

Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae

Abstract

Background

Pineapple (*Ananas comosus* var. *comosus*) is known as the king of fruits for its crown and is the third most important tropical fruit after banana and citrus. The plant, which is indigenous to South America, is the most important species in the Bromeliaceae family and is largely traded for fresh fruit consumption. Here, we report the complete chloroplast sequence of the MD-2 pineapple that was sequenced using the PacBio sequencing technology.

Result

In this study, the high error rate of PacBio long sequence reads of *A. comosus*'s total genomic DNA were improved by leveraging on the high accuracy but short Illumina reads for error-correction via the latest error correction module from Novocraft. Error corrected long PacBio reads were assembled by using a single tool to produce a contig representing the pineapple chloroplast genome. The genome of 159,636 bp in length is featured with the conserved quadripartite structure of chloroplast containing a large single copy region (LSC) with a size of 87,482 bp, a small single copy region (SSC) with a size of 18,622 bp and two inverted repeat regions (IRA and IRB) each with the size of 26,766 bp. Overall, the genome contained 117 unique coding regions and 30 were repeated in the IR region with its genes contents, structure and arrangement similar to its sister taxon, *Typha latifolia*. A total of 35 repeats structure were detected in both the coding and non-coding regions with a majority being tandem repeats. In addition, 205 SSRs were detected in the genome with six protein-coding genes contained more than two SSRs. Comparative chloroplast genomes from the subclass Commelinidae revealed a conservative protein coding gene albeit located in a highly divergence region. Analysis of selection pressure on protein-coding genes using Ka/Ks ratio showed significant

positive selection exerted on the *rps7* gene of the pineapple chloroplast with *P* less than 0.05. Phylogenetic analysis confirmed the recent taxonomical relation among the member of commelinids which support the monophyly relationship between Arecales and Dasypogonaceae and between Zingiberales to the Poales, which includes the *A. comosus*.

Conclusions

The complete sequence of the chloroplast of pineapple provides insights to the divergence of genic chloroplast sequences from the members of the subclass Commelinidae. The complete pineapple chloroplast will serve as a reference for in-depth taxonomical studies in the Bromeliaceae family when more species under the family are sequenced in the future. The genetic sequence information will also make feasible other molecular applications of the pineapple chloroplast for plant genetic improvement.