E

BMT/6/12
ORIGINAL: English
DATE: February 21, 2000

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA-PROFILING IN PARTICULAR

Sixth Session
Angers, France, March 1 to 3, 2000

STANDARDISATION OF MOLECULAR MARKER SYSTEMS FOR VARIETY TESTING

prepared by experts from the Netherlands, the United Kingdom, Germany and France

Ben Vosman[1*], Robert Cooke[2], Martin Ganal[3], Roger Peeters[4], Peter Isaac[5] , Marion Röder[3], Jenny Jackson[2], Sarah Rendell[2], Marcel Dijcks[4] , Yolanda Kleyn [1], Dirk Visser[1] , Katja Wendehake[3], Tatyana Areshchenkova[3], Victor Korzun[3], Marc Amelaine[5], Vincent Wickaert [5] and Gerard Bredemeijer[1*]

[1]Plant Research International (*authors for correspondence: e-mail B.Vosman@plant.wag-ur.nl or G.M.M.Bredemeijer@plant.wag-ur.nl
[2]National Institute of Agricultural Botany (NIAB)
[3]Institute for Plant Genetics and Crop Plant Research (IPK)
[4]Nunhems Zaden b.v. (Nunhems)
[5]Agrogene SA (Agrogene)

***** This document will be published in the Acta Horticulturae
Key words: Microsatellite, STMS markers, tomato, wheat, standardisation, database, network

Abstract

The present study is part of an EU project that aims to demonstrate the technical viability of STMS markers for variety identification. As examples two important European crop species, tomato and wheat were chosen. Initially, about 30-40 STMS markers were used to identify a set of 20 good markers per crop and to standardise the methodology and the interpretation of the results in different laboratories. Several systems were used for the detection of STMS polymorphisms.

The selected STMS markers are being tested on 500 varieties of each species and databases are being constructed. The first comparisons of data generated by the different laboratories revealed a high degree of agreement. The causes of discrepancies between duplicate samples analysed in different laboratories and precautions to prevent them, are discussed.

1. Introduction

Recent developments in molecular biology have resulted in novel techniques of DNA profiling that can be used for the characterisation of plant material. Molecular markers have many advantages for plant variety identification over the more traditionally used morphological and biochemical markers because of their independence from environmental influences, high level of polymorphism, and their almost unlimited availability. Molecular techniques are also likely to be extremely discriminating and much more rapid. There are several approaches to DNA profiling. The Sequence Tagged Microsatellite Site (STMS) approach yields highly informative and discriminative markers, is suitable for automation and the results obtained can easily be stored in an electronic database, which facilitates comparison of results and collaboration between laboratories.

Two years ago, the European Union has agreed to fund a Biotechnology Demonstration project entitled ‘Molecular markers for variety testing’. This multi-national project aims to demonstrate to potential users the many advantages of STMS markers for variety testing. The technical viability of the STMS approach for variety identification and discrimination will be demonstrated in two important European crop species: tomato (*Lycopersicon esculentum*) and wheat (*Triticum aestivum*). Databases containing the molecular description of the most common varieties of tomato and wheat grown during the last 10 years in Europe will be constructed and tested. It will be demonstrated that the markers and databases can be utilised for a range of applications, including all aspects of variety identification, quality control and genetic diversity measurements.

Utility of the technique in wheat and tomatoes has been reported by Plaschke *et al*. (1995) and Smulders *et al*. (1997) respectively. Initially, a set of approximately 30-40 STMS primer pairs was selected for each crop. These STMS markers have been previously shown to reveal polymorphisms within small collections of wheat varieties (Plaschke *et al*. 1995; Röder *et al*. 1995 and 1998) and tomato varieties (Smulders *et al*. 1997; Bredemeijer *et al*. 1998; Areshchenkova and Ganal 1999). As several systems for the detection of STMS polymorphism were used it was important to standardise both the methodology (including sampling of material, DNA extraction and estimation, PCR conditions) and interpretation of the results. The detection systems differ with respect to their mode of action, the costs involved in setting up the detection system and the type of laboratory facilities needed. Instead of choosing one system to be used by all participants within this project, the participants utilise the STMS detection system that was already in use in their laboratory.

