Existing potato markers and marker conversions

Walter De Jong
PAA Workshop
August 2009
What makes for a good marker?

- diagnostic for trait of interest
- robust – works even with DNA of poor quality or low quantity
- cheap
- easy to use
- co-dominant
- dosage sensitive
Agarose Gel Electrophoresis

Key assumption of this presentation: most of us will continue to use agarose gels for years to come

😊 Easy
😊 Universal
😊 Expensive per data point
😊 Low throughput
Example of a potato marker

PVY \( (Ry_{adg}) \) – Kasai et al. 2000 Genome 43:1-8

allele specific amplification of a diagnostic product
- potatoes with the product are resistant
- potatoes without the product are susceptible**

** or the PCR failed for some reason

This may be the most useful potato marker developed to date
How are allele specific markers made?

<table>
<thead>
<tr>
<th>Haplotype 1</th>
<th>Haplotype 2</th>
<th>Haplotype 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAAAAC TCTTGGTTAT AGCCTA-&gt;</td>
<td>AATGAGAGAA GACATAAAAAC TCTTGGTTAT AGCCTATAGG AGCCAGCTAC 100</td>
<td></td>
</tr>
<tr>
<td>haplotype 1</td>
<td>AATGAGAGAA GACACAAAAAC TCTTGGT-&gt; ---- TATAGG AGCCAGATAC 93</td>
<td></td>
</tr>
<tr>
<td>haplotype 2</td>
<td>AATGAGAGAA GACACAAAAAC TCTTGGT-&gt; ---- TATAGG AGCCAGCTAC 93</td>
<td></td>
</tr>
<tr>
<td>haplotype 3</td>
<td>TTTACCGGAA AGTTT-CCCA GAATACTTAT AGGATATCTT ATGTCAAATT 149</td>
<td></td>
</tr>
<tr>
<td>haplotype 1</td>
<td>TTTACCGGAG AGTT-CCCCA TAATACTTAT GTCAA----- ----- AATTTT 132</td>
<td></td>
</tr>
<tr>
<td>haplotype 2</td>
<td>TTTACCGGAA AATTTCCCAG AATACTT-AT AGGATATCTT ATGTCAAATT 142</td>
<td></td>
</tr>
<tr>
<td>haplotype 3</td>
<td>TCATTGAAGC AGTATTATTT TTTTTCTTGG AATTGGTGAA ACAACACATT 199</td>
<td></td>
</tr>
<tr>
<td>haplotype 1</td>
<td>TCA-TGATGC AGTATTATTT TTTC--TTGG AATTGGTGAA ACAACACATT 179</td>
<td></td>
</tr>
<tr>
<td>haplotype 2</td>
<td>TCATTGATGC AGTATTTTTT TTTC--TTGG AATTGGTGAA ACAACACATT 190</td>
<td></td>
</tr>
<tr>
<td>haplotype 3</td>
<td>&lt;-TAACCACCT TGTTGTGTA</td>
<td></td>
</tr>
</tbody>
</table>

One or both amplification primers anneal only to allele of interest.
A co-dominant CAPS marker:

- presence of upper band diagnostic for resistance
- absence of upper band diagnostic for susceptibility
- lower bands represent one or more additional alleles, and are present in most progeny
How do CAPS markers work?
(CAPS = cleaved amplified polymorphic sequence)

1. Amplify all alleles by PCR
2. Digest with a restriction enzyme that generates unique pattern with allele of interest

an unpublished marker linked to H1 (golden nematode resistance gene)

-- an allele-specific band is diagnostic for resistance
-- an unrelated band is used as a control for success of PCR
Is the marker robust?

two markers linked to root-knot resistance:

which would you rather use?
Not all DNA isolation methods yield pristine DNA

Cheap, inexpensive methods like Edwards quickprep result in DNA of variable quality.

Many commercial kits cost 1$/prep and yield exceptionally clean DNA.

In a breeding program, more likely to use quick-and-dirty methods – and primers that amplify well with clean DNA won’t necessarily amplify dirty DNA very well.

Try many primer pairs when designing assay you will use a lot – some primer pairs are more robust than others.
markers for resistance to potato pathogens

- PVY ($Ry_{adg}$) – Kasai et al. 2000 Genome 43:1-8
- PVY ($Ry_{sto}$) – Flis et al. 2005 Molecular Breeding 15: 95-101
- PVX resistance (Rx1) – Gebhardt et al. 2006 Theor Appl Genet 112:1458-1464
- Golden nematode (Gro1-4) – Gebhardt et al. 2006 Theor Appl Genet 112:1458-1464
- Golden nematode ($H1$) – ask Walter De Jong for a protocol (unpublished)
- White cyst nematode race Pa2/Pa3 – Sattarzadeh et al. 2006 Mol Breeding 18:301-312
- Root-knot nematode ($R_{Mc1(blb)}$) – Zhang et al. 2007 Crop Science 47:2021-2026
- Wart ($Sen1$) – Gebhardt et al. 2006 Theor Appl Genet 112:1458-1464
- Late blight ($RB$) – Colton et al. 2006 Crop Sci 46:589-594
some markers for potato quality traits

- Dominant allele at the red ($R$) locus – De Jong et al 2003 Theor Appl Genet 107:1375-1383


- Allele of invertase associated with chip color – Li et al 2005 Genetics 170:813-821

Useful website for markers in 19 crops

Generation Challenge Program Toolbox

http://s2.generationcp.org/gcp-tmm/web/

Many potato markers are described here
Why marker conversion?

