Marker Conversion:
From SNP to CAPS and dCAPS markers

Sungwoo Lee

lee.3546@osu.edu

Dept. of Horticulture and Crop Science,
The Ohio State University

Introduction

- Single nucleotide polymorphisms (SNPs) are a source of abundant genetic variation among cultivars or genotypes.
- Due to their abundance and use in automated high-throughput genotyping, SNPs are commonly used molecular markers in genetic mapping and marker-assisted selection.
- This tutorial is about alternative ways to utilize SNPs, which are CAPS and dCAPS, besides the conventional SNP genotyping method.

This tutorial aims

- To introduce basics of CAPS (cleaved amplified polymorphic sequence) and dCAPS (derived CAPS) markers.
- To demonstrate the application of these markers to marker-assisted selection (MAS)

General SNP genotyping methods

- Four SNP genotyping assays (Lee et al., 2004)
  - Single-base extension
  - Allele-specific primer extension
  - Oligonucleotide ligation
  - Direct hybridization
- These SNP genotyping assays require fluorescence-labeled ddNTPs or probes (primers) and specialized detection equipment.

Let's assume that we already have SNP information between two parents of a mapping population or a breeding population. Four assays are generally used in SNP genotyping; single-base extension, allele-specific primer extension, oligonucleotide ligation, and direct hybridization. These PCR-based methods are fast, convenient, and also possible to multiplex. However, they require either fluorescence-labeled ddNTPs or probes (primers) for the SNP and specialized detection equipment. These can be limitations of SNP genotyping.

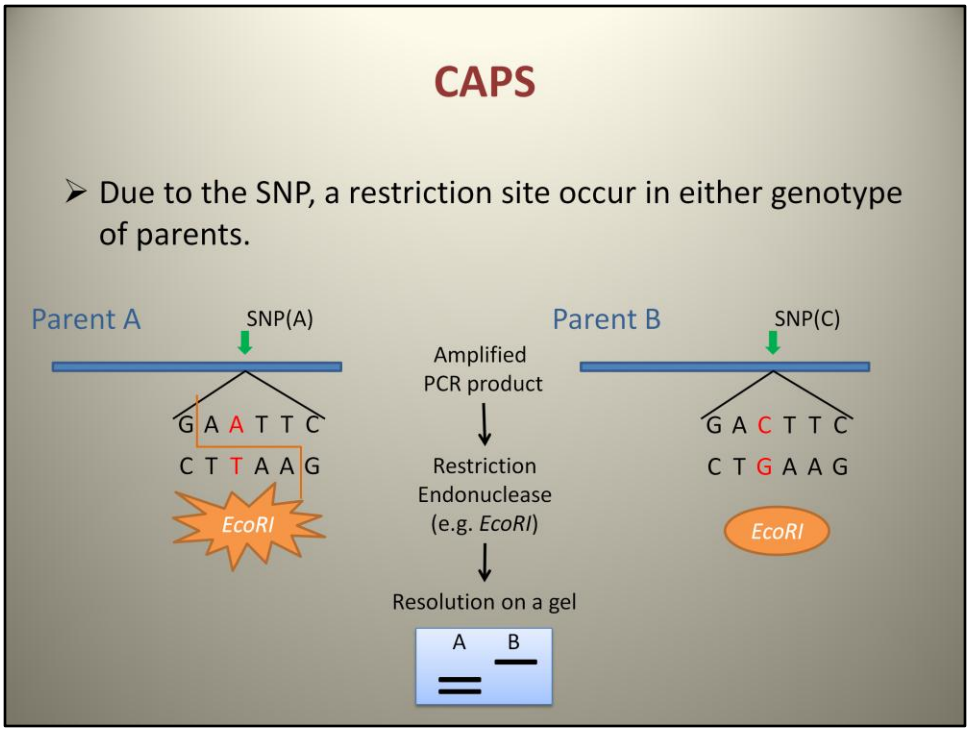
Concept of CAPS and dCAPS

- CAPS and dCAPS convert SNPs (sequence-based polymorphisms) into size-based polymorphisms.
- In the case of dCAPs, the SNP creates or removes a restriction endonuclease recognition site.

