

Orthologous Gene and Phylogenetic analysis

Hansheng Zhao

Abstract

Citation: Hansheng Zhao Orthologous Gene and Phylogenetic analysis. **protocols.io**

dx.doi.org/10.17504/protocols.io.phpdj5n

Published: 16 Apr 2018

Protocol

Single-copy gene families and multi-gene families were identified by orthMCL (version 2.0.9) among *Ph. edulis* and other 7 plant species, including *Amborella trichopoda* (version 1.0) from Amborella Genome Database (amborella.huck.psu.edu), *Elaeis guineensis* (GCF_000442705.1) from NCBI database, *Arabidopsis thaliana* (TAIR10), *Brachypodium distachyon* (version 3.1), *Oryza sativa* (version 7.0), *Spirodela polyrhiza* (version 2) and *Sorghum bicolor* (version 3.1) from ENSEMBL database.

Step 1.

☰ SOFTWARE PACKAGE (Linux -)

orthMCL, 2.0.9

Protein alignment was conducted by MUSCLE (version 3.8.31), and CDS alignment was constructed based on the result of protein alignment. Then, all aligned CDS sequences were concatenated to generate a super gene for each specie using an in-house Perl script.

Step 2.

☰ SOFTWARE PACKAGE (Linux -)

MUSCLE, 3.8.31

All single-copy genes were used to construct the phylogenetic tree by PhyML (version 3.0) specifying a HKY85 substitution model with a gamma distribution across sites

Step 3.

☰ SOFTWARE PACKAGE (Linux -)

PhyML, 3.0 [↗](http://www.atgc-montpellier.fr/phyml/)

<http://www.atgc-montpellier.fr/phyml/>

We used the MCMCTree program of PAML (version 4.5) to estimate the divergence time with the following parameters: “-nsample 200000 -burnin 40000”.

Step 4.

☰ SOFTWARE PACKAGE (Linux -)

PAML, 4.5 [↗](#)