

# When Does Phylogenetic Analysis Go Wrong?

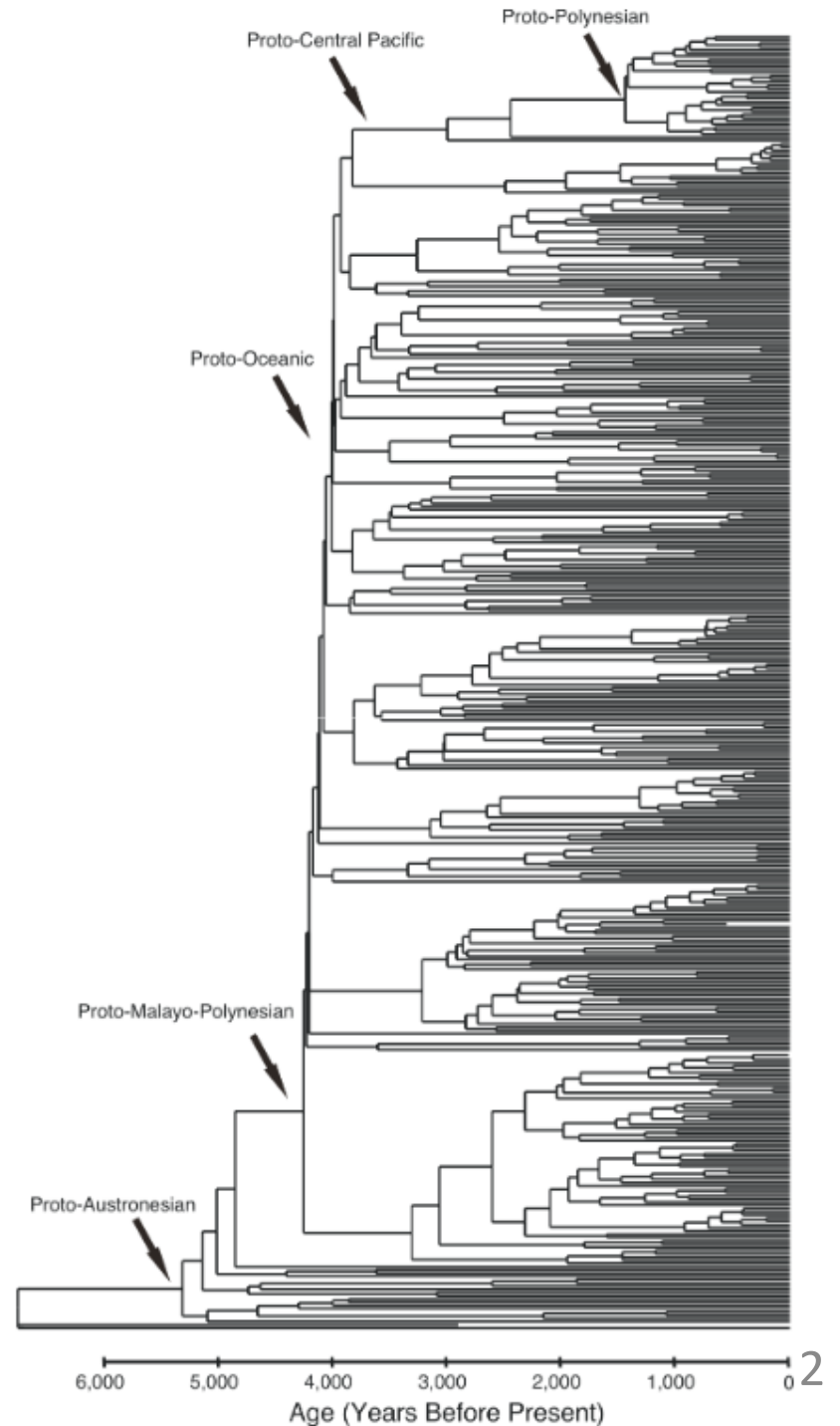
MURAWAKI Yūgo  
Kyoto University



# Phylogenetic Analysis

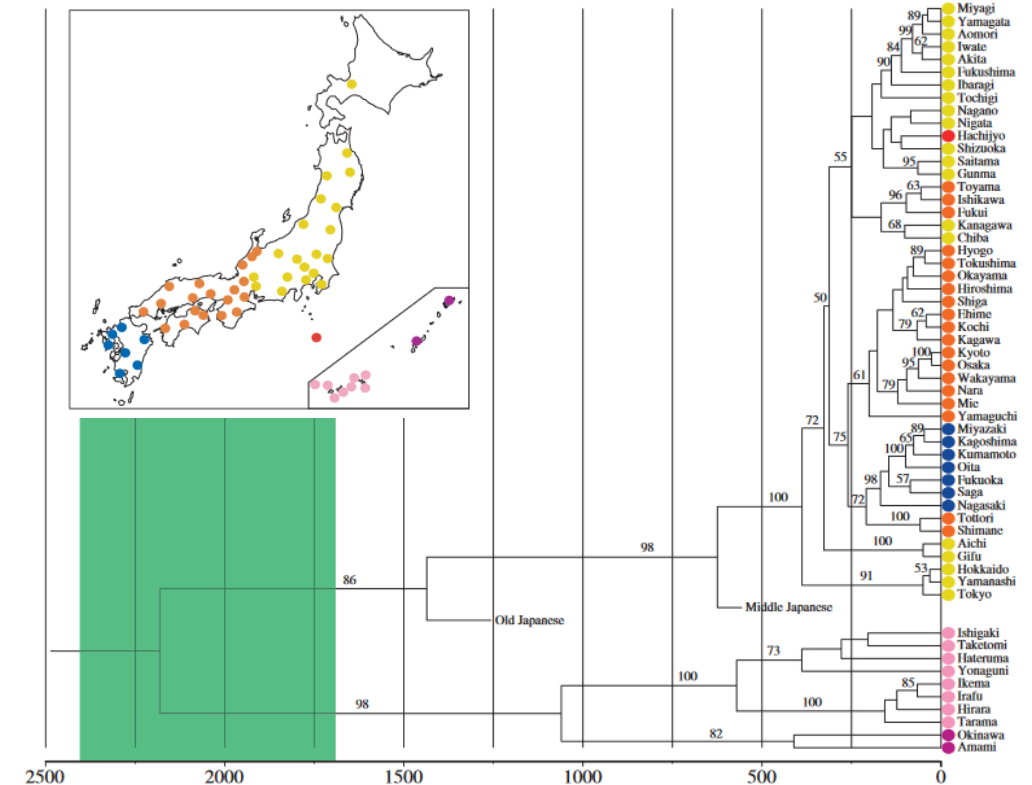
- Tree representation of the evolutionary relationships among languages
- Traditionally done manually by linguists
- Recent applications of compute-heavy statistical methods to basic vocabulary data

