

MARCADORES MOLECULARES E SUAS APLICAÇÕES NO MELHORAMENTO ANIMAL

Luciana Correia de Almeida Regitano¹, Gisele Batista Veneroni²

¹Pesquisadora, Embrapa Pecuária Sudeste, Rodovia Washington Luiz, km 234, Caixa Postal 339, CEP 13560-970, São Carlos, SP, Brasil. Endereço eletrônico: <luciana@cnpse.embrapa.br>. ²Bióloga, MsC, Universidade Federal de São Carlos, Programa de Pós-graduação em Genética e Evolução.

Introdução

Marcadores genéticos têm sido elementos fundamentais nos estudos de segregação de caracteres hereditários, na análise do comportamento de genes em populações e na reconstrução da história evolutiva de populações, entre outras aplicações.

Desde a descoberta do DNA, a genética molecular experimentou extraordinário avanço, que passou pelo desenvolvimento de métodos de análise da estrutura e da função do material genético, e de equipamentos com capacidade para análise automatizada de grande quantidade de amostras, até o desenvolvimento de métodos estatísticos e de ferramentas de informática, resultando na ciência conhecida como genômica.

O sequenciamento do genoma de grande número de espécies produziu volume razoável de informações, que, aliadas à evolução metodológica, permitiram o desenvolvimento de uma variedade de marcadores genéticos moleculares.

O sequenciamento do genoma do bovino teve sua primeira versão em julho de 2005, ano em que um mapa genético com aproximadamente 4.000 marcadores foi também publicado.

Em 2008, painéis constituídos de mais de 50 mil marcadores foram empregados na produção de um mapa da diversidade dos bovinos. Painéis de alta densidade de marcadores de outras espécies domésticas estão também disponíveis, tais como o *OvineSNP50 BeadChip* – da empresa norte-americana Illumina, com mais de 50 mil marcadores de polimorfismos de um nucleotídeo (*single nucleotide*

polymorphism- SNP), que cobrem intervalos de aproximadamente 46 kb do genoma dos ovinos –, o de suínos (~ 60,000 SNPs) e o de equinos (~50 mil SNPs).

Com a disponibilidade dessas ferramentas, a geração maciça de dados populacionais tem permitido compreender a estrutura genômica de populações, tais como extensão dos blocos de desequilíbrio de ligação (SELLNER et al., 2007), a variabilidade entre e dentro de populações (BOVINE HAPMAP CONSORTIUM et al., 2009) e os processos de domesticação pelos quais as espécies passaram.

As informações geradas pelos projetos de seqüenciamento do genoma permitiram, por exemplo, o mapeamento de mais de 1.600 mutações responsáveis por características de herança mendeliana em seres humanos. Em suínos e em bovinos, dezenas de doenças hereditárias tiveram seu mecanismo molecular desvendado. Permanece, porém, o desafio de elucidar os mecanismos determinantes da variação genética de características mais complexas, entre as quais figura a maioria das características de interesse econômico dos animais domésticos (GEORGES e ANDERSSON, 2003). Geralmente, essas características são controladas por muitos genes, que podem resultar em complexas interações alélicas e interações não-alélicas, e são influenciadas pelo ambiente. Diversos estudos têm demonstrado a possibilidade de mapear genes ou blocos de genes adjacentes que influenciam uma característica quantitativa, denominados *quantitative trait loci* (QTL).

Além da análise do genoma estrutural, a análise de expressão em escala genômica, também denominada genética genômica ou mapeamento de transcriptoma, tem permitido a identificação de variação no padrão de expressão de conjuntos de genes expressos; a análise dessa variação na expressão, aliada à análise de marcadores, possibilita identificar regiões do genoma responsáveis pelas mudanças na expressão desses genes (KADARMIDEEN et al., 2006). A integração de informações estruturais e funcionais deverá permitir melhor conhecimento do controle genético das características de interesse econômico e auxiliar as decisões do melhoramento.

Marcadores moleculares

Desde a redescoberta dos princípios de Mendel, no início do século XX, o foco da atenção dos geneticistas passou a ser o gene como unidade fundamental da variação biológica. Com o desenvolvimento da genética de populações, surgiu o

conceito de utilização de genes individuais como marcadores, com a finalidade de fazer inferências sobre as características de uma população, tais como o conteúdo de variabilidade, os padrões de migração, a seleção e a deriva genética.

Marcadores genéticos são caracteres de herança mendeliana simples, em que o padrão fenotípico individual pode ser experimentalmente determinado, de modo a permitir que a segregação do locus marcador responsável pelo caráter seja acompanhada. A análise de segregação requer a existência de pelo menos duas formas alélicas, ou seja, a existência de polimorfismo no locus marcador.

Inicialmente, as características disponíveis para esse tipo de análise eram as mutações que produziam alterações morfológicas, como o nanismo, a ausência de asas em *Drosophila* ou a ausência de pelos em camundongos. Entretanto, tais mutações são pouco frequentes nas populações naturais, em que a maior parte da variação genética é de caráter contínuo (TANKSLEY, 1993). Além disso, mutações que provocam alterações fenotípicas drásticas, como as citadas, geralmente comprometem a adaptação do indivíduo, o que reduz sua utilidade em estudos de comparação entre populações.

As primeiras contribuições ao estudo de marcadores genéticos em animais foram as descobertas dos polimorfismos de antígenos eritrocitários (STORMONT e CUMLEY, 1943). A análise de marcadores foi ampliada com o desenvolvimento de técnicas de eletroforese de proteínas associadas a métodos de coloração histoquímica, as quais permitiram que a variação genética das proteínas passasse a ser estudada (SMITHIES, 1955).

