

Carlos M. Rodríguez López · Andrew C. Wetten
Michael J. Wilkinson

Detection and quantification of in vitro-culture induced chimerism using simple sequence repeat (SSR) analysis in *Theobroma cacao* (L.)

Received: 18 May 2004 / Accepted: 22 September 2004 / Published online: 23 October 2004
© Springer-Verlag 2004

Abstract Mutation rates are often elevated in plants regenerated from in vitro culture, giving rise to so-called ‘somaclonal variation’. Detailed characterisation of mutation profiles that arise during culture should improve our understanding of processes influencing mutation and allow the selection of protocols yielding the fewest/least severe changes. Somatic mutations will usually produce genetic chimeras where unchanged alleles are retained by some cells. Such chimeras are difficult to detect but likely to form a significant proportion of any regenerant population. We present a simple protocol that enables the provisional diagnosis of both homogenous and chimeric mutants among large regenerant populations, together with a semi-quantitative means of estimating the proportion of mutant cells. The assay exploits consistent differential amplification of alternate simple sequence repeat alleles at heterozygous loci. Calibration of the relative amplification of alleles from two genotypes—and the synthetic chimeras created from them—revealed a strong linear relationship between ‘peak heights’ representing alternate alleles following capillary electrophoresis. The assay predicts chimeric composition to a reasonable level of confidence ($\pm 5\%$) so long as the infrequent allele exceeds 15% of the template. The system was applied to 233 regenerants of cocoa somatic embryogenesis and identified 72 (31%) putative chimeric mutants for slippage mutation or allele loss across two loci.

seed. The crop is relatively poorly developed from its wild relatives (Cuatrecasas 1964) and is susceptible to many pests and diseases (Kennedy and Mooleedhar 1992). Furthermore, small plantations of selected genotypes are generally established from clonal cuttings. Whilst it is possible to generate commercial-scale propagation systems using budwood (Traore et al. 2003), there is increasing interest in the application of somatic embryogenesis systems (Lopez-Baez et al. 1993; Li et al. 1998; Traore et al. 2003). These protocols offer a greater capacity for the rapid propagation of elite trees but carry the risk of producing mutant regenerants in such numbers that economic viability of the approach is compromised (Plader et al. 1998). This so-called ‘somaclonal variation’ (Larkin and Scowcroft 1981) has been defined as genetic and phenotypic variation among clonally propagated plants of a single donor clone (Kaepler et al. 2000) and includes genomic (large-scale deletions and gross changes in chromosome structure/number and directed and undirected point mutations) (Joyce et al. 2003) and epigenetic changes (histone acetylation, DNA methylation, chromatin remodelling, etc). Somaclonal variation can thus be broadly divided into genetic and epigenetic changes: both forms can cause profound phenotypic change, but the epigenetic changes are often heritable only through somatic division and normally revert to parental phenotypes following meiosis. For this reason, genetic changes have the greater relevance for breeding (Kaepler et al. 2000), although epigenetic changes are important with respect to plant distribution and storage. It should be noted, however, that certain epigenetic phenomena that occur following tissue culture can remain quite stable for many generations (Kaepler et al. 2000). The precedence for this scenario is provided by the ‘mantled inflorescence’ somaclonal variants that compromised the commercial propagation of oilpalm clones by somatic embryogenesis in the 1990s (Jaligot et al. 2000).

Phenotypic ‘off-type’ regenerants of cocoa have been reported following somatic embryogenesis (Alemanno et al. 1996) although it was not clear whether the

Introduction

Cocoa (*Theobroma cacao*) is a tropical tree crop with a typical juvenility period of 3 years when grown from

Communicated by J.W. Snape

C. M. Rodríguez López · A. C. Wetten · M. J. Wilkinson (✉)
School of Plant Sciences, The University of Reading,
Whiteknights, Reading, RG66AS, UK
E-mail: m.j.wilkinson@reading.ac.uk

abnormal phenotypes arose only from genetic causes and whether some of the somaclonal variants were chimeric for the change. In any event, the appearance of morphological abnormalities arising from somaclonal variation represents a potential problem that may threaten the commercial viability of cocoa somatic embryogenesis for clonal multiplication purposes (Mohan 2001). There is therefore a need to provide some quantitative basis on which to compare the incidence of somaclonal variation from different micropropagation protocols, thereby enabling breeders to identify those giving rise to the fewest variants. Additionally, it is important to characterise the main causes of variation generated by the system under study. There is a wide range of possible approaches for this purpose. At the simplest level, the tree's extended juvenile period renders the phenotypic screening of regenerants impractical. More information can be gained from cytological studies, where gross genomic changes and the presence of cytological chimerism have been inferred from chromosome counts (Tremblay et al. 1999; Cammareri et al. 2002), karyotype analysis (Fourré et al. 1997) and flow cytometry (De Schepper et al. 2001; Gimenez et al. 2001; Lim and Loh, 2003; von Aderkas et al. 2003). However, these approaches lack the resolution to characterise small-scale mutations and are generally poorly suited to quantitative analysis of genetic change. We have therefore adopted a molecular strategy to screen regenerants for genetic change.

Several studies have used molecular approaches to detect somaclonal variation, with most applying multi-locus systems such as random amplified polymorphic DNAs (RAPDs; Rani et al. 2000; Feuser et al. 2003), inter-simple sequence repeats (ISSRs; Albani et al. 1998; Leroy et al. 2001) and amplified fragment length polymorphisms (AFLPs; Matthes et al. 2001; Joyce and Cassells 2002). These systems potentially provide wide genome coverage but do not readily allow diagnosis of the cause of any polymorphism detected. Significantly, the dominant nature of most markers generated by these systems means that it is not possible to detect mutation chimeras. Conversely, Simple sequence repeat (SSR) markers are codominantly inherited and so could detect mutations that are present in only a small proportion of cells. SSR markers have been used previously to detect somaclonal variation (Diwan and Cregan 1997) and even chimerism for the presence of a novel allele that arose through strand slippage during DNA replication (slippage mutation) (Franks et al. 2002). Chromosome and locus loss represents an additional source of genetic change for which SSRs could also have value. During PCR, alternate alleles at a heterozygous locus have been found to amplify in repeatable proportions with the resultant allelic peaks often being of approximately the same size (Clayton et al. 1998). Reproducibility in the relative amplification efficiencies of alternate alleles may allow for the detection of allele loss or the presence of a new allele within a mutation chimera, even if it is present in relatively few cells. Given that parental and

mutant allele amplifications are performed in the same reaction, and with the same primers, direct comparison of the allele peak profiles of the parental clone with that of the chimeric mutant may even allow for a semi-quantitative assessment of mutant cell abundance. The validity of such an approach would be dependent on an accurate measure of the amount of amplified DNA and on a strong relationship existing between the relative abundance of the template from the mutant and parental cell types. In this study, we examine the feasibility of using sets of mapped, fluorescently labelled SSR markers (Lanaud et al. 1999; Risterucci et al. 2000) to detect mutation chimeras and evaluate the scope for semi-quantitative assessment of cell mixtures within them. We then aim to characterise the extent of genetic mutation among 233 cocoa regenerants from somatic embryogenesis.