Automated DNA sequencers are used by CPRO (ALF express), IPK (ALF and ALFexpress) and Nunhems (ABI). NIAB uses IR-labelled primers in combination with a LI-COR DNA Analyser 4200 whilst Agrogene uses a combined fluoro-phosphoimager in conjunction with fluorescently labelled primers. In the present paper, standardisation of methodology and the first results regarding the construction of the databases are presented.

2. Material and methods

2.1 Plant material

Approximately 500 varieties of each crop were assembled from the variety lists of all European countries. For this, national agencies and breeding companies have provided most of the seeds. Four or six individuals were sampled as a bulk to represent each genotype. The wheat genotypes were sampled directly from seeds, and the tomato genotypes from seedlings. Each partner analysed an agreed number of varieties (tomato: CPRO 400, Agrogene 300, IPK 200, NIAB 50, and Nunhems 50; wheat: IPK 500, Agrogene 400, NIAB 50, CPRO 100).

2.2 DNA extraction

Tomato DNA was extracted from seedlings essentially as described by Fulton *et al*. (1995) with some slight modifications (chloroform-isoamyl mixture was replaced by chloroform). Wheat DNA was extracted from seeds according to Plaschke *et al*. (1995).

2.3 PCR

Standard set of PCR conditions (reaction volume 25Ïl, 0.2 ÏM of each primer, 0.25 mM of each dNTP, 2.5 mM $MgCl_2$, 1 unit of AmpliTaq Gold (Perkin Elmer) and 5 Ïl wheat DNA or tomato DNA) plus specific conditions for each primer pair (annealing temperature, number of cycles, multiplexing). Standard cycling conditions were: 45 cycles of 94°C for 1 min, 50, 55, or 60°C for 1 min and 72°C for 1 min. After the 45 cycles, one cycle of 72°C for 5 min was added.

2.4 Detection of STMS

For detection of the PCR products 3 participants used fluorescently labelled primers in conjunction with an automated DNA sequencer (CPRO: ALFexpress, Pharmacia; IPK: ALF- and ALFexpress sequencer, Pharmacia; Nunhems: ABI prism 377, Perkin Elmer). Agrogene used end labelling of one of each pair of PCR primers with fluorescein in combination with conventional sequencing gels and a Molecular Dynamics Storm 860 imager. NIAB used IR-labelled primers in combination with a LI-COR DNA analyser 4200 (MWG).

3. Results

3.1 Standardisation of detection systems

For each crop species a minimum of 20 primer pairs had to be selected from existing collections of CPRO and IPK in a way that the microsatellite technologies used in different laboratories were able to identify alleles in the same way. Two standard sets, of 8 varieties each, were analysed with the 30-40 microsatellites and the best primer pairs were selected. Typical examples of banding/peak patterns generated by different detection techniques are shown in Figure 1. Scoring matrices were constructed for each variety and marker.

Comparison of the data produced by different partners resulted in a preliminary selection of 27 wheat and 23 tomato markers. The criteria for choosing a marker included scorability of the patterns, map position, reproducibility of scoring between laboratories, and the level of polymorphism detected between varieties.

Subsequently, an inter-laboratory ring test was conducted to test the methodology and interpretation of the results. All participants analysed the same collection of 16 varieties of both species with the selected markers. In spite of the use of different technologies allele scoring by the partners resulted mostly in the same allele classification for the 16 varieties. In general, data from the wheat samples were relatively easy to compare between groups. In some cases the polymerase added an extra A to the fragment resulting in n and n+1 peaks. To avoid this pigtail primers were used.

Comparing data sets for tomato was more difficult due to the occurrence of heterozygotes in hybrids, often in combination with differences in the amount of amplification product for the two alleles. The latter phenomenon may also be due to heterogeneity of the seed samples used or to differential amplification of the alleles. To distinguish between these two possibilities DNA from individual plants have been tested in some cases when the partners scored differently (see examples in Fig. 2).