Austronesian language family  
[\[Greenhill and Gray, 2009\]](#)



# The Tree Assumption Can be Violated

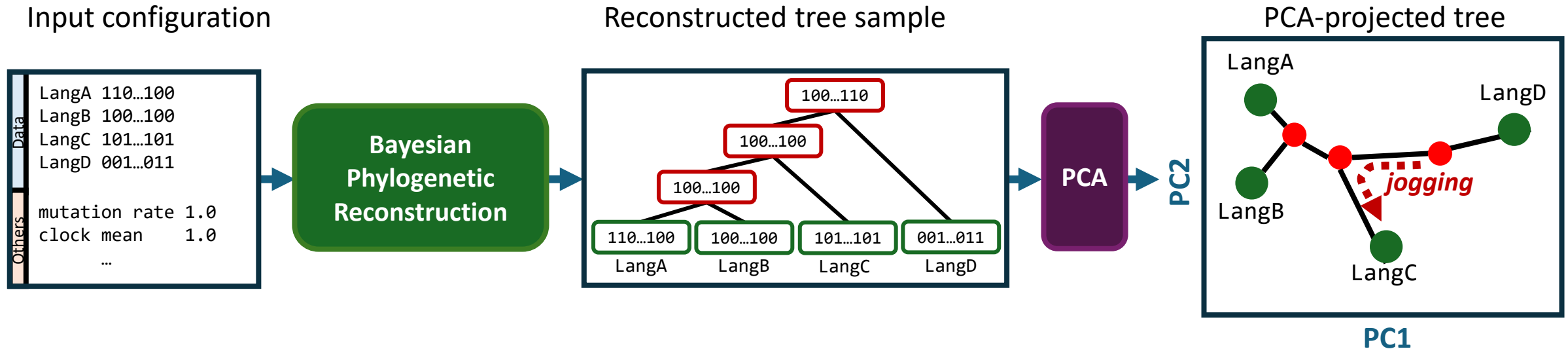
- Misapplication of the model is common
  - Assumption: Independent evolution following a branching event
  - Reality: Borrowings obscure vertical inheritance signals
- Question: Can we intuitively visualize these violations?