Materials and methods

Plant material

Source material was obtained from cocoa trees in the Intermediate Quarantine Facility, The University of Reading. Closed immature flower buds (4–5 mm) were collected from genotypes LCT EEN 37/I, LCT EEN 162/S-1010, SC 3 and SIAL 93 between 0800 hours and 1200 hours in February 2001. Staminodes isolated from these buds were used to induce and regenerate somatic embryos according to Li et al. (1998).

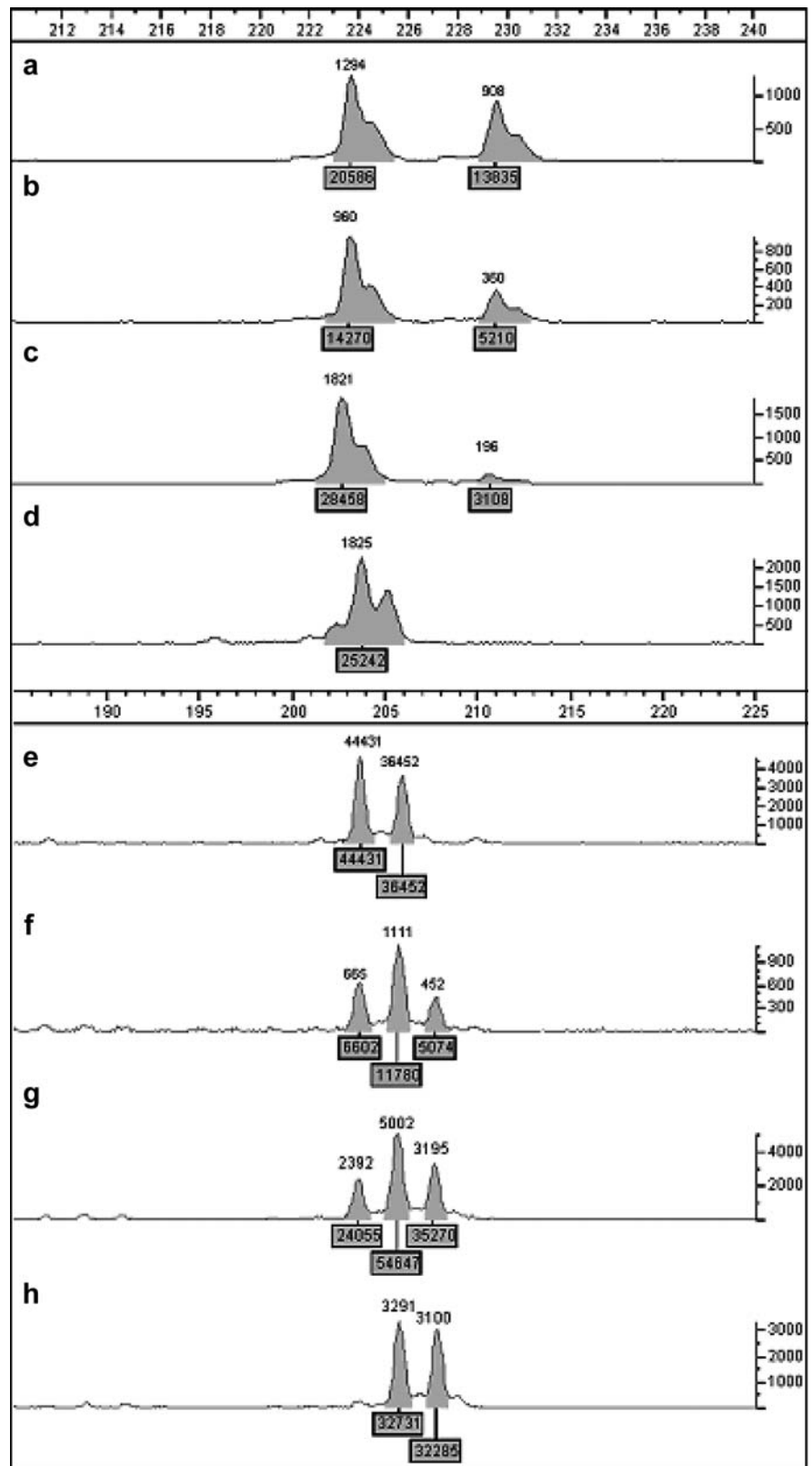
DNA isolation

DNA was extracted from the cotyledons and leaves of *in vitro* plantlets and from recently expanded leaves of trees using the DNeasy 96 Plant kit (Qiagen, Valencia, Calif.) and the Mixer Mill MM 300 (Retsch, Germany). Isolated DNA was diluted in nanopure water to produce working stocks of 5 ng/μl. There were 20 DNA replicate extractions performed from each tree and two replicate extractions for each regenerant plantlet (13 regenerants from LCT EEN 37/I; 180 from LCT EEN 162/S-1010; 20 from SC 3; 20 from SIAL 93).

Amplification of target sequences with PCR

Fifteen fluorescently labelled SSR primers were tested (mTcCIR1, mTcCIR2, mTcCIR3, mTcCIR6, mTcCIR7, mTcCIR8, mTcCIR10, mTcCIR15, mTcCIR17, mTcCIR18, mTcCIR19, mTcCIR24, mTcCIR25, mTcCIR26, and mTcCIR61) across all plant material. The primer sequences, map positions and conditions for amplification are described by Lanaud et al. (1999) and Risterucci et al. (2000). Labelled PCR products were diluted by 1/10, and 1-μl samples were mixed with 1 μl ROX/HiDi mix (50 μl ROX plus 1 ml HiDi formamide). The

