; WANG; JIN, 2011; HAN et al., 2012; LAURI et al., 2013; SARGOLZAEI et al., 2013).

A amplificação total do genoma (WGA) é uma técnica utilizada para aumentar a quantidade de gDNA de amostras com quantidade limitada de DNA original (COSKUN; ALSMADI, 2007), possibilitando a obtenção de microgramas de DNA a partir de amostras com baixa quantidade de gDNA inicial ou mesmo a partir de uma única célula, com aproximadamente 5-6pg de gDNA (ZHENG; WANG; JIN, 2011). Nos últimos anos, diversas metodologias de WGA têm sido descritas por diferentes autores e com diferentes aplicações.

2.6.1 Primer extension preamplification (PEP)

Uma das primeiras metodologias de WGA foi descrita por Zhang (1992) que utilizou o protocolo denominado “Primer Extension Preamplification” (PEP) para análises haplotípicas de uma única célula espermática. O PEP envolve ciclos da Reação em Cadeia pela Polymerase (PCR) e o uso da Taq Polymerase com oligonucleotídeos randômicos de 15 bases. Entretanto, diferentemente da PCR padrão, após a desnaturação, a temperatura de anelamento dos primers é mantida a 37°C, possibilitando o anelamento inespecífico dos primers ao molde de DNA, seguida pela elevação gradual da temperatura na taxa de 1°C a cada 10 segundos até 55°C, e finalmente, manutenção a 55°C por 4 minutos para a polimerização das fitas de DNA (ZHANG, 1992). De acordo com Paunio et al. (1996) essa metodologia permite a amplificação de 96% do genoma, com cobertura de no mínimo 1000 vezes.

A técnica de PEP-PCR foi utilizada por diversos autores que demonstraram a possibilidade de avaliações genéticas múltiplas a partir de poucas células.

Kristjansson et al. (1994) utilizaram linfoblastos isolados de pacientes com diagnóstico de distrofia muscular de Duchenne, causada por mutações conhecidas nos exons 4, 8, 12, 45, e 48. Estes linfoblastos foram separados individualmente e submetidos à reação de PEP, sendo o produto subdividido em seis alíquotas, possibilitando a realização de PCR específico para cada exon separadamente. Foram analisados os 5 exons para a distrofia e os *loci* ZFX/ZFY. O mesmo procedimento foi efetuado com amniócitos, células das vilosidades coriônicas e blastômeros, obtendo-se 93% de exatidão no diagnóstico, indicando a capacidade desta técnica para a realização de diagnósticos genéticos pré-implantacionais precisos sem comprometimento do embrião, e para mais de um locus simultaneamente.

Schaaff et al. (1996) utilizaram 102 amniócitos, efetuando sequencialmente PEP, PCR e Nested PCR, com oligonucleotídeos homólogos aos cromossomos X e Y do gene da amelogenina (AMGX) e do pseudogene amelogenina (AMGY), respectivamente, que envolvem a deleção de 177 pares de bases (pb) no gene AMGY, e amplificaram o *locus* onde ocorre a mutação da fibrose cística ($\Delta F508$). Foram alcançados 93% de sensibilidade e 100% de exatidão na determinação do

sexo e 95% de eficácia do diagnóstico da mutação da fibrose cística, indicando a importância desta técnica para a determinação do sexo em doenças recessivas ligadas ao cromossomo X.

Gibbons et al. (1995) buscaram alternativas para melhorar a eficiência do PGD em humanos a partir de uma única célula. Segundo os autores, muitos processos são ineficientes devido à diferença na amplificação entre os diferentes genes, como por exemplo: 1) região determinante do sexo no cromossomo Y, 85% (77/90); 2) doença de Tay-Sachs, 51% (97/187); e 3) fibrose cística, 85% (87/107), por fatores, como a estrutura secundária do gene que poderia antagonizar a ligação do oligonucleotídeo ou a extensão da polimerase, ou por sítios de mutação que podem estar presentes nas regiões a serem amplificadas. Para contornar problemas gerados pela obtenção de resultados incorretos por falhas na amplificação, os autores incluíram um passo de reamplificação pela técnica de PEP nas amostras não diagnosticadas, obtendo desta maneira maior número de amostras com diagnóstico correto.

Casas e Kirkpatrick (1996), a fim de melhorar a eficácia na amplificação do gDNA a partir de uma única célula, utilizaram diferentes protocolos de amplificação por PEP, empregando diferentes polimerases (*Taq* DNA polimerase, *Pfu* polimerase e fragmento Stoffel), observando que o uso associado da *Taq* DNA polimerase e da *Pfu* polimerase na relação de 5U/ 0,3125U, respectivamente, melhorou a eficiência da amplificação do DNA genômico.

Xu et al. (1993), na tentativa de otimizar a reação de PEP, fizeram modificações a partir do procedimento convencional descrito por ZHANG et al. (1992), diminuindo o tempo (de 8h para 4h e 40min.); o número de ciclos (de 50 para 40); e o volume final (de 60µl para 40µl), encontrando eficiência semelhante ao procedimento convencional.

Hochman et al. (1995) afirmam que o DNA contido em uma biópsia embrionária pode ser usado para um número limitado de *loci* pela Multiplex PCR e que diversas replicações são necessárias para minimizar erros randômicos que ocorrem na PCR. Para aumentar o número de possíveis análises genéticas, utilizaram o PEP em combinação com o PCR a partir de um quarto de embrião bovino, demonstrando que podem ser efetuadas múltiplas análises genéticas para *loci* marcadores como *kappa* caseína, deficiência da adesão leucocitária bovina (BLAD) e microssatélite D9S1, e para *loci* determinante do sexo como BRY 1,

Bov97M e ZFX/ZFY, concluindo que a técnica de PEP permite a análise de aproximadamente 40 regiões do genoma por quarto de embrião e o uso combinado dos *loci* Bov97M e ZFX/ZFY oferecem alta segurança na rotina de sexagem em embriões bovinos.

De acordo com Zapata (1998) um fator limitante para o PGD pela técnica de PCR é a possibilidade de diagnóstico de somente uma característica a partir da biópsia embrionária, uma vez que se consome todo o material retirado do embrião em cada teste. O autor afirma ainda que a técnica de PEP foi desenvolvida com o objetivo de permitir a múltipla detecção de diferentes *loci* a partir de uma única célula ou biópsia embrionária de até, potencialmente, uma célula. Em seu trabalho foram utilizadas três regiões previamente descritas do genoma do camundongo (SRY, HPRT e DNXds3) com o intuito de testar a viabilidade da pré-amplificação seguida de amplificações específicas, em várias concentrações de DNA purificado de leucócitos. Os resultados permitiram ao autor afirmar que a técnica de PEP-PCR aumenta em torno de 10 vezes a sensibilidade do diagnóstico em comparação à técnica de PCR isoladamente e que apesar de apresentar porcentagens de não amplificação variáveis, a técnica de PEP-PCR é uma boa alternativa ao PCR isoladamente, para a realização do diagnóstico genético pré-implantacional de diferentes *loci* gênicos simultaneamente, sendo esta técnica uma ferramenta nova e poderosa nesse campo, e que seu desenvolvimento e padronização deverão proporcionar inúmeros avanços na aplicação das biotécnicas da reprodução tanto em humanos quanto em animais.

Apesar das diversas aplicações demonstradas, a metodologia de PEP apresenta algumas desvantagens, como a pequena quantidade de produto amplificado que varia de 10^2 a 10^4 (PAUNIO et al., 1996) e a produção de fragmentos menores do que 1.500 pares de bases (WELLS et al., 1999), o que tem limitado a sua aplicação no diagnóstico genético pré-implantacional.

Trabalhos mais recentes têm demonstrado metodologias alternativas para a amplificação total do genoma a partir de baixas concentrações de gDNA (BARKER et al., 2004), poucas células (HANDYSIDE et al., 2004), ou até mesmo uma única célula (TREFF et al., 2011).

2.6.2 Metodologia OmniPlex

De acordo com Barker et al. (2004), a metodologia OmniPlex pode ser utilizada com o objetivo de aumentar a quantidade de DNA em amostras com baixa concentração de gDNA. O GenomePlex (Sigma-Aldrich, St. Lois, MO, USA) é um método alternativo de WGA baseado na fragmentação randômica do gDNA molde e na conversão dos fragmentos em biblioteca de DNA denominada OmniPlex, cujos fragmentos possuem alta representatividade do genoma. Posteriormente, as bibliotecas OmniPlex são amplificadas com a utilização de primers universais, possibilitando a amplificação do molde de DNA milhares de vezes, com a ação de DNA Polimerase de alta fidelidade, produzindo fragmentos de 100pb a 2kb, com média de ~450pb. Dependendo da quantidade inicial de gDNA podem ser produzidas quantidades finais de 5-10µg de DNA, em menos de 4 horas de amplificação. A figura 1 esquematiza a reação de amplificação com a metodologia GenomePlex (Sigma-Aldrich).

Figura 1 - Diagrama esquemático das etapas de amplificação pela metodologia OmniPlex: Fragmentação do gDNA molde; anelamento dos primers nas extremidades dos fragmentos gerados; extensão e formação da biblioteca OmniPlex, representando todo o genoma do gDNA molde; amplificação exponencial da biblioteca OmniPlex, com alta quantidade e representatividade do genoma

Fonte: www.sigmaaldrich.com

2.6.3 Amplificação por Deslocamento Múltiplo (MDA)

A metodologia de Amplificação por Deslocamento Múltiplo (do inglês: Multiple Displacement Amplification – MDA) foi descrita por Dean et al. (2002) e é baseada na amplificação isotérmica que ocorre de 30 a 31°C com o bacteriófago Phi29 DNA Polimerase e com primers randômicos que se ligam na amostra de gDNA inicial. Essa enzima tem a capacidade de adicionar em média 70.000 nucleotídeos no sentido 3'-5', gerando fragmentos entre 2 e 100kb (média de ~10kb), resultando em amplificação com maior fidelidade do que os métodos que utilizam a Taq Polimerase. A Phi29 DNA Polimerase possui forte atividade de deslocamento, propiciando a amplificação exponencial por mecanismos de ramificação da fita de

gDNA, resultando em alta quantidade de DNA amplificado (HAN et al., 2012). A quantidade de DNA final varia de 10-40µg de DNA, dependendo da quantidade de gDNA inicial, com o tempo de amplificação de 8 a 16 horas. Atualmente existem dois Kits disponíveis comercialmente que utilizam a metodologia MDA: 1) GenomiPhi DNA Amplification Kit (GE Healthcare, Piscataway, NJ, EUA) e 2) Repli-g Kit (Qiagen, Hilden, Alemanha). A figura 2 esquematiza a reação de amplificação pela metodologia MDA, com a Phi29 DNA Polimerase (MAMONE, 2003).

Figura 2 - Diagrama esquemático das etapas de Amplificação por Deslocamento Múltiplo (MDA): Anelamento dos primers randômicos aos múltiplos sítios do gDNA molde desnaturado; a Phi29 DNA polimerase inicia o processo de polimerização e extensão simultânea nos múltiplos sítios do gDNA molde; o processo de síntese, deslocamento e complementaridade do DNA gera uma nova fita simples de DNA; a subsequente ligação de um novo primer e a replicação por deslocamento da fita simples resulta na formação de uma fita de DNA complementar

Fonte: Mamone (2003).

2.6.4 WGA com pequenas quantidades de gDNA

Barker et al. (2004) utilizaram dois métodos de WGA: 1) Amplificação por Deslocamento Múltiplo (MDA) e 2) Tecnologia OmniPlex, utilizando 10ng e 100ng de gDNA inicial para as respectivas metodologias, atingindo quantidade final de 75-79µg com a metodologia MDA e de 25-75 µg com a Omniplex. Todas as amostras foram genotipadas utilizando o “Illumina linkage panel” versão 1, contendo 2.320 SNP distribuídos no genoma. Os genótipos obtidos com as amostras de WGA apresentaram Call Rate acima de 98% e a comparação com genótipos obtidos de gDNA (sem amplificação) apresentou concordância acima de 99,8%, para as duas metodologias de WGA analisadas.

Mais recentemente, Han et al. (2012) compararam a uniformidade de amplificação pela metodologia MDA, utilizando arranjo comparativo de hibridização genômica (do inglês: “array-based comparative genomic hybridization” – aCGH) para avaliar a variação do número de cópias (do inglês: “Copy Number Variations” – CNVs) de amostras submetidas ao protocolo de WGA com os Kits Repli-g (Qiagen) e GenomePhi (GE Healthcare), utilizando 20ng de gDNA inicial. Foi utilizado o ensaio “Agilent Human CGH” contendo aproximadamente um milhão de probes, em um paciente normal e dois com fibrose cística. Os resultados com ambos os métodos mostraram alta uniformidade e consistência no processo de WGA quando comparados com gDNA não amplificado, apresentando baixas taxas de erro na replicação e sugerindo que a metodologia WGA/MDA pode ser utilizada para viabilizar a genotipagem de amostras com baixa concentração de gDNA.

Apesar dos bons resultados apresentados por Barker et al. (2004) e Han et al. (2012), ambos utilizaram quantidade mínima de 10ng como gDNA inicial para os protocolos de WGA. Entretanto, a quantidade de gDNA disponível para o diagnóstico genético pré-implantacional é limitada a pequena quantidade de células obtidas pela micromanipulação embrionária sem prejudicar a viabilidade do embrião (CENARIU et al., 2012).

2.6.5 WGA no diagnóstico genético pré-implantacional

De acordo com Handyside et al. (2004) a metodologia de amplificação total do genoma marcou uma nova era para o diagnóstico genético pré-implantacional, pois possibilita a realização de análises genéticas múltiplas a partir de uma única célula. Os autores utilizaram a metodologia MDA para amplificar o gDNA de um único linfócito, comparando com grupos de 2, 5, 10 e 20 linfócitos, seguido pela análise de 20 diferentes sequencias específicas de DNA. Embora o resultado com uma única célula tenha apresentado falhas de amplificação, resultados mais consistentes foram obtidos com 2-5 células. Os grupos com 10 e 20 células apresentaram resultado idêntico ao controle positivo (gDNA), concluindo que a metodologia de WGA/MDA foi eficiente para aumentar a quantidade de gDNA da amostra inicial e possibilitar a realização de diferentes análises genéticas pela técnica de PCR convencional.

Alonso et al. (2007) avaliaram a metodologia OmniPlex, utilizando o Kit GenomePlex Single Cell (Sigma-Aldrich, St. Lois, MO, USA) para amplificação de gDNA obtido a partir de biópsia de 28 embriões bovinos produzidos *in vitro*. A concentração do produto após a amplificação foi avaliada por espectrofotometria (NanoDrop) revelando quantidade média de $5,7 \pm 1,9\mu\text{g}$ de DNA.

Mais recentemente, Chen et al. (2011) relataram bons resultados utilizando a metodologia MDA, com o GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Piscataway, NJ, USA), para realização do diagnóstico genético pré-implantacional da neurofibromatose tipo 1 em embriões humanos.

Atualmente, diferentes grupos têm desenvolvido metodologias de análise de CNVs, utilizando aCGH, para diagnóstico de aneuploidias a partir de uma única célula embrionária humana. Entretanto, uma célula humana possui aproximadamente de 6-7pg de gDNA e os microarranjos exigem quantidades em níveis de nanogramas para realização adequada da genotipagem, sendo necessária a amplificação do gDNA inicial em mais de 1.000 vezes. De acordo com Treff et al. (2011), as metodologias de WGA com uma única célula ainda são uma limitação para a realização do diagnóstico genético pré-implantacional utilizando plataformas de alta densidade de SNP.

O objetivo do trabalho de Treff et al. (2011) foi de realizar a comparação direta de três Kits comerciais de WGA (Repli-g, GenomePhi e GenomePlex) a partir

de uma única célula embrionária humana, analisando a qualidade e acurácia da amplificação em plataforma de genotipagem de 250K SNP (Affymetrix) e pela análise de CNVs. Os autores utilizaram células fibroblásticas humanas de 4 linhagens diferentes para o número de cópias do cromossomo X, comparando o resultado da genotipagem do produto de WGA de 1 célula com gDNA extraído de grande quantidade de células de cada linhagem. A eficiência da amplificação foi avaliada pela quantificação das amostras por espectrofotometria (NanoDrop) e eletroforese, considerando como bem sucedidas aquelas com quantidade acima de 250ng. Três amostras de cada linhagem celular foram submetidas ao protocolo de WGA com cada um dos Kits testados, sendo posteriormente utilizados 250ng do WGA e 250ng do gDNA para a genotipagem. A análise do Call Rate (CR) foi baseada no número de SNP identificados adequadamente, dividido pelo número total de SNP presentes no Chip (250K). A acurácia da genotipagem foi avaliada pelo percentual de genótipos concordantes entre as amostras de WGA comparadas com o padrão ouro (gDNA) de cada grupo. As taxas de “Allele Drop Out” (ADO) foram definidas pelos SNP que foram identificados como homozigotos nas amostras de WGA e heterozigotos nas de gDNA dentro da mesma linhagem celular. Os mesmo dados foram utilizados para as análises de variação do número de cópias, possibilitando a comparação da amplificação dos 3 diferentes Kits, com suas respectivas amostras de gDNA e com o cariótipo já conhecido de cada linhagem celular.

Os resultados apresentados por Treff et al. (2011) demonstraram similaridade do produto amplificado com os Kits Repli-g e o GenomiPhi, com fragmentos de ~10kb para ambos, enquanto que o produto obtido com o GenomePlex foi de 100 a 1000pb. O controle negativo obtido com o Repli-g apresentou banda similar ao material amplificado com uma célula, demonstrando que a eletroforese em gel de agarose não é uma boa opção para a avaliação da qualidade de amplificação com essa metodologia. A média de produto de WGA obtida com os diferentes Kits foi de $0,57 \pm 0,1\mu\text{g}$, $7,63 \pm 0,4\mu\text{g}$ e $18,93 \pm 0,8\mu\text{g}$, para os Kits GenomiPhi, GenomePlex e Repli-g, respectivamente. A média da quantificação observada no controle negativo com o Repli-g foi de $15,62 \pm 2,3\mu\text{g}$, revelando que a quantificação por espectrofotometria também não é um método adequado para avaliar a eficiência da amplificação desse produto. O resultado médio do CR na genotipagem de 250K SNP foi de 74%, 78% e 88%, para o GenomiPhi, GenomePlex e Repli-g,

respectivamente. A taxa de acurácia na genotipagem foi de 86% para o GenomePhi, 89% para o GenomePlex e 96% para o Repli-g, sendo esse último o Kit de WGA mais indicado para genotipagem de SNP a partir de uma única célula.

