Genome-wide identification of genes involved in the lignin biosynthetic pathway

Hansheng Zhao

Abstract

Protocol

Data downloading: The five genome sequences of A. thaliana (TAIR10), B. distachyon (v3.1), O. sativa (v7.0), Populus trichocarpa (JGI2.0.31), and S. bicolor (v3.1) were downloaded from the ENSEMBL database

Step 1.

Literature-based investigating: 140 genes from the lignin biosynthetic pathway was experimentally validated from previous studies, and then, these known genes were collected and used as the query sequences for further identification.

Step 2.

we performed BLASTN searches against the six genome sequences including moso bamboo using the coding sequence of the known genes with the following cut-off values: E-value <1e-10; identity >95%; and coverage rate >40% query sequence.

Step 3.

≋ SOFTWARE PACKAGE (Linux -)

Standard Nucleotide BLAST, 2.2.26 ↗

The filtered sequences were subsequently analyzed by hmmsearch using the Pfam-A.hmm database (released 2017/03/31).

Step 4.

≋ SOFTWARE PACKAGE (Linux -)

hmmsearch, 3.1b2 ↗
http://hmmer.org/download.html

Consequently, unclear sequences with incomplete domains were discarded by manual correction

Step 5.