Fig. 1 a–d Simple sequence repeat (SSR) profiles generated by mTcCIR6 using the following template mixtures to simulate a genetic chimera for allele loss: **a** 100% genotype SIAL 93 (wild type), **b** DNA mix of 50% SIAL 93 and 50% LCT EEN 162/S-1010 (simulates 66% mutant cells in chimera), **c** DNA mix of 33% SIAL 93 and 66% LCT EEN 162/S-1010 (80% mutant cells), **d** 100% LCT EEN 162/S-1010 (100% mutant). **e–h** SSR profiles from mTcCIR10 using the following template mixtures to simulate genetic chimeras where slippage has created mutant cells with a novel allele: **e** 100% genotype SIAL 93 (wild type), **f** DNA mix 50% LCT EEN 162/S-1010 and 50% SIAL 93 (simulates 50% mutant cells in chimera), **g** DNA mix of 33% SIAL 93 and 66% LCT EEN 162/S-1010 (66% mutant cells), **h** 100% LCT EEN 162/S-1010 (100% mutant). *Horizontal axis* indicates allele size in base pairs, *vertical axis* indicates peak intensity (peak height). *Numbers above peaks* indicate peak height, *numbers in boxes* indicate peak area



samples were denatured at 95°C for 3–5 min and snap-cooled on ice for 2 min. PCR products were subsequently fractionated on an ABI PRISM 3100 (PE Applied Biosystems, Foster City, Calif.) at 3 kV for 22 s and at 15 kV

for 45 min, at 60°C. Markers producing two alleles in the parental tree were selected when screening regenerants for chimeric somaclonal variants arising from allele loss (e.g. chimeras caused by the deletion within or elimination of

one homologous chromosome) or allele duplication and slippage. In this way, the change to either allele could be detected and characterised. Accordingly, markers mTcCIR1 (HEX), mTcCIR6 (6-FAM), mTcCIR7 (HEX), mTcCIR8 (6-FAM), mTcCIR15 (NED) and mTcCIR61 (6-FAM) were used for the analysis of LCT EEN 37/I regenerants; markers mTcCIR10 (6-FAM) and mTcCIR15 (NED) were selected for genotype LCT EEN 162/S-1010 regenerants; mTcCIR1 (HEX), mTcCIR3 (NED), mTcCIR6 (6-FAM), mTcCIR7 (HEX), mTcCIR8 (6-FAM), mTcCIR10 (6-FAM), mTcCIR15 (NED) and mTcCIR61 (6-FAM) were used for genotype SC 3 regenerants; mTcCIR3 (NED), mTcCIR6 (HEX), mTcCIR8 (6-FAM), mTcCIR10 (6-FAM), mTcCIR15 (NED), mTcCIR17 (NED) and mTcCIR61 (6-FAM) were used for SIAL 93 regenerants.

Relationship between peak heights or peak areas of alternate SSR alleles

Analysis of the relative peak heights of alternate SSR alleles in the parental control tree profiles (mTcCIR10 and mTcCIR15 were used for the analysis of LCT EEN 162/S-1010, and mTcCIR6, mTcCIR8, mTcCIR10, mTcCIR15, mTcCIR17 and mTcCIR61 for SIAL 93) was executed with GENOTYPER 3.7 NT software (Applera Corp). As well as sizing the detected products for allelic designation, this software also displays peak heights and automatically calculates the area under each peak. Thus, the peak-area and peak-height values accompanying the allele sizes provide valuable quantitative information about the profile. Standard curves were generated by plotting peak-height or peak-area values of the earliest peak (representing the alleles with the lowest molecular weight) against that of the later peak (representing the allele with the highest molecular weight). A linear regression line was then produced for each group of samples and variable using MICROSOFT EXCEL.

Creation of synthetic chimeras using template mixes

DNA from genotypes LCT EEN 162/S-1010 and SIAL 93 was diluted to create working stocks of 5 ng/ μ l. Uniformity of stock concentrations was assessed with the analysis programme LABWORKS (Analisis software, USA) following fractionation by electrophoresis through a 1.5% (w/v) agarose gel at 120 V for 1 h. These solutions were used to create artificial chimeric standards with different proportions of each genotype (1:0; 5.7:1; 3:1; 2:1; 1:1; 1:2; 1:3; 1:5.7; 0:1). We used 5 μ l of the artificial chimeric template in all subsequent PCR reactions.

Primers mTcCIR6 and mTcCIR10 were used to attempt semi-quantitative analysis of the artificial chimeric templates. These markers map to cocoa linkage groups 6 and 5, respectively (Lanaud et al. 1999; Risterucci et al. 2000). In both cases, the forward primers were fluorescently labelled at the 5'-end; mTcCIR6Forward contains a 6-FAM reporter molecule and mTcCIR10Forward

contains a HEX reporter molecule. When used on a SIAL 93 template, the mTcCIR6 primers amplifies two alleles (224-nt and 230-nt fragments) (Fig. 1a), whereas the same primers used on LCT EEN 162/S-1010 yield only the 224-nt allele (Fig. 1d). However, because LCT EEN 162/S-1010 contains two copies of the constant allele, compensation was made when calculating the relationship between peak ratio and simulated chimerism (e.g. a 1:1 DNA mix ratio corresponded to a 3:1 224:230 allele ratio). Accordingly, artificial chimeras were generated by mixing DNA templates to simulate a SIAL 93 genetic chimera for loss of the 230-base allele (Fig. 1b, c). In comparison, primer mTcCIR10 generates two alleles for both genotypes (SIAL 93: 203 bp and 205 bp; LCT EEN 162/S-1010: 205 bp and 207 bp) (Fig. 1e, h). As both genotypes are heterozygous and possess only one common allele, the resultant DNA pool can be used to simulate a genetic chimera for allele loss in SIAL 93 (loss of the 203-base allele) or LCT EEN 162/S-1010 (loss of the 207-base allele) or a chimera for a novel allele by slippage mutation (207-base allele appearing in SIAL 93 or the 203-base allele appearing in LCT EEN 162/S-1010) (Fig. 1f, g). Here, no correction for allele duplication was necessary.

