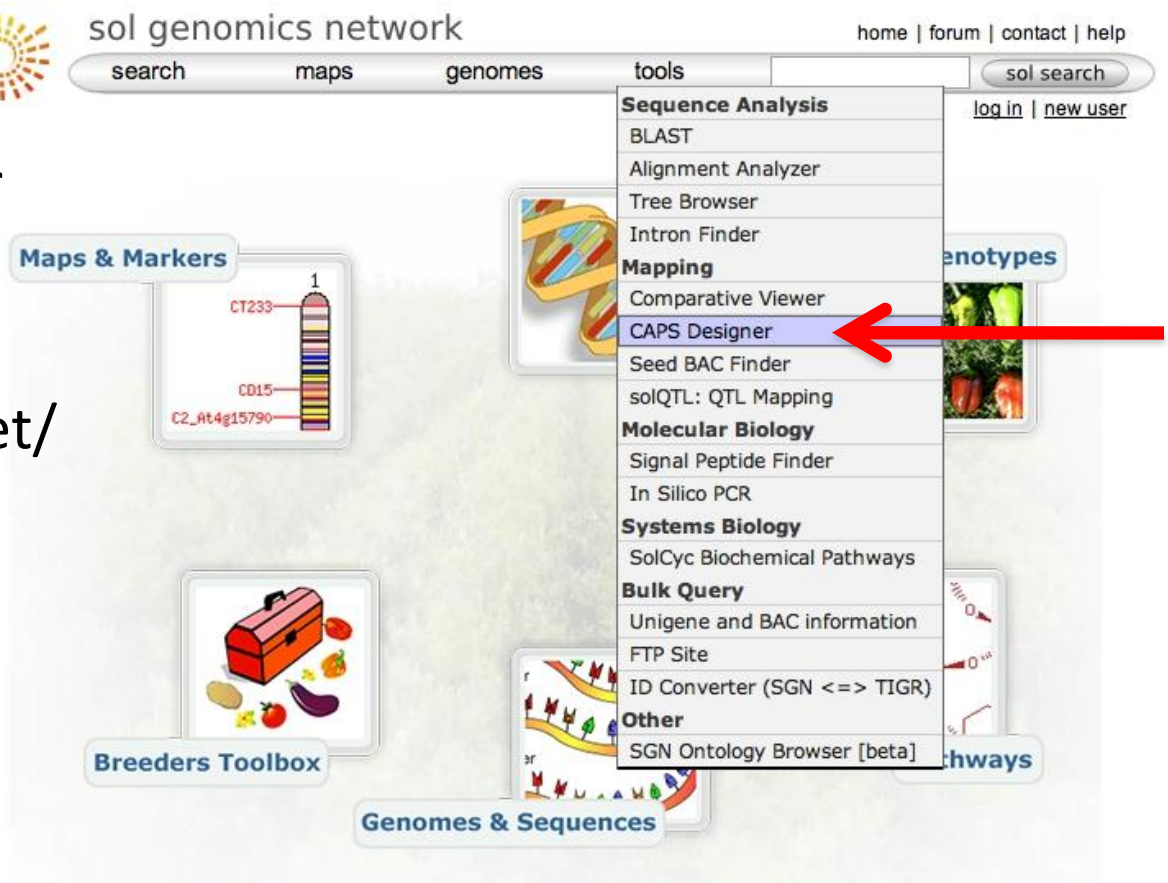
Designing CAPS markers using SGN CAPS Designer

Matthew Robbins and Heather Merk

The Ohio State University, OARDC

Objective

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



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 an restriction endonuclease (RE, also restriction enzyme) recognition site

