

# Designing CAPS markers using SGN CAPS Designer

Matthew Robbins and Heather Merk  
The Ohio State University, OARDC

# Objective

- Design a CAPS marker using SGN CAPS designer
- <http://solgenomics.net/>

The image shows a screenshot of the Sol Genomics Network (SGN) website. At the top, the logo is a sun-like icon with rays. The text 'sol genomics network' is displayed in the center, with navigation links for 'home | forum | contact | help' on the right. Below this is a search bar and a menu with options: 'search', 'maps', 'genomes', 'tools', and 'sol search'. The 'tools' menu is expanded, showing a list of tools categorized into 'Sequence Analysis', 'Mapping', 'Molecular Biology', 'Systems Biology', 'Bulk Query', and 'Other'. The 'CAPS Designer' tool is highlighted in blue, and a red arrow points to it from the right. Other tools listed include BLAST, Alignment Analyzer, Tree Browser, Intron Finder, Comparative Viewer, Seed BAC Finder, solQTL: QTL Mapping, Signal Peptide Finder, In Silico PCR, SolCyc Biochemical Pathways, Unigene and BAC information, FTP Site, ID Converter (SGN <=> TIGR), and SGN Ontology Browser [beta]. The background of the website features several icons: 'Maps & Markers' (a chromosome map), 'Breeder's Toolbox' (a toolbox with vegetables), 'Genomes & Sequences' (a DNA double helix), 'Phenotypes' (a tomato), and 'Pathways' (a metabolic pathway diagram).

# This tutorial requires:

- Background information
  - Minimum: 20 bp of DNA sequence flanking a SNP; recommended entire sequence between PCR primers that amplify a region flanking a SNP
  - PCR primer design is not part of this tutorial, but primers are required to detect the SNP
- A computer with internet access

# Introduction to CAPS

- CAP(S): Cleaved/cut amplified polymorphic (sequences)
  - (Konieczny and Ausubel, 1993) A CAP is based on a sequence polymorphism that creates or eliminates a restriction endonuclease (RE, also restriction enzyme) recognition site

# CAPS Marker Example

Individual A

...GAGCG**CCGG**AA...

Individual B

...GAGCG**TGG**AA...

MseI restriction enzyme

recognition sequence: **CCGG**

- Individual A has an MseI recognition site (**blue**)
- The SNP between individuals A & B (**red**) eliminates the recognition site in individual B

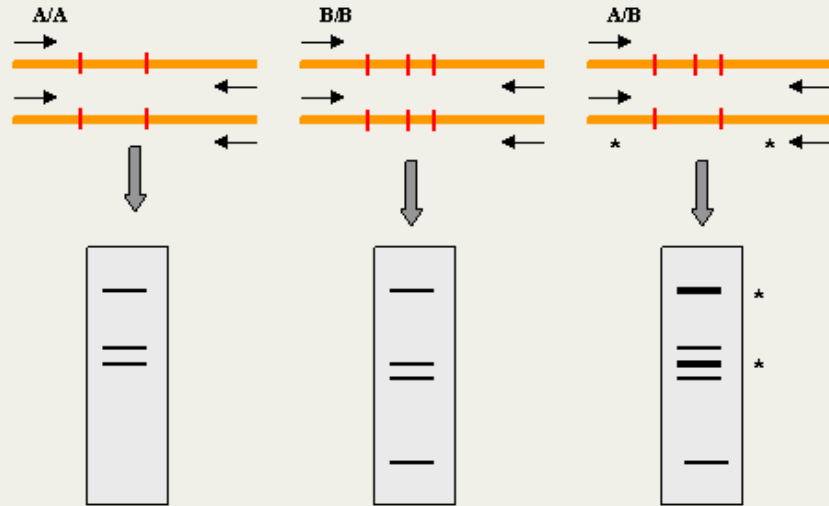
# Steps to Detect CAPS Markers

1. PCR amplification with primers flanking the SNP
2. Digestion of PCR products by the appropriate restriction enzyme
3. Gel electrophoresis to detect fragment length polymorphisms

## How It Works

The CAPS assay uses amplified DNA fragments that are digested with a restriction endonuclease to display RFLP.

