

HYBRIDNUMBER

An algorithm to calculate the exact number of hybridization events for two given rooted binary phylogenetic trees on the same set of taxa.

Magnus Bordewich, Simone Linz, Katherine St. John, and Charles Semple

March 9, 2007

Requirements

The Perl command-line script `HybridNumber.pl` requires the following:

- Perl 5.8 (tested under Windows and Linux)
- depending on your current Perl installation the following libraries might be necessary and can be downloaded via CPAN (<http://www.cpan.org/>):
 - `Array::Compare` by Magnum Solutions Ltd.
 - `Graph` by Jarkko Hietaniemi
 - `Heap` by O'Reilly & Associates
 - `List::Compare` by James E. Keenan

Installation

HYBRIDNUMBER consists of the main Perl script `HybridNumber.pl` and some Perl modules which are listed in the following:

- `tree.pm`: basic routines to handle a tree
- `subtree.pm`: routines to apply the subtree reduction rule
- `chain.pm`: routines to apply the chain reduction rule
- `cluster.pm`: routines to apply the cluster reduction rule
- `forest.pm`: routines to calculate an acyclic-agreement forest of minimum size, including the exhaustive search part of the algorithm

To run `HybridNumber.pl`, store all above mentioned Perl modules in the same directory as `HybridNumber.pl` itself. Otherwise, if you prefer a separate directory containing all `*.pm` files, you need to adjust line 17 of the `HybridNumber.pl` script to the right path. For example, if you copy the modules to

```
/home/user/PerlModules/
```

you have to edit line 17 in the following way:

```
use lib "/home/user/PerlModules/";
```

To calculate the hybridization number for two given rooted binary phylogenetic trees, you need to execute the following command:

```
perl HybridNumber.pl <tree1> <tree2>.
```

Both treefiles need to be in the string format which is described in the next section. To check if the program runs properly, you may want to use the two sample trees provided in the `sample_trees` directory with a hybridization number of 3.

The string format

As an input, `HybridNumber.pl` requires the two trees to be given in a special string format. For example, consider the 5-species tree $((((1,2),(3,4)),5))$; given in the Newick Format. In the corresponding string format all leaves are represented by their present-day species and all inner vertices (parentheses in the Newick Format) are represented by a concatenation of their descendant species. Therefore, the string format of the above mentioned tree, starting at the root, is

```
1,2,3,4,5 1,2,3,4 1,2 3,4 1 2 3 4 5.
```

Each species in the tree must be represented by an integer, more than one species per vertex are separated by a comma, and two vertices are separated by a single space. Note that the whole tree string should fit in one single line.

In case you already have a tree given in the Newick Format with species names represented by numbers, you could also use the provided Java Program `newick2string/stringTree.jar` to convert the Newick into the string format. To start the program, execute the following command:

```
java -jar stringTree.jar <infile>.
```

Any feedback is most welcomed to: linz@cs.uni-duesseldorf.de