An Improved Method of Cleaved Amplified Polymorphic Sequence (CAPS) Markers in *Arabidopsis thaliana*  

YUAN Ke-Jun**  
(Shandong Institute of Pomology, Taian 271000, China)  

Abstract: The marker enzyme site for each of most CAPS (Cleaved amplified polymorphic sequence) markers could be found in one *Arabidopsis* ecotype Col (Columbia), only 37.5%~47.5% of analysed enzymes had cut sites in the amplification region of Col genome. These findings indicated that over half of analysed enzymes were not useful in developing most CAPS markers. Based on these findings, an improved CAPS method for development of new markers was established, that is, restriction analysis was conducted at first to find candidate enzymes that have cut sites in Col amplification region, these candidate enzymes were used to screen CAPS marker enzyme, not using other enzymes unless no polymorphism between ecotypes was detected. Application experiment results showed that the improved method using Col genomic sequence could be used to develop CAPS markers for other two ecotypes Ler (Landsberg erecta) and Cvi (Cape Verde Islands), even if Ler and Cvi genomic sequences were not known; the cost for over 50% enzymes could be saved. Application of this genome-based economic CAPS (geCAPS) method in other plants such as *Oryza sativa* and *Camellia sinensis* were discussed.  

Keywords: Plants; *Arabidopsis thaliana*; Genome; CAPS (Cleaved amplified polymorphic sequence); Marker; Restriction enzymes

The analysis of DNA polymorphisms is critical to modern genetic research. One of the most common classes of DNA polymorphisms present both in natural populations and after induced mutation is single nucleotide polymorphisms (SNPs). Many methods have been developed for the detection of SNPs (Myers et al., 1985; Labrune et al., 1991; Konieczny and Ausubel, 1993; Neff et al., 1998). Among them, the most widely used method is cleaved amplified polymorphic sequence (CAPS) analysis (Konieczny and Ausubel, 1993). It has been widely used in the detection of mutations(Konieczny and Ausubel, 1993), generation of new markers (Caranta et al., 1999), marker-assisted breeding of plants (Moury et al., 2000), genetic mapping (Mou et al., 2000; Brugmans et al., 2002) and map-based gene cloning (Jander et al., 2002), and also used in the genetic analyses of dog (http://www.ornl.gov/TechResources/Human_Genome/publicat/94SANTA/mapping/neff.html, 1994) and insects (Hasselman et al., 2001), as it is simple, co-dominant, relatively inexpensive, and utilizes the ubiquitous technologies of PCR, restriction digestion and standard agarose slab gel electrophoresis (Neff et al., 1998). It was reported that the original CAPS article ranked the fourth among the 25 most highly-cited papers on *Arabidopsis* and plant disease resistance (http://www.esitopics.com/arab/papers/a1.html, 2003). Obviously, any progress in such a CAPS technique is important.

* This work was supported by the research foundation of China Education Ministry for returned overseas Chinese (No.2003（10）14) and the research foundation of Human Resource Department of Shandong province for returned overseas Chinese （No.2004（527））. Received: September 16, 2004; Accepted: December 10, 2004.

**E-mail: yuankj@vip.sina.com
In the CAPS technique, gene-specific primers are used to amplify template DNA and polymorphic nucleotides are detected by the loss or gain of a restriction enzyme site (Konieczny and Ausubel, 1993; Neff et al, 1998). With the completion of the Arabidopsis Genome Initiative (The Arabidopsis Genome Initiative,2000), a wealth of sequence data are available for design of primers. Ideally, these sequence data allow the most economic development of new reliable PCR-based markers (Hauser et al, 1998). How to get sufficient and most economic utilization of genome or gene sequence and save experiment cost in developing most widely used CAPS markers is a novel and interesting topics. But, even with the entire sequence in hand, it has not been paid attention to this respect. Only several modified CAPS methods with some improvements in other respects, such as dCAPS and ECAPS established by Beye, have been developed (Michaels et al, 1998; Baumbusch et al, 2001). We consider this to be a significant oversight. Hence, we try to establish an improved CAPS method that aims at the sufficient utilization of Columbia (Col) genome sequence in developing new markers of Arabidopsis thaliana, that is, restriction analysis in the amplification region of Col genome is conducted at first to find candidate enzymes that have cut sites in the Col amplification region and may generate ecotype-specific digestion patterns, and then experiment work is performed to screen CAPS marker enzyme with these candidate enzymes, not using other enzymes unless no polymorphism between ecotypes is detected. This method is named geCAPS (genome-based economic CAPS) method, as you will find that it is established based on genome information and can save a part of enzyme cost.