### CAPS assay: amplification – digestion – gel separation



Unique sequence primers are used to amplify a mapped DNA sequence from two related individuals (for example, from two different inbred ecotypes), A/A and B/B, and from the heterozygote A/B. The amplified fragments from A/A and B/B contain two and three RE recognition sites, respectively. In the case of the heterozygote A/B, two different PCR products will be obtained, one which is cleaved three times and one which is cleaved twice. When fractionated by agarose or acrylamide gel electrophoresis, the PCR products digested by the RE will give readily distinguishable patterns. Some bands will appear as doublets.

#### Advantages

- Most CAPS markers are co-dominant and locus-specific.
- Most CAPS genotypes are easily scored and interpreted.
- CAPS markers are easily shared between laboratories.
- CAPS assay does not require the use of radioactive isotopes, and it is more amenable, therefore, to analyses in clinical settings.

[www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml](http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml)

# Identifying Restriction Enzymes to detect CAPS

- Several applications automatically identify which restriction enzymes can be used to detect a SNP as a CAPS marker
  - SGN CAPS designer – focus of this tutorial
    - [http://sgn.cornell.edu/tools/caps\\_designer/caps\\_input.pl](http://sgn.cornell.edu/tools/caps_designer/caps_input.pl)
  - SNPS2CAPS
    - <http://pgrc.ipk-gatersleben.de/snp2caps/>
    - [\(Thiel et al, 2004\)](#)
  - Blastdigester
    - [http://bar.utoronto.ca/ntools/cgi-bin/ntools\\_blast\\_digester.cgi](http://bar.utoronto.ca/ntools/cgi-bin/ntools_blast_digester.cgi)
    - [\(Ilic et al, 2004\)](#)



# SGN CAPS designer

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



sol genomics network

home | forum | contact | help

search

maps

sequencing

tools

sol search

[log in](#)

## CAPS Designer

### 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.

[http://sgn.cornell.edu/tools/caps\\_designer/caps\\_input.pl](http://sgn.cornell.edu/tools/caps_designer/caps_input.pl)

# Step 1: Organize sequences in FASTA format

locus name

SNP base

>CT10649\_C

```
AATAGCAGCATGGTGGGCATCCTCTGTCTCCACTGCTCGAATCCTTCTCCGGCAATGCACCATGTTTTATTAAGAAGCTCTCCATCTTCTTCGTGAATATATCATTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTC
ATTCTACATCTTAGAACAAAAACGTGAAGATTCAAGATCATAGACTTGGAAAAGTGATGAAAGAGCG C CGGAATCCAGAACACCAAATACATGCTTGCACATAACGTATTTCTGTCAAACCTAACAAACGTCAAGTAACTCCCAA
AACCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTAATGATTAACAAAGCAAGAACATGACCTATTGAGATAATCTCAAAGGCAATTGTGTATAAAGAATGTTGTTATTCCACTTGAATGACTCA
ATCAAACGGAAAAATTGCATGTAACACCCATCTTCATCTAGAAATTTCAAATGATGAGGCAGAGAAAGATATGAAAAACCAATAAACTTGAGTTCCATTTTCAAACCACAAAAGTGAATCCACCAGTTCCAAAAAATTGTGCAGACT
AAACTTATTCATGAAGTAAGTATGTCACAATGGCAAAAGAAAAGAAATCATTGGTGGTACATACTGCTGCAAGACTGTATTTTCCCTCAAGATTTACATAATGCCAGCATTAGAGCTGTTTTCTGCATATAAAGAACAGTTAGTC
CCTCAAATCTCAACATGTTCAACAAAATTTTACAAGCTAATAAACAAGAAAAGAACTGCGATTGGAGGAAAAGCAAAAGCAAACACTAATACTACAAAAACAATAACATACCTGTGAAGTCACTCCAGTACATTATGAACCTTGA
ACTGAAAAGGAAAATGCTCTAGCACACTCATAACACATTACACAGTCAGATATGTGTCTAATGGAACAATTGTGTTCTATATGCAGAGTTCTAAAGATTCAATTTTTTATCAATAAAAAATGGTCCCTTTTGTTCCTACTTGGGTT
GCTGCAGCTAAAAGAAAATCCTACTTACAACAGATACCAAAGCTACTAAATATCATCTCCCTTCTACTTTTCTCAAAGATTGAATTTTTCTTCTCAAATACTGAAAACCTTTCTACTGAAACACACATCCCAAGACATAAAT
TAAGAAAAATTGAGGAAAAGAAAATACCCAGTACCAACAGGGCCACCAATTCGAATAGTAAAGGCTTTTCACTGAAATTCCTGTCTTAAGTGGAGGTGCCTTCTGCTAAAGTAGCCAGGTGAATAAATAGG
```

