E

BMT/8/6
ORIGINAL: English
DATE: August 22, 2003

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

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

Eighth Session
Tsukuba, Japan, September 3 to 5, 2003

SNPS FOR VARIETY IDENTIFICATION IN TOMATO

Document prepared by experts from the Netherlands

SNPS FOR VARIETY IDENTIFICATION IN TOMATO

Ben Vosman*, Danny Esselink and Gerard van der Linden

Plant Research International b.v., Wageningen, the Netherlands
*author for correspondence; e-mail Ben.Vosman@wur.nl

Introduction

1. Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation. SNPs are found in every part of the genome; their frequency depending on the type of DNA and the genetic distance between the genotypes that are compared. The genome-wide occurrence would make the SNP an ideal marker for variety identification. Moreover, a large part of the phenotypic variation present in a set of varieties is likely to be based on SNPs at critical positions in and around genes that encode key proteins. Therefore, some SNP markers have the potential of being descriptive for the phenotype as well as the genotype.

2. For years, the potential of SNPs was recognized, but the lack of a large number of SNPs as well as an affordable system to score these in a large number of genotypes prevented the adoption of this marker system of the future. Technical developments in DNA sequencing and SNP detection during the last decade as well as the population of databases with an enormous body of sequence information have largely taken away these obstacles.

3. This study was undertaken to demonstrate the feasibility of SNPs as a molecular marker system for variety identification and breeding in tomato.

SNP discovery

4. SNPs were collected from various sources using several strategies:

    1. The nucleotide databases were searched for genomic sequences of tomato genes, and introns within these genes were amplified (with the primers located as much as possible in the flanking exons) and the amplicons sequenced directly in a set of nine tomato varieties and a wild relative (*L. pennellii*). Alignments of the sequences allowed identification of the SNPs.
    2. cDNA sequences (mRNAs from the nucleotide databases, ESTs) were used as a source for SNPs in expressed genes. The sequences were amplified in several tomato varieties similar to strategy 1.
    3. We used the Tanksley RFLP probes as starting points for SNP discovery. The probes were selected on genomic location and suitability of the sequence for amplification and SNP detection, amplified from a set of six tomato varieties, and SNPs extracted from the alignments.

5. Using strategy 1, nearly all primer pairs designed for locus amplification were successful. On average, one SNP was found in every 500 bp.

6. With strategy 2, approximately 30% of the primer pairs designed on EST and mRNA sequences failed to produce an amplification product with genomic DNA, most likely due to the presence of large introns in the target sequence. Several other products were not of the predicted length based on the mRNA sequence. Sequencing of these fragments demonstrated that in most instances introns were present in the target sequence. Hardly any SNPs were detected in coding sequences. The SNPs that were discovered based on exonic EST and mRNA sequences were nearly all located in introns in the genomic sequence.

7. For strategy 3, a total of 141 Tanksley probes were sequenced, and 133 putative SNPs were discovered in 68 loci. A large portion of these SNPs (83) was unique to one of the tested tomato varieties, which was shown with other markers to genetically differ considerably from most cultivated tomato. These may therefore not be highly informative for identification of commercially grown tomato varieties.

8. The different strategies produced over 200 putative SNPs for tomato.

SNP detection platform

9. A large variety of SNP detection platforms is currently available (Syvanen, A.–C. 2001) These differ widely with respect to number of samples that can be analyzed, the level of multiplexing possible, type of equipment needed and cost involved. Available platforms range from very simple systems (such as Tetra Arms PCR) allowing the detection of SNPs at low cost in a simple PCR reaction followed by agarose gel electrophoresis up until highly advanced systems (such as DNA-microarrays). The ideal platform should:

- allow medium to high throughput analysis
- be flexible towards number of SNPs and genotypes
- allow multiplex analysis of SNPs
- be Cost-effective
- be easy to implement, e.g. using equipment already available in the lab.

