

ohnologuous genes parser

Aim

The aim of this procedure is to retrieve pairs of ohnologuous genes from a Synmap analysis between two of the same genome.

Procedure

1- Synmap analysis

The first step is to perform a SynMap analysis (<https://genomeevolution.org/CoGe/SynMap.pl>) and inputing two times the same genome (see SynMap doc). Then download the following file :

Download Results		
General	Homolog search	Diagonals
Analysis Log	Fasta file for Malus x domestica (apple) (v1.1): CDS	DAGChainer Initial Input file
Image File	Fasta file for Malus x domestica (apple) (v1.1): CDS	DAGChainer Input file converted to gene order
	Unfiltered Last results	DAGChainer Input file post repetitive matches filtered
	Filtered Last results (no tandem duplicates)	
	Tandem Duplicates for Malus x domestica (apple) (v1.1)	
	Tandem Duplicates for Malus x domestica (apple) (v1.1)	
		Results
		DAGChainer Output
		DAGChainer output in genomic coordinates
		Final syntenic gene-set output with GEvo links
		Condensed syntelog file with GEvo links
		SVG Version of Syntenic Dotplot

2- *filter_ohnologuous_genes.py* use

Description

This script parses the DAGChainer output in order to retrieve pairs of ohnologuous genes which are inside large enough synteny windows to minimize noise.

Usage

```
python filter_ohnologuous_genes.py input_file output_file min_window_size
```

- **input_file** : the DAGchainer output file shown above
- **output_file** : path to the output file
- **min_window_size** : minimum synteny window size (in **base pairs**) in which genes will be considered. On at least one chromosome, the size of the synteny link has to be superior or equal to this value.

Output

The output (**output_file** parameter) is a tab-separated file with the two following columns :

- **1** : first gene of the ohnologuous pair
- **2** : second gene of the ohnologuous pair