

GENETREE 1.0

A program for comparing gene and species phylogenies
on computers running Mac OS or Microsoft® Windows™ 95/NT

User's Guide

by

Roderic D. M. Page
University of Glasgow

Permission to use and distribute this software and its documentation for any purpose is hereby granted without fee, provided the above copyright notice, author statement and this permission notice appear in all copies of this software and related documentation.

THE SOFTWARE IS PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EXPRESS, IMPLIED OR OTHERWISE, INCLUDING WITHOUT LIMITATION, ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL THE AUTHOR, THE DIVISION OF ENVIRONMENTAL AND EVOLUTIONARY BIOLOGY OR THE UNIVERSITY OF GLASGOW BE LIABLE FOR ANY SPECIAL, INCIDENTAL, INDIRECT OR CONSEQUENTIAL DAMAGES OF ANY KIND, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER OR NOT ADVISED OF THE POSSIBILITY OF DAMAGE, AND ON ANY THEORY OF LIABILITY, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

Contents

1. PREFACE.....	5
1.1 CONTACTING THE AUTHOR.....	5
2. INTRODUCTION.....	6
2.1 TERMINOLOGY.....	6
2.2 RECONCILED TREES.....	7
2.2.1 <i>Searching for optimal species trees.....</i>	7
2.2.2 <i>Missing genes.....</i>	8
3. USING THE PROGRAM.....	9
3.1 OBTAINING GENETREE.....	9
3.2 INSTALLATION.....	9
3.2.1 <i>Mac OS.....</i>	9
3.2.2 <i>Windows.....</i>	9
3.3 INTERNET ACCESS.....	9
3.4 FILE FORMAT.....	9
3.5 GETTING DATA FROM THE INTERNET.....	10
3.5.1 <i>Using the Bork BLAST server.....</i>	10
3.6 USER INTERFACE.....	12
3.6.1 <i>Log window.....</i>	12
3.6.2 <i>Text editor.....</i>	13
3.6.3 <i>Main window.....</i>	13
3.6.4 <i>Reconciled tree windows.....</i>	14
3.6.5 <i>Graphical output.....</i>	14
3.6.6 <i>Tree editor.....</i>	14
3.6.7 <i>Online help.....</i>	14
4. COMPARING GENE AND SPECIES TREES.....	15
4.1 OPTIMALITY CRITERIA.....	15
4.1.1 <i>Duplications and losses.....</i>	15
4.1.2 <i>Duplications only.....</i>	15
4.1.3 <i>Deep coalescences.....</i>	15
4.2 SPECIES TREE COST.....	16
4.3 RANDOM TREES.....	16
4.4 HEURISTIC SEARCH.....	16
4.4.1 <i>Starting trees.....</i>	17
4.4.2 <i>Tree perturbation.....</i>	17
4.4.3 <i>Constraints.....</i>	17
4.4.4 <i>Tree storage.....</i>	17
4.4.5 <i>Deciding when to accept an improvement.....</i>	17
4.4.6 <i>Results.....</i>	18
4.4.7 <i>Excluding blocks from the analysis.....</i>	18
4.5 SHOW SPECIES TREE COSTS.....	18
4.6 SHOW GENE TREE COSTS.....	18
4.7 PRINT SPECIES TREES.....	18
4.8 SAVE SPECIES TREES.....	19
4.9 EDIT CONSTRAINT.....	19
4.10 SHOW CONSTRAINT.....	19
5. REFERENCES.....	20
6. APPENDICES.....	21
6.1 ACCESSING YOUR WEB BROWSER FROM WITHIN GENETREE.....	21
6.1.1 <i>Windows95/NT 4.0.....</i>	21

6.1.2 <i>Mac OS</i>	21
6.2 TREE EDITOR	22
6.2.1 <i>Using the tools</i>	22
6.2.2 <i>Undoing a tree manipulation</i>	23
6.3 GENETREE AND COMPONENT 2.0.....	24

1. Preface

GENETREE has its origins as part of my tree comparison program COMPONENT (Page, 1993), written while I was a post doc at the Natural History Museum in London. However, the code has been completely rewritten in C++ (using Borland C++ and Metrowerks CodeWarrior), and has been ported to the Macintosh. Writing the cross-platform interface to the program was made easier by using ideas from the National Center for Biotechnology's VIBRANT library written by Jonathon Kans, and Julian Smart's wxWin C++ library. Implementing tabbed dialog boxes and other features was made easier by public domain code from Jim Stout (Jim's CDEFs) and Gilles Volland (MRU library). The installation program for Windows 95/NT comes courtesy of InstallShield Corporation. The original impetus for writing GENETREE came from work by Joe Slowinski and Boris Mirkin. The algorithm for constructing a reconciled tree is described in Page (1994) . The heuristic search strategy was influenced by many discussions with Mike Charleston (Charleston, 1995), and the program uses Oliver Eulenstein's unpublished algorithm for computing the cost of mapping a gene tree onto a species in linear time.

1.1 Contacting the author

Roderic D. M. Page
Division of Environmental and Evolutionary Biology
Institute of Biomedical and Life Sciences
University of Glasgow
Glasgow G12 8QQ
Scotland, United Kingdom