Here we show its scientific grounds by investigating the enzyme digestion patterns for 132 plus 28 CAPS markers on the AGI map (http://www.arabidopsis.org) and conducting restriction analysis in the Col amplification region for 50 markers of Col and Ler, report its successful application in another pair of Arabidopsis ecotypes Ler and Cvi, discuss the feasibility of its application in other plants and offer the detailed experiment method we once used, which facilitates the use of geCAPS method.

1 Materials and Methods

1.1 Enzyme digestion pattern investigation

The TAIR Marker Search tool was used to find all CAPS markers that were polymorphic between Col and Ler on the AGI map (http://www.arabidopsis.org), each of these markers was checked, their enzyme digestion patterns were investigated to find the number of markers with no enzyme site and the number of markers with enzyme sites in the Col amplification region. Markers that were polymorphic between Cvi and Ler were also investigated.

1.2 Restriction analysis (F11A12 marker as an example)

Forward primer was used to find marker’s starting place and reverse primer was used to find its end place in the BAC of the Col genome, using AGI whole genome (BAC clones, DNA) dataset and BLASTN program of TAIR BLAST 2.0. After forward primer sequence (ggtgg tgaag aaggt gctcg) was input and BLAST was run, BLAST query on Arabidopsis sequences was obtained. BLAST query for reverse primer sequence (attgt aattt gggag ctgga cc) could also be obtained. Then F11A12 marker could be found at BAC F11A12 and its position at 13428-15127. TAIR Quick Search was used to find F11A12 BAC clone in TAIR Database and its GenBank accession AC068900. After the DNA sequence at 13428-15127 region was copied, restriction analysis tool of TAIR BLAST 2.0 was used to conduct analysis and display restriction map. Then, we found 123 candidate enzymes that have cut sites in the region 13428-15127 (including Rsal enzyme for F11A12 marker), 117 enzymes that have no cut site in the region 13428-15127, the name of analysed enzymes and the TAIR restriction map of 123 candidate enzymes. Restriction analysis tool on the TAIR website can also be used in other plants. Some software with such function can also be used to conduct restriction analysis.

1.3 Application experiment

The Col genomic sequence (http://www.arabidopsis.org) was used to design primers for 6 CAPS markers. Most new marker specification can be found in our previous publication (Bentsink et al, 2003), except for Marker F11A12 whose forward primer is ggtgg tgaag aaggt gctcg and reverse primer is attgt aattt gggag ctgga cc, marker enzyme is Rsal.

