Discovery, mapping, and application of single nucleotide polymorphisms in Lycopersicon esculentum

Wencai Yang 1, Xiaodong Bai2, Christina Eaton3, Sophie Kamoun4, Esther van der Knaap5, and David Francis 1

1Department of Horticulture and Crop Science, 2Department of Entomology, 3Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center, 1680 Madison Ave., Wooster, OH 44691, USA

Summary

Single nucleotide polymorphism (SNP) discovery through de novo sequencing is inefficient within cultivated tomato (Lycopersicon esculentum Mill) because the polymorphism rate is lower than the sequencing error rate. The availability of expressed sequence tag (EST) data has made it feasible to discover potential SNPs. We aimed to apply the sequencing-by-synthesis Verilog technology to experimentally verify SNPs by dividing the tomato ESTs by variety of origin, selecting contigs with a minimum of three sequences to correct for sequence error, and comparing sequences from different varieties. We have identified candidate SNPs for use within cultivated germplasm pools. 1245 contigs have three EST sequences of Rio Grande and three EST sequences of TA496 were used for SNP discovery. We detected 1 SNP for every 8,500 bases analyzed, with 101 candidate SNPs in 44 genes identified. Experimental verification using restriction digestion or Cel I digestion confirmed 83% of the polymorphic tests presented. SNPs between Rio Grande and TA496 have a high probability (53%) of detecting SNPs between other L. esculentum varieties. Twenty-six SNPs in 14 genes were mapped to specific chromosomes. SNPs LEOH32 and LEOH35 were linked to quantitative trait loci contributing to fruit color in crosses between elite varieties.

Introduction

A limitation to applying marker-assisted selection is the practice of breeding tomato varieties in that the level of polymorphism between elite varieties is very low. The objective of this research was to assess the potential of using existing polymorphisms to facilitate marker-assisted selection. The Ohio Agricultural Research and Development Center (OARDC) has made available at least 138,100 Expressed sequence tags (ESTs). Of these, approximately 15% were derived from the variety Rio Grande or from Rio Grand x Monarch crosses. The remaining sequences were derived from TA496, which has a processing tomato pedigree trace to IE205. By comparing sequence data from Rio Grande and TA496 we identified genetic differences between these varieties. polymorphisms were discovered from this data mining were then applied to genetic studies within breeding populations, and markers linked to quantitative trait loci (QTL) contributing to fruit color were identified.

Materials and methods

Identifying single nucleotide polymorphisms (SNPs)

The approach to discovering SNPs in ESTs from Lycopersicon esculentum is outlined in Figure 1. Briefly, the OARDC EST database was parsed into database consisting of only TA496 and Rio Grande, R11-12 and R11-13. Sequence comparisons between these data sets were used to identify potential SNPs.

Confirmation of Candidate SNPs

Primers were designed flanking the putative polymorphism with the optimal PCR product length set between 150 and 600 bp. Four variations TA496, E8203, Rio Grande and R11-12 were used to verify amplification products. Sixty-six SNPs were used to verify the utility of these SNPs. The polymorphism rate was determined by scoring amplified products, heating to denature, and re-annealing. Cell 1 was purified according to published methods and DNA digestion was performed at 45°C in 20 ml Tris-HCl pH 7.4, 25 mM KC1 and 10 mM MgC12, for 30 min. Single stranded Cel I digestion products were separated using 10% TBIurea/polyacrylamide and visualized by staining with SYBR Gold.

SNPs polymorphic in other L. esculentum wild species/germplasm

To test if the SNPs identified between TA496 and Rio Grande are also polymorphic in other L. esculentum varieties, 67 potential SNPs were used for further verification. Of the 43 confirmed SNPs, 27 also polymorphisms among the selected varieties. The SNPs discovered in the EST database had a high probability (53%) of detecting SNPs between other varieties.

SNPs were also tested for polymorphism in wild tomato species LA716, (L. pennelli), LA407, and Money maker. These varieties were used to test polymorphisms using PCR followed by restriction enzyme digestion or Cel I digestion after partial digestion. The polymorphism rate was between 19% and 22% of the observed variation for color, and do not correspond to previously described genes (for example op5 on chromosome 6) known to affect color in tomato.

SNPs polymorphic in other L. esculentum wild species/germplasm

Polymorphism of SNPs

Two populations were used to map the SNPs. The first was a set of L. esculentum cultivars MR2 (Rishk and Zanin), 1998 Genetics 141:1147-1152. The second population was an F2 derived from a cross of LA1589 and Sun6 and LA407. The F2 coordinate indicates darkness or lightness of color. Chroma (saturation or vividness of color) is calculated from a* and b* as (a* 2 + b* 2) ½. As chromaticity increases, a color becomes more intense; as it decreases a color becomes more dull. Data were collected for 24 fruit from individual F2 progeny in two breeding populations. Populations consisting of 160 individuals and 1023 x Ohio 7814 consisting of 60 individuals.

Statistical analysis

Linkage relationships among genotypic classes of SNPs with L. and Chama were determined by ANOVA using single marker-trait analysis. Markers were considered as fixed effects, replicates as random effects, and the F2 was performed using the genotype within marker variation as the error term. Total phenotypic variation explained by markers identified for L. and Chroma was calculated with variance components estimated using restricted maximum likelihood (REML).

Result

SNPs between TA496 and Rio Grande: 1245 contigs with at least three ESTs for each variety were available for identifying potential SNPs. Forty-four unigenes showing 101 potential polymorphisms were identified. Only two polymorphisms were insertions/deletions (indels). Sixty-six potential SNPs. Forty-four unigenes showing 101 potential polymorphisms were identified.

SNPs polymorphic in other L. esculentum wild species/germplasm

Table 1. Summary of EST derived SNPs confirmed in Lycopersicon esculentum. Forty-three of out 55 SNPs tested (82%) were confirmed.

Table 2. SNPs associated with lightness-darkness of color (L) and intensity of color (Chroma) in two elite breeding populations.

Table 3. SNPs associated with fruit color

Table 4. SNPs associated with fruit color

Conclusion

•The frequency of single nucleotide polymorphisms in L. esculentum ESTs is lower than in other plant species. The polymorphism rate is approximately ten fold lower than the sequencing error rate.

•Public access to EST sequence trace files or Phred quality value data would allow for more efficient SNP discovery by permitting the use of quality information as a substitute for sequence redundancy.

Based on the estimated number of genes in tomato (35,000; Van der Hoeven et al., 2002). The Plant Cell 14:1141-1152) we estimate that there are 1000 SNPs between TA496 and Rio Grande.

SNPs between Rio Grande and TA496 have a high probability (53%) of detecting SNPs between other L. esculentum wild species/germplasm.