Telephone: (0141) 330 4778
Fax: (0141) 330 5971
Email: r.page@bio.gla.ac.uk

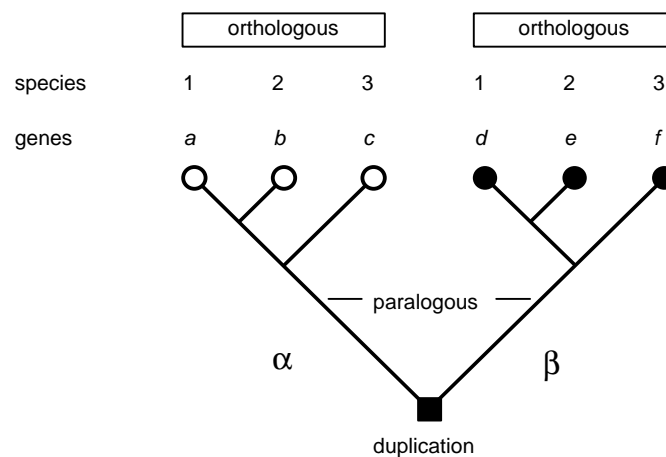
2. Introduction

GENETREE is an experimental tool for visualising and analysing the relationship between gene and species trees using reconciled trees. If you are interested in using GENETREE I strongly suggest that you read the papers by Page and Charleston (1997a; 1997b) which describe the concepts behind the program. Preprints of these papers can be downloaded from my Web site (<http://taxonomy.zoology.gla.ac.uk/rod/pubs.html>). A quick overview of reconciled trees is provided below.

2.1 Terminology

Some terminology used in this manual is defined below.

- Duplication** May literally be a gene duplication resulting in paralogous copies of the gene, but is also used here to describe cladogenesis of genes independent of speciation such that a species has more than one copy of the gene. Hence, “duplication” is also used here to describe interspecific coalescence of alleles.
- Orthologous** A pair of genes that are descendants of the same copy of a gene are orthologous.
- Paralogous** A pair of genes separated by at least one gene duplication are paralogous.

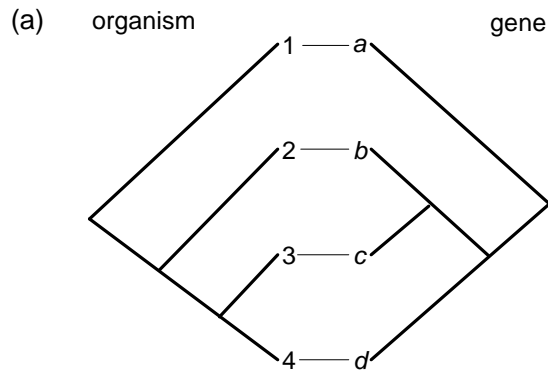


Phylogeny for three species 1-3 and six (*a-f*) genes that stem from a gene duplication resulting in two paralogous clades of genes, α and β . The α genes *a-c* are orthologous with each other, as are the β genes *d-f*, however any α and any β gene is paralogous as they are separated by a gene duplication.

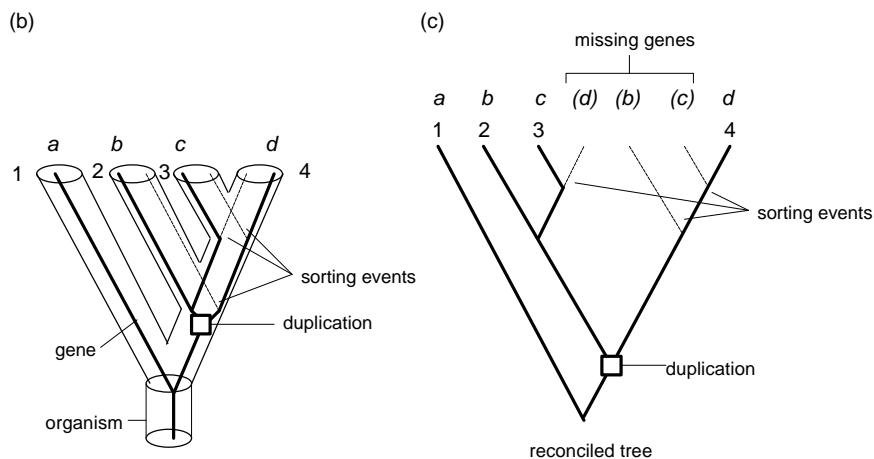
- Reconciled tree** The simplest embedding of an associate tree inside its host tree.
- Sorting event** The (apparent) absence of a gene in the descendants of a species which itself had that gene. Sorting events include gene loss, but may also simply reflect failure to detect a copy of the gene in a species.
- Tanglegram** A diagram showing a species and gene tree facing each other with pairs of species and genes linked.

2.2 Reconciled trees

The concept of reconciled trees dates from Goodman *et al.*'s (1979) attempts to reconcile disagreements between then accepted mammalian evolutionary relationships and those obtained from haemoglobin genes. Largely neglected until recently, reconciled trees are now receiving renewed attention from biologists and mathematicians (Eulenstein *et al.*, 1997; Guigó *et al.*, 1996; Mirkin *et al.*, 1995; Page and Charleston, 1997a; Page and Charleston, 1997b; Yuan *et al.*, in press). Suppose we have a phylogeny for four species, and four genes sampled from those species, and that the two trees, which we believe to be correct, disagree (a).



The question is, how can the trees both be true, and yet be discordant? One explanation is to embed the gene tree in the species tree (b), which requires us to postulate a number of gene duplications and subsequent losses (in this instance one duplication and three losses). This embedding can also be represented using a reconciled tree (c), which simply takes the embedded gene tree and “unfolds” it so that it lies flat on the page. The reconciled tree would depict the complete history of the gene if there had been no gene losses (the three sorting events). As a consequence of the gene duplication we would expect species 2, 3 and 4 to each have two copies of the gene. That they do not requires us to postulate three gene losses. Alternatively, the gene copies may be present but undetected. Hence the reconciled tree makes predictions about the existence of undiscovered genes. It also suggests that genes *b* and *c* are paralogous to gene *d*, which is not apparent from the gene phylogeny alone.



2.2.1 Searching for optimal species trees

Reconciled trees can be used to find the best species tree for more than one gene. Given n gene trees G_1, G_2, \dots, G_n , and species tree T , the cost of reconciling all n gene trees is simply the sum of the costs of reconciling the individual gene trees, $c(G_1, T) + c(G_2, T) + \dots + c(G_n, T)$. An analogy that may be helpful when using GENETREE is to think of each gene as a character, each terminal sequence

as a character state, and the gene tree is the character state tree. Mapping the gene tree onto the species tree is equivalent to optimising a character on a tree. The cost of the reconciled tree is equivalent to the “length” or number of steps a character requires.