3.2 Construction of the STMS databases

For the construction of the STMS databases of wheat and tomato 500 varieties were collected for each crop. These varieties were analysed in duplicate (at 2 laboratories) using the final selection of markers shown in Tables 1 and 2. At the moment, most analyses have been finished and for part of the data the allele scores were inserted in a database.

In general, the patterns generated in the different laboratories using the different technologies were well scorable. An example of a multiplex pattern of a series of wheat varieties is shown in Figure 3. Although allele designation for the duplicate samples was in most cases the same, several discrepancies were observed. The number of discrepancies were strongly dependent on the marker. For example for three tomato loci, the number of discrepancies between duplicate samples was 10 for LELEUZIP, 21 for TMS1 and 32 for LE21085. For each of the 3 loci, 6 of these discrepancies were due to lack of germination at one of the two laboratories. In the worst case (LE21085), 24 varieties (5%) of the duplicate samples were scored differently: i.e in one laboratory as a homozygote , and in the other as a heterozygote or scored as different alleles. In the best case (LELEUZIP) only 4 discrepancies were detected (0.8%).

The discrepancies found between duplicate samples were analysed by rescoring the peak and banding patternsrevealing that they were caused by:

1. Methodological problems

- differences in resolution capacity of individual gel systems. The use of short gels often resulted in incomplete separation of the fragments in hybrids.
- echo bands were observed in the phospho-imagersystem. This complicated scoring of some alleles.
- differences in the setting of thresholds for the definition of allelic peaks caused differences between labs

- missing data, caused by the absence of PCR products
- mis scoring caused by overloading and underloading of gels

2. Data entry errors (mistyping, misscoring). In wheat some markers showed high numbers of alleles, making correct scoring difficult when not all reference alleles were on the same gel

3. Heterogeneity of samples (residual or hybrid heterozygosity).

4. Discussion

The overall objective of the project is to demonstrate the utility of STMS markers for variety identification in tomato and wheat. The present paper deals with the selection of primer pairs, the standardisation of methodology and the first results of genotyping a large number of modern varieties for the construction of STMS databases.

Ideally, markers are freely available, highly polymorphic, mapped, evenly distributed over the genome, suitable for multiplexing, and easy and reproducibly to score in different laboratories. The present study showed that only a few of the available markers fulfilled all these criteria especially for those used in tomato. A number of markers were not suitable for further use because of low levels of polymorphism, the generation of complicated patterns (eg. high stutters, nonspecific peaks) or non-reproducible results.

In spite of the use of different detection systems, allele scoring by the partners resulted in most cases in a reproducible classification of the tested varieties. In general, wheat samples had a low number of heterozygotes but a higher number of alleles. In a few cases, scoring of the wheat patterns was complicated by additional low peaks that might represent amplification products of a locus on one of the two other genomes. Comparing data sets from tomato was sometimes difficult due to the occurrence of many heterozygotes in hybrids often in combination with differences in the amount of amplification product for the two alleles.

Strategies for dealing with the difficulties have been agreed on between the partners including the use of specific selection criteria for some markers (eg. minimum or maximum size of the fragments), the use of ‘pig tail’ primers for some loci to circumvent the problem of extra base additions (Brownstein *et al*. 1996), testing of individuals to distinguish between heterogeneity and differential amplification, setting thresholds above which a peak should be scored, and scoring heterozygote alleles on short gels.

The differences in scoring markers between laboratories strongly depended on the marker used. Some markers gave no problems at all while others, like LE21085, had a considerable number of problematic cases that could not be resolved until now. Once again stressing the need for high quality markers. Also the equipment used had a strong effect on the scoring quality (e.g. the length of the separation gel).