Porém, com o desenvolvimento das técnicas de análise de DNA, a possibilidade de analisar variações individuais nas sequências de DNA, independentemente de corresponderem a um peptídeo ou não, permitiu o desenvolvimento de diversos tipos de marcadores moleculares. Nesse caso, cada segmento do DNA constitui um locus e os padrões correspondentes às variações de sequência nesse segmento são fenótipos moleculares. A utilização de muitos desses marcadores se baseia na técnica de reação em cadeia da polimerase (PCR), que permitiu a automação e a simplificação das etapas de obtenção dos padrões moleculares. Nessa técnica, segmentos de DNA específicos são replicados *in vitro* e isso resulta na produção de milhares de cópias da sequência desejada, em quantidade suficiente para permitir a pronta visualização do DNA, sem a necessidade de métodos indiretos. Além disso, os produtos da PCR podem ser utilizados na construção de sondas e de moléculas recombinantes.

Os diversos tipos de marcadores moleculares revelam variações de sequência oriundas de diferentes mecanismos de mutação e a escolha do tipo de marcador a ser utilizado depende essencialmente da aplicação, mas deve levar em conta dificuldade técnica, custo para a obtenção do número necessário de genótipos, informatividade, distribuição pelo genoma e tipo de interação alélica.

Polimorfismos de comprimento de fragmentos de restrição

A ocorrência de variação individual no número e no tamanho dos fragmentos formados pela digestão do DNA com uma enzima de restrição foi demonstrada por Grodzicker et al. (1975), em adenovírus. Essa variação é o resultado de mutações de ponto, que eliminam ou criam sítios de restrição para determinada enzima, podendo também originar-se de eventos de inserção ou de deleção entre dois sítios de restrição adjacentes. Essa variação foi denominada polimorfismo de comprimento de fragmentos de restrição (*restriction fragment length polymorphism* – RFLP).

O método desenvolvido para adenovírus não produziu resultados satisfatórios em eucariotos superiores, pois, em virtude do tamanho do genoma dessas espécies, o número de fragmentos resultantes da digestão com enzimas de restrição é muito grande. Conseqüentemente, tentativas de separar esses fragmentos por eletroforese em gel produziram um rastro em vez de bandas discretas. A solução para esse problema surgiu com o desenvolvimento da técnica de *Southern blot* (SOUTHERN, 1975), na qual os fragmentos separados por eletroforese são transferidos para um suporte sólido – usualmente uma membrana de nitrocelulose ou de náilon –, imobilizados e desnaturados. Uma vez que a molécula de DNA tem tendência a formar fitas duplas complementares, um segmento de DNA de fita simples é utilizado como sonda. A sonda é marcada com um isótopo radioativo, como o ^{32}P , ou, mais recentemente, com moléculas antigênicas que são detectadas por anticorpos associados a enzimas que catalisam reações cromogênicas ou quimiluminescentes. Como resultado, apenas os fragmentos complementares à sequência da sonda são revelados.

As sondas utilizadas para a obtenção de marcadores do tipo RFLP podem ser ou não originadas de regiões codificadoras. O emprego de sonda de DNA complementar ao RNA mensageiro (cDNA) é o mais comum e apresenta a vantagem de não conter sequências repetitivas. Sondas anônimas, constituídas por fragmentos de DNA genômico formados ao acaso, possibilitam a rápida obtenção de novos

marcadores, além de permitirem a análise de regiões não codificadoras do genoma. Porém, em virtude da abundância de DNA repetitivo no genoma dos eucariotos, quando o objetivo é a obtenção de marcadores unilocais, essas sondas devem ser submetidas a um processo de seleção, de forma a garantir que apenas aquelas que representam sequências de cópia única sejam empregadas.

Os marcadores de RFLP são co-dominantes e muitos dos loci descritos são dialélicos, quer originados de mutações de ponto quer de mutações estruturais. Marcadores dialélicos são bastante limitados para o mapeamento genético, em consequência do número de cruzamentos informativos. O número desses cruzamentos é função do número de progenitores que segregam para o marcador (heterozigotos) e do número de progênes às quais se pode atribuir um dos alelos ao progenitor. Essa relação pode ser quantificada pelo valor de conteúdo de informação polimórfica (PIC), definido por Botstein et al. (1980) como

$$PIC = \left(1 - \sum_{i=1}^{n-1} p_i^2\right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 p_i^2 p_j^2.$$

Nesta equação, o primeiro membro corresponde à proporção de indivíduos heterozigotos na população, pressupondo-se equilíbrio de Hardy-Weinberg. Essa proporção é também definida como diversidade gênica (WEIR, 1996). Dessa diversidade se subtrai a proporção esperada de progênes não informativas, isto é, aquelas em que não é possível diferenciar o alelo paterno do materno. Um exemplo seria o acasalamento de dois indivíduos heterozigotos idênticos ($A_1A_2 \times A_1A_2$): em aproximadamente metade de suas progênes, as progênes heterozigotas, o alelo A_1 poderia ter origem tanto paterna quanto materna. O mesmo raciocínio seria válido para a análise da origem do alelo A_2 .

Outras desvantagens dos marcadores de RFLP são o grau de dificuldade técnica e a presença de alelos raros, os quais fazem com que muitas populações apresentem padrão monomórfico. Como foi mencionado, a técnica original depende tanto do desenvolvimento e da seleção de sondas com base em bibliotecas de genoma ou de cDNA como da identificação das combinações de sonda e de endonuclease de restrição que revelam polimorfismos.

Para a análise de RFLP com a PCR, as extremidades da sonda que revelou o polimorfismo são sequenciadas e as informações são utilizadas para a síntese de pequenas moléculas de DNA complementares às extremidades da região que se

pretende amplificar. Essas moléculas, denominadas *primers*, funcionam como ponto de origem para a replicação do DNA *in vitro* (FERREIRA e GRATTAPAGLIA, 1995). Quando a posição do sítio polimórfico é conhecida, por exemplo, dentro de um gene bem caracterizado, os *primers* podem ser desenvolvidos por meio de sequências publicadas ou catalogadas em bases de dados, tais como o *GenBank* (NCBI, 2009). A região que contém os sítios polimórficos de restrição é amplificada e os fragmentos produzidos pela digestão do produto de amplificação podem ser então analisados em géis de agarose. A grande vantagem dos marcadores de RFLP para estudos de população é a reprodutibilidade dos padrões fenotípicos; isso permite a designação de alelos correspondentes a cada padrão de fragmentos de restrição, que podem ser facilmente comparados entre as populações. São muito utilizados como método diagnóstico, quando, por exemplo, um locus de RFLP é associado a uma característica monogênica, tal como uma doença hereditária.