2.2.2 Missing genes

If we are comparing more than one gene tree with a species tree then it will often be the case that not all the genes are known in all the species of interest. While the algorithms for mapping two trees and for constructing the reconciled tree are still applicable in this case, the number of losses computed needs to be interpreted carefully. For example, in the algorithm described above a gene tree for four genes may be perfectly consistent with a larger species tree (on $n > 4$ species say), but the lack of genes in the remaining $(n - 4)$ of those species will be counted as losses. Given the uneven taxonomic sampling in the sequence data bases a more reasonable interpretation may be that these species simply have not been sequenced for that gene. This is the approach adopted in GENETREE. For more details see Page and Charleston (1997a; 1997b).

3. Using the program

This section gives an overview of how to use GENETREE. Topics covered include installation, configuration, file formats, using the Bork BLAST server, and more.

3.1 Obtaining GENETREE

GeneTree can be freely downloaded from the Internet by visiting <http://taxonomy.zoology.gla.ac.uk/rod/genetree/genetree.html>. The program is free, but please register your copy by sending an email to r.page@bio.gla.ac.uk. This will enable me to keep you informed of bug fixes and updates.

3.2 Installation

3.2.1 Mac OS

The Macintosh version requires a Power Macintosh running System 7.5 or later, and the free Internet Config system extension. This extension stores information about your Internet preferences and applications, and GENETREE uses it to communicate with your Web browser. To install GENETREE simply double click the self extracting file **genetreeppc.sea**. You will need to install the font "COMPONENT Monaco" on your system. To do this, drag the COMPONENT Font File onto your System folder.



3.2.2 Windows

GENETREE is a 32 bit program that requires Windows 95 or NT 4.0. To install the program unzip the file **genetree.zip** and run the installation program **setup.exe**. This program installs GENETREE on your computer, and supports the Uninstall feature of Windows 95/NT so that you can remove GENETREE from your computer using the Add/Remove Programs control panel application.

3.3 Internet access

Some GENETREE features require access to the Internet. If properly configured, GENETREE can communicate directly with your Web browser. See the Appendices for details about configuring Internet access.

3.4 File format

GENETREE uses essentially the same variant of the NEXUS format (Maddison *et al.*, 1997) used by COMPONENT 2.0 (Page, 1993). The input file comprises a TAXA block, one or more DISTRIBUTION blocks, and a TREES block (unlike COMPONENT the TREES block **must** be included). You are referred to the COMPONENT manual (or that program's online help) for details. The basic syntax is:

```

#NEXUS

[species names]
BEGIN TAXA;
    DIMENSIONS NTAX=number-of-species;
    TAXLABELS species1 [species2 ...];
ENDBLOCK;

[Gene tree(s)]
BEGIN DISTRIBUTION;
    [TITLE = 'title-string'];
    NTAX = number-of-sequences;
    [For each gene sequence list the species it comes from]
    RANGE
        Sequence_1 : species_x,
        [sequence_2 : species_x,]
        sequence_ntax : species_x;
    [one or more trees for these sequences]
    TREE [*] tree-name = rooted-tree-specification;
    [TREE [*] tree-name = rooted-tree-specification;]
ENDBLOCK;

[Additional DISTRIBUTION blocks may go here...]

[Optional constraint on species tree]
[BEGIN ASSUMPTIONS;
    CONSTRAINT constraint-name = rooted-tree-specification;
ENDBLOCK;]

[ species tree descriptions go here]
BEGIN TREES;
    TREE [*] tree-name = rooted-tree-specification;
    [TREE [*] tree-name = rooted-tree-specification;]
ENDBLOCK;

```

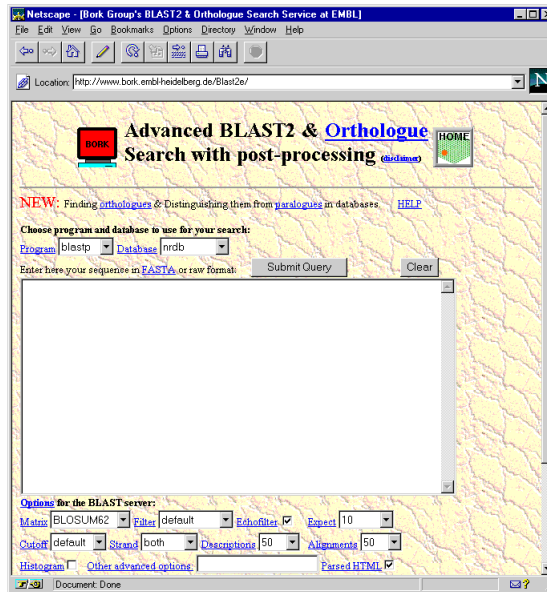
At present all trees must be rooted and binary (fully resolved).

3.5 Getting data from the Internet

As well as reading NEXUS files, GENETREE can read HTML files produced by the Bork group's advanced BLAST server. This server (<http://www.bork.embl-heidelberg.de/Blast2e/>) automates analysis of a set of sequences using reconciled trees. The server is experimental, but its output can be a useful starting point for further analysis.