The markers used to map traits are generally not the same markers that breeders are willing/able to use for MAS

In potato, RFLP, AFLP and SSR markers have most often been used for mapping. SolCAP will be using Infinium to detect SNPs.
- none of these are simple PCR markers well-suited for agarose gels

In the future, most of us will likely move to SNP-based platforms. We’re just not there yet.
SNP to PCR marker conversion

SolCAP will be testing 7600 SNP markers

Example of a SNP:

AAGTCTGTGT
AAGACTGTGT

SolCAP will convert SNP markers to simple PCR markers (small grants program)

Following slides discuss marker conversions in detail
- SNP
- RFLP
- AFLP
Creating a CAPS marker from a SNP requires:

- 25 bp of DNA sequence on each side of the SNP (for PCR primers to anneal to)
  - For each SolCAP SNP, a lot of flanking sequence will generally be known
- a computer with internet access
Identifying a restriction enzyme for a CAPS assay

Several applications automatically identify restriction enzymes that can be used to detect a SNP with a CAPS assay

- **Blastdigester**
  - [http://bar.utoronto.ca/ntools/cgi-bin/ntools_blast_digester.cgi](http://bar.utoronto.ca/ntools/cgi-bin/ntools_blast_digester.cgi)

- **SNPS2CAPS**
  - [http://pgrc.ipk-gatersleben.de/snp2caps/](http://pgrc.ipk-gatersleben.de/snp2caps/)

- **SGN CAPS designer**
  - [http://sgn.cornell.edu/tools/caps_designer/caps_input.pl](http://sgn.cornell.edu/tools/caps_designer/caps_input.pl)
Software comparison

- SGN CAPS is the easiest and simplest to use.
- However, it does not evaluate all RE’s commercially available.
- SNPS2CAPS is more comprehensive and flexible, but not as user friendly.
- Blastdigester finds SNPs in BLAST output
SGN CAPS designer

- This web-based tool accepts sequence input as aligned sequences (clustal format) or individual FASTA sequences

http://sgn.cornell.edu/tools/caps_designer/caps_input.pl

Introduction

This tool designs CAPS assays for two sequences. Two types of nucleotide inputs are accepted: fasta sequences and clustal alignment. It generates a list of polymorphic enzymes that cut the sequences into different length products. The user is allowed to input up to three sequences, however each time only two are analyzed.

Suggestions:

1. Low quality nucleotides and "n"s at both ends of a sequence generate ambiguity. Please remove them from the input sequences.
2. Polymorphic digested fragments of too small sizes are hard to visualize on 1-4% agarose gel, thus not suitable for CAPS experiment. Please exclude some nucleotides (for example 20) at both ends to avoid the problem.
SGN CAPS designer

First step: organize two sequences (one for each allele) in FASTA format

>allele1
AATAGCAGCATGTTGGGCATCCCTCCTCCTCACTGCCAATGCACTGTTGATTTTAATAGGAACACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTCT
ATTCTCATTTAGAAACAAAAACGTGAAAGTATTTGAAAGGAACACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTCT

>allele2
AATAGCAGCATGTTGGGCATCCCTCCTCCTCACTGCCAATGCACTGTTGATTTTAATAGGAACACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTCT
ATTCTCATTTAGAAACAAAAACGTGAAAGTATTTGAAAGGAACACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTCT
SGN CAPS designer

Second step: paste FASTA sequences into SGN CAPS designer, select desired options, then click “Find Caps”
If the first option is checked, the output will only display inexpensive RE’s that can be used for a CAPS marker.

Entering the default of 20 bp insures that the RE site is not within 20 bp of the end of the fragment. so the marker can be resolved by agarose gel electrophoresis.

**Options**

- Find enzymes priced less than $65/1000u.
- Exclude 0 nucleotides at both ends
- Don't show enzymes that cut both parents more than 4 times

If there are too many sites for an enzyme, the fragment will be cut into several small pieces, which will be difficult to resolve and score on agarose gels.
SGN CAPS designer

Output (truncated): This is the top of the output after clicking on “Find Caps”

It is a good idea to click on the link to view the clustal alignment to make sure results are reasonable.
A reasonable alignment, showing a single SNP

Example is from tomato, where SNPs are relatively uncommon
Factors to consider when choosing a RE:
1. Price
2. The number of fragments produced
3. Is the RE already in use in your lab?
4. Does the enzyme cut OK in PCR buffer?
5. Especially for an autotetraploid – the ideal enzyme cleaves all alleles except the allele you want to track – see next two slides
When designing CAPS markers, if possible, choose an enzyme that doesn’t cut allele of interest

1. Larger bands are easier to score on an agarose gel (more ethidium binding sites per molecule = brighter)
When designing CAPS markers, if possible, choose an enzyme that doesn’t cut allele of interest.

