

Identitag documentation

Using Identitag for comparative transcriptomic analysis

1 Introduction

This document describes how to use Identitag for comparing SAGE libraries from two different species, thus allowing to do comparative transcriptomic analysis.

Before doing comparative transcriptomic analysis, two Identitag databases for two different species must be built (the section "Use of Identitag for tag-to-gene mapping" of Identitag website describes how to construct an Identitag database). Then you can connect these two Identitag databases using orthology relationship. The scripts provided here search for putative orthologous sequences between the two sets of transcript sequences from these two different species. It takes into account the fact that the two sets of transcript sequences could be redundant and could not represent the entire transcriptome of the two species considered (for more details about this method see Identitag website). Then these putative orthologous relationships are loaded into Identitag database (the structure of this database is described on Identitag website). For questions and comments on Identitag please contact Céline Keime : keime@cgmc.univ-lyon1.fr.

2 What is necessary to connect two Identitag databases using orthology relationship ?

2.1 Software requirements

- A Bourne Shell (/bin/sh) and Perl interpreter (/usr/bin/perl)
- A MySQL server and client
(LOAD DATA LOCAL must be active on MySQL client)
- blastall, blastclust and formatdb executables must be operational on your environment

2.2 Data requirements

- A database containing tables corresponding to two Identitag databases for two different species.

- A file containing the proteins used for BLASTX (last step of the search for orthologous sequences), in Fasta format, e.g. sptrembl.fas (this file could be found on EBI ftp site : ftp.ebi.ac.uk/pub/databases/sp_tr_nrdb/fasta/).

3 Connecting two Identitag databases using orthology relationship

- Uncompress the file orthology.tgz : this produces a directory called "orthology" which contains all scripts needed to create and load all Identitag tables.
- In orthology directory, run sh Orthology.sh.
- Answer to all questions, this will search for orthologous transcript sequences between the two species considered, then create the Species1_Species2_Transcript table in your database and fill this table with putative orthology relationships.