Microssatélites

Os genomas dos eucariotos abrigam grande quantidade de DNA repetitivo, classificado de acordo com o número de nucleotídeos e a complexidade da sequência que compõe as repetições. As diferentes classes são também caracterizadas por comportamento típico quanto à distribuição no genoma e quanto ao mecanismo envolvido em suas origens (JOBSE et al., 1995).

A variabilidade de sequências repetitivas foi utilizada para a produção de perfis característicos de cada indivíduo pela técnica de *DNA fingerprinting*, na qual um complexo padrão de bandas é obtido da hibridização de sondas homólogas à regiões repetitivas. Essa técnica é de grande utilidade na identificação de indivíduos ou de linhagens e em testes de paternidade. Entretanto, a aplicação em estudos de população pode ser dificultada pela elevada variação intrapopulacional, não sendo possível estabelecer um perfil característico da população, principalmente naquelas populações que possuem base genética ampla. Além disso, sua natureza multilocal não permite admitir, sem detalhada análise de segregação, que fragmentos de mesma taxa de migração sejam alelos idênticos de um mesmo locus.

Avanço importante na obtenção de marcadores hipervariáveis unilocais veio da utilização de uma classe de DNA repetitivo, as sequências de microssatélites (*simple sequence repeats* – SSR). Os microssatélites são caracterizados por repetições em tandem de um mono-, di-, tri- ou tetranucleotídeo, localizadas dentro de regiões de

sequência única. Cada bloco de repetições tem geralmente menos do que 100 pares de nucleotídeos (TAUTZ, 1989). Do mesmo modo que outras regiões repetitivas do genoma, a variação do número de repetições em cada locus é provavelmente resultante de erros no deslocamento da DNA-polimerase durante a replicação do DNA.

Os microssatélites poli(G) e poli(A) são os mais simples, enquanto os poli(GT) são os mais frequentes, aparecendo em aproximadamente 5 a 10 x 10⁴ loci individuais no genoma dos mamíferos. Muitos outros microssatélites foram identificados até o momento e é possível que qualquer sequência de poucos nucleotídeos represente um microssatélite no genoma de eucariotos.

Os microssatélites podem ser amplificados de maneira específica pela técnica de PCR, utilizando *primers* que contenham parte da sequência flanqueadora. A amplificação resulta em produtos de diferente tamanho, em função do número de cópias da sequência repetitiva delimitada pelos *primers*. A maior dificuldade técnica está relacionada à identificação dos genótipos, principalmente nos loci de microssatélites constituídos por repetições de um a dois nucleotídeos, nos quais dois alelos adjacentes diferem entre si por apenas um ou dois pares de bases, respectivamente. Há necessidade, portanto, de separar os produtos de PCR por eletroforese em gel de poliacrilamida de alta definição. A identificação dos produtos pode se valer de três estratégias:

1) Utilização de um dos *primers* marcados com isótopo radioativo. Essa marcação pode ser feita pela adição de [³²P] catalisada pela enzima T4-polinucleotideoquinase.

2) Marcação de um dos *primers* com fluoróforos e leitura em equipamento sequenciador automático.

3) Coloração do gel por impregnação com prata.

O desenvolvimento de marcadores de microssatélites pode ser feito pela análise de sequências contidas em bancos de dados, localizando-se aquelas que contêm SSR e delineando-se *primers* para a região que flanqueia a repetição. Outra estratégia é a seleção de fragmentos contidos em uma biblioteca genômica. A seleção dos clones que contêm microssatélites pode ser realizada pela hibridização das colônias que constituem a biblioteca com um oligonucleotídeo sintético marcado, composto pela repetição que se procura, de adenina (A), citosina (C), guanina (G) ou timina (T), por exemplo (dG-dT)₁₅ ou (dA-dA-dT)₁₀, para localizar as repetições de dinucleotídeos CA ou de trinucleotídeo TTA, respectivamente. Algumas metodologias utilizam o prévio

enriquecimento das bibliotecas por meio de métodos que promovam a clonagem preferencial de fragmentos formados por sequências repetitivas.

Os clones positivos, isto é, aqueles que ficaram marcados após a hibridização, são isolados e sequenciados para o desenvolvimento de *primers* complementares às sequências flanqueadoras das SSR.

Uma vez que os *primers* utilizados são complementares às sequências de cópia única, obtém-se marcadores unilocais, altamente polimórficos e de herança co-dominante. Esses atributos são de grande valor para a construção de mapas genéticos.

Distorções da segregação podem surgir em decorrência da “expansão” ou da “contração” do microssatélite durante a meiose. A frequência desses eventos é da ordem de 10^{-4} a 10^{-5} por locus, por gameta (HOLMES, 1994). Entretanto, em presença de alterações do sistema de reparo, a taxa de mutação pode aumentar de 100 a 1.000 vezes. A instabilidade resultante tem sido associada com o desenvolvimento de alguns tipos de câncer (SIMPSON, 1996) e com doenças hereditárias, como a síndrome do cromossomo X frágil.

Ao considerar a frequência dessas sequências e a média do tamanho do genoma de mamífero de 3×10^9 pares de nucleotídeos, seria possível construir mapas genéticos das espécies domésticas com aproximadamente 10.000 locus de microssatélites. Em bovinos e em outras espécies domésticas, como suínos, galinhas, ovinos, caprinos e equinos, mapas genéticos saturados foram obtidos com marcadores de microssatélites (VIGNAL et al., 2002).