In the near future, the remaining experiments for the construction of the databases will be finished. Subsequently, the patterns will be analysed and the scoring data will be compared by the partners. After performing a number of replicate experiments and testing of individuals a consensus database will be constructed. Finally, a series of ‘blind’ tests will be carried out in which 8 unlabelled samples of each crop will be identified by the partners using their own

profiling system and the databases. This test is important because it has to be shown that the methodology and databases could be used successfully by everybody.

Acknowledgements

The authors acknowledge the financial support of the European Commission DGXII Biotechnology grant BIO4-CT97-2377.
We also thank the breeding companies for providing the tomato and wheat seeds.

References

Areshchenkova, T. and Ganal, M.W., 1999. Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42: 536-544.

Bredemeijer, G.M.M., Arens, P., Wouters, D., Visser, D. and Vosman, B., 1998. The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor.Appl. Genet. 97: 584-590.

Broun, P. and Tanksley, S.D., 1996. Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol. Gen. Genet. 250: 39-49.

Brownstein, M. J., Carpten, J. D. and Smith, J.R., 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004-1010.

Fulton, T.M., Chunwongse, J. and Tanksley, S.D., 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol. Biol. Reporter 13: 207-209.

Phillips, W.J., Chapman, C.G.D. and Jack, P.L., 1994. Molecular cloning and analysis of one member of a polymorphic family of GACA-hybridising DNA repeats in tomato. Theor. Appl. Genet. 88: 845-851.

Plaschke, J., Ganal, M. W. and Röder, M. S., 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91: 1001-1007.

Röder, M. S., Plaschke, J., König, S. U., Börner, A., Sorrells, M. E., Tanksley, S. D. and Ganal, M. W., 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol. Gen. Genet. 246: 327-333.

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P. and Ganal, M.W., 1998. A microsatellite map of the wheat genome. Genetics 149: 2007-2023.

Smulders, M. J. M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P. and Vosman, B., 1997. Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor. Appl. Genet. 94: 264-272.

Table 1 Characteristics of the tomato microsatellites selected for the construction of the database. Data are based on fragments detected with an ALFexpress DNA sequencer (CPRO).

Marker	Repeat type	product sizes (bp)	chromosomal location	number of alleles[f]
TMS9[a]	(GATA)26 imperfect	337-354	12	4
LE20592[b]	(TAT)15-1(TGT)4	158-167	11	4
LEE6[c]	(GTT)28-3	201-207	1	2
LEMDDNa[b]	(TA)9	204-221	5	3
TMS34	(GA)19	180-205	9	4
LED4[c]	(TCT)32-1	150-188	10	4
LED10[c]	(TCT)29-2	197-307	6	3
LE21085[b]	(TA)2(TAT)9-1	98-113	4	3
LELEUZIP[b]	(AGG)6-1TT(GAT)7	96-98	8	2
TMS1	(GT)n	130-132	2	2
ATTa[d]	(TTA)5CT(ATT)8...	218-221	3	2
LEE102[c]	(GTT)88 imperfect	283-307	12	3
LELE25[b]	(TA)11	211-217	10	4
TMS33[a]	(GA)26 imperfect	268-276	12	3
LED112A[c]	(GAA)32-2	282-328	8	4
LEWIPIG[b]	(CT)4(AT)4	255-263	9	2
LESATTAGA[b]	(TA)11(GA)11	167-171	?	3
JACKP1[e]	(GATA)n,(GACA)n	371-389	11	3
TMS22[a]	(GT)9(AT)8(AC)13(GA)12 imperfect	152-156	4	2[g]
LED1A[c]	(TCT)21TCCTTCC(TCT)6	145-169	10	3
LEH228[c]	(TGT)n	150-156	?	3

[a] Areshchenkova and Ganal (1998)
[b] Smulders et al. (1997)
[c] STMS isolated by Arens, P. (CPRO)
[d] Broun and Tanksley (1996)
[e] Phillips et al. (1994)
[f] Number of alleles found in the 8 standard and 14 ring test varieties
[g] Alleles of the locus generating short fragments

Table 2. Characteristics of the wheat microsatellites selected for the construction of the database.

