



Empresa Brasileira de Pesquisa Agropecuária
Centro de Pesquisa Agropecuária de Clima Temperado
Ministério da Agricultura, Pecuária e Abastecimento

ISSN 1806-9193

Dezembro, 2008

versão

ON LINE

Documentos 248

**Mapeamento por
associação em plantas:
conceitos básicos e
perspectivas de uso no
melhoramento de batata**

Pelotas, RS
2008

Exemplares desta publicação podem ser adquiridos na:

Embrapa Clima Temperado

Endereço: BR 392, km 78
Caixa Postal 403, CEP 96001-970 - Pelotas, RS
Fone: (53) 3275 8199
Fax: (53) 3275 8219 - 3275 8221
Home page: www.cpact.embrapa.br
E-mail: sac@cpact.embrapa.br

Comitê de Publicações da Unidade

Presidente: Walkyria Bueno Scivittaro
Secretária-Executiva: Joseane M. Lopes Garcia
Membros: Cláudio Alberto Souza da Silva, Lígia Margareth Cantarelli Pegoraro, Isabel Helena Verneti Azambuja, Luís Antônio Suita de Castro, Sadi Macedo Sapper, Regina das Graças V. dos Santos
Suplentes: Daniela Lopes Leite e Luís Eduardo Corrêa Antunes

Revisor de texto: Sadi Macedo Sapper
Normalização bibliográfica: Regina das Graças Vasconcelos dos Santos
Editoração eletrônica e capa: Oscar Castro
Fotos da capa: Gilnei Wanke
Arte da capa: Miguel Ângelo (estagiário)

1ª edição

1ª impressão 2008: 100 exemplares

Todos os direitos reservados

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

Castro, Caroline Marques

Mapeamento por associação em plantas: conceitos básicos e perspectivas de uso no melhoramento de batata / Caroline Marques Castro, Arione da Silva Pereira. - Pelotas: Embrapa Clima Temperado, 2008.
30 p. - (Embrapa Clima Temperado. Documentos, 248).

ISSN 1516-8840

Batata — *Solanum tuberosum* – Marcador molecular – QTL. I. Pereira, Arione da Silva. II. Título. III. Série.

CDD 635.21

Autor

Caroline Marques Castro

Eng. Agrôn., Dra.
Embrapa ClimaTemperado
Cx. Postal 403, 96001-970 - Pelotas, RS
(caroline@cpact.embrapa.br)

Arione da Silva Pereira

Eng. Agrôn., Ph.D.
Embrapa ClimaTemperado
Cx. Postal 403, 96001-970 - Pelotas, RS
(arione@cpact.embrapa.br)

Apresentação

Nas últimas décadas, os avanços obtidos na área de genômica foram expressivos. O seqüenciamento completo do genoma de algumas espécies se tornou uma realidade. Vários genes foram identificados. Entretanto, o uso desta importante área da ciência para explorar a diversidade genética conservada nos bancos ou coleções de germoplasma, na busca por genes de importância agrônômica, ainda é um desafio. Nesse sentido, o mapeamento por associação, também conhecido como o mapeamento por desequilíbrio de ligação, desponta como um importante método para relacionar genes e alelos a caracteres agrônômicos de interesse. Inicialmente usado em genética humana, para mapear genes relacionados a doenças, nos últimos anos esta técnica vem sendo aplicada em plantas. Neste documento são abordados os conceitos básicos relacionados ao mapeamento por associação. São discutidos os principais pontos para o sucesso deste método em explorar a diversidade genética e são mostrados exemplos e perspectivas do uso desta metodologia no melhoramento genético da batata.

Waldyr Stumpf Junior

Chefe-Geral
Embrapa Clima Temperado

Sumário

Mapeamento por associação em plantas: conceitos básicos e perspectivas de uso no melhoramento de batata	9
1. Introdução	9
2. Desequilíbrio de ligação	11
2.1. Dinâmica do desequilíbrio de ligação	14
3. Métodos para o mapeamento por associação em plantas	16
4. Relevância do desequilíbrio de ligação para o mapeamento de genes	19
5. Mapeamento por associação: exemplos e perspectivas de uso no melhoramento de batata .	21
6. Considerações finais	24
Referências	25

Mapeamento por associação em plantas: conceitos básicos e perspectivas de uso no melhoramento de batata

Caroline Marques Castro
Arione da Silva Pereira

1. Introdução

Nos últimos 20 anos, desde o começo da era PCR (*Polymerase Chain Reaction*), em meados de 1980 (SAIKI et al., 1985; MULLIS e FALOONA, 1987), o uso de marcadores de DNA para estudos de diversidade genética em plantas tornaram-se freqüentes em grande parte dos programas de melhoramento genético. Um número cada vez maior de técnicas moleculares foi desenvolvido, visando acessar a variabilidade de forma mais precisa, rápida e com custo mais baixo (SPOONER et al., 2005). Uma grande quantidade de dados genotípicos, originados da caracterização molecular das coleções de germoplasma, estão disponíveis, seja na literatura e/ou em base de dados. Indiscutivelmente, estes estudos tiveram um importante papel no manejo dos recursos genéticos conservados nos bancos e coleções de germoplasma. Entretanto, os frutos da revolução molecular, no sentido de promover o uso da riqueza de alelos conservadas nos bancos de germoplasma, serão provavelmente usufruídos neste século, por meio do mapeamento por associação (GEBHARDT et al., 2004).