CAPS Marker Example

Individual A …GAGCGCCGGAA…
Individual B …GAGCGCTGGAA…

MseI restriction enzyme recognition sequence: CCGG

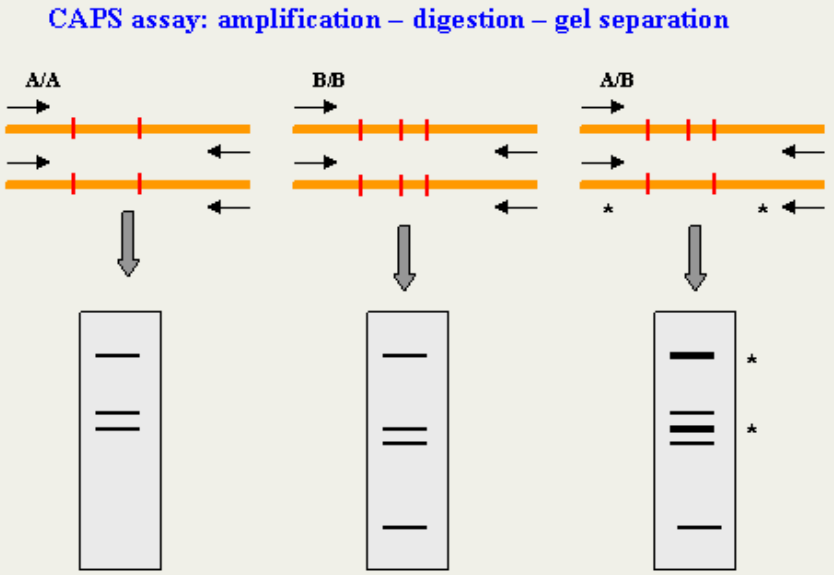
- Individual A has an MseI recognition site (blue)
- The SNP between individuals A & B (red) eliminates the recognition site in individual B

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.



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.

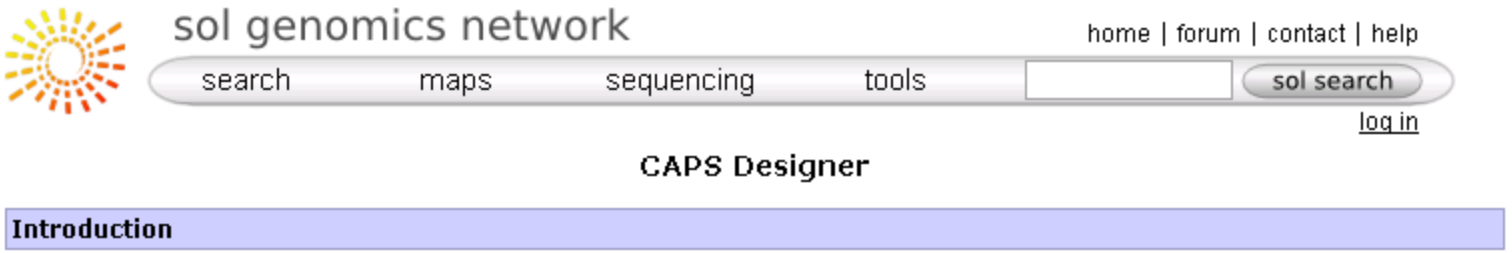
www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml

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
  - SNPS2CAPS
    - http://pgrc.ipk-gatersleben.de/snp2caps/
    - (Thiel et al, 2004)
  - Blastdigester
    - 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



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

Step 1: Organize sequences in FASTA format

locus name

SNP base

>CT10649_C

AATAGCAGCATGGTGGGCATCCTCCTGTCTCCACTGCTCGAATCCTTTCCTCCGGCAATGCACCATGTTTTATTAAGAACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTC

ATTCTACATCTTAGAACAAAAACGTGAAGATTTCAAGATCATAGACTTGGAAAGTGATGAAAGAGCGCCGGAATTCCAGAACACCAAATACATGCTTGCACATAACGTATTTCCTGTCAAACTAACAAACGTCAAGTAAACTCCCAA

AACCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTTAATGATTAAAAAACAAGAACATGACCTATTGAGATAATCTCCAAAGGCAATTGTGTATAAAGAATGTTGTTATTCCACTTGAAATGACTCA
ATCAAACGGAAAATTGCATGTAACACCCATCTTCATCTAGAATTTCAAATGATGAGGCAGAGAAAGATATGAAAACCACCAATAAACTTGAGTTCCATTTTCCAAACCACAAAAGTGAATCCACCAGTTCCAAAAAATTGTGCAGACT
AAACTTATTCATGAAGTAAGTATGTCACAATGGCAAAAGAAAGAAATCATTTGGTGGTACATACTGCTGCAAGACTGTATTTTTCCCTCAAGATTTTACATAATGCCAGCATTAGAGCTGTTTTCCTGCATATAAAAGAACAGTTAGTC
CCTCAAAATCTCAACATGTTCAACCAAAATTTTACAAGCTAATAAACAAGAAAGAAACTGCGATTGGAGGAAAAGCAAAAGCAAACACTAATACTACAAAAACAATAACATACCCTGTGAAGTCACTCCAGTACATTATGAACCTTGA
ACTGAAAAGGAAAATGCTCTAGCACACTCATACACACATTCACACAGTCAGATATGTGTCTAATGGAACAATTGTGTTCCTATATGCAGAGTTCTAAAGATTCAATTTTTTTATCAATAAAAATGGTCCCTTTTGTTTCTTACTTGGGTT
GCTGCAGCTAAAAGAAAATCCTACTTACAACAGATACCAAAAGCTACTAAATATCATCCTCCCCCTTCTACTTTCATTTCTCAAAGATTGAATTTTTCTTCTCAAATACTGAAAACCCTTTCACTTGAACACACATCCCAAGACATAAATT
TAAGAAAAATTGAGGGAAAAGAAGAAATACCCAGTACCAACAGGGCCACCAATTCCAATAGTAAAGGCTCTTTCACTGAAATTCCTGTCATTAAGTGGAGGTGCCCTTCTGCTAAAGTAGCCCAGGTGAATAAATAGG