Locus	Repeat type	Product sizes (bp)	Chromosome	Number of Alleles in ringtest
Secalin	-	100	1R	2
Taglgap	(CAA)15	209-281	1B	5
WMS003	(CA)18	75-83	3D	3
WMS018	(CA)17GA(TA)4	178-194	1B	5
WMS046	(GA)2GC(GA)33	145-183	7B	6
WMS095	(AC)16	115-134	2A	5
WMS155	(CT)19	132-153	3A	4
WMS160	(GA)21	171-186	4A	4
WMS165	(GA)20	185-191,193-205	4A,4D	2,5
WMS190	(CT)22	198-214	5D	4
WMS261	(CT)21	160-209	2D	3
WMS325	(CT)16	133-149	6D	6
WMS357	(GA)18	119-25	1A	3
WMS389	(CT)14(GT)16	116-150	3B	7
WMS408	(CA)>22(TA)(CA)7(TA)9	149-199	5B	5
WMS437	(CT)24	91-130	7D	7
WMS458	(CA)13	109-115	1D	3
WMS513	(CA)12	140-150	4B	4
WMS577	(CA)14(TA)6	126-214	7B	12
WMS619	(CT)19	135-173	2B	6
WMS631	(GT)23	187-212	7A	4
WMS680	(TG)9(AG)24 imp	110-141	6B	2