O mapeamento por associação, também conhecido como mapeamento por desequilíbrio de ligação, ou mapeamento por desequilíbrio de fase gamética, baseia-se em conceitos de genética de populações para identificar relações entre marcadores genéticos e caracteres fenotípicos. O princípio fundamental do mapeamento por associação é o desequilíbrio de ligação (WRAY e VISSCHER, 2008).

Os termos *desequilíbrio de ligação* e *ligação genética* são freqüentemente confundidos. Embora relacionados, são termos distintos. A *ligação genética* refere-se à herança correlacionada de *loci* através da sua conexão física no cromossomo. Por outro lado, o *desequilíbrio de ligação* refere-se à correlação de alelos em uma população. O conflito entre os dois termos ocorre porque a ligação genética pode resultar em altos níveis de desequilíbrio de ligação. Por exemplo, se duas mutações ocorrem em intervalos de poucas bases uma da outra, elas vão sofrer a mesma pressão de seleção e deriva ao longo do tempo e, devido ao fato de que a recombinação entre estas duas bases é rara, a presença destes polimorfismos vai ser altamente correlacionada e a ligação genética irá resultar em grande desequilíbrio de ligação. Diferentemente, polimorfismos localizados em cromossomos distintos passam por pressões de seleção e segregação independente, assim, estes polimorfismos vão ter um menor nível de desequilíbrio de ligação (FLINT-GARCIA et al., 2003).

Historicamente, estudos de ligação genética foram usados para medir a proximidade genética de um *locus* com outro com o objetivo de mapear caracteres qualitativos e quantitativos. Ao contrário dos mapas de ligação clássicos, os quais são baseadas em estrutura de família e requerem o desenvolvimento de populações onde as relações intra-familiar, por meio da análise dos seus descendentes, são utilizadas para prever correlações entre fenótipo e genótipo, os métodos associativos não requerem o desenvolvimento de populações de mapeamento. Baseiam-se em desequilíbrios de ligação já existentes nas populações, que ocorreram previamente,

historicamente sem registro (GEBHARDT et al., 2004). A diversidade genética é avaliada em populações naturais para identificar os polimorfismos que se correlacionam com a variação fenotípica e, assim, criar associações entre marcadores genéticos e fenótipo, permitindo explorar a diversidade alélica de cada espécie (MACKAY e POWELL, 2007). O mapeamento por desequilíbrio de ligação pode ser aplicado a uma série de populações experimentais ou não experimentais. Por exemplo, genótipos elite de programas de melhoramento genético, ou coleções de germoplasma podem ser diretamente usadas para mapear QTLs (*Quantitative Trait Loci*), explorando dados que foram historicamente acumulados nos programas de pesquisa (GEBHARDT et al., 2004).

Neste documento são discutidos alguns conceitos básicos do mapeamento por associação em plantas e são mostrados exemplos do uso desta metodologia para mapear caracteres de interesse agrônomo em batata (*Solanum tuberosum* L.), uma espécie tetraplóide, com herança tetrassômica.

2. Desequilíbrio de ligação

O desequilíbrio de ligação é a associação não ao acaso de alelos em diferentes *loci*, é a correlação entre polimorfismos que é causada pela sua história compartilhada de mutação e recombinação. Inicialmente, o desequilíbrio de ligação está presente em uma população em uma taxa determinada pela distância genética entre os dois *loci* e o número de gerações desde que ele surgiu (FLINT-GARCIA et al., 2003).

Considerando um par de alelos no *locus* 1, *A* e *a*, e no *locus* 2, *B* e *b*, com as respectivas frequências p_A , $p_{a'}$, p_B e $p_{b'}$ as probabilidades de encontrar os alelos destes dois *loci* juntos, no mesmo cromossomo, são: p_{AB} , $p_{Ab'}$, p_{aB} e $p_{ab'}$. Se estiverem em equilíbrio de ligação, os alelos do *locus* 2 (*B* e *b*) são independentes dos alelos do *locus* 1 (*A* e *a*), ou seja, $p_{AB} = p_A p_B$. Uma medida direta de medir o desequilíbrio de ligação (*D*) é por

meio da diferença entre as frequências alélicas p_{AB} e $p_A p_B$ ou seja, $D_{ab} = (\pi_{AB} - \pi_A \pi_B)$.

Havendo equilíbrio de ligação, $D = 0$. Entretanto, com desequilíbrio, D pode ser negativo ou positivo, sendo dependente das frequências alélicas encontradas na população em estudo (FALCONER e MACKAY, 1996).

Uma série de estatísticas tem sido usadas para medir o desequilíbrio de ligação, sendo que todas têm como componente básico a diferença entre as frequências alélicas observada e esperada (D_{ab}). As duas estatísticas mais usadas para mensurar o desequilíbrio de ligação são r^2 e D' (FLINT-GARCIA et al., 2003).

A estatística r^2 , também descrita na literatura como D^2 , é a correlação ao quadrado entre a presença e ausência de alelos em diferentes *loci* (Hill e Robertson, 1968):

$$r^2 = \frac{(D_{ab})^2}{\pi_A \pi_a \pi_B \pi_b}$$

A estatística D' é calculada da seguinte forma:

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_b, \pi_a \pi_B)} \quad , \text{ se } D_{ab} < 0;$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_B, \pi_a \pi_b)} \quad , \text{ se } D_{ab} > 0;$$

A estatística D' varia entre -1 e 1, enquanto o seu valor absoluto $|D'|$ varia de 0 a 1. Somente quando forem observados todos os quatro possíveis haplótipos (p_{AB} , $p_{Ab'}$, p_{aB} e $p_{ab'}$) é que D' será inferior a 1. Nesse caso, conseqüentemente, um suposto evento de recombinação terá ocorrido entre os dois *loci*. Sempre que uma das quatro frequências haplotípicas for zero, $|D'|$ será igual a 1, o que ocorre com frequência, quando se trabalha com populações pequenas (WRAY e VISSCHER, 2008).

