Plant genomic instability detected by microsatellite-primers

Xavier, J. Leroy*
Laboratoire de Biotechnologie et d’Amélioration des Plantes
ISAMOR, Technopôle Brest-Iroise
Université de Bretagne Occidentale, 29280 Plouzané, France
E-mail : xavier.leroy@univ-brest.fr

Karine Leon
Aventis Crop Science
France
E-mail : Karine.Leon@aventis.com

Michel Branchard
Laboratoire de Biotechnologie et d’Amélioration des Plantes
ISAMOR, Technopôle Brest-Iroise
Université de Bretagne Occidentale, 29280 Plouzané, France
E-mail : michel.branchard@univ-brest.fr

Financial support: Xavier Leroy is supported by a grant of the Brittany Regional Council.

Keywords: Brassica, Cell proliferation, DNA-damage, ISSR, Somaclonal variation.

This report describes a new application of the Inter-Simple Sequence Repeat (ISSR) technique. This technology based on the amplification of regions between microsatellites was applied to different calli from the same cauliflower mother plant. One of the tested ISSR primers, (GATA)_4, generated great polymorphism. Twelve different markers were detected on polyacrylamide gels. After sequencing, one sequence showed homology with a predicted A. thaliana gene closely related to genes involved in the regulation of cell proliferation in mammalians. This marker is characterised by three microsatellites and a palindromic sequence. Possible causes of mutations in this marker are discussed and will be investigated. ISSR amplification appears as a reliable method in the determination of genetic instability at early stages in in vitro culture.

Over the last years several techniques for identifying genetic polymorphisms have been described. Restriction Fragment Length Polymorphism (RFLP, Botstein et al, 1980) analysis is one of the first techniques widely used to detect variation at the sequence level. Later, PCR-based techniques were developed, and among them one should mention Random Amplified Polymorphic DNA (RAPD, Williams et al, 1990), Arbitrarily Primed Polymerase Chain Reaction (AP-PCR, Welsh and McClelland, 1990), DNA Amplification Fingerprinting (DAF, Caetano-Anolles et al, 1991). Then, a collective term, the Multiple Arbitrary Amplicon Profiling (MAAP), was suggested to describe all the characteristics common to these closely related techniques (Caetano-Anolles et al, 1992). A more reproducible method, based on the detection of genomic restriction fragments by PCR amplification, the Amplified Fragment Length Polymorphism (AFLP, Vos et al, 1995), was described. The major drawback of these techniques is their sensitivity to DNA quality or reaction conditions.

Another method demonstrated that microsatellites, or Simple Sequence Repeats (SSRs), are useful as genetic markers for comparative analysis and mapping of genomes (Weber et al, 1989). However, the whole procedure from the construction of a genomic library to the synthesis of specific primers from the flanking sequences is time-consuming and expensive. Then, a variant of the SSRs technique has been described (Zietkiewicz et al, 1994); it uses as primers microsatellite oligonucleotides that amplify genomic segments different from the repeat region itself.

This approach, named Inter-SSR (ISSR), employs oligonucleotides based on a simple sequence repeat anchored or not at their 5’- or 3’-end by two to four arbitrarily chosen nucleotides. This triggers site-specific annealing and initiates PCR amplification of genomic segments which are flanked by inversely orientated and closely spaced repeat sequences. The study reported here was based on an inter-SSR PCR method to illustrate somaclonal variation leading to genomic instability.

The process of somaclonal variation is believed to require multiple genetic and/or epigenetic events which affect patterns of expression, or result in mutational alteration of genes. Various molecular mechanisms may be responsible for the DNA mutation and genetic instability leading to the development of variations. Such molecular mechanisms would include DNA damage and mutation, alteration of cell ability to repair damaged or mutated DNA, alteration...
of genes for cell-cycle control mechanisms, DNA methylation (Merlo et al, 1995). However, only few mechanisms of somaclonal variation have been described in the literature.

DNA damage can result from a spontaneous alteration of the DNA molecule, e.g. a mutation caused throughout DNA replication, DNA repair, gene arrangement, or chemical alteration resulting from oxidation and methylation. An individual degree of gene sensitivity to DNA damaging agents depends on factors such as the sequence of the gene, and whether it is actively transcribed or replicated (Markowitz et al, 1995). Various manifestations of patterns of genomic instability have been found in the human genome; several of them have been successfully used for the examination of human tumours. Microsatellite instability is one among these manifestations of genomic instability (Thibodeau et al, 1993). It corresponds to an alteration in size of simple repeat sequences. A finding of microsatellite instability implies the presence of mutations in at least one gene involved in DNA mismatch repair mechanisms (Strand et al, 1993). Thus, though a method to detect microsatellite length constitutes an important tool in genetic characterisation of cultivars, it does not allow one to detect major forms of genomic instability such as molecular aneuploidy arising from deletions, amplifications, translocations, insertions, recombination and chemical alteration (Stoler et al, 1999). The study reported here provides one with a method to detect genomic instability. ISSR technique should be useful to detect major types of somaclonal variation.

