

User guide for EpiCs

September 25, 2015

Contents

1	Implementation	3
2	Input format	3
3	Usage	4
3.1	Choice of the matrix	4
3.2	Running EpiCs	5
4	Output format	6
5	Contact	6
6	Citing EpiCs	6

This manual provides a tutorial for EpiCs 1.0. EpiCs (Evolutionary processes independence Computed statistics) calculates the p-value of the co-repartition of the occurrences of a pair of events on a phylogenetic tree under a model of independence. The method is described in Behdenna et al. (submitted).

1 Implementation

EpiCs is an open-source program written in C language and compiled with gcc 4.2.1. It can only be run as a command line. It is available at:

<http://wwwabi.snv.jussieu.fr/public/EpiCs/>.

2 Input format

EpiCs calculates the p-value of the co-repartitions of the occurrences of a pair of events on a phylogenetic tree under a model of independence. It takes into account the topology of the tree and the location of the occurrences of each events on this tree.

The input format for EpiCs is a rooted phylogenetic tree in the Newick tree format, branch lengths included. We add the positions of events by inserting at each node of the tree a list of integers corresponding to the number of occurrences of each event on the corresponding branch. Each list has the following format:

$[n_1/n_2/\dots/n_m]$

where m is the number of different events and n_i , for $i = 1\dots m$ is the number of occurrences of the event E_i on the branch ($n_i = -1$ if no occurrence of E_i is located on the branch). m can be greater than 2: in this case, EpiCs will calculate the p-values for each pair of events.

For internal nodes, the list is inserted as a bootstrap (*i.e.* between the closing parenthesis and the colon preceding the branch length). For external nodes, it is directly put after the name of the leaf.

Example

$(A[2/-1/3] : 10, (B[2/1/3] : 5, C[1/2/-1] : 7) [4/3/-1] : 4) : 0$ corresponds to the following tree:

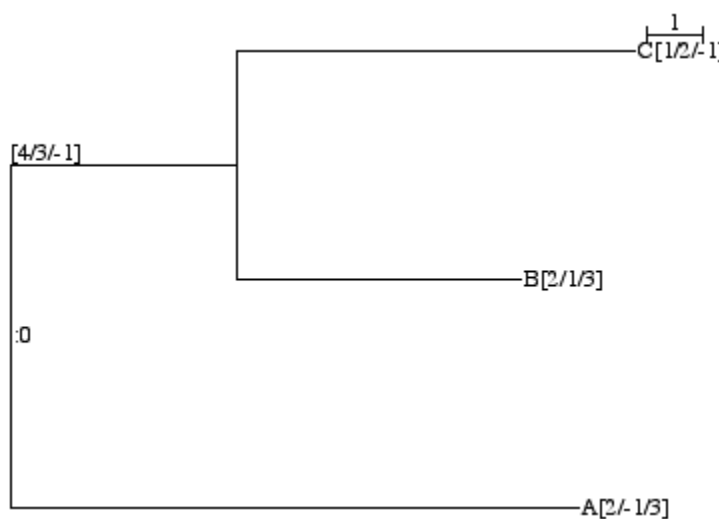


Figure 1: Example tree

In this example, the tree has 3 leaves A, B and C. To each (internal as external) node, a list is attached, each of those lists being of length 3, thus the positions of the occurrences of 3 events (here called E_1 , E_2 and E_3) are described. For example, the event E_1 has two occurrences on the external branch A, two on the external branch B, one on the branch C and finally four on the ancestral branch of the subtree (B, C). E_2 is not represented on the branch A.

3 Usage

3.1 Choice of the matrix

For a pair of events (E_1, E_2) , we distinguish two types of interaction between evolutionary events:

- *Coordinated pairs*, or *inseparable pairs*, when an E_1 event occurs in the same branch as an E_2 event.

- *Chronologically (or genealogically) ordered pairs*, when an E_2 event is in the subtree defined by the node under an E_1 event (*i.e.* the E_1 event precedes an E_2 event).

In particular, genealogically ordered pairs are ordered, whereas we can not state any order between two inseparable events.

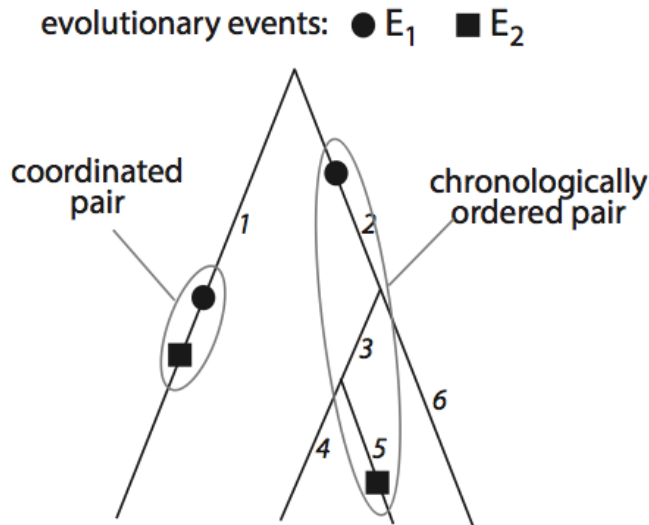


Figure 2: Example tree and the location of the occurrences of two events

By default, EpiCs only considers the genealogically ordered pairs, see the section Options below for more details about testing inseparable pairs or both inseparable and genealogically ordered pairs.

3.2 Running EpiCs

To run EpiCs, simply type:

```
./epics [OPTIONS]... < TREE_FILE
```

Options are the following:

- I: inseparable pairs only will be computed (Identity matrix)
- S: genealogically ordered pairs only will be computed (S matrix)
- B: (both) inseparable and genealogically ordered pairs will be computed (S+Id matrix)

(default: *-S* option is used, chronologies only will be computed (*S* matrix))

- b: outputs branches number, length and associated probabilities
- d: (distribution) output on stderr the chronologies and co-occurrences probabilities distribution
- m: (matrix) output on stderr (the *S* or *S+Id* matrix)
- t: (tree) output the tree on stderr
- e: (e1,e2) output on stderr the e1 and e2 vectors (of mutations)
- v: (verbose) outputs running progression of the program
- a: select all options b,d,m,t,e

4 Output format

By default, the standard output reports, for each pair of events (E_i, E_j) , the number of interactions counted (*i.e.* number of inseparable pairs, genealogically ordered pairs, or both) and the associated p-value, for both orders $E_1 \rightarrow E_2$ and $E_2 \rightarrow E_1$.

In the case of more than two events, the standard output will list all the possible pairs, reporting the counts and the associated p-values.

5 Contact

To ask a question on EpiCs, report a suggestion (e.g. why not including other options) or if you think you have discovered a bug (if any ?), please contact:

Abdelkader Behdenna at kaderbehdenna@gmail.com.

6 Citing EpiCs

A. Behdenna, J. Pothier, S. Abby, A. Lambert and G. Achaz, Testing for independence between evolutionary processes (submitted).