**Nick’s revised PYR suggestions**

1. **BEAST2 inputs / outputs - What are the kinds of date calibration points**

In Bayesian phylogenetic inference using BEAST2, the analysis relies on several key input and output components. The primary inputs include an aligned sequence dataset, typically in FASTA or NEXUS format, and an XML configuration file generated via BEAUti. This XML file encodes model specifications, prior distributions, and calibration details. Optional but often critical are calibration data, which may include fossil-derived node calibrations or tip dates for sequences with known sampling times. BEAST2 produces several output files, including a .log file recording sampled parameter values (analyzed in Tracer), a .trees file containing the posterior distribution of trees, and a .state file that preserves the MCMC state for resumption. Post-analysis, users can generate a Maximum Clade Credibility (MCC) tree using TreeAnnotator and visualize results in programs such as FigTree. Regarding temporal calibration, BEAST2 supports three main types: node calibrations based on fossil data, which involve placing priors (e.g., lognormal or uniform distributions) on the age of clades; tip calibrations, which utilize sampling dates for sequences—commonly used in viral evolution studies; and secondary calibrations, which incorporate divergence estimates from prior research rather than direct fossil or sequence age data.

1. **Philosophy of AIC/AICc/BIC model selection – read e.g. Anderson 2006 or Burnham & Anderson articles**

“All models are wrong, but some are useful”. (George Box)

The philosophy of model selection using criteria such as AIC, AICc, and BIC centers on balancing model fit with model complexity to avoid overfitting and identify models that generalize well to new data. As outlined by Burnham & Anderson (2004), these information-theoretic approaches emphasize the trade-off between explaining the data and maintaining parsimony. AIC (Akaike Information Criterion) is grounded in information theory and estimates the relative quality of models by minimizing information loss. "Information loss" here refers to how much detail or accuracy about the **true underlying process** is lost when we use a **simplified model** to describe reality. AICc is a corrected version of AIC, particularly suitable for small sample sizes, while BIC (Bayesian Information Criterion) incorporates a stronger penalty for model complexity and is rooted in Bayesian principles. These criteria do not test hypotheses or claim a model is “true” but rather evaluate which model is most efficient at approximating the “true” model.

1. **What is an SSE model, how is it different from e.g. a biogeobears biogeog model**

Tree models capture how lineages diversify over time. The Yule model assumes a constant speciation rate without extinction (a "pure birth" process), is suitable for modelling early stages of diversification. The birth-death model, by contrast, incorporates both speciation and extinction rates.

Character models describe how genetic or morphological traits evolve along the branches of a phylogenetic tree. For DNA sequences, commonly used substitution models include HKY (Hasegawa–Kishino–Yano) and GTR (General Time Reversible). The HKY model accounts for differences between transition (A↔G, C↔T) and transversion rates and allows unequal base frequencies, providing a biologically meaningful improvement over simpler models like Jukes-Cantor. The GTR model is more general, allowing each substitution type (e.g., A→G, C→T) to have its own rate

State-dependent speciation and extinction (SSE) models integrate phylogenetic tree models with character evolution by allowing traits such as ecological or morphological features to directly influence rates of speciation and extinction. A key example is ClaSSE (Cladogenetic State Speciation and Extinction), which models changes in traits during speciation events and allows for interactions between character states and lineage diversification. In contrast, BioGeoBEARS focuses on modelling the evolution of geographic ranges rather than general traits, and it offers a suite of models for historical biogeography. The core models include:

* DEC (Dispersal–Extinction–Cladogenesis): Models range evolution as a continuous-time process involving dispersal (range expansion), extinction (range contraction), and cladogenesis (range division at speciation). It assumes ranges change gradually along branches and then are divided among descendant lineages at nodes.
* DEC+J: Extends DEC by adding a “J” parameter for founder-event speciation, allowing a lineage to "jump" to a new area during speciation. This cladogenetic jump dispersal captures rare but important biogeographic events, especially relevant in island systems.
* DIVALIKE: Inspired by the DIVA model, this assumes vicariance (range splitting) and subset sympatry are common at speciation events, but it does not include explicit extinction. It allows ancestral ranges to split in a way that emphasizes geographic continuity.
* DIVALIKE+J: Adds founder-event speciation to the DIVALIKE framework.
* BAYAREALIKE: Based on the Bayesian model BAYAREA, it models range evolution as anagenetic changes along branches, without any cladogenetic range inheritance scenarios. It assumes a high likelihood of widespread ranges but lacks explicit extinction and vicariance processes.
* BAYAREALIKE+J: Adds the jump-dispersal parameter to BAYAREALIKE.

Although BioGeoBEARS models resemble SSE models conceptually—since they model the evolution of discrete states (geographic ranges)—they simplify the full SSE framework. For example, extinction rates are typically fixed to zero in BioGeoBEARS to reduce computational complexity, and the number of transition parameters is limited. As Matzke (2022) notes, BioGeoBEARS can be viewed as implementing a constrained version of ClaSSE, where certain diversification processes are limited or held constant.

**References**

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>

Matzke, N. J. (2022). Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. *Journal of Biogeography*, *49*(10), 1805–1824. <https://doi.org/10.1111/jbi.14346>