

Christian Zmasek

Publications

Software

Archaeopteryx

applets manual

documentation

examples

amphibians

apaf1

archaeopteryx_a

gsdi

gsdi_a

mollusca taxonomy

msa_display

multiple support

ncbi_taxonomy

signed_applets

simple_a

simple_e

species images

stripped

visual_styles

wnt

test

Test

version history

forester

confadd

decorator

GSDI

msa_compactor

NHX

pccx

phyloxml converter

RIO

surfacing

Sitemap

GSDI: Generalized Speciation Duplication Inference

Purpose

To infer duplication events on a gene tree given a trusted species tree.

Download

Current version: 1.001 (2013-03-25)

- [forester.jar](#)
- Source code is available through Google code at: <http://code.google.com/p/forester/>

Usage

```
% java -Xmx1024m -cp path/to/forester.jar  
org.forester.application.gsdi [-options] <gene  
tree in phyloXML format> <species tree> <outfile>
```

Options

- -g: to allow stripping of gene tree nodes without a matching species in the species tree
- -m: use most parsimonious duplication model for GSDI: assign nodes as speciations which would otherwise be assigned as potential duplications due to polytomies in the species tree
- -q: to allow species tree in other formats than phyloXML (i.e. Newick, NHX, Nexus)
- -b: to use SDIse algorithm instead of GSDI algorithm (for binary species trees)

Gene tree

Must be in phyloXML format, with taxonomy and sequence data in appropriate fields ([example](#)).

Species tree

Must be in phyloXML format unless option -q is used ([example](#)).

Output

Besides the main output of a gene tree with duplications and speciations assigned to all of its internal nodes, this program also produces the following:

- a log file, ending in "_gsdi_log.txt" ([example](#))
- a species tree file which only contains external nodes with were needed for the reconciliation, ending in "_species_tree_used.xml"
- if the gene tree contains species with scientific species names such as "Pyrococcus horikoshii strain ATCC 700860" and if a mapping cannot be establish based on these, GSDI will attempt to map by removing the "strain" (or "subspecies") information, these will be listed in a file ending in "_gsdi_remapped.txt".

Taxonomic mapping between gene and species tree

GSDI can establish a taxonomic mapping between gene and species tree on the following three data fields

- scientific names (e.g. "Pyrococcus horikoshii")
- taxonomic identifiers (e.g. "35932" from uniprot or ncbi)
- taxonomy codes (e.g. "PYRHO")

Example

```
% gsdi -g -q gene_tree.xml tree_of_life.nwk
out.xml
```

Example files

- [gene tree](#)
- [species tree](#)
- [log file \(output\)](#)

References

- Zmasek CM and Eddy SR "A simple algorithm to infer gene duplication and speciation events on a gene tree" [Bioinformatics](#), 17, 821-828
- Zmasek CM and Eddy SR "RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs" [BMC Bioinformatics](#) 2002, 3:14
- Han M and Zmasek CM "phyloXML: XML for evolutionary biology and comparative genomics" [BMC Bioinformatics](#) 2009, 10:356

Last modified: 2015-01-21