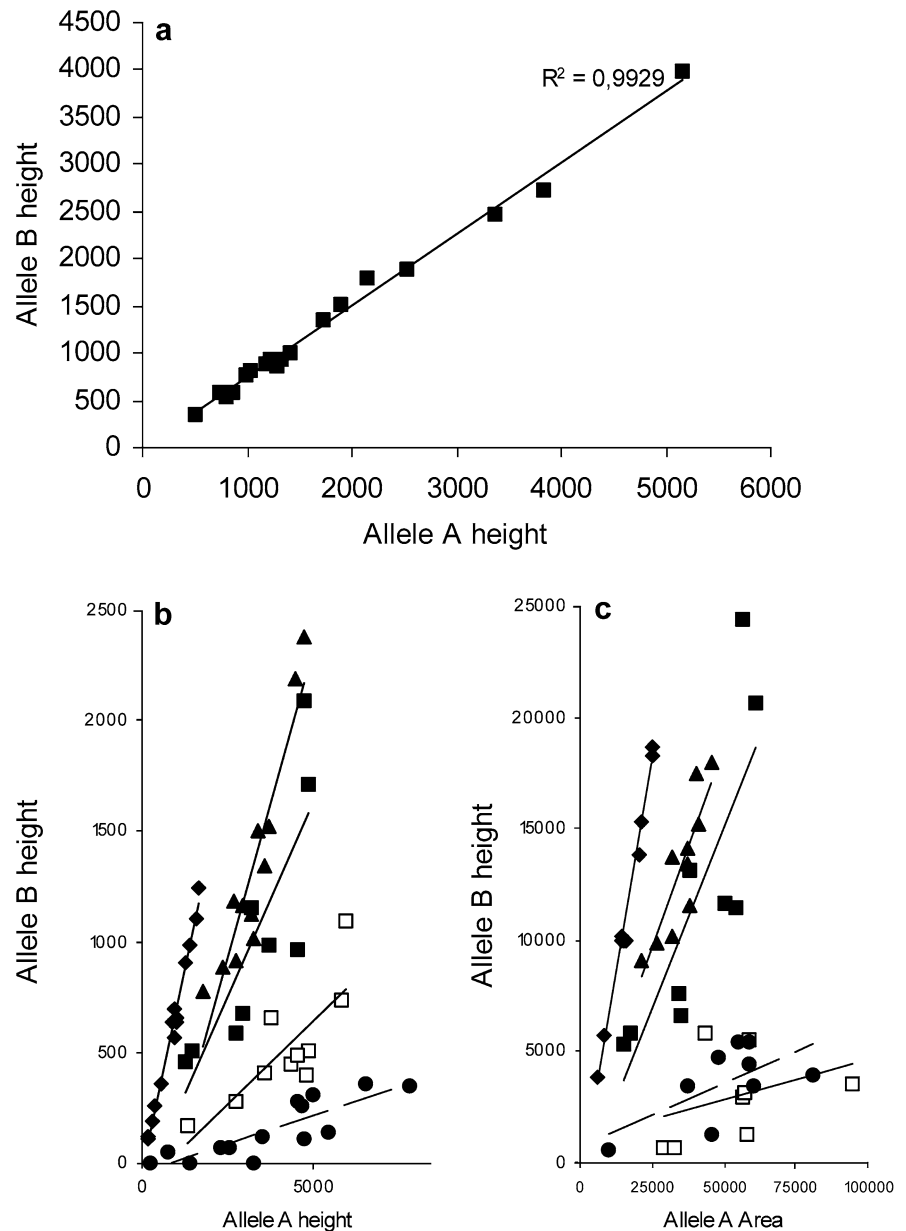
Data analysis

The analysis and interpretation of a somaclonal variant SSR profile involves a series of steps. Relative peak heights of alleles in the parental control and regenerant profiles were measured using GENOTYPER 3.7 NT as described above. Standard curves were generated by plotting peak-height values for alternate alleles. Linear regression was applied for each marker using MICROSOFT EXCEL. A simple linear relation with groups was produced, and GENSTAT 6 (Lawes Agricultural Trust, Oxford, UK) was used with the relative peak heights to compare the different relative DNA concentration groups. $P < 0.01$ constituted statistical significance for the general model comparing the variables and $P < 0.05$ constituted statistical significance for the groups.

Markers applied to regenerants were tested for the presence of aberrant genotypes possessing significantly deviant peak-height ratios. Samples were designated as 'off-type' when both DNA extractions presented a peak ratio outside the standard deviation range of the mean ratio for the entire population. The remaining samples were deemed to have 'wild-type' peak ratios. The significance of the aberrant profiles was tested using 95% confidence interval for each sample obtained from regression analysis computed using MINITAB 13.1 (MINITAB, State College, Pa.). Groups obtained were compared with the parentals by linear regression analysis, with groups assigning 'wild-types' and 'off-types' as two groups with different lines.

By definition, chimeras contain some 'wild-type' cell lineages while the rest are aberrant. It is therefore expected that the profiles from chimera cell mixes will vary between two extremes: one where the template contains mostly unchanged cells and so yields a 'wild-type' pro-

Fig. 2 Alternate SSR allele amplification from pure and mixed template DNA.
a Relative efficiency of alternate SSR allele amplification using pure template DNA: linear regression analyses of alternate allele peak height from 20 PCRs (independent DNA extractions) of genotype SIAL 93 using the SSR marker mTcCIR8 (*solid squares*).
b,c Effect of template mix composition (synthetic chimeras) on the relationship between alternate allele peak heights (**b**) or peak areas (**c**) generated using marker mTcCIR6. Rhomboids represent 100% template of the (wild type) heterozygous genotype, SIAL 93. *Triangles* represent a 66% SIAL 93–33% LCT EEN 162/S-1010 mix, simulating a chimera in which two-thirds of the cells are wild type and one-third have lost one allele, *solid squares* represent a 50% SIAL 93–50% LCT EEN 162/S-1010 mix, *squares* represent a 25% SIAL 93–75% LCT EEN 162/S-1010 mix, *circles* represent a 15% SIAL 93–85% LCT EEN 162/S-1010 mix, *circles* represent a 15% SIAL 93–85% LCT EEN 162/S-1010 chimera is indicated by a *broken line*



file, and the other where the extraction is dominated by mutant cells, and so produces an aberrant profile. Clearly, such variation is not expected from genetically homogenous samples. Thus, a second analysis was performed to identify putative chimeras based on excessive variance between replicates. Here, a sample was deemed to be chimeric when the difference between the larger and smaller peak-height ratios exceeded the mean difference between replicates in the total population plus the standard deviation for this difference.

Results

DNA analyses were carried out on genetic material obtained from PCR products of 20 individual DNA

extractions from each genotype using 15 primer pairs (mTcCIR1, mTcCIR2, mTcCIR3, mTcCIR6, mTcCIR7, mTcCIR8, mTcCIR10, mTcCIR15, mTcCIR17, mTcCIR18, mTcCIR19, mTcCIR24, mTcCIR25, mTcCIR26 and mTcCIR61). These primers generated 77 alleles in the four genotypes tested. Profiles were conserved between replicate DNA extractions of all parents except for locus MTcCIR15 (linkage group 1) in SIAL 93. This microsatellite showed an off-type in which one of the parental alleles (238 bp) had apparently been reduced in size by a single base (237 bp). Sequencing of the three alleles showed a two-base variation from the expected size and a sequence difference between the wild-type alleles and the off-type allele for mTcCIR15 unexpectedly missing a dinucleotide GA repeat compared to its wild-type homologue (Data not shown).

All of the heterozygous SSR markers analysed for genotypes SIAL 93 and LCT EEN 162/S-1010 exhibited a strong linear relationship between peak heights of alternate alleles (Fig. 2a, b) and between peak areas of alternate alleles (Fig. 2c). When the strength of the relationship of the two variables was compared, peak-height ratios consistently produced the higher correlations (Table 1).

