PALAEOVERSE LECTURE SERIES

Accounting for rate heterotachy in morphological phylogenetics

An overview in RevBayes

Basanta Khakurel

Department of Earth and Environmental Sciences
Ludwig-Maximilians-Universität München

2026-10-29

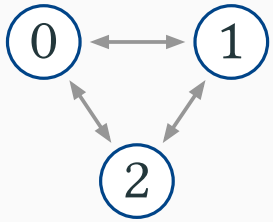
b.khakurel@lmu.de basantakhakurel.github.io basantakhakurel Slides: basantakhakurel.github.io/presentations

Inferring time-calibrated phylogenies

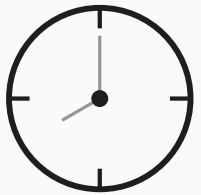
The data

	characters	age
Taxon A	0 1 1 0 2	—
Taxon B	0 1 0 0 2	—
Taxon C	1 0 1 1 1	—
† Taxon D	1 0 ? 1 ?	52 Ma
† Taxon E	? 0 1 ? 0	110 Ma

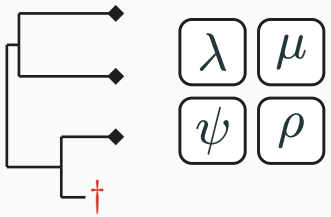
Model components



Substitution model



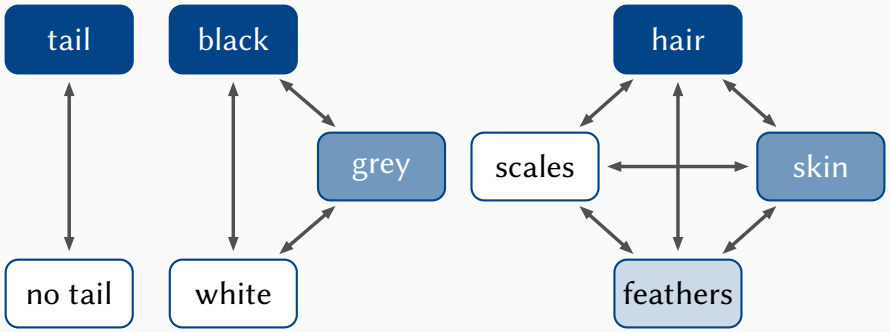
Clock model



Tree model

Adapted from Warnock & Wright (2020)

The Mk model



$k = 2$ $k = 3$ $k = 4$

How it works (Lewis, 2001)

- Jukes Cantor model for DNA, generalised to k states.
- Transitions among states follow a Markov process.
- **Branch length** = expected changes per character.

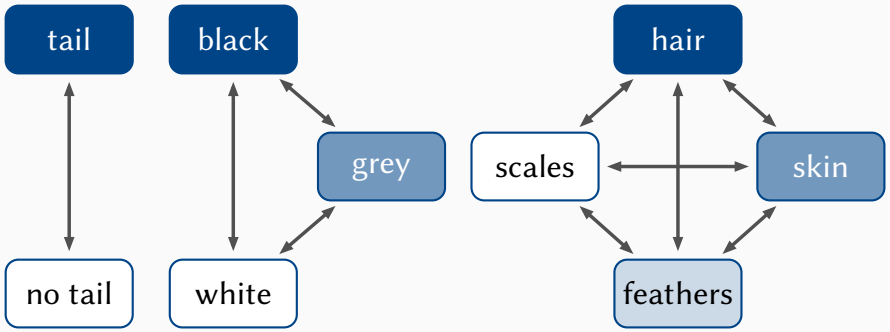
Assumptions

$$\mathcal{Q}_{\text{Mk}} = \begin{pmatrix} * & 1/2 & 1/2 \\ 1/2 & * & 1/2 \\ 1/2 & 1/2 & * \end{pmatrix}$$

Rate matrix for $k = 3$

Adapted from Harmon (2022)

The Mk model



$k = 2$ $k = 3$ $k = 4$

$$\mathcal{Q}_{\text{Mk}} = \begin{pmatrix} * & 1/2 & 1/2 \\ 1/2 & * & 1/2 \\ 1/2 & 1/2 & * \end{pmatrix}$$

