

# Molecular markers for DNA-fingerprinting in cotton

*B. Bojinov<sup>1</sup> and J.M. Lacape<sup>2</sup>*

*<sup>1</sup> Department of Genetics and Plant Breeding, Agricultural University of Plovdiv, Plovdiv  
BULGARIA*

*<sup>2</sup> Programme Coton, CIRAD-CA, Montpellier FRANCE*

*Correspondence author [bojinov@writeme.com](mailto:bojinov@writeme.com); [bojinov@email.com](mailto:bojinov@email.com)*

## ABSTRACT

Molecular markers are long proposed as suitable for fingerprinting of crop species. With the development of the techniques the attention has shifted from protein to DNA-based markers as age-, tissue- and status-independent and, therefore, more reliable and reproducible. Several such markers (RFLP, RAPD, AFLP, SSR) were used in cotton for developing a high density genetic map for genomic studies and with the prospect for marker-assisted selection. The availability of these markers, together with the existing possibility for genetic transformation of the crop, make the development of DNA "fingerprints" for the varieties an issue of rising interest among breeders and seed companies. To the best of our knowledge no concentrated effort for developing such fingerprints has been made up to now in spite of the several reports proposing the potential usefulness of microsatellites and AFLPs in this respect. Here we present our work on comparing applicability of three types of molecular markers (microsatellites, AFLPs and SAMPLs) towards establishing a reliable, reproducible and relatively inexpensive protocol for DNA-fingerprinting for cotton. Furthermore, the intended transferability to the widest possible range of laboratories and, therefore, bringing it the closest to the breeder's fields, was attempted through the replacement of the standard radioactive DNA labeling procedure with the less hazardous "silver staining" protocol. The outcomes and the possible extensions of the present work are discussed.

## Introduction

Molecular markers are long proposed as suitable for fingerprinting of crop species. With the development of the techniques the attention has shifted from protein to DNA-based markers as age-, tissue- and status-independent and, therefore, more reliable and reproducible. Several such markers (RFLP, RAPD, AFLP, SSR) were used in cotton for developing a high density genetic map (Reinisch *et al.*, 1994; Shapley *et al.*, 1996; Shapley *et al.*, 1998; Brubaker *et al.*, 1999; Rungis *et al.*, 2002; Ulloa *et al.*, 2002), genetic diversity evaluation (Franco *et al.*, 2000; Abdalla *et al.*, 2001; Brubaker and Wendel, 2001; Kaur and Chahal, 2001; Rana and Bhat, 2002) and with the prospect for marker-assisted selection (Kohel *et al.*, 2001; Reddy *et al.*, 2001; Rungis *et al.*, 2002). The wide availability of these markers, together with the existing possibility for genetic transformation of the crop, make the development of DNA "fingerprints" for the varieties an issue of rising interest among breeders and seed companies.

To the best of our knowledge no concentrated effort for developing such fingerprints has been made up to now in spite of the several reports proposing the potential usefulness of microsatellites and AFLPs in this respect (Multani and Lyon, 1995; Wang *et al.*, 1997; Agrawal *et al.*, 1999; Song GuoLi *et al.*, 1999). No publication is available so far on the use of SAMPL markers in cotton. This technique is considered to have a high multiplex ratio like AFLP, and some degree of locus specificity like SSRs, which would make it the most efficient of all the molecular markers known so far (except of course SNPs, that have yet to be used in plant systems in any significant measure) (Roy *et al.*, 2002). Here we present a comparative study of the three marker systems (SSR, AFLP and SAMPL) with regard to their application as DNA-fingerprinting technique. Predominantly the genotypes from *G. hirsutum* and *G. barbadense* were considered, as these are the most widely used species.

## Experimental procedure

A working collection of 141 *G. hirsutum*, 22 *G. barbadense* and 19 other (representing cultivated and some wild diploid species) accessions from the CIRAD gene bank was chosen for this study. DNA from each accession was extracted from young fully expanded leaves according to the supplier recommendations of the commercially available DNA-extraction kit used (Qiagen Plant DNeasy Kit).

The SSR markers used were developed either at Brookhaven National Laboratory (further referred to as "BNL") or CIRAD (further referred to as "CIR"). The BNL microsatellites are all derived from *G. hirsutum* cv. Acala Maxxa genomic libraries enriched for GA and CT inserts according to Ostrander *et al.* (1992). Primer sequences used are available at <http://demeter.bio.bnl.gov/acecot.html>. CIR microsatellites are derived from genomic libraries enriched for CA repeats as described in Lacape *et al.* (2003). SSR analysis conditions were as described in Risterucci *et al.* (2000).