We next investigated the possibility of exploiting the consistency in these relationships to predict the composition of cell lineages in genetic chimeras. To do this, we created synthetic chimeras by mixing DNA from one heterozygous genotype with that from another that shares one of the two alleles. The first aim was to mimic a genetic chimera where one allele had been lost from some cells. For this, we used SSR marker mTcCIR6 that generates two alleles (224 and 230) in genotype SIAL 93 (Fig. 1a) and only one allele (224) in genotype LCT EEN 162/S-1010 (Fig. 1d). Manipulation of the relative proportions of each genotype simulated a genetic chimera in which a proportion of cells had lost the high-molecular-weight '230-base' allele (Fig. 1a–d). There was a strong linear relationship between the relative peak heights of alternate alleles in the control samples (one genotype only) and those in the variously mixed template samples (Fig. 2a–c). Moreover, there was a marked difference between the synthetic chimeras, with the slope of the linear relationships declining as the proportion of the genotype lacking one allele increased in the template mixture (i.e. as the synthetic chimera mimicked an increased incidence of allele loss). Slightly higher deviations and weaker correlations were also noted as the relative contribution of the genotype with the lost allele increased (Table 2). A

similar picture emerged when the peak areas of alternate alleles were examined (Table 2), except that the relationship between relative peak areas and template mix composition was strong only for samples comprising less than 66% of the genotype lacking one allele. Indeed, in the most extreme mix (92% simulated allele loss) the slope of the relationship was greater than that in the mix containing 84% simulated allele loss (Fig. 2c). Accordingly, all subsequent analyses focussed on the use of peak heights as the predictor of genetic composition of the synthetic chimeras.

Direct comparison of alternate alleles from mTcCIR6 in the simulated chimeras and parental genotypes revealed that overall, the mean peak-height ratio decreases with increasing abundance of the 'lost' allele and follows a declining sigmoid curve (Fig. 3a) described by the formula:

$$y = 1.2014x^3 - 2.2308x^2 + 0.3393x + 0.6886$$

where y is measured peak-height ratio (230/224 allele peak height) and x is the percentage of 'mutant' DNA in the sample.

Profiles of SSR marker mTcCIR10 yielded three alleles among the parental genotypes [203 bases and 205 bases in SIAL 93 (Fig. 1e), 205 bases and 207 bases in LCT EEN 162/1010 (Fig. 1h)] of which only one was conserved (205 allele). In the simulated chimeras, therefore, there was a gradual decline in allele 203 and a concurrent increase in allele 207 as the proportion of LCT EEN 162/1010 increased in the mix (Fig. 1e–h). When all three peaks were considered collectively, there was a wide standard deviation observed in the 'triple' peak ratios of all mixes (data not shown), and so we

Table 1 Linear correlation between alternate SSR alleles

r^2 value	SSR							
	Genotype LCT EEN 162/S-1010		Genotype SIAL 93					
	mTcCIR10	mTcCIR15	mTcCIR6	MTcCIR8	mTcCIR10	mTcCIR15	mTcCIR17	mTcCIR61
Height ^a	0.9764	0.7982	0.9915	0.9929	0.9698	0.9837	0.9944	0.9249
Area ^a	0.9248	0.7966	0.9911	0.9924	0.9227	0.9832	0.9911	0.9238

^aPeak-height and peak-area correlations are compared using r^2 values obtained from marker linear correlations from LCT EEN 162/S-1010 and SIAL 93 donor plants for each SSR

Table 2 Linear correlation between alternate SSR alleles in synthetic chimeras

r^2 value	Proportion of mutant DNA (%)						
	0	26	40	50	66	84	92
Height ^a	0.9915	0.9325	0.7512	0.8653	0.6918	0.6425	0.7198
Area ^a	0.9911	0.9043	0.7193	0.7373	0.6701	0.1379	0.4134

^aPeak-height and area correlations of different DNA mixes are compared using r^2 values obtained from marker linear correlations from synthetic chimeras for SSR marker mTcCIR6

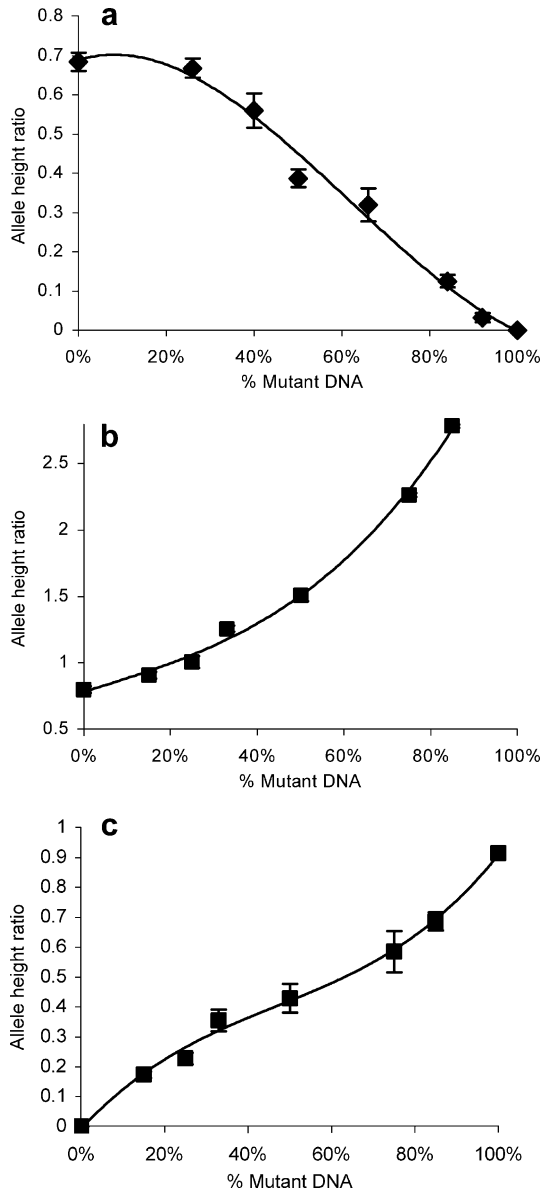


Fig. 3 Effect of template mix composition on mean peak heights of alternate SSR alleles. Mean value of ten replicates for peak-height ratio of alternate alleles are plotted against template mix composition, with standard deviations indicated. *Rhomboids* represent locus mTcCIR6 ($r^2=0.9882$), squares represent locus mTcCIR10. **a** Change in mean peak ratio from a 0% content of ‘mutated’ template (i.e. simulating pure ‘wild-type’ genotype SIAL 93) to 100% LCT EEN 162/S-1010 (simulating homogenous mutant tissue). The graph simulates increasing loss of the second allele of a pair. **b** Changing peak-height ratio for mTcCIR10 as the template mix contains an increasing proportion of mutant cells that lack one allele ($r^2=0.9963$). The graph simulates increasing loss of the first alternate allele of a pair. **c** Changing peak-height ratio for locus mTcCIR10 as template mix contains an increasing proportion of a novel allele from LCT EEN 162/S-1010 ($r^2=0.9955$). This graph simulates increasing abundance of a novel allele in the chimera

elected to compare pairs of alleles. As before, we arbitrarily set the low-molecular-weight allele in the pair as the divider. Considering only the loss of the SIAL 93-specific ‘203 allele’ relative to the conserved ‘205 allele’,

the peak-height ratio increased as the percentage of LCT EEN 162/1010 (mutant template) increased according to an exponential curve (Fig. 3b) that followed the formula:

$$y = 1.9564x^3 - 0.1306x^2 + 1.0252x + 0.7814$$

where y is measured peak-height ratio (205/203) and x is the percentage of ‘mutant’ DNA