Apesar da distribuição regular dos microssatélites, a padronização da designação dos alelos é complexa, uma vez que na maioria dos locus, a diferença entre alelos é de apenas dois pares de bases e as estimativas de tamanho dos fragmentos pode variar entre equipamentos e entre técnicas. Essa variação dificulta a comparação de frequências alélicas obtidas em experimentos independentes.

Polimorfismos de um nucleotídeo

Os avanços nos sistemas de sequenciamento automatizado de ácidos nucleicos têm permitido que as sequências de diferentes indivíduos sejam comparadas. Dessa forma, variações individuais, resultantes de mutações de ponto, podem ser identificadas. Essas mutações podem ser substituições, deleções e adições de nucleotídeos, que, uma vez caracterizadas, constituem os marcadores

de polimorfismos de nucleotídeos (*single nucleotide polymorphisms* – SNPs). Apesar de tanto substituições quanto inserções ou deleções serem consideradas SNPs, é importante ter em mente que os dois tipos se originam de mecanismos de mutação distintos (VIGNAL et al., 2002). Apesar de haver muito mais possibilidades de transversões (substituições de uma purina por uma pirimidina ou vice-versa) do que transições (substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina), a razão observada entre o número de transições sobre o de transversões varia de 1,4 em seres humanos a 2,36 em galinhas (VIGNAL et al., 2002).

Os SNPs são marcadores atraentes para a análise genética, por se encontrarem em praticamente qualquer região do genoma ou em qualquer sequência de interesse, por exemplo, em éxons de genes; isso pode ter implicações diretas nas funções da proteína correspondente. Outras características importantes são a maior estabilidade, quando comparada à dos marcadores de microssatélites, além da possibilidade de automação. Essa última é uma vantagem considerável em situações que exigem a análise de grande número de indivíduos e de marcadores, como no caso da utilização de marcadores aleatórios para varreduras completas de genomas para localização de QTL.

A maioria dos marcadores do tipo SNP são dialélicos, que resultam, como discutido anteriormente, em pouca informação por locus. Do ponto de vista de mapeamento, cinco marcadores desse tipo fornecem aproximadamente a mesma informação que um locus de microssatélite, de tal forma que os mapas de SNPs devem ter densidade maior do que os de microssatélites para se obter o mesmo poder de detecção de QTL. Porém, a densidade necessária ao mapeamento fino de QTL só pode ser obtida com esse tipo de marcador (VIGNAL et al., 2002).

Métodos de prospecção de SNPs

Muitas técnicas de detecção de SNPs têm sido descritas com base em comparação de sequências loco-específicas. Atualmente há muitas tecnologias que podem ser utilizadas na prospecção de SNPs; dentre elas podemos destacar a eletroforese em gradiente desnaturante, a clivagem química de pareamento imperfeito, o método de polimorfismo de conformação de fita simples e a cromatografia líquida desnaturante de alta performance.

A eletroforese em gradiente desnaturante (DGGE) é uma técnica baseada no princípio de separação de produto de PCR e tem como base diferenças na sequência do DNA que resultem em alterações nas propriedades de desnaturação da molécula.

Aplica-se um agente químico desnaturante em forma de gradiente em um gel de poliacrilamida. Sequências de DNA diferentes migrarão através desse gel e serão desnaturadas em concentrações específicas de agente desnaturante. A desnaturação altera a velocidade de migração e faz com que moléculas de tamanho semelhante apresentem taxas de migração diferentes, resultando em padrões de bandas específicos quando há polimorfismo (FISCHER e LERMAN, 1983).

A clivagem química de pareamento imperfeito (CCM) é uma tecnologia baseada no fato de que moléculas de citosina e de timina, quando em banda simples, são oxidadas por hidroxilamina e KMnO_4 , respectivamente. Para detecção de *mismatch* (polimorfismo) basta expor as sequências a esses agentes. A base modificada quimicamente é depois clivada com piperidina e o ponto de *mismatch* pode ser detectado pelo tamanho do produto gerado pela clivagem em eletroforese em gel (COTTON et al., 1988; COTTON, 1999). As variações poderão ser detectadas com o uso de sondas complementares.

O princípio do método de polimorfismo de conformação de fita simples (SSCP) está baseado na característica de que fitas simples de DNA, ao renaturarem, adquirem conformações espaciais específicas (estrutura terciária) dependentes de suas sequências de bases (estrutura primária); a taxa de migração de um segmento de DNA quando submetido a uma eletroforese depende dessa estrutura terciária (LIU et al., 1999). Assim, fragmentos de DNA que diferirem em uma única base (SNP) apresentarão diferenças em suas estruturas terciárias e migrarão em velocidades distintas (ORTÍ et al., 1997; SUNNUCKS et al., 2000).

Na cromatografia líquida desnaturante de alta performance (DHPLC) dois fragmentos são gerados por PCR, um no qual se deseja investigar a presença de SNPs e outro dito normal. Os fragmentos são desnaturados e submetidos a reanelamento gradual. Os produtos reanelados são aplicados em uma coluna de DHPLC. Se não houver SNP serão formados apenas “homoduplex”. Mas se o DNA contiver SNP, serão formados “heteroduplex” de DNA normal e mutado. Assim, “homoduplex” e “heteroduplex” apresentarão temperaturas de anelamento diferentes

e os últimos não serão retidos na coluna, gerando um padrão cromatográfico distinto do gerado quando não há SNP (OEFNER e UNDERHILL, 1995).

A espectrometria de massa de fragmentos de DNA pode ser empregada para detecção de SNPs, pois qualquer alteração de base mudará a massa do fragmento. No entanto, fragmentos de DNA grandes sofrem fragmentação no processo, o que reduz a resolução e a sensibilidade (NORDHOFF et al., 1994). Assim, diferenças de uma única base podem ser adequadamente detectadas em fragmentos menores. No caso de fragmento grande, será necessária digestão com enzimas de restrição (CHIU et al., 2000), o que pode tornar o processo caro e trabalhoso.