[Lee and Hasegawa, 2011]

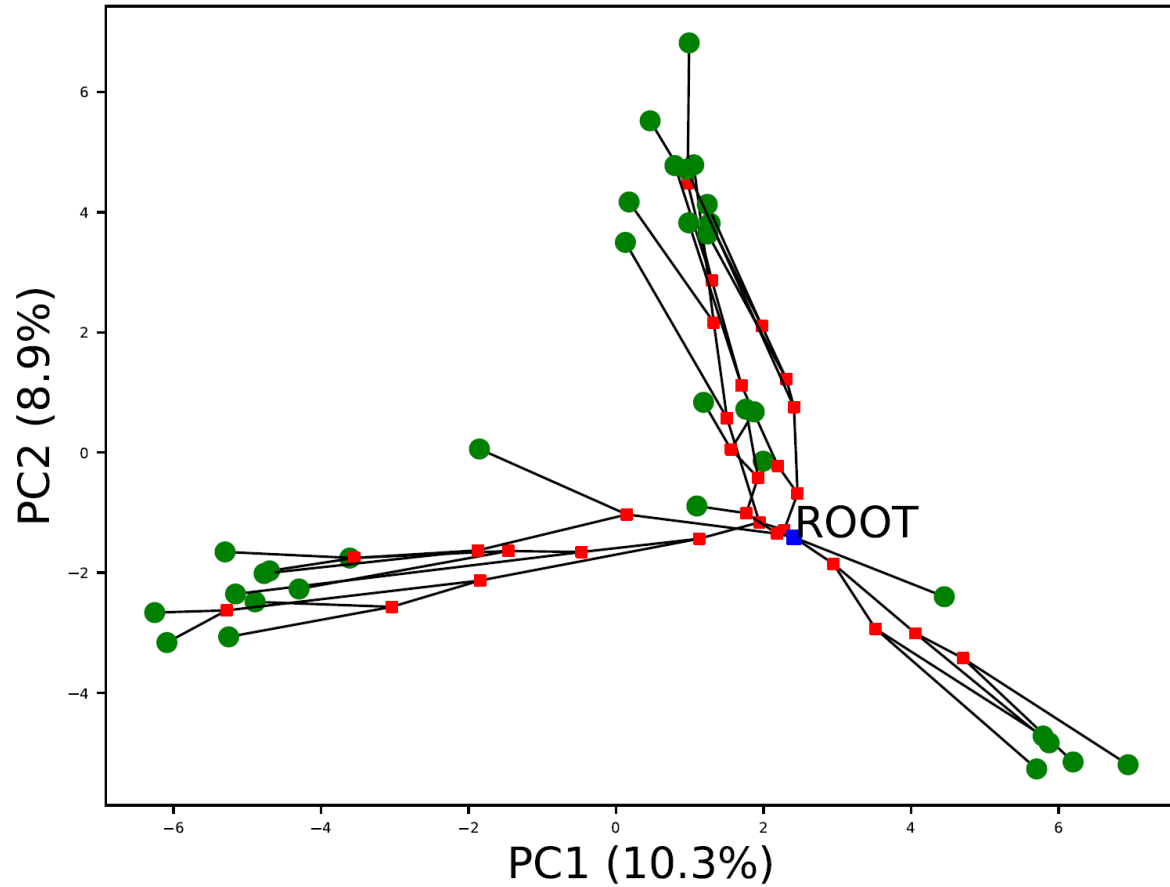
# Principal Component Analysis as a Sanity Check

[Murawaki, 2024]