Rate matrix for $k = 3$

How it works (Lewis, 2001)

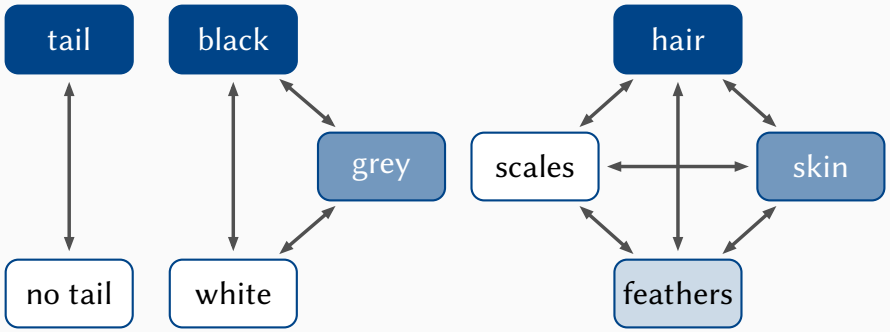
- Jukes Cantor model for DNA, generalised to k states.
- Transitions among states follow a Markov process.
- **Branch length** = expected changes per character.

Assumptions

- All changes equally likely.

Adapted from Harmon (2022)

The Mk model



$k = 2$ $k = 3$ $k = 4$

How it works (Lewis, 2001)

- Jukes Cantor model for DNA, generalised to k states.
- Transitions among states follow a Markov process.
- **Branch length** = expected changes per character.

Assumptions

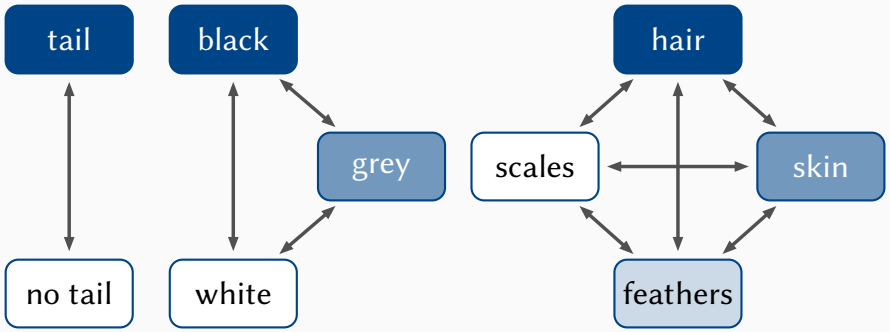
- All changes equally likely.
- Characters evolve independently.

$$\mathcal{Q}_{\text{Mk}} = \begin{pmatrix} * & 1/2 & 1/2 \\ 1/2 & * & 1/2 \\ 1/2 & 1/2 & * \end{pmatrix}$$

Rate matrix for $k = 3$

Adapted from Harmon (2022)

The Mk model



$k = 2$ $k = 3$ $k = 4$

$$\mathcal{Q}_{\text{Mk}} = \begin{pmatrix} * & 1/2 & 1/2 \\ 1/2 & * & 1/2 \\ 1/2 & 1/2 & * \end{pmatrix}$$

Rate matrix for $k = 3$

How it works (Lewis, 2001)

- Jukes Cantor model for DNA, generalised to k states.
- Transitions among states follow a Markov process.
- **Branch length** = expected changes per character.

Assumptions

- All changes equally likely.
- Characters evolve independently.
- **Homogeneous rates**: every character, every branch, throughout time.