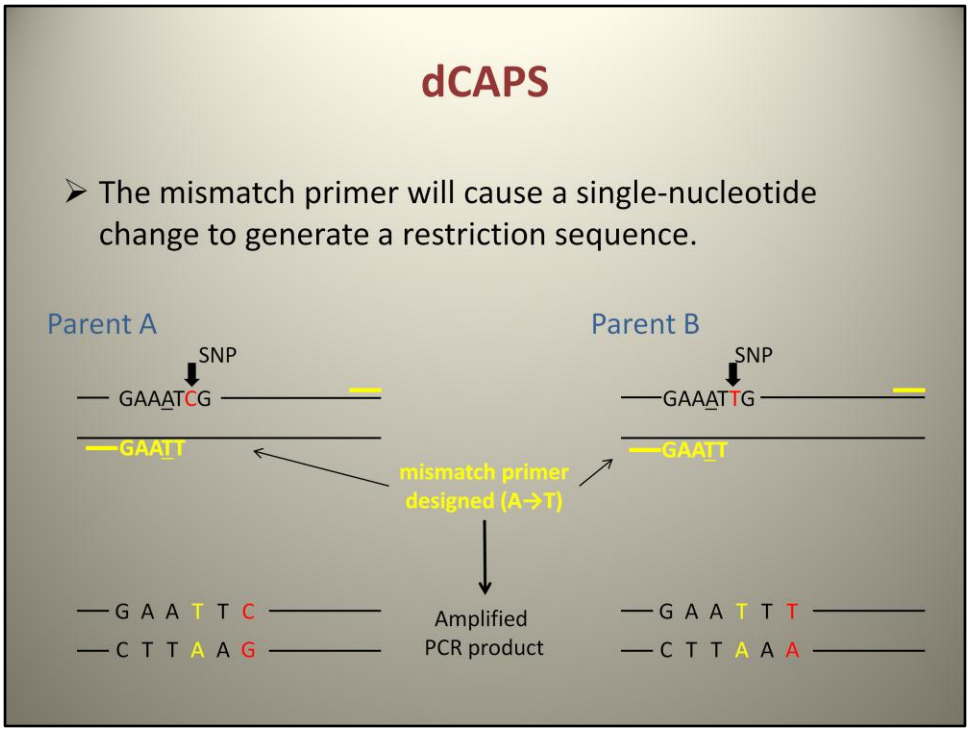
CAPS and dCAPS markers are based on SNP but don't require those fluorescence-labeled probes (or ddNTPs) and specialized detection equipment. The concept of these markers is to convert SNPs into size-based polymorphism by utilizing the feature of either presence or absence of the restriction site between parents due to the SNP. One assumption here is that SNP should be located on the pre-existing or post-generated restriction sequence.

Pros. and *Cons.*

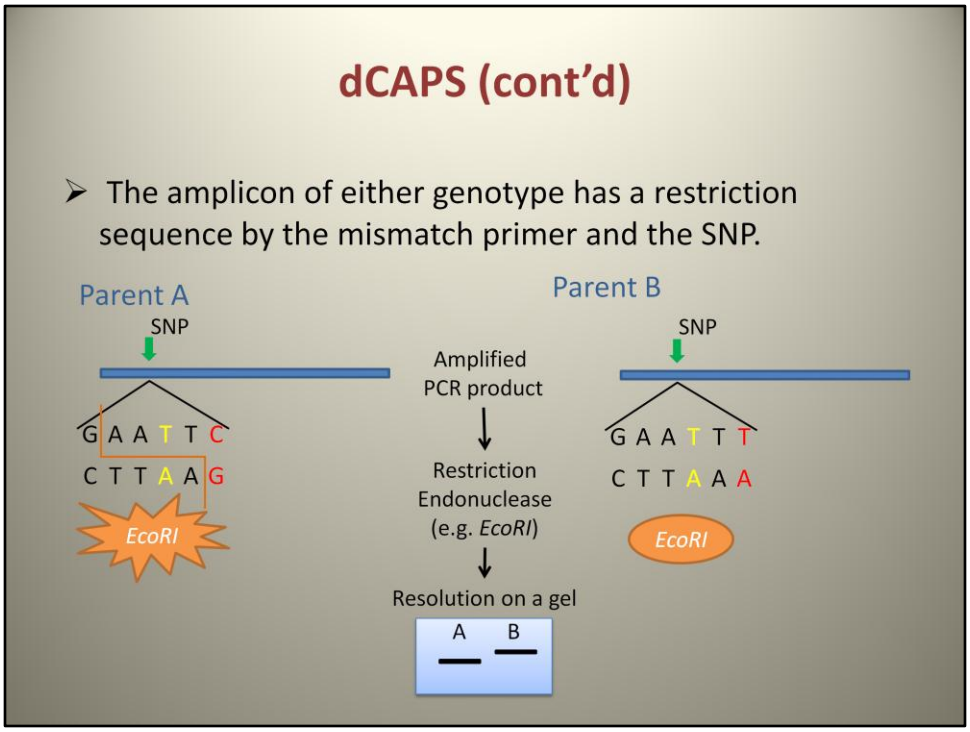
- *Pros.*
  - Fluorescence-labeled ddNTPs or primers as well as specialized detection equipments are not required.
- *Cons.*
  - Restriction enzyme can be expensive.
  - SNP does not always change a restriction site (CAPS only).