>CT10649_T

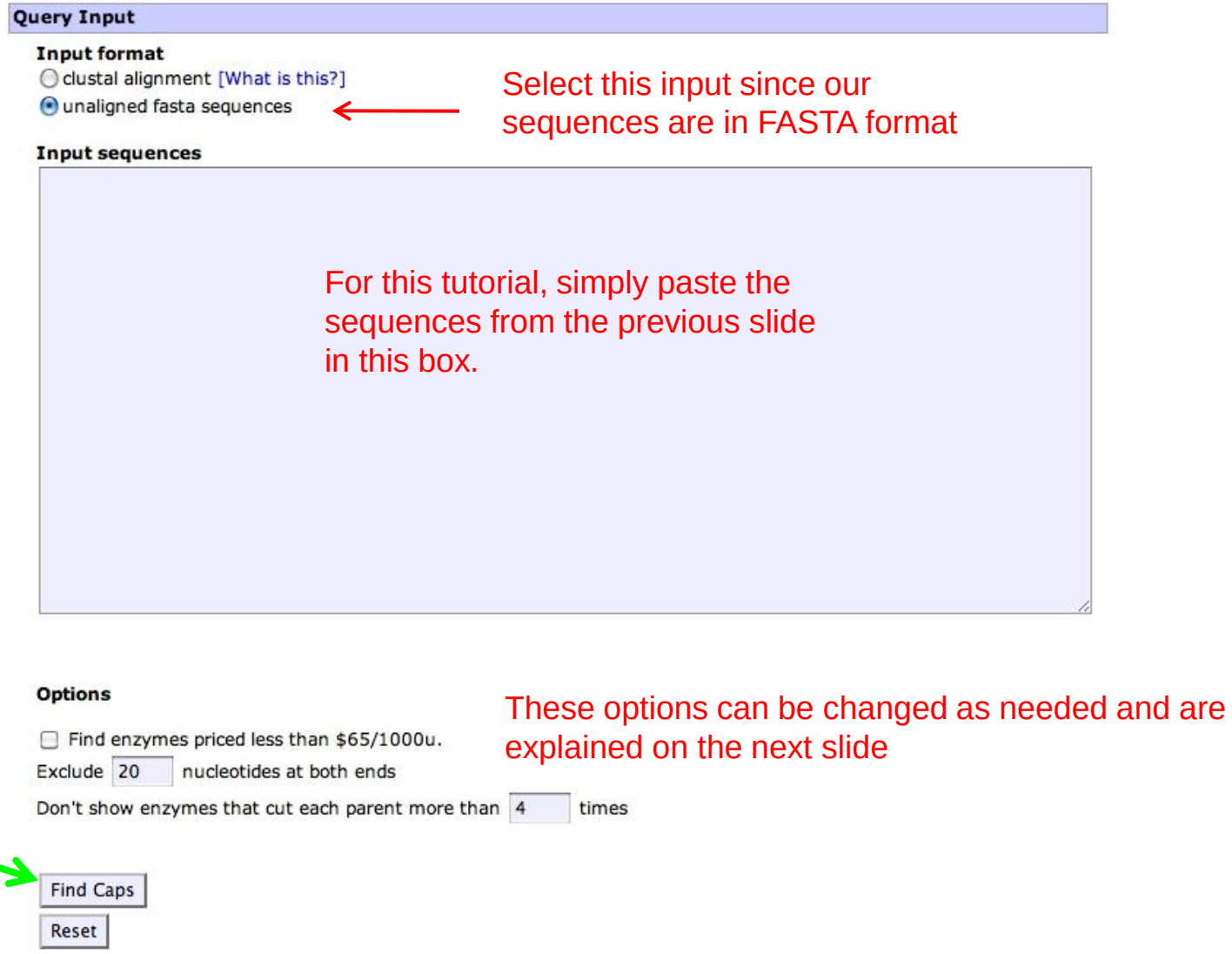
AATAGCAGCATGGTGGGCATCCTCCTGTCTCCACTGCTCGAATCCTTTCCTCCGGCAATGCACCATGTTTTATTAAGAACTCTCCATCTTCTTTCGTGAATATATCATTTGTTACCTACAAAGAAATGCATCTTAACATGGACATTAATTC