Lauri et al. (2013) avaliaram a eficiência da metodologia MDA para genotipagem de marcadores do tipo microsatélites e SNP (Bovine SNP50 Genotyping BeadChip – Illumina) utilizando biópsias de embriões bovinos produzidos pela técnica de clonagem. Os autores avaliaram a correlação entre a quantidade de gDNA inicial e a quantidade de células obtida na biópsia embrionária com o sucesso da metodologia de WGA/MDA realizada com o Repli-g Mini Kit (Qiagen). Os experimentos foram realizados com biópsias de 10 células (16 replicatas), 30 células (16 replicatas) e 100-2000 células (12 replicatas) obtidas por cultivo *in vitro* de células embrionárias e com biópsias de 10 células de embriões bovinos produzidos por fertilização *in vitro* (PIV) com genótipo desconhecido (16 replicatas). A qualidade e a eficiência da amplificação foram avaliadas pela análise do Call Rate (CR), Allele Drop Out (ADO) e Allele Drop In (ADI). O CR foi definido pelo número de SNP identificados em relação ao total de SNP presentes no Bovine SNP50 Genotyping BeadChip. O ADI foi calculado com base nos *loci* homozigotos dos pais em comparação aos embriões ($AA \times AA = AA$), enquanto o ADO foi calculado na expectativa dos *loci* heterozigotos dos embriões em relação aos pais ($AA \times BB = AB$). O grupo de biópsias de 10 células apresentou CR médio de 72,45%, com variação de 29,30% a 93,73%. A variação diminuiu no grupo de biópsias de 30 células, com média de 92,54%, mínima e máxima de 73,46% e 96,78%, respectivamente. A maior taxa de CR foi obtida com o grupo de 100-2000 células, com média de 96,04%, mínima de 85,75% e máxima de 98,20%. O resultado do grupo de embriões PIV foi estatisticamente igual ao grupo de 10 células de embriões clones, com média de 81,42% de CR, mas apresentando grande variação entre as amostras, sendo o mínimo de 21,36% e o máximo de 95,45%. O percentual de ADO foi inversamente proporcional ao número de células de cada grupo, apresentando médias de 12,60%, 1,41% e 0,48%, para os grupos de 10, 30 e 100-2000 células, respectivamente. O mesmo foi observado para a taxa de ADI, com médias de 2,87%, 0,33% e 0,07% para a mesma sequência dos grupos. Assim como na análise do CR, o grupo dos embriões PIV teve o mesmo resultado estatístico do grupo com 10 células de embriões clones, com média de ADO de 13,75% e de ADI de 3,34%.

Sargolzaei et al. (2013) realizaram a genotipagem a partir de biópsias de embriões bovinos, utilizando a metodologia de WGA seguida pela análise em plataforma de alta densidade de SNP (BovineSNP50 BeadChip – Illumina), obtendo 96% de sucesso na genotipagem (681/709). Os genótipos foram submetidos à análise em programa computacional desenvolvido especificamente para realização de imputação para correção dos Allele Drop Out. Parte dos embriões foi transferida, produzindo o nascimento de 56 bezerros até o momento, possibilitando a comparação do genótipo obtido na fase embrionária com o genótipo do bezerro nascido. A média percentual de inconsistências entre o genótipo do embrião e do bezerro foi de 1,6%, enquanto que a média percentual de perda de SNP foi de 11,4%. Entretanto, após a realização da imputação esses valores foram corrigidos para 0,94% e 0,01%, respectivamente. Foi realizado o cálculo do valor genômico dos animais nascidos, obtendo-se correlação entre o valor genômico dos bezerros e dos respectivos embriões de 0,975 antes da imputação e de 0,991 depois da imputação, indicando que a Seleção Genômica pode ser aplicada com eficiência em fase embrionária.

Os recentes avanços referentes às tecnologias de reprodução assistida (produção de embriões *in vitro*), acopladas às técnicas de diagnóstico genético pré-implantacional (micromanipulação embrionária e WGA) e à seleção assistida por marcadores moleculares, constituirão importantes ferramentas para que, em futuro próximo, o processo de seleção genética seja aplicado em fases embrionárias precoces. De acordo com Chrenek et al. (2001), a caracterização genética de embriões é o caminho mais rápido para obtenção de gerações com o genótipo desejado, possibilitando a redução entre intervalos de gerações e a diminuição de custos nos processos de melhoramento genético.

3 HIPÓTESE E OBJETIVOS

3.1 HIPÓTESE

A qualidade da genotipagem em plataforma de alta densidade de SNP independe da quantidade de células obtidas na biópsia de embriões bovinos produzidos *in vitro* e submetidas à amplificação por deslocamento múltiplo.

3.2 OBJETIVOS

- Avaliar a metodologia de amplificação por deslocamento múltiplo (WGA/MDA) a partir de quantidades decrescentes de gDNA e de diferentes quantidades de células obtidas pela biópsia de embriões bovinos produzidos *in vitro*.
- Realizar a genotipagem das amostras submetidas à metodologia de WGA/MDA em plataforma de alta densidade de SNP.
- Avaliar a eficiência e qualidade da genotipagem pela análise do Call Rate, GenCall Score, Allele Drop In e Allele Drop Out.

4 MATERIAL E MÉTODOS

4.1 PURIFICAÇÃO, QUANTIFICAÇÃO E DILUIÇÃO DE gDNA

As 8 doadoras e 3 touros utilizados no presente projeto tiveram amostra de sangue periférico e/ou de bulbos capilares coletados para obtenção de DNA genômico. As amostras de sangue foram submetidas ao protocolo de extração e purificação de gDNA utilizando o Kit MiniPrep (Axygen Bioscience), enquanto que as amostras de bulbos capilares foram processadas com o Kit NucleoSpin Tissue (Macherey-Nagel), seguindo as orientações dos fabricantes.

Todas as amostras foram quantificadas por espectrofotometria (NanoDrop 2000) para avaliação da quantidade (ng/μL) e da qualidade do gDNA, sendo utilizadas nas análises somente amostras com absorbância (razão 260/280) entre 1,7 e 2,0. A concentração das amostras foi ajustada para 50ng/μL, sendo que 3 delas (2 machos e 1 fêmea) foram submetidas à diluição seriada, nas concentrações de 5ng/μL, 500pg/μL, 100pg/μL, 50pg/μL e 5pg/μL.

4.2 PRODUÇÃO *IN VITRO* DE EMBRIÕES BOVINOS

Nesse estudo foram utilizados 88 embriões bovinos produzidos *in vitro*, a partir da aspiração folicular transvaginal guiada por ultrassonografia de 8 doadoras (F1 a F8), acasaladas com 3 diferentes touros (M1 a M3). A aspiração folicular foi realizada a campo, seguida pela seleção e classificação dos complexos *cumulus*-oócitos (COCs), que foram lavados em TCM-199 (GIBCO BRL; Grand Island, NY, EUA) tamponado com Hepes, suplementado com 10% de soro fetal bovino (SFB) (GIBCO BRL; Grand Island, NY, EUA), 0,20 mM de piruvato de sódio e 83,4 μg/mL de amicacina (Instituto Biochimico, Rio de Janeiro, Brasil). Os COCs foram transportados em incubadora portátil até o laboratório, onde foram mantidos durante 24h em incubadora (38,5°C com 5% CO₂ em ar e umidade máxima), sob óleo mineral em gotas de 100 μL de meio TCM-199 suplementado com 10% de SFB, um

$\mu\text{g/mL}$ FSH (FolltropinTM, Bioniche Animal Health, Belleville, Ont., Canadá), 50 $\mu\text{g/mL}$ hCG (ProfasiTM, Serono, São Paulo, Brasil) e estradiol (1 $\mu\text{g/mL}$), 0,20mM de piruvato de sódio e 83,4 $\mu\text{g/mL}$ de amicacina.

A fertilização *in vitro* (FIV) foi realizada com sêmen convencional de 3 touros, descongelado (35°C por 30 seg) e lavado duas vezes por centrifugação (600 rpm por 5 min) em 2 mL de meio TALP suplementado com 0,2 mM de piruvato e 83,4 g/mL de amicacina, tamponado com 10 mM de Hepes. A concentração do sêmen foi ajustada para 25×10^6 espermatozóides (sptz) móveis/mL. O volume de 4 μL de sêmen (10^5 sptz) foi adicionado em cada gota de 90 μL de TALP-FIV (TALP suplementado com 10 g/mL de heparina, 18 M de penicilamina, 10 M de hipotaurina e 8 M de epinefrina) sob óleo mineral. Posteriormente, foram adicionados 25 a 30 oócitos em cada gota, seguido por incubação de 20-24h em incubadora a 38,5°C com 5% CO₂ em ar e umidade máxima.

Os prováveis zigotos foram co-cultivados (grupos de 25) na incubadora (38,5°C com 5% CO₂ em ar e umidade máxima) com células da granulosa em gotas de 100 μL de SOF (WELLS et al., 1999), 2,5% SFB + 0,5% de albumina sérica bovina (BSA) sob óleo mineral. No dia três (D3) e no dia cinco (D5) do cultivo *in vitro* (CIV) foi realizada a substituição de 50% do volume de meio das gotas por meio novo (“feeding”). O meio de cultivo do “feeding” foi o mesmo meio do início do cultivo. A taxa de clivagem foi avaliada no terceiro dia de cultivo (D3) e a previsão da conversão oócitos/embriões foi avaliada no sexto dia de cultivo (D6), quando se iniciou o processo de micromanipulação embrionária.

4.3 MICROMANIPULAÇÃO EMBRIONÁRIA

A técnica de micromanipulação utilizada foi a de microaspiração de células viáveis de embriões bovinos produzidos *in vitro*, utilizando dois micromanipuladores mecânicos e dois microinjetores (Narishige) acoplados a microscópio invertido (Olympus, modelo IX51-III). Cada embrião foi devidamente avaliado de acordo com o estágio de desenvolvimento: Mórula (MO), Blastocisto inicial (BI), Blastocisto (BL), Blastocisto expandido (BX), Blastocisto em eclosão (BN) e Blastocisto eclodido (BE);

e com o grau de qualidade (1, 2 ou 3) de acordo com os padrões estabelecidos pela IETS. Os embriões foram transferidos individualmente para gotas identificadas de 100µL de meio HSOF para realização da micromanipulação.

As micropipetas utilizadas na micromanipulação embrionária (Holding e Biópsia) foram fabricadas utilizando capilares de borossilicado (Word Precision Instrument) e os equipamentos Puller (Campden Instrument Limited, Moddel 773), esmerilhador (Narishige) e microforja (Narishige). A Holding foi produzida com diâmetro externo de ~180µm e diâmetro interno de ~35µm, enquanto que a pipeta de biópsia foi produzida com diâmetro externo entre 20-25µm. Nos embriões em que foi possível identificar a massa celular interna (BL e BX), a biópsia foi realizada preferencialmente pela aspiração de células do trofoblasto, evitando a manipulação do embrioblasto (Figura 3).

As biópsias foram classificadas de acordo com a quantidade e qualidade de material aspirado. Devido à impossibilidade de contar com exatidão o número de células obtidas em cada biópsia, o critério utilizado nesse experimento foi baseado na proporção de material aspirado em relação à quantidade total de células embrionárias viáveis, bem como no estágio e qualidade embrionária. Dessa forma, todos os embriões foram submetidos à microaspiração de aproximadamente 10% da massa embrionária viável, possibilitando a classificação do número de células obtidas na biópsia de acordo com o estágio embrionário. As biópsias de mórulas e blastocistos iniciais, que possuem entre 50 e 100 células, foram classificadas como Grupo 1 (5 a 10 células) e blastocistos e blastocistos expandidos, que possuem entre 100 e 200 células, foram classificadas como Grupo 2 (10 a 20 células).

Figura 3 - Sequencia de micromanipulação de blastocisto bovino produzido *in vitro*, pela técnica de microaspiração

Legenda: A = Posicionamento da massa celular interna no sentido oposto à biópsia; B = Perfuração da zona pelúcida e aspiração de células do trofoblasto; C = Material aspirado na biópsia (10% da massa celular embrionária viável); D = Embrião após a biópsia. Fonte: Alonso (2013).

As células biopsiadas foram transferidas para tubo de polipropileno (0,2mL), centrifugadas e congeladas a -20°C , até o momento do início do protocolo de WGA. Após a biópsia, os embriões foram cultivados individualmente em gotas identificadas por no mínimo 12 horas, permitindo a avaliação da viabilidade embrionária pelo seu desenvolvimento e taxa de eclosão. A figura 4 mostra grupo de embriões com 12 horas de cultivo após a realização da biópsia.

Os embriões eclodidos foram transferidos individualmente para tubos de polipropileno (0,2mL), permitindo a formação do Grupo 3 do experimento, representando amostra com mais de 100 células embrionárias.

Figura 4 - Embriões bovinos produzidos *in vitro* com 12 horas de cultivo após a realização da biópsia por microaspiração

Fonte: Alonso (2013).

4.4 AMPLIFICAÇÃO TOTAL DO GENOMA (WGA)

A técnica de amplificação total do genoma (WGA) foi utilizada com o objetivo de aumentar a quantidade de gDNA obtido na biópsia embrionária, viabilizando a realização da genotipagem em plataforma de alta densidade de marcadores do tipo SNP.

A metodologia utilizada foi a de Amplificação por Deslocamento Múltiplo (MDA), utilizando o Repli-g Midi Kit (Qiagen, Hilden, Alemanha) seguindo o protocolo descrito por Handyside et al. (2004). As biópsias embrionárias foram transferidas para tubo de polipropileno de 0,2mL contendo 2 μ L de tampão PBS, totalizando volume aproximado de 3,5 μ L. Foi adicionado 3,5 μ L do tampão D2, seguido por 10 minutos de incubação em gelo e 5 minutos a 65°C em termociclador (MJ Research

PTC 100). Foram adicionados 3,5µL da solução “Stop Solution” para finalização do procedimento de lise celular. Logo em seguida foram adicionados 40µL do “Master Mix”, constituído por 10µL de água ultra pura, 29µL do “Reaction Buffer” e 1µL de DNA polimerase (Phi29). A reação de amplificação isotérmica ocorreu durante 16 horas a 30°C, seguida por 3 minutos a 65°C para inativação da Phi29 DNA polimerase.

O produto de WGA foi avaliado por eletroforese em gel de agarose (1%) corado com brometo de etídio, para avaliação de amplificação de fragmento de aproximadamente 10kb e as alíquotas foram congeladas a -20°C até o momento da realização das análises genéticas.

4.5 PLATAFORMA DE GENOTIPAGEM – ISCAN/ILLUMINA

A avaliação da qualidade do material obtido pela metodologia de WGA/MDA foi realizada pela genotipagem de marcadores moleculares do tipo SNP, utilizando o ensaio Infinium em plataforma iScan/Illumina, com volume padrão de 4µL de cada amostra.

A plataforma iScan utiliza a tecnologia “BeadArray” e oferece flexibilidade para genotipar de centenas a milhares de marcadores do tipo SNP por amostra de DNA. A tecnologia “BeadArray” é baseada na presença de beads de ~3µM, alocados em micropoços existentes numa lâmina de sílica. Cada bead possui de centenas a milhares de oligonucleotídeos responsáveis por capturar as sequencias específicas de cada SNP presente no microarranjo.

Esta plataforma apresenta alta sensibilidade e excelente limite de detecção, gerando dados de alta qualidade, caracterizados por resultados próximos de 99% de “Call Rate” e de 99,9% de reprodutibilidade. Nesse estudo, foi utilizado produto específico para a espécie bovina denominado “BovineLD BeadChip”, com capacidade de avaliação de 6.909 SNP (24 amostras/Chip), como mostrado na figura 5.

Figura 5 - Imagem do BeadChip BovineLD e representação esquemática da tecnologia BeadArray

Fonte: www.illumina.com

O protocolo de análise do perfil de SNP foi realizado como o preconizado pelo fabricante e está disponível em detalhes no endereço eletrônico www.illumina.com. Em suma, o protocolo em questão consistiu de seis etapas principais distintas:

1) Amplificação: o DNA das amostras foi desnaturado, neutralizado e submetido a um processo de amplificação isotérmico. Essa etapa objetivou aumentar a concentração de DNA centenas de vezes para que isso não fosse um fator limitante no processo, reduzindo a qualidade dos resultados.

2) Fragmentação: o produto de amplificação foi fragmentado por um processo enzimático do tipo *end-point*, isto é, os sítios de fragmentação são determinados por atributos específicos da sequência de DNA, o que evita a ocorrência de fragmentação excessiva.

3) Hibridização: o DNA amplificado e fragmentado se anela (hibridiza) à sequência de 50 bases que está covalentemente ligada à nano partícula *bead*, sendo que cada *bead* corresponde a um locus específico do DNA.

4) Extensão de base-única: ocorre a extensão de uma única base nucleotídica da sequência, que irá incorporar os corantes que permitirão a determinação dos genótipos.

5) Coloração: coloração da lâmina (micro-arranjo). Diversos corantes foram incorporados aos *beads* sendo responsáveis pela emissão de fluorescência.

6) Visualização das Imagens: Um conjunto de laser foi utilizado para excitar os fluoróforos que estavam ligados aos *beads* na lâmina, permitindo que o sinal de fluorescência fosse captado e analisado pelo iScan (Figura 6).