Materials and Methods

Plant material

Meurz, Jakez and H524 are Brassica oleracea var. botrytis cultivars obtained from Organisation Bretonne de Sélection (OBS, Plougoum, France). Seeds were sterilised and germinated on MS medium (Murashige and Skoog, 1962). Three-week-old hypocotyls were excised and placed onto a basal MS medium containing various Thidiazuron and Dicamba concentrations for three weeks. The experiments were conducted at 25 ± 1°C under a 16-h photoperiod of cool white fluorescent light (25 µE.m⁻².s⁻¹). The calli formed on hypocotyl explants were collected, termed CJ for Jakez calli, CH for H524 calli, CM for Meurz calli and stored individually at -80°C until DNA extraction. The results reported here deal with the calli obtained on Dicamba (10 µg.l⁻¹) for H524, or Dicamba (5 µg.l⁻¹) and Thidiazuron (10 µg.l⁻¹) for Jakez and Meurz.

DNA extraction, amplification and electrophoresis conditions

DNA was extracted from young leaves and 38 calli according to Leroy et al (1997). Four ISSR primers, (CAA)₃, (CAG)₃, (GACA)₃ and (GATA)₃, initially tested on 54 Brassica oleracea genotypes were applied to studies on clonal micropropagation.

Amplification products were performed in a GeneAmp PCR system 2400 (PE Applied Biosystems) using 20 ng genomic DNA, 2.5 mM MgCl₂, 75 mM Tris-HCl (pH 9.0), 1.5 U Taq DNA polymerase, 400 µM of each dNTP, 100 pM primer and deionized water (to a final volume of 25 µl). The amplification profile consisted of a first step at 94°C (1 min) followed with 30 cycles of 94°C (1 min) / 42°C (GATA₃), 54°C (GACA₃), 52°C (CAA₃), 64°C (CAG₃) (1 min) / 72°C (4 min), and 1 cycle at 72°C (7 min).

PCR products were electrophoresed either in 2% agarose gels and stained with ethidium bromide or in 6% non-denaturing polyacrylamide gels using Bioblock apparatus, then silver-stained (Gootlieb and Chavko, 1987). PCR fragments were eluted in 50 µl Tris EDTA (TE), purified by centrifugation at 10 000 g for 6 min with a mixture of chloroform/isoamyl-alcohol (50 µl) and precipitated by glycogen (1 µl) and ammonium acetate 3 M (50 µl). Ligations were carried out according to the Manufacturer's instructions with Promega T4 DNA ligase using 25 ng pGEM-T Easy Vector (Promega).

Markers characterisation

E. coli DH5α strain (supE44 Δlac U169 (F80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) (Sambrook et al, 1989) was used for transformation experiments and propagation of plasmids. Plasmid DNA was extracted from the transformed colonies according to Sambrook et al (1989) and re-amplified by specific T7 and SP6 primers (Promega). For optimum results, the PCR products were purified and concentrated by ultrafiltration through anisotropic membranes (Centricon YM-100 columns, Millipore).

The amount of purified DNA was quantified by measuring the absorbance at 260 nm. Cycle sequencing was performed in a GeneAmp PCR system 9600 (PE Applied Biosystems) using 10 ng PCR product DNA, 3.2 pmol primer, 8 µl Terminator Ready Reaction Mix and deionized water to a total volume of 20 µl. Optimised thermal cycling conditions were as follows: 25 cycles of 96°C for 10 sec, then 50°C for 5 sec, and then 60°C for 4 min. Once DNA sequencing reactions were completed, the excess of BigDye terminators was removed by purification of extension products on Centi-Sep spin columns (Princeton Separations).

PCR products electrophoresis was performed on the ABI Prism 310 Genetic Analyser. Putative genes were identified by combination of four gene prediction programs including ORF Finder (http://www.ncbi.nlm.nih.gov/orffinder), GRAIL (http://avalon.epm.ornl.gov/Grail-bin), GenScan (http://gnomic.stanford.edu), and NetPlantGene
Plant genomic instability detected by microsatellite-primers

Searches of the complete sequence against a peptide database and the Arabidopsis EST database at NCBI and at TIGR.

