

Transcriptome annotation

Bing Cheng; Agnelo Furtado; Robert Henry

Abstract

This protocol is to annotate the tetraploid coffee long read transcriptome.

Citation: Bing Cheng; Agnelo Furtado; Robert Henry Transcriptome annotation. **protocols.io**

dx.doi.org/10.17504/protocols.io.ja4cigw

Published: 10 Aug 2017

Guidelines

Warning

BLAST output format should be type 5 (.xml).

Protocol

Download plant proteins GI list

Step 1.

Download the plant geninfo identifier (GI) list from NCBI Protein Entrez.

Create a non-redandant plant proteins database (nr-plant)

Step 2.

Download the NCBI non-redentant proteins from NCBI FTP site. Then retrieve the plant proteins with the plant GI list.

NOTES

GigaScience Database 08 Aug 2017

Use of high performance computers (HPC).

BLASTx against nr-plant

Step 3.

E-value set at 1e-10.

NOTES

GigaScience Database 08 Aug 2017

Use of HPC.

Download non-redundant nucleotide (nt) database

Step 4.

Download the NCBI non-redundant proteins from NCBI FTP site.

📌 NOTES

GigaScience Database 08 Aug 2017

Use of HPC.

BLASTn against nt database

Step 5.

Submit sequences without hit to nr-plant to BLASTn to nt database

📌 NOTES

GigaScience Database 08 Aug 2017

Use of HPC.

BLASTn to C. canephora coding sequences with UTR and C. arabica EST database.

Step 6.

E-value set at 1e-20.

📌 NOTES

GigaScience Database 08 Aug 2017

Use of HPC.

Filter the BLAST output

Step 7.

The output of BLAST result was filtered with query coverage, cumulative identity and sequence length. High, medium and low quality were clustered.

📌 NOTES

GigaScience Database 08 Aug 2017

Use of HPC. The script to filter the result is available at:
https://github.com/chengbing0404/BLAST5_result_handle

Function annotation

Step 8.

InterProScan, mapping, annotation, annex, GO-Slim and KEGG pathway was loaded.

📌 **NOTES**

GigaScience Database 08 Aug 2017

BLAST2GO software was used.

Search for non-coding RNAs

Step 9.

Sequences without hit to all the above databases were processed further to search the non-coding RNAs.

📌 **NOTES**

GigaScience Database 08 Aug 2017

Run Rfam from BLAST2GO