Figura 6 – Plataforma de genotipagem de alta densidade de SNP – iScan/Illumina

5 ANÁLISES ESTATÍSTICAS

5.1 DEFINIÇÃO DAS VARIÁVEIS DE QUALIDADE DA GENOTIPAGEM

5.1.1 GenCall Score e Call Rate

O GenCall Score (GC) é uma medida relativa que indica o grau de confiabilidade do genótipo atribuído a um dado SNP para um determinado indivíduo. Este valor é calculado com base na dispersão da amostra em relação ao centro da área de distribuição esperada para intensidade e frequência luminosa normalizada daquele genótipo (AA, AB ou BB). O valor do parâmetro varia de 0 (ponto mais distante do centro) a 1 (centro do cluster), sendo 0.15 o limiar mínimo para a designação do genótipo na plataforma Infinium (Figura 7). Valores acima de 0.7 indicam genótipos excelentes, enquanto o intervalo de 0.15 a 0.7 é considerado uma faixa sensível. O 10º percentil do GC (GC10) é um parâmetro que indica, para cada amostra, a qualidade global dos genótipos determinados.

Todo genótipo com designação bem sucedida ($GC > 0.15$) é denominado *Call*, e de maneira oposta temos o *No Call*. Assim, um parâmetro que indica a taxa de determinação de genótipos para uma determinada amostra é o *Call Rate* (CR), o qual consiste na simples divisão de *Calls* pelo total de SNPs ($Calls + No Calls$). Esta proporção é então transformada em porcentagem. Para genotipagens realizadas de acordo com as instruções do protocolo Infinium (Illumina), utilizando gDNA dentro dos padrões de quantidade e qualidade recomendados, amostras com CR acima de 90% são consideradas como genotipagem bem sucedida, embora 99% ser indicativo de excelência. Valores de CR e GC10 para cada amostra foram obtidas através do módulo *Genotyping* v.1.9.4 do software *GenomeStudio* v.2011.1.

Figura 7 - Determinação de genótipos e GenCall Score (GC).

Legenda: Eixos x e y representam a razão de intensidade alélica (Norm Theta, i.e., frequência da fluorescência total) e a intensidade de fluorescência total (Norm R), respectivamente. Elipsóides descrevem a inércia esperada para 95% dos pontos com genótipos AA, AB e BB (i.e., clusters), e cruzes no interior dos elipsóides representam a expectativa de Norm Theta e Norm R (i.e., centróides). A: Determinação do GC (distância relativa entre um ponto projetado e o centróide mais próximo). B: Inclusão de intervalos de confiança ('faixa sensível') para o GC (áreas sombreadas escuras, delimitadas por flechas), formando o 'Genoplot'. Pontos projetados dentro do intervalo de confiança são denominados *Calls*, e pontos fora do intervalo são denominados *No Calls*.

Fonte: Alonso (2013).

5.1.2 Allele Drop In (ADI) e Allele Drop Out (ADO)

As taxas de erro de genotipagem (ADI e ADO) foram calculadas baseadas somente nos alelos autossômicos identificados corretamente tanto no pai como na mãe de cada embrião. O ADI foi calculado com base na expectativa de homozigose, devido ao loci homozigoto em ambos os pais, enquanto que o ADO foi calculado na expectativa de heterozigose, resultante de loci homozigoto de um dos pais (Quadro 1).

O cálculo da frequência absoluta de casos de ADI e ADO por amostra foi realizado com auxílio do software *PLINK* v1.07 (PURCELL et al., 2007), utilizando os procedimentos implementados na opção mendel e scripts customizados na linguagem AWK. Os valores absolutos foram transformados em porcentagem, pela divisão do número de ADI e ADO pelo total de Calls de uma determinada amostra.

Quadro 1 – Definição de ADI e ADO do software PLINK v1.07

Tipo de erro Mendeliano	Pai	Mãe	Filho
ADI	AA	AA	AB
	BB	BB	AB
	BB	BB	AA
	AA	AA	BB
ADO	BB	**	AA
	**	BB	AA
	AA	**	BB
	**	AA	BB

Legenda: ADI = Allele Drop In / ADO = Allele Drop Out

5.2 ESTATÍSTICA DESCRITIVA

Medidas resumo foram utilizadas para descrever a distribuição dos dados para cada variável estudada. Como medidas de posição, foram reportadas as estatísticas de tendência central média, mediana e moda, e as separatrizes 1º quartil (Q1), 3º quartil (Q3), 95º percentil (P95) e 5º percentil (P5). Para representar a dispersão dos dados, foram utilizadas as medidas: máximo, mínimo, amplitude total, amplitude interquartil, variância, desvio padrão e coeficiente de variação. A forma da distribuição de cada variável foi sumarizada através do coeficiente de assimetria e curtose. Adicionalmente, três testes para a aderência dos dados à distribuição normal foram aplicados: Shapiro-Wilk, Kolmogorov-Smirnov e Anderson-Darling. Essas estatísticas, e outras descritas a seguir, foram calculadas no ambiente *R* v.2.15.0 (R DEVELOPMENT CORE TEAM, 2008), utilizando pacotes da base do software e rotinas customizadas. O conjunto de dados (dataset) utilizado nas análises está descrito nos apêndices desse trabalho.

5.3 CORRELAÇÃO ENTRE AS VARIÁVEIS

Para caracterizar o grau de dependência entre CR, GC10, ADI e ADO, o coeficiente de correlação de postos de Spearman (ρ) foi calculado para cada possível par de variáveis. Este coeficiente foi adotado por possuir propriedades estatísticas desejáveis ao tipo de associação investigada: 1) a medida é não paramétrica, ou seja, não assume ou requer conhecimento sobre os parâmetros da distribuição de probabilidade da qual as variáveis comparadas foram amostradas; e 2) ao contrário da correlação de Pearson, que investiga a relação entre as variáveis por meio de uma função estritamente linear, ρ busca qualquer relação monotônica entre elas, ou seja, investiga se valores ordenados em uma variável se refletem em valores ordenados na outra.

Valores de ρ distribuem-se entre -1 e 1. O sinal de ρ indica a direção da associação entre as variáveis: positivo se os valores de uma variável aumentam em

função da outra e negativo se os valores de uma variável diminuem em função da outra. Uma correlação igual a zero denota independência entre as variáveis. Desta forma, um teste de hipóteses foi aplicado, para cada comparação, para determinar se ρ é significativamente diferente de zero, adotando $\alpha = 0.05$. Além disso, gráficos de dispersão foram plotados para cada par de variáveis.

5.4 AVALIAÇÃO DA QUALIDADE DA GENOTIPAGEM COM DIFERENTES QUANTIDADES DE CÉLULAS OBTIDAS NA BIÓPSIA EMBRIONÁRIA

A família de modelos para Análise de Variância (ANOVA) compõe a ferramenta de escolha no teste de diferença de médias de uma variável entre grupos submetidos a tratamentos distintos. Entretanto, pressuposições destes modelos, tais como independência das observações, homogeneidade de variância entre grupos e normalidade da população, limitam sua aplicação em conjuntos de dados que não atendem estas exigências. Quando estas pressuposições são violadas, métodos de transformação de dados são empregados para melhorar a aderência das variáveis ao modelo, porém estes procedimentos nem sempre resultam em ajuste. Assim, o análogo não paramétrico da ANOVA, o teste de Kruskal-Wallis para análise de variância de fator único, foi utilizado para investigar diferenças na distribuição das variáveis CR, GC10, ADI e ADO entre grupos de biópsia com diferentes números de células. A hipótese nula foi rejeitada para P -valores inferiores a 0.05. Em seguida, a versão adaptada para dados em postos do teste de Tukey para diferenças honestamente significantes foi aplicada para a identificação *post hoc* de quais grupos apresentavam diferenças. Os resultados foram reportados em tabelas e gráficos de boxplot.

5.5 INFERÊNCIA DE GENÓTIPOS NÃO DETERMINADOS NAS BIÓPSIAS EMBRIONÁRIAS

Uma vez que os genótipos das doadoras e touros utilizados nos acasalamentos para a produção dos embriões *in vitro* eram conhecidos, uma parcela dos genótipos não determinados (NoCall) nas amostras de biópsia de embrião foi reconstruída com base na primeira lei de Mendel, ou “Lei da Segregação”. Assumindo um genoma diploide, o genótipo de um indivíduo para um determinado SNP é composto por um alelo de origem paterna e outro de origem materna. Quando ambos os pais são homozigotos, o genótipo do embrião pode ser facilmente inferido. Entretanto, quando um dos pais é heterozigoto, a inferência do genótipo não observado requer a aplicação de análises de imputação mais sofisticadas. Assim, foi desenvolvido um algoritmo (mendelFix) para a inferência dos genótipos não determinados, possibilitando aumentar o CR final das amostras obtidas a partir de biópsias embrionárias. Os passos do programa computacional “mendelFix” estão descritos nos apêndice desse trabalho.

6 RESULTADOS

6.1 AVALIAÇÃO DA METODOLOGIA WGA/MDA COM QUANTIDADES DECRESCENTES DE gDNA

O desempenho da metodologia de WGA, pelo método MDA (Repli-g Mini Kit – Qiagen) foi avaliado de acordo com a quantidade de gDNA inicial, utilizando quantidades decrescentes de três amostras de gDNA bovino (2 machos e 1 fêmea): 10ng, 1ng, 500pg, 200pg, 100pg e 10pg. O produto do WGA foi quantificado por espectrofotometria (NanoDrop 2000) e avaliado por eletroforese em gel de agarose. A quantificação por espectrofotometria mostrou-se ineficiente para a avaliação dos produtos de WGA entre as diferentes quantidades de gDNA (Tabela 1), com altas concentrações independentemente da quantidade de gDNA inicial, bem como nas amostras de controle negativo (NO DNA).

Tabela 1 - Quantificação por espectrofotometria de amostras submetidas à metodologia WGA/MDA com quantidades decrescentes de gDNA inicial

AMOSTRA	QUANTIFICAÇÃO DO PRODUTO DE WGA (ng/ μ L)						
	10ng	1ng	500pg	200pg	100pg	10pg	NO DNA
M1	330.5	361.1	464.9	466.7	362.0	454.5	465.8
M2	362.4	452.7	453.3	460.3	458.9	461.4	457.2
F8	387.2	435.8	442.0	457.8	463.6	463.8	574.3
MÉDIA	360.0	416.5	453.4	461.6	428.2	459.9	499.1

Legenda: M1 e M2 = gDNA de macho bovino; F8 = gDNA de fêmea bovina

A avaliação por eletroforese em gel de agarose do produto de WGA com quantidades decrescentes de gDNA inicial mostrou a presença de amplificação de fragmento de aproximadamente 10kb, com evidente diminuição de intensidade da banda com menores quantidades de gDNA (Figura 8). Foi possível a identificação de produto amplificado até a quantidade 100pg de gDNA inicial.

Figura 8 - Eletroforese em gel de agarose (1%) corado com brometo de etídio com 10µL do produto de WGA com quantidades decrescentes de gDNA inicial

Legenda: P = Peso molecular (1kb); F8D1 = 10ng; F8D2 = 1ng; F8D3 = 500pg; F8D4 = 200pg; F8D5 = 100pg; F8D6 = 10pg.

Fonte: Alonso (2013).

A avaliação da eficiência de amplificação pela metodologia MDA foi realizada pela genotipagem em plataforma iScan/Illumina, utilizando o BeadChip BovineLD, com 6.909 SNP. Uma amostra com 200ng de gDNA de cada animal foi utilizada como padrão ouro, seguindo o protocolo recomendado pela Illumina, possibilitando a comparação do CR e GC10 entre a amostra controle (gDNA) e as quantidades decrescentes de gDNA submetidas ao protocolo de WGA, conforme descrito na tabela 2.

Tanto o CR como o GC10 foram diretamente proporcionais com a quantidade de gDNA inicial para o WGA. Os resultados médios (CR/GC10) com as quantidades de 10ng (97,84/0,79) e 1ng (96,03/0,77) foram muito próximos dos obtidos com gDNA (98,86/0,81), com pequena variação entre as diferentes amostras. A variação entre amostras aumentou com as quantidades de 500pg, 200pg e 100pg, embora a média de CR tenha sido mantida acima de 80% e a média de GC10 acima de 0,55 para esses grupos. Os valores de CR (43,61%) e GC10 (0,28) diminuíram consideravelmente nas amostras de WGA com 10pg de gDNA inicial. A distribuição do CR das 3 amostras analisadas nas quantidades decrescentes de gDNA pode ser observada no gráfico dotplot (Figura 9).

Tabela 2 – Resultados de Call Rate (CR) e 10º percentil do GCScore (GC10) de 3 amostras de gDNA bovino (M1, M2 e F8) e das respectivas amostras submetidas ao protocolo de WGA/MDA com quantidades decrescentes de gDNA inicial

AMOSTRA	gDNA		WGA (4µL)											
	(200ng)		10ng		1ng		500pg		200pg		100pg		10pg	
	CR	GC10	CR	GC10	CR	GC10	CR	GC10	CR	GC10	CR	GC10	CR	GC10
M1	98,7	0,81	98,7	0,81	97,0	0,78	80,0	0,51	71,3	0,45	75,6	0,51	39,5	0,24
M2	99,1	0,81	98,0	0,80	95,8	0,78	94,3	0,76	89,1	0,70	84,9	0,60	51,8	0,33
F8	98,6	0,81	96,8	0,77	95,2	0,76	91,4	0,72	85,8	0,60	81,3	0,53	39,5	0,27
MÉDIA	98,9	0,81	97,8	0,79	96,0	0,77	88,6	0,66	82,1	0,58	80,6	0,55	43,6	0,28

Figura 9 – Dotplot da distribuição do Call Rate (CR) de 3 amostras de gDNA bovino (M1, M2, F8) com quantidades decrescentes de gDNA molde submetidas ao protocolo de amplificação total do genoma (WGA) pela metodologia de amplificação por deslocamento múltiplo (MDA)

Fonte: Alonso (2013).

6.2 AVALIAÇÃO DA GENOTIPAGEM DE BIÓPSIAS EMBRIONÁRIAS

Embriões bovinos produzidos *in vitro* foram submetidos a seções de micromanipulação pela técnica de microaspiração para obtenção de biópsias embrionárias. A quantidade de células obtidas na biópsia foi estimada com base no estágio e qualidade do embrião, sendo aspirado aproximadamente 10% da massa embrionária viável, conforme descrito no item 4.3, permitindo a formação de 2 grupos com diferentes quantidades de células: Grupo 1) 5-10 células (28 replicatas) e Grupo 2) 10-20 células (37 replicatas). Os embriões eclodidos foram transferidos individualmente para tubos de PCR (0,2mL), possibilitando a formação de um terceiro grupo: Grupo 3) >100 células (23 replicatas). Todas as amostras foram submetidas à metodologia de WGA/MDA e posteriormente, 4μL de cada amostra foi usado para genotipagem em plataforma iScan/Illumina com o BeadChip BovineLD (6.909 SNP). Tanto as 8 doadoras como os 3 touros utilizados para a produção *in vitro* dos embriões também foram genotipados com o mesmo conjunto de SNP.

O produto do WGA/MDA obtido a partir das biópsias embrionárias foi avaliado por eletroforese em gel de agarose (1%), apresentando grande similaridade na intensidade das bandas, em amostras com a mesma quantidade de células, mas com diferentes taxas de CR (Figura 10).

Figura 10 - Eletroforese em gel de agarose (1%) corado com brometo de etídio, com 10μL do produto da amplificação por deslocamento múltiplo (MDA) obtido de amostras com 10-20 células (Grupo 2), com diferentes taxas de Call Rate (CR)

Legenda: Amarelo: CR>91%; Verde: CR 88%-80%; Vermelho: CR<51%

Fonte: Alonso (2013)

A qualidade da genotipagem foi avaliada pela análise das variáveis de CR, GC10, ADI e ADO entre os grupos 1, 2 e 3. A forma de distribuição dos dados foi analisada para cada variável e resumida através do coeficiente de assimetria e curtose. A tabela 3 apresenta os valores da estatística descritiva.

Tabela 3 – Estatística descritiva das variáveis utilizadas para avaliação da qualidade da genotipagem de biópsias embrionárias.

Estatística Descritiva	Variável			
	CR	GC10	ADI	ADO
n	88.00	88.00	88.00	88.00
Média	76.43	0.56	4.81	2.92
Mediana	84.87	0.64	0.05	0.78
Moda	90.98	0.74	0.10	0.28
Q1	54.11	0.31	0.00	0.06
Q3	94.66	0.76	2.80	3.26
P95	97.12	0.79	28.90	15.02
P5	40.51	0.25	0.00	0.00
Máximo	97.45	0.79	42.73	19.19
Mínimo	35.59	0.22	0.00	0.00
Amplitude total	61.86	0.57	42.73	19.19
Amplitude interquartil	40.56	0.45	2.80	3.21
Variância	424.250	0.040	99.200	22.210
Desvio padrão	20.60	0.21	9.96	4.71
Coeficiente de variação	26.95	37.50	207.07	161.30
Coeficiente de assimetria	-0.67	-0.44	2.16	2.07
Curtose	-1.15	-1.50	3.74	3.44
Shapiro-Wilk	1.89×10^{-08}	1.32×10^{-08}	6.86×10^{-15}	4.83×10^{-13}
Kolmogorov-Smirnov	1.54×10^{-03}	8.36×10^{-03}	4.00×10^{-11}	6.56×10^{-06}
Anderson-Darling	6.00×10^{-14}	1.13×10^{-13}	1.52×10^{-41}	2.30×10^{-27}

Legenda: CR = Call Rate - GC10 = 10º percentil do GCScore;
ADI = Allele Drop In - ADO = Allele Drop Out

A média do CR apresentou grande variação entre os grupos, com valores de 57,66%, 78,47% e 95,97%, respectivamente para os Grupos 1, 2 e 3. Entretanto, a variação do CR dentro de cada grupo foi inversamente proporcional ao número de células obtidas na biópsia embrionária, com máximas e mínimas de 90,29% e 35,59% (Grupo 1), 95,96% e 42,28% (Grupo 2) e 97,45% e 84,88% (Grupo 3). No grupo 1, apenas 21,4% (6/28) das amostras apresentaram CR acima de 80%, enquanto que no grupo 2 a proporção foi de 70,3% (26/37) de CR>80% (Tabela 4).