One flower bundle or leaf for each plant of two Arabidopsis ecotypes Ler and Cvi was sampled. Experiment work were conducted using a once briefly described method (Bentsink et al, 2000). DNA extraction temperature condition was similar to that of a published method (Kaundun et al, 2003), a mixture of 125μL extraction buffer (0.35M Sorbitol, 100mM Tris, 5mM EDTA, pH adjust 7.5 with HCl), 175μL lysis buffer (200mM Tris, 50mM EDTA, 2% NaCl, 2%(W/V)CTAB) and 30μL 10%(W/V) Sarkosyl(10g/100mL N-lauroyl sarcosine) was preheated at 65°C for 5min, and then used to extract DNA, from each sample, at 65°C water-bath for 30min, the sample tube was gently shaken 3 times. 400μL (or 700μL) chloroform/isoamylcohol (24/1,v/v) was added and sufficiently mixed. About 270μL (or 500μL) of supernatant in sample tube was transferred to a new tube after centrifugation.
at 15 000 g for 5 min. 270 µL (or 500 µL) of cold isopropanal (-20°C) was added in new tube and mixed by gentle shaking, nucleic acids were precipitated with an equal volume of cold isopropanal. The supernatant was decanted after centrifugation at 15 000 g for 5 min. The pellet obtained was rinsed with 200 µL (or 500 µL) 70% ethanol (-20°C), dried in vacuum excicator for 10 min, added 25 µL MQ water, and then kept at 4°C for use. PCR reactions (50 µL) contained approximately 50 ng genomic DNA and 49 µL reaction mixture, one reaction mixture composed of 2 µL dNTPs (5 mM), 5 µL taq buffer, 2 µL primer (0.1 µg/µL), 0.2 units of taq polymerase and 40 µL water. A total of 35 PCR cycles were performed. Each cycle included a 30 sec melt at 94°C, 1 min annealing step at appropriate temperature (50-55°C) for each specific marker followed by 2 min of primer extension at 72°C. A mixture of 2 µL PCR products, 8 µL water and 2 µL loading dye was used to check PCR products by electrophoresis at 80 V on 1% agarose gel. If PCR products were good, 5 µL of the PCR products were digested for 2-3 h or overnight at 37°C in 20 µL total volume that contained 2 µL restriction buffer (10x), 10 units of the appropriate restriction endonuclease and 12 µL water. After digestion, a mixture of 20 µL digestion products plus 5 µL loading dye was analysed by electrophoresis at 80 V on 1.5% agarose gel, stained with ethidium bromide.

In the geCAPS method, a set of candidate enzymes among 20-30 available enzymes or 10-15 candidate enzymes was used to screen specific enzyme for CAPS marker. If no specific enzyme among candidate enzymes was found, a set of different candidate enzymes was used until no candidate enzymes was available. Then, other enzymes were used to screen CAPS marker. Take F11A12 marker as an example, when general CAPS protocol was used to develop F11A12 marker in the year of 2001, we did experiment using 20 enzymes, i.e. BsaAl, BsaBl, Hgl, Dral, Clal, KpnI, SstI, CtoI, Xhol, Xbal, Accl, Bfal, Scrfl, Mbol, Mspl, HinfI, Ndel, HindIII, EcoRV, RsaI. If geCAPS method were used, prior restriction analysis in the region 13428-15127 of F11A12 BAC were conducted to find candidate enzymes, only 9 candidate enzymes (Bfal, Scrfl, Mbol, Mspl, HinfI, Ndel, HindIII, EcoRV, RsaI) would be needed to conduct experiment.

In our experiment, in order to find the percentage of candidate enzymes among available restriction enzymes and test polymorphisms between Arabidopsis ecotypes Ler and Cvi, DNA fragments were amplified from the two ecotypes using the designed primers and digested with a set of 12-30 restriction enzymes.

2 Results

2.1 Enzyme digestion pattern investigation

When TAIR Marker Search tool was used to investigate the CAPS markers of Ler and Col on the AGI map (http://www.arabidopsis.org), 168 CAPS markers, which were polymorphic between Ler and Col, were found. 132 CAPS markers among them provided enough information to analyse their enzyme digestion patterns. Among 132 CAPS markers, 23 (17.4%) markers were found only with marker enzyme sites in the Ler amplification region but no in the Col amplification region, 109 (82.6%) markers were found with marker enzyme sites in the Col amplification region (Table 1). Therefore, mutations between Ler and Col happened at the enzyme sites of Col for most markers, there were only a few markers whose mutations were caused by creating an enzyme site in Ler, marker enzyme site for each of most CAPS markers could be found in the Col amplification region. This is one reason why restriction analysis is conducted at first in the Col amplification region to find candidate enzymes for developing a CAPS marker of Ler and Col.

To understand the feasibility of geCAPS method being used in other ecotypes Ler and Cvi, we investigated the CAPS markers of Ler and Cvi on the AGI map (http://www.arabidopsis.org). 28 CAPS markers that were polymorphic between Ler and Cvi were found, it was also found that the enzyme digestion pattern of Ler or Cvi was the same as that of Col for 27 CAPS markers.