>CT10649\_T

```
AATAGCAGCATGGTGGGCATCCTCTGTCTCCACTGCTCGAATCCTTCTCCGGCAATGCACCATGTTTTATTAAGAAGCTCTCCATCTTCTTCGTGAATATATCATTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTC
ATTCTACATCTTAGAACAAAAACGTGAAGATTCAAGATCATAGACTTGGAAAAGTGATGAAAGAGCG T CGGAATCCAGAACACCAAATACATGCTTGCACATAACGTATTTCTGTCAAACCTAACAAACGTCAAGTAACTCCCAA
AACCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTAATGATTAACAAAGCAAGAACATGACCTATTGAGATAATCTCAAAGGCAATTGTGTATAAAGAATGTTGTTATTCCACTTGAATGACTCA
ATCAAACGGAAAAATTGCATGTAACACCCATCTTCATCTAGAAATTTCAAATGATGAGGCAGAGAAAGATATGAAAAACCAATAAACTTGAGTTCCATTTTCAAACCACAAAAGTGAATCCACCAGTTCCAAAAAATTGTGCAGACT
AAACTTATTCATGAAGTAAGTATGTCACAATGGCAAAAGAAAAGAAATCATTGGTGGTACATACTGCTGCAAGACTGTATTTTCCCTCAAGATTTACATAATGCCAGCATTAGAGCTGTTTTCTGCATATAAAGAACAGTTAGTC
CCTCAAATCTCAACATGTTCAACAAAATTTTACAAGCTAATAAACAAGAAAAGAACTGCGATTGGAGGAAAAGCAAAAGCAAACACTAATACTACAAAAACAATAACATACCTGTGAAGTCACTCCAGTACATTATGAACCTTGA
ACTGAAAAGGAAAATGCTCTAGCACACTCATAACACATTACACAGTCAGATATGTGTCTAATGGAACAATTGTGTTCTATATGCAGAGTTCTAAAGATTCAATTTTTTATCAATAAAAAATGGTCCCTTTTGTTCCTACTTGGGTT
GCTGCAGCTAAAAGAAAATCCTACTTACAACAGATACCAAAGCTACTAAATATCATCTCCCTTCTACTTTTCTCAAAGATTGAATTTTTCTTCTCAAATACTGAAAACCTTTCTACTGAAACACACATCCCAAGACATAAAT
TAAGAAAAATTGAGGAAAAGAAAATACCCAGTACCAACAGGGCCACCAATTCGAATAGTAAAGGCTTTTCACTGAAATTCCTGTCTTAAGTGGAGGTGCCTTCTGCTAAAGTAGCCAGGTGAATAAATAGG
```

Sequence names (after the “>” symbol) consist of a locus name (CT10649), an underscore (\_), and the SNP base (C or T)

# Step 2: Input Sequences

## Query Input

### Input format

- clustal alignment [What is this?]  
 unaligned fasta sequences

Select this input since our sequences are in FASTA format



### Input sequences

For this tutorial, simply paste the sequences from the previous slide in this box.

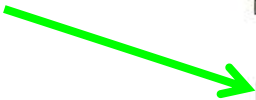
### Options

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

These options can be changed as needed and are explained on the next slide

Find Caps

Reset



# CAPS Designer Options

- If this option is checked, the output will only display *inexpensive* RE's that can be used for a CAPS marker.
- Limits the number of RE's that could be used, but produces a more cost-effective CAPS marker.

