Development of ESTs and data mining of pineapple EST-SSRs

Abstract

Improving the quality of the non-climacteric fruit, pineapple, is possible with information on the expression of genes that occur during the process of fruit ripening. This can be made known through the generation of partial mRNA transcript sequences known as expressed sequence tags (ESTs). ESTs are useful not only for gene discovery but also function as a resource for the identification of molecular markers, such as simple sequence repeats (SSRs). This paper reports on firstly, the construction of a normalized library of the mature green pineapple fruit and secondly, the mining of EST-SSRs markers using the newly obtained pineapple ESTs as well as publically available pineapple ESTs deposited in GenBank. Sequencing of the clones from the EST library resulted in 282 good sequences. Assembly of sequences generated 168 unique transcripts (UTs) consisting of 34 contigs and 134 singletons with an average length of \( \approx 500 \) bp. Annotation of the UTs categorized the known proteins transcripts into the three ontologies as: molecular function (34.88\%), biological process (38.43\%), and cellular component (26.69\%). Approximately 7\% (416) of the pineapple ESTs contained SSRs with an abundance of trinucleotide SSRs (48.3\%) being identified. This was followed by dinucleotide and tetranucleotide SSRs with frequency of 46 and 57\%, respectively. From these EST-containing SSRs, 355 (85.3\%) matched to known proteins while 133 contained flanking regions for primer design. Both the ESTs were sequenced and the mined EST-SSRs will be useful in the understanding of non-climacteric ripening and the screening of biomarkers linked to fruit quality traits. © 2011 Springer Science+Business Media B.V.