AFLP analysis was performed using the Life Technology AFLP™ analysis system I (Gibco BRL, Gathersburg, Md., USA) using the two step amplification as described by Vos *et al.* (1995). The protocol was as recommended by the supplier with minor modifications as detailed in Risterucci *et al.* (2000). Preliminary screening for determination of the most suitable primer couples for fingerprinting was performed using 22 *G. hirsutum* accessions and the *G. barbadense* variety VH8 as an outlier. Eight *EcoRI*+3 and 8 *MseI*+3 primers were used in total in the selective amplification, making 64 *EcoRI*/*MseI* combinations. After determination of the primer couples producing the highest number of polymorphic bands, they were further tested on the entire working collection.

SAMPL is a modified AFLP technique in which adapter-ligated restriction fragments of a conventional

AFLP are used, but the final PCR amplification is achieved using primers which differ from those employed in AFLP. One of the primers is chosen to anneal at fixed point of a SSR motive, and the other is an AFLP primer that is designed on the basis of the sequence of the synthetic adapter for the MseI restriction site, and carries 2–3 selective nucleotides (Witsenboer *et al.*, 1997). SAMPL analyses were based on the AFLP technology detailed above, with one modification - SAMPL primers, labeled with  $\gamma$ -<sup>32</sup>P-ATP, were used in place of the selective EcoRI-adaptor annealing AFLP primers in the final amplification step. A total of 10 SAMPL primers (Table 1) were tested for their ability to produce polymorphic bands, each in combination with the 8 MseI+3 primers from the same AFLP kit as above, thus resulting in a total of 80 primer couple combinations. These were tested on the same small group of 23 accessions as in the initial testing of the AFLPs to which 7 Bulgarian genotypes (all from *G. hirsutum*) were added.

## Results

Some preliminary data (Lacape *et al.*, unpublished) showed, that out of 28 SSRs only one had 10 alleles, and one had six alleles in a tested *G. hirsutum* population. All other microsatellite markers had 4 and fewer alleles. This indicated that a possibility to identify sufficiently polymorphic SSRs exists (for the purposes of present study, we would consider as “highly polymorphic” the SSRs with six and more alleles relatively evenly distributed in *G. hirsutum*), but the cases of such microsatellites would be quite rare. Our attempts to identify more such markers failed, as all nine additional SSRs we tested on 141 *G. hirsutum* genotypes had four or less alleles, often with quite skewed distribution (Figure 1). Testing the same SSRs in the *G. barbadense* background, showed that in spite of the significantly smaller number of the *G. barbadense* accessions used (22 in total), the overall number of observed alleles per SSR marker was not inferior (Table 2). The statistical analysis of the data showed that the polymorphism in *G. barbadense* is significantly higher as revealed by this type of markers (data not shown).

As expected, the AFLPs revealed significantly higher number of polymorphisms as compared to SSRs. The polymorphic bands observed per reaction varied to a great extent between the primer couples (data not shown). This allowed for the selection of three primer couples to be tested on the entire working collection of 182 accessions. Running these primer couples on the 141 *G. hirsutum* accessions allowed to identify not only markers that are rather evenly distributed, but also the ones that are specific to just one (or a very few) genotypes as well (Figure 2). Such markers are highly informative in achieving the intended fingerprinting of particular genotypes. Once again, the accessions of *G. barbadense* origin showed higher polymorphism than those belonging to *G. hirsutum*.

As expected from the technique design, the SAMPL markers showed intermediate polymorphism to that of the AFLPs and SSRs (Figure 3). Similarly to the AFLPs, SAMPLs revealed a number of polymorphisms, relatively equally distributed between the *G. hirsutum* accessions as well as some unique bands, present in single accessions of the set. Unfortunately, none of the 80 primer couples tested showed sufficiently high number of polymorphisms to be compatible to AFLPs, or the typical co-localization (for at least some) of the bands as usually observed in a SSR reaction (data not shown).

