

# ASR Methods

Tony Ni, December 2022

## Obtain and process metadata

- Use scripts from `miscellaneous_format+metadata.py` to obtain genome names from tree file, obtain accession number (substring of the names), and samples of which metadata is missing. This script also contains other blocks for formatting and troubleshooting.
- Use `downloadMetadata.sh` to download location data from EBI's BioSamples database.
- Use `getLocationFromJSON.py` to obtain location data from the JSON files downloaded. Adjust `keys` according to the naming of attributes.
- Use `groupLocation.py` to assign more general states to the locations obtained. Adjust the lists for categorization according to the location values.
- Use `convertToSimmapInput.py` to associate states with tip names in tree through the shared BioSamples accession number.

## Reconstruct ancestral state

- Use `ASR_Simmap.Rmd` for stochastic mapping using SIMMAP.