Tabela 4 – Avaliação da genotipagem entre grupos com diferentes número de células obtidas na biópsia embrionária, pela análise do Call Rate (CR), 10º percentil do GCScore (GC10), Allele Drop In (ADI) e Allele Drop Out (ADO)

GRUPO 1 (5-10 células)					GRUPO 2 (10-20 células)					GRUPO 3 (>100 células)				
N	CR(%)	GC10	ADI (%)	ADO (%)	N	CR(%)	GC10	ADI(%)	ADO (%)	N	CR(%)	GC10	ADI(%)	ADO (%)
1	90.29	0.73	0.048	0.161	1	95.96	0.78	0.000	0.015	1	97.45	0.79	0.000	0.015
2	83.84	0.59	0.363	2.679	2	94.52	0.77	0.000	0.015	2	97.45	0.79	0.000	0.000
3	83.46	0.59	0.052	0.886	3	93.13	0.76	0.000	0.202	3	97.32	0.79	0.000	0.015
4	82.03	0.54	0.071	1.413	4	91.93	0.76	0.016	0.047	4	97.20	0.78	0.000	0.000
5	81.87	0.58	0.266	4.089	5	91.23	0.75	0.000	0.032	5	97.16	0.79	0.000	0.000
6	81.13	0.57	0.322	3.394	6	90.97	0.74	0.016	0.255	6	97.04	0.79	0.000	0.000
7	79.49	0.53	0.547	11.996	7	90.32	0.72	0.000	0.209	7	96.99	0.79	0.015	0.179
8	73.26	0.43	0.495	3.996	8	90.04	0.74	0.016	0.080	8	96.91	0.79	0.000	0.000
9	70.86	0.51	0.123	9.695	9	89.49	0.71	0.000	0.113	9	96.88	0.79	0.000	0.000
10	64.03	0.37	0.747	3.984	10	89.04	0.71	0.000	0.342	10	96.88	0.79	0.000	0.000
11	63.33	0.35	1.304	5.904	11	88.68	0.71	0.016	0.523	11	96.74	0.79	0.000	0.000
12	58.32	0.32	18.216	10.860	12	88.38	0.73	0.000	0.377	12	96.74	0.79	0.000	0.000
13	52.86	0.28	38.634	17.795	13	88.20	0.71	0.000	0.164	13	96.70	0.79	0.000	0.015
14	49.61	0.26	15.483	0.935	14	87.61	0.67	0.033	0.860	14	96.59	0.79	0.015	0.090
15	47.75	0.28	18.361	1.639	15	87.23	0.67	0.050	2.160	15	96.44	0.78	0.030	0.000
16	47.48	0.26	24.542	1.465	16	87.20	0.69	0.000	0.199	16	96.36	0.78	0.000	0.000
17	46.70	0.28	17.970	0.993	17	87.01	0.70	0.000	2.715	17	96.23	0.79	0.000	0.000
18	45.65	0.26	26.254	16.413	18	86.23	0.71	0.000	0.471	18	96.19	0.76	0.045	0.060
19	45.55	0.26	42.730	2.100	19	85.99	0.64	0.000	0.303	19	96.03	0.76	0.000	0.045
20	45.38	0.24	5.493	0.926	20	85.33	0.61	0.000	0.577	20	95.44	0.77	0.000	0.030
21	45.06	0.24	3.988	1.866	21	84.86	0.67	0.017	0.427	21	95.09	0.76	0.000	0.061
22	44.94	0.26	19.252	1.774	22	84.61	0.61	0.171	3.443	22	92.71	0.74	0.000	0.016
23	41.72	0.26	16.360	1.806	23	84.39	0.60	0.069	3.108	23	84.88	0.68	0.000	0.068
24	39.86	0.24	20.836	4.982	24	84.33	0.65	0.000	0.687	-	-	-	-	-
25	39.55	0.26	1.392	12.422	25	80.99	0.58	0.072	3.221	-	-	-	-	-
26	38.17	0.22	35.687	11.503	26	80.41	0.56	0.054	2.740	-	-	-	-	-
27	36.81	0.25	3.740	17.717	27	73.00	0.47	0.099	2.363	-	-	-	-	-
28	35.59	0.28	0.733	9.731	28	71.96	0.45	0.101	2.739	-	-	-	-	-
-	-	-	-	-	29	68.49	0.42	0.592	1.735	-	-	-	-	-
-	-	-	-	-	30	59.93	0.35	2.563	5.030	-	-	-	-	-
-	-	-	-	-	31	54.52	0.33	30.728	19.192	-	-	-	-	-
-	-	-	-	-	32	51.46	0.27	8.927	2.872	-	-	-	-	-
-	-	-	-	-	33	51.22	0.26	3.509	0.707	-	-	-	-	-
-	-	-	-	-	34	49.42	0.27	30.323	7.683	-	-	-	-	-
-	-	-	-	-	35	47.93	0.27	19.323	5.624	-	-	-	-	-
-	-	-	-	-	36	45.17	0.29	6.288	4.235	-	-	-	-	-
-	-	-	-	-	37	42.28	0.25	5.896	17.655	-	-	-	-	-
Média	57.66	0.37	11.215	5.826	Média	78.47	0.58	2.943	2.517	Média	95.97	0.78	0.005	0.026

A taxa de ADI diminuiu de 11,2% para 2,9% entre os grupos 1 (5-10 células) e 2 (10-20 células), atingindo taxa praticamente nula (0,005%) no grupo com mais de 100 células. Os valores médios de ADO foram de 5,8%, 2,5% e 0,026% para os grupos 1, 2 e 3, respectivamente.

A aderência dos dados à distribuição normal foi analisada por três diferentes testes, conforme descrito no item 5.2, entretanto nenhum deles apontou para a normalidade dos dados (Tabela 3). Os seguintes métodos de transformação de dados foram empregados na tentativa de melhorar a aderência das variáveis aos modelos para Análise de Variância (ANOVA): Raiz Quadrada, Transformação Logarítmica, Arco-seno e Box-Cox. Como nenhum dos métodos apresentou resultado satisfatório, a investigação das diferenças na distribuição das variáveis foi realizada pelo análogo não paramétrico da ANOVA, o teste de Kruskal-Wallis, revelando diferenças altamente significativas em todas as variáveis analisadas entre os 3 grupos com diferentes número de células (Tabela 5).

Tabela 5 - Análise de variância não paramétrica (Kruskal-Wallis) para as variáveis dependentes investigadas. CR = Call Rate, GC10 = 10º percentil do GC Score, ADI = Allele Drop In, ADO = Allele Drop Out

VARIÁVEL	χ^2*	p-valor
CR	59,079	1.48×10^{-13}
GC10	58,380	2.10×10^{-13}
ADI	46,773	6.97×10^{-11}
ADO	52,784	3.45×10^{-12}

Legenda: *2 Graus de liberdade (número de grupos -1)

A análise para identificação das diferenças entre os grupos 1, 2 e 3 foi realizada pela versão adaptada para dados em postos do teste de Tukey para diferenças honestamente significantes, tendo os resultados sido reportados em gráfico boxplot (Figura 11). Tanto as taxas de CR como de GC10 foram significativamente diferentes entre os 3 grupos.

Figura 11 – Representação gráfica (Boxplot) das variáveis de genotipagem analisadas entre os 3 grupos com diferentes quantidades de células obtidas na biópsia embrionária

Legenda: C1 = 5-10 células; C2 = 10-20 células; C3 = >100 Células. CR = Call Rate; GC10 = 10º percentil do GC Score; ADI = Allele Drop In; ADO = Allele Drop Out. Letras diferentes (a, b ou c) representam diferença significativa entre os grupos. Fonte: Alonso (2013).

O grau de dependência entre as variáveis foi calculado pelo coeficiente de correlação de postos de Spearman (ρ), sendo que os valores de ρ distribuem-se entre -1 e 1. Os resultados foram descritos na tabela 6 e gráficos de dispersão foram plotados para cada par de variáveis (Figura 12). Os resultados apresentaram correlação positiva entre o CR e GC10 ($0,99/P < 0,001$). As taxas de ADI e ADO apresentaram correlação negativa com o CR e CG10 (ADI/CR: -0,87; ADI/GC10: -0,88; ADO/CR: -0,87; ADO/GC10: -0,86), com $P < 0,001$ para todas as variáveis.

Tabela 6 - Coeficientes de correlação de postos de Spearman para as variáveis dependentes investigadas. CR = Call Rate, GC10 = 10º percentil do GCScore, ADI = Allele Drop In, ADO = Allele Drop Out

VARIÁVEIS	CR	GC10	ADI	ADO
CR	-	0.99***	-0.87***	-0.87***
GC10	0.99***	-	-0.88***	-0.86***
ADI	-0.87***	-0.88***	-	0.82***
ADO	-0.87***	-0.86***	0.82***	-
*P < 0.05 **P < 0.01 ***P < 0.001				

Figura 12 – Gráfico de dispersão (Scatterplot) entre os possíveis pares de variáveis

Legenda: CR = Call Rate; GC10 = 10º percentil do GCScore; ADI = Allele Drop In; ADO = Allele Drop Out. Fonte: Alonso (2013).

Conforme os resultados apresentados, a taxa de insucesso da genotipagem foi avaliada pela proporção de SNP não determinados (NoCall) em cada amostra. A taxa de NoCall foi inversamente proporcional ao número de células obtidas na biópsia embrionária, ou seja, quanto menor o número de células maior será o número de SNP não determinados. Enquanto o Grupo 1 (5-10 células) apresentou média de NoCall de 42,34%, esse valor diminuiu para 21,53% no Grupo 2 (10-20 células) e para 4,03% no Grupo 3 (>100 células).

No presente estudo, a avaliação das taxas de erro de genotipagem (ADI e ADO) foi realizada com base no padrão de herança mendeliana, ou seja, baseado nos genótipos das doadoras e touros utilizados no acasalamento de cada embrião bovino produzido *in vitro*. Seguindo o mesmo raciocínio e de posse do genótipo dos pais, foi possível desenvolver um algoritmo para a inferência de genótipos não determinados, possibilitando aumentar o CR final das amostras obtidas a partir de biópsias embrionárias.

A tabela 7 apresenta a comparação do CR original (CR1 - obtido na genotipagem dos grupos com diferentes quantidades de células embrionárias) com o respectivo valor do CR corrigido pela inferência de genótipos dos pais (CR2). Dessa forma, o aumento de CR1 para CR2 e o ganho médio (CR2 – CR1) entre os Grupos 1, 2 e 3, foi respectivamente de 57,64% para 79,69% (22,05%), 78,43% para 88,20% (9,77%) e 95,92% para 97,28% (1,37%). O CR mínimo antes e depois da correção por inferência dentro dos grupos aumentou de 35,60% para 70,60% (Grupo 1), de 42,30% para 74,80% (Grupo 2) e de 84,90% para 88,90% (Grupo 3).

Tabela 7 – Comparação do Call Rate original (CR1) obtido nos grupos com diferente número de células (1, 2 e 3) com os respectivos valores após a correção dos genótipos pela inferência baseada na Lei da Segregação (CR2) pelo “mendelFix”

	GRUPO 1			GRUPO 2			GRUPO 3		
	CR	MÍN	MÁX	CR	MÍN	MÁX	CR	MÍN	MÁX
CR1 (%)	57.64	35.60	90.20	78.43	42.30	95.90	95.92	84.90	97.50
CR2 (%)	79.69	70.60	92.60	88.20	74.80	97.10	97.28	88.90	98.60
CR2 - CR1	22.05			9.77			1.37		

7 DISCUSSÃO

O Brasil é líder mundial na produção *in vitro* de embriões bovinos. Atualmente há empresas distribuídas nos diferentes estados do país que realizam todas as etapas do trabalho, incluindo a seleção e sincronização das receptoras, aspiração folicular, produção *in vitro*, transferência de embriões, diagnóstico de gestação e sexagem fetal, fazendo com que essa biotecnologia esteja cada vez mais acessível aos criadores.

Paralelamente à rápida evolução e aplicação das biotecnologias da reprodução animal, os recentes avanços na área da genômica tem permitido a associação das características fenotípicas com informações moleculares, presentes no DNA. Dessa forma, a seleção animal tem caminhado para a associação das características quantitativas, tradicionalmente avaliadas pelas DEPs, com as informações genômicas, possibilitando o desenvolvimento e aplicação do conceito de seleção assistida por marcadores e/ou seleção genômica.

Ao contrário do melhoramento genético clássico, que depende da avaliação da progênie de um determinado indivíduo, a seleção genômica permite a obtenção de informações logo após o seu nascimento, diminuindo o custo e o tempo do processo de seleção.

As atuais plataformas de genotipagem permitem a determinação de centenas a milhares de marcadores do tipo SNP, simultaneamente, a partir de amostra com 200ng de gDNA. Essa quantidade de gDNA pode ser facilmente obtida pelos métodos de extração e purificação de amostras biológicas como sangue, bulbos capilares, sêmen e diferentes tecidos.

Nos mamíferos, o genoma de um indivíduo é definido a partir da fecundação dos gametas haploides, sendo metade da informação proveniente do oócito e metade do espermatozoide. Após a fecundação e formação do zigoto (diploide), os processos de multiplicação e diferenciação celular são os responsáveis pelo desenvolvimento do embrião, cuja sequencia de gDNA será mantida praticamente a mesma após o nascimento e por toda a vida do animal.

O diagnóstico genético pré-implantacional compreende um conjunto de técnicas que possibilitam a realização de análises genéticas ainda na fase embrionária, permitindo a identificação e prevenção de doenças genéticas, bem

como a seleção dos embriões com características zootécnicas desejadas. Entretanto, o processo de micromanipulação e a quantidade de células obtidas na biópsia pode comprometer a viabilidade embrionária. Em trabalho realizado com 4.650 embriões bovinos PIV, submetidos à micromanipulação pela técnica de microaspiração, observamos que nas biópsias de até 15% da massa celular embrionária viável a taxa de sobrevivência dos embriões foi de 90,1%, enquanto que no grupo com biópsias com mais de 15%, a viabilidade diminuiu para 72,6% (Alonso, 2008). Lauri et al. (2013) relataram resultado semelhante com embriões bovinos produzidos *in vivo* e biopsiados pela técnica de microseção, com sobrevivência de 89% dos embriões com biópsia de 10 células e de 62% para biópsias de 30 células.

Dessa forma, uma das maiores limitações para a realização da genotipagem de embriões em plataforma de alta densidade de SNP é a pequena quantidade de gDNA obtida na biópsia embrionária. Cada célula bovina possui ~5pg de gDNA (Lauri et al., 2013), ou seja, uma biópsia com 10-20 células terá no máximo entre 50-100pg, sendo necessário aumentar essa quantidade em mais de 2.000 vezes para atingir os valores recomendados para as atuais plataformas de genotipagem de SNP.

Diversos estudos tem apontado para a metodologia de amplificação total do genoma (WGA) pela técnica de amplificação por deslocamento múltiplo (MDA) como sendo a melhor alternativa para aumentar a quantidade de gDNA das células obtidas na biópsia embrionária. Nesse trabalho, a metodologia de WGA/MDA (Repli-g Mini Kit - Qiagen) foi aplicada tanto em amostras com quantidades decrescentes de gDNA inicial como em biópsias com diferentes quantidades de células obtidas de embriões bovinos PIV. Todas as amostras foram genotipadas com microarranjo de 6.909 SNP (BeadChip Bovine LD), em plataforma iScan/Illumina, permitindo a análise do CR, GC10, ADI e ADO e a comparação entre os grupos com diferentes quantidades de células.

O experimento inicial foi realizado para avaliar o resultado do WGA/MDA com quantidades decrescentes de gDNA inicial. Considerando que uma célula bovina possui ~5pg de gDNA, as quantidades de 10ng, 1ng, 500pg, 200pg, 100pg e 10 pg, representam quantidades equivalentes a 2.000, 200, 100, 40, 20 e 2 células bovinas, respectivamente. O produto do WGA/MDA de todas as amostras foi quantificado por espectrofotometria (Nanodrop 2000) e avaliado em gel de agarose 1% para visualização dos fragmentos amplificados. Embora alguns autores tenham relatado a

eficiência da metodologia MDA pela quantificação das amostras por espectrofotometria, os resultados desse estudo demonstraram que essa não é uma metodologia indicada para isso. Assim como descrito por Treff et al. (2011), todas as amostras, incluído os controles negativos, apresentaram altos valores na quantificação por espectrofotometria.

A avaliação do produto de WGA/MDA também foi realizada por eletroforese em gel de agarose. Os resultados visualizados com quantidades decrescentes de gDNA (Figura 8) demonstraram bandas com intensidade decrescente, proporcional à quantidade de gDNA inicial, sendo possível visualizar a presença de fragmento com aproximadamente 10kb nas amostras com até 100pg de gDNA inicial. A intensidade das bandas também foi proporcional aos resultados obtidos na genotipagem, sugerindo que essa poderia ser uma metodologia de triagem para selecionar as amostras bem sucedidas no processo de WGA/MDA e, conseqüentemente, com maiores chances de obtenção de resultados satisfatórios na genotipagem.

As taxas de CR e GC10 foram diretamente proporcionais à quantidade de gDNA molde submetido ao protocolo de WGA/MDA (Tabela 2). A média do CR das amostras genotipadas com 200ng de gDNA (98,86%) foi muito similar a dos grupos submetidos ao WGA/MDA com 10ng (97,84%) e 1 ng (96,06%). Vale ressaltar que o protocolo original do Repli-g Mini Kit (Qiagen) recomenda a utilização mínima de 10ng, embora alguns autores obtiveram bons resultados com quantidades menores, chegando até uma única célula (Treff et al. 2011). A variação do CR entre amostras com a mesma quantidade de gDNA inicial aumentou com amostras ≤ 500 pg, embora a média do CR tenha se mantido acima de 80% até a quantidade de 100pg, o que equivale a ~20 células bovinas, sugerindo que essa metodologia pode ser aplicada com relativo sucesso em biópsias embrionárias. Embora Treff et al. (2011) tenha reportado CR de 88% na genotipagem em plataforma com 250K SNP (Affymetrix) a partir de uma única célula submetida ao mesmo protocolo de amplificação utilizado nesse estudo (Repli-g Mini Kit – Qiagen), a média de CR obtida na plataforma iScan/Illumina utilizada nesse trabalho, utilizando 10pg de gDNA inicial (~2 células) foi de 43,6%, não apresentando resultados satisfatórios.

