Outline of the course

- Introduction about Compara
  - Resources
  - API

- Base objects
  - Genes, peptides, RNAs
  - Multiple / pairwise alignments

- Data objects
  - Families
  - Gene trees
  - Homologies
GeneTree example on the website

Tree

Multiple alignment
Protein-Tree pipeline overview

All e! genes – canonical prot.

BLAST

hcluster_sg

MCoffee: MSA

TreeBeST: (+ reconciliation)

Ortholog/Paralog inference

Vilella et al., Genome Res. 2009
ncRNA-Tree pipeline overview

All e! ncRNA genes

Grouped in Family Models - RFAM

Infernal alignment + RaxML trees

PRANK alignment + NJ/ML trees

TreeBeST (tree reconciliation)

Ortholog/Paralog inference

Pignatelli et al., in preparation
GeneTree object / GeneTreeAdaptor

- Represents a set of members, in a phylogenetic tree

```perl
$genetree_adaptor->fetch_by_stable_id(...)
$genetree_adaptor->fetch_default_for_Member(...)
```

- `fetch_all*` methods require some more arguments:
  - `clusterset_id => 'default'
  - `tree_type => 'tree'
  - `member_type => 'protein' or 'ncrna'

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment</td>
<td><code>$tree-&gt;get_SimpleAlign()</code></td>
</tr>
<tr>
<td>Stable ID</td>
<td><code>$tree-&gt;stable_id()</code></td>
</tr>
<tr>
<td>Tree export</td>
<td><code>$tree-&gt;newick_format('simple')</code></td>
</tr>
<tr>
<td></td>
<td><code>$tree-&gt;nhx_format('full')</code></td>
</tr>
<tr>
<td></td>
<td><code>$tree-&gt;print_tree()</code></td>
</tr>
</tbody>
</table>
GeneTreeNode object

The actual tree structure is a hierarchy of GeneTreeNode objects.
**GeneTreeNode** object

The actual tree structure is a hierarchy of **GeneTreeNode** objects.

Additional information is stored with "tags"

```php
$node->get_all_tags()
$node->get_tagvalue('node_type') or 'taxon_name', 'bootstrap'
```
Exercises – Protein and ncRNA trees

- Print the protein tree with the stable id ENSGT00390000003602

- Print all the members of the tree containing the human ncRNA gene ENSG00000238344

- Count the number of duplication events in the tree of the zebrafish protein-coding gene ENSDARG000000003399