The New Hampshire X Format (NHX)

This document is available at: http://www.genetics.wustl.edu/eddy/forester/NHX.html

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 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
- number of children per node is at least two (allows polytomous trees)
- the tree is assumed to be rooted if the deepest node is a bifurcation
- the order of the tags does not matter, with the exception that the sequence name must be first (if assigned)
- the length of all character string based data is unlimited (name, species, EC number)
- 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 sequence name and branch length (in other words, all fields eXtending NH) must be wrapped by "[&&NHX" and "]". E.g. "ADH1:0.11[&&NHX:S=human:E=1.1.1.1]".

Remark. Currently, ATV and FORESTER can still read files which lack these brackets, but this is deprecated.

In contrast to its name, NHX also has restrictions compared to Felsenstein's definition of the NH format:
"Empty" nodes are not allowed (e.g. "(,(,),)" is not acceptable).
The following characters can not be part of names: '(' ')' '[' ']' ',' ':' as well as white spaces.

The tags are as follows.

TAG	**VALUE**	**MEANING**
no tag	String	sequence name of this node (MUST BE FIRST, IF ASSIGNED)
:	double	branch length to parent node (MUST BE SECOND, IF ASSIGNED)
:B=	integer	bootstrap value at this node (does not apply to external nodes)
:S=	String	species name of the species/phylum at this node
:T=	integer	NCBI taxonomy ID of the species/phylum at this node
:E=	String	EC number at this node
:D=	'Y' or 'N'	‘Y’ if this node represents a duplication event – ‘N’ if this node represents a speciation event (does not apply to ext nodes)
:O=	integer	orthologous to this external node
:SO=	integer	"super orthologous" (no duplications on paths) to this external node
:L=	float	log likelihood value on parent branch
:Sw=	'Y' or 'N'	placing a subtree on the parent branch of this node makes the tree significantly worse according to Kishino/Hasegawa test (or similar)
:Co=	'Y' or 'N'	collapse this node when drawing the tree (default is not to collapse)

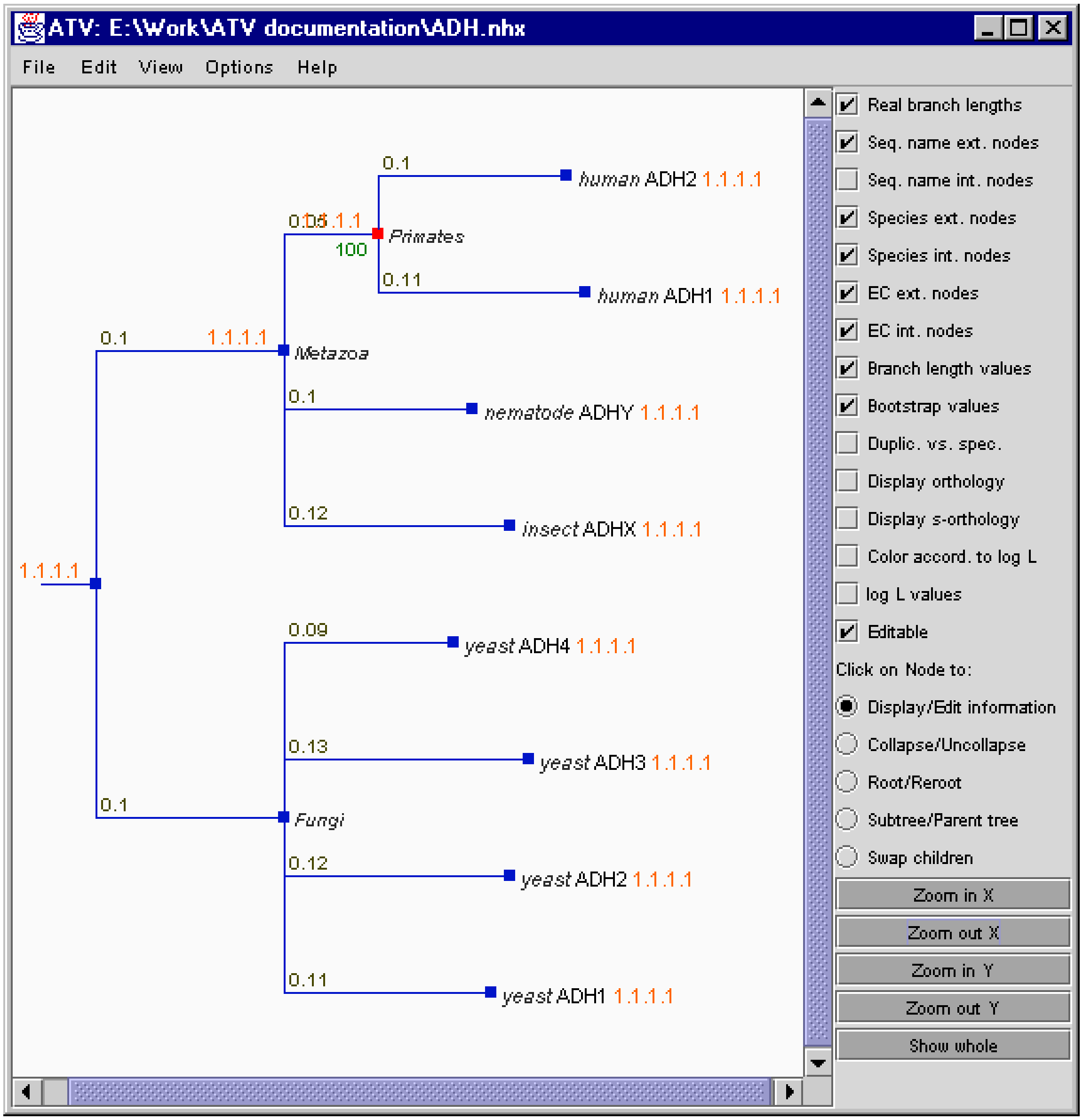
In Java, the data types are defined as follows:

String:	character string of arbitrary length
double:	64bit signed floating point number
float:	32bit signed floating point number
integer:	32bit signed integer number

An example of a (rooted) 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];
```

This tree would look as follows in ATV:



☑ Real branch lengths
☑ Seq. name ext. nodes
□ Seq. name int. nodes
☑ Species ext. nodes
☑ Species int. nodes
☑ EC ext. nodes
☑ EC int. nodes
☑ Branch length values
☑ Bootstrap values
□ Duplic. vs. spec.
□ Display orthology
□ Display s-orthology
□ Color accord. to log L
□ log L values
☑ Editable

Click on Node to:

● Display/Edit information
○ Collapse/Uncollapse
○ Root/Reroot
○ Subtree/Parent tree
○ Swap children