O experimento realizado a partir de biópsias com diferentes quantidades de células de embriões bovinos PIV revelou que a qualidade da genotipagem está diretamente relacionada com a quantidade de células obtidas na biópsia, e consequentemente, com a quantidade de gDNA inicial ao processo de WGA/MDA.

Uma das principais dificuldades no delineamento experimental do presente estudo foi definir a quantidade de células obtidas na biópsia de mórulas e blastocistos pela técnica de microaspiração, devido a impossibilidades de determinar com exatidão a quantidade de células sem auxílio de técnicas de coloração. Lauri et al. (2013) mencionaram a realização de biópsias com 75, 40, 30, 20 e 10 células, obtidas pela técnica de microseção, entretanto os autores não descreveram a metodologia de contagem de células. Cenariu et al. (2012) descreveram diferentes métodos para a realização da biópsia de mórulas e blastocistos bovinos produzidos *in vivo*, relatando a obtenção de 2-3 células das mórulas e de 4-5 células do trofoblasto dos blastocistos, mas também não citaram a metodologia de contagem de células. Lacaze et al. (2008) compararam a taxa de concepção e a acurácia da identificação do sexo de embriões bovinos com biópsias de 3, 3-7 e >7 células, mas assim como os demais autores, não descreveram a metodologia de contagem de células.

No presente estudo, o número de células foi estimado com base no percentual de material aspirado em relação ao total da massa embrionária viável, sendo padronizada a aspiração de aproximadamente 10% do total de células de cada embrião. Além disso, o número total de células dos embriões foi estimado com base em referências de autores que utilizaram métodos de coloração para correlacionar o número de células com o estágio embrionário (XU et al., 1992; SOOM et al., 1997; MORI et al., 2002). Dessa forma, consideramos que mórulas e blastocistos iniciais possuem entre 50-100 células, e que blastocistos e blastocistos expandidos têm entre 100 e 200 células, permitindo a classificação das biópsias em número estimado de células de acordo com o estágio de desenvolvimento embrionário. Assim, a aspiração de 10% da massa celular de embriões menos desenvolvidos (Mo e Bi) formaram o grupo de 5-10 células (Grupo 1), enquanto que os mais desenvolvidos (Bl e Bx) formaram o grupo com 10-20 células (Grupo 2). O grupo 3 (>100 células) foi formado por blastocistos eclodidos, que foram transferidos individualmente para tubos de PCR.

Todas as amostras foram submetidas ao mesmo protocolo de WGA/MDA e o produto foi avaliado por eletroforese em gel de agarose, seguindo a mesma estratégia descrita com quantidades decrescentes de gDNA. Essa metodologia foi utilizada para avaliar o produto de WGA/MDA obtido com diferentes quantidades de células embrionárias, bem como associar a intensidade da banda com o resultado da genotipagem de cada amostra. Entretanto, o padrão de intensidade das bandas foi muito similar entre os grupos com diferentes quantidades de células. Além disso, não foi possível observar qualquer tipo de associação da intensidade das bandas com os respectivos resultados da genotipagem. Conforme apresentado na figura 12, amostras com CR>90% apresentaram grande similaridade no padrão de intensidade das bandas das amostras com CR<51%, mostrando que essa metodologia não é eficiente para a triagem de amostras com maiores chances de sucesso na genotipagem, como sugerido anteriormente. Embora metodologias mais avançadas de quantificação não tenham sido utilizadas nesse trabalho, a concentração e a quantidade total de material produzido pela MDA pode ser mensurada com maior fidelidade através de técnicas de quantificação de fita dupla de DNA, como o PicoGreen, ou pela PCR em tempo real. A determinação da concentração obtida no processo de WGA/MDA poderia viabilizar avaliação mais precisa do processo de amplificação e, eventualmente, diminuir a variação dos resultados da genotipagem obtidas com biópsias embrionárias.

Todas as variáveis de genotipagem analisadas no presente estudo (CR, GC10, ADI e ADO) apresentaram diferença altamente significativa entre os grupos 1, 2 e 3, comprovando que a quantidade de células obtida na biópsia influencia diretamente o resultado da metodologia de WGA/MDA e, conseqüentemente, a qualidade da genotipagem em plataforma de alta densidade de SNP.

O CR foi diretamente proporcional ao número de células obtidas na biópsia e submetidas ao protocolo de WGA/MDA, ou seja, quanto maior o número de células maior o CR. A variação do CR entre os grupos foi mais alta nas amostras com menor número de células. Esses resultados são semelhantes aos reportados por Lauri et al. (2013), como demonstrado na tabela 8.

Tabela 8 – Resultados de CR (Call Rate), Allele Drop In (ADI) e Allele Drop Out (ADO) do presente estudo (Alonso, 2013) e dos reportados por Lauri et. al. (2013)

Referência	Embrião	WGA	Genotipagem	N° células	N	Call Rate	CR Máx	CR Mín	ADO	ADI
Alonso, 2013	Bovino PIV	Repli-g	BovineLD	5-10	28	57,66	90,29	35,59	5,83	11,22
		Mini Kit	Illumina	10-20	37	78,47	95,96	42,28	2,52	2,94
		Qiagen	6.909 SNP	>100	23	95,97	97,45	84,88	0,03	0,005
Lauri et al., 2013	Bovino Clone	Repli-g	BovineSNP50	10	16	72,45	93,73	29,30	12,60	2,87
		Mini Kit	Illumina	30	16	92,54	96,78	73,46	1,41	0,33
		Qiagen	54.609 SNP	200-1000	12	96,04	98,20	85,75	0,48	0,07

Sargolzaei et al. (2013) reportaram CR médio de 88,6% na genotipagem de 681 embriões bovinos produzidos *in vivo*, enquanto Le Bourhis et al. (2011) relataram CR de $85 \pm 8\%$ com biópsias de 5-7 células e de $90 \pm 4\%$ com biópsia de 8-10 células, ambos com o BovineSNP50 (Illumina). Esses valores de CR, obtidos com embriões bovinos produzidos *in vivo*, foram maiores do que os obtidos no presente estudo com embriões PIV e do que os descritos por Lauri et al. (2013), com embriões clones, com a mesma quantidade de células. Isso sugere que a qualidade do DNA presente nas células dos embriões produzidos *in vivo* pode ser melhor do que a dos embriões PIV, sendo necessária maior quantidade de células desses últimos para obtenção de resultados satisfatórios na genotipagem.

No presente trabalho, a qualidade da determinação dos SNP foi avaliada pelo GC10 e pelas taxas de ADI e ADO, analisadas conforme descrito no item 5.1.2. As taxas de GC10 foram diretamente proporcionais ao CR, com diferenças significativas entre os 3 grupos com diferentes quantidades de células obtidas na biópsia (Figura 11) e alta correlação positiva entre as duas variáveis (Tabela 5). De acordo com os padrões de qualidade estabelecidos pela Illumina, o ensaio Infinium, quando realizado com gDNA, apresentam CR>99% e as taxas de GC10 são consideradas excelentes quando >0,70. Entretanto, CR>90% e GC10>0,15 são valores aceitáveis. Nos resultados obtidos com amostras de WGA/MDA, observa-se que todas as amostras com CR>90% apresentaram GC10>0,70, confirmando que os critérios de avaliação podem ser mantidos com amostras amplificadas pela metodologia MDA.

A acurácia da genotipagem dos embriões foi avaliada pela análise das taxas de ADI e ADO, realizadas com base nos genótipos das 8 doadoras e dos 3 touros utilizados para a produção *in vitro* dos embriões. O coeficiente de correlação de postos de Spearman evidenciou correlação negativa das taxas de ADI e ADO com

as de CR e GC10, ou seja, quanto maiores os valores de CR e GC10, menores serão os erros de genotipagem.

Nesse estudo, as taxas de ADI para os grupos 1, 2 e 3 foram de 11,22%, 2,94% e 0,005%, respectivamente. As taxas de ADO, na mesma ordem dos grupos, foram de 5,83%, 2,52% e 0,03%. Conforme descrito na tabela 8, esses resultados são semelhantes aos reportados por Lauri et al. (2013). Sargolzaei et al. (2013) encontraram taxa de 1,6% de inconsistências dos genótipos obtidos a partir de células biopsiadas quando comparado com os respectivos bezerros nascidos. De acordo com Le Bourhis et al. (2011), amostras com CR>85% apresentaram taxa de ADO<1%, considerando CR de 85% como um valor corte aceitável para a genotipagem de embriões.

O cálculo do coeficiente de correlação de postos de Spearman apontou para correlação altamente significativa ($P<0,001$) entre todas as variáveis de genotipagem. Os resultados foram apresentados na forma de gráfico de dispersão (Figura 12), facilitando a visualização e interpretação da correlação entre os possíveis pares de variáveis. As taxas de ADI permaneceram abaixo de 1% nas amostras com CR>70% e GC10>0,4, sugerindo que esse tipo de erro é menos comum no processo de WGA/MDA. Entretanto, falhas do tipo ADO passaram a ocorrer com maior frequência nas amostras com CR<85% e GC10<0,7.

Os resultados desse estudo demonstram claramente que, apesar de ser possível a realização da genotipagem a partir de células embrionárias, a quantidade de células obtidas na biópsia influencia diretamente a qualidade da genotipagem. Entretanto, a taxa de mortalidade embrionária é diretamente proporcional à quantidade de células obtidas no processo de micromanipulação (ALONSO, 2008; LAURI et al., 2013), ou seja, quanto maior o número de células biopsiadas, menor a chance de sobrevivência do embrião. Dessa forma, a micromanipulação de embriões em estádios de desenvolvimento mais avançados (BI e Bx) pode ser mais recomendada do que embriões nos estádios de Mo e Bi, possibilitando a obtenção de maior número de células com menores prejuízos ao embrião.

Metodologias alternativas tem sido propostas na tentativa de aumentar a quantidade de células obtidas a partir da biópsia embrionária. Gamarra et al. (2009) utilizaram 3 diferentes tipos de meio para o cultivo *in vitro* de células obtidas pela biópsia de embriões bovinos, mas não houve formação de colônia e replicação celular em nenhum dos grupos avaliados. Seguindo a mesma estratégia, Moros-

mora et al. (2011) obteve replicação celular em apenas 50% dos casos, após cultivo *in vitro* de células do trofoblasto de embriões bovinos, durante 10 dias. Melhores resultados foram reportados por Lauri et al. (2013) que realizaram o cultivo *in vitro* de 20-30 células biopsiadas de blastocisto bovino, durante 4-5 dias, atingindo quantidade final de 100-2000 células. Entretanto, os próprios autores apontaram que essa estratégia pode não ser viável comercialmente pelo aumento de custos e do tempo necessário para no cultivo celular.

Mais recentemente, metodologias computacionais de imputação têm sido utilizadas para viabilizar a seleção genômica a partir de ensaios de baixa densidade e custos mais acessíveis, permitindo a determinação *in silico* de haplotipos baseada na determinação de SNP específicos. De acordo com Sargolzaei et al. (2013) a correção dos erros de genotipagem (ADO) podem ser realizadas com uso de software específico de imputação, diminuindo as falhas de genotipagem (NoCall) de 11,4% para 0,01% e as taxas de erro (ADO) de 1,6% para 0,94%.

No presente estudo, foi desenvolvido um algoritmo para a inferência de genótipos baseado na Lei da Segregação (mendelFix), possibilitando aumentar o CR final das amostras genotipadas a partir de biópsias de embriões. Utilizando essa estratégia, foi possível aumentar a média do CR do Grupo 1 (5-10 células) de 57,64% para 79,69%, do Grupo 2 (10-20 células) de 78,43% para 88,20% e do Grupo 3 (>100 células) de 95,92% para 97,28%, de tal forma que o efeito da correção foi maior quando aplicado em amostras com CR mais baixo. No grupo 2, após a correção do CR, 86,5% das amostras (32/37) atingiram CR>80%, demonstrando que a genotipagem de embriões bovinos submetidos à biópsia de 10-20 células, conforme descrito no presente trabalho, pode ser realizada de forma satisfatória.

8 CONCLUSÕES

Os resultados do presente estudo permitem concluir que, embora seja possível realizar a genotipagem de embriões bovinos em plataforma de alta densidade de SNP com amostras submetidas ao protocolo WGA/MDA, o número de células obtidas na biópsia embrionária e, conseqüentemente, a quantidade de gDNA inicial submetida à amplificação, afeta a qualidade da genotipagem. Resultados satisfatórios de genotipagem podem ser obtidos a partir de biópsia com 10-20 células de embriões bovinos produzidos *in vitro*, com a possibilidade de aumentar o CR final pela inferência de parte dos genótipos não determinados.

Finalmente, a associação das biotecnologias da reprodução animal, como a produção *in vitro* de embriões bovinos e a micromanipulação embrionária, com metodologias genômicas, como a amplificação total do genoma e a genotipagem em plataforma de alta densidade de SNP, possibilita a realização do diagnóstico genético pré-implantacional e viabiliza a análise de características genéticas associadas com características produtivas ainda em fase embrionária. O desenvolvimento e aplicação dessas biotecnologias representam grande avanço científico e tecnológico, contribuindo para acelerar ainda mais o processo de melhoramento genético animal.

REFERÊNCIAS

- ABE, T.; SABURI, J.; HASEBE, H.; NAKAGAWA, T.; KAWAMURA, T.; SAITO, K.; NADE, T.; MISUMI, S.; OKUMURA, T.; KUCHIDA, K.; HAYASHI, T.; NAKANE, S.; MITSUHASHI, T.; NIRASAWA, K.; SUGIMOTO, Y.; KOBAYASHI, E. Bovine quantitative trait loci analysis for growth, carcass and meat quality traits in an F2 population from a cross between Japanese Black and Limousin. **Journal of Animal Science**, v. 86, p. 2821-2832, 2008.
- ABIEC. ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNE. **Pecuária Brasileira**. 2011. Disponível em: <http://www.abiec.com.br/3_pecuaria.asp>. Acesso em: 02 out. 2011.
- ALMODIN, C. G.; MORON, A. F.; KULAY JUNIOR, L.; KRUGER, E.; PEREIRA, L. A. C.; MINGUETTI-CÂMARA, V. C. Avaliação do desenvolvimento de embriões bovinos pós-biópsia: um modelo para treinamento. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 21, n. 9, p. 533-538, 1999.
- ALONSO, R. V.; PERRI, S. H. V.; GARCIA, J. F. Preimplantation genetic diagnosis of bovine embryos using whole genome amplification. **Journal Venomous Animals Toxins including Tropical Diseases**, v. 13, n. 4, p. 951, 2007.
- ALONSO, R. V. **Fatores que afetam a viabilidade e a proporção do sexo de embriões bovinos produzidos *in vitro* em programa comercial de sexagem**. 2008. 59 f. Dissertação (Mestrado em Ciência Animal) – Universidade Estadual Paulista, Araçatuba, 2008.
- ALONSO, R. V.; HELLÚ, J. A. A.; PERRI, S. H. V.; VISINTIN, J. A.; GARCIA, J. F. Factors Affecting commercial sexing program and multiple genetic analysis perspectives of in vitro-procedure bovine embryos. **Reproduction, Fertility and Development**, v. 21, n. 1, p. 227, 2008.
- ASSIS NETO, A. C.; GALDOS-RIVEROS, A. C. R.; MANCANARES, A. C. F.; ALBERTO, M. L. V.; AMBRÓSIO, C. E.; PHAVARON, P. O.; MEIRELLES, F. V.; MIGLINO, M. A. Abnormalities in bovine conceptus development during the embryonic phase after in vitro fertilization (IVF) and cloning by nuclear transfer (NT). **Acta Scientiae Veterinariae**, v. 39, n. 1, p. 223-226, 2011.
- BARKER, D. L.; HANSEN, M. S. T.; FARUQI, A. F.; GIANNOLA, D.; IRSULA, O. R.; LASKEN, M. L.; MAKAROV, A.; OLIPHANT, A.; PINTER, J. H.; SHEN, R.; SLEPTSOVA, I.; ZIEHLER, W.; LAI, E. Genotyping across a 2320-SNP linkage panel two methods of Whole-Genome Amplification enable accurate. **Genome Research**, v. 14, n. 1, p. 901-907, 2004.
- BROOKES, A. J. The essence of SNPs. **Gene**, n. 234, p. 177-186, 1999.
- CAETANO, A. R. Marcadores SNP: Conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, p. 64-71, 2009.

CARVALHO, J. O.; SILVA, L. P.; OLIVEIRA, T. A.; SARTORI, R.; DODE, M. A. N. Sexing by flow cytometry changes morphometric characteristics and surface membrane of bovine spermatozoa. **Acta Scientiae Veterinariae**, v. 38, n. 2, p. 678, 2010.

CASAS, E.; KIRKPATRICK, B. W. Evaluation of different amplification protocols for use in primer extension preamplification. **Biotechniques**, v. 20, n. 2, p. 219-225, 1996.

CENARIU, M.; PALL, E.; CERNEA, C.; GROZA, I. Evaluation of bovine embryo biopsy techniques according to their ability to preserve embryo viability. **Journal of Biomedicine and Biotechnology**, v. 2012, p. 1-5, 2012.

CHEN, Y. L.; HUNG, C. C.; LIN, S. Y.; FANG, M. Y.; TSAI, Y. Y.; CHANG, L. J.; LEE, C. N.; SU, Y. N.; CHEN, S. U.; YANG, Y. S. Successful application of the strategy of blastocyst biopsy, vitrification, whole genome amplification, and thawed embryo transfer for preimplantation genetic diagnosis of neurofibromatosis type 1. **Taiwanese Journal of Obstetrics & Gynecology**, v. 50, n. 1, p. 74-78, 2011.

CHRENCK, P.; BOULANGER, L.; HEYMAN, Y.; UHRIN, P.; LAURINCIK, J.; BULLLA, J.; RENARD, J. P. Sexing and multiple genotype analysis from a single cell of bovine embryo. **Theriogenology**, v. 55, n. 1, p. 1071-1081, 2001.

COSKUN, S.; ALSMADI, O. Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. **Prenatal Diagnosis**, v. 27, n. 1, p. 297-302, 2007.

