

## Summary

TreeMaker is a python program to convert a text-based classification schema into a Newick file for use in phylogenetic and bioinformatic programs.

Often research in linguistics or cultural evolution produces tree taxonomies or classifications. However, these are often not in a format readily available for use in programs that can understand and manipulate trees. For example, the global taxonomy of languages published by the Ethnologue (???) classifies languages into families and subgroups using a taxonomy string e.g. the language Kalam is classified as “Trans-New Guinea, Madang, Kalam-Kobon”, while [Mauwake] is classified as “Trans-New Guinea, Madang, Croisilles, Pihom”, and [Kare] is “Trans-New Guinea, Madang, Croisilles, Kare”. This classification indicates that while all these languages are part of the Madang subgroup of the Trans-New Guinea language family, Kare and Mauwake are more closely related (as they belong to the Croisilles subgroup).

Other publications use a tabular indented format to demarcate relationships, such as this example from Stephen Wurm’s classification of his proposed Yele-Solomons language phylum:

```

    This gives the following picture of the composition of the Yele-
    Solomons Stock (93501):
    1) The Central Solomon Family 6850
        Bilua          4300
        Baniata        900
        Lavukaleve     700
        Savosavo       9502
    2) The Kazukuru Family }
        Kazukuru           }
        Guliguli           }
        Dororo             }
    3) The Yele family-level Isolate 2500
```

Figure 1: Example of a language taxonomy in indented format

Both the taxonomy string and tabular format however are hard to load into software packages that can analyse, compare, visualise and manipulate trees. *TreeMaker* aims to make this easy by converting taxonomic data into Newick format and Nexus (Maddison 1997), popular formats commonly usable by phylogenetic manipulation programs.

## Converting a Taxonomy to a Tree:

TreeMaker can convert a text file with a taxonomy (easily obtained from Ethnologue or manually entered) like this:

```
Bilua      Yele-Solomons, Central Solomon
Baniata    Yele-Solomons, Central Solomon
Lavukaleve Yele-Solomons, Central Solomon
Savosavo   Yele-Solomons, Central Solomon
Kazukuru   Yele-Solomons, Kazukuru
Guliguli   Yele-Solomons, Kazukuru
Dororo     Yele-Solomons, Kazukuru
Yele       Yele-Solomons
```

Into a Newick tree representation:

```
((Baniata,Bilua,Lavukaleve,Savosavo),(Dororo,Guliguli,Kazukuru),Yele)
```

... which can then be loaded into phylogenetic programs to visualise or manipulate:

TreeMaker has been used to enable the analyses in (Bromham et al. 2018), and a number of forthcoming articles.

## References

- Bromham, Lindell, Xia Hua, Marcel Cardillo, Hilde Schneemann, and Simon J. Greenhill. 2018. "Parasites and Politics: Why Cross-Cultural Studies Must Control for Relatedness, Proximity and Covariation." *Open Science* 5 (8). The Royal Society. <https://doi.org/10.1098/rsos.181100>.
- Maddison, Wayne P. 1997. "Gene Trees in Species Trees." *Systematic Biology* 46 (3). Oxford University Press: 523. <https://doi.org/10.2307/2413694>.

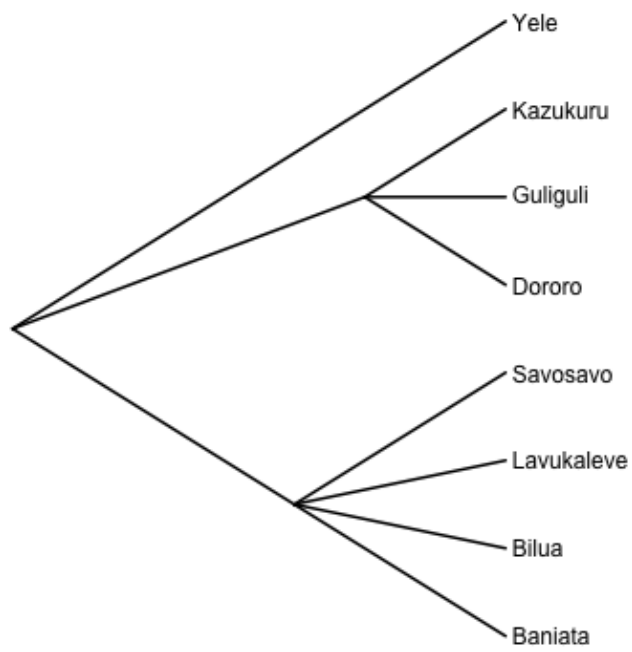


Figure 2: Tree visualisation of the relationships between these languages