Characters changing speed: the covariomorph model and its impact on phylogenetic trees

Basanta Khakurel[1,2], and Sebastian Höhna[1,2]

[1] GeoBio-Center, LMU München, Germany
[2] Department of Earth and Environmental Sciences, Paleontology & Geobiology, LMU München, Germany
`b.khakurel@lmu.de`

Why uniform-rate models fall short for morphology?

- The Mk model [1] assumes a **uniform evolutionary rate for all characters across the tree**.
- In reality, selective pressures vary across characters, lineages, and time; **lineage-specific rate variation for individual characters is rarely modeled** for morphology.
- We implemented a **covarion model** [2] for morphology in **`RevBayes`**.
- Our **Covariomorph model** captures both among-character rate variation and lineage-specific rate changes.

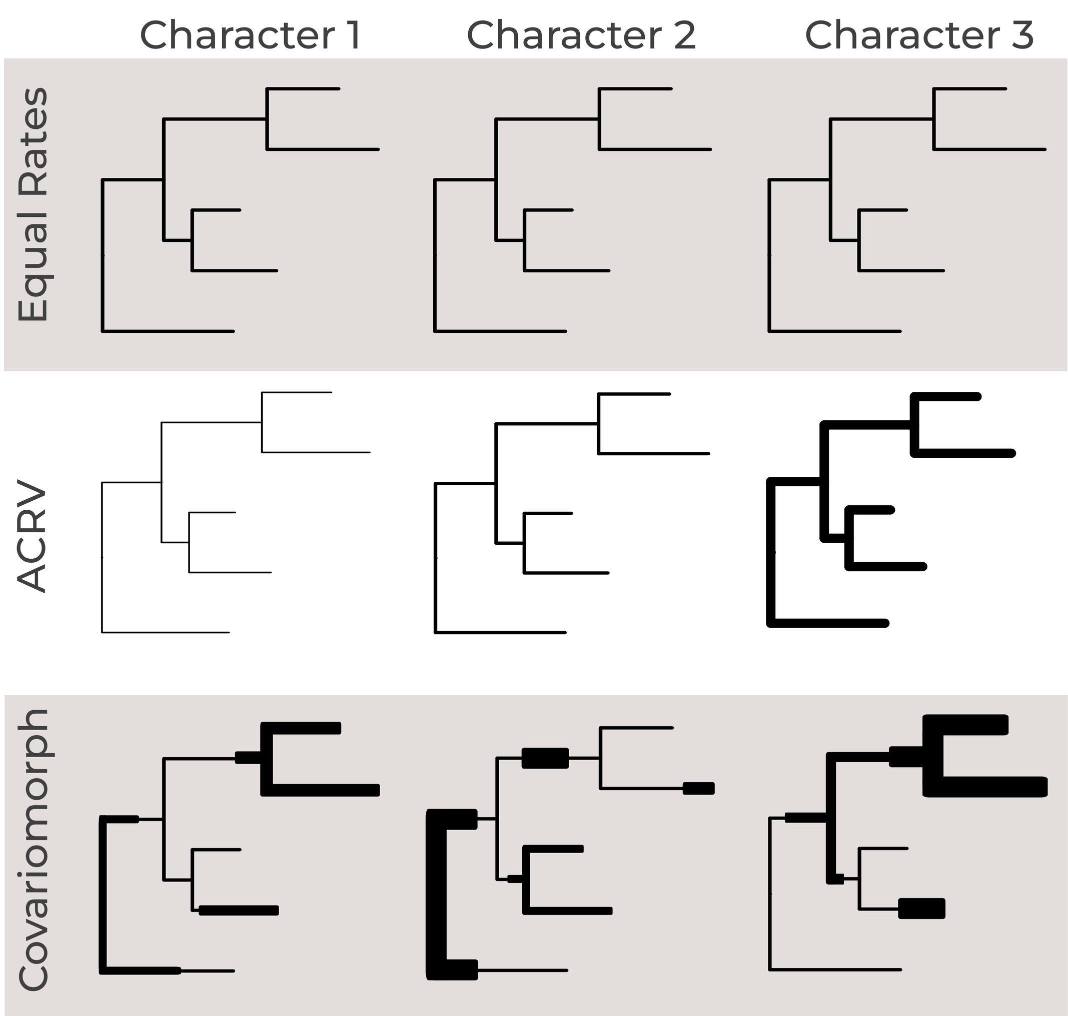

Figure 1: **Rate models in morphology.** Equal rates (Mk): uniform rates for all characters across lineages. ACRV: rates differ across characters, not lineages. Covariomorph: rates differ both across characters and lineages.

The Covariomorph model

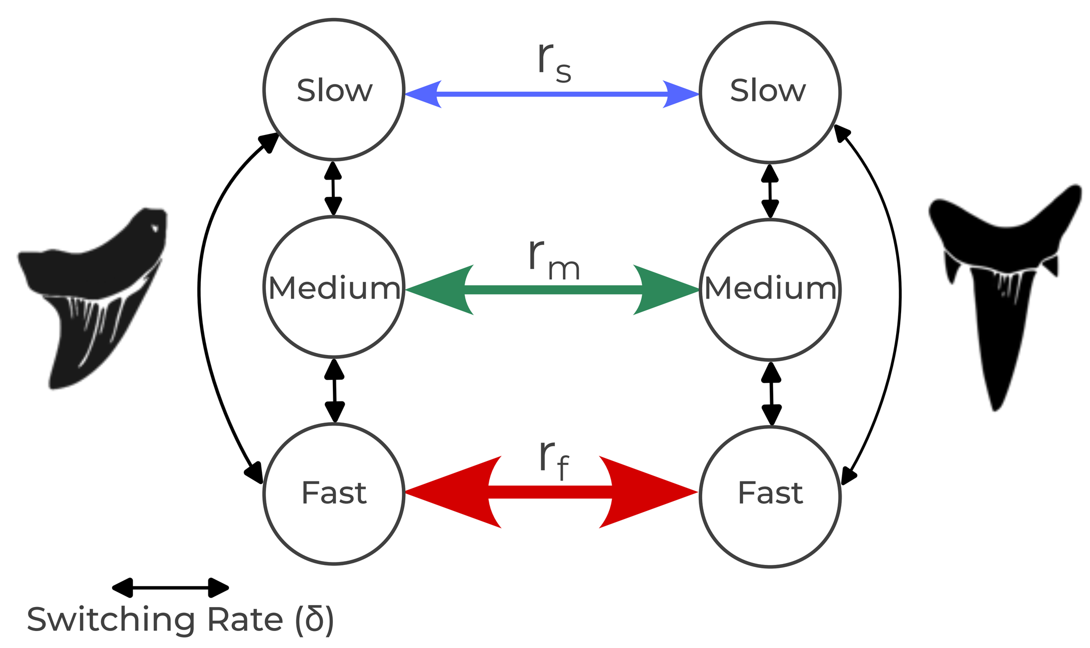

Figure 2: **Schematic of a 3-rate binary covariomorph model.** A character (e.g., tooth shapes) can be in two states and evolve at slow (r_s), medium (r_m), or fast (r_f) rates (colored arrows). The rates can be drawn from a probability distribution. In our examples, we use a lognormal distribution with mean $\mu = 0$ and standard deviation σ. Rate-category switches (black arrows) occur at rate δ, allowing rates to vary among characters and across lineages.