As estatísticas r^2 e D' refletem diferentes aspectos de desequilíbrio de ligação e apresentam comportamentos diversos em condições distintas. Na estatística r^2 , somente se encontra um valor igual a 1, ou seja, desequilíbrio de ligação completo, quando os dois *loci* em questão apresentam frequências alélicas idênticas. Por outro lado, na estatística D' , sempre que a combinação de um par de alelos não estiver presente na população em estudo, D' será igual a 1 (FLINT-GARCIA et al., 2003). Hedrick (1987), em um estudo baseado em simulações, mostrou que a correlação entre as duas medidas (r^2 e D') é muitas vezes baixa.

Esta diferença nos valores encontrados no desequilíbrio de ligação quando se adota r^2 ou D' é resultado da história evolutiva que cada uma destas estatísticas incorpora para mensurar o desequilíbrio de ligação, sendo que cada método tem suas vantagens e desvantagens. A estatística r^2 resume as histórias de recombinação e de mutação que estes *loci* sofreram, enquanto que D' mede apenas a história de recombinação e, sendo assim, fornece uma medida mais exata para estimar diferenças de eventos de recombinação. Entretanto, D' é fortemente afetada por amostras de tamanho pequeno, resultando em um comportamento errático quando se compara *loci* com baixas frequências alélicas. Por outro lado, quando se tem o objetivo de examinar a resolução de estudos associativos, geralmente é conveniente a estatística r^2 , uma vez que ela é um indicativo de como os marcadores podem se correlacionar com o QTL de interesse (FLINT-GARCIA et al., 2003).

Muitos estudos que descrevem o desequilíbrio de ligação em uma determinada população estimam ambas as medidas (r^2 e D'), o que permite um julgamento imediato do desequilíbrio de ligação junto com a diferença das frequências alélicas. Por exemplo, alto $|D'|$ e alto r^2 implica em uma tendência de apenas dois haplótipos estarem presentes e com pequena diferença na frequência do par de alelos; alto $|D'|$ e baixo r^2 implica em uma

tendência de estarem presentes apenas três haplótipos, com pares de alelos com frequências distintas; baixo $|D|$ e baixo r^2 implica na tendência de estarem presentes os quatro haplótipos e que a ligação dos alelos é ao acaso (WRAY e VISSCHER, 2008).

Há duas formas comuns de visualizar a extensão do desequilíbrio de ligação entre pares de *loci*. Os gráficos de declínio do desequilíbrio de ligação são usados para visualizar a razão na qual o desequilíbrio de ligação diminui em função da distância genética ou física. São construídos gráficos de dispersão (*scattered plots*) dos valores r^2 versus a distância genética ou física entre todos os pares de alelos. Alternativamente, também podem ser usadas matrizes de desequilíbrio, as quais são eficazes para visualizar o arranjo linear do desequilíbrio de ligação entre *loci* ao longo do cromossomo. Ambos os tipos de gráficos destacam a variação ao acaso do desequilíbrio de ligação devido a uma série de forças evolucionárias (FLINT-GARCIA et al., 2003).

2.1 Dinâmica do desequilíbrio de ligação

Uma vez que a frequência alélica e a recombinação entre *loci* afeta o desequilíbrio de ligação, a maioria dos processos observados em genética de populações refletem nos padrões do desequilíbrio de ligação. Há uma série de forças evolucionárias que ocasionam o desequilíbrio de ligação, incluindo mutação, deriva genética, migração e seleção (WRAY e VISSCHER, 2008).

A mutação fornece o material bruto para produzir o polimorfismo que estará em desequilíbrio de ligação. O tamanho da população também tem um papel importante. O desequilíbrio de ligação pode ser originado em populações que sofreram recentemente uma redução no tamanho (*bottleneck*), o que resultou em deriva genética e, conseqüentemente, poucas combinações alélicas são passadas para as gerações futuras. Já a migração, ou seja, a introdução de novos

indivíduos na população, resulta no fluxo de genes entre genótipos de populações geneticamente distintas seguido por inter-cruzamentos, o que resulta na introdução de cromossomos de diferentes ancestrais, com frequências alélicas distintas. Frequentemente o desequilíbrio de ligação resultante da migração se estende por *loci* que não estão ligados, entretanto, nesse caso, esse desequilíbrio é derrubado rapidamente com os cruzamentos ao acaso (FLINT-GARCIA et al., 2003).

A principal força que enfraquece o desequilíbrio de ligação é a recombinação. Alelos que estão fisicamente ligados nos pais (haplótipos parentais), são transmitidos juntos para a progênie, até que ocorra um evento de recombinação que separe os dois alelos. A cada geração, a transferência de alelos fisicamente ligados é reduzida na proporção da fração de recombinação, até que seja adquirido o equilíbrio de ligação, onde os haplótipos recombinantes e não recombinantes são igualmente distribuídos na população (GEBHARDT, 2007). Em uma população grande, de cruzamentos ao acaso, a relação entre desequilíbrio de ligação no tempo t relativo ao desequilíbrio no tempo 0 pode ser derivada da seguinte forma (WRAY e VISSCHER, 2008):

No tempo 0 , o desequilíbrio inicial é:

$$D(0) = p_{AB} - p_A p_B$$

Estendendo para qualquer geração futura (t):

$D(t) = \frac{1}{4}(1 - 2c)(1 - c)^{t-1}$, sendo c a fração recombinante entre os *loci* ligados.