2. In a heterodimer, the absence of a restriction site is dominant over the presence of a restriction site.

In last cycle of PCR, many heterodimers are formed as strands of different alleles anneal to each other.

1/4  
1/4  
1/4  
1/4  

1/16  
6/16  
9/16  

7/16 don’t cut

Eco RI
Sometimes the SGN CAPS designer does not identify any useful REs. This might be because there are no RE sites created by any SNPs present.

However, note that the SGN CAPS designer does not test all commercially available RE’s. For a more comprehensive analysis, you can use SNPS2CAPS.

Restriction sites can also be created by designing primers with one or two mismatches that anneal in the vicinity of the SNP. This is the “dCAPs” approach; for a reference see Neff et al 2002 Trends in Genetics 18:613-615
How RFLP markers work

cDNA or genomic DNA probe
Two sources of RFLP marker polymorphism:
- polymorphic restriction sites
- large deletions/insertions in sequence between restriction sites

If you sequence an RFLP probe, you can’t tell what the polymorphism was that gave rise to the polymorphic band
- all you have is the sequence of a small part of the genome, one allele of which happens to be (tightly) linked to trait of interest
alleles, not genes, are key in breeding

How a molecular biologist thinks about the relationship between an RFLP probe and a gene:

- one gene
- one probe
- chromosome
alleles, not genes, are key in breeding

How a breeder thinks about the relationship between an RFLP probe and a gene:

many alleles of the gene
many alleles of the probe

chromosome diversity across species
important issue in marker conversion: re-establishing the relationship between PCR marker and trait alleles

➢ a simple way to do this:
  • identify all alleles of new PCR marker
  • test each allele in turn, to see which (if any) correlates with desired trait allele

➢ sometimes no correlation is found
  • can occur if primers don’t amplify necessary marker allele (lots of SNPs in potato; can occur at primer binding site)
  • can also occur if primers amplify a different part of the genome entirely (problem in repetitive DNA)
**typical sequence variation / haplotype structure in potato (observed in pooled DNA, 10 cultivars)**

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Polymorphic positions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28 68 125 132 134 141 151 157 183 207 214</td>
</tr>
<tr>
<td>a</td>
<td>G A C T G A - - A - G</td>
</tr>
<tr>
<td>b</td>
<td>G A C C A G - - A G G</td>
</tr>
<tr>
<td>c</td>
<td>A G C C G A - - A - G</td>
</tr>
<tr>
<td>d</td>
<td>G G C C A A - - T - G</td>
</tr>
<tr>
<td>e</td>
<td>G G A C G A A G C - A</td>
</tr>
</tbody>
</table>

**COSII_At3g54470; 19 clones sequenced; first 215 bp; orotate phosphoribosyltransferase**

**COSII_At4g26750; 17 clones sequenced; 367 bp unknown protein; matches tomato EST DB722697 over entire length**

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Polymorphic positions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>47 48 76 101 111 128 188 229 308 309 310</td>
</tr>
<tr>
<td>a</td>
<td>C A A G A A T T - - -</td>
</tr>
<tr>
<td>b</td>
<td>T T A G G C A G G T G</td>
</tr>
<tr>
<td>c</td>
<td>T T G A G A T T G T G</td>
</tr>
</tbody>
</table>
How AFLP markers work

- Genomic DNA
- Restriction enzyme digestion
- Adaptor ligation
- Selective PCR amplification
- AFLP fingerprint
AFLP to PCR marker conversion

- Two sources of AFLP marker polymorphism:
  - polymorphic restriction sites
  - polymorphic SNPs in sequence adjacent to adaptors (affect primer binding)

- Just as for an RFLP probe - if you clone and sequence an AFLP band, you can’t immediately ID the polymorphism that gave rise to the band
  - Unlike an RFLP probe, AFLP bands are generally small – this provides less sequence to design primers against, and less sequence polymorphism to work with
  - It is often necessary to clone and sequence flanking sequence to obtain useful polymorphism (e.g. by inverse PCR)
If you want to clone and sequence an AFLP band, make sure the clone is the correct size before sequencing it.
once potato genome sequence becomes available...

...there will be huge amounts of flanking sequence around each SNP, providing essentially infinite ways to design primers for useful PCR assays
assessing allele dosage

➢ for the more advanced – two references that describe general approaches for measuring dosage of specific alleles in potato:
  • TaqMan (De Jong et al 2003 Theor Appl Genetics 107:1384-1390)
  • High Resolution Melting (De Koeyer et al 2009 Molecular Breeding DOI 10.1007/s11032-009-9309-4)