ATTCTACATCTTAGAACAAAAACGTGAAGATTTCAAGATCATAGACTTGGAAAGTGATGAAAGAGCGTCGGAATTCCAGAACACCAAATACATGCTTGCACATAACGTATTTCCTGTCAAACTAACAAACGTCAAGTAAACTCCCAA

AACCTCGGTTTGAAGTTTTGGCAGATGCCAATATGTGATGTTTAGGAAGGAGTCCTTAATGATTAAAAAACAAGAACATGACCTATTGAGATAATCTCCAAAGGCAATTGTGTATAAAGAATGTTGTTATTCCACTTGAAATGACTCA
ATCAAACGGAAAATTGCATGTAACACCCATCTTCATCTAGAATTTCAAATGATGAGGCAGAGAAAGATATGAAAACCACCAATAAACTTGAGTTCCATTTTCCAAACCACAAAAGTGAATCCACCAGTTCCAAAAAATTGTGCAGACT
AAACTTATTCATGAAGTAAGTATGTCACAATGGCAAAAGAAAGAAATCATTTGGTGGTACATACTGCTGCAAGACTGTATTTTTCCCTCAAGATTTTACATAATGCCAGCATTAGAGCTGTTTTCCTGCATATAAAAGAACAGTTAGTC
CCTCAAAATCTCAACATGTTCAACCAAAATTTTACAAGCTAATAAACAAGAAAGAAACTGCGATTGGAGGAAAAGCAAAAGCAAACACTAATACTACAAAAACAATAACATACCCTGTGAAGTCACTCCAGTACATTATGAACCTTGA
ACTGAAAAGGAAAATGCTCTAGCACACTCATACACACATTCACACAGTCAGATATGTGTCTAATGGAACAATTGTGTTCCTATATGCAGAGTTCTAAAGATTCAATTTTTTTATCAATAAAAATGGTCCCTTTTGTTTCTTACTTGGGTT
GCTGCAGCTAAAAGAAAATCCTACTTACAACAGATACCAAAAGCTACTAAATATCATCCTCCCCCTTCTACTTTCATTTCTCAAAGATTGAATTTTTCTTCTCAAATACTGAAAACCCTTTCACTTGAACACACATCCCAAGACATAAATT
TAAGAAAAATTGAGGGAAAAGAAGAAATACCCAGTACCAACAGGGCCACCAATTCCAATAGTAAAGGCTCTTTCACTGAAATTCCTGTCATTAAGTGGAGGTGCCCTTCTGCTAAAGTAGCCCAGGTGAATAAATAGG

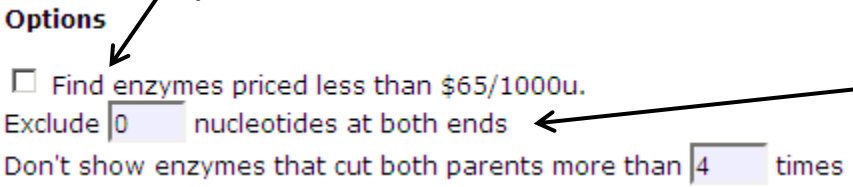
Sequence names (after the “>” symbol) consist of a locus name (CT10649), an underscore (_), and the SNP base (C or T)

Step 2: Input Sequences



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.



• 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/dowload 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).