<table>
<thead>
<tr>
<th>Marker enzyme</th>
<th>Marker numbers</th>
<th>Main enzyme (numbers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme only with cut sites in Ler, no in Col</td>
<td>23</td>
<td>HindIII(5), HincII(3), ACCI(2), Xbal(2), EcoRl(2)</td>
</tr>
<tr>
<td>Enzyme with cut sites in Col</td>
<td>109</td>
<td>DdeI(16), Rsal(11), HindIII(7), HaellII(5), HincII(5), ACCI(5), Alul(5), Mbol(5), Xbal(5), Mspl(5)</td>
</tr>
</tbody>
</table>

Note: Ler and Col stand for Landsberg erecta and Columbia, respectively. All 132 markers obtained from TAIR (http://www.arabidopsis.org).
Although the enzyme digestion patterns of Ler and Cvi were not the same as that of Col for marker NIT4, its marker enzyme MboII was found among the candidate enzymes by restriction analysis in the amplification region of Col genome. Therefore, marker enzyme site for each of these CAPS markers can also be found in Col, the Col genomic sequence can also be used to develop CAPS markers that were polymorphic between other ecotypes such as Ler and Cvi.

DdeI, Rsal and HindIII enzyme sites were found easy to cause mutations, 34 (25.8%) markers of Ler and Col and 8 (28.6%) markers of Ler and Cvi were found mutations at these 3 enzyme sites. In addition, 35 (26.5%) markers of Ler and Col caused mutations at the sites of other 7 enzymes HaeIII, HincII, Accl, Alul, Mbol, Xbal and MspI, 10 (7.6%) markers of Ler and Col were found mutations at least two different enzymes. Furthermore, Accl, HindIII, HincII, Xbal and EcoRI were enzymes easy to create cut sites in Ler when no marker enzyme site presented in Col, 14 (60.9%) markers created cut sites in Ler for these 5 enzymes among 23 markers with no marker enzyme site in Col. These results are also very useful for development of new CAPS markers, especially when no polymorphism between ecotypes is detected using a set of candidate enzymes.

### 2.2 Restriction analysis for 50 markers of Ler and Col

Restriction analysis was mainly used to find the percentage of candidate enzymes among available restriction enzymes, and then to see how many restriction enzymes could be saved in developing a CAPS marker. When 243 restriction enzymes were used to conduct restriction analysis in the Col amplification region for 50 markers of Ler and Col on five chromosomes, the results revealed that there were 4497 candidate enzymes with cut sites and 7496 enzymes with no cut site in the Col amplification region. The candidate enzymes accounted for 37.5% of total enzymes, 4497/(7496+4497)=0.375, the enzymes with no cut site accounted for 62.5% (Table 2). The results also showed that at least one marker enzyme was found among the candidate enzymes for each of 43(86%) CAPS markers (Table 2). Since the enzymes with no cut site accounted for 62.5%, and 86% or 82.6% markers were found with marker enzyme sites in the Col amplification region (Table 2, Table 1), it actually meant that 62.5% of analysed enzymes were not useful in developing 86% (or 82.6%) CAPS markers. This is another reason why restriction analysis is conducted at first to find candidate enzymes. It also gives the reason why less restriction enzymes are needed for screening a CAPS marker and the cost for enzymes can be reduced in the geCAPS method. Theoretically, the cost for over 50% enzymes can be saved in developing a geCAPS marker, 86%x62.5%=53.8%, 82.6%x62.5%=51.6%.

As is well known, the more enzymes have restriction sites in the amplification region, the more possibility a specific marker enzyme can be found. Our results showed that 4497 candidate enzymes had restriction sites in the 61406 nucleotides (the Col amplification region of 50 markers), on average, there was one candidate enzyme for every 14bp. Although this is an oversimplification, the restriction sites of candidate enzymes are not randomly distributed, it really implied that there were many potential enzyme restriction sites in the amplification region. This is the third reason for establishing geCAPS method and why we can use the candidate enzymes to screen CAPS marker enzyme.