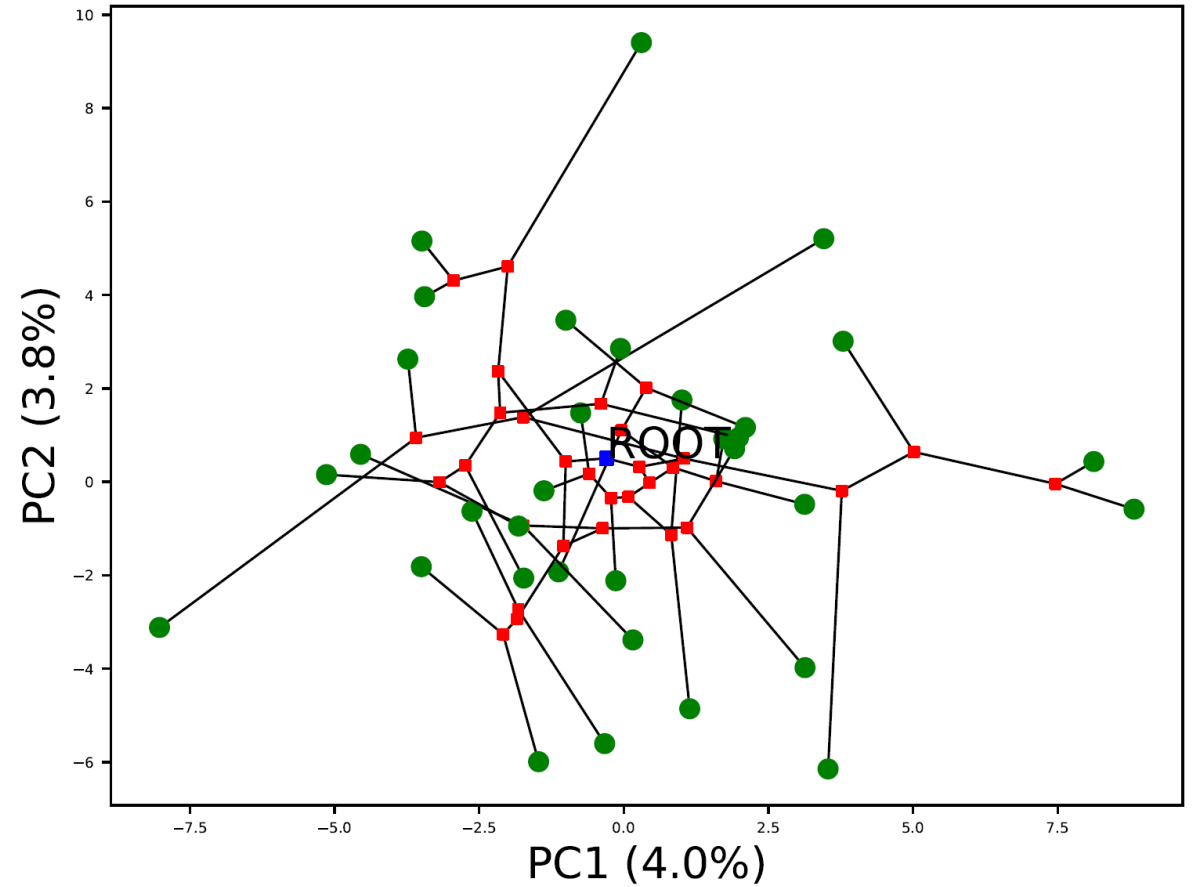


- Apply linear dimensionality reduction (PCA) to the state of a tree sample
- Lexical data must roughly follow a unidirectional pattern along PC1
- Violations are visualized as *jogging* patterns