## Discussion

The results from testing 37 microsatellite markers demonstrated that for the moment they should not be advised for use for fingerprinting, especially if large numbers of accessions have to be identified. Having revealed in too many cases too few polymorphisms within the *G. hirsutum* genotypes, high number of these markers will be needed to resolve any collection of such genotypes that is of reasonable size for breeding for example. In the present study, where the accessions tested were chosen with the intention to represent the broadest spectrum of genotypes from around the world, the use of almost entire set of tested SSRs (excluding the monomorphic ones in this background) was needed to resolve 141 *G. hirsutum* accessions. This means that the use of SSR markers is impractical at this stage and will stay such unless more highly polymorphic (with more than 6 evenly distributed in the studied collection alleles) SSRs are identified. Due to a much higher polymorphism these markers could still be considered for use in *G. barbadense*, especially when silver staining would come out as the only revealing technique of choice, due to practical reasons discussed in more detail below.

The fact that in the initial screening of the 64 AFLP primer couple combinations it appeared possible to resolve the 23 accessions used at that stage with just one primer couple (data not shown), demonstrated the power of this technique for use in DNA fingerprinting. The preliminary evaluation of the ability of the AFLP primer couples to reveal available polymorphism allowed to significantly reduce the number of tests to be run on the larger set of working collection accessions. It is the combination of a significant number of evenly distributed through the *G. hirsutum* accessions markers together with the presence of unique (or very rare) ones (Figure 2), revealed in the same reaction, that made the AFLPs our technique of choice for achieving fingerprinting of the genotypes with a minimal number of PCR reactions. The theoretical expectations (if a random appearance of the observed markers would have been presumed) were pointing to the possibility to resolve our working collection of *G. hirsutum* accessions with just one primer pair. However the well established genetic bottleneck in the origin of the modern upland cotton varieties (Small *et al.*, 1999; Iqbal *et al.*, 2001)

did not allow for such assumption. Therefore three of the primer couples that revealed the highest number of polymorphisms in the preliminary tests were run on the working collection. After analyzing DNA patterns obtained we have found that the entire set of 141 *G. hirsutum* accessions could be resolved with just two primer couple combinations. This means that it is possible to fingerprint any of these accessions with using just two of the 64 primer couples (from a commercially available AFLP kit), therefore significantly reducing the labor and consumables needed for this purpose. With the higher polymorphism rate observed in *G. barbadense* the same resolution might be possible for similar number of accessions with a single primer couple, but our present set of just 22 accessions from this species does not allow us to be more affirmative to that point. Even if that is not the case, we believe the results of the present study indicate that developing of DNA "ID cards" for the cotton varieties would be relatively quickly and inexpensively achieved with the use of AFLPs, if such is to become a standard practice in the future.

Attempts to use SAMPL technique did not result in the satisfactory levels of revealed polymorphisms in the tested set of 30 accessions (Figure 3). This of course is true only in comparison with the results obtained using AFLPs. Our expectations that through using this technique combining of the high number of observed polymorphisms with the ability to identify increased number of co-dominant markers will become possible were not confirmed in spite of rather large number of primer combinations used. As demonstrated in Table 1, we tried to use 3 types of SAMPL primers. These were anchored either on the compound motive (as described originally in Morgante and Vogel, 1994 – primers SAMPL1 to SAMPL6), on the 5' end of the repeat sequence (as described originally in Vivek and Simon, 1999 – primers SAMPL7 and SAMPL8) or on the 3' end of the repeat (primers SAMPL9 and SAMPL10). Unfortunately none of these primer designs resulted in the presence of co-dominant markers in the tested group of 30 genotypes. As the number of polymorphisms revealed per primer couple in the present study was inferior to that of AFLP technique we did not consider SAMPL primers for testing on the entire working collection as it seemed obvious that more PCR reactions would be needed to achieve the same level of resolution as with AFLPs.

For the purposes of making DNA fingerprinting accessible to as wide range of laboratories as possible a comparison of the radioactive labeling and the "silver staining" of the PCR products from the same reaction was made. It demonstrated that a number of fainter (as revealed by radioactive labeling) bands are lost when the second approach is used due to its lower sensitivity. The resolution was even further decreased when attempts were made to photograph the silver stained gels for long term preservation of the obtained patterns. Altogether that results in significant decrease of

the number of scorable bands from the stored images of the gels. Depending on the relatedness and size of the genotype collection to be fingerprinted this could result in the need to significantly increase the number of PCR reactions, therefore increasing the cost per "fingerprint". Our second concern with this revealing technique is that only processed images of the original gels can be stored long-term, as keeping the stained gels is impractical. However this would mean losing of the raw data from the assays, which might appear as crucial difference between this and radioactive revealing of that data in some cases (for example when variety identification will be needed for breeder's rights protection). Of course where the laboratories do not have the authorization to work with radioactive material, silver staining will still be the only technique of choice. As discussed above, it might as well be equally sensitive to the use of the radioactivity, especially in the cases where the use of SSRs will be justified. Studying a small number of genotypes, groups with higher diversity (as seems to be the case for *G. barbadense* collections), or identification of additional highly polymorphic SSRs (with more than 6 alleles in the studied group) would be just a few examples when it might prove fruitful.