CURI, R. A.; LOPES, C. R. Teste de paternidade em bovinos: Teste de paternidade por microssatélites em bovinos da raça Gir. **Biotecnologia Ciência & Desenvolvimento**, v. 21, p. 40-45, 2001.

DEKKERS, J. C. M. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. **Journal Animal Science**, v. 82, n. 2, p. 313-328, 2004.

FAO. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **The state of food and agriculture, 2010-2011: World food and agriculture in review (Part II)**. Roma: FAO, 2011. p. 65-82. Disponível em: <<http://www.fao.org/docrep/013/i2050e/i2050e.pdf>> Acesso em: 12 dez. 2012.

GAMARRA, G.; LE BOURHIS, D.; GALL, L.; LAFFONT, L.; RUFFINI, S.; HUMBLLOT, P. Attempts to culture biopsied cells from in vitro bovine blastocysts for genotyping. **Reproduction, Fertility and Development**, v. 21, p. 238, 2009.

GARCIA, J. F. Practical considerations of embryo manipulation: Preimplantation genetic typing. **Theriogenology**, v. 56, p. 1393-1399, 2001.

GARCIA, J. F.; PORTO NETO, L. P. Uso de marcadores moleculares em programas de transferência de embriões. **Acta Scientiae Veterinariae**, v. 34, n. 1, p. 197-203, 2006.

GE HEALTHCARE. **Illustra GenomePhi V2 DNA Amplification Kit**. 25-6600-30wp, Rev B, 2006. Disponível em: <[http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/CDA1124D7ECD2298C1257628001D2E3B/\\$file/GPHI_V2_25660030_revB.pdf](http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/CDA1124D7ECD2298C1257628001D2E3B/$file/GPHI_V2_25660030_revB.pdf)>. Acesso em: 08 dez. 2012.

GIBBONS, W. E.; GITLIN, S. A.; LANZENDORF, S. E. Strategies to respond to polymerase chain reaction deoxyribonucleic acid amplification failure in a preimplantation genetic diagnosis program. **American Journal of Obstetrics and Gynecology**, v. 172, n. 4, p. 1088-1096, 1995.

GUTIÉRREZ-GIL, B.; WIENER, P.; NUTE, G.R.; BURTON, D.; GILL, J. L.; WOOD, J.D.; WILLIAMS, J. L. Detection of quantitative trait loci for meat quality traits in cattle. **Animal Genetics**, v. 39, p. 51-61, 2008.

HAN, T.; CHANG, C. W.; KWEKEL, J. C.; CHEN, Y.; GE, Y.; MARTINEZ-MURILLO, F.; ROSCOE, D.; TEZAK, Z.; PHILIP, R.; BIJWAARD, K.; FUSCOE, J. C. Characterization of whole genome amplified (WGA) DNA for use in genotyping assay development. **BMC Genomics**, v. 13, n. 217, p. 1-15, 2012.

HANDYSIDE, A. H.; ROBINSON, M. D.; SIMPSON, R. J.; OMAR, M. B.; SHAW, M. A.; GRUDZINSKAS, J. G.; RUTHERFORD, A. Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. **Molecular Human Reproduction**, v. 10, n. 10, p. 1-6, 2004.

HAYES, B. J. **QTL mapping, MAS, and genomic selection**. Estados Unidos.: Iowa State University, 2007. Disponível em: <<http://www.ans.iastate.edu/section/abg/shortcourse/notes.pdf>> Acesso em: 03 out. 2012.

HOLBROOK, J. F.; STABLEY, D.; SOL-CHURCH, K. Exploring Whole Genome Amplification as a DNA recovery tool for molecular genetic studies. **Journal of Biomolecular Techniques**, v. 16, n. 1, p. 125-133, 2005.

HOCHMAN, D.; ZARON, Y.; DEKEL, I.; FELDMESSER, E.; MEDRANO, J. F.; SHANI, M.; RON, M. Multiple genotype analysis and sexing of IVF bovine embryos. **Theriogenology**, v. 7, n. 3, p. 1063-75, 1995.

HUMBLLOT, P.; LE BOURHIS, D.; FRITZ, S.; COLLEAU, J. J.; GONZALES, C.; GUYADER-JOLY, C.; MALAFOSSE, A.; HEYMAN, Y.; AMIGUES, Y.; TISSIER, M.; PONSART, C. Reproductive technologies and genomic selection in cattle, **Veterinary Medicine International**, 2010. 8 p.
Disponível em: <<http://www.hindawi.com/journals/vmi/2010/192787/>>. Acesso em: 07 mar. 2013

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção da Pecuária Municipal**, Rio de Janeiro, v. 39, p. 1-63, 2011. Disponível em: <ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2011/ppm2011.pdf> Acesso em: 02 nov. 2012.

ILLUMINA. **Revolutionizing genetics in the cattle industry**. Disponível em: <<http://www.illumina.com/Documents/products/brochures/BovineProductBrochure.pdf>> Acesso em: 20 abr. 2013.

IMAI, K.; MATSUGUIGE, T.; WATANABE, T.; SUGIMOTO, Y.; IHARA, N. Mapping of quantitative trait locus for beef marbling on bovine chromosome 9 in purebred Japanese black cattle. **Animal Biotechnology**, v. 18, p. 75-80, 2007.

USDA. INTERAGENCY AGRICULTURAL PROJECTIONS COMMITTEE. **Agricultural Projections to 2017**. Office of the Chief Economist, World Agricultural Outlook Board, U.S. Department of Agriculture. Long-term Projections Report OCE-2008-1. 2008. 104 p. Disponível em: <http://www.ers.usda.gov/media/274754/oce20081_1_.pdf>. Acesso em: 05 out. 2012.

JIANG, Z.; MICHAEL, J. J.; CHEN, J.; DANIELS, T. F.; KUNEJ, T.; GARCIA, M. D.; GASKINS, C. T.; BUSBOOM, J. R.; ALEXANDER, L. J.; WRIGHT JR, R. W.; MACNEIL, M. D. Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. **International Journal of Biological Sciences**, v. 5, n. 6, p. 528-542, 2009.

KLEPER FILHO, E. **Melhoramento genético animal no Brasil: fundamentos, história e importância**. Campo Grande: EMBRAPA – Gado de Corte, 1999. 63 f. Disponível em: <http://ag20.cnptia.embrapa.br/Repositorio/DOC075_000fm0sx1wy02wyiv80kxlb36m6fxzfm.pdf>. Acesso em 02 mar. 2013.

KRISTJANSSON, K.; CHONG, S. S.; VEYVER, V.; SUBRAMANIAN, S.; SNABES, M. C.; HUGHES, M. R. Preimplantation single cell analyses of dystrophin gene deletions using whole genome amplification. **Nature Genetics**, v. 6, p. 19-23, 1994.

LAURI, A.; LAZZARI, G.; GALLI, C.; LAGUTINA, I.; GENZINI, E.; BRAGA, F.; MARIANI, P.; WILLIAMS, J. L. Assessment of MDA efficiency for genotyping using cloned embryo biopsies. **Genomics**, v. 101, p. 24-29, 2013.

LE BOURHIS, A. D.; MULLAART, E.; HUMBLLOT, P.; COPPIETERS, W.; PONSART, C. Bovine embryo genotyping using 50K single nucleotide polymorphism chip. **Reproduction, Fertility and Development**, v. 23, n. 1, p. 197, 2010.

LE BOURHIS, A. D.; MULLAART, E.; SCHROOTEN, C.; FRITZ, S.; COPPIETERS, W.; PONSART, C. Breeding values concordance between embryos and corresponding calves. **Reproduction, Fertility and Development**, v. 24, n. 1, p. 180, 2011.

LI, W. H.; SADLER, L. A. Low nucleotide diversity in man. **Genetics**, v. 129, p. 513–523, 1991.

LÔBO, R. B.; BEZERRA, L. A. F.; MAGNABOSCO, C. U.; OLIVEIRA, H. N.; GROSSI, D. A. **Sumário de touros das raças nelore, guzerá, brahman e tabapuã 2011**. Ribeirão Preto: ANCP, 2011.

LÔBO, R. B.; NKRUMAH, D.; GROSSI, D. A.; BARROS, P. S.; PAIVA, P.; BEZERRA, L. A. F.; OLIVEIRA, H. N.; SILVA, M. V. H. Implementation of DNA markers to produce genomically-enhanced EPDs in Nelore cattle. **Acta Scientiae Veterinariae**, v. 39, n. 1, p. 23-27, 2011.

MAMONE, A. **Representational amplification of genomic DNA**. Genetic Sequencing and Variation. 2003. Disponível em: <http://www.biocompare.com/Application-Notes/42533-GENETIC-SEQUENCING-AND-VARIATION-Representational-Amplification-Of-Genomic-DNA/>. Acesso em: 02 jan. 2013.

BRASIL. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO – Assessoria de Gestão Estratégica. **Projeções do agronegócio Brasil 2010/11 a 2020/21**. 2011. Disponível em: http://www.agricultura.gov.br/arq_editor/file/Ministerio/gestao/projecao/PROJECOE%20DO%20AGRONEGOCIO%202010-11%20a%202020-21%20-%202_0.pdf. Acesso em: 05 maio 2013.

MIZOGUCHI, Y.; WATANABE, T.; FUJINAKA, K.; IWAMOTO, E.; SUGIMOTO, Y. Mapping of quantitative trait loci for carcass traits in a Japanese Black (Wagyu) cattle population. **Animal Genetics**, v. 37, p. 51-54, 2006.

MORI, M.; OTOI, T.; SUZUKI, T. Correlation between the cell number and diameter in bovine embryos produced in vitro. **Reproduction in Domestic Animals**, v. 37, p. 181-184, 2002.

MOROS-MORA, R.; LAGUNA-BARRAZA, R.; BELTRAN-BRENA, P.; GUTIERREZ-ADAN, A.; RIZOS, D.; RAMIREZ, M. A. **A system to establish a biopsy-derived trophoblast cell line for bovine embryo genotyping**. In: REUNIÃO ANUAL DA ASSOCIAÇÃO EUROPEIA DE TRANSFERÊNCIA DE EMBRIÃO, 27., 2011. Chester, Inglaterra, 2011. p. 208.

NASSER, L. F.; PENTEADO, L.; REZENDE, C. R.; SÁ FILHO, M. F.; BARUSELLI, P. S. Fixed time artificial insemination and embryo transfer programs in Brazil. **Acta Scientiae Veterinariae**, v. 39, n.1, p. 15-22, 2011.

OKADA, A.; TAMATE, H. B. Pedigree analysis of the Sika Deer (*Cervus nippon*) using microsatellite markers. **Zoological Science**, v. 17, p. 335-340, 2000.

PAUNIO, T.; REIMA, I.; SYVANEN A. C. Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solidphase minisequencing of blastomere DNA. **Clinical Chemistry**, v. 42, n. 9, p. 1382-1390, 1996.

POLISSENI, J. M. S; FERREIRA DE SÁ, W.; GUERRA, M. O.; MACHADO, M. A.; SERAPIÃO, R. V.; CARVALHO, B. C.; CAMARGO, L. S. A.; PETERS, V. M. Post-biopsy bovine embryo viability and whole genome amplification in preimplantation genetic diagnosis. **Fertility and Sterility**, v. 93, n. 3, p. 783-788, 2010.

PONSART, C.; LE BOURHIS, D.; KNIJN, H.; FRITZ, S.; GUYADER-JOLY, C.; OTTER, T.; LACAZE, S.; CHARREAUX, F.; DUPASSIEUX, D.; MULLAART, E. **Embryo genotyping: from DNA amplification to field implementation**. In: REUNIÃO ANUAL DA ASSOCIAÇÃO EUROPEIA DE TRANSFERÊNCIA DE EMBRIÃO, 28., 2012. Saint Malo, França, 2012. p. 83-93.

PURCELL S.; NEALE B.; TODD-BROWN K.; THOMAS L.; FERREIRA M. A.; BENDER D.; MALLER J.; SKLAR P.; DE BAKKER P. I.; DALY M. J.; SHAM P. C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, v. 81, n. 3, p. 559-575, 2007.

QIAGEN. **Whole genome amplification from small number of cells or single cells**, 2011. Disponível em: <<http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDoQFjAB&url=http%3A%2F%2Fwww.qiagen.com%2Fresources%2FDownload.aspx%3Fid%3D%257B843654E0-2CCB-474B-B4B8-8744453ED5CB%257D%26lang%3Den%26ver%3D1&ei=LJuMUZXGHvC50AGujYGACw&usq=AFQjCNGRfb1U2DeTp3ZHREBUVGwdG2P7fw>> Acesso em: 25/02/2013

R DEVELOPMENT CORE TEAM: **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.R-project.org>>. Acesso em: 08 fev. 2013.

SARGOLZAEI, M.; VIGNEAULT, C.; BLONDIN, P.; SCHENKEL, F.; CHESNAIS, J. **Genotyping of bovine embryo biopsies with a high density panel for use in genome selection**. In: PLANT & ANIMAL GENOME, 21., 2013, San Diego. 2013. p. 0541. Disponível em: <<https://pag.confex.com/pag/xxi/webprogram/Paper7976.html>>. Acesso em: 15 mar. 2013

SCHAAFF, F.; WEDEMANN, H.; SCHWINGER, E. Analysis of sex and $\Delta F508$ in single amniocytes using primer extension preamplification. **Human Genetic**, v. 98, p. 158-61, 1996.

SMITH, L. C.; SUZUKI, J. R. J.; PERECIN, F.; MEIRELLES, F. V. Reprogramming of genomic imprints by in vitro culture and cloning procedures in cattle. **Acta Scientiae Veterinariae**, v. 39, n. 1, p.103, 2011.

SPITS, C.; LE CAIGNEC, C.; DE RYCKE, M.; VAN HAUTE, L.; VAN STEIRTEGHEM, A.; LIEBAERS, I.; SERMON, K. Optimization and evaluation of single-cell whole genome multiple displacement amplification. **Human Mutation**, v. 27, n. 5, p. 496-503, 2006.

SOOM, A. V.; BOERJAN, M. L.; BOLS, P. E. J.; VANROOSE, G.; LEIN, A.; CORYN, M.; DE KRUIF, A. Timing of compaction and inner cell allocation in bovine embryos produced in vivo after superovulation. **Biology of Reproduction**, v. 57, p. 1041-1049, 1997.

STROUD, B. **The year 2011 worldwide statistics of embryo transfer in domestic animals**. Arizona, Estados Unidos. Statistics and date retrieval committee report, IETS, 2012.

TAKASUGA, A.; WATANABE, T.; MIZOGUCHI, Y.; HIRANO, T.; IHARA, N.; TAKANO, A.; YOKOUCHI, K.; FUJIKAWA, A.; CHIBA, K.; KOBAYASHI, N.; TATSUDA, K.; OE, T.; FURUKAWA-KUROIWA, M.; NISHIMURA-ABE, A.; FUJITA, T.; INOUE, K.; MIZOSHITA, K.; OGINO, A.; SUGIMOTO, Y. Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. **Mammalian Genome**, v. 18, p. 125-136, 2007.

THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v. 324, n. 5926. p. 522-528, 2009.

THE BOVINE HAPMAP CONSORTIUM. Genome-Wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, v. 324, n. 5926, p. 528-532, 2009.

TREFF, N. R.; SU, J.; TAO, X.; NORTHROP, L. E.; SCOTT JR, R. T. Single-cell whole-genome amplification technique impacts the accuracy of SNP microarray-based genotyping and copy number analyses. **Molecular Human Reproduction**, v. 17, n. 6, p. 335-343, 2011.

VAN EENENNAAM, A. **DNA-Based technologies**. Davis, California, Estados Unidos: National Beef Cattle Evaluation Consortium. Beef sire selection manual, Colorado State University, Cornell University e University of Georgia, 2005. p. 66-73. Disponível em: <<http://www.uky.edu/Ag/AnimalSciences/pubs/asc169-13.pdf>>. Acesso em: 10 fev. 2012.

VAN TASSELL, C. P.; SMITH, T. P.; MATUKUMALLI, L. K.; TAYLOR, J. F.; SCHNABEL, R. D.; LAWLEY, C. T.; HAUDENSCHILD, C. D.; MOORE, S. S.; WARREN, W. C.; SONSTEGARD, T. S. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature Methods**, v. 5, n. 3, p. 247-252, 2008.

VIGNAL, A.; MILAN, D.; SAN CRISTOBAL, M. E.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 3, p. 275-305, 2002.

XU, K. P.; YADAV, B. R.; RORIE, R. W.; PLANTE, L.; BETTERIDGE, K. J.; KING, W. A. Development and viability of bovine embryos derived from oocytes matured and fertilized in vitro and co-cultured with bovine oviducal epithelial cells. **Journal of Reproduction & Fertility**, v. 94, p. 33-43, 1992.

XU, K.; TANG, Y.; GRIFO, J. A.; ROSENWAKS, Z.; COHEN, J. Primer extension preamplification for detection of multiple genetic loci from single human blastomeres. **Human Reproduction**, v. 8, n. 12, p. 2206-22, 1993.

WELLS, D.; SHERLOCK, J. K.; HANDYSIDE, A. H.; DELHANTY, J. D. Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridization. **Nucleic Acids Research**, v. 27, n. 4, p. 1214-1218, 1999.

WU, X. L.; MACNEIL, M. D.; DE, S.; XIAO, Q. J.; MICHAL, J. J.; GASKINS, C. T.; REEVES, J. J.; BUSBOOM, J. R.; WRIGHT JR, R. W.; JIANG, Z. Evaluation of candidate gene effects for beef back fat via Bayesian model selection. **Genetics**, v. 125, p. 103-113, 2005.

ZAPATA, S. T. **Micromanipulação e diagnóstico genético pré-implantacional em embriões murinos pela técnica de PEP-PCR ("Primer extension preamplification- polymerase chain reaction")**. 1998. 50 p. Dissertação (Mestrado) – Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 1998.

ZHANG, L.; CUI, X.; SCHIMITT, K.; HUBERT, R.; NAVIDI, W.; ARNHEIM, N. Whole genome amplification from a single cell: implications for genetic analysis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, p. 5847-5851, 1992.