# Simulation Experiments

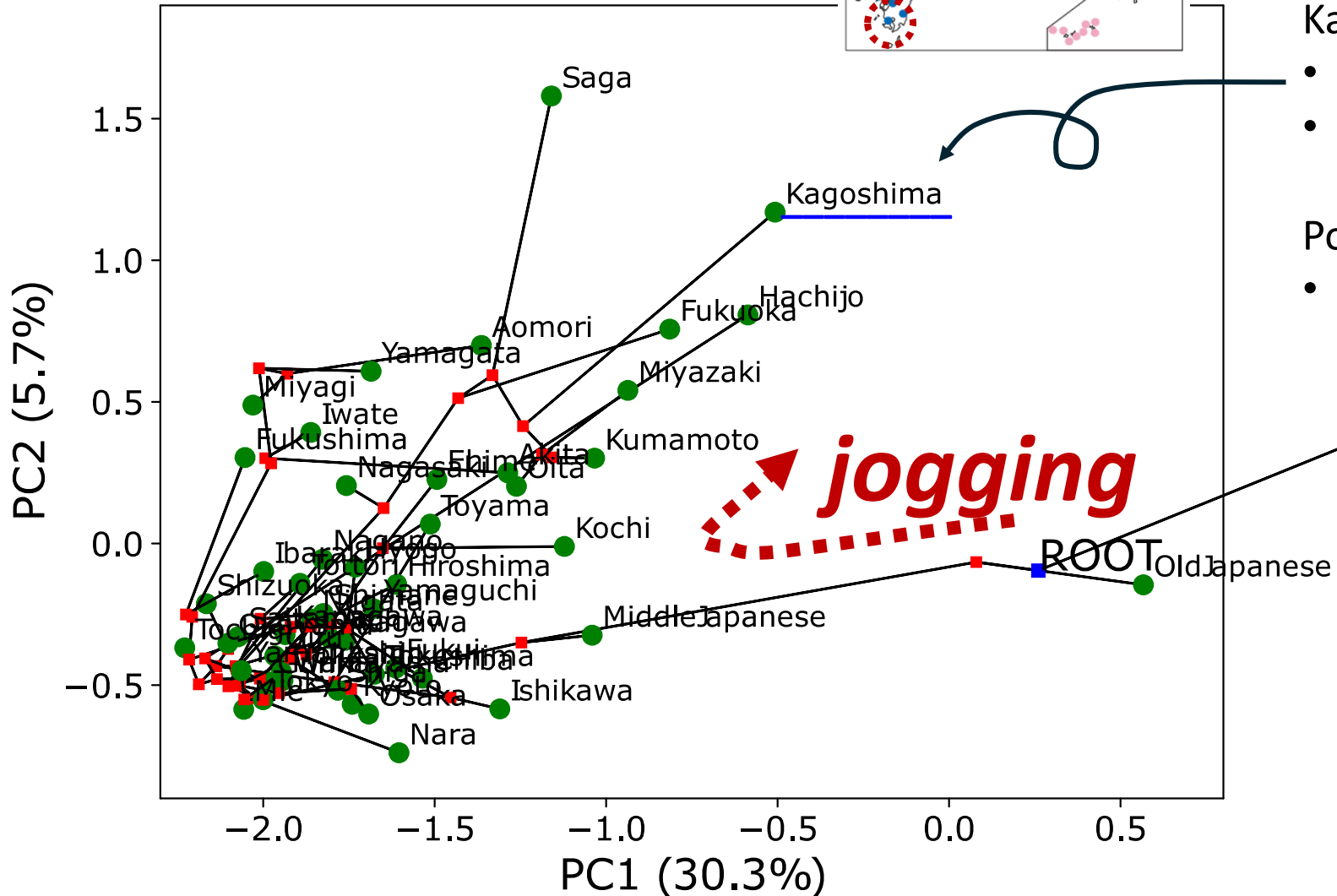
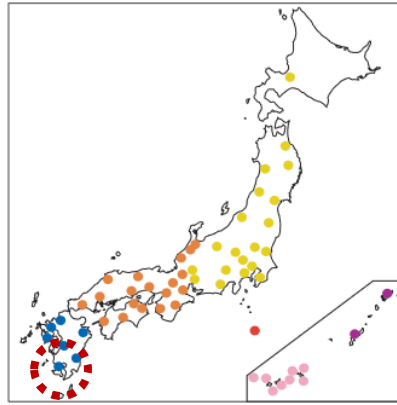


No-borrowing scenario



Heavy-borrowing scenario

# Japonic Example



Kagoshima is

- Closest to Old Japanese along PC1
- But second-to-last in terms of overall similarities

Possible explanation:

- Less affected by dialect leveling
  - Relatively rapid overall changes
  - Retained features indicate archaism