Quanto mais próximos estiverem fisicamente os dois *loci*, o desequilíbrio de ligação persistirá por múltiplas gerações meióticas na população devido à baixa frequência de recombinação (GEBHARDT, 2007).

3. Métodos para o mapeamento por associação em plantas

Duas estratégias são adotadas em estudos associativos, uma aborda o genoma todo (*genome-wide*) e a outra é uma abordagem por meio de genes candidatos (*candidate gene*). Na primeira estratégia, um grande número de marcadores, distribuídos ao longo do genoma e preferencialmente com posição conhecida em um mapa físico da espécie, é usado na genotipagem da população. Os mesmos indivíduos são fenotipados para os caracteres de interesse. Os marcadores são usados para avaliar todos os genes simultaneamente. A segunda estratégia, gene-candidato, foca em genotipar regiões do genoma que aportam os genes candidatos (GEBHARDT, 2007). A análise funcional de genes que atuam na resistência a pestes ou patógenos, ou em qualquer outra rota metabólica importante, sejam estes genes identificados na própria espécie em estudo, ou em espécies modelo, como *Arabidopsis thailana*, abastecem a oferta de genes-candidatos para estudos associativos (GEBHARDT et al., 2007).

Ambas as estratégias tiram proveito do nível de desequilíbrio de ligação presente na população (MORGANTE e SALAMINI, 2003). A natureza do desequilíbrio de ligação determina qual o tipo de abordagem associativa que deve ser adotada. A razão com que ocorre o declínio do desequilíbrio de ligação em função da distância genética ou física na população em estudo é que define se deve ser usada a estratégia de avaliação do genoma todo ou de genes-candidatos (FLINT-GARCIA et al., 2003). Quando o desequilíbrio de ligação decai em uma distância curta ao longo do cromossomo, o número de marcadores ao acaso necessários para encontrar associações significativas pode ser substancialmente grande (MALOSETTI et al., 2007).

De modo geral, quando o desequilíbrio de ligação é pequeno na população, um grande número de marcadores será necessário

para detectar uma associação, porém, tem-se o potencial de atingir um mapa com alta resolução. Inversamente, quando o desequilíbrio de ligação é grande na população em estudo, um número menor de marcadores será necessário para detectar associações. Entretanto, obviamente, o mapa terá uma resolução inferior (MORGANTE e SALAMINI, 2003).

Estudos associativos têm sido a principal aplicação do desequilíbrio de ligação. Antes de examinar associações entre marcadores e caracteres fenotípicos, é calculado o desequilíbrio de ligação entre pares de marcadores seguindo uma das metodologia previamente descrita neste documento (item 2). Quando existe desequilíbrio de ligação entre um marcador molecular e um *locus* que controla um caráter de interesse, os marcadores podem estar associados aos valores fenotípicos mensurados na população de indivíduos (LANDER e SCHORCK, 1994). Ao longo de um série de gerações, em uma população não estruturada, ou seja, em uma população de cruzamentos ao acaso, livre de sub-divisão ou migração, apenas as correlações entre o QTL e os marcadores proximamente ligados a ele é que irão permanecer (MACKAY e POWELL, 2007).

A metodologia estatística mais utilizada para acessar associações entre marcadores e características fenotípicas consiste principalmente na comparação entre as médias da característica fenotípica para os dois estados do alelo por meio de algum tipo de teste-*t*, seja o clássico teste-*t*, ou um método alternativo mais robusto, como por exemplo o Mann-Whitney teste-*U* (MALOSETTI et al., 2007).

Entretanto, o sucesso do mapeamento por associação depende da possibilidade de separar o desequilíbrio de ligação oriundo do efeito da recombinação, a ligação genética propriamente dita, do desequilíbrio de ligação decorrente de outras forças evolucionárias (REMINGTON et al., 2001). A estrutura populacional é vista como a segunda principal causa de associação entre marcadores e caracteres fenotípicos (FLINT-

GARCIA et al., 2003). A principal restrição ao uso do mapeamento por associação em espécies vegetais ocorre em função do grande risco de identificar associações falso-positivas em função da estrutura populacional (PRITCHARD, 2001). A maioria das populações vegetais apresenta algum grau de estruturação ou subdivisão, como conseqüência, com freqüência são encontradas correlações entre *loci* não ligados (MACKAY e POWELL, 2007). A história de melhoramento genético de muitas das principais espécies vegetais de importância agrônômica, seguido pelo limitado fluxo gênico entre as espécies cultivadas e os seus parentes silvestres, criou complexas estratificações no germoplasma da maioria das espécies de importância agrônômica, o que complica os estudos associativos (SHARBEL et al., 2000).

Ignorar as correlações entre os genótipos em estudo induz a testes liberais de associação entre marcadores e caracteres fenotípicos, o que aumenta drasticamente o número de falsos QTLs detectados (PARISSEAUX e BERNARDO, 2004). Métodos estatísticos foram desenvolvidos visando corrigir o problema da estrutura populacional e assim diminuir a identificação de falsas associações. Uma das estratégias mais utilizadas no mapeamento por associação é inicialmente examinar se podem ser distinguidos grupos dentro da população para que, somente após a correção para o efeito de grupo intra-populacional, sejam testadas as associações marcador-caráter fenotípico (PRITCHARD et al., 2000). Dessa forma, são feitos testes-*t* dentro dos grupos identificados (SIMKO et al., 2004a). O princípio básico é que após remover os efeitos de grupo, apenas as associações causadas pela ligação física entre os alelos é que irão permanecer (MALOSETTI et al., 2007).