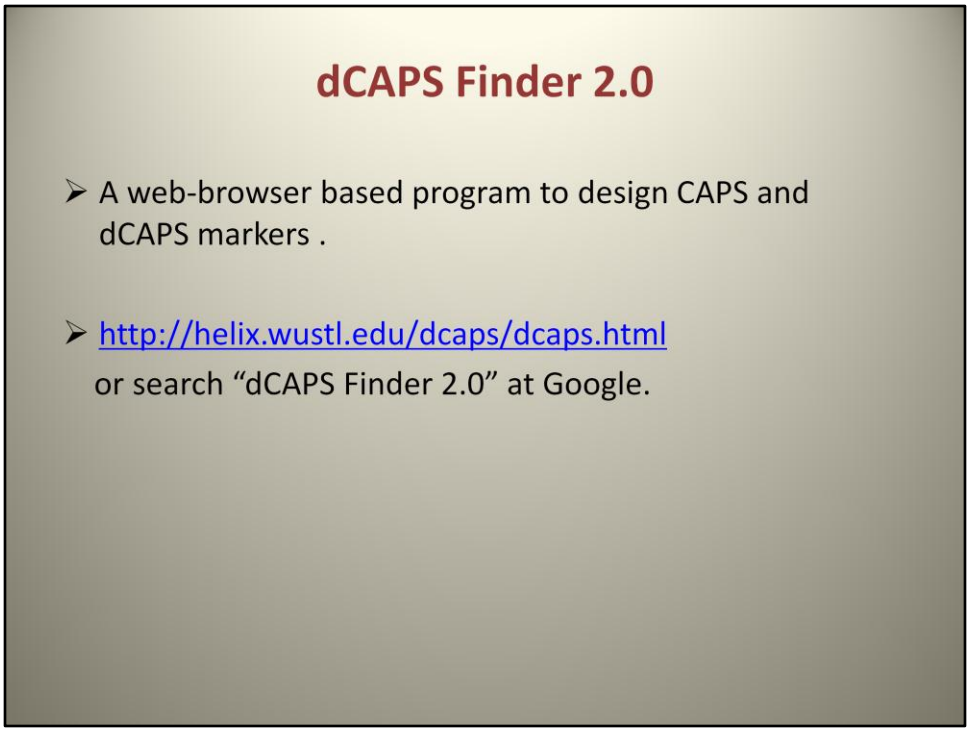
CAPS means cleaved amplified polymorphic sequence. This is used when the restriction sequence is naturally present because of the SNP between both parents. The PCR amplicon of parent A has "*EcoRI*" site while parent B has nothing. When the PCR products are digested with the restriction enzyme, *EcoRI* , the enzyme will cut the amplicon in A but cannot cut in B. When resolving PCR products on a 4% agarose gel, two bands will be shown for A while one band for B, presenting polymorphism between A and B.



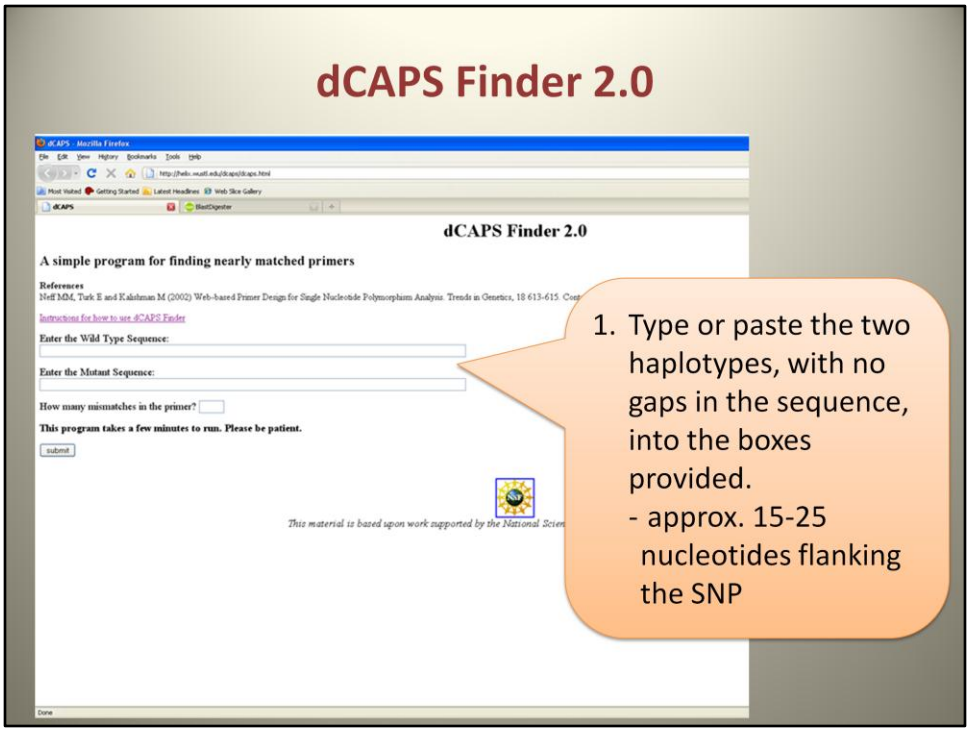
When SNPs do not change a restriction site, dCAPS can overcome the limitation. dCAPS is available when a restriction site can be produced by one or two base changes using a mismatch primer. When designing PCR primers, there is one mismatch embedded in the primer sequence annealing to the restriction sequence. For example, actual sequence is GAAAT but the primer is designed to end with GAATT. Once the target region is amplified with this primer pair, the amplicons get a single nucleotide converted due to the mismatch. *EcoRI* site is produced in parent A.