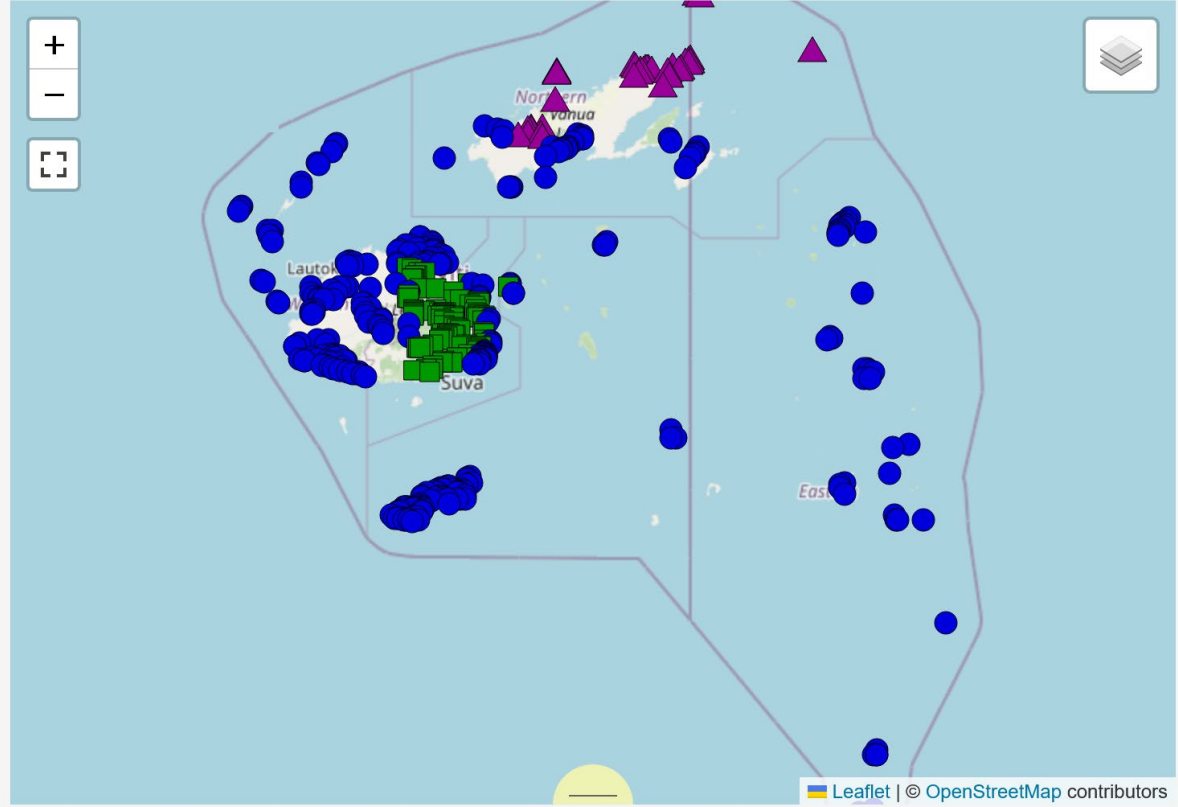


# Aside: The c11d Toolkit

- Converting raw linguistic data into an interactive web app is now fairly easy
  - Thanks to the open-source c11d toolkit

**Concept: ika\_\_fish\_\_**  
Standard Fijian: ika (fish)

Legend ▾ Icon size ▾  Show/hide Labels GeoJSON ▾



Showing 1 to 100 of 476 entries

← Previous 1 2 3 4 5 Next → ⓘ

Details	Language	Value	Source
	<input type="text" value="Search"/>	--any-- ▾	
<a href="#">more</a>	Matokana	● ika	<a href="#">↻</a>
<a href="#">more</a>	Nukuni	● ika	<a href="#">↻</a>



# Conclusions

- Computational phylogenetic analysis is powerful but prone to misuse
- Proposed a simple and effective sanity check to detect anomalies in phylogenetic analysis
- In a separate line of research, I have focused extensively on statistically modeling vertical inheritance in the presence of significant borrowings
  - If you are interested in, please check [my publications](#)