We then considered the novel allele (207 bases) that simulated one that had been generated by the slippage mutation of the ‘203 allele’ in some cells. Once again, we divided the second of the two alleles (207) by the first (205). As expected, the peak ratio exhibited an inverse pattern to the simulated lost allele (see above), with the mean peak ratio increasing with increasing novel allele abundance according to a sigmoid curve (Fig. 3c) described by the formula:

$$y = 1.4031x^3 - 1.9824x^2 + 1.4956x - 0.0068$$

Overall, therefore, robust and repetition-consistent peak-height-ratio relationships were obtained for different amplification efficiencies performed on the same genotype. The slopes of the peak-ratio relationship generated by synthetic chimeras differ from that observed in the parental genotypes in a fashion that closely correlated with mix composition. Furthermore, the relationship between ratios remained constant when the simulated chimeras mimicked the loss of a high-molecular-weight allele (Fig. 3a), the loss of a low-molecular-weight allele (Fig. 3b) or the appearance of a novel allele by slippage mutation (Fig. 3c). Therefore this analysis was applied to cocoa regenerant material recovered from somatic embryogenesis.

Analysis of somatic embryo-derived populations

Relative allele-peak heights were recorded in 233 cocoa regenerants. For this, we applied labelled SSR markers to study regenerant populations from the four cocoa genotypes. Each genotype regenerant population exhibited a strong relationship between peak heights similar to that observed in the parental clones (Fig. 4a–c). However, some regenerants fell outside this relationship, and these were deemed to be chimeric mutants when the profiles from both replicated DNA extractions fell outside the 95% confidence intervals obtained by regression analysis (Fig. 4b, c). Yet this system proved to be ineffective in discriminating aberrant peak ratios among samples when the amplification was weak, even when peak-height ratios were widely divergent from those of the parent (Fig. 4c). For this reason, regenerants were re-examined and deemed to possess aberrant ratios (putative chimeras) when height ratios fell outside the confidence limits of the wild-type group set by linear regression analysis with groups at 95% confidence intervals. When this approach was applied to genotype LCT EEN 162/S-1010 regenerants, 17.75% of the plants

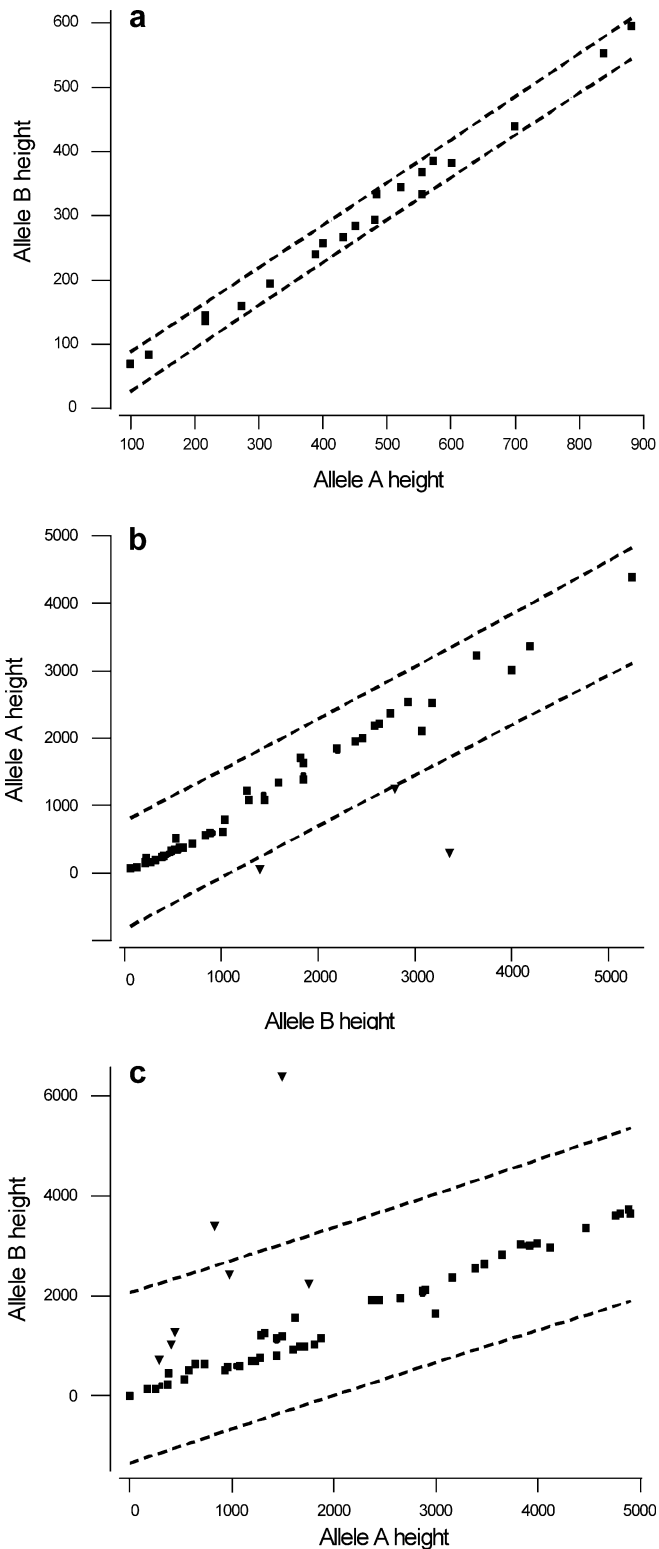


Fig. 4 Regression analyses of alternate allele-peak heights from parental and regenerant material. *Dashed lines* represent the 95% prediction interval, *squares* represent wild-type samples, *inverted triangles* represent samples presenting off-type peak intensity ratios. **a** Fitted line plot for SSR marker mTcCIR6 for the SIAL 93 parental sample population. **b** Fitted line plot for SSR marker mTcCIR6 for the SIAL 93 regenerant population. Note three regenerant samples presenting off-type peak-height ratios outside of the 95% prediction intervals. **c** Fitted plot line for SSR marker mTcCIR17 for SIAL 93 regenerant population. Note that some samples with low peak intensities appear deviant but still lie within the 95% prediction intervals

on the marker used (Table 3). When replicate ratios from each sample were compared and evaluated for excessive variability, the percentage of putative mutants increased accordingly (Table 3). The total number of mutants detected after combining both analyses was 72, which represents 31% of the regenerant population.