## Options

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

- If there are too many RE sites for the same enzyme, the fragment will be cut into several small pieces
- This will produce a complex pattern of bands that will be difficult to resolve and score on agarose gels.

- If the RE site is close to the edge of the fragment, digestion will produce a very short fragment and a long fragment almost the same size as the undigested fragment.
- It is difficult to resolve the long piece of the digested fragment and the undigested fragment on agarose gels unless there is > 20 bp difference.
- Entering the default of 20 bp ensures that the RE site is not within 20 bp of the end of the fragment.
- This option is applicable only if the position of the SNP is unknown or near the edge of the sequence.
- For this tutorial, we know that the SNP in the CT10649 locus is not near the edge of the sequence, so we enter a "0".

# Output

Output can be copied and pasted directly into any word processing program that supports HTML. Alternatively, the results can be obtained in plain text format by clicking the top link (1).

Click on link (2) to view the clustal alignment to make sure our results are based on the true SNP.


## CAPS Designer Result

### For experienced users

- 1 [View/download plain text result file](#)
- 2 [View/download alignment in clustal format](#)  
[View/download alignment in fasta format](#)

[Back to input](#)

### Notes

1. Polymorphism caused by an ambiguous nucleotide 'N' is not considered.
2. Please check the provided local alignment around the predicted CAPs site in order to make sure it's not caused by alignment gaps.
3. Analysis is based on sequenced parts of the PCR products. Additional cutting sites and digested fragments may exist.
4. Enzymes separated by a slash are isoschizimers.
5. Enzyme price is based on NEB catalogue ([http://www.neb.com/nebecomm/price\\_list.pdf](http://www.neb.com/nebecomm/price_list.pdf)) .

### Query Summary

<b>Aligned Sequences</b>	CT10649_C, CT10649_T		
<b>Alignment Length(w/ gaps)</b>	1330 bp	<b>Search Range</b>	1 - 1330 bp
<b>Cutting Sites Limit</b>	4		
<b>Enzyme Selection</b>	All		

# CLUSTAL Alignment

CLUSTAL W (1.83) multiple sequence alignment

```
CT10649_C      AATAGCAGCATGGTGGGCATCCTCCTGTCTCCACTGCTCGAATCCTTTCCTCCGGCAATG
CT10649_T      AATAGCAGCATGGTGGGCATCCTCCTGTCTCCACTGCTCGAATCCTTTCCTCCGGCAATG
*****
```

The alignment illustrates that the correct SNP was identified (the missing asterisk).

```
CT10649_C      CACCATGTTTTATTAAGAACTCTCCATCTTCTTTTCGTGAATATATCATTGTACCTACA
CT10649_T      CACCATGTTTTATTAAGAACTCTCCATCTTCTTTTCGTGAATATATCATTGTACCTACA
*****
```

This alignment can be copied and pasted into a word processing or a text document for archival purposes.

```
CT10649_C      AAGAAATGCATCTTAACATGGACATTAATTCATTCTACATCTTAGAACAAAAACGTGAAG
CT10649_T      AAGAAATGCATCTTAACATGGACATTAATTCATTCTACATCTTAGAACAAAAACGTGAAG
*****
```

```
CT10649_C      ATTTCAAGATCATAGACTTGGAAAAGTGATGAAAGAGCGCCGGAATTCAGAACACCAAAAT
CT10649_T      ATTTCAAGATCATAGACTTGGAAAAGTGATGAAAGAGCGTTCGGAATTCAGAACACCAAAAT
*****
```