O principal método de sequenciamento utilizado para prospecção de SNP é o sequenciamento direto (método didesoxi ou de terminação da cadeia). Esse método se baseia na incorporação de desoxinucleotídeos (dNTPs) e de didesoxinucleotídeos (ddNTPs) a uma fita de DNA que está sendo formada e utiliza como molde o DNA de interesse. Assim que os ddNTPs são adicionados à cadeia, a extensão é interrompida, pois os ddNTPs não possuem o grupo hidroxila (OH) 3', necessário para a ligação do próximo desoxinucleotídeo (dNTP). Esses ddNTPs são marcados com fluorescência. Dessa forma, podem ser detectados e a sequência dos nucleotídeos, identificada. Ao final de vários ciclos são obtidas diversas cadeias de DNA, de tamanhos variáveis e terminadas em diferentes ddNTPs que, posteriormente, serão identificados em sequenciador automático (REGITANO et al., 2007).

Métodos de genotipagem de SNPs

Muitos métodos podem ser utilizados para genotipar SNPs. A maioria se baseia na obtenção e na separação de produtos de PCR. Nesse processo geralmente ocorre geração de um produto alelo específico para o SNP de interesse, com posterior detecção do genótipo. As metodologias de discriminação alélica de SNPs são realizadas por meio de reações bioquímicas alelo-específicas. Segundo Kim e Misra (2007), os métodos de extensão de *primer*, hibridização, ligação e clivagem enzimática são os mais utilizados. Após a reação de discriminação alélica os genótipos podem ser identificados com base em sua massa, em sua fluorescência ou em sua quimiluminescência.

A metodologia de extensão de *primer* é baseada na incorporação de deoxirribonucleotídeos específicos complementares à sequência de DNA-alvo. Há

muitas variações dessa tecnologia, no entanto, segundo Kwok (2001), elas podem ser agrupadas em duas classes. Na primeira é utilizada uma adaptação do sequenciamento, com o emprego de ddNTPs, para identificação do nucleotídeo polimórfico com um *primer* comum para amplificação de ambos os alelos. A segunda classe consiste numa PCR alelo-específica, em que a DNA polimerase amplifica o DNA-alvo somente se os *primers* forem perfeitamente complementares ao sítio do DNA-alvo, utilizando-se assim *primers* específicos para detectar cada alelo. Os genótipos do SNP podem ser revelados segundo a massa ou segundo a fluorescência.

A hibridização tem como base a estabilidade térmica do DNA de dupla fita. Em condições de alta estringência, essa estabilidade é suficiente para distinguir entre pares de sonda – DNA-alvo perfeitos e imperfeitos. A hibridização só acontece se ocorrer pareamento perfeito entre sonda e DNA-alvo. Assim, podem ser construídas sondas específicas para cada alelo. Uma ferramenta que utiliza esse princípio é o ensaio para genotipagem de SNP TaqMan® (Applied Biosystems, CA), que constitui uma combinação da hibridização e da atividade exonucleásica 5' da DNA-polimerase, acoplada à detecção de fluorescência (VEGA et al., 2005).

A terceira metodologia de discriminação alélica mais popular é a ligação. É sabido que quando dois oligonucleotídeos hibridizam perfeitamente a um DNA de fita simples, ficando um adjacente ao outro, a enzima DNA-ligase tende a uni-los e formar um único nucleotídeo. É nessa característica que essa metodologia se baseia. A reação consta de três sondas de oligonucleotídeos, sendo duas alelo-específicas marcadas cada uma com um corante fluorescente distinto e a terceira com biotina ligada, que se combinará adjacientemente à sonda alelo-específica (LANDEGREN et al., 1988). Quando ocorre ligação perfeitamente complementar de uma sonda alelo-específica e, em sua adjacência, a ligação da sonda com biotina, a ligase as une, formando um único oligonucleotídeo. Essa reação pode ser analisada, por exemplo, em eletroforese capilar para detecção dos alelos do SNP. Uma ferramenta que se utiliza dessa metodologia é o CFET (combinatorial fluorescence energy transfer) (TONG et al., 2001).

A habilidade de certas enzimas reconhecerem sequências específicas do DNA e efetuarem clivagem nesse ponto constitui outra forma de discriminação alélica. Assim, um alelo de um SNP pode criar ou destruir um sítio de restrição diferenciando o padrão eletroforético entre alelos de um SNP. Uma característica

marcante de muitos desses sítios de clivagem é que eles possuem dupla simetria rotacional, isto é, a sequência de reconhecimento é palindrômica e os sítios de clivagem são simetricamente posicionados (BERGER et al., 2002) Uma técnica que utiliza tais enzimas para discriminar alelos de SNPs é o RFLP.

Outra técnica que vem sendo empregada na discriminação alélica de SNPs são os *amplification refractory mutation systems* (ARMS). Há muitas variações dessa técnica. A mais tradicional (ARMS-PCR) é realizada com a utilização de quatro *primers*, dois internos alelo-específicos (denominados *inner primers*) e dois complementares à sequência que flanqueia o SNP, externos (denominados *outer primers*). Como os dois *primers* internos possuem sequência complementar aos alelo-específicos do SNP, cada *primer* se anelará apenas à região genômica que contém o alelo correspondente. Esses *primers* alelo-específicos contêm um *mismatch* no antepenúltimo nucleotídeo da extremidade 3' para aumentar a especificidade alélica. Os *primers* externos se ligam às regiões flanqueadoras em distâncias distintas em relação ao SNP, de forma a amplificar regiões de diferentes tamanhos, dependendo do alelo presente (YE et al., 2001).

Buitkamp e Semmer (2004) desenvolveram a metodologia que denominaram de ARMS PRNP, baseada no ARMS-PCR. Essa metodologia é muito semelhante ao ARMS-PCR, exceto pelo fato de que possui apenas três *primers*, sendo um externo e dois específicos. Outra diferença está no fato de que um dos *primers* alelo-específicos contém uma cauda degenerada de quatro pares de bases. Esse fato gerará *amplicons* de tamanhos diferentes, específicos para cada alelo.

