Generation of SNPs in eggplant (Solanum melongena L.)

Barchi L[1], Lanteri S[1], Portis E[1], Acquadro A[1], Valè G[2], Toppino L[3], Rotino GL[3]

Motivation

Single nucleotide polymorphisms (SNPs) are the most abundant types of DNA sequence polymorphism. Their high availability provide enhanced possibilities for genetic and breeding applications such as cultivar identification, construction of genetic maps and marker-assisted breeding. Furthermore, the development of high throughput genotyping methods make SNPs highly attractive as genetic markers. Several methodologies are available for SNPs discovery. Recently, Miller et al. (Genome Research, 17, 2007) proposed the use of Restriction-site Associated DNA (RAD) method in association with next generation sequencing (Illumina) for rapid and cost effective SNP mining. RAD technology creates a reduced representation of the genome (RAD tags) allowing for the identification of high number of SNPs suitable for genetic studies. At present in eggplant (Solanum melongena L.) no information are reported on SNP marker development. Here we report on the identification of a wide set of SNPs by sequencing RAD tags derived from two eggplant inbred lines: '305E40' and '67/3', which we used as parents of an F2 intra-specific mapping population. The newly identified SNPs derived markers will consistently contribute in saturating the already developed intra-specific genetic map and will provide powerful tools for comparative genetics studies within the Solanaceae family.

Methods

Plant materials The parental line '305E40', characterized by elongated fruit, is a doubled haploid (DH) obtained through anther culture of an advanced introgression line (BC7) derivative of an interspecific hybrid somatic Solanum aethiopicum gr. Gilo x S. melongena cv. Dourga. The parental line '67/3' is a selection from the intraspecific cross cv. 'Purpura' x cv.'CIN2' followed by seven cycles of selfing, and is characterized by round fruit type. These parents were crossed to obtain F1 hybrids which, after selfing, generated a F2 population of 141 individuals. RAD library preparation, sequencing, assembly and SNP discovery Genomic DNA from the parental lines '305E40' and '67/3' was digested with restriction endonucleases SgrAI (SgrAI round) and PstI (PstI round) and an adapter (P1) was ligated to the

[1] DIVAPRA Plant Genetics and Breeding, University of Torino, 10095 Grugliasco (Torino) Italy; [2] CRA-GPG Genomic Research Centre, 29017 Fiorenzuola d'Arda (Piacenza) Italy [3] CRA-ORL Research Unit for Vegetable Crops, 26836 Montanaso Lombardo (Lodi) Italy

generated fragments. The P1 adapter contained a forward amplification primer site, an Illumina sequencing primer site, and a barcode. P1 Adapter-ligated fragments were combined, sheared and ligated to a second adapter (P2). RAD tags, carrying P1 and P2 adapters, were selectively PCR amplified. RADs from each parent were sequenced on the Genome Analyzer II (Illumina) platform using paired end 54 bp sequence reads (2 x 54 bp). Paired end sequences from each parent were pooled and segregated by single read RAD sequences. Velvet was used to assemble consensus LongRead contigs from the paired end data. For SgrAI round, SNPs were called at minimum of 4x coverage while for PstI round SNPs were called at minimum of 6x coverage. Sequences annotation CAP3 algorithm was used for identifying sequences in common between parents. A new dataset (A) was constituted for further analyses: it included singlets from '67/3' and '305E40' as well as contigs deriving from both RAD rounds. Standalone BLAST tool was used to identify best annotation for each dataset. A BLASTX algorithm was carried out against TAIR9, adopting as a threshold E-value 1e-09, while BLASTN algorithm was performed against SGN Cornell unigene database with a cut-off !1e-20. Gene Ontology categorization of non redundant sequences was inferred by TAIR9 best hits.

Results

SNPs discovery A total of 38,941 and 38,935 contigs were obtained for parent '67/3' and '305E40', respectively. Globally, almost 15.45 Mb of high-quality de novo eggplant sequences were obtained. The average contig length was 353.9 bp for '67/3' and 340.8 bp for '305E40', with an overall value of 347.3 bp. A total of 11,580 SNPs and 1,664 InDels were identified considering both lines. The complete SNP panel was screened to identify those alleles free of flanking polymorphisms and 2,435 SNPs were found as candidates for genotyping assays. Sequences annotation The dataset A, constituted of 43,795 sequences (30,187 in common from both parents, 5,535 singlets from '67/3' and 5,603 singlets from '305E40'), was searched against TAIR9 protein database using BLASTX. Globally, 11,588 sequences (26.45%) matched at E-value !1E-09, clustering with 6,875 Arabidopsis unigenes. The latter were successfully GO categorised. BLASTN search against Cornell unigene database revealed a total of 17,631 sequences having a significant hit (E-value !1E-20), clustering with 14,338 SGN unigenes: these data will be useful for anchoring eggplant genetic maps to the tomato genome scaffold. In conclusion, the de novo sequencing of 15.45 Mb of eggplant genome, via RAD technology, provided 2,435 SNPs suitable for a wide range of genomic studies.

Contact e-mail

alberto.acquadro@unito.it