### 2.3 Restriction analysis for 6 new markers of Ler and Cvi

As shown in Table 3, at least one marker enzyme could be found among the candidate enzymes for each of 6 markers of Ler and Cvi. This further confirmed that restriction analysis in the amplification region of Col genome could be used to find candidate enzymes for other ecotypes such as Ler and Cvi.

When 243 restriction enzymes were used to conduct restriction analysis in the relevant Col amplification region for these 6 markers, the results revealed that there were 688 candidate enzymes with cut sites and 761 enzymes with no cut site.

<table>
<thead>
<tr>
<th>Chrom.</th>
<th>Markers(Sizes in bp)</th>
<th>Enzymes with no cut site</th>
<th>Enzymes with cut sites</th>
<th>Marker enzymes (from candidate enzymes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>11 (12495)</td>
<td>1735</td>
<td>842</td>
<td>13 (10)</td>
</tr>
<tr>
<td>II</td>
<td>9 (9785)</td>
<td>1416</td>
<td>758</td>
<td>10 (8)</td>
</tr>
<tr>
<td>III</td>
<td>10 (10899)</td>
<td>1513</td>
<td>902</td>
<td>10 (8)</td>
</tr>
<tr>
<td>IV</td>
<td>10 (14649)</td>
<td>1424</td>
<td>988</td>
<td>10 (9)</td>
</tr>
<tr>
<td>V</td>
<td>10 (13578)</td>
<td>1408</td>
<td>1007</td>
<td>11 (10)</td>
</tr>
<tr>
<td>Total</td>
<td>50(61406)</td>
<td>7496</td>
<td>4497</td>
<td>54 (43+2)</td>
</tr>
</tbody>
</table>

Note: All 50 markers and analysis tool obtained from TAIR (http://www.arabidopsis.org), restriction analysis was conducted in the amplification region of Col genome. Candidate enzymes were enzymes with cut sites in the Col amplification region.
site, the candidate enzymes accounted for 47.5% of total enzymes, 688/(761+688)=0.475, the enzymes with no cut site accounted for 52.5% (Table 3). As at least one marker enzyme could be found among the candidate enzymes for each of 6 markers, theoretically, the cost for 52.5% enzymes can be saved in developing these 6 CAPS markers.

It was also found that 688 candidate enzymes had restriction sites in the 9757 nucleotides of 6 markers (Table 3), on average, there was one candidate enzyme for every 14bp. It actually implied that there were many potential enzyme restriction sites in the amplification region. This further explains why we can use the candidate enzymes to screen CAPS marker enzyme.

2.4 Application analysis and experiment

When a set of 12-30 available enzymes were used to conduct restriction analysis in the relevant Col amplification region for 6 markers of Ler and Cvi, the results revealed that the candidate enzymes accounted for 46.5% of total enzymes, 66/142=0.465, the enzymes with no cut site accounted for 53.5% (Table 4). As at least one marker enzyme could be found among the candidate enzymes for each of 6 markers, theoretically, the cost for 53.5% enzymes can be saved in developing these CAPS markers.

The experiment results showed that the enzymes with cut sites in the amplified products accounted for 43%, 61/142=0.43, the enzymes with no cut site in the amplified products accounted for 57%, among each set of 12-30 available enzymes. Although the result was a little different from that of restriction analysis, the experiment results really confirmed that the selection and utilization of candidate enzymes were feasible and the cost for over 50% enzymes could be saved in developing CAPS markers.

To have a better understanding of the geCAPS method, the electrophoresis photograph of Ler and Cvi PCR products digested with different restriction enzymes was given in Figure 1. We once mentioned in the methods section, when geCAPS method was used to develop F11A12 marker, among 20 restriction enzymes, only 9 candidate enzymes (BfAl, ScrFl, MboI, Mspl, HinFl, Ndel, HindIII, EcoRV, RsaI) were needed to conduct experiment. Figure 1 showed that 8 candidate enzymes (BfAl, MboI, Mspl, HinFl, Ndel, HindIII, EcoRV, RsaI) really had cut sites in the PCR products, enzyme Rsa I among 8 candidate enzymes could be used as marker enzyme of F11A12 marker. This meant that the cost for 55% restriction enzymes could be saved in developing F11A12 marker, (20-9)/20=0.55.