Discussion

Over the past two decades, somaclonal variation has been extensively investigated using cytological, biochemical and molecular techniques (Harding 2004), and the basis and uses of such variation have been extensively discussed (see Evans 1998; Martin et al. 2001). We have shown that a strong, linear relationship exists between SSR allele peak-height ratios in heterozygous genotypes. This supports previous studies showing that alleles at a heterozygous locus are amplified in approximately equal proportions (Lygo et al. 1994; Gill et al. 1997). As fluorescent signals are known to show a linear correlation to DNA concentrations over a great range of intensities (Ziegle et al. 1992), any variation in the relationship between alternate peak-height ratios can therefore be attributed to slight differences between alleles in the relative efficiency of amplification. For example, very low levels of template DNA might cause stochastic variation in early cycles of PCR (Clayton et al. 1998), or somatic mosaicism (Rolf et al. 2002) can lead to preferential amplification of one allele. In our experiments, robust and repetition-consistent peak-height-ratio relationships were found for different amplifications performed on the same genotype. Amplification efficiencies of alternate alleles are relatively independent of template DNA concentration since we found that relative peak height of the different alleles follows a linear relationship as long as all samples are compared in a single experiment, making screening for genetic chimeras among regenerant clones a practical proposition. In general, however, a low-molecular-weight SSR allele will tend to amplify with greater efficiency than a high-molecular-weight allele (Clayton et al. 1998). It should therefore be remembered that strongly preferential amplification of one of a pair of alleles could occur if the two alternate alleles have markedly different sizes and that, in such instances, the peak ratios may exhibit greater variance. Nevertheless,

were identified as chimeric for marker MTcCIR15 compared with 7.75% of the regenerants identified as chimeric for marker MTcCIR10. LCT EEN 37/I, SIAL 93 and SC3 regenerants yielded a much lower rate of apparent genetic chimerism, with estimates from individual markers ranging between 0 and 15% depending

Table 3 Percentage of regenerants designated as exhibiting 'off-type' SSR peak-height ratios

Genotype	Putative chimeric samples for each SSR marker ^a (%)									Total ^b
	mTcCIR 1	mTcCIR 3	mTcCIR 6	mTcCIR 7	mTcCIR 8	mTcCIR 10	mTcCIR 15	mTcCIR 17	mTcCIR 61	
LCT EEN 37/I	7.7 (0)	–	0(0)	0 (7.7)	0 (0)	–	0 (7.7)	–	7.70(0)	31
LCT EEN 162/S-1010	–	–	–	–	–	7.8 (3.9)	17.8 (7.8)	–	–	31
SC 3	5 (15)	15 (0)	5 (10)	0 (10)	5 (10)	0 (5)	10 (0)	–	0 (5)	45
SIAL 93	–	10 (10)	0 (15)	–	10 (5)	5 (15)	5 (15)	15 (10)	10 (0)	55
Population percentage ^c	6.1 (9.1)	12.5 (5)	2 (9.4)	0 (9.1)	6 (6)	6.8 (5)	15.2 (7.4)	15 (10)	5.6 (1.9)	31

^aTwo values are indicated: the value outside the brackets represents the percentage of regenerants in which both replicates exhibited an off-type peak ratio; values given within the brackets represent the percentage of regenerants where variation between replicates exceeded the standard deviation of the entire population

^bCumulative totals (expressed as a percentage of the total population)

^cVariation between microsatellite loci expressed as the percentage of putative chimeras identified

the principle of comparing alternate allele amplification efficiency in order to draw inferences about template composition appears to be sound and has a precedence in forensic studies where the approach has been applied to detect and quantify the presence of mixed templates (Clayton et al. 1998). The forensic analysis of mixed samples is directly analogous to the effect of genetic chimerism studied here.

In our study the relationship between peak-height ratios and template composition was notably stronger than that between peak area and template composition. This is of interest since a previous report compared amplification of alternate alleles on the basis of peak area (Clayton et al. 1998). The lower correlations with peak area can partly be attributed to the influence of background signal (Clayton et al. 1998), which becomes more important as the concentration of the least represented allele diminishes. Similarly, the reliability of the peak-height ratios in predicting template mix composition also declined with reduced efficiency of amplification, although the decline was smaller and might be explained by sensitivity to disproportionate effects of minor pipetting or DNA quantification errors. We therefore elected to use peak-height ratios for the detection of chimeric somaclonal variants.

The slopes of the peak-height relationship generated by synthetic chimeras fell outside of that observed in the parental genotypes in a fashion that correlated with mix composition. This is supported by previous studies in which the relative amplification of heterozygotic alleles varied in proportion with the composition of the DNA template (Lygo et al. 1994; Sparkes et al. 1996). Our results clearly indicate that variation in peak-height ratio can be used to monitor allele loss or abundance of a novel allele created by slippage mutation. In the case of slippage, the simultaneous analysis of all three alleles created unwanted additional variance, and it was more reliable to assess loss of the mutated allele and gain of the new allele independently.

The total number of chimeric mutants provisionally detected by combining both analyses was 72, some 31%

of the total population. Interestingly, this is approximately the same frequency as apparently homogenous genetic mutants from the same regenerant population (data not shown) and demonstrates that chimeras can be abundant. Chimeras result from a combination of events, each of which has a small probability of occurrence (Franks et al. 2002). A somatic mutation must first occur in a cell, which is, or will become, part of a shoot meristem and the cell lineage then replicate to the point of being detectable among the tissue sampled. The probability of these events occurring in cocoa primary somatic embryos is apparently elevated by the enhanced rates of somatic mutation in plants regenerated from *in vitro* culture (Larkin and Scowcroft 1981) and by the multicellular origin of such embryos (Alemanno et al. 1996; Maximova et al. 2002).

The ability to identify putative genetic chimeras as described here should be followed by a more detailed study to fully characterise these plants. This would involve a broader sampling of tissues from mature regenerant plants to establish the extent and nature of the genetic chimerism. This would necessarily involve DNA extraction from plant organs and different cell layers. The comparison of such samples would provide valuable insight into the nature of genetic chimera (e.g. periclinal and sectorial chimeras) and may help indicate when the mutations occurred during somatic embryogenesis. The use of labelled primers allows multiplexing of many products to widen genome coverage. At the same time, the application of SSR markers provides scope to distinguish between the different forms of genetic change, including chromosome or chromosome arm loss (using several linked SSRs), point mutations leading to the failure of amplification (loss of one SSR allele) through to slippage mutation (leading to the creation of a new allele). Taken collectively, such information would allow improvements to be made to existing protocols that would minimise the incidence of somaclonal variation.

Acknowledgements We thank Cocoa Research UK for funding this study.