Figure 1. Banding profiles of cultivar ‘Jakez’ by (GATA)$_4$ with plant (JP), callus (JC), variant callus (JC2), and weight marker (WM, Smartladder, Eurogentec). Five different JC2 PCR products were loaded. Arrows indicate the absence of amplicons from the profile of no congruent callus and the presence of one additional band.

Results and Discussion

For analysing the domain arrangements of proteins, we applied the ProDom-CG, accessed at http://www.toulouse.inra.fr/prodomCG.html.

Assays were repeated 5 times for each sample from plant, normal and variant calli to ensure reproducibility. Alterations that recurred in the 5 assays were taken into account in the determination of the instability index. Testing of 38 calli on agarose gels revealed polymorphic markers for three calli named CH2, CJ2 and CM4.

Variations in fingerprint patterns resulted in the occurrence of a band of 1480 bp (CAA)$_5$ for the first variant (CH2), of two fragments at 2000 bp (CAA)$_5$ and 2200 bp (GATA)$_4$ for CJ2; they also led to the loss of one fragment at 800 bp (GACA)$_4$ for CJ2 and of one marker at 1150 bp (CAA)$_5$ for
the third callus CM4. No differences were observed when using \((CAG)_5, (CAA)_5\) detected polymorphism from three calli, whereas \((GACA)_4\) and \((GATA)_4\) confirmed the genetic variation of the same callus.

The polymorphism within calli was obvious presenting distinct fingerprints from each of them with five clear polymorphic markers. Previous results (Leroy et al., 1997) disclosed high levels of polymorphism between different Brassica clones on agarose gels with ethidium bromide staining. However, further investigations revealed a higher polymorphism on polyacrylamide with silver staining by allowing discrimination of monomorphic bands on agarose or detection of faint or not visible bands. As the detection of polymorphic markers was enhanced on polyacrylamide gels, GATA PCR products were electrophoresed on 6% polyacrylamide gels. The altered products of interest were excised from the gel, eluted, and re-amplified.

The re-amplified product was ligated as an insert into a vector. Further results only detailed variant GATA clone. The results from hybridisation-based fingerprinting with \((GATA)_4\) probes proved no intraclonal variability for Achillea millefolium (Weising et al., 1994), sugarbeet (Schmidt et al., 1993) and raspberry (Hoepfner et al., 1993).

In our study, \((GATA)_4\) primer generated one polymorphic band and 12 altered markers between variant (JC2) and normal calli (JC) on agarose and polyacrylamide gels, respectively (Figure 1). The variant was characterised by the loss of 11 amplicons all present in the control plant and the other calli. The same variant produced a profile containing an additional band at 2210 bp, identical to the 2200 bp agarose marker, that was missing in the other profiles (Table 1). Both polymorphisms were reproduced in other PCRs and gels.

The medium genomic instability index for this variant callus with \((GATA)_4\) was 35.3%, that is 35.3% of the bands generated from normal calli were altered in somaclonal variant DNA. These data suggest an initiation of genomic instability at an early stage of calli formation. It is likely that the number of genomic lesions will be greater in calli of a higher stage if the occurrence of such alterations increases over cellular proliferation.

After cloning and sequencing, one of these polymorphic markers, GATA291, was found very interesting further to its characterisation by simple perfect SSRs defined by Weber (1990) at its ends: indeed, GATA showed at least 4 repetitions and, at its 3’-end, it was adjacent to a block of imperfect tandem repeats of a (GT)$_{12}$ sequence (Figure 2). As a rule, compound repeats have not been well characterised, and their potential informativeness has not yet been established. However, the polymorphism information content (PIC value) in compound SSR sequences has been shown to be generally high. Weber (1990) noted that compound repeats account for about 10% of the total sequences cloned from the human genome and theorised that repeats composed of at least 12 units are consistently more polymorphic than shorter ones.

In our investigations homologies were found between the presented genomic sequence and Arabidopsis thaliana BAC and cDNA. Identities were of 77.1 and 78% (176/225), and the probabilities of 3.3e$^{-27}$ and 7.0e$^{-24}$ for the BAC and cDNA, respectively (Figure 3). The GENSCAN output predicted a peptide sequence of 68 aa, with a probability of 0.968 for correct exon. The comparison of the amino acid sequence against the protein sequence database confirmed predictions with an identity of (54/68) with a protein kinase of Arabidopsis thaliana (Figure 4) and with human and mouse ttk genes coding protein kinase.