### Table 3 Restriction analysis for 6 markers of Ler and Cvi

<table>
<thead>
<tr>
<th>Marker</th>
<th>Markers sizes (bp)</th>
<th>Enzymes with no cut site</th>
<th>Enzymes with cut sites</th>
<th>Marker enzymes (from candidate enzymes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4P13b</td>
<td>1385</td>
<td>118</td>
<td>125</td>
<td>1 (1)</td>
</tr>
<tr>
<td>cT22N4a</td>
<td>1759</td>
<td>127</td>
<td>114</td>
<td>2 (1)</td>
</tr>
<tr>
<td>cT13O15b</td>
<td>1803</td>
<td>140</td>
<td>102</td>
<td>2 (2)</td>
</tr>
<tr>
<td>cT13O15c</td>
<td>1417</td>
<td>124</td>
<td>117</td>
<td>1 (1)</td>
</tr>
<tr>
<td>cF4P13a</td>
<td>1694</td>
<td>135</td>
<td>107</td>
<td>1 (1)</td>
</tr>
<tr>
<td>F11A12</td>
<td>1699</td>
<td>117</td>
<td>123</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Total</td>
<td>9757</td>
<td>761</td>
<td>688</td>
<td>8 (6+1)</td>
</tr>
</tbody>
</table>

Note: Marker sizes were calculated based on the relevant Columbia amplification region for these markers, analysis tool obtained from TAIR (http://www.arabidopsis.org) and restriction analysis was conducted in the relevant Columbia amplification region for these markers. Candidate enzymes were enzymes with cut sites in the Columbia amplification region.

### Table 4 Application analysis and experiment results for 6 markers of Ler and Cvi

<table>
<thead>
<tr>
<th>Marker</th>
<th>Restriction analysis in Col genome Enzymes with no cut site</th>
<th>Enzymes with cut sites</th>
<th>Check in Ler and Cvi PCR products Enzymes with no cut site</th>
<th>Enzymes with cut sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4P13b</td>
<td>7</td>
<td>13</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>cT22N4a</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>cT13O15b</td>
<td>19</td>
<td>11</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>cT13O15c</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>cF4P13a</td>
<td>17</td>
<td>13</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>F11A12</td>
<td>11</td>
<td>9</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>66</td>
<td>81</td>
<td>61</td>
</tr>
</tbody>
</table>

Note: Analysis tool obtained from TAIR (http://www.arabidopsis.org) and restriction analysis was conducted in the relevant Col amplification region for these markers of Ler and Cvi. Enzymes with cut sites were candidate enzymes.
3 Discussion

Although a number of new technologies for high-throughput detection of DNA polymorphisms, such as DHPLC, Taq-Man PCR assay and pyrosequencing method, have been developed in the past few years (Jander et al, 2002), due to their high initial equipment cost, CAPS technique is still the first choice for most plant research units. In the original CAPS protocol (Konieczny and Ausubel, 1993), the primary PCR products were subjected to digestion with a panel of restriction enzymes. Enzymes were chosen based on the DNA sequence of amplified product. Typically, each PCR product was digested with 25 restriction enzymes. If none of enzymes detected a polymorphism, a second series of digestions was performed with 25 different enzymes until a polymorphism was found or no more enzymes were available. Obviously, it did not mention how to choose enzymes before enzyme digestion experiment. Our results showed that marker enzyme site for each of most CAPS markers could be found in one Arabidopsis ecotype Col, only 37.5%-47.5% of analysed enzymes had cut sites in the amplification region of Col genome, on average, there was one candidate enzyme for every 14 nucleotides. Based on these findings, here we establish geCAPS method for development of new markers in Arabidopsis thaliana and confirm its scientific grounds. This method not only can be used to choose candidate enzymes before experiment, but also can utilize much more genome or gene sequence information and save the cost for over half of available enzymes, it is really a genome-based economic CAPS (geCAPS) method. In addition, as it uses inexpensive technologies and reagents common to most molecular biology laboratories and the related software are free available on the TAIR website (http://www.arabidopsis.org), it can be used immediately and is a very practical method, especially useful for those laboratories that cannot afford to utilize expensive technologies and generate new SNP and InDel markers, which results in a high per-assay cost for a lab that does not need to perform large (Jander et al, 2002).