Based on our experience with these three marker types and two band revealing techniques at present we would recommend the use of radiolabelled AFLPs where large numbers of accessions are to be fingerprinted and all the needed authorizations are obtained.

## Acknowledgements

The current research was supported by one year EU "Marie-Curie" Fellowship to Bojin Bojinov. Valuable assistance in the choice of SAMPL markers from B. Burr from the Brookhaven National Laboratory, USA and by B. Hau from CIRAD, France for the selection of the genotypes for the working collection is highly appreciated.

## References

- Abdalla, A.M., Reddy, O.U.K., El Zik, K.M. and Pepper, A.E. (2001). Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theoretical and Applied Genetics*, **102**: 222-229.
- Agrawal, D.C., Saha, S., Jenkins, J.N., Zipf, A. and May, L. (1999). Use of cross-species simple sequence repeat (SSR) primers for developing polymorphic DNA markers. *Journal of New Seeds*, **1**: 25-37.
- Brubaker, C.L. and Wendel, J.F. (2001) RFLP diversity in cotton. *Genetic Improvement of Cotton*, Science Publishers Inc., 81-102.
- Brubaker, C.L., Paterson, A.H. and Wendel, J.F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome*, **42**: 184-203.
- Franco, E.A., Cabaltea, M.T., Abrina, Y.L., Barroga, M.E., Pascua, L.T., Damo, M.C., Rinen, E.C., Cosico, V.B. and Orpian E.D., Jr. (2000). Current status and

- prospects of cotton biodiversity in the Philippines. *Philippine Journal of Crop Science*, **25**: 61.
- Iqbal, M.J., Reddy, O.U.K., El-Zik, K.M., and Pepper, A.E. (2001). A genetic bottleneck in the 'evolution under domestication' of Upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics*, **103**: 547-554.
  - Kaur, A.P. and Chahal, G.S. (2001). Genetic diversity at molecular and morphological levels in Upland cotton (*Gossypium hirsutum* L.). *Crop Improvement*, **28**: 26-31.
  - Kohel, R.J., Yu, J., Park, Y.H. and Lazo, G.R. (2001). Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica*, **121**: 163-172.
  - Lacape, J.M., Nguyen, T.B., Courtois, B., Bojinov, B. and Hau, B. (2003). Genome mapping of tetraploid cotton: towards a saturated and unified map. This volume.
  - Morgante, M. and Vogel, J. (1994). Compound microsatellite primers for the detection of genetic polymorphisms. U S patent application no. 08/326456.
  - Multani, D.S. and Lyon, B.R. (1995). Genetic fingerprinting of Australian cotton cultivars with RAPD markers. *Genome*, **38**: 1005-1008.
  - Ostrander, E.A., Jong, P.M., Rine, J. and Duyk, G. (1992). Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences of the USA*, **89**: 3419-3423.
  - Rana, M.K. and Bhat, K.V. (2002). Genetic diversity analysis in Indian diploid cotton (*Gossypium* spp.) using RAPD markers. *Indian Journal of Genetics and Plant Breeding*, **62**: 11-14.
  - Reddy, O.U.K., Pepper, A.E., Abdurakhmonov, I., Saha, S., Jenkins, J.N., Brooks, T., Bolek, Y. and El-Zik, K.M. (2001). New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *Journal of Cotton Science*, **5**: 103-113.
  - Reinisch, A.J., Dong, J.M., Brubaker, C.L., Stelly, D.M., Wendel, J.F. and Paterson, A.H. (1994). A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. *Genetics*, **138**: 829-847.
  - Risterucci, A.M., Grivet, L., N'Goran, J.A.K., Pieretti, I., Flament, M.H. and Lanaud, C. (2000). A high density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, **101**: 948-955.
  - Roy, J. K., Balyan, H.S., Prasad, M. and Gupta, P.K. (2002). Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat. *Theoretical and Applied Genetics*, **104**: 465-472.
  - Rungis, D., Llewellyn, D., Dennis, E.S. and Lyon, B.R. (2002). Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. *Australian Journal of Agricultural Research*, **53**: 551-560.
  - Shappley, Z.W., Jenkins, J.N., Meredith, W.R. and McCarty, J.C. (1998). An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. *Theoretical and Applied Genetics*, **97**: 756-761.
  - Shappley, Z.W., Jenkins, J.N., Watson, C.E. Jr., Kahler, A.L. and Meredith, W.R. (1996). Establishment of molecular markers and linkage groups in two F<sub>2</sub> populations of Upland cotton. *Theoretical and Applied Genetics*, **92**: 915-919.
  - Small, R.L., Ryburn, J.A. and Wende, J.F. (1999). Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). *Molecular Biology and Evolution*, **16**: 491-501.
  - Song Guoli, Zhang ChunQing, Jia JiZeng, Wang KunBo and Cui RongXia (1999). Cotton AFLP analysis with silver-staining and preliminary report of variety fingerprinting based on it. *Acta Gossypii Sinica*, **11**: 281-283.
  - Ulloa, M., Meredith, W.R. Jr., Shappley, Z.W. and Kahler, A.L. (2002). RFLP genetic linkage maps from four F<sub>2.3</sub> populations and a joinmap of *Gossypium hirsutum* L. *Theoretical and Applied Genetics*, **104**: 200-208.
  - Vivek, B.S. and Simon, P.W. (1999). Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). *Theoretical and Applied Genetics*, **99**: 58 - 64.
  - Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van d. Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**: 4407-4414.
  - Wang, X., Guo, W., Zhang, T., Pan, J., Wang, X.Y., Guo, W.Z., Zhang, T.Z., and Pan, J.J. (1997). Analysis of RAPD fingerprinting on short-seasoned cotton cultivars in China. *Acta Agronomica Sinica*, **23**: 669-676.
  - Witsenboer, H., Vogel, J. and Michelmore, R.W. (1997). Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). *Genome*, **40**: 923-936.