10. We have evaluated several SNP detection platforms for medium throughput SNP detection and tested two of them in tomato, including minisequencing on micro arrays, and the SNaPshot genotyping kit for fluorescent detection on automated sequencers.

11. Micro array-based detection with on-chip single base extension proved to be a reliable platform allowing multiplexing of a large number of SNPs. However, the micro-array approach is not flexible with respect to number of SNPs evaluated, all markers should be available at the start of the experiment. In addition, the throughput of this platform with regard to the number of genotypes is as yet low, and it is not easily implemented (requiring special equipment and expertise).

12. The SNaPshot multiplex genotyping kit for analysis on automated sequencers fulfills most of the stated requirements. It is flexible, allows multiplexing up to 10 SNPs, is cost-effective and can handle a large number of genotypes efficiently. Moreover, it can be used with standard sequencing equipment and implementation is relatively easy.
From the set of Tanksley probe-based SNPs, 35 SNPs (on 20 loci) were used for SNaPshot multiplexed genotyping of a collection of varieties and related materials

Tomato SNP genotyping with SNaPshot technology

13. The set of 35 polymorphic SNPs could uniquely identify most of the varieties tested. Ten varieties formed five pairs with unique fingerprints (fig. 1), the sixth pair consists of a duplicate sample. As expected, some individuals resulting from a segregating population grouped nicely together. The wild *lycopersicon* species were clearly different from the cultivated varieties. Two-thirds of the SNPs produced a scorable result in the SNaPshot analysis with two *Solanum tuberosum* varieties, but only a few were polymorphic. Further testing of these SNPs may reveal whether these SNPs predate potato and tomato speciation. The SNPs did not produce a result in the more distant *Capsicum* varieties.

14. In half of the loci that contained more than one SNP, several haplotypes could be distinguished within the variety set, indicating recombination events within the few hundred base pairs separating these markers.

Conclusions

15. SNPs can be detected by comparative sequencing. For tomato, finding SNPs useful for variety identification in this way is not very efficient and cost effective. Although sequencing of 140 RFLP probes resulted in over 130 SNPs on 68 loci, only 35 SNPs on 20 loci appeared to be useful for identification. However, it is not unlikely that informative SNPs for tomato will be available through other sources in the near future. Tomato sequence data are accumulating fast, and efficient mining of the growing sequence databases for putative SNPs is rapidly becoming a feasible and efficient alternative for SNP discovery.

16. The SNaPshot multiplex genotyping technology is a suitable platform, which can be implemented easily and allows multiplexing of 8-10 SNPs, with only limited optimisation efforts required.

17. Identification of varieties using SNPs seems to be relatively efficient and it is likely that a small extension of the set of SNPs will allow the unique identification of a large group of varieties (up to 95-98%). As with all techniques, the last few percents will be difficult due to the fact that the varieties are so similar. But, the advantage of using SNPs in this case is that one can carefully select the markers to be used on the basis of map position and level of discrimination. Also, for specific groups of varieties the most informative SNPs can be combined in a multiplex for efficient and cost-effective genotyping.