Adapted from Harmon (2022)

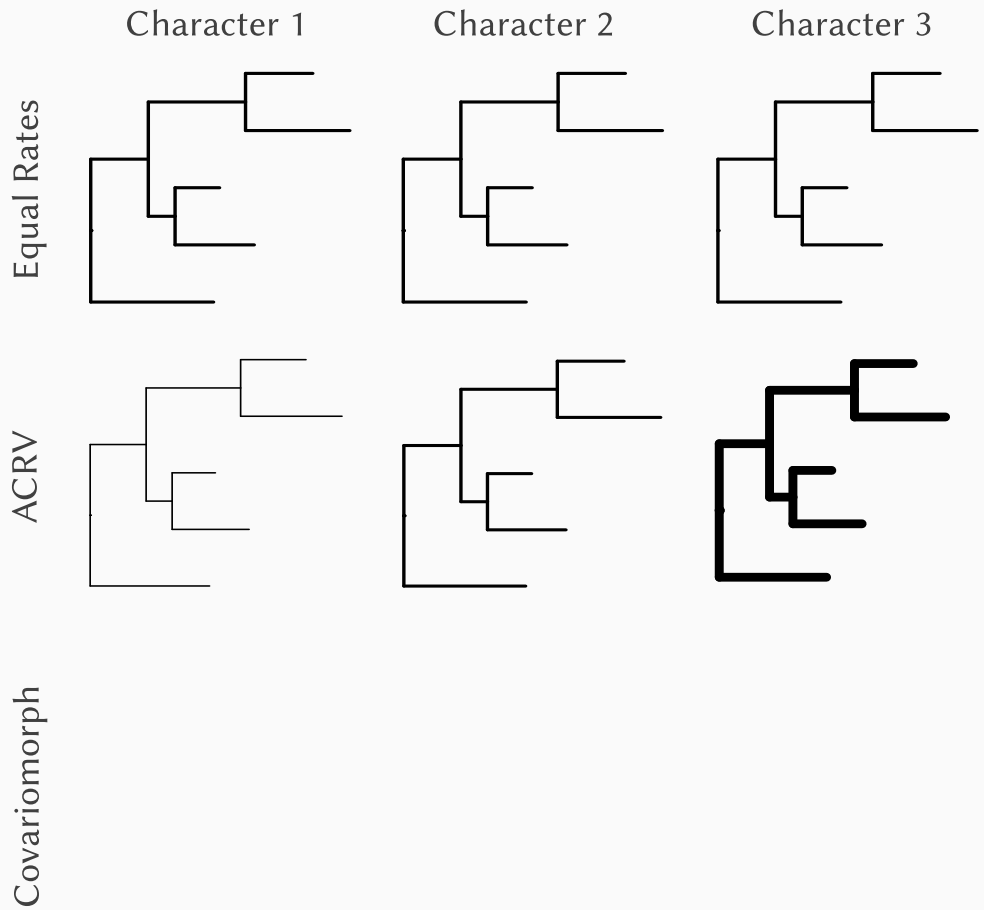
The morphological modeling landscape



Thick branch = high rate

- **Mk**: same rate everywhere, for every character, on every branch (Lewis, 2001).