Query Summary

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

CLUSTAL Alignment

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

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

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

Candidate CAPS

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

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	

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	

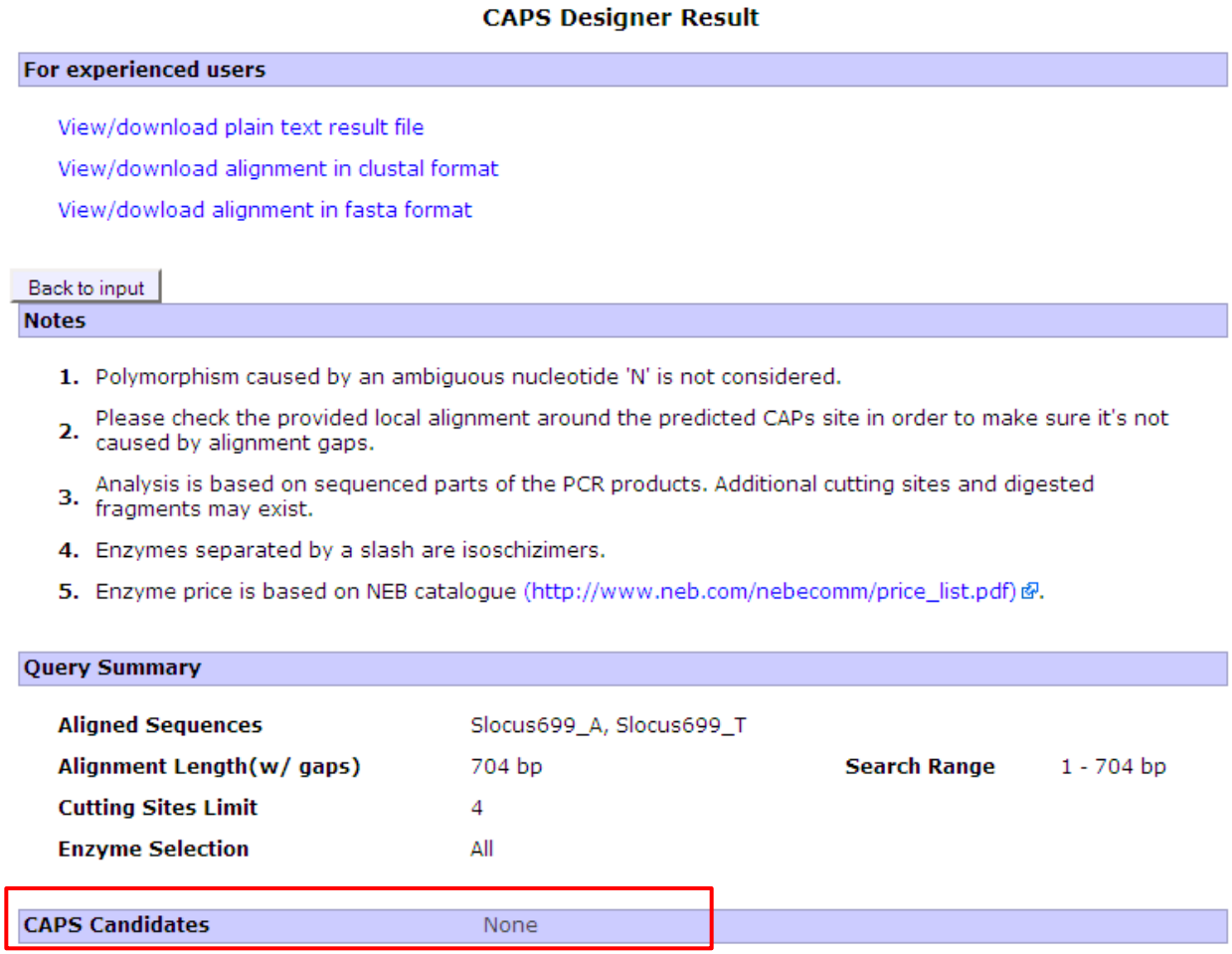
Enzyme	Hpy188I	**Price**	Hpy188I,1000u/$58
Recognition Sequence	TC.GA		
CT10649_C Cutting Site(s)	943	**CT10649_C Fragments(s),bp**	943 387
CT10649_T Cutting Site(s)	223 943	**CT10649_T Fragments(s),bp**	223 720 387
CAPS Site	223	**CT10649_C** agtgatgaaagagcgCCGGAattccagaac **CT10649_T** agtgatgaaagagcgTCGGAattccagaac	

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	

What if there are no Candidate CAPS?

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/)



CAPS Designer Result

For experienced users

View/download plain text result file

View/download alignment in clustal format

View/dowload 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).

Query Summary

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

CAPS Candidates None

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 (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 (verified 7 Dec 2010).
- Provart, 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 (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/ (verified 7 Dec 2010).