ZHENG, Y.; WANG, N.; LI, L.; JIN, F. Whole genome amplification in preimplantation genetic diagnosis. **Journal of Zhejiang University Science B (Biomedicine & Biotechnology)**, v. 12, n. 1, p. 1-11, 2011.

APÊNDICES: CONJUNTO DE DADOS

Continua

IDENTIFICAÇÃO				CALL RATE (CR)				VARIÁVEIS QUALITATIVAS						INFERÊNCIA DE CR				
Amostra	Touro	Doadora	Grupo	CR (%)	MISS (%)	Call (n)	No Call (n)	GC10	ADI (n)	ADO (n)	ERROS (n)	ADI (%)	ADO (%)	Call 1 (n)	Call 2 (n)	Inferidos	CR 1 (%)	CR 2 (%)
E01	M1	F5	G1	36,81	63,19	2540	4360	0,252	95	450	545	3,74	17,72	2545	4949	2404	36,80	71,60
E04	M1	F7	G1	35,59	64,41	2456	4444	0,282	18	239	257	0,73	9,73	2459	4890	2431	35,60	70,80
E05	M1	F4	G1	82,03	17,97	5660	1240	0,541	4	80	84	0,07	1,41	5665	5977	312	82,00	86,50
E06	M1	F4	G1	45,06	54,94	3109	3791	0,243	124	58	182	3,99	1,87	3111	5010	1899	45,00	72,50
E07	M1	F4	G1	45,38	54,62	3131	3769	0,237	172	29	201	5,49	0,93	3136	5054	1918	45,40	73,20
E08	M1	F2	G1	79,49	20,51	5485	1415	0,532	30	658	688	0,55	12,00	5492	6069	577	79,50	87,80
E09	M1	F6	G1	39,55	60,45	2729	4171	0,256	38	339	377	1,39	12,42	2733	4880	2147	39,60	70,60
E10	M1	F6	G1	70,86	29,14	4889	2011	0,508	6	474	480	0,12	9,70	4890	5885	995	70,80	85,20
E11	M1	F6	G1	64,03	35,97	4418	2482	0,373	33	176	209	0,75	3,98	4421	5574	1153	64,00	80,70
E12	M3	F8	G1	83,84	16,16	5785	1115	0,589	21	155	176	0,36	2,68	5789	6229	440	83,80	90,20
E13	M1	F1	G1	38,17	61,83	2634	4266	0,22	940	303	1243	35,69	11,50	2639	4965	2326	38,20	71,90
E14	M1	F1	G1	81,13	18,87	5598	1302	0,572	18	190	208	0,32	3,39	5598	5933	335	81,00	85,90
E15	M1	F1	G1	81,87	18,13	5649	1251	0,575	15	231	246	0,27	4,09	5655	6098	443	81,80	88,30
E16	M3	F8	G1	52,86	47,14	3647	3253	0,284	1409	649	2058	38,63	17,80	3650	5595	1945	52,80	81,00
E17	M3	F8	G1	45,55	54,45	3143	3757	0,26	1343	66	1409	42,73	2,10	3145	5370	2225	45,50	77,70
E18	M3	F8	G1	47,48	52,52	3276	3624	0,264	804	48	852	24,54	1,47	3280	5470	2190	47,50	79,20
E19	M3	F8	G1	44,94	55,06	3101	3799	0,259	597	55	652	19,25	1,77	3103	5388	2285	44,90	78,00
E20	M1	F4	G1	45,65	54,35	3150	3750	0,261	827	517	1344	26,25	16,41	3151	5206	2055	45,60	75,40
E21	M1	F4	G1	63,33	36,67	4370	2530	0,352	57	258	315	1,30	5,90	4377	5622	1245	63,40	81,40
E22	M1	F4	G1	39,86	60,14	2750	4150	0,236	573	137	710	20,84	4,98	2752	5000	2248	39,80	72,40
E23	M1	F4	G1	90,29	9,71	6230	670	0,727	3	10	13	0,05	0,16	6230	6399	169	90,20	92,60
E24	M1	F4	G1	73,26	26,74	5055	1845	0,432	25	202	227	0,50	4,00	5056	5880	824	73,20	85,10
E25	M1	F4	G1	58,32	41,68	4024	2876	0,323	733	437	1170	18,22	10,86	4027	5622	1595	58,30	81,40
E26	M1	F4	G1	46,7	53,3	3222	3678	0,278	579	32	611	17,97	0,99	3225	5267	2042	46,70	76,20
E27	M1	F4	G1	47,75	52,25	3295	3605	0,275	605	54	659	18,36	1,64	3300	5265	1965	47,80	76,20
E28	M1	F4	G1	83,46	16,54	5759	1141	0,586	3	51	54	0,05	0,89	5768	6128	360	83,50	88,70
E29	M1	F4	G1	49,61	50,39	3423	3477	0,264	530	32	562	15,48	0,94	3427	5336	1909	49,60	77,20
E30	M1	F4	G1	41,72	58,28	2879	4021	0,26	471	52	523	16,36	1,81	2883	5084	2201	41,70	73,60
E31	M2	F5	G2	95,96	4,04	6621	279	0,78	0	1	1	0,00	0,02	6629	6708	79	95,90	97,10
E32	M2	F5	G2	91,23	8,77	6295	605	0,745	0	2	2	0,00	0,03	6295	6441	146	91,10	93,20
E33	M2	F5	G2	90,04	9,96	6213	687	0,738	1	5	6	0,02	0,08	6213	6421	208	89,90	92,90

Continua

IDENTIFICAÇÃO				CALL RATE (CR)				VARIÁVEIS QUALITATIVAS						INFERÊNCIA DE CR				
Amostra	Touro	Doadora	Grupo	CR (%)	MISS (%)	Call (n)	No Call (n)	GC10	ADI (n)	ADO (n)	ERROS (n)	ADI (%)	ADO (%)	Call 1 (n)	Call 2 (n)	Inferidos	CR 1 (%)	CR 2 (%)
E34	M2	F2	G2	84,39	15,61	5823	1077	0,599	4	181	185	0,07	3,11	5823	6195	372	84,30	89,70
E35	M2	F2	G2	87,23	12,77	6019	881	0,672	3	130	133	0,05	2,16	6026	6279	253	87,20	90,90
E36	M2	F2	G2	54,52	45,48	3762	3138	0,334	1156	722	1878	30,73	19,19	3767	5629	1862	54,50	81,50
E37	M2	F2	G2	93,13	6,87	6426	474	0,761	0	13	13	0,00	0,20	6433	6524	91	93,10	94,40
E38	M2	F2	G2	88,68	11,32	6119	781	0,705	1	32	33	0,02	0,52	6127	6314	187	88,70	91,40
E39	M2	F2	G2	87,61	12,39	6045	855	0,674	2	52	54	0,03	0,86	6045	6282	237	87,50	90,90
E40	M2	F2	G2	45,17	54,83	3117	3783	0,289	196	132	328	6,29	4,24	3122	5329	2207	45,20	77,10
E41	M2	F2	G2	84,61	15,39	5838	1062	0,613	10	201	211	0,17	3,44	5838	6101	263	84,50	88,30
E42	M2	F2	G2	84,86	15,14	5855	1045	0,668	1	25	26	0,02	0,43	5856	6108	252	84,80	88,40
E43	M2	F2	G2	68,49	31,51	4726	2174	0,42	28	82	110	0,59	1,74	4733	5878	1145	68,50	85,10
E44	M2	F2	G2	91,93	8,07	6343	557	0,756	1	3	4	0,02	0,05	6344	6456	112	91,80	93,40
E45	M2	F3	G2	59,93	40,07	4135	2765	0,351	106	208	314	2,56	5,03	4140	5732	1592	59,90	83,00
E46	M2	F3	G2	49,42	50,58	3410	3490	0,268	1034	262	1296	30,32	7,68	3413	5500	2087	49,40	79,60
E47	M2	F3	G2	80,41	19,59	5548	1352	0,555	3	152	155	0,05	2,74	5555	6131	576	80,40	88,70
E48	M2	F3	G2	47,93	52,07	3307	3593	0,27	639	186	825	19,32	5,62	3308	5449	2141	47,90	78,90
E49	M2	F6	G2	71,96	28,04	4965	1935	0,453	5	136	141	0,10	2,74	4971	5888	917	71,90	85,20
E50	M2	F5	G2	80,99	19,01	5588	1312	0,579	4	180	184	0,07	3,22	5595	6173	578	81,00	89,30
E51	M2	F6	G2	73	27	5037	1863	0,471	5	119	124	0,10	2,36	5038	5850	812	72,90	84,70
E52	M2	F6	G2	89,04	10,96	6144	756	0,711	0	21	21	0,00	0,34	6151	6328	177	89,00	91,60
E53	M2	F6	G2	88,38	11,62	6098	802	0,727	0	23	23	0,00	0,38	6105	6291	186	88,40	91,10
E54	M2	F6	G2	87,01	12,99	6004	896	0,698	0	163	163	0,00	2,72	6008	6196	188	87,00	89,70
E55	M2	F6	G2	87,2	12,8	6017	883	0,69	0	12	12	0,00	0,20	6017	6223	206	87,10	90,10
E56	M2	F5	G2	51,46	48,54	3551	3349	0,273	317	102	419	8,93	2,87	3553	5497	1944	51,40	79,60
E57	M2	F6	G2	42,28	57,72	2917	3983	0,25	172	515	687	5,90	17,66	2921	5166	2245	42,30	74,80
E58	M2	F6	G2	89,49	10,51	6175	725	0,708	0	7	7	0,00	0,11	6181	6353	172	89,50	92,00
E59	M2	F6	G2	85,33	14,67	5888	1012	0,605	0	34	34	0,00	0,58	5894	6225	331	85,30	90,10
E60	M2	F6	G2	85,99	14,01	5933	967	0,638	0	18	18	0,00	0,30	5933	6220	287	85,90	90,00
E61	M2	F6	G2	88,2	11,8	6086	814	0,705	0	10	10	0,00	0,16	6094	6316	222	88,20	91,40
E62	M2	F6	G2	94,52	5,48	6522	378	0,768	0	1	1	0,00	0,02	6522	6613	91	94,40	95,70
E63	M2	F5	G2	90,97	9,03	6277	623	0,74	1	16	17	0,02	0,26	6285	6443	158	91,00	93,30
E64	M2	F5	G2	51,22	48,78	3534	3366	0,258	124	25	149	3,51	0,71	3538	5523	1985	51,20	79,90

IDENTIFICAÇÃO				CALL RATE (CR)				VARIÁVEIS QUALITATIVAS						INFERÊNCIA DE CR				
Amostra	Touro	Doadora	Grupo	CR (%)	MISS (%)	Call (n)	No Call (n)	GC10	ADI (n)	ADO (n)	ERROS (n)	ADI (%)	ADO (%)	Call 1 (n)	Call 2 (n)	Inferidos	CR 1 (%)	CR 2 (%)
E65	M2	F5	G2	86,23	13,77	5950	950	0,714	0	28	28	0,00	0,47	5959	6170	211	86,20	89,30
E66	M2	F5	G2	90,32	9,68	6232	668	0,718	0	13	13	0,00	0,21	6240	6427	187	90,30	93,00
E67	M2	F5	G2	84,33	15,67	5819	1081	0,646	0	40	40	0,00	0,69	5826	6092	266	84,30	88,20
E68	M2	F5	G3	97,2	2,8	6707	193	0,784	0	0	0	0,00	0,00	6715	6799	84	97,20	98,40
E69	M2	F2	G3	96,44	3,57	6654	246	0,781	2	0	2	0,03	0,00	6654	6731	77	96,30	97,40
E70	M2	F2	G3	96,7	3,3	6672	228	0,787	0	1	1	0,00	0,02	6680	6749	69	96,70	97,70
E71	M2	F2	G3	97,45	2,55	6724	176	0,791	0	1	1	0,00	0,02	6732	6791	59	97,40	98,30
E72	M2	F2	G3	97,16	2,84	6704	196	0,791	0	0	0	0,00	0,00	6704	6754	50	97,00	97,80
E73	M2	F3	G3	92,71	7,29	6397	503	0,744	0	1	1	0,00	0,02	6406	6515	109	92,70	94,30
E74	M2	F3	G3	95,09	4,91	6561	339	0,762	0	4	4	0,00	0,06	6561	6712	151	95,00	97,10
E75	M2	F3	G3	96,03	3,97	6626	274	0,763	0	3	3	0,00	0,05	6635	6752	117	96,00	97,70
E76	M2	F3	G3	95,44	4,57	6585	315	0,774	0	2	2	0,00	0,03	6593	6730	137	95,40	97,40
E77	M2	F5	G3	97,45	2,55	6724	176	0,791	0	0	0	0,00	0,00	6733	6814	81	97,50	98,60
E78	M2	F6	G3	96,36	3,64	6649	251	0,78	0	0	0	0,00	0,00	6649	6737	88	96,20	97,50
E79	M2	F6	G3	96,23	3,77	6640	260	0,787	0	0	0	0,00	0,00	6648	6739	91	96,20	97,50
E80	M2	F6	G3	96,88	3,12	6685	215	0,79	0	0	0	0,00	0,00	6693	6765	72	96,90	97,90
E81	M2	F6	G3	96,74	3,26	6675	225	0,787	0	0	0	0,00	0,00	6683	6750	67	96,70	97,70
E82	M2	F6	G3	96,74	3,26	6675	225	0,786	0	0	0	0,00	0,00	6675	6748	73	96,60	97,70
E83	M2	F6	G3	96,88	3,12	6685	215	0,786	0	0	0	0,00	0,00	6693	6764	71	96,90	97,90
E84	M2	F6	G3	96,91	3,09	6687	213	0,786	0	0	0	0,00	0,00	6687	6757	70	96,80	97,80
E85	M2	F6	G3	97,04	2,96	6696	204	0,79	0	0	0	0,00	0,00	6704	6782	78	97,00	98,20
E86	M2	F1	G3	96,99	3,01	6692	208	0,79	1	12	13	0,02	0,18	6692	6772	80	96,90	98,00
E87	M2	F1	G3	96,59	3,41	6665	235	0,785	1	6	7	0,02	0,09	6665	6756	91	96,50	97,80
E88	M2	F1	G3	96,19	3,81	6637	263	0,761	3	4	7	0,05	0,06	6637	6734	97	96,10	97,50
E89	M2	F5	G3	97,32	2,68	6715	185	0,789	0	1	1	0,00	0,02	6716	6799	83	97,20	98,40
E90	M2	F5	G3	84,88	15,12	5857	1043	0,676	0	4	4	0,00	0,07	5865	6139	274	84,90	88,90
F1	NA	NA	NA	97,67	2,33	6739	161	0,802	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F2	NA	NA	NA	97,12	2,88	6701	199	0,797	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F3	NA	NA	NA	98,8	1,2	6817	83	0,811	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F4	NA	NA	NA	98,52	1,48	6798	102	0,808	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
F5	NA	NA	NA	98,88	1,12	6823	77	0,811	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Conclusão

[illegible]

Programa computacional (script) para inferência de genótipos não determinados

05/10/13

mendelfix.pl

```

1  #!/usr/bin/perl -w
2
3  #mendelFix
4  #Version: v1.0
5  #Date: 10 May 2013
6  #Description: script to correct Mendelian errors in trio designs
7  #Written by: Yuri Tani Utsunomiya - ytutsunomiya@gmail.com
8  #Input: 12-recoded PLINK PED and MAP files
9  #Usage: perl mendelfix.pl pedfile mapfile outfile
10
11 # Copyright (C) 2013 Yuri Tani Utsunomiya
12 #
13 #   This program is free software: you can redistribute it and/or modify
14 #   it under the terms of the GNU General Public License as published by
15 #   the Free Software Foundation, either version 3 of the License, or
16 #   any later version.
17
18 #   This program is distributed in the hope that it will be useful,
19 #   but WITHOUT ANY WARRANTY; without even the implied warranty of
20 #   MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
21 #   GNU General Public License for more details.
22
23 #   You should have received a copy of the GNU General Public License
24 #   along with this program. If not, see <http://www.gnu.org/licenses/>.
25
26
27 #Welcome header
28 print "\n\n\n";
29 print "0-----0\n";
30 print "|               mendelFix v1.0 (May 2013)           |\n";
31 print "|-----|\n";
32 print "|   Correction of SNP Mendelian errors in trio data   |\n";
33 print "|-----|\n";
34 print "|               (C) 2013 Yuri Tani Utsunomiya | GPL v3   |\n";
35 print "0-----0\n\n";
36
37 print "Hello there! Thanks for using mendelFix!\n\n";
38 print "It is easy to cite us:\n";
39 print "Utsunomiya YT, Alonso RV, Carmo AS, Campagnari F, Visintin J, Garcia JF (2013).\n";
40 print "mendelFix: a Perl script for checking Mendelian errors in high density SNP data of trio\n";
41 print "designs.\n";
42 print "http://arxiv.org/.\n";
43 print "Questions or bugs? Contact Yuri: ytutsunomiya@gmail.com\n\n";
44 print "Analysis started: " . (localtime) . "\n\n";
45
46 #Check options
47 die "FATAL ERROR\nYour command line seems to have an incorrect number of provided arguments...\n";
48 Usage: perl mendelfix.pl pedfile mapfile outfile [chromosomes to ignore]\n\n", if (@ARGV < 3);
49
50 my $ped = $ARGV[0];
51 my $map = $ARGV[1];
52 my $pmis = $ARGV[2] . ".pmis";
53 my $stats = $ARGV[2] . ".stats";
54 my $outped = $ARGV[2] . ".ped";
55 my $outmap = $ARGV[2] . ".map";
56
57 #Subroutine prototypes
58 sub getMap;
59 sub getIDs;
60 sub getGenotypes;
61 sub mendelFix;
62
63 #Open outfiles
64 open (PMIS,">$pmis");
65 open (STATS,">$stats");
66 print STATS "ID\tNCALL1\tCR1\t";
67 print STATS "FATHER\tFOCALL\tFOIBS0\tFOIBS1\tFOIBS2\tFOERROR\t";
68 print STATS "MOTHER\tMOCALL\tMOIBS0\tMOIBS1\tMOIBS2\tMOERROR\t";

