

**Genome-wide *in silico* analysis and characterization of simple sequence repeat loci in
Coconut (*Cocos nucifera* L.)**

W. O. K. Weerakkodi¹, H. D. D. Bandupriya^{1*}

¹Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka

* dbandupriya@pts.cmb.ac.lk

Coconut (*Cocos nucifera* L.) is an important oil crop that can play a major role in the food security of the people in most of the Asian countries including Sri Lanka. The genomic knowledge on coconut is extremely important to increase productivity by incorporating superior traits targeting quality end products. Simple Sequence Repeats (SSR) are DNA-based markers which are informative for basic and applied studies such as molecular breeding. Coconut genetic and genomic resources, especially molecular markers, have been scarce until recently, impeding coconut breeding efforts. Comprehensive analysis of SSR has become possible now because several coconut genomes have been sequenced. In this study the draft genome of *C. nucifera* Hainan Tall (HAT) variety (accession number GCA_008124465.1) was surveyed to determine the distribution and frequency of SSRs. The polymorphism of the available SSRs of the HAT variety was investigated using an electronic PCR approach and, manual analysis of each PCR product marker in five genomes: *C. nucifera* Chowghat Green Dwarf variety, *C. nucifera* Catigan Green Dwarf variety, *Phoenix dactylifera* and *Elaeis guineensis* and *C. nucifera* L. ESTs. A total of 522,524 SSRs (repeat units 1-6 bp) from the draft genome of variety HAT with an overall density of 258.675 (SSRs/Mbp) was characterized using the GMATA tool. Dinucleotide was the most common repeat motif with a frequency of 58.87%, followed by 28.01% mononucleotides, 7.15% trinucleotide, 4.31% tetranucleotide, 1.50% pentanucleotide and 0.16% hexanucleotide. The motif AG/CT was the most abundant and AT/TA was the second most abundant among all identified SSR motifs by accounting for 42.01% and 32.79% respectively. A total of 245,048 unique SSR markers were developed from the total SSR loci and a high degree of polymorphism (nearly 51 %) was found between different genotypes. Polymorphic SSR sites of *C. nucifera* were mostly comprised of dinucleotide motifs (82.91%), followed by trinucleotide motifs (10.39%), then tetranucleotide motifs (4.15%). The findings of the present study indicate that whole-genome sequencing is an excellent resource for developing SSR markers, and the newly identified large numbers of SSR markers could make an important contribution to the coconut research community

Keywords: *C. nucifera* L. , microsatellites, polymorphism