Simulation results

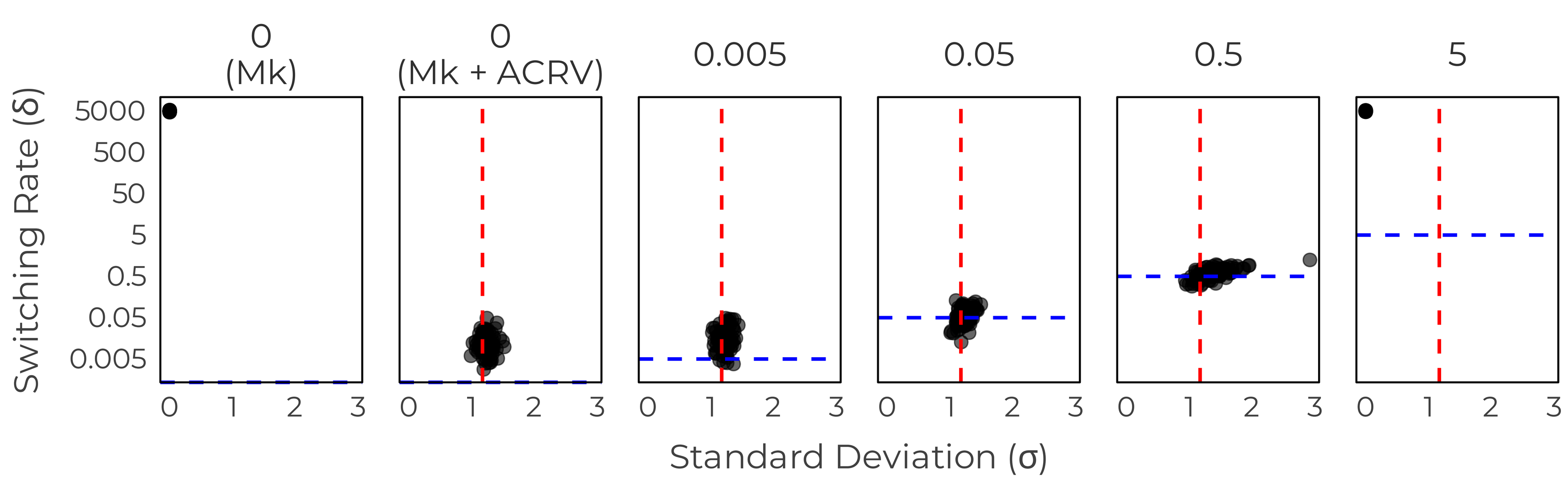

Figure 3: **Posterior median switching rates (δ) and standard deviations (σ) from 100 covariomorph simulations with four rate categories.** Panels show different true δ values (labels). Points are posterior medians; dashed lines mark true δ and σ values. The model accurately recovers parameters across a reasonable range.