Métodos que utilizam a hibridização alelo-específica, ou uma combinação de hibridização com extensão de molécula, em arranjos de alta densidade, surgiram como alternativa para a genotipagem de SNPs em escala genômica. Apesar do elevado custo dessa metodologia, a utilização de *pools* de DNA tem sido proposta como alternativa para redução do custo em estudos de associação (PEARSON et al., 2007).

Mapeamento de QTL

O mapeamento de QTL, ou seja, a detecção, a localização e a estimativa do efeito de regiões do genoma associadas a uma característica quantitativa requer grande número de animais para os quais estejam disponíveis dados de avaliação do fenótipo de interesse. O número de indivíduos necessários depende da magnitude

do efeito que se deseja identificar, da herdabilidade do caráter e da estrutura da população, entre outros fatores, porém, esse número é frequentemente da ordem de milhares. Esse requisito é particularmente restritivo em animais de grande porte, nos quais o custo de produção e de manutenção de grandes populações experimentais é elevado.

As populações comerciais têm sido úteis para o mapeamento de QTL em bovinos de leite, dos quais grandes progênies de touro, resultantes do uso intensivo de inseminação artificial, podem ser encontradas. Georges e Andersson (2003) apontaram como vantagem do mapeamento de QTL em gado leiteiro a intensa anotação de dados referentes a avaliações fenotípicas, tais como medidas de qualidade do leite, avaliações da saúde dos animais, avaliações morfométricas e informações de *pedigree* e de manejo, principalmente nos países mais desenvolvidos.

Essa situação, entretanto, não se aplica aos bovinos de corte, cujas famílias são menores e cuja quantidade de características fenotípicas avaliadas é restrita. Além disso, características de difícil avaliação, tais como taxa de ovulação, maciez da carne e resistência a doenças, raramente são consideradas em programas de melhoramento, requerendo a utilização de populações experimentais, as quais devem ser obtidas mediante a aplicação de algum delineamento genético, tais como F_2 , retrocruzamento ou famílias de meios-irmãos. Os resultados mais promissores de identificação de marcadores associados a caracteres de interesse econômico são os relatados com suínos, de que 1.675 QTL foram descritos em mais de 100 publicações (ROTHSCHILD et al., 2007). Algumas mutações em genes, ou em desequilíbrio de ligação com a mutação causal, têm sido regularmente utilizadas pela indústria de melhoramento de suínos.

O mapeamento de QTL em espécies domésticas beneficia-se do reduzido tamanho efetivo das populações, pois a maioria das raças é formada com base em pequeno número de animais fundadores. Esse fato leva à redução da complexidade das características dentro de raças, uma vez que poucos alelos estão representados na população.

A utilização de pequeno número de indivíduos selecionados para dar origem à próxima geração leva ao aumento de endogamia e ao conseqüente aumento da probabilidade de a progênie receber dos parentais regiões do genoma que são idênticas por descendência. Essas regiões podem conter alelos recessivos

deletérios, ou seja, que causam doenças, malformações ou anomalias hereditárias. O emprego de marcadores como método diagnóstico para eliminar animais portadores de alelos deletérios está entre as mais promissoras aplicações dessa tecnologia. Como exemplo, pode-se citar a mutação responsável pela síndrome do estresse dos suínos (ROTHSCHILD et al., 2007), a mutação no gene que codifica a proteína CD18, responsável pela síndrome de deficiência de adesão dos leucócitos, em bovinos da raça Holandesa (SHUSTER et al., 1992), e a recente descrição de mutações independentes no gene *lipoprotein receptor-related protein 4* relacionadas à sindactilia dos bovinos (DROGEMULLER et al., 2007).

Seleção assistida por marcadores

A seleção assistida por marcadores (MAS) tem por objetivo aumentar a acurácia da seleção. Porém, quando se utiliza MAS, apenas alguns dos genes que contribuem para a variação do caráter selecionado são avaliados. Assim, o emprego de informação de genótipo para o marcador, sem considerar as avaliações de diferença esperada na progênie, que fornecem uma visão do componente quantitativo não explicado pelo marcador, pode conduzir à rápida fixação do alelo de QTL selecionado e à maior perda de variabilidade durante o processo de seleção. A MAS deve ser realizada concomitantemente à seleção tradicional, pois dar grande ênfase a somente um marcador poderá ocasionar perda da variabilidade em outros loci, restringir a seleção no longo prazo e resultar em efeito carona ou na fixação de alelos desfavoráveis por deriva na porção do genoma não associada ao marcador. Segundo Van Eenennaam (2007), os benefícios da MAS são maiores nas características que têm baixa herdabilidade, nas que são difíceis ou caras para mensurar e nas que não podem ser medidas antes que o animal contribua para dar origem à próxima geração. Outras características que podem ser beneficentemente selecionadas com MAS são produção de leite, habilidade materna, resistência a doenças, rendimento de carcaça, desempenho de crescimento, fertilidade, eficiência reprodutiva e características limitadas pelo sexo.

Os maiores avanços na área de MAS deverão ser experimentados quando as novas tecnologias de análise de marcadores, tais como chips de SNPs, que permitem a investigação simultânea de grande número de marcadores, alcançarem custo compatível com a aplicação. Nesse caso, a seleção será feita com base no genoma do animal e não mais em informações pontuais de poucos genes. Essa

realidade já pode ser observada nos sumários de touros da raça Holandesa, nos quais machos jovens já possuem estimativa de valor genético com base exclusivamente na informação genômica.

Aliadas às técnicas de análise do genoma funcional e da genômica comparada, as informações sobre QTL deverão rapidamente conduzir à descoberta de genes e de mutações causais, aquelas que determinam a variação fenotípica. A compreensão dos mecanismos biológicos e do controle genético das características de interesse será um campo aberto para a manipulação da expressão de genes, quer via transgênese quer via técnicas de expressão transitória, tais como o RNA de interferência.