References

- Aderkas P von, Pattanavibool R, Hristoforoglu K, Ma Y (2003) Embryogenesis and genetic stability in long term megagametophyte-derived cultures of larch. *Plant Cell Tissue Organ Cult* 75:27–34
- Albani MC, Wilkinson MJ (1998) Inter-simple sequence repeat polymerase chain reaction for the detection of somaclonal variation. *Plant Breed* 117:573–575
- Alemanno L, Berthouly M, Michaux-Ferriere N (1996) Histology of somatic embryogenesis from floral tissues in cocoa. *Plant Cell Tissue Organ Cult* 46:187–194
- Cammarerri M, Errico A, Filippone E, Esposito S, Conicella C (2002) Induction of variability in chimeric *Aster cordifolius* ‘White Elegans’ through somaclonal variation. *Euphytica* 128:19–25
- Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci Int* 91:55–70
- Cuatrecasas J (1964) Cacao and its allies: a taxonomic review of the genus *Theobroma*. *Contrib U S Nat Herb* 35:379–614
- De Schepper S, Leus L, Mertens M, Debergh P, Van Bockstaele E, De Loose M (2001) Somatic polyploidy and its consequences for flower coloration and flower morphology in azalea. *Plant Cell Rep* 20:583–590
- Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soy bean. *Theor Appl Genet* 95:723–733
- Evans DA (1998) Somaclonal variation—genetic basis and breeding applications. *Trends Genet* 5:46–50
- Feuser S, Meler K, Daquinta M, Guerra MP, Nodari RO (2003) Genotypic fidelity of micropropagated pineapple (*Ananas comosus*) plantlets assessed by isozyme and RAPD markers. *Plant Cell Tissue Organ Cult* 72:221–227
- Fourré JL, Berger P, Niquet L, André P (1997) Somatic embryogenesis and somaclonal variation in Norway spruce: morphogenetic, cytogenetic and molecular approaches. *Theor Appl Genet* 94:159–169
- Franks T, Botta R, Thomas MR (2002) Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *Theor Appl Genet* 104:192–199
- Gill PD, Sparkes R, Kimpton CP (1997) Development of the guidelines to designate alleles using an STR multiplex system. *Forensic Sci Int* 89:185–197
- Gimenez C, De Garcia E, De Enrech NX, Blanca I (2001) Somaclonal variation in banana: cytogenetic and molecular characterization of the somaclonal variant CIEN BTA-03. *In Vitro Cell Dev Biol Plant* 37:217–222
- Harding K (2004) Genetic integrity of cryopreserved plant cells: a review. *Cryoletters* 1:3–22
- Jaligot E, Rival A, Beule T, Dussert S, Verdeil JL (2000) Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. *Plant Cell Rep* 19:684–690
- Joyce SM, Cassells AC (2002) Variation in potato microplant morphology *in vitro* and DNA methylation. *Plant Cell Tissue Organ Cult* 70:125–137
- Joyce SM, Cassells AC, Jain SM (2003) Stress and aberrant phenotypes in *in vitro* culture. *Plant Cell Tissue Organ Cult* 74:103–121
- Kaeppler SM, Kaeppler HF, Rhee Y (2000) Epigenetic aspects of somaclonal variation in plants. *Plant Mol Biol* 43:179–188
- Kennedy AJ, Mooleedhar V (1992) Conservation of cocoa in field genebanks—the International Cocoa Genebank, Trinidad. In: ICGT (ed) *Proc Int Workshop Conservation Characterization Utilization Cocoa Genet Resources in the 21st Century*. The University of The West Indies, St. Augustine, Trinidad, pp 21–26
- Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda P (1999) Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* 8:2141–2143
- Larkin PJ, Scowcroft WR (1981) Somaclonal variation: a novel source of variability from cell cultures for plant improvement. *Theor Appl Genet* 60:197–214
- Leroy XJ, Leon K, Hilyc JM, Chaumeil P, Branchard M (2001) Detection of *in vitro* culture-induced instability through inter-simple sequence repeat analysis. *Theor Appl Genet* 102:885–891
- Li Z, Traore A, Maximova S, Gupta PK (1998) Somatic embryogenesis and plant regeneration from floral explants of cacao (*Theobroma cacao* L.) using thidiazuron. *In Vitro Cell Dev Biol Plant* 34:293–299
- Lim WL, Loh CS (2003) Endopolyploidy in vanda miss Joaquim (Orchidaceae). *New Phytol* 159:279–287
- Lopez-Baez O, Bollon H, Eskes A (1993). Embryogénèse somatique de cacaoyer *Theobroma cacao* L. à partir de pièces florales. *C R Acad Sci Paris* 316:579–584
- Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, Kimpton CP, Gill P (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. *Int J Legal Med* 107:77–89
- Martin C, Gonzalez-Benito E, Iriando JM, Perez C (2001) Somaclonal variation in tissue cultures: implications in plant breeding and genetic conservation. *Agro Food Industry Hi Tech* 12:11–13
- Matthes M, Singh R, Cheah SC, Karp A (2001) Variation in oil palm (*Elaeis guineensis* Jacq.) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. *Theor Appl Genet* 102:971–979
- Maximova SN, Alemanno L, Young A, Ferriere N, Traore A, Guiltinan MJ (2002) Efficiency, genotypic variability, and cellular origin of primary and secondary somatic embryogenesis of *Theobroma cacao* L. *In Vitro Cell Dev Biol Plant* 38:252–259
- Mohan JS (2001) Tissue culture-derived variation in crop improvement. *Euphytica* 118:153–166
- Plader W, Malepszy S, Burza W, Rusinowski Z (1998) The relationship between the regeneration system and genetic variability in the cucumber (*Cucumis sativus* L.). *Euphytica* 103:9–15
- Rani V, Singh KP, Shiran B, Nandy S, Goel S, Devarumath RM, Sreenath HL, Raina SN (2000) Evidence for new nuclear and mitochondrial genome organizations among high-frequency somatic embryogenesis derived plants of allotetraploid *Coffea arabica* L. (Rubiaceae). *Plant Cell Rep* 19:1013–1020
- Risterucci AM, Grivet L, N’Goran JAK, Pieretti I, Flament MH, Lanaud C (2000) A high density linkage map of *Theobroma cacao* L. *Theor Appl Genet* 101:948–955
- Rolf B, Wiegand P, Brinkmann B (2002) Somatic mutations at STR loci—a reason for three-allele pattern and mosaicism. *Forensic Sci Int* 126:200–202
- Sparkes R, Kimpton CP, Gilbard S, Carne P, Anderson J, Oldroyd N, Thomas D, Urquhart A, Gill PD (1996) The validation of 7-locus multiplex STR test for use in forensic casework. (I) Mixtures, degradation and species studies. *Int J Legal Med* 109:186–194
- Traore A, Maximova SN, Guiltinan MJ (2003) Micropropagation of *Theobroma cacao* L. using somatic embryo-derived plants. *In Vitro Cell Dev Biol Plant* 39:332–337
- Tremblay L, Levasseur C, Tremblay FM (1999) Frequency of somaclonal variation in plants of black spruce (*Picea mariana*, Pinaceae) and white spruce (*P. glauca*, Pinaceae) derived from somatic embryogenesis and identification of some factors involved in genetic instability. *Am J Bot* 86:1373
- Ziegler JS, Su Y, Corcoran KP, Nie L, Mayrand PE, Hoff LB, McBride LJ, Kronick MN, Diehl SR (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* 14:1026–1031