Entretanto, o uso de testes-*t* para estabelecer associações entre marcadores e caracteres fenotípicos impõe algumas restrições às análises e inferências. No mapeamento associativo é interessante que se possa utilizar dados fenotípicos acumulados ao longo dos anos nos programas de melhoramento genético, os quais são, freqüentemente,

altamente desbalanceados. Nesse sentido, a estrutura de modelos-mistos mostra-se como uma ferramenta poderosa para ser empregada no mapeamento associativo (MALOSETTI et al., 2007).

O emprego de modelos-mistos em estudos associativos em plantas tem se mostrado um dos métodos mais eficientes para diminuir falsas associações entre marcadores e caracteres fenotípicos. Nos modelos-mistos são incorporadas as relações entre os genótipos, seja por meio de informações de *pedigree*, de sub-estrutura populacional (grupos) ou de qualquer outra informação que esteja disponível sobre o germoplasma, a fim de controlar as possíveis falsas associações (MALOSETTI, 2007; YU et al., 2006).

4. Relevância do desequilíbrio de ligação para estudos de mapeamento de genes

Até o início deste século, a maioria dos estudos de mapeamento de caracteres qualitativos e quantitativos em plantas era baseada em mapas de ligação dependentes do desenvolvimento de populações estruturadas. Nestes estudos, um dos tipos mais comuns de populações de mapeamento utilizados são as populações F_2 , originalmente derivadas da auto-fecundação do híbrido F_1 oriundo do cruzamento de dois indivíduos contrastantes para a característica de interesse, e as populações de retro-cruzamento, derivadas do cruzamento do híbrido F_1 com um dos pais (COLLARD et al., 2005).

Vários QTLs foram mapeados com sucesso em diversas espécies de plantas por meio do mapeamento de ligação clássico, baseado em populações estruturadas (STUBER et al., 1999). Entretanto, duas grandes limitações ao uso deste tipo de população para mapear QTLs podem ser citadas. Uma é o fato de que um limitado número de eventos de recombinação ocorreu nestas populações, o que resulta em baixa resolução

para o mapeamento de QTLs e, a outra, o fato de que apenas poucos alelos por *locus* podem ser estudados simultaneamente; por exemplo, em espécies diplóides, apenas dois alelos são amostrados por *locus* em cada estudo (FLINT-GARCIA et al., 2003).

Apesar dos esforços para aumentar a resolução das populações de mapeamento, como o desenvolvimento de grandes populações de linhagens puras recombinantes (RILs - *Recombinant Inbred Lines*), populações as quais passaram por vários ciclos de cruzamento ao acaso, o que aumenta o número potencial de eventos de recombinação, a resolução para muitos QTLs ainda é de muitos centimorgans (cM), o que corresponde a centenas de genes (FLINT-GARCIA et al., 2003). Com o uso de RILs, QTLs foram mapeados em regiões com distâncias genéticas variando de 10 a 30 cM (ALPERT e TANKSLEY, 1996; STUBER et al., 1999). Entretanto, para o uso de QTLs por programas de melhoramento de plantas, seja por meio da seleção assistida por marcadores, ou para que sejam identificados genes funcionais, é necessário que a posição destes QTLs seja estimada com grande resolução (GUPTA et al.; 2005).

Alternativamente, o uso de populações naturais para mapear caracteres de interesse, por meio do mapeamento por associação, tem como grande vantagem o potencial de mapear QTLs com maior resolução, uma vez que se aproveita dos eventos de mutação e recombinação acumulados ao longo do tempo na população para analisar um fenótipo de interesse (WRAY e VISSCHER, 2008). Adicionalmente, em estudos associativos, dezenas de alelos podem ser acessados simultaneamente (WHITT e BUCKLER, 2003).

Além disso, deve-se destacar que uma das principais razões de interesse no mapeamento associativo em plantas está no fato de que, uma vez que não há necessidade de desenvolvimento de populações estruturadas para o mapeamento, um grande número de avaliações fenotípicas rotineiramente feitas nos

programas de melhoramento de plantas ao longo dos vários anos, como os próprios ensaios de VCU (Valor de Cultivo e Uso), podem ser usados na busca por identificar genes ou marcadores associados aos fenótipos de interesse (MALOSETTI et al., 2007).

Entretanto, é importante enfatizar que ambos os métodos de mapeamento são importantes e apresentam suas vantagens e limitações. Um ponto importante que deve ser salientado é que no mapeamento por desequilíbrio de ligação, no caso de haver interesse em identificar associações entre alelos raros, ou seja, alelos que ocorrem em baixa frequência na população, e um determinado fenótipo específico, pelas ferramentas estatísticas disponíveis, dificilmente por meio de mapeamento associativo vão ser detectadas estas associações (SIMKO, 2004).

Recentemente, alguns estudos vêm sendo desenvolvidos visando desenvolver metodologias que utilizem de forma conjunta o mapeamento de ligação clássico e o mapeamento por associação, a fim de gerar mapas altamente refinados que tiram proveito das vantagens de ambos os tipos de mapeamento (WU et al., 2002).