3.5.1 Using the Bork BLAST server

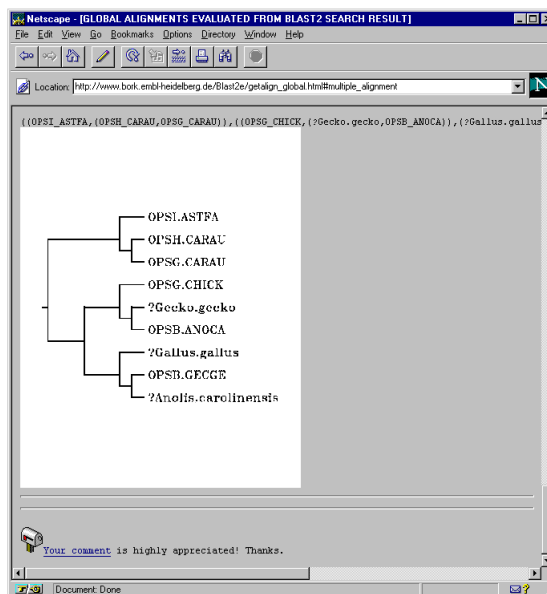
From within GENETREE you can go to the Bork BLAST server using **the Bork BLAST server** command on the **Internet** menu. Full details on using the Bork BLAST server are given in (Yuan *et al.*, in press) and in documentation at the site itself. The first step in the analysis is to paste in a query sequence and click on **Submit Query**.



This returns a page listing the sequences most similar to your query sequence in typical BLAST format. The next step in the analysis is to select the sequences you wish to analyse further. Once the sequences are selected, clicking on either the **Local Alignments** or **Global Alignments** button starts the following analysis:

- the sequences are aligned using ClustalW
- a neighbor-joining tree is constructed for the aligned sequences
- a species tree is constructed for the taxa the genes were obtained from using the current taxonomy of the species
- a reconciled tree is constructed for the gene and species trees

The results of the analysis are displayed in a single Web page, part of which appears below.



If you save this page to disk as an HTML file, you can import it into GENETREE using the **Import BLAST HTML** command on the **Internet** menu. This procedure can be a bit fraught for the following reasons:

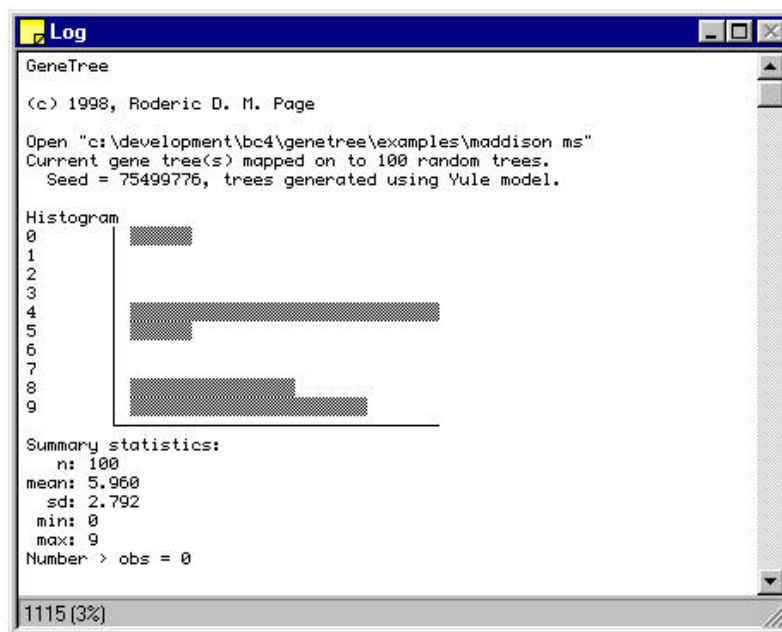
- The HTML file may be missing some information (for example, the species from which a sequence was obtained).
- The species tree may be incorrect and/or missing one or more taxa.
- The gene tree is not a fully resolved, rooted tree.

Because of this, GENETREE converts the HTML into a temporary NEXUS file which it displays in it's text editor. You can edit the tree descriptions by selecting them using the cursor, then choosing the command **Edit tree** from the **Edit** menu. This opens the tree in a tree editor. Once you've edited the tree, closing the tree editor replaces the original tree description with one corresponding to the edited tree. To process the file chose the **Execute** command on the **File** menu. If successful, GENETREE will display the trees; otherwise it will report an error and display its location in the edit window.

3.6 User Interface

The user interface to GeneTree includes a Log window to record results of analyses, text edit windows, and various tree and chart display windows. Note that each window has it's own specific menu commands, and that some commands are only available when the appropriate window is active.

3.6.1 Log window



The Log window displays the results of your analyses. The status bar at the bottom of the window displays the number of bytes in the buffer, and how full the buffer is. Currently the buffer is limited to 32K, and will not store any information when that limit is reached. You can use the **Clear** command on the **Edit** menu to clear the Log window buffer. The contents of the Log window can be saved to a text file using the **Save as text** command, and they can be copied to the Clipboard using the **Copy** command.

3.6.2 Text editor

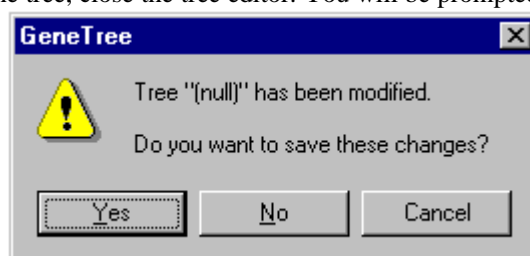
GENETREE has a simple text editor that can be used to create and edit data files. In addition to standard editing features such as copying and pasting text from the Clipboard, and search and replace functions, the editor also enables you to edit tree descriptions in situ. To do this:

- Select the tree description (including the semicolon) in the editor window

```
BEGIN TREES;
  [Species tree(s)]
  TREE * T1= ((Rabbit,(Human,(Pig,(Bovine,Sheep))), (Rat,Mouse));
ENDBLOCK;
```

- Choose the **Edit tree** command, or press **Ctrl+E** (Windows) or **⌘ E** (Macintosh). The program will open the tree description in the tree editor.