The geCAPS method actually identifies SNPs by restriction enzyme digestion of specific PCR products. On average, there is one predicted SNP every 3.3 kb between Col and Ler. However, this number may be an underestimate of the true frequency (Jander et al, 2002; The Arabidopsis Genome Initiitive, 2000). The estimated frequency between each pair of ecotypes Ler, Col and Ws (Wassilewskija) was one SNP in every 207-1027 nucleotides (Konieczny and Ausubel, 1993; Hauser et al, 1998; Cho et al, 1999). As 28% of the mutations (SNPs) can be usable directly as CAPS markers(Michaels et al, 1998), the designed size of a geCAPS marker should be among 739-3668bp, 207÷28%=739, 1027÷28%=3668. In fact, the actual sizes of most CAPS markers were in this range. In our experiment, the sizes of 6 CAPS markers for Ler and Cvi were 1385-1803bp. These were different from that of a dCAPS marker, which was among 80-300bp as its marker enzyme site was arbitrarily placed in one of its primers (Neff et al, 1998; Michaels et al, 1998). Certainly, when the sequences of both ecotypes such as the genome sequences of Col and Ler are known, small CAPS markers of less than 739bp can also be designed and developed, which is much easier. In such a case, if 243 restriction enzymes or many available enzymes are used to conduct restriction analysis in the amplification regions of Col and Ler, we may find a unique enzyme that have cut site(s) in one ecotype and no cut site in another ecotype, this unique enzyme can be directly used as marker enzyme.

Of course, in most case, the genome or gene sequence of one ecotype is enough for developing a CAPS marker in the geCAPS method. For example, when developed 6 new CAPS markers of Ler and Cvi in the year of 2001, we just used the Col genome sequence of Arabidopsis thaliana. This is very important because we do not know the genome sequences of two ecotypes (or cultivars) for most of other plants (including economic plants). Here is an example for economic plants. When we tested the feasibility of geCAPS method application in Camellia sinensis, only three gene sequences of one cultivar
Yabukita were used. This means that as well as the genome or gene sequence of one ecotype or crop cultivar is known, the geCAPS method can be used to develop CAPS marker, it is very useful for developing CAPS markers of other plants. After inspecting many CAPS markers of Oryza sativa and Camellia sinensis (Barlaan et al., 2001; Kaundun et al., 2003), we found that it was really feasible to use geCAPS method in other plants. The genome sequence data are now available for Oryza sativa (Yu et al., 2002). With the completion of more plant and animal genome projects or more gene sequence being available, this method must be very useful in a variety of genetic analysis applications of plants as well as animals.

Acknowledgements

Experiment work for 6 new markers were conducted when Yuan K J officially collaborated with Vreugdenhil D in Laboratory of Plant Physiology, Wageningen University, the Netherlands, under support of China Scholarship Council (CSC No.99937032) during 2000 to 2001, Bentsink L in Laboratory of Genetics, Wageningen University designed the primers for 6 markers, and with Koornneef M in the same laboratory provided technical assistance in experiment, supplied experiment material and facility. The author thanks for their suggestion on the manuscript and appreciates help from three of them as well as Jamar D, Ruys G, Aarts M, Hanhart C, Clerkx E and Salah E during his stay in Wageningen University.

References


Bentsink L, Yuan K, Koornneef M, Vreugdenhil D. The genetics of phytate and phosphate accumulation in seeds and leaves of Arabidopsis thaliana, using natural variation. Theoretical and Applied Genetics, 2003, 106: 1234-1243


Kaundun S S, Matsumoto S. Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, Camellia sinensis (L.) O. Kuntze, and differentiation between assamica and sinensis varieties. Theoretical and Applied Genetics, 2003, 106: 375-383