5. Mapeamento por associação: exemplos e perspectivas de uso no melhoramento de batata

O primeiro estudo associativo em plantas foi realizado em milho, adotando a estratégia gene-candidato. Neste estudo, a variação do gene *dwarf8* (*d8*) foi avaliada para a associação com a época de florescimento e estatura de planta em linhagens de milho (THORNSBERRY et al., 2001). Desde então, uma gama de espécies passaram a ser estudadas visando o mapeamento por associação. Gupta et al. (2005) fizeram uma revisão sobre os estudos associativos em plantas superiores onde são relatados estudos em mais de 15 espécies. Em batata,

os primeiros estudos sobre desequilíbrio de ligação e mapeamento associativo foram publicados em 2004 (SIMKO, 2004; SIMKO et al.; 2004a, b; GEBHARDT et al., 2004).

A batata (*Solanum tuberosum*) é uma espécie tetraplóide com herança tetrassômica. A complexidade dos estudos clássicos de mapeamento genético em populações tetraplóides, associado ao fato de que muitas fontes de resistência foram introduzidas das espécies silvestres diplóides para o germoplasma cultivado (4x), motivou os estudos de identificação de QTLs em populações diplóides de batata derivadas de cruzamentos inter-específicos (EWING et al., 2000; VISKER et al., 2003; OS et al., 2006). Mapas de ligação para identificação de QTLs em populações tetraplóides também foram desenvolvidos com sucesso, embora mostrem e confirmem a complexidade de construir mapas e avaliar QTLs em populações tetraplóides (BRADSHAW et al., 2004; BRYAN et al., 2004). Ainda que o mapeamento clássico de QTLs em populações diplóides e tetraplóides de batata tenha sido bem sucedido, o uso do mapeamento associativo no germoplasma tetraplóide de batata tem como grande vantagem a avaliação da presença de QTLs diretamente no mesmo nível de ploidia da batata cultivada e dentro de uma composição genética mais representativa para a espécie, com a possibilidade de usar o próprio germoplasma elite dos programas de melhoramento para compor a população em estudo (MALOSETTI et al., 2007).

Motivados pelas vantagens do mapeamento associativo, os primeiros estudos associativos em batata visaram avaliar o potencial de uso desta metodologia no germoplasma tetraplóide, o qual é composto fundamentalmente por variedades desenvolvidas pelos programas de melhoramento. Essas variedades são genótipos heterozigotos, fixados pela propagação vegetativa. Resultados iniciais sugerem que o declínio do desequilíbrio de ligação em função da distância genética em *Solanum tuberosum* é menos prolongado do que em espécies autógamas, como a soja, porém é mais prolongado do que em espécies alógamas, como o milho (SIMKO et al.,

2006).

Os resultados de estudos associativos são promissores em batata. Genes de resistência a *Verticillium* foram mapeados com sucesso em populações tetraplóides de batata, tanto usando a abordagem de gene-candidato para mapear QTLs de resistência a *Verticillium dahliae* (SIMKO et al., 2004a), como adotando a abordagem do genoma todo para mapear genes de resistência a *Verticillium albo-atrum* (SIMKO et al., 2004b). Gebhardt et al. (2004) identificaram por meio do mapeamento associativo usando a abordagem de gene-candidato, QTLs relacionados com a resistência a *Phytophthora infestans* e com o ciclo, ao avaliar 600 cultivares de batata desenvolvidas entre 1850 e 1990. Utilizando também a abordagem de genes-candidato, Malosetti et al. (2007) identificaram dois marcadores moleculares com potencial de uso para seleção de genótipos de batata com resistência a *Phytophthora infestans*, ao explorarem os dados fenotípicos coletados ao longo de 25 anos nos ensaios de VCU de batata da Holanda.

D'hoop et al. (2008), ao analisarem cultivares e alguns progenitores tetraplóides de batata, no total 221 genótipos, usando a abordagem do genoma todo com marcadores AFLP, identificaram associações interessantes entre marcadores e caracteres fenotípicos relacionados com a qualidade do tubérculo, mesmo tendo utilizado um número relativamente pequeno de marcadores no estudo, no total 250, comprovando o grande potencial dessa metodologia para identificar QTLs no germoplasma tetraplóide de batata.

Li et al. (2008), ao avaliarem o germoplasma tetraplóide de três programas de melhoramento de batata, adotando a estratégia de genes-candidato, identificaram associações entre marcadores moleculares e os caracteres fenotípicos coloração do *chips* após a fritura, conteúdo de amido, rendimento de tubérculo e rendimento de amido. Neste trabalho é mostrado o potencial do uso do mapeamento associativo para dissecar simultaneamente múltiplos caracteres complexos, altamente

influenciados pelo ambiente, em populações avançadas de melhoramento de batata.

Os exemplos de mapeamento associativo em batata têm mostrado resultados bastante interessantes, com grande capacidade de identificar associações entre marcadores moleculares e fenótipos de interesse para os programas de melhoramento (SIMKO et al., 2004a, b; GEBHARDT et al., 2004; MALOSETTI et al., 2007; D'HOOP et al., 2008). O mapeamento associativo surge como uma potencial solução aos problemas até então encontrados para o emprego de seleção assistida por marcadores moleculares em batata, técnica que ainda é pouca utilizada nos programas de melhoramento devido, principalmente, ao fato de que os estudos de QTL nesta espécie foram predominantemente desenvolvidos em populações diplóides e, freqüentemente, envolvendo cruzamentos interespecíficos, resultando na identificação de QTLs que não são imediatamente diagnosticados nas populações de melhoramento (GEBHARDT et al., 2007).