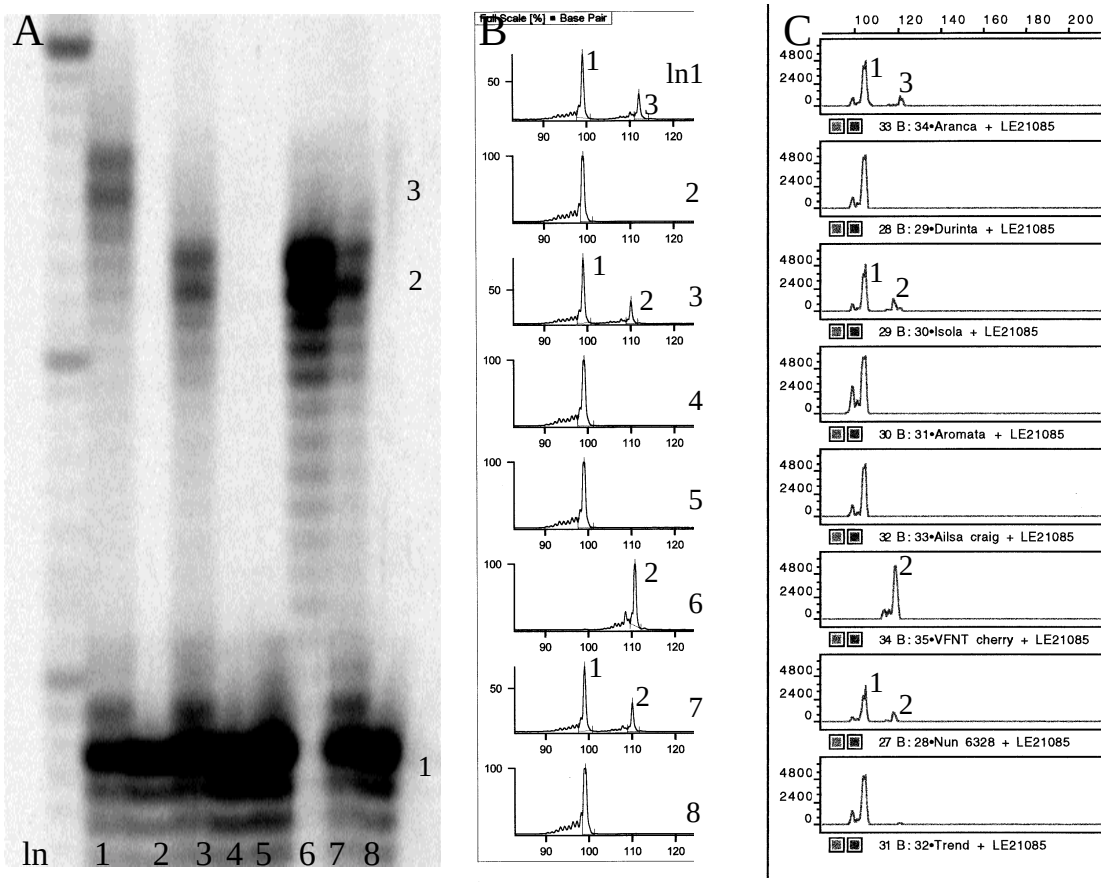
Legends

Fig. 1 LE21085 patterns of 8 tomato varieties generated by using fluorescein labelled primers in combination with a fluoro-phosphoimager (A. Agrogene image) and by using fluorescently labelled primers in combination with a DNA sequencer (B. ALFexpress/ CPRO image, C. ABI/ Nunhems image). Sample order lane 1. Aranca, 2. Durinta, 3. Isola, 4. Aromata, 5. Ailsa Craig, 6. VFNT Cherry, 7. Nun 6328, 8. Trend.

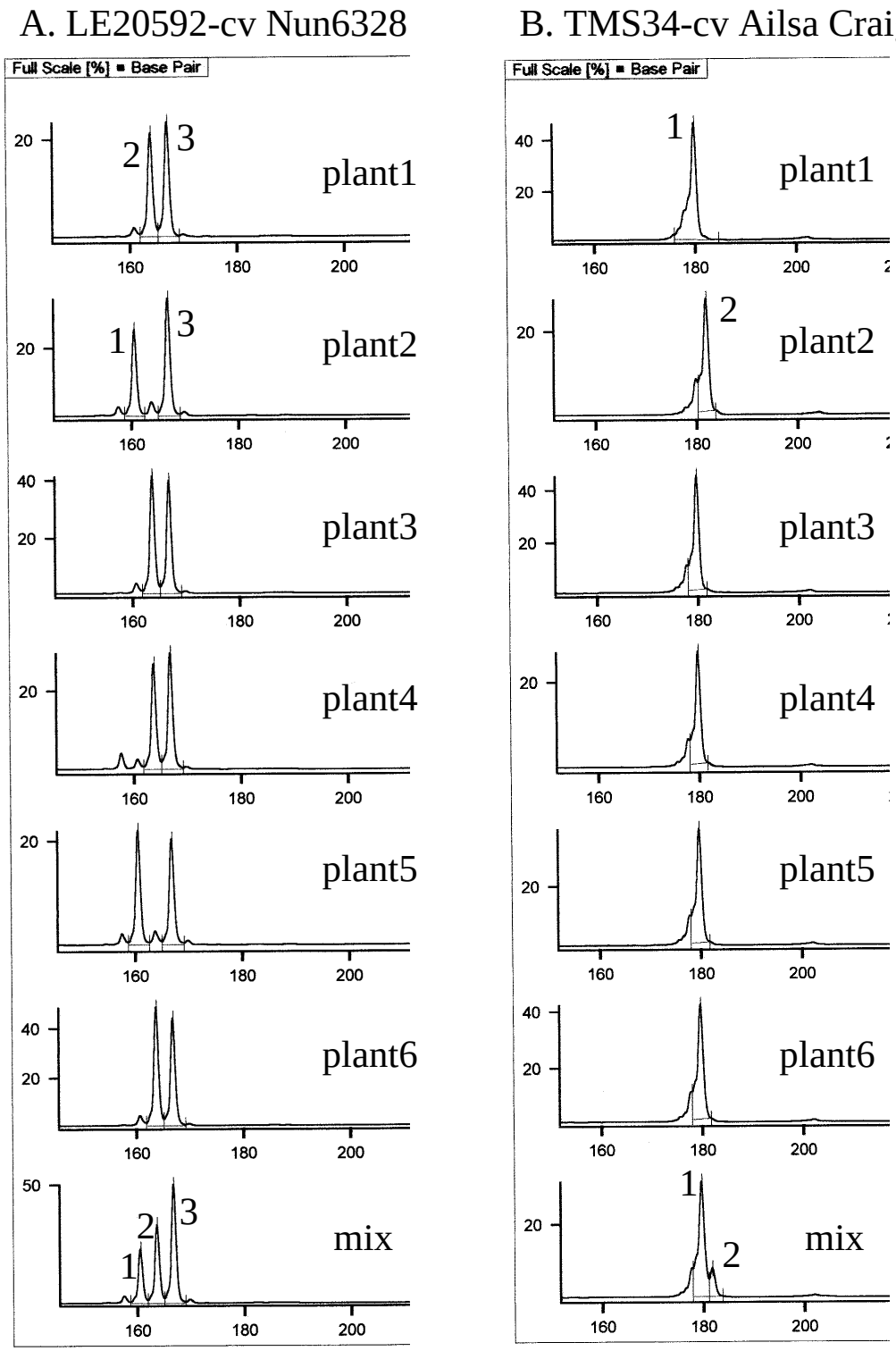
Fig. 2. Heterogeneity of tomato samples. ALFexpress patterns (CPRO) of 6 individual plants and a mix of 6 plants of cv Nun6328 amplified with LE20592 (A) and cv Ailsa Craig amplified with TMS34 (B).