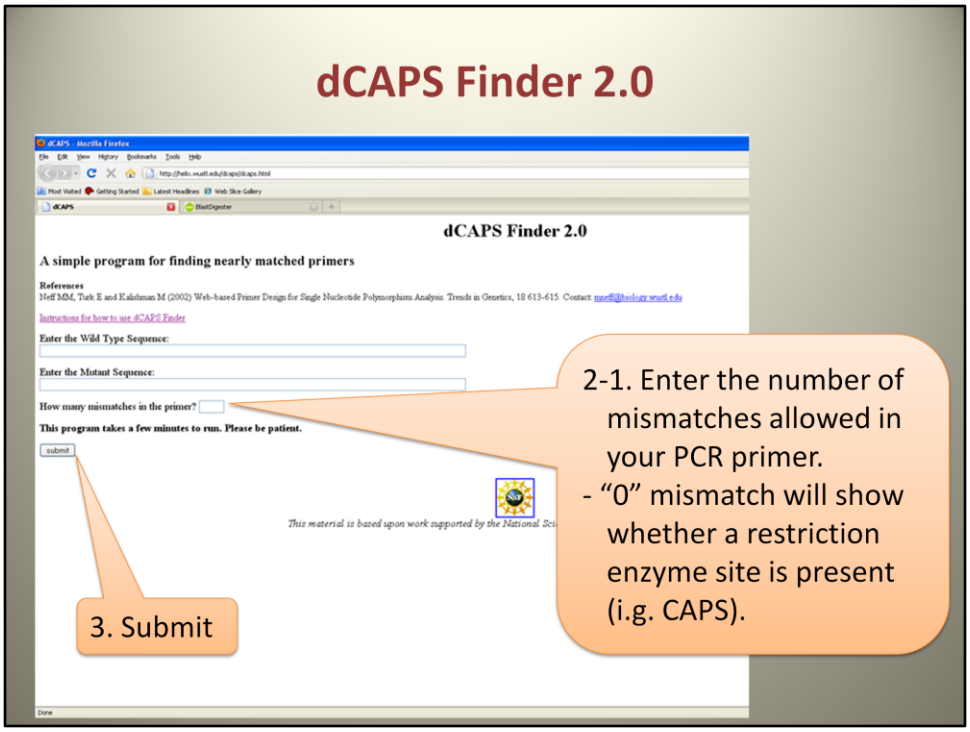
In the same manner as CAPS, polymorphism can be observed after the *EcoRI* digestion on an agarose gel.