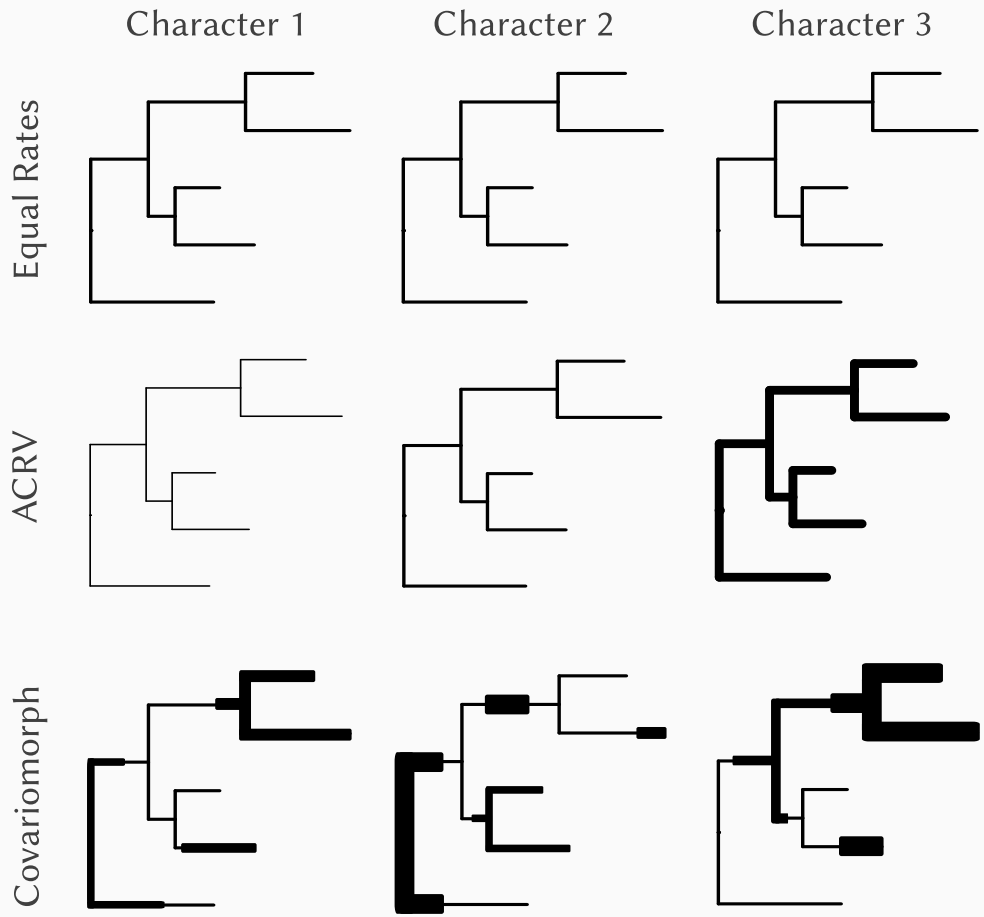
The morphological modeling landscape



Thick branch = high rate

- **Mk**: same rate everywhere, for every character, on every branch (Lewis, 2001).
- **ACRV**: characters differ in rate, but each character's rate is frozen for the whole tree (Harrison & Larsson, 2015; Yang, 1994).