Empirical datasets

Evidence of lineage-specific rate variation

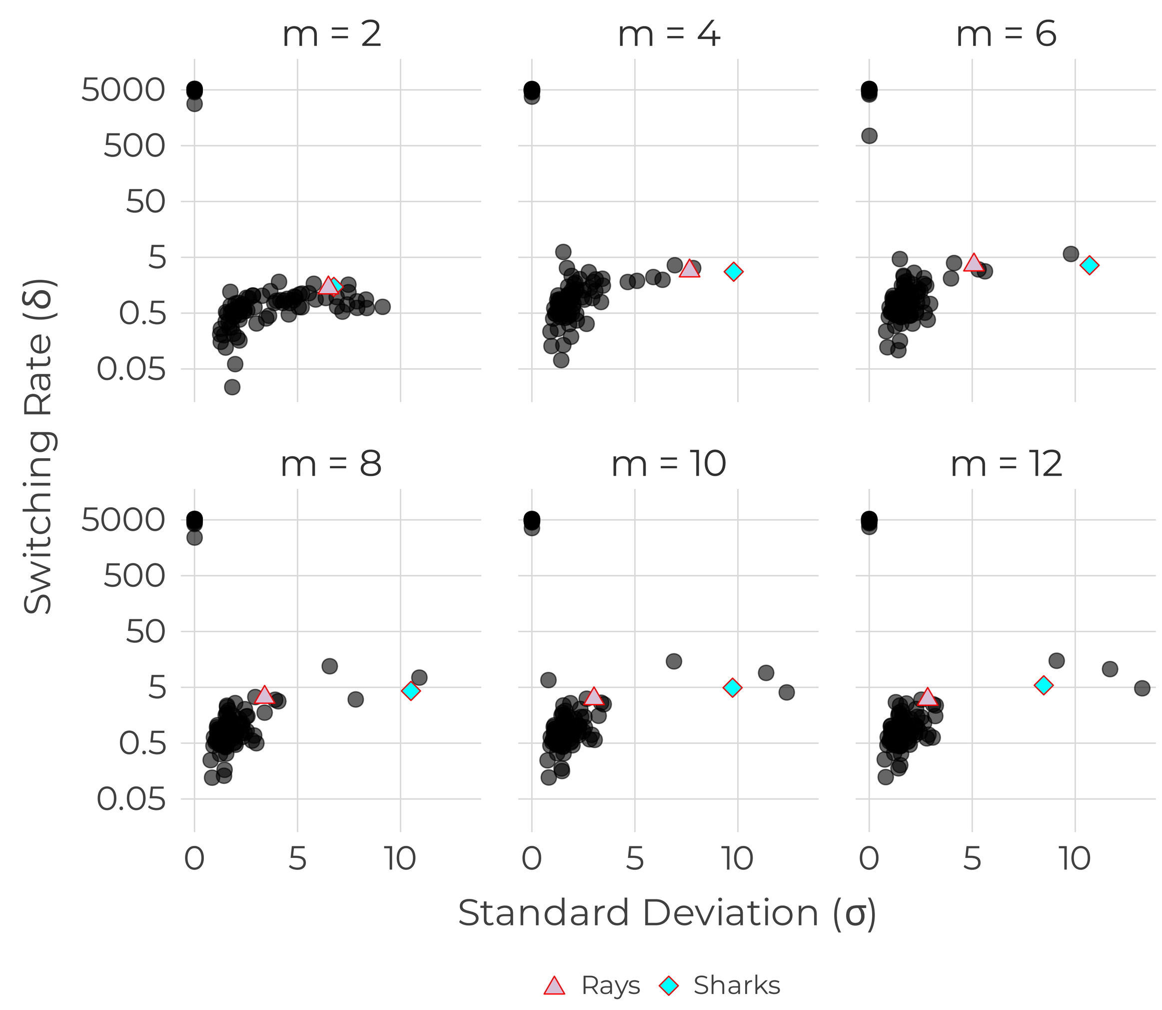

Figure 4: **Inferred switching rates vs. standard deviation from the covariomorph model with varying number of rates m.** Each point is the posterior median for one of 164 empirical datasets. Two patterns emerge: datasets with equal rates for all characters (top-left cluster) and those with rate variation among characters and lineages.