Caracterização de recursos genéticos

Uma das premissas fundamentais do melhoramento é a existência de variabilidade genética nas características a serem selecionadas. O processo de melhoramento leva à fixação de caracteres e à perda de variabilidade, quer intencionalmente, por meio de seleção, quer acidentalmente, por meio de deriva genética, resultante do reduzido tamanho efetivo das populações. A conservação de recursos genéticos animais tem por objetivo manter os repositórios de alelos e de combinações alélicas que possam vir a suprir necessidades do melhoramento animal, até mesmo de nichos pouco privilegiados pelos objetivos da indústria agropecuária, como genes relacionados à adaptação a condições tropicais e de baixo investimento tecnológico (ILRI, 2007).

Entretanto, quando se trata de animais domésticos, a manutenção de recursos genéticos requer extensas áreas e representa custo razoável. A utilização de marcadores moleculares permite o monitoramento da variabilidade genética e a estimativa de parâmetros populacionais, tais como o coeficiente de endogamia. Nas populações em que não há registro genealógico, é possível identificar a paternidade, apesar de esse processo requerer a análise de grande número de marcadores ou de padrões de *DNA-fingerprints*.

A análise de diversidade intrapopulacional pode também ser utilizada na identificação dos acessos que constituem a menor coleção em que está representada a máxima variabilidade genética de uma unidade de preservação (*core collection*). Esse conceito tem sido amplamente empregado em recursos genéticos

vegetais e foi recentemente proposto para o manejo de coleções de animais domésticos (EGITO, 2007).

Considerações finais

Os avanços metodológicos dos últimos 50 anos permitiram o desenvolvimento de um número ilimitado de marcadores moleculares. A escolha do marcador adequado a cada situação depende principalmente do objetivo do trabalho. Para construção de mapas genéticos e para mapeamento de QTL, marcadores altamente informativos podem ser melhores, porém, quando se deseja a construção de mapas de alta densidade, mesmo que menos informativos, os SNPs são mais adequados, pois se apresentam em maior frequência nos genomas. Para fins de comparação de populações, além do conteúdo de informação polimórfica, a reprodutibilidade dos fenótipos moleculares é um dos fatores a ser considerado.

Referências bibliográficas

BERGER, J. M.; TYMOCZKO, J. L.; STRYER, L.; CLARKE, N. D. **Biochemistry 6.1: the basic tools of gene exploration**. 5^{ed}. New York: W. H. Freeman, 2002.

BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal of Human Genetics**, v. 32, n. 3, p. 314-331, 1980.

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

BUITKAMP, J.; SEMMER, J. A robust, low- to medium-throughput prnp genotyping system in sheep. **BioMed Central Infectious Diseases**, v. 4, p. 30, 2004.

CHIU, N. H.; TANG, K.; YIP, P.; BRAUN, A.; KOSTER, H.; CANTOR, C. R. Mass spectrometry of single-stranded restriction fragments captured by an undigested complementary sequence. **Nucleic Acids Research**, v. 28, n. 8, p. E31, 2000.

COTTON, R. G. H. Mutation detection by chemical cleavage. **Genetic Analysis: Biomolecular Engineering**, v. 14, n. 5-6, p. 165-168, 1999.

COTTON, R. G.; RODRIGUES, N. R.; CAMPBELL, R. D. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. **Proceedings of the National Academy of Sciences of the United States of América**, v. 85, n. 12, p. 4397-4401, 1988.

DROGEMULLER, C.; LEEB, T.; HARLIZIUS, B.; TAMMEN, I.; DISTL, O.; HOLTERSHINKEN, M.; GENTILE, A.; DUCHESNE, A.; EGGEN, A. Congenital syndactyly in cattle: four novel mutations in the low density lipoprotein receptor-related protein 4 gene (LRP4). **BioMed Central Genetics**, v. 8, p. 5, 2007.

EGITO, A. A. **Diversidade genética, ancestralidade individual e miscigenação nas raças bovinas no Brasil com base em microssatélites e haplótipos de DNA mitocondrial**: subsídios para a conservação. 2007. 246f. Tese (Doutorado em Biologia Celular) – Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, 2007.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores RAPD e RFLP em análise genética**. Brasília: EMBRAPA-CENARGEN, 1995. 220 p.

FISCHER, S. G.; LERMAN, L. S. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. **Proceedings of the National Academy of Sciences of the United States of America**, v. 80, n. 6, p. 1579-1583, 1983.

GEORGES, M.; ANDERSSON, L. Positional identification of structural and regulatory quantitative trait nucleotides in domestic animal species. In: COLD SPRING HARBOR SYMPOSIA ON QUANTITATIVE BIOLOGY, 68., 2003. **Anais...** Cold Spring: Harbor Laboratory Press, p. 179-188, 2003.

GRODZICKER, T.; WILLIAMS, J.; SHARP, P.; SAMBROOK, J. Physical mapping of temperature-sensitive mutations of adenoviruses. **Cold Spring Harbor Symposium on Quantitative Biology**, v. 39, v. 1, p. 439-446, 1975.

HOLMES, N. G. Microsatellite markers and the analysis of genetic disease. **British Veterinary Journal**, v. 150, n. 5, p. 411-421, 1994.

ILRI – INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE. **Improving utilization of farm animal genetic resources**. 2007. Disponível em: <<http://www.ilri.org/research/Content.asp?SID=15&CCID=41>>. Acesso em: 10 jun. 2009.

JOBSE, C.; BUNTJER, J. B.; HAAGSMA, N.; BREUKELMAN, H. J.; BEINTEMA, J. J.; LENSTRA, J. A. Evolution and recombination of bovine DNA repeats. **Journal of Molecular Evolution**, v. 41, n. 3, p. 277-283, 1995.

KADARMIDEEN, H. N.; ROHR, P. von; JANSSE, L. L. G. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. **Mammalian Genome**, v. 17, n. 6, p. 548-564, 2006.

KIM, S.; MISRA, A. SNP genotyping: technologies and biomedical applications. **Annual Review of Biomedical Engineering**, v. 9, p. 289-320, 2007.