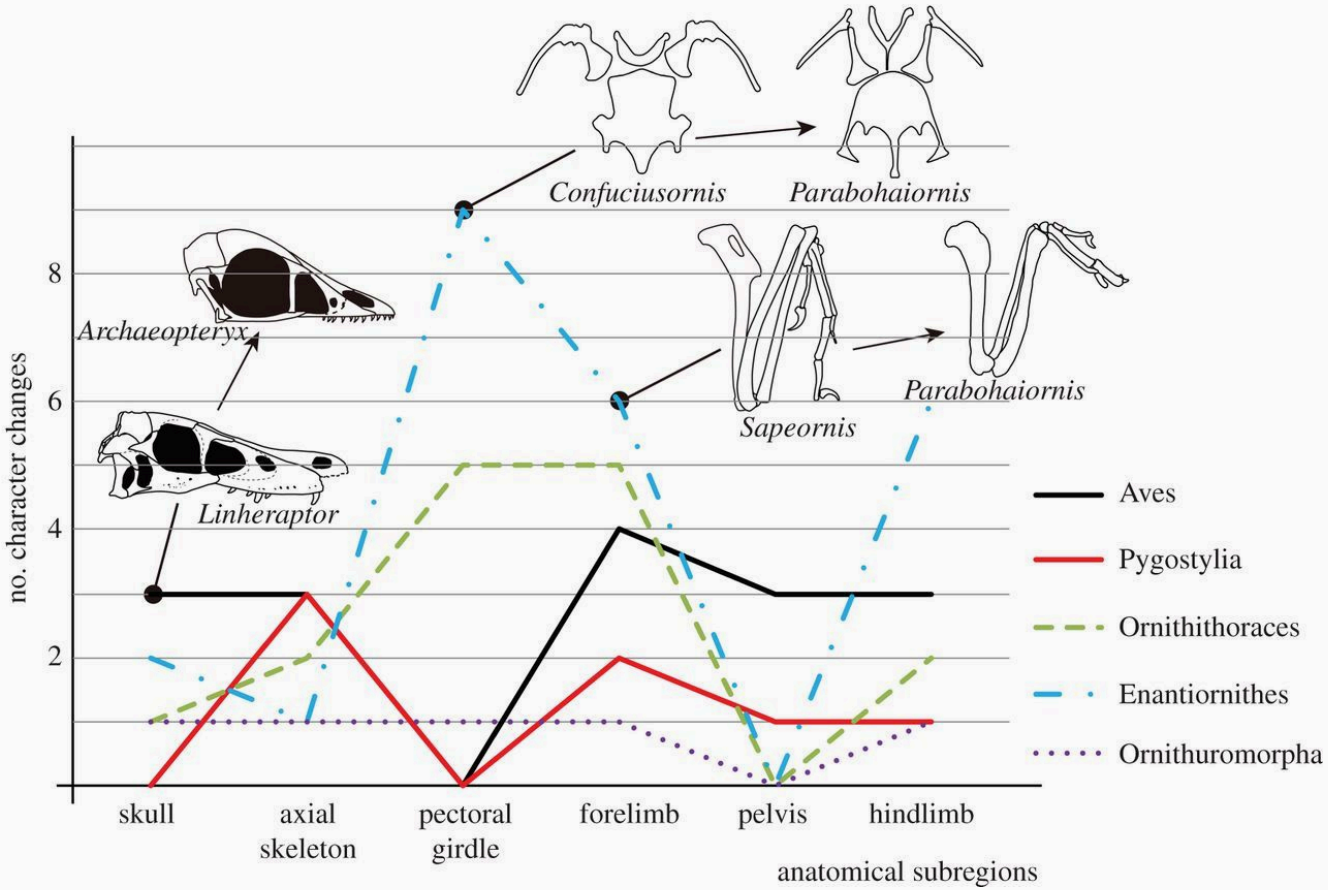
The morphological modeling landscape



Thick branch = high rate

- **Mk**: same rate everywhere, for every character, on every branch (Lewis, 2001).
- **ACRV**: characters differ in rate, but each character's rate is frozen for the whole tree (Harrison & Larsson, 2015; Yang, 1994).
- **Heterotachy**: rates for characters differ and can **change as they evolve along the tree**.

Avian evolution: an example



- **Selective constraints** change as lineages traverse ecological and developmental space.
- A character under **purifying selection** in one clade can be free to vary in another.

Wang & Lloyd (2016)

How can we model changing rates?

Approach	Rate variation	Character-specific rate changes along the tree?
Mk (Lewis, 2001)	none	✗
ACRV (Harrison & Larsson, 2015; Yang, 1994)	among characters, fixed across the tree	✗
Relaxed morphological clock (Drummond et al., 2006; Lee et al., 2014)	among lineages, shared by all characters	✗
Partitions (e.g. by body region)	among characters; sets defined in advance	◐
Covarion-like (Galtier, 2001; Tuffley & Steel, 1998) → **covariomorph**	among characters *and* along the tree; informed by the data	✓

Covarion model

- **Co**ncomitantly **Vari**able Cod**ons**
- Originally for molecular data (Fitch & Markowitz, 1970; Tuffley & Steel, 1998).

Covarion model

- **Co**ncomitantly **Vari**able Cod**ons**
- Originally for molecular data (Fitch & Markowitz, 1970; Tuffley & Steel, 1998).
- “On” and “off” states let a site stop or resume evolving.

substitutions don’t occur	switch to **on**
switch to **off**	substitutions occur

Covarion model

- **Co**ncomitantly **Vari**able Cod**ons**
- Originally for molecular data (Fitch & Markowitz, 1970; Tuffley & Steel, 1998).
- “On” and “off” states let a site stop or resume evolving.
- Used repeatedly for molecular data (Galtier, 2001; Huelsenbeck, 2002; Lockhart et al., 1998)

substitutions don’t occur	switch to **on**
switch to **off**	substitutions occur

Covarion model

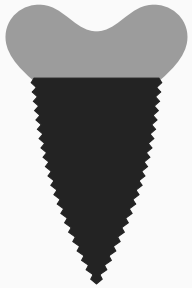
- **Co**ncomitantly **Vari**able Cod**ons**
- Originally for molecular data (Fitch & Markowitz, 1970; Tuffley & Steel, 1998).
- “On” and “off” states let a site stop or resume evolving.
- Used repeatedly for molecular data (Galtier, 2001; Huelsenbeck, 2002; Lockhart et al., 1998)
- **A promising mechanism for morphology to model heterotachy.**

substitutions don’t occur	switch to **on**
switch to **off**	substitutions occur

Example: three-rate binary covariomorph model



serrations
absent (0)