![Figure 2. Nucleotidic sequence of the B. oleracea GATA291 marker. Nucleotides are numbered with reference to the initiation microsatellite.](image)

<table>
<thead>
<tr>
<th>1</th>
<th>11</th>
<th>21</th>
<th>31</th>
<th>41</th>
<th>51</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATAGATAGA TAGATAATAG TTAACTTGTG TGTGTGTTG TATGTATGT AACGTTCATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTCTTCATAG GTGACAGATA AGGCTTGGT TAAGGAAGTT ATGAGTGCA GTATGACTAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAAAGATGGG AAATGGAAG AAGACGGGTG TATATATATG GTACTCGAGT ATGGTGAAAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGATTTAGCT CACATGCTTT CACAAAAATG GCGGGAACTC GATGATTCCA AGTCGACAAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTTAGATGAA AACTGCGCTC GTTTTTACTG GCAGGTATCT ATCTATCTAT C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Protein kinases are generally divided into 2 broad classes based on their ability to phosphorylate either tyrosine (tyr) or serine/threonine (ser/thr) residues. However, several protein kinases have dual specificity and are able to...
GSDB:S:5158489|AC012193|ATAC012193 Arabidopsis thaliana (79697 nt)
initn: 563 init1: 450 opt: 643 Z-score: 573.6 expect 3.3e-27
77.093% identity in 227 nt overlap (50-276:2048-20270)

<table>
<thead>
<tr>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTTAACTTGTGTTGTGTTATGTGTATTACGTTCACTTTCAATAGGTGAGAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>::</td>
<td>::</td>
<td>::</td>
<td>::</td>
<td>::</td>
<td>::</td>
</tr>
</tbody>
</table>

GSDB:S GTTGGACATTATTTGCTCAACGAAACACTAATTCTCTGGTATTAGGTGACGAAT
20020 20030 20040 20050 20060 20070

GSDB:S AGGCTTTTTAAAGGAATTTAAGCTGCTGTATGACTAACAAAGATGGAAAGTAGTTAAG
20080 20090 20100 20110 20120 20130

Figure 3a. Sequence alignment between genomic sequence of B. oleracea and BAC.

phosphorylate the 3 amino acids. Because E. coli lacks phosphotyrosine, Mills et al. (1992) searched for novel tyrosine kinases by using an antibody against phosphotyrosine to screen a T-cell expression library in E. coli. They isolated cDNAs encoding a predicted 857-amino acid protein that they designated TTK. Northern blot analysis revealed that the TTK gene was expressed as a 4-kb mRNA in tissues with a large number of rapidly proliferating cells such as testis and thymus. When expressed in E. coli, TTK phosphorylated ser, thr, and tyr residues, although the level of tyr phosphorylation was low.

Lindberg et al. (1993) isolated a partial cDNA encoding the catalytic domain of TTK, which they referred to as PYT (phosphotyrosine-picked threonine kinase). They found that the catalytic domain of PYT acts primarily as a protein threonine kinase, but that the enzyme can phosphorylate tyr to a small extent, making it a potential dual-specificity protein kinase. Lindberg et al. (1993) stated that the Esk gene is the mouse ortholog of PYT. Mutations in the essential S. cerevisiae Mps1 (monopolar spindle-1) gene cause defects in the duplication of spindle pole body and cell cycle control. Lauze et al. (1995) cloned the Mps1 gene and found that it encodes a dual-specificity kinase with similarities to the mammalian TTK and Esk kinases. The present paper reports on an alteration of a fingerprinting pattern with loss of a marker corresponding to a part of a gene involved in the regulation of cell proliferation. This change would be the consequence of microsatellite instability at the primer fixative site, or of modification(s) either at the poly (GT) SSR or between the microsatellites. Strand et al. (1993) examined the genetic control of the stability of poly (GT) tracts and showed that the instability of poly (GT) tracts in yeast is increased by mutations in the mismatch repair genes (PMS1, MLH1, and MSH2); on the other hand, mutations affecting the proof-reading functions of DNA polymerases have no influence on this instability. Thus, an instability of SSRs may result from either an increased rate of DNA polymerase slippage or a decreased efficiency of mismatch pair. Liu et al. (1999) reported on the identification and characterisation of cDNAs containing dinucleotide microsatellite repeats. Further to the construction and screening of a microsatellite-enriched cDNA library using(CA)15 and (GA) probes, they
isolated 47 polyadenylated clones and found 15 unique cDNAs. Nine of them contained microsatellites in the 3′-nontranslated region (NTR), 5 others had microsatellites in 5′-NTR and one contained CA dinucleotide repeats in its coding region.

