

# NHX - New Hampshire eXtended version 2.0

Copyright (C) 1999 - 2009 by Christian M. Zmasek. Written by Christian M. Zmasek. Permission is granted to copy this document provided that this copyright notice is not removed and that the contents of this document are not altered in any way.

This document is available at: <http://www.phylosoft.org/NHX>

Notice. This (version 2.0) is expected to be the final version of NHX.

It is recommended to use **phyloXML** (<http://www.phyloxml.org>) instead of NHX.

**Archaeopteryx** (available at: <http://phylosoft.org/archaeopteryx>) is an example of an application program implementing the NHX format.

## Description

NHX is a format for describing annotated phylogenetic trees. NHX is based on the New Hampshire (NH) standard (also called "Newick tree format"). It has the following extensions (compared to NH as used in the Joseph Felsenstein's PHYLIP package):

- it introduces tags to associate various data fields with a node of a phylogenetic tree
- both internal and external nodes can be tagged
- arbitrary number of children per node
- the tree is assumed to be rooted if the deepest node is a bifurcation
- the order of the tags does not matter
- the length of all character string based data is unlimited (e.g. name, species)
- Comments between '[' and ']' are removed (unless the opening bracket is followed by "&&NHX")

In order to remain compatible with the NEXUS format, all fields except name and branch length (in other words, all fields extending NH) must be wrapped by "[&&NHX"and"]". For example: "ADH1:0.11[&&NHX:S=human:E=1.1.1.1]".

The following characters can not be part of names: ( ) [ ] , : ; as well as white space.

### An example of a (rooted) phylogenetic tree in NHX:

```
((((ADH2:0.1[&&NHX:S=human:E=1.1.1.1],  
ADH1:0.11[&&NHX:S=human:E=1.1.1.1]):0.05[&&NHX:S=Primates:E=1.1.1.1:D=Y:B=100],  
ADHY:0.1[&&NHX:S=nematode:E=1.1.1.1],  
ADHX:0.12[&&NHX:S=insect:E=1.1.1.1]):0.1[&&NHX:S=Metazoa:E=1.1.1.1:D=N],  
(ADH4:0.09[&&NHX:S=yeast:E=1.1.1.1],ADH3:0.13[&&NHX:S=yeast:E=1.1.1.1],  
ADH2:0.12[&&NHX:S=yeast:E=1.1.1.1],  
ADH1:0.11[&&NHX:S=yeast:E=1.1.1.1]):0.1[&&NHX:S=Fungi])[&&NHX:E=1.1.1.1:D=N];
```

## NHX version 2.0 elements

Element	Type	Description
no tag	string	name of this node/clade (MUST BE FIRST, IF ASSIGNED)
:	decimal	branch length to parent node (MUST BE SECOND, IF ASSIGNED)
:GN=	string	gene name
:AC=	string	sequence accession
:ND=	string	node identifier - if this is being used, it has to be unique within each phylogeny
:B=	decimal	confidence value for parent branch
:D=	'T', 'F', or '?'	'T' if this node represents a duplication event - 'F' if this node represents a speciation event, '?' if this node represents an unknown event (D= tag should be replaced by Ev= tag)
: Ev=duplications>specia tions>gene losses>event type>duplication type	int int int string string	event (replaces the =D tag), number of duplication, speciation, and gene loss events, type of event (transfer, fusion, root, unknown, other, speciation_duplication_loss, unassigned)
:E=	string	EC number at this node
:Fu=	string	function at this node
:DS=protein- length>from>to>suppor t>name>from>...	int int int double string int ...	domain structure at this node
:S=	string	species name of the species/phylum at this node
:T=	integer	taxonomy ID of the species/phylum at this node
:W=	integer	width of parent branch
:C=rrr.ggg.bbb	integer.integer.integer	color of parent branch
:Co=	'Y' or 'N'	collapse this node when drawing the tree (default is not to collapse)
:XB=	string	custom data associated with a branch
:XN=	string	custom data associated with a node
:O=	integer	orthologous to this external node
:SN=	integer	subtree neighbors
:SO=	integer	super orthologous (no duplications on paths) to this external node