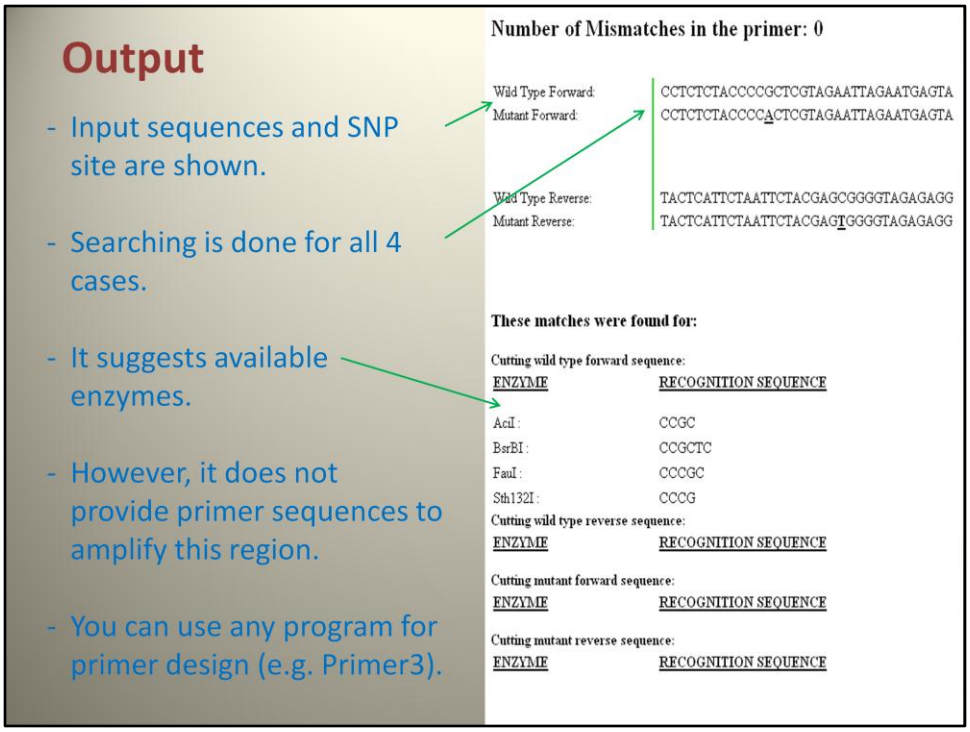
dCAPS Finder 2.0 is an easy and simple program to find CAPS and dCAPS markers. This is a web-browser based program, thus installation and sign-up are not necessary. It is also free. You can be connected at this address or search it at Google.



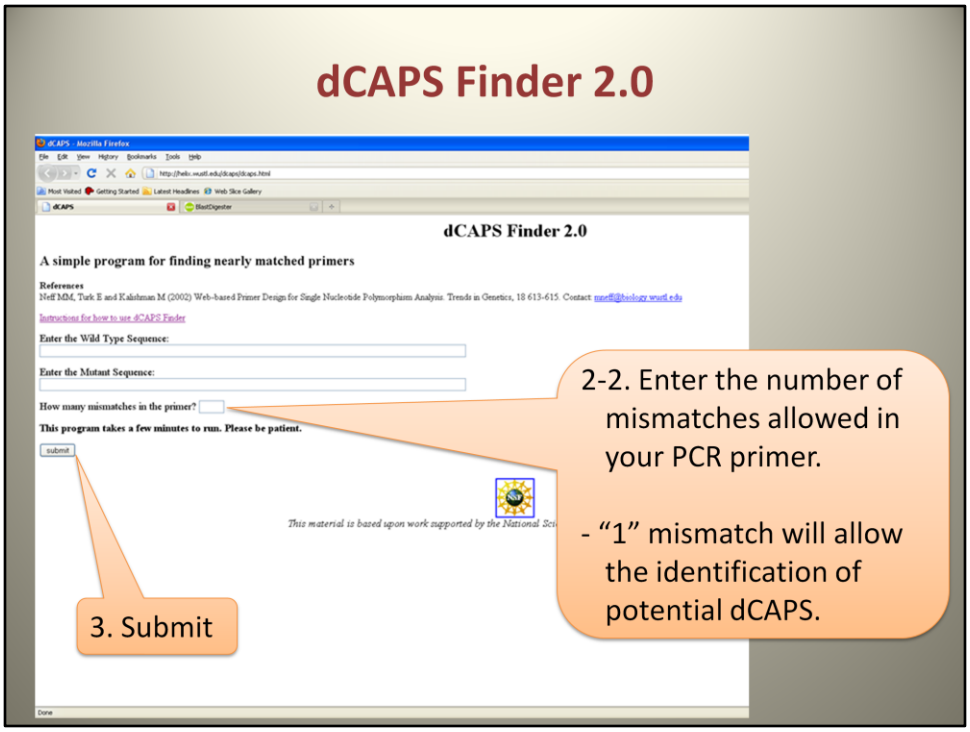
This is the first page. As you see it is very simple. You just need to input three things. First, you should input sequences. You can type or paste two haplotypes with 15-25 nucleotide sequences flanking the SNP, without any gaps.