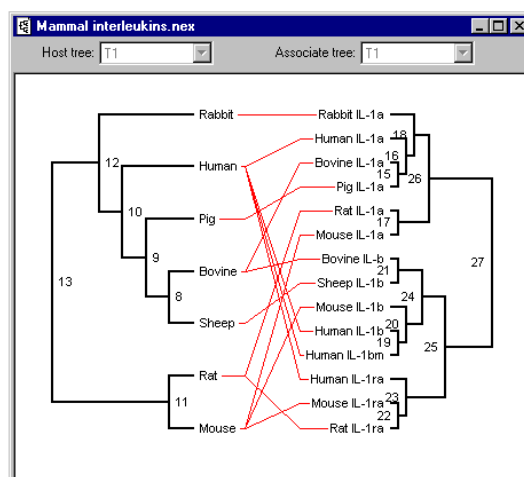
- Once you've edited the tree, close the tree editor. You will be prompted to save the tree



- If you click on the **Yes** button the tree description highlighted in the editor window will be replaced with the tree description corresponding to the edited tree.

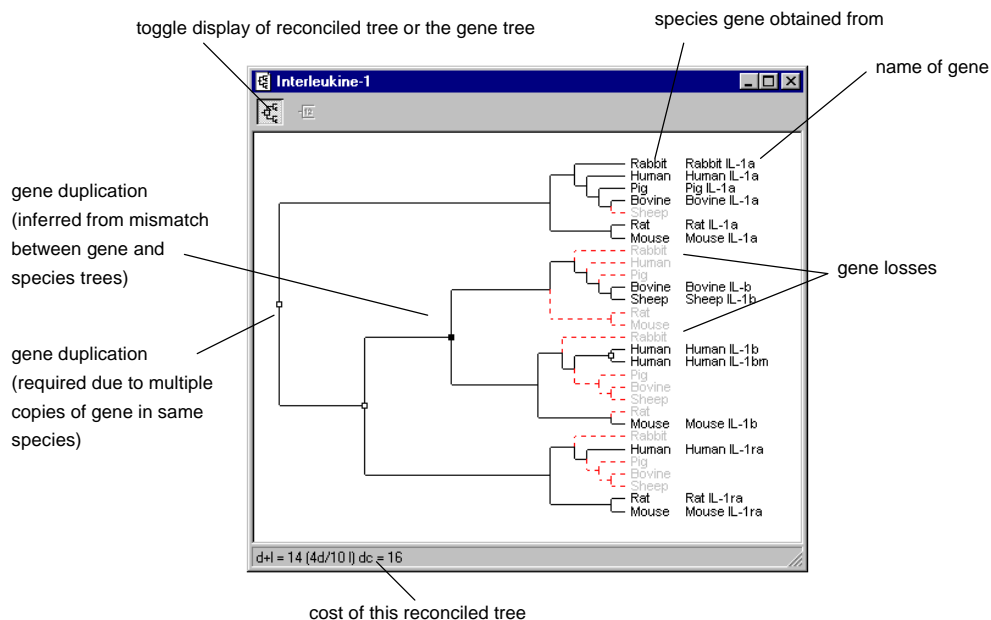
3.6.3 Main window

The main window displays a tanglegram showing the host tree and an associate tree. If you have more than one associate block you can change the block displayed in the tanglegram by making the corresponding block window active. The tanglegram window displays the total cost of all the reconciled trees in the active blocks. If you have more than one host or associate tree you can select alternative trees to display using the pop-up menus in the tool bar.



3.6.4 Reconciled tree windows

Each DISTRIBUTION block has a corresponding window which displays the reconciled tree for the current gene tree in that block and the current species tree.



The cost of the reconciled tree is displayed in the status bar in the form " $d+l=c1 (xd/yl) dc=c2$ ", where $c1$ is the sum of the number of duplications (x) and losses (y), and $c2$ is the number of deep coalescences. Hence, $d+l=14 (4d/10l) dc = 16$, means that the reconciled tree requires four gene duplications and ten gene losses, for a total cost of 14, and requires 16 deep coalescences.

If the DISTRIBUTION block contains a TITLE command then that text will appear in the title bar of the window. You can change the window title by using the **Rename** command. Note that if you have more than one gene tree in a DISTRIBUTION block the reconciled tree is computed for the gene tree that is currently displayed in the tanglegram window. To change the gene tree select the desired tree from the popup menu in the tanglegram window.


3.6.5 Graphical output

The tree(s) displayed in the tanglegram and reconciled tree windows in GENETREE can be printed, copied to the Clipboard, or saved as a graphics file. TO DO: printing big reconciled trees.

3.6.6 Tree editor

GENETREE has a built-in tree editor which can be used to edit tree descriptions in NEXUS text files, as well as construct constraint trees. The editor is described in more detail in the Appendices.

3.6.7 Online help

You can access the GENETREE online help through your Web browser using the **Help** command. On the Macintosh this command appears under the  menu, in Windows it appears on the **Help** menu. Choosing this command will launch your Web browser and display the help pages for GENETREE. This command requires that you have a properly configured Web browser and an Internet connection on your machine (see the Appendix for more details). The Macintosh version of the program also supports Balloon Help.

4. Comparing gene and species trees

GENETREE computes the cost between one or more gene trees and a species tree that you have supplied (or inferred using the program). The cost for each gene is displayed in the corresponding reconciled tree window. The total cost is displayed in the tanglegram window. Using the **Analysis** menu commands you can perform a range of further analyses. You can

- select the optimality criterion for scoring reconciled trees
- compute the total cost of all species trees
- compute the total cost for a set of random species trees
- perform a heuristic search for the species tree (or trees) that yield the lowest cost
- save any species trees you have computed
- construct constraint trees

4.1 Optimality criteria

The first three commands set the optimality criterion used for scoring reconciled trees.

4.1.1 Duplications and losses

The default optimality criterion is the number of duplications and losses required to reconcile the gene tree with the species tree.