6. Considerações finais

O mapeamento por associação, bastante usado em genética humana e animal, surge como uma importante metodologia para mapear genes de interesse em plantas. Entretanto, alguns pontos devem ser ressaltados para que esta técnica seja usada com sucesso na identificação de genes de interesse nas coleções de germoplasma. É importante destacar que, sendo o desequilíbrio de ligação a base do mapeamento por associação, entender a estrutura do desequilíbrio de ligação na população em estudo é o ponto inicial básico e crucial para desenvolver estudos associativos. Como questões fundamentais para obter sucesso no mapeamento por associação, salienta-se a importância da escolha do germoplasma a ser estudado. Este deve ser o mais diverso possível, para cobrir a amplitude de variação fenotípica do caráter de interesse; a qualidade dos dados fenotípicos, que devem usufruir de desenhos

experimentais que permitam controlar o erro experimental; e o uso de métodos estatísticos que possibilitem controlar o problema de estrutura populacional, a fim de eliminar associações falso-positivas. A crescente facilidade de acesso a técnicas moleculares de grande cobertura do genoma em várias espécies vegetais, assim como a crescente identificação e disponibilidade de genes envolvidos em importantes rotas metabólicas, faz com que o uso de mapeamento associativo, seja por meio da abordagem do genoma como um todo, ou por meio de estudos gene-candidato, tenha um papel fundamental em facilitar o uso da diversidade genética pelos programas de melhoramento, usufruindo dos benefícios resultantes da era genômica. Os resultados encontrados em estudos associativos com batata, uma espécie com padrão de segregação complexo devido à herança tetrassômica, mostram-se extremamente promissores para elucidar a genética de caracteres agrônômicos complexos e desenvolver marcadores moleculares ligados a genes de interesse para serem usados nos programas de melhoramento genético de batata, visando acelerar o desenvolvimento de novas variedades.

7. Referências

ALPERT, K. B.; TANKSLEY, S. D. High-resolution mapping and isolation of yeast artificial chromosome contig containing *fw2.2*: A major fruit weight quantitative locus in tomato. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, p. 15503-15507, 1996.

BRADSHAW, J. E.; PANDE, B.; BRYAN, G. J.; HACKETT, C. A.; MCLEAN, K.; STEWART, H. E.; WAUGH, R. Interval mapping of quantitative trait loci for resistance to late blight *Phytophthora infestans* (Mont.) de Bary, height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). **Genetics**, Pittsburgh, v. 168, p. 983-995, 2004.

BRYAN, G. J.; MCLEAN, K.; PANDE, B.; PURVIS, A.; HACKETT,

C. A.; BRADSHAW, J. E.; WAUGH, R. Genetic dissection of H3-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. **Molecular Breeding**, Dordrecht, v. 14, p. 105-106, 2004.

COLLARD, B. C. Y.; JAHUFER, M. Z. Z.; BROUWER, J. B.; PANG, E. C. K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, Wageningen, v. 142, p. 169-196, 2005.

D'HOOP, B. B.; PAULO, M. J.; MANK, R. A.; van ECK, H. J.; van EEUWIJK, F. A. Association mapping of quality traits in potato (*Solanum tuberosum* L.). **Euphytica**, Wageningen, v. 161, p. 47-60, 2008.

EWING, E. E.; SIMKO, I.; SMART, C. D.; BONIERBALE, M. W.; MIZUBUTI, E. S. G.; MAY, G. D.; FRY, W. E. Genetic mapping from field tests of qualitative and quantitative resistance to *Phytophthora infestans* in a population derived from *Solanum tuberosum* and *Solanum berthaultii*. **Molecular Breeding**, Dordrecht, v. 6, p. 25-36, 2000.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4 ed. Essex: Longman Group, 1996. 464 p.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, Palo Alto, v. 54, p. 357-374, 2003.

GEBHARDT, C.; BALLVORA, A.; WALKEMEIER, B.; OBERHAGEMANN, P.; SCHULER, K. Assessing genetic potential in germplasm collection of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. **Molecular Breeding**, Dordrecht, v. 13, p. 93-102, 2004.

GEBHARDT, C. Molecular markers, maps and population genetics. In.: VREUGDENHIL, D. (Ed.). **Potato biology and biotechnology advances and perspectives**. Amsterdam: Elsevier, 2007. p. 77-89.

GEBHARDT, C.; LI, L.; PAJEROWSKA-MUKTHAR, K.; ACHENBACH, U.; SATTARZADEH, A.; BORMANN, C.; ILARIONOVA, E.; BALLVORA, A. Candidate gene approach to identify genes underlying quantitative traits and develop diagnostic markers in potato. **Crop Science**, Madison, v. 47(S3), p. S106-S111, 2007.

GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, Dordrecht, v. 57, p. 461-485, 2005.

HEDRICK, P. W. Gametic disequilibrium measures: proceed with caution. **Genetics**, Pittsburgh, v. 117, p. 331-341, 1987.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Berlin, v. 38, p. 226-231, 1968.

LANDER, E. S.; SCHORK, N. J. Genetic dissection of complex traits. **Science**, New York, v. 265, p. 2037-2048, 1994.

LI, L.; PAULO, M. J.; STRAHWALD, J.; LÜBECK, J.; HOFFERBERT, H. R.; TACKE, E.; JUNGHANS, H.; WUNDER, J.; DRAFFEHN, A.; EEUWIJK, F.; GEBHARDT, C. Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. **Theoretical and Applied Genetics**, Berlin, v. 116, p. 1167-1181, 2008.