Posterior clade support under different models

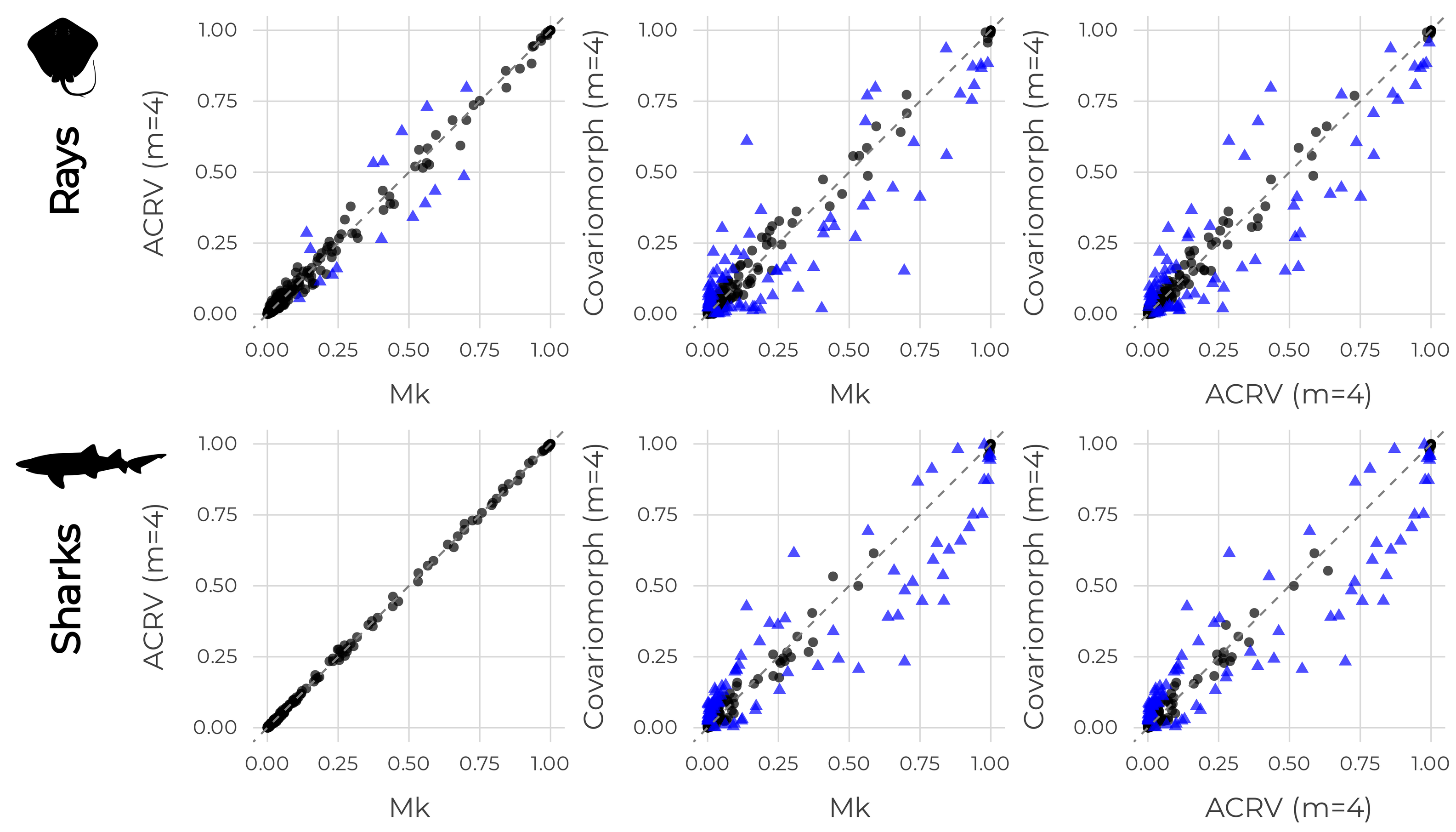

Figure 5: **Differences in clade support between rate-variation models for Sharks and Rays.** Scatter plots compare posterior clade probabilities for two models (one from each type in Fig. 1). Circles mark clades with no significant difference (ESS = 200, split-frequency test); triangles mark clades with significant differences.

Take home messages

- We **implemented the covarion model for morphology** in `RevBayes`.
- Simulations show it **accurately recovers switching rates and rate variation**.
- The **model collapses** to the Mk model when no rate variation is present and to ACRV model when no switching is present in the data.
- Empirical analyses reveal **widespread rate variation** in real datasets.
- Accounting for this heterogeneity **can improve tree topology and branch lengths**.

References

[1] Paul O Lewis. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6):913–925, 2001.
[2] Chris Tuffley and Mike Steel. Modeling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences*, 147(1):63–91, 1998.

Acknowledgements

This work was supported by the European Union (ERC, MacDrive, GA 101043187).