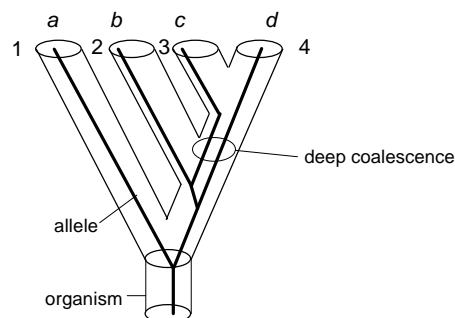
4.1.2 Duplications only

Choosing this criterion means that the number of gene losses is ignored. This may be appropriate if the sampling of different taxa is uneven (Page and Charleston, 1997a).

4.1.3 Deep coalescences

Counting duplications and losses is appropriate if gene duplication is the cause of multiple gene lineages. However, at lower taxonomic levels multiple sequences may be alleles rather than paralogous loci. GENETREE uses Maddison's (Maddison, 1997) concept of deep coalescence as a criterion of fit between allele phylogenies and species trees.

In the figure below, alleles c and d coexist in the common ancestor of species 3 and 4, but do not coalesce until we go down the tree to the common ancestor of species 2, 3, and 4. This failure of two allele lineages to coalesce in the species lineage in which they first co-occur counts as a single deep coalescence event, and hence the cost of the reconciled tree is 1.



GENETREE computes the deep coalescence cost of a reconciled tree by counting the number of gene lineages that do not coalesce on each branch of the species tree.

4.2 Species tree cost

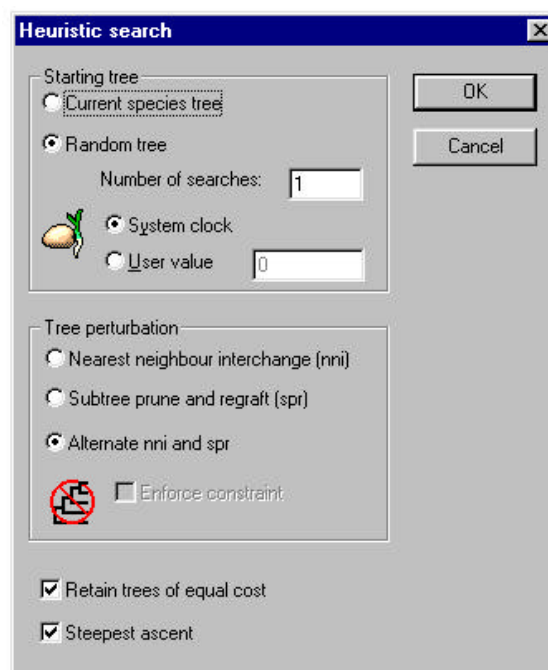
The Species Trees command allows you to compute the total cost of all species trees when reconciled with the current gene tree in each active DISTRIBUTION block. The resulting distribution is displayed in a chart window. The cost computed depends on which optimality criterion you have selected (see previous section).

4.3 Random trees

The **Random Species Trees** command allows you to compute the total cost for a set of random species trees. The current gene tree in each active DISTRIBUTION block is reconciled with each random species tree, and the resulting distribution of total cost is displayed in a chart window. The cost computed depends on which optimality criterion you have selected (see previous section).

4.4 Heuristic search

One use of GENETREE is to find the species tree that the gene trees best fit (i.e., the species tree with the smallest total cost when reconciled with all gene trees). This tree is most parsimonious in the sense of minimising the total number of events required to explain the evolutionary history of the genes. The problem of finding the optimal species tree is NP-complete, so GENETREE uses a heuristic approach. There are a number of options available which affect how the search is undertaken, and these are set in the Heuristic search dialog box:



4.4.1 Starting trees

The search can either start from the current species tree, or from one or more random trees (the later is the default). Typically a search should use a number of random starting points, however there may be times that you want to explore the landscape around a species tree. For example, after doing a search without retaining equally good solutions, you may wish to see whether more than one tree with that cost exists.

Random starting trees tend to be poor estimates of the optimal species tree, but using multiple random trees gives information on the search landscape of the problem (Charleston, 1995). In particular, convergence on the same cost value from multiple starting points suggest that that cost value may be optimal. Multiple starting points also increase the chance of detecting multiple, equally good solutions.

4.4.2 Tree perturbation

GENETREE offers three kinds of tree perturbations or “branch (Swofford and Olsen, 1990): nearest neighbour interchanges (nni), subtree pruning and regrafting (spr) and alternate nni's and spr's. Nni's offer quick but rather limited perturbation, whereas spr is much more extensive and, as a consequence, more time consuming. The alternate strategy (alt) combines both nni and spr by first trying nni's then, if no nni produces an improvement, sprs are tried.

In practice, a single run of nni's from a random starting tree is not very effective, although multiple runs may encounter the optimal solution. Spr tends to be time consuming but effective. Alt tends to be quicker than spr and as effective (Page and Charleston, 1997a), and hence is the default.

4.4.3 Constraints

In some cases it may be unnecessary to search the space of all possible species trees. For example, if we assume that mammals are monophyletic then species trees in which mammals do not form a clade can be safely ignored. GENETREE uses a constraint tree (Constantinescu and Sankoff, 1986) to represent the set of possible species trees. If you create a constraint tree and check the **Enforce constraint** check box, then GENETREE will only consider species trees that are compatible with the constraint tree.

4.4.4 Tree storage

You can elect to store all the species trees of equal cost that are found by GENETREE (**up to the program's limit of**), or just one. If you chose the former (by ticking the **Retain trees of equal cost** check box) and GENETREE finds more than one tree, each tree in turn will act as the starting tree for a heuristic search.