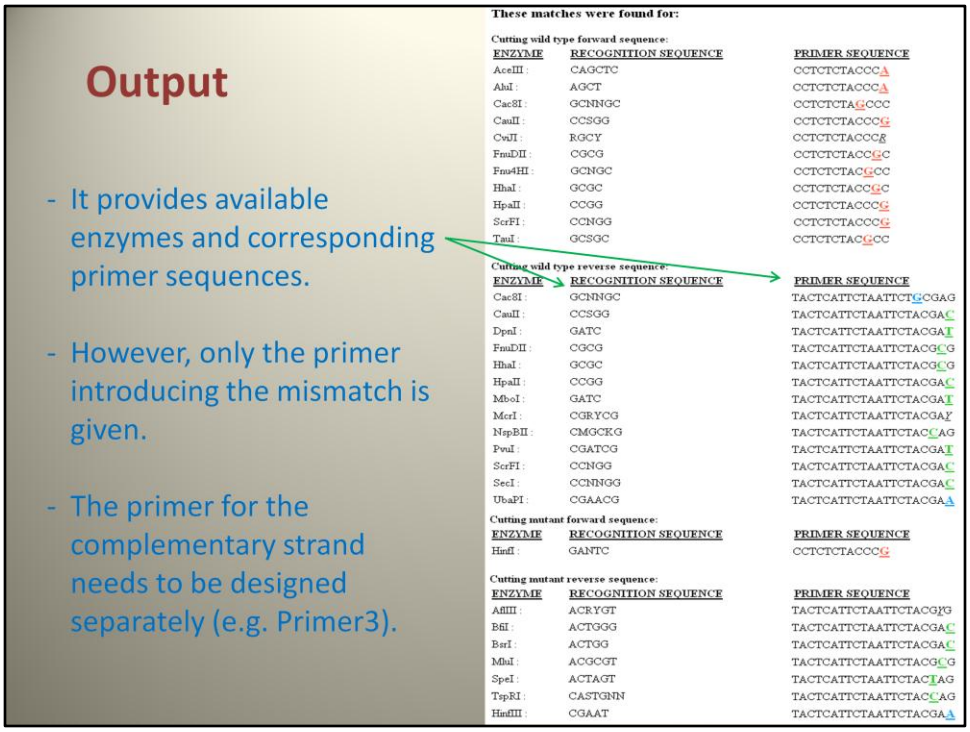
Next, you should enter the number of mismatches you want. If you want CAPS, enter "0". Then, the output will display whether any restriction enzyme site is present in the query sequences. Finally, submit it.



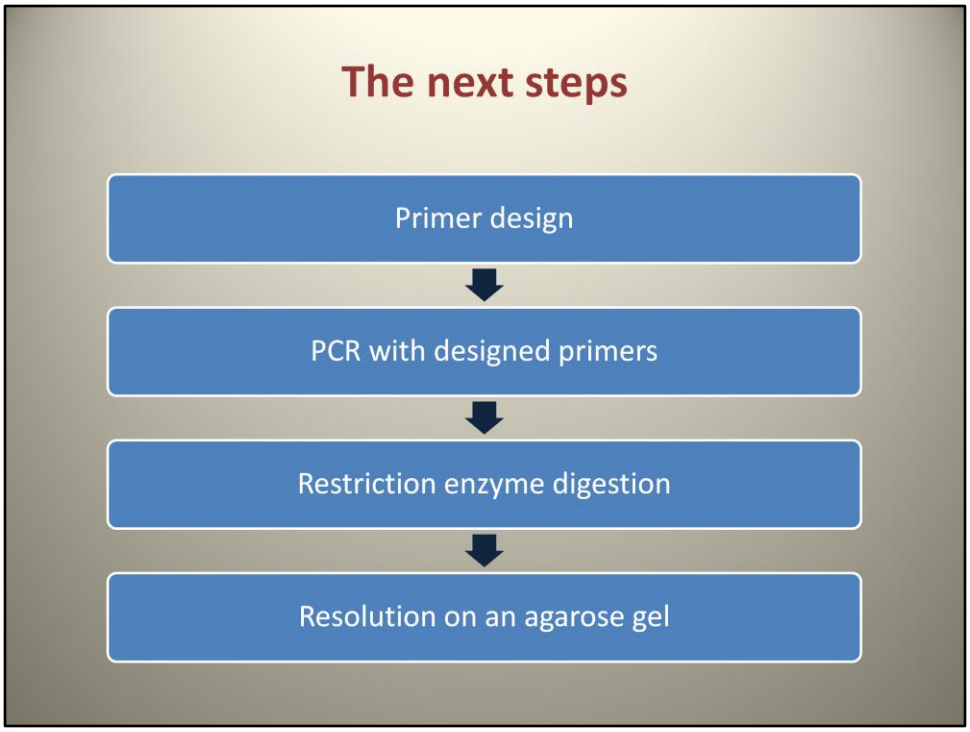
This is the output. Input sequences will be shown. This program searches for any restriction enzyme sites based on four cases of sequences, which are two query sequences and two reverse complementary sequences of the query sequences.
For “0” mismatches, it only gives the enzyme site without primer information.
In this case, you need to design a pair of primers to amplify the pertinent sequence containing a target restriction sequence using a primer-design program, like Primer3.