Fig. 3. Multiplex pattern of a series of wheat varieties generated with the markers WMS680 and WMS3. Amplification products were detected on 6% Sequagel, using the LI-COR DNA 4200 (NIAB).

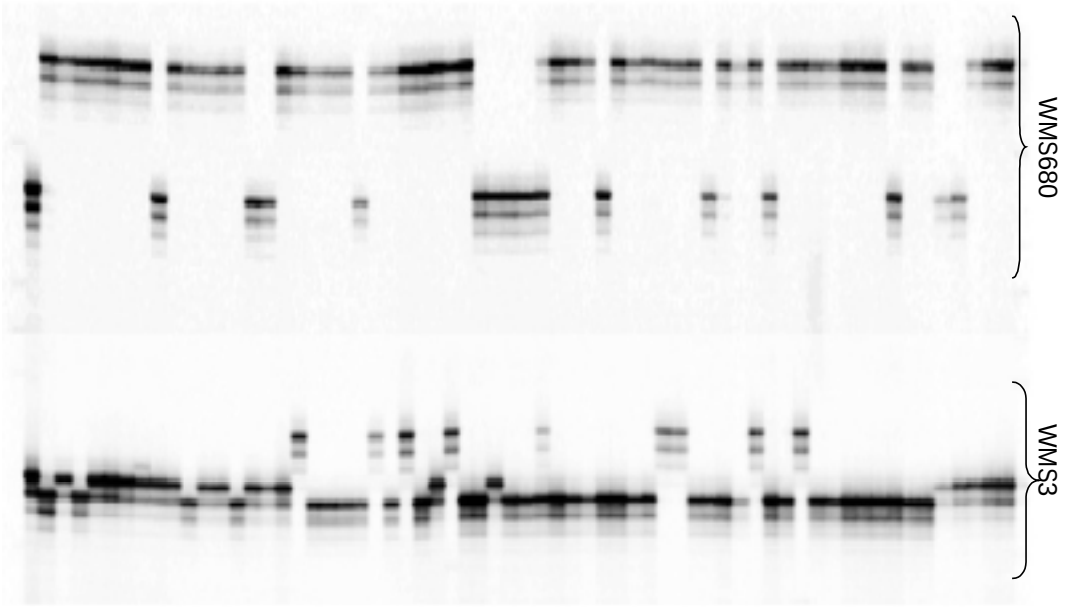
(figure 1)



(figure 2)


A. LE20592-cv Nun6328
B. TMS34-cv Ailsa Craig
Full Scale [%] ■ Base Pair
plant1
plant2
plant3
plant4
plant5
plant6
mix
160
180
200

(figure 3)



[End of document]