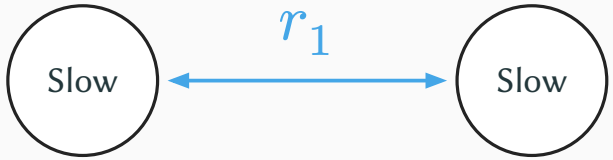


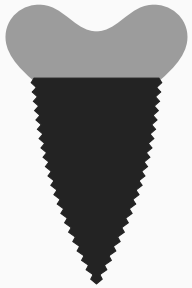
serrations
present (1)

Example: three-rate binary covariomorph model



serrations
absent (0)





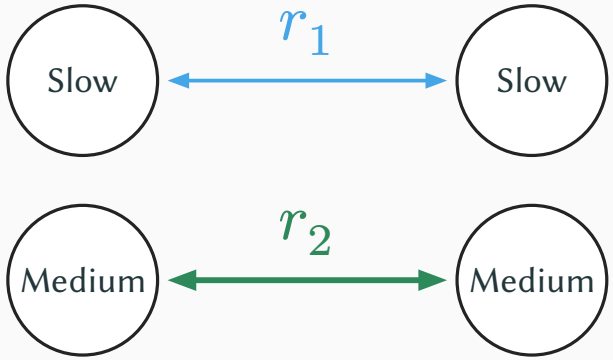
serrations
present (1)

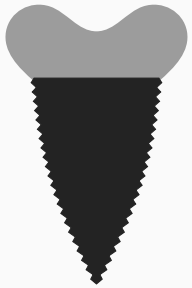
$$\tilde{Q} = \begin{pmatrix} * & r_1 \\ r_1 & * \end{pmatrix}$$

Example: three-rate binary covariomorph model



serrations absent (0)





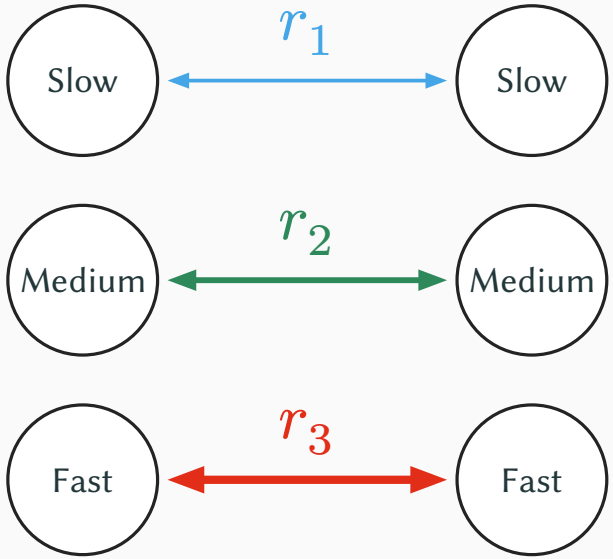
serrations present (1)

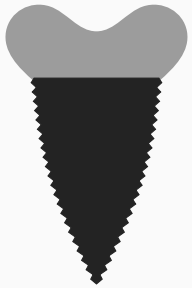
$$\tilde{Q} = \begin{pmatrix} * & r_1 & & \\ r_1 & * & & \\ & & * & r_2 \\ & & r_2 & * \end{pmatrix}$$

Example: three-rate binary covariomorph model



serrations absent (0)