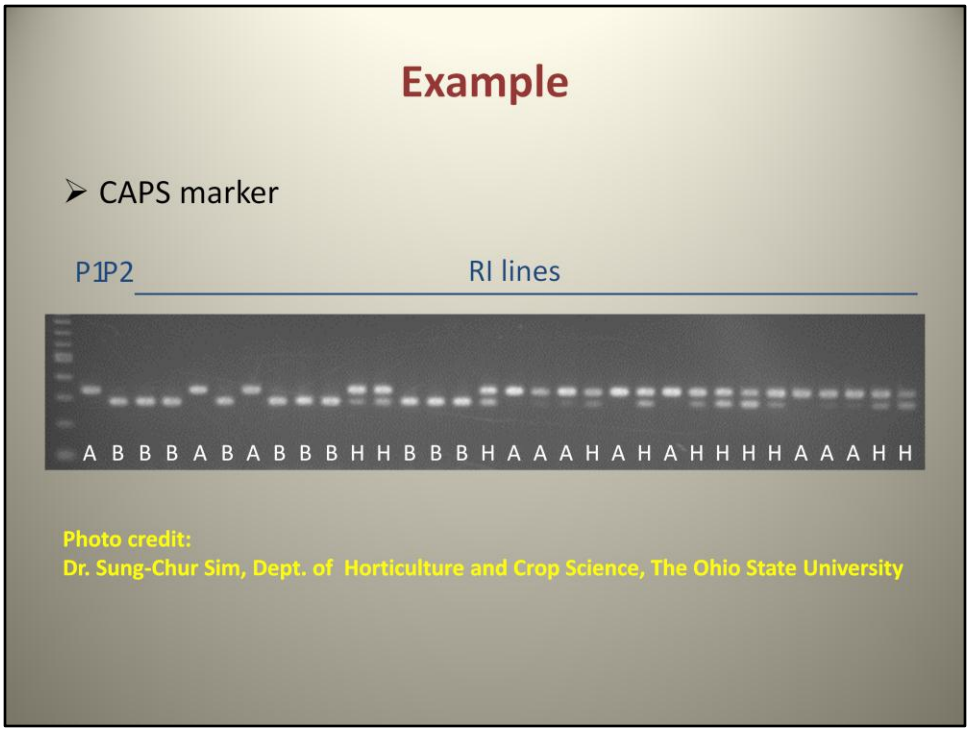
If you couldn't find any enzyme sites within the query, we can try dCAPS.
Put in "1" for the mismatches, it will show potential dCAPS.



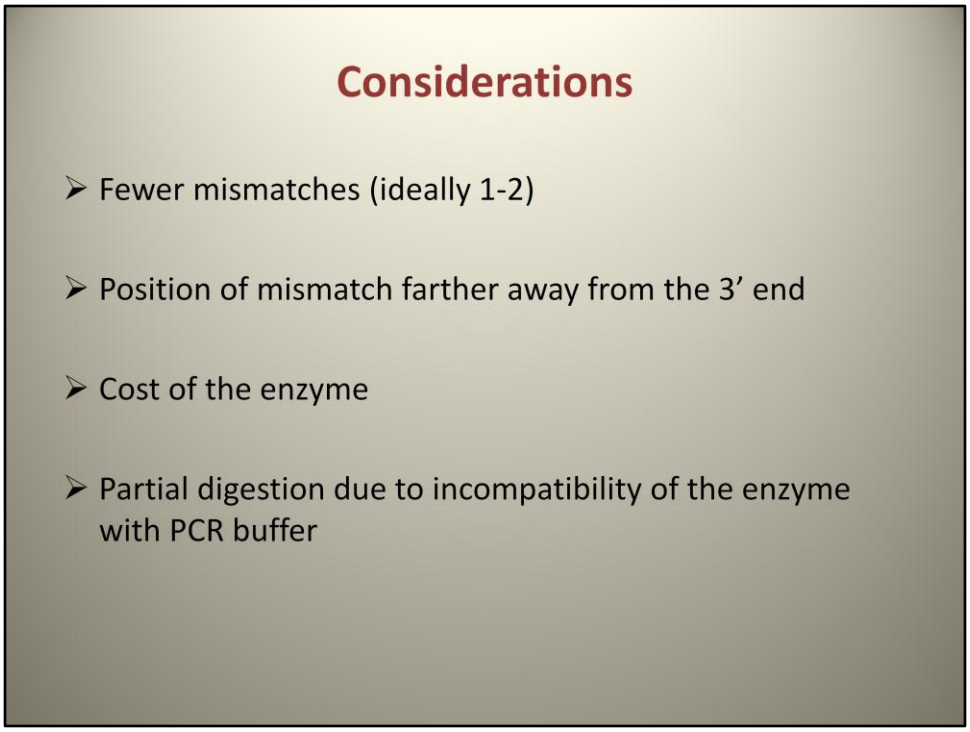
This also searches forward and reverse complementary sequences of two queries for restriction enzyme site, respectively.
The output will display available restriction enzyme sites within the queries and the possible primer sequences with one mismatch. The colored sequences indicate mismatches. The output provides the primer sequence to introduce the restriction site. The primer for the other strand needs to be designed separately.