**Table 1.** Sequences of adaptors and primers used in the SAMPL reactions<sup>1</sup>.

Adapters / Primers	Sequence
Adapters	
EcoRI adaptor	5'-CTCGTAGACTGCGTACC CATCTGACGCATGGTTAA-5'
MseI adaptor	5'-GACGATGAGTCCTGAG TACTCAGGACTCAT-5'
Selective amplification primers	
MseI primers	
MseI+3 primer 1	5'-GATGAGTCCTGAGTAACAA-3'
MseI+3 primer 2	5'-GATGAGTCCTGAGTAACAC-3'
MseI+3 primer 3	5'-GATGAGTCCTGAGTAACAG-3'
MseI+3 primer 4	5'-GATGAGTCCTGAGTAACAT-3'
MseI+3 primer 5	5'-GATGAGTCCTGAGTAACTA-3'
MseI+3 primer 6	5'-GATGAGTCCTGAGTAACTC-3'
MseI+3 primer 7	5'-GATGAGTCCTGAGTAACTG-3'
MseI+3 primer 8	5'-GATGAGTCCTGAGTAACTT-3'
SAMPL primers <sup>2</sup>	
SAMPL-1	5'-TGTGTGTGTGTGTGTATATA-3'
SAMPL-2	5'-ACACACACACACACATATAT-3'
SAMPL-3	5'-TCTCTCTCTCACACACAC-3'
SAMPL-4	5'-AGAGAGAGAGTGTGTGTG-3'
SAMPL-5	5'-ACACACACACTCTCTCTC-3'
SAMPL-6	5'-ACACACACAGAGAGAGAG-3'
SAMPL-7	5'-ATGCACACACACACACACA-3'
SAMPL-8	5'-CAGCTCTCTCTCTCTCTCT-3'
SAMPL-9	5'-CACACACACACACACAGTA-3'
SAMPL-10	5'-CTCTCTCTCTCTCTGAC-3'

<sup>1</sup> Sequences of the primers for the SSR and AFLP reactions are not given as they are available elsewhere – see text.