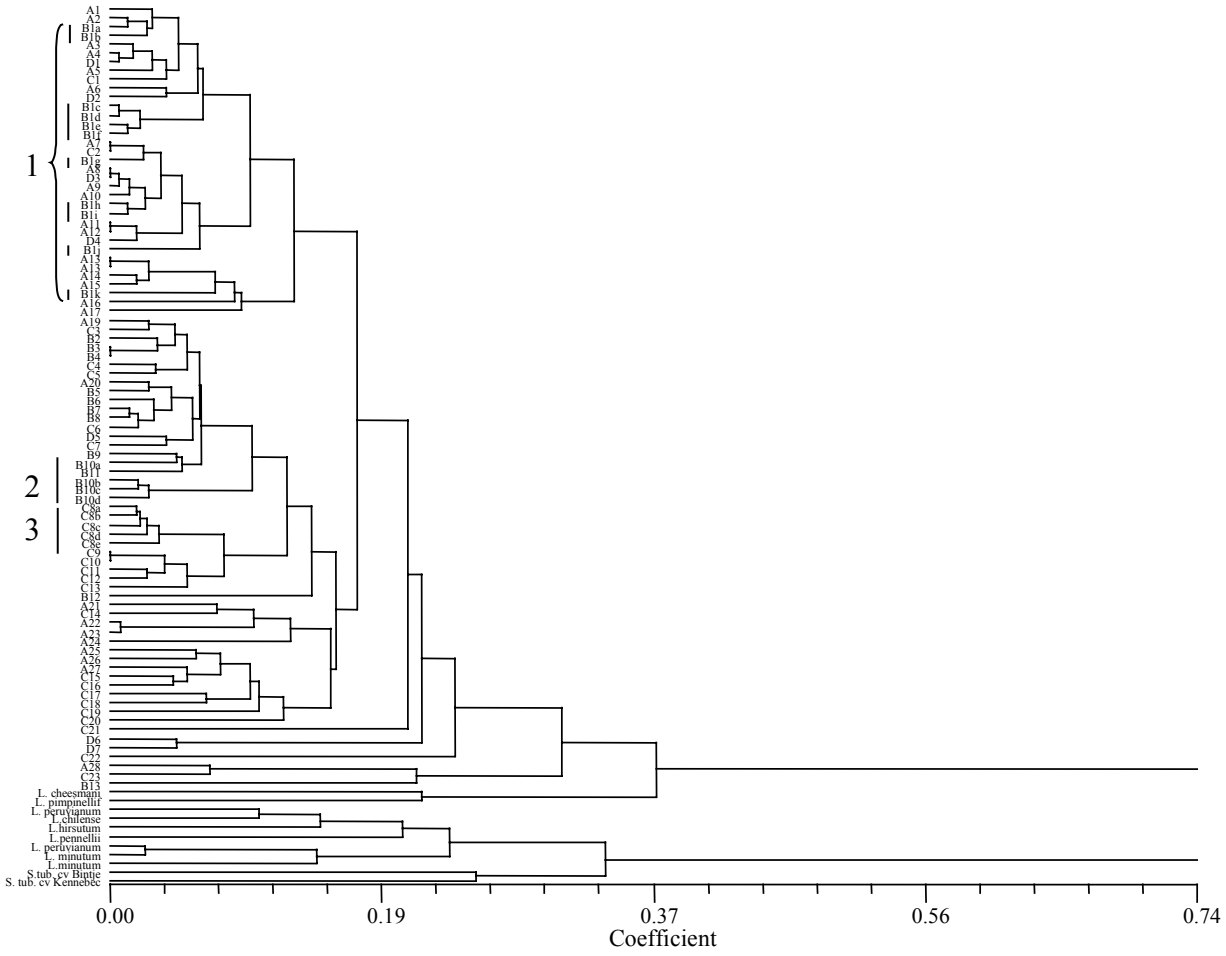
18. Finally, SNPs have the advantage of easy databasing and results obtained are independent of the detection platform.

Acknowledgements
Part of this project was funded by a consortium of ENZA zaden (Enkhuizen), Syngenta Seeds b.v. (Enkhuizen), Nunhems zaden (Haelen) and Plant Research International (Wageningen). The other part was funded by the Ministry of Agriculture, Nature management and Fisheries of the Netherlands.

Literature:
Syvanen, A. –C. (2001) Accessing genetic variation: Genotyping single nucleotide polymorphisms. Nature Genetics 2: 931-942

Legend to Fig. 1:

Genetic distance analysis of Lycopersicon genotypes and related species with SNP markers. Varieties are coded, with A, B and C groups representing varieties from three different breeders, and D varieties from several other breeders. The numbers (1, 2 and 3) indicate groups of genotypes (B1a-k, B10a-d and C8a-e respectively) that are part of progenies of single crosses from breeding programs.



[End of document]