One among the nine cDNAs mentioned above contained trinucleotide repeats in the coding region additional to CA repeats in its 3′-NTR. Some of these genes are known to be involved in regulatory functions such as DNA repair protein and cyclin-dependent kinase. Further to the analysis of (CAA)$_n$ and (CAG)$_n$ repeats (corresponding to glutamine residues) in *Triticum aestivum*, Morgante and Olivieri (1993) highlighted their presence in many sequenced genes belonging to the family of gliadins.

In our case, the inactivation (or invalidation) of the protein kinase gene would generate a specific phenotype that was observed because of the sacrifice of the calli. However, the plant homologue *ttk* gene can be characterised by at least 3 microsatellites. The discovery of mutation in this gene could be the start point of further investigations aimed at a better understanding of cell proliferation and differentiation. In a previous study (Leroy et al, 1997) mutations were characterised with other microsatellites, including (CAA)$_5$. Tracts of polyglutamine have been found in transcription factors and play a role in normal protein function shown by their ability to activate transcription when present in DNA-binding proteins. At the present time unstable mutations by trinucleotidic expansions are known to be involved in 12 human diseases.
Figure 4. Sequence alignment between deduced amino acid sequences of B. oleracea predicted gene bok1 and A. thaliana protein kinase.

As suggested by others results (XJL and KL, manuscript in preparation), homologies of ISSRs with transcription factors may play a major role in transcriptional activity by modifications of polyglutamine tracts. One should notice in the genomic sequence presented above an inverted sequence from –223 to –245 with a palindromic motif of 6 bp in its centre. Inverted repeats can base pair to the complementary strands or to themselves. Since single-stranded DNA cannot bend enough, adjacent nucleotides on one strand can create hydrogen bond like base pairs. There is a mandatory 3-6 bp spacer or loop between the 5' and 3' stems that allow the DNA to fold back on itself. These stem-loop structures lead to the formation of hairpins (single-stranded DNA) or cruciforms (both strands of the duplex DNA).

<table>
<thead>
<tr>
<th>Marker names</th>
<th>Band sizes (bp)</th>
<th>JP/JC</th>
<th>JCNC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA2220</td>
<td>2218</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>GATA840</td>
<td>837</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA560</td>
<td>564</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA510</td>
<td>512</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA500</td>
<td>500</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA435</td>
<td>434</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA380</td>
<td>379</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA370</td>
<td>368</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA360</td>
<td>359</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA330</td>
<td>333</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA291</td>
<td>291</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>GATA280</td>
<td>282</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Polymorphic fragments from somaclonal variant. This table shows the sizes of polymorphic PCR-products between plant and variant clonal callus. Presence (+) or absence (-) of bands are indicated.

Regions of DNA alternative secondary structure pose a barrier to replication fidelity. Previous studies in E. coli have demonstrated that the extruded cruciform facilitates frameshift mutations by bringing the DNA slippage sites (direct repeats) into close proximity (Glickman and Ripley, 1984; Schaaper et al, 1986). This can lead to deletion- as well as insertion-mutations. The replication of the strand is likely interrupted at the meeting point with the replication fork and the extruded cruciform.

Theoretically, like perfect palindromes imperfect ones can also mediate similar mutations and, in addition, be involved in intra- and inter-strand switching, which increases the spectrum of potential mutation (De and Ripley, 1984, Rosche et al, 1997). Intrastrand switching requires that the polymerase replicates through the centre of the palindrome and, with the nascent chain, disassociates from the template strand to engage itself in intrastrand annealing into a hairpin structure and continue to polymerise (Papanicolaou and Ripley, 1989; 1991). This situation is transient, and following additional polymerisation, the nascent strand and polymerase switch back to the original template strand. If the template is a quasipalindrome, intrastrand switching will convert the quasipalindrome into a perfect inverted repeat. The mutation can then be fixed by DNA repair or a second round of replication.

Inverted repeats have several biological roles, pose a special impediment to DNA replication fidelity and are associated with several human disease-related genes.
In summary, cloning and sequencing of the altered ISSRs bands can be used in the identification and mapping of altered regions of the genome. A precise definition of the kind of mutations detected may further yield important clues to novel somaclonal phenotypes.

Mutations detected by microsatellites primers suggest several possible alterations in the genome rapidly accessible by ISSR. Researches based on this model may be used to assess the genotoxic effects of various compounds such as herbicides. With a same approach, genomic DNA would be subjected to analysis by restriction enzymes to detect, for example, methylation status; moreover, the whole gene could be analysed through the development of amplified consensus gene markers (ACGM).

**Acknowledgements**

The authors thank Dr Waly Dioh at Monsanto for helpful discussions and Dr M.-P. Friocourt for English corrections.

**References**


Plant genomic instability detected by microsatellite-primers by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6231-6235.