4.4.5 Deciding when to accept an improvement

GENETREE takes a tree and starts to perturb that tree in search of a better tree. If a better tree is found, the search begins again using the better tree, until no perturbation produces an improvement. If a perturbation produces an improvement, we could chose to either (a) wait until all perturbations have been made and use the best tree found as the start of the next round of perturbations, or (b) immediately restart the search using the better tree. Option (a) means that we will always be using the best possible tree as the starting point for the search, at the cost of the time to perform all rearrangements on each tree, whereas option (b) will almost always be quicker to start a new round of perturbations (the first tree to show an improvement is accepted), although an even better tree may have been obtained had we waited. Option (a) is in effect when the **Steepest ascent** check box is ticked, otherwise option (b) is employed.

4.4.6 Results

The species trees found by heuristic search replace those already in memory (for example, those read in from the data file). You can save the species trees to a NEXUS tree file using the **Save Species Trees** command.

4.4.7 Excluding blocks from the analysis

If you have more than one set of genes (i.e., more than one DISTRIBUTION block) then you may wish to analyse some of the genes separately. By default, all the **Analysis** menu commands operate on all the genes together. You can change this by using the **Include/Exclude** command. By moving associates between the Include and Exclude lists you can specify which associates GENETREE uses when computing the total cost.

4.5 Show species tree costs

For each species tree

This command shows the cost of reconciling the current gene tree for current DISTRIBUTION block gene with all the species trees. All the optimality criteria (duplications, losses, and deep coalescence) are computed.

```
Species tree   d/l   dc
-----
      1         0/0   0
      2         1/3   1
      3         1/3   1
      4         1/3   1
      5         1/4   2
      6         1/4   2
      7         2/7   3
      8         2/7   3
      9         2/7   3
     10         2/7   3
     11         2/7   3
     12         2/7   3
     13         1/3   1
     14         2/6   2
     15         2/6   2
-----
```

d/l = duplications and losses
dc = deep coalescences

4.6 Show gene tree costs

This command shows the cost of reconciling the current gene tree for each DISTRIBUTION block with the current species tree. All the optimality criteria (duplications, losses, and deep coalescences) are computed.

4.7 Print species trees

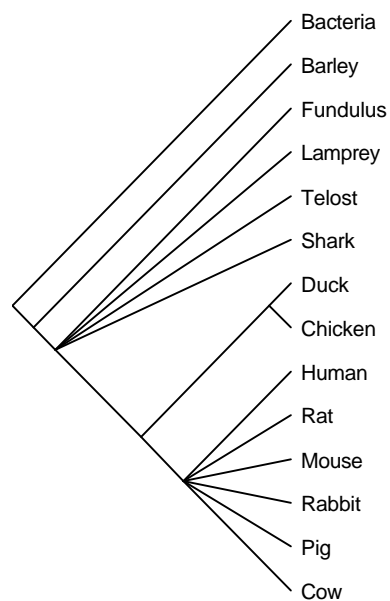
Prints the species trees. You can print more than one tree per page.

4.8 Save species trees

Saves the species trees to a NEXUS format tree file.

4.9 Edit constraint

Opens the tree editor (see Appendix) and displays the current constraint tree. If you have not defined a constraint tree, the current species tree will be used. By editing this tree you can prevent the heuristic search from considering certain trees. For example, the tree below constrains the heuristic search to consider only species trees that have mammals, birds, amniotes, and eukaryotes as clades.



4.10 Show constraint

Displays the current constraint tree (if any) in the Log window.

5. References

- Charleston, M. A. (1995). Towards a characterization of landscapes of combinatorial optimisation problems, with special reference to the phylogeny problem. *J. Comput. Biol.* **2**: 439-450.
- Constantinescu, M. and Sankoff, D. (1986). Tree enumeration modulo a consensus. *J. Classif* **3**: 349-356.
- Eulenstein, O., Mirkin, B. and Vingron, M. (1997). Comparison of annotating duplications, tree mapping, and copying as methods to compare gene trees with species trees. In "Mathematical Hierarchies in Biology" (B. Mirkin, F. R. McMorris, F. S. Roberts and A. Rzhetsky, Eds.), Vol. 37, pp. 71-93, American Mathematical Society, Providence, Rhode Island.
- Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E. and Matsuda, G. (1979). Fitting the gene lineage into its species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.* **28**: 132-168.
- Guigó, R., Muchnik, I. and Smith, T. F. (1996). Reconstruction of ancient molecular phylogeny. *Mol. Phylog. Evol.* **6**: 189-213.
- Maddison, D. R., Swofford, D. L. and Maddison, W. P. (1997). NEXUS: An extensible file format for systematic information. *Syst. Biol.* **46**: 590-621.
- Maddison, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46**: 523-536.
- Mirkin, B., Muchnik, I. and Smith, T. F. (1995). A biologically consistent model for comparing molecular phylogenies. *J. Comput. Biol.* **2**: 493-507.
- Page, R. D. M. (1993). "COMPONENT, Tree comparison software for Microsoft® Windows™," Version 2.0, The Natural History Museum, London.
- Page, R. D. M. (1994). Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* **43**: 58-77.
- Page, R. D. M. and Charleston, M. A. (1997a). From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol. Phylog. Evol.* **7**: 231-240.
- Page, R. D. M. and Charleston, M. A. (1997b). Reconciled trees and incongruent gene and species trees. In "Mathematical Hierarchies in Biology" (B. Mirkin, F. R. McMorris, F. S. Roberts and A. Rzhetsky, Eds.), Vol. 37, pp. 57-70, American Mathematical Society, Providence, Rhode Island.
- Swofford, D. L. and Olsen, G. J. (1990). Phylogeny reconstruction. In "Molecular systematics" (D. M. Hillis and C. Moritz, Eds.), pp. 411-501, Sinauer Associates, Sunderland.
- Yuan, Y. P., Eulenstein, O., Vingron, M. and Bork, P. (in press). Towards detection of orthologues in sequence databases. *Comput. Applic. Biosci.*

6. Appendices

6.1 Accessing your Web browser from within GENETREE

The online help and Bork server access features of GENETREE require that GENETREE can locate and talk to your Web browser program.