This is the correct SNP ↗

Closing the new window or tab will return us to the full results.

```
CT10649_C      ACATGCTTGCACATAACGTATTTCTGTCAAACCTAACAAAACGTCAAGTAAACTCCCAAAA
CT10649_T      ACATGCTTGCACATAACGTATTTCTGTCAAACCTAACAAAACGTCAAGTAAACTCCCAAAA
*****
```

```
CT10649_C      CCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTTAATGA
CT10649_T      CCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTTAATGA
*****
```

```
CT10649_C      TTAAAAAACAAGAACATGACCTATTGAGATAATCTCCAAAGGCAATTGTGTATAAAGAAT
CT10649_T      TTAAAAAACAAGAACATGACCTATTGAGATAATCTCCAAAGGCAATTGTGTATAAAGAAT
*****
```

```
CT10649_C      GTTGTTATTCACCTTGAAATGACTCAATCAAACGGAAAATTGCATGTAACACCCATCTTC
CT10649_T      GTTGTTATTCACCTTGAAATGACTCAATCAAACGGAAAATTGCATGTAACACCCATCTTC
*****
```

# Candidate CAPS

CAPS Candidates

Enzyme	CfoI/HhaI	Price	HhaI,2000u/\$53
Recognition Sequence	GCGC		
CT10649_C Cutting Site(s)	219	CT10649_C Fragments(s),bp	219 1111
CT10649_T Cutting Site(s)	None	CT10649_T Fragments(s),bp	1330
CAPS Site	219	CT10649_C ggaaagtgatgaaagaGCGCcggaattcca CT10649_T ggaaagtgatgaaagaGCGTcggaattcca	

Enzyme	Hpy188I	Price	Hpy188I,1000u/\$58
Recognition Sequence	TC.GA		
CT10649_C Cutting Site(s)	943	CT10649_C Fragments(s),bp	943 387
CT10649_T Cutting Site(s)	223 943	CT10649_T Fragments(s),bp	223 720 387
CAPS Site	223	CT10649_C agtgatgaaagagcgCCGGAattccagaac CT10649_T agtgatgaaagagcgTCGGAattccagaac	

Enzyme	Hpy99I	Price	over \$65/1000u
Recognition Sequence	CG[T]A]CG		
CT10649_C Cutting Site(s)	None	CT10649_C Fragments(s),bp	1330
CT10649_T Cutting Site(s)	221	CT10649_T Fragments(s),bp	221 1109
CAPS Site	221	CT10649_C aaagtgatgaaagagCGCCGgaattccaga CT10649_T aaagtgatgaaagagCGTCGgaattccaga	

Enzyme	LpnI/Bsp143II/BstH2I /HaeII	Price	over \$65/1000u
Recognition Sequence	[A]G]GCGC[C]T]		
CT10649_C Cutting Site(s)	220	CT10649_C Fragments(s),bp	220 1110
CT10649_T Cutting Site(s)	None	CT10649_T Fragments(s),bp	1330
CAPS Site	220	CT10649_C gaaagtgatgaaagAGCGCCggaattccag CT10649_T gaaagtgatgaaagAGCGTCggaattccag	

Enzyme	MspI	Price	MspI,5000u/\$53
Recognition Sequence	CCGG		
CT10649_C Cutting Site(s)	55 222	CT10649_C Fragments(s),bp	55 167 1108
CT10649_T Cutting Site(s)	55	CT10649_T Fragments(s),bp	55 1275
CAPS Site	222	CT10649_C aagtgatgaaagagcgCCGGAattccagaa CT10649_T aagtgatgaaagagcgTCGGAattccagaa	

For this sequence, we have a choice of five RE's.

# How to Choose a Restriction Enzyme

1. Price. Prices given by SGN CAPS designer are typically close to the current value and are useful for comparison among candidate RE's.

2. The number of fragments produced by digestion of the PCR products. More fragments means a more complex banding pattern on a gel, which may be more difficult to interpret.

3. If any of the RE's are already in use in your lab. Familiarity with RE's is an advantage.

Once you have chosen which RE to use, follow the manufacturer's recommendations for digestion of PCR products.