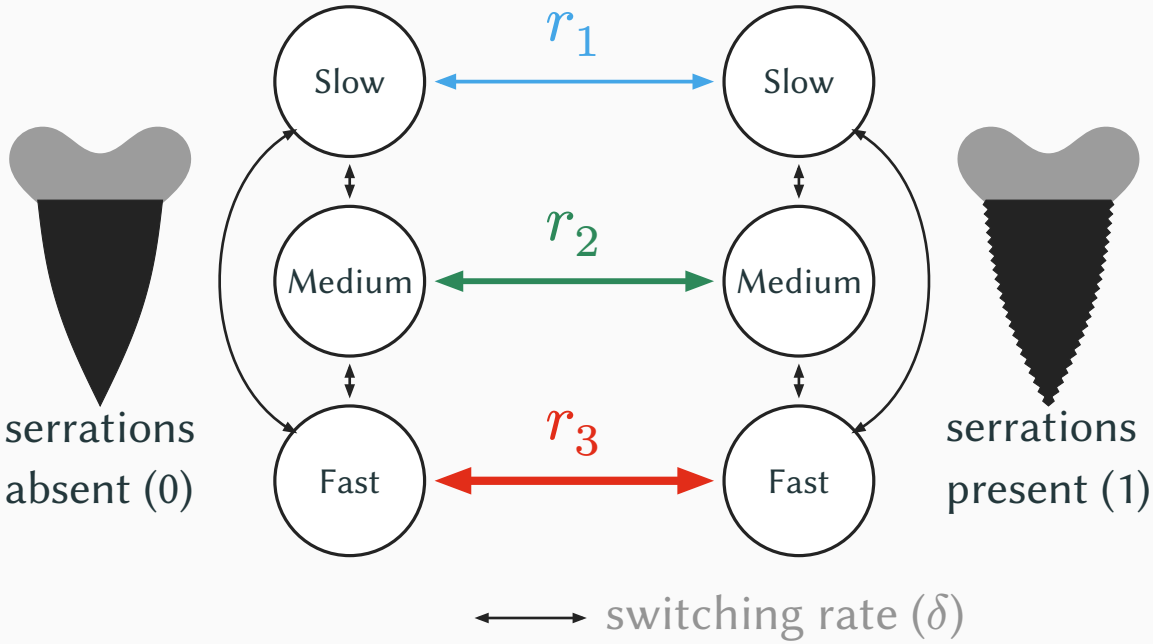




serrations present (1)

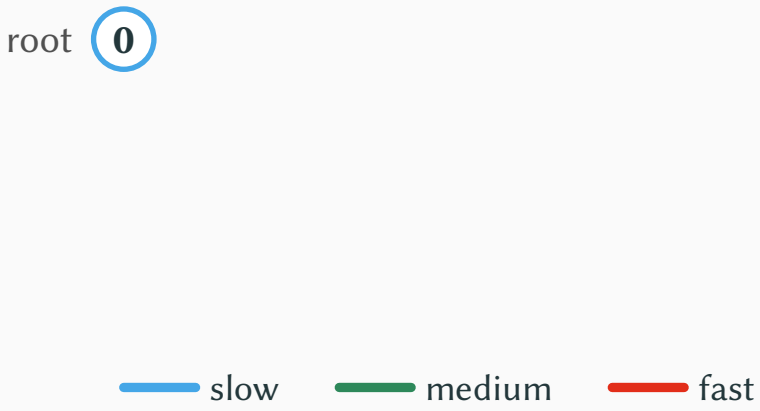
$$\tilde{Q} = \begin{pmatrix} * & r_1 & & & & \\ r_1 & * & & & & \\ & & * & r_2 & & \\ & & r_2 & * & & \\ & & & & * & r_3 \\ & & & & r_3 & * \end{pmatrix}$$