MACKAY, I.; POWEL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, London, v. 12, p. 57-63, 2007.

MALOSETTI, M.; van der LINDEN, C. G.; VOSMAN, B.; van

EEUWIJK, F. A. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. **Genetics**, Pittsburgh, v. 175, p. 879-889, 2007.

MORGANTE, M.; SALAMINI, F. From plant genomics to breeding practice. **Current Opinion in Biotechnology**, London, v. 14, p. 214-219, 2003.

MULLIS, K. B.; FALOONA, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. **Methods in Enzymology**, San. Diego, v. 155, p. 335-350, 1987.

OS, H.; ANDRZEJEWSKI, S.; BAKKER, E.; BARRENA, I.; BRYAN, G. J.; CAROMEL, B.; GHAREEB, B.; ISIDORE, E.; DE JONG, W.; KOERT, P.; LEFEBVRE, V.; MILBOURNE, D.; RITTER, E.; VOORT, J. N.; ROUSSELLE-BOURGEOIS, F.; VLIET, J.; WAUGH, R.; VISSER, R.; BAKKER, J.; ECK, H. J. Construction of a 10,000-marker ultradense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genomewide physical map. **Genetics**, Pittsburgh, v. 173, p. 1075-1087, 2006.

PARISSEAUX, B.; BERNARDO, R. In silico mapping of quantitative trait loci in maize. **Theoretical and Applied Genetics**, Wageningen, v. 109, p. 508-514, 2004.

PRITCHARD, J. K.; STEPHENS, M.; ROSENBERG, N. A.; DONNELLY, P. Association mapping in structured populations. **American Journal of Human Genetics**, Pittsburgh, v. 67, p. 170-181, 2000.

PRITCHARD, J. K. Deconstructing maize population structure. **Nature Genetics**, Inglaterra, v. 28, p. 203-204, 2001

REMLINGTON, D. L.; THORNSBERRY, J. M.; MATSUOKA, Y.; WILSON, L. M.; WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M.; BUCKLER, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize

genome. **Genetics**, Pittsburgh, v. 98, p. 11479-11484, 2001.

SAIKI, R. K.; SCHARF, S.; FALOONA, F.; MULLIS, K. B.; HORN, G. T.; ERLICH, H. A.; ARNHEIM, N. Enzymatic amplification of beta-globin genomic sequences and restriction site analyses for diagnosis of sickle cell anemia. **Science**, New York, v. 230, p. 1350-1354, 1985.

SHARBEL, T. F.; HAUBOLD, B.; MITCHELL-OLDS, T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. **Molecular Ecology**, Hoboken, v. 9, p. 2109-2118, 2000.

SIMKO, I. One potato, two potato: haplotype association mapping in autotetraploids. **Trends in Plant Science**, London, v. 9, p. 441-448, 2004.

SIMKO, I.; COSTANZO, S.; HAYNES, K. G.; CHRIST, B. J.; JONES, R. W. Linkage disequilibrium mapping of *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. **Theoretical and Applied Genetics**, Berlin, v. 108, p. 217-224, 2004a.

SIMKO, I.; HAYNES, K.G.; EWING, E. E.; COSTANZO, S.; CHRIST, B. J.; JONES, R. W. Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association test and genetic linkage analysis. **Molecular Genetics and Genomics**, Berlin, v. 271, p. 522-531, 2004b.

SIMKO, I. HAYNES, K. G.; JONES, R. W. Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. **Genetics**, Pittsburgh, v. 173, p. 2237-2245, 2006.

SPOONER, D.; van TREUREN, R.; VICENTE de, M. C. **Molecular markers for genebank management**. Rome: International

Plant Genetic Resources Institute, 2005. 126 p. (IPGRI. Technical bulletin, 10).

STUBER, C. W.; POLACCO, M.; SENIOR, M. L. Synergy of empirical breeding, marker assisted selection and genomics to increase crop yield potential. **Crop Science**, Madison, v. 39, p. 1571-1583, 1999.

THORNSBERRY, J. M.; GOODMAN, M. M.; DOEBLEY, J. KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. S. *Dwarf8* polymorphisms associate with variation in flowering time. **Nature Genetics**, Inglaterra, v. 28, p. 286-289, 2001.

VISKER, M. H. P. W.; KEIZER, L. C. P.; VAN ECK, H. J.; JACOBSEN, E.; COLON, L. T.; STRUIK, P. Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? **Theoretical and Applied Genetics**, Berlin, v. 106, p. 317-325, 2003.

WHITT, S. R.; BUCKLER, E. S. Using natural allelic diversity to evaluate gene function. **Methods in Molecular Biology**, v. 236, p. 123-139, 2003.

WRAY, N. R.; VISSCHER, P. M. Population genetics and its relevance to gene mapping. In: NEALE, B. N. M.; FERREIRA, M. A. R.; MEDLAND, S. E.; POSTHUMA, D. (Ed.). **Statistical genetics: Gene mapping through linkage and association**. New York: Taylor & Francis Group, 2008. p. 87-112.

WU, R.; MA, C.; CASELLA, G. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. **Genetics**, Pittsburgh, v. 160, p. 779-792, 2002.

YU, J.; PRESSOIR, G.; BRIGGS, W. H.; BI, I. V.; YAMASAKI, M.; DOEBLEY, J. F.; MCMULLEN, M. D.; GAUT, B. S.; NIELSEN, D. M.; HOLLAND, J. B.; KRESOVICH, S.; BUCKLER, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, Inglaterra, v. 38, p. 203-208, 2006.