```

05/10/13

mendelfix.pl

```

68 print STATS "TRIOCALL\tADI\tPADI\tADO\tPADO\tNERROR\tPPCERROR\t";
69 print STATS "NFI\tNCALL2\tCR2\n";
70 open (OUTPED,">$outped");
71 open (OUTMAP,">$outmap");
72
73
74 #Main script
75 my ($chr, $snp, $cM, $bp) = getMap($map);
76 my @chr = @$chr;
77 my @snp = @$snp;
78 my @cM = @$cM;
79 my @bp = @$bp;
80 my %seen={};
81 @uniq = grep { ! $seen{$_}++ } @chr;
82 print "Read SNP map information from ".$map.". Found ".$#snp+1." SNPs in ".$#uniq+1."
      chromosomes.\n";
83 if($#ARGV == 3){
84     print "Chromosome ".$ARGV[3]." will not be checked for Mendelian errors.\n";
85 }elsif($#ARGV > 3){
86     @chr_out=@ARGV[3..$#ARGV];
87     print "The following chromosomes will not be checked for Mendelian errors:\n".join("
      ",@chr_out[0..$#chr_out-1])) and ";
88     print $chr_out[$#chr_out]."\n";
89 }else{
90     print "All chromosomes will be analyzed.\n";
91 }
92 my ($fam, $id, $sire, $dam, $sex, $phen) = getIDs($ped,$#snp+1);
93 my @fam = @$fam;
94 my @id = @$id;
95 my @sire = @$sire;
96 my @dam = @$dam;
97 my @sex = @$sex;
98 my @phen = @$phen;
99 print "Read pedigree data from ".$ped.". Found ".$#id+1." individuals.\n";
100 print "Individuals with missing parent will be reported in ".$pmiss.", and will not be fixed.\n";
101 print "Excuting the main algorithm...\n";
102 print "1st step: offspring genotypes with Mendelian errors will be set to missing.\n";
103 print "2nd step: parent information will be used to infer missing genotypes in the offspring
      (following Mendel's 1st law).\n";
104 print "Summary statistics will be reported in ".$stats.". \n";
105 for ($count = 0; $count <= $#id; $count++){
106     my $sire_index = "";
107     my $dam_index = "";
108     if($sire[$count] eq "0" or $dam[$count] eq "0"){
109         print PMISS $id[$count]."\n";
110         my ($id_geno,$id_cr,$id_ncalls) = getGenotypes($ped,$count+1);
111         $id_geno=~s/ //g;
112         my @a=($id_geno=~m/./g);
113         $id_geno = join(" ",@a);
114         print OUTPED $fam[$count]." ".$id[$count]." ".$sire[$count]." ".$dam[$count]." ".$sex[$count]."
      ".$phen[$count]." ".$id_geno."\n";
115     }else{
116         my ($sire_index) = grep {$id[$_] eq $sire[$count]}(0..$#id);
117         my ($dam_index) = grep {$id[$_] eq $dam[$count]}(0..$#id);
118         if(defined $sire_index and defined $dam_index and $sire_index ne "" and $dam_index ne ""){
119             my ($id_geno,$id_cr,$id_ncalls) = getGenotypes($ped,$count+1);
120             my ($sire_geno,$sire_cr, $sire_ncalls) = getGenotypes($ped,$sire_index+1);
121             my ($dam_geno,$dam_cr, $dam_ncalls) = getGenotypes($ped,$dam_index+1);
122             my ($id_geno_fixed,$id_cr_fixed,$id_ncalls_fixed,$id_adi,$id_ado,$id_nfixed,$id_nerrors,
      $fibs_zero,$fibs_one,$fibs_two,$mibs_zero,$mibs_one,$mibs_two,$trio_ncalls) =
      mendelFix($sire_geno,$dam_geno,$id_geno,\@chr,\@chr_out);
123             if($id_ncalls >= 1){
124                 $fo_ncalls=$fibs_zero+$fibs_one+$fibs_two;
125                 if($fo_ncalls < 1){
126                     $fibs_zero="NA";
127                     $fibs_one="NA";
128                     $fibs_two="NA";
129                 }
130                 $mo_ncalls=$mibs_zero+$mibs_one+$mibs_two;
131                 if($mo_ncalls < 1){

```

05/10/13

mendelfix.pl

```

132     $mibs_zero="NA";
133     $mibs_one="NA";
134     $mibs_two="NA";
135 }
136 if($fibs_zero ne "NA"){
137     $foerror=sprintf("%e",$fibs_zero/$fo_ncalls);
138 }else{
139     $foerror="NA";
140 }
141 if($mibs_zero ne "NA"){
142     $moerror=sprintf("%e",$mibs_zero/$mo_ncalls);
143 }else{
144     $moerror="NA"
145 }
146 if($trio_ncalls >= 1){
147     $id_p_adi=sprintf("%e",$id_adi/$trio_ncalls);
148     $id_p_ado=sprintf("%e",$id_ado/$trio_ncalls);
149     $id_errorrate=sprintf("%e",$id_nerrors/$trio_ncalls);
150 }else{
151     $id_adi="NA";
152     $id_p_adi="NA";
153     $id_ado="NA";
154     $id_p_ado="NA";
155     $id_errorrate="NA";
156     $id_nerrors="NA";
157 }
158 }else{
159     $fo_ncalls=0;
160     $fibs_zero="NA";
161     $fibs_one="NA";
162     $fibs_two="NA";
163     $foerror="NA";
164     $mo_ncalls=0;
165     $mibs_zero="NA";
166     $mibs_one="NA";
167     $mibs_two="NA";
168     $moerror="NA";
169     $id_adi="NA";
170     $id_p_adi="NA";
171     $id_ado="NA";
172     $id_p_ado="NA";
173     $id_nerrors="NA";
174     $id_errorrate="NA";
175 }
176 print STATS $id[$count]."\t".$id_ncalls."\t".$id_cr."\t";
177 print STATS $sire[$count]."\t".$fo_ncalls."\t".$fibs_zero."\t".$fibs_one."\t".$fibs_two."\t".
    $foerror."\t";
178 print STATS $dam[$count]."\t".$mo_ncalls."\t".$mibs_zero."\t".$mibs_one."\t".$mibs_two."\t".
    $moerror."\t";
179 print STATS $trio_ncalls."\t".$id_adi."\t".$id_p_adi."\t".$id_ado."\t".$id_p_ado."\t".
    $id_nerrors."\t".$id_errorrate."\t";
180 print STATS $id_nfixed."\t".$id_ncalls_fixed."\t".$id_cr_fixed."\n";
181 print OUTPED $fam[$count]. " ".$id[$count]. " ".$sire[$count]. " ".$dam[$count]. " ".
    $sex[$count]. " ".$phen[$count]. " ".$id geno_fixed."\n";
182 }
183 }
184 }
185 print "Wrote fixed genotypes to ".$outped."\n";
186 for ($count = 0; $count <= $#snp; $count++){
187     print OUTMAP $chr[$count]."\t".$snp[$count]."\t".$cM[$count]."\t".$bp[$count]."\n";
188 }
189 print "Wrote new mapfile to ".$outmap."\n";
190
191 #Close outfiles
192 close PMISS;
193 close STATS;
194 close OUTPED;
195 close OUTMAP;
196
197 #Goodbye message

```

05/10/13

mendelfix.pl

```

198 print "\n\nAnalysis done: " . (localtime) . "\nSee you!!!!\n\n\n";
199
200
201 #Subroutines used
202 sub getMap{
203     my @line="";
204     my @chr="";
205     my @snp="";
206     my @cM="";
207     my @bp="";
208     open (IN, $_[0]);
209     while(<IN>){
210         $_=~s/\n//;
211         $_=~s/\r//;
212         next if($_ eq "");
213         @line=split(/\t/, $_);
214         if ($#line != 3){
215             print "\nFATAL ERROR\n";
216             print "Your map file " . $_[0] . " does not seem to have 4 tab-delimited columns at line " . $_ . ":
217             \n\" . $_ . "\n\n";
218             die "\nmendelFix will now halt!\n\n\n";
219         }
220         push(@chr,$line[0]);
221         push(@snp,$line[1]);
222         push(@cM,$line[2]);
223         push(@bp,$line[3]);
224     }
225     @chr = @chr[1..$#chr];
226     @snp = @snp[1..$#snp];
227     @cM = @cM[1..$#cM];
228     @bp = @bp[1..$#bp];
229     close IN;
230     return(\@chr,\@snp,\@cM,\@bp);
231 }
232
233 sub getIDs{
234     my $ncol = ((2*$_[1])+6);
235     my @line="";
236     my @fam="";
237     my @id="";
238     my @sire="";
239     my @dam="";
240     my @sex="";
241     my @phen="";
242     open (IN, $_[0]);
243     while(<IN>){
244         $_=~s/\n//;
245         $_=~s/\r//;
246         next if($_ eq "");
247         @line=split(/ /, $_);
248         if($#line != $ncol-1){
249             print "\nFATAL ERROR\n";
250             print "Your ped file " . $_[0] . " does not seem to have " . $ncol . " (2*$_[1]. SNPs + 6) space-
251             delimited columns at line " . $_ . ":
252             \n\" . substr($_, 0, 50) . "... \n\n";
253             die "\nmendelFix will now halt!\n\n\n";
254         }
255         push(@fam,$line[0]);
256         push(@id,$line[1]);
257         push(@sire,$line[2]);
258         push(@dam,$line[3]);
259         push(@sex,$line[4]);
260         push(@phen,$line[5]);
261     }
262     @fam = @fam[1..$#fam];
263     @id = @id[1..$#id];
264     @sire = @sire[1..$#sire];
265     @dam = @dam[1..$#dam];
266     @sex = @sex[1..$#sex];
267     @phen = @phen[1..$#phen];

```


05/10/13

mendelfix.pl

```

266 close IN;
267 return(\@fam,\@id,\@sire,\@dam,\@sex,\@phen);
268 }
269
270 sub getGenotypes{
271 my @geno="";
272 my $a="";
273 my $nsnps="";
274 my $ncalls="";
275 my $genotypes="";
276 my $callrate="";
277 open (IN, $_[0]);
278 while(<IN>){
279     if($_ == $_[1]){
280         $_ =~ s/\n//;
281         $_ =~ s/\r//;
282         @geno=split(/ /,$_);
283         @geno = @geno[6..$#geno];
284         $a = join(" ",@geno);
285         @geno=( $a =~ m/..g );
286         if (grep{!/(11|12|22|00)/, @geno}){
287             print "\nFATAL ERROR\n";
288             print "Your ped file \"$_[0].\" seems to have an unsupported allele coding at line \"$_\":\n";
289             print "\"".substr($_, 0, 50)."...\n";
290             print "Coding should be 0 = missing, 1 = reference (or minor) allele, and 2 = alternative (or
major) allele\n";
291             print "Tip: you can easily obtain this type of coding using the --recodel2 option in
PLINK\n";
292             die "\nmendelFix will now halt!\n\n\n\n";
293         }
294         my @call = grep { $geno[$_] ne "00" }(0..$#geno);
295         $nsnps=$#geno+1;
296         $ncalls=$#call+1;
297         $callrate=sprintf("%.3f",$ncalls/$nsnps);
298         $genotypes = join(" ",@geno);
299     }
300 }
301 close IN;
302 return($genotypes,$callrate,$ncalls);
303 }
304
305
306 sub mendelFix{
307 my @snpcr=@{$_[3]};
308 my @snpcout=@{$_[4]};
309 my @dad="";
310 my @mom="";
311 my @son="";
312 my $a="";
313 my $nsnps="";
314 my $ncalls="";
315 my $fixed_genotypes="";
316 my $callrate="";
317 my $nerrors=0;
318 my $adi=0;
319 my $ado=0;
320 my $nfixed=0;
321 my $p_fixed=0;
322 my $analyze = 1;
323 my $fibs_zero = 0;
324 my $fibs_one = 0;
325 my $fibs_two = 0;
326 my $mibs_zero = 0;
327 my $mibs_one = 0;
328 my $mibs_two = 0;
329 my $trio_ncalls = 0;
330 @dad = split(/ /,$_[0]);
331 @mom = split(/ /,$_[1]);
332 @son = split(/ /,$_[2]);
333 for ($c = 0; $c <= $#son; $c++){

```

05/10/13

mendelfix.pl

```

334     if(@snp_out and $snp_chr[$c] == @snp_out){
335         $analyze=0;
336     }else{
337         $analyze=1;
338     }
339     if($analyze == 1){
340         if($son[$c] ne "00"){
341             #Father-Offspring IBS checks
342             $fibs_two++, if ($dad[$c] eq "11" and $son[$c] eq "11");
343             $fibs_one++, if ($dad[$c] eq "11" and $son[$c] eq "12");
344             $fibs_zero++, if ($dad[$c] eq "11" and $son[$c] eq "22");
345             $fibs_two++, if ($dad[$c] eq "22" and $son[$c] eq "22");
346             $fibs_one++, if ($dad[$c] eq "22" and $son[$c] eq "12");
347             $fibs_zero++, if ($dad[$c] eq "22" and $son[$c] eq "11");
348             $fibs_two++, if ($dad[$c] eq "12" and $son[$c] eq "12");
349             $fibs_one++, if ($dad[$c] eq "12" and $son[$c] eq "11");
350             $fibs_one++, if ($dad[$c] eq "12" and $son[$c] eq "22");
351             $son[$c] = "00", if($dad[$c] eq "11" and $son[$c] eq "22" and $mom[$c] eq "00");
352             $son[$c] = "00", if($dad[$c] eq "22" and $son[$c] eq "11" and $mom[$c] eq "00");
353             #Mother-Offspring IBS checks
354             $mibs_two++, if ($mom[$c] eq "11" and $son[$c] eq "11");
355             $mibs_one++, if ($mom[$c] eq "11" and $son[$c] eq "12");
356             $mibs_zero++, if ($mom[$c] eq "11" and $son[$c] eq "22");
357             $mibs_two++, if ($mom[$c] eq "22" and $son[$c] eq "22");
358             $mibs_one++, if ($mom[$c] eq "22" and $son[$c] eq "12");
359             $mibs_zero++, if ($mom[$c] eq "22" and $son[$c] eq "11");
360             $mibs_two++, if ($mom[$c] eq "12" and $son[$c] eq "12");
361             $mibs_one++, if ($mom[$c] eq "12" and $son[$c] eq "11");
362             $mibs_one++, if ($mom[$c] eq "12" and $son[$c] eq "22");
363             $son[$c] = "00", if($mom[$c] eq "11" and $son[$c] eq "22" and $dad[$c] eq "00");
364             $son[$c] = "00", if($mom[$c] eq "22" and $son[$c] eq "11" and $dad[$c] eq "00");
365             #Set to missing ADI/ADO occurrences
366             $trio_ncalls++, if($dad[$c] ne "00" and $mom[$c] ne "00");
367             if($dad[$c] eq "11" and $mom[$c] eq "11" and $son[$c] eq "12"){
368                 $son[$c] = "00";
369                 $adi++;
370             }elseif($dad[$c] eq "11" and $mom[$c] eq "11" and $son[$c] eq "22"){
371                 $son[$c] = "00";
372                 $adi++;
373             }elseif($dad[$c] eq "22" and $mom[$c] eq "22" and $son[$c] eq "12"){
374                 $son[$c] = "00";
375                 $adi++;
376             }elseif($dad[$c] eq "22" and $mom[$c] eq "22" and $son[$c] eq "11"){
377                 $son[$c] = "00";
378                 $adi++;
379             }elseif($dad[$c] eq "11" and $mom[$c] eq "12" and $son[$c] eq "22"){
380                 $son[$c] = "00";
381                 $ado++;
382             }elseif($dad[$c] eq "11" and $mom[$c] eq "22" and $son[$c] eq "22"){
383                 $son[$c] = "00";
384                 $ado++;
385             }elseif($dad[$c] eq "12" and $mom[$c] eq "11" and $son[$c] eq "22"){
386                 $son[$c] = "00";
387                 $ado++;
388             }elseif($dad[$c] eq "22" and $mom[$c] eq "11" and $son[$c] eq "22"){
389                 $son[$c] = "00";
390                 $ado++;
391             }elseif($dad[$c] eq "22" and $mom[$c] eq "12" and $son[$c] eq "11"){
392                 $son[$c] = "00";
393                 $ado++;
394             }elseif($dad[$c] eq "22" and $mom[$c] eq "11" and $son[$c] eq "11"){
395                 $son[$c] = "00";
396                 $ado++;
397             }elseif($dad[$c] eq "12" and $mom[$c] eq "22" and $son[$c] eq "11"){
398                 $son[$c] = "00";
399                 $ado++;
400             }elseif($dad[$c] eq "11" and $mom[$c] eq "22" and $son[$c] eq "11"){
401                 $son[$c] = "00";
402                 $ado++;
403         }

```


05/10/13

mendelfix.pl

```

404     }
405     if($son[$c] eq "00"){
406         if($dad[$c] eq "12" or $mom[$c] eq "12"){
407             next;
408         }elsif($dad[$c] eq "11" and $mom[$c] eq "11"){
409             $son[$c]="11";
410             $nfixed++;
411         }elsif($dad[$c] eq "22" and $mom[$c] eq "22"){
412             $son[$c]="22";
413             $nfixed++;
414         }elsif($dad[$c] eq "11" and $mom[$c] eq "22"){
415             $son[$c]="12";
416             $nfixed++;
417         }elsif($dad[$c] eq "22" and $mom[$c] eq "11"){
418             $son[$c]="12";
419             $nfixed++;
420         }
421     }
422 }
423 }
424 my @call = grep {$son[$_] ne "00"}(0..$#son);
425 $nsnps=$#son+1;
426 $ncalls=$#call+1;
427 $nerrors=$adi+$ado;
428 $callrate=sprintf("%.3f",$ncalls/$nsnps);
429 $a=join(" ",@son);
430 @a=($a=~m/./g);
431 $fixed_genotypes = join(" ",@a);
432 return($fixed_genotypes,$callrate,$ncalls,$adi,$ado,$nfixed,$nerrors,$fibs_zero,$fibs_one,
433 $fibs_two,$mibs_zero,$mibs_one,$mibs_two,$trio_ncalls);
434 }
435
436 exit;
437
438

```