If you try a CAPS marker, which is 0 mismatches, you are required to design a pair of primers to amplify the region surrounding the target SNP separately. If you try a dCAPS marker, you need to find a reverse primer that will be used to amplify the region with the dCAPS primer output. Next, you need to perform PCR with the designed primers, followed by restriction enzyme digestion. Finally, the enzyme-digested PCR fragments resolved on a 4% agarose gel will reveal polymorphic bands between two parents.



This is an example of two parents and an RIL population genotyped with a CAPS marker. PCR was performed, followed by *BsrBI* digestion. Parent 1 was not cut by the enzyme, but parent 2 was cut. The digested PCR product were resolved on a 4% agarose gel. The genotype of parent 1 was scored A, the one of parent 2 was scored B, and the heterozygous was scored H.



There are some considerations in using CAPS or dCAPS markers. For dCAPS, fewer mismatches in the primer sequence are favorable and having mismatches away from the 3' end is helpful for the primer's affinity to the template. Next, the cost of the enzyme should be a consideration. Some restriction enzymes can be expensive. You should consider whether a cheaper enzyme is available when choosing the CAPS or dCAPS primers. The restriction enzyme activity can be affected by PCR buffer. Partial digestion caused by incomplete reaction of a restriction enzyme is a potential obstacle in scoring of genotypes. The concentration of salt (e.g. $MgCl_2$) can also cause partial digestion. It can lead homozygous genotypes being mis-scored as heterozygous.

Partial digestion

- Partial digestion can confuse scoring of genotypes

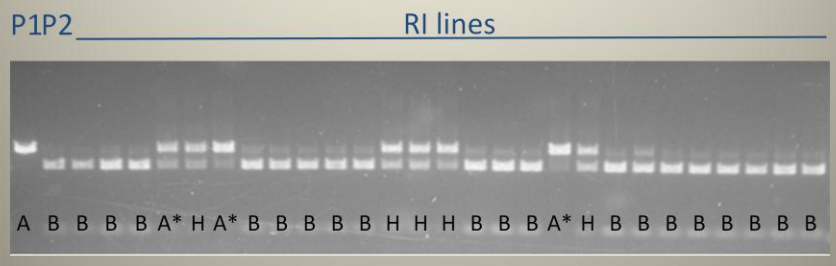


Photo credit:
Dr. Sung-Chur Sim, Dept. of Horticulture and Crop Science, The Ohio State University

This is an example of partial digestion. The genotypes with a 'star' are vague. They look like genotype A but also show another faint band. Most cases of genotype B show a faint A allele, which is the remainder of PCR product. This can confuse us in scoring; whether the genotype is A, and should not be cut, or B, and should be cut further. So we may mis-read them as H instead of either A or B. One way to distinguish partial digestion is to compare the intensity of each band. If two bands are equally stained, the individual is more likely heterozygous. Loading a heterozygous control as well as the parents can be an alternative way to prevent mistakes in scoring. In spite of these alternatives, it can be still hard to discriminate from time to time.

References Cited

- Lee, S. H., D.R. Walker, P. B. Cregan and H. R. Boerma. 2004. Comparison of four flow cytometric SNP detection assays and their use in plant improvement. Theoretical and Applied Genetics 110:167-174. (Available online at: http://dx.doi.org/10.1007/s00122-004-1827-1) (verified 10 Jan 2011).
- Neff, M.M., E. Turk and M. Kalishman. 2002. Web-based primer design for single nucleotide polymorphism analysis. Trends in Genetics 18:613-615. (Available online at: http://dx.doi.org/10.1016/S0168-9525(02)02820-2) (verified 10 Jan 2011).
- Neff, M.M., J.D. Neff, J. Chory and A.E. Pepper. 1998. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in arabidopsis thaliana genetics. Plant Journal 14:387-392. (Available online at: dx.doi.org/10.1046/j.1365-313X.1998.00124.x) (verified 10 Jan 2011).

External Link

- dCAPS finder. [Online] Washington University Biology Department. Available at: http://helix.wustl.edu/dcaps/dcaps.html (verified 10 Jan 2011).

Acknowledgement

- I would like to thank Dr. Sung-Chur Sim for the gel images and some advice, as well as Dr. Heather Merk for grammatical revision.