KWOK, P. Y. Methods for genotyping single nucleotide polymorphisms. **Annual Review of Genomics and Human Genetics**, v. 2, p. 235-258, 2001.

LANDEGREN, U.; KAISER, R.; SANDERS, J.; HOOD, L. A ligase-mediated gene detection technique. **Science**, v. 241, n. 4869, p. 1077-1080, 1988.

LIU, Q.; FENG, J.; BUZIN, C.; WEN, C.; NOZARI, G.; MENGOS, A.; NGUYEN, V.; LIU, J.; CRAWFORD, L.; FUJIMURA, F. K.; SOMMER, S. S. Detection of virtually all mutations-SSCP (DOVAM-S): a rapid method for mutation scanning with virtually 100% sensitivity. **Biotechniques**, v. 26, n. 5, p. 932-942, 1999.

NCBI – NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **GenBank overview**. Disponível em: <<http://www.ncbi.nlm.nih.gov/Genbank/index.htm>>. Acesso em: 26 jun. 2009.

NORDHOFF, E.; KIRPEKAR, F.; KARAS, M.; CRAMER, R.; HAHNER, S.; HILLENKAMP, F.; KRISTIANSEN, K.; ROEPSTROFF, P.; LEZIUS, A. Comparison of IR- and UV-matrix-assisted laser desorption/ionization mass spectrometry of oligodeoxynucleotides. **Nucleic Acids Research**, v. 22, n. 13, p. 2460-2465, 1994.

OEFFNER, P. J.; UNDERHILL, P. A. Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). **American Journal of Human Genetics**, v. 57, p. 266, 1995.

ORTÍ, G.; HARE, M. P.; AVISE, J. C. Detection and isolation of nuclear haplotypes by PCR-SSCP. **Molecular Ecology**, v. 6, n. 6, p. 575-580, 1997.

PEARSON, J. V.; HUENTELMAN, M. J.; HALPERIN, R. F.; TEMBE, W. D.; MELQUIST, S.; HOMER, N.; BRUN, M.; SZELINGER, S.; COON, K. D.; ZISMANN, V. L.; WEBSTER, J. A.; BEACH, T.; SANDO, S. B.; AASLY, J. O.; HEUN, R.; JESSEN, F.; KOLSCH, H.; TSOLAKI, M.; DANILIDOU, M.; REIMAN, E. M.; PAPASSOTIROPOULOS, A.; HUTTON, M. L.; STEPHAN, D. A.; CRAIG, D. W. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies. **American Journal of Human Genetics**, v. 80, n. 1, p. 126-139, 2007.

REGITANO, L. C. A.; MÉO-NICIURA, S. C.; IBELLI, A. M. G.; GOUVEIA, J. J. S. **Protocolos de biologia molecular aplicada à produção animal**. 1. ed. São Carlos: Embrapa Pecuária Sudeste, 2007. 66 p. Disponível em: <<http://www.cppse.embrapa.br/080servicos/070publicacaogratis/e-book/LivroProtMol.pdf>>. Acesso em: 10 jun. 2009.

ROTHSCHILD, M. F.; HU, Z. L.; JIANG, Z. Advances in QTL mapping in pigs. **International Journal of Biological Sciences**, v. 3, n. 3, p. 192-197, 2007.

SELLNER, E. M.; KIM, J. W.; MCCLURE, M. C.; TAYLOR, K. H.; SCHNABEL, R. D.; TAYLOR, J. F. Board-invited review: Applications of genomic information in livestock. **Journal of Animal Science**, v. 85, n. 12, p. 3148-3158, 2007.

SHUSTER, D. E.; KEHRLI Jr, M. E.; ACKERMANN, M. R.; GILBERT, O. R. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 19, p. 9225-9229, 1992.

SIMPSON, A. J. G. Mini review: microsatellite instability in human cancer. **Brazilian Journal of Genetics**, v. 19, p. 171-174, 1996.

SMITHIES, O. Zone electrophoresis in starch gel: group variation in the serum proteins of normal human adults. **Biochemical Journal**, v. 61, n. 4, p. 629-641, 1955.

SOUTHERN, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. **Journal of Molecular Biology**, v. 98, n. 3, p. 503-517, 1975.

STORMONT, C.; CUMLEY, R. W. Cellular antigens in cattle blood. **Journal of Heredity**, v. 34, p. 35-41, 1943.

SUNNUCKS, P.; WILSON, A. C. C.; BEHEREGARAY L. B.; ZENGER, K.; FRENCH, J.; TAYLOR, A. C. SSCP is not difficult: the application of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. **Molecular Ecology**, v. 9, n. 11, p. 1699-1710, 2000.

TANKSLEY, S. D. Mapping polygenes. **Annual Review of Genetics**, v. 27, p. 205-233, 1993.

TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, v. 17, n. 16, p. 6463-6471, 1989.

TONG, A. K.; LI, Z.; JONES, G. S.; RUSSO, J. J.; JU, J. Combinatorial fluorescence energy transfer tags for multiplex biological assays. **Nature Biotechnology**, v. 19, n. 8, p. 756-759, 2001.

VAN EENENNAAM, A. Marker assisted selection in beef cattle. Disponível em: <http://animalscience.ucdavis.edu/animalbiotech/Outreach/Marker_Assisted_Selection_in_Beef_Cattle.pdf>. Acesso em: 05 de junho 2007.

VEGA, F. M.; LAZARUK, K. D.; RHODES, M. D.; WENZ, M. H. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP genotyping assays and the SNPLex genotyping system. **Mutation Research**, v. 573, n. 1-2, p. 111-135, 2005.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 34, n. 3, p. 275-305, 2002.

WEIR, B. S. **Genetic data analysis**: methods for discrete population genetic data. 2. ed. Massachusetts: Sinauer Associates, 1996. 445 p.

YE, S.; DHILLON, S.; KE, X.; COLLINS, A. R.; DAY, I. N. M. An efficient procedure for genotyping single nucleotide polymorphisms. **Nucleic Acids Research**, v. 29, n. 17, p. E88, 2001.