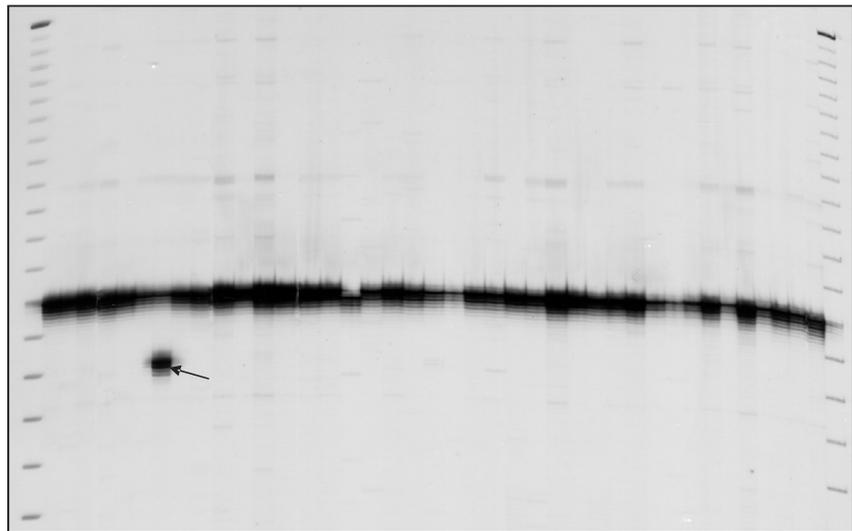
<sup>2</sup> Primer anchoring nucleotides are in bold.

**Table 2.** Range of observed alleles of the SSR markers in the two most common commercial cotton species. The number of tested accessions is 141 for *G. hirsutum* and 22 for *G. barbadense*.

No. of observed alleles	% of total	
	<i>G. hirsutum</i>	<i>G. barbadense</i>
1	7.14	3.33
2	28.57	26.66
3	35.71	36.66
4	10.71	16.66
5	10.71	6.66
6	3.57	0
7	0	3.33
8	0	3.33
9	0	0
10	3.57	0
11	0	3.33

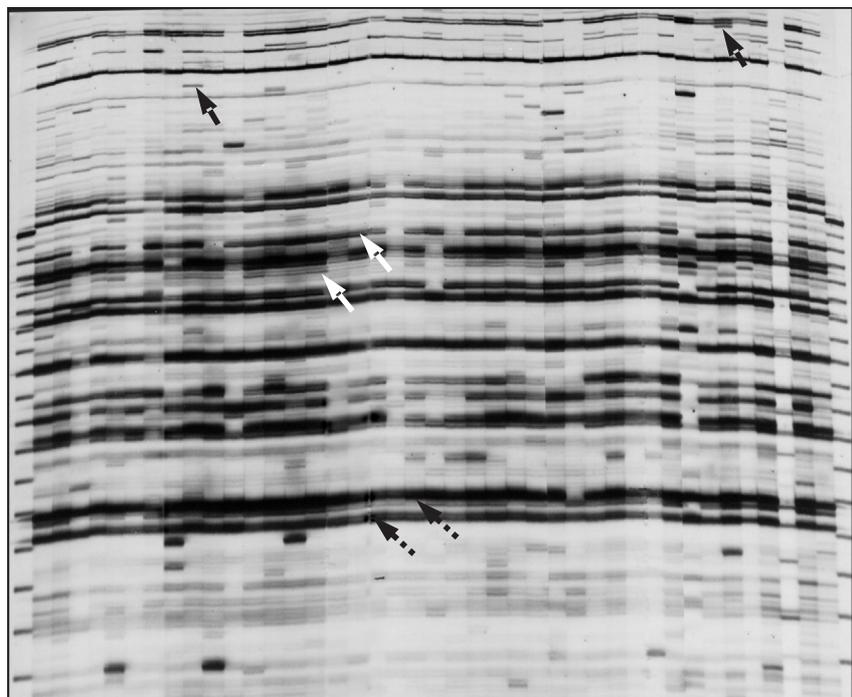
**Figure 1.**

Example of skewed microsatellite allele distribution. A single polymorphic band (arrow) is observed in a subset of *G. hirsutum* accessions from the working collection.



**Figure 2.**

An example of AFLP marker distribution in a subset of *G. hirsutum* accessions from the working collection. Different types of markers are clearly distinguishable: monomorphic (dotted black arrows), polymorphic (solid white arrows), and unique (solid black arrows).



**Figure 3.**

An example of SAMPL marker distribution in the tested subset of 29 *G. hirsutum* accessions. Lane 12 is *G. barbadense* variety VH8, lanes 14 to 30 represent two groups of presumably closely related genotypes (lanes 14-23 are varieties and lines of Brazilian origin; lanes 24-30 represent seven varieties and lines of Bulgarian origin).