6.1.1 Windows95/NT 4.0

Under Windows 95/NT 4.0 this should be automatic: GENETREE uses the Windows registry to locate the browser. If this fails, then it is possible that the browser is not correctly installed. If this is not the case, contact me to resolve the problem.

6.1.2 Mac OS

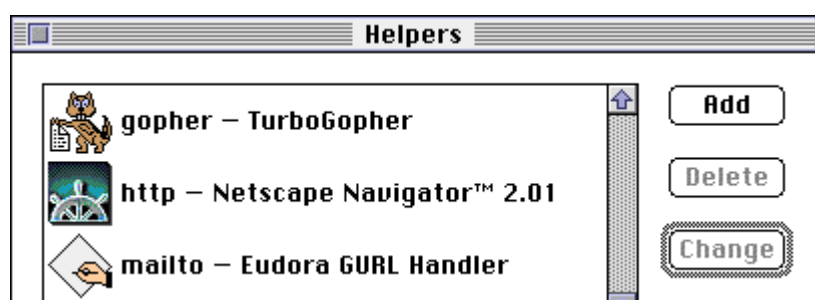
You must have Internet Config installed and configured on your machine in order for GENETREE to be able to talk to your Web browser. Internet Config is a freeware system extension that allows applications to share the same preferences. You can add your e-mail address, news information, and Helper configuration information into Internet Config once, instead of in each of your Internet applications. Applications that are Internet Config aware include Microsoft Internet Explorer, among others. You can download Internet Config 1.3 from most Mac OS software collections, or from the Glasgow Taxonomy Server (<http://taxonomy.zoology.gla.ac.uk/rod/treeview/help/InternetConfig1.3.sea.hqx>).

Once you have Internet Config installed you need to ensure that your Web browser is listed as your “http” helper in Internet Config's settings. To do this follow these steps:

1. Click on the **Helpers** button in the Internet Preferences dialog.



2. In the Helpers dialog box set the http helper to your browser by clicking the **Add** button. Once you've added your browser you should see something like this:



3. Quit Internet Config, remembering to save your settings.

When you start GENETREE the program will use Internet Config's settings to locate your Web browser, and will communicate with your browser via the Internet Config extension.

6.2 Tree Editor

The tree editor provides tools for moving branches, rerooting, polytomy formation, and rearranging the appearance of the tree. The tools available are displayed in the tool bar:



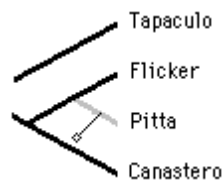
To select a tool click on the icon corresponding to the tool you want to use. Once you have selected a tool the cursor will change to the appropriate shape. To manipulate the tree click the cursor on a branch; the corresponding operation will be performed on that branch (unless the desired operation is invalid, in which case the tree editor gently “beeps” and ignores the command). The default tool is the arrow tool, which is used for moving branches.

6.2.1 Using the tools



Move branch

A branch can be picked up and moved to another part of the tree by clicking on the branch and dragging it to the new position. The branch being moved will be highlighted with a different colour, and the cursor will change from an arrow to a small diamond:



Drag the diamond cursor to the position you want to move the branch, and release the mouse button. If the new placement is invalid (for example, moving a branch onto one of its descendants) the tree editor will gently “beep” and ignore your request.



Collapse branch

Clicking with this tool on an internal branch deletes that branch producing a polytomy. The editor ignores attempts to delete terminal branches.



Collapse clade

Clicking with this tool on an internal branch collapses the clade above that branch



Reroot

Clicking on a branch with this tool reroots the tree on that branch.

The next four tools change the appearance of the tree in the window, but do not affect the cladistic topology. They cannot be undone.



Rotate branches

This tool rotates the descendants of a binary node, so that the left descendant is now the right descendant, and visa-versa. You can only use this tool on binary internal nodes. If you want to rearrange the descendants of a polytomy you must use the next tool.



Exchange branches

To rearrange the order of the descendants of a polytomy use this tool to interchange pairs of branches. Click on one descendant, drag the cursor onto the other descendant, and release the mouse button.



Ladderize left

This tool orders the subtree rooted at the branch clicked on so that the “heavier” nodes (those nodes with the largest clusters) appear leftmost (uppermost on the screen).



Ladderize right

This tool orders the subtree rooted at the branch clicked on so that the “heavier” nodes (those nodes with the largest clusters) appear rightmost (lowermost on the screen).

6.2.2 Undoing a tree manipulation

The tree editor remembers the previous tree, so that you can undo a change and recover the old tree by choosing **Undo** from the **Edit** menu, or by using these keys:

⌘ **Z** (MacOS)

Ctrl + Z (Windows)

Choosing the **Undo** command again restores the tree you just undid. **Undo** is only available for the operations that materially affect the tree topology: **move branch**, **collapse branch**, **collapse clade**, and **reroot**. Cosmetic changes (such as **rotate branches**) cannot be undone.

The **Undo** menu item will reflect which type of move can be undone; for example the menu item will read “**Undo move branch**” if you are using the move branch tool.

6.3 GeneTree and COMPONENT 2.0

This is intended for users of COMPONENT 2.0 (Page, 1993). Several commands in the **Map trees** submenu in COMPONENT have exact equivalents in GeneTree. These are listed below.

COMPONENT command	GENETREE equivalent
Reconcile with tree	GENETREE automatically displays the reconciled tree
Map onto all trees	Species tree fit
Options	GENETREE always treats species without genes as “missing” information
Heuristic search	Heuristic search

Some features of COMPONENT are not in GENETREE. These include the ability to arbitrarily resolve polytomies in input trees, and the **Prune each leaf** command.