CAPS Candidates

Enzyme	CfoI/HhaI	Price	HhaI,2000u/\$53
Recognition Sequence	GCGC		
CT10649_C Cutting Site(s)	219	CT10649_C Fragments(s),bp	219 1111
CT10649_T Cutting Site(s)	None	CT10649_T Fragments(s),bp	1330
CAPS Site	219	CT10649_C ggaaagtgatgaaagaGCGCcggaattcca CT10649_T ggaaagtgatgaaagaGCGTcggaattcca	

Enzyme	Hpy188I	Price	Hpy188I,1000u/\$58
Recognition Sequence	TC.GA		
CT10649_C Cutting Site(s)	943	CT10649_C Fragments(s),bp	943 387
CT10649_T Cutting Site(s)	223 943	CT10649_T Fragments(s),bp	223 720 387
CAPS Site	223	CT10649_C agtgatgaaagagcgCGGAattccagaac CT10649_T agtgatgaaagagcgTCGGAattccagaac	

Enzyme	Hpy99I	Price	over \$65/1000u
Recognition Sequence	CG[T A]CG		
CT10649_C Cutting Site(s)	None	CT10649_C Fragments(s),bp	1330
CT10649_T Cutting Site(s)	221	CT10649_T Fragments(s),bp	221 1109
CAPS Site	221	CT10649_C aaagtgatgaaagagCGCCGgaattccaga CT10649_T aaagtgatgaaagagCGTCGgaattccaga	



# What if there are no Candidate CAPS?

## CAPS Designer Result

### For experienced users

[View/download plain text result file](#)

[View/download alignment in clustal format](#)

[View/download alignment in fasta format](#)

[Back to input](#)

### Notes

1. Polymorphism caused by an ambiguous nucleotide 'N' is not considered.
2. Please check the provided local alignment around the predicted CAPs site in order to make sure it's not caused by alignment gaps.
3. Analysis is based on sequenced parts of the PCR products. Additional cutting sites and digested fragments may exist.
4. Enzymes separated by a slash are isoschizimers.
5. Enzyme price is based on NEB catalogue ([http://www.neb.com/nebecomm/price\\_list.pdf](http://www.neb.com/nebecomm/price_list.pdf)) [↗](#).

### Query Summary

Aligned Sequences	Slocus699_A, Slocus699_T	Search Range	1 - 704 bp
Alignment Length(w/ gaps)	704 bp		
Cutting Sites Limit	4		
Enzyme Selection	All		

**CAPS Candidates** None

This may occur if there is no RE available to recognize the SNP sequence.

However, the SGN CAPS designer does not test all commercially available RE's. For a more comprehensive analysis, the SNPS2CAPS program may be used ([pgrc.ipk-gatersleben.de/snp2caps/](http://pgrc.ipk-gatersleben.de/snp2caps/))

# References

- Ilic, K., T. Berleth and N.J. Provart. 2004. BlastDigester - a web-based program for efficient CAPS marker design. Trends in Genetics 20:280-283.
- Konieczny, A. and F.M. Ausubel. 1993. A procedure for mapping arabidopsis mutations using codominant ecotype-specific pcr-based markers. Plant Journal 4:403-410.
- Thiel, T., R. Kota, I. Grosse, N. Stein and A. Graner. 2004. SNP2CAPS: A SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res. 32:e5.

# External Links

- CAPS Designer [Online]. Sol Genomics Network. Boyce Thompson Institute. Available at: [solgenomics.net/tools/caps\\_designer/caps\\_input.pl](http://solgenomics.net/tools/caps_designer/caps_input.pl) (verified: 6 Dec 2010).
- Cleaved amplified polymorphic sequences [Online]. U.S. National Library of Medicine, National Institutes of Health. Available at: [www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml](http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml) (verified 7 Dec 2010).
- Provar, N. BlastDigester. [Online]. The Bio-Array Resource for Plant Biology, University of Toronto. Available at: [bar.utoronto.ca/ntools/cgi-bin/ntools\\_blast\\_digester.cgi](http://bar.utoronto.ca/ntools/cgi-bin/ntools_blast_digester.cgi) (verified 7 Dec 2010).
- SNP2CAPS [Online]. Plant Genome Resources Center, Leibniz Institute of Plant Genetics and Crop Plant Research. Available at: [pgrc.ipk-gatersleben.de/snp2caps/](http://pgrc.ipk-gatersleben.de/snp2caps/) (verified 7 Dec 2010).