Example: three-rate binary covariomorph model



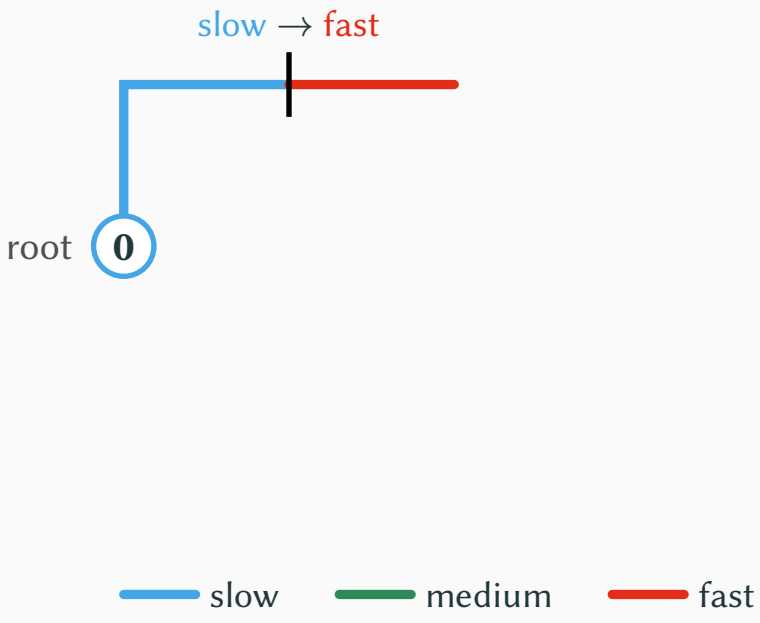
$$\tilde{Q} = \begin{pmatrix} * & r_1 & \delta/2 & 0 & \delta/2 & 0 \\ r_1 & * & 0 & \delta/2 & 0 & \delta/2 \\ \delta/2 & 0 & * & r_2 & \delta/2 & 0 \\ 0 & \delta/2 & r_2 & * & 0 & \delta/2 \\ \delta/2 & 0 & \delta/2 & 0 & * & r_3 \\ 0 & \delta/2 & 0 & \delta/2 & r_3 & * \end{pmatrix}$$

Covariomorph Generative Process



At the root the process starts with a state (0) and a rate category (slow).

Covariomorph Generative Process



Here the character switches from **slow to fast**.

Two possible events along a branch

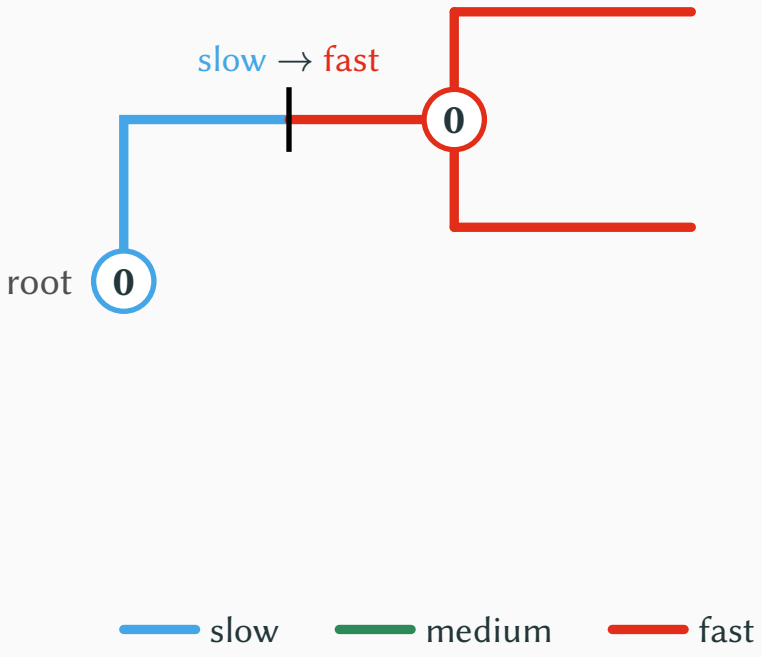
In state j and category i, the waiting time to the next event is $\mathrm{Exp}\big(\delta - r_i \mathcal{Q}_{jj}\big)$. An event can be a:

$$\textbf{state change:} \quad \frac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$$

$$\textbf{category switch:} \quad \frac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph Generative Process



At a speciation event, the descendant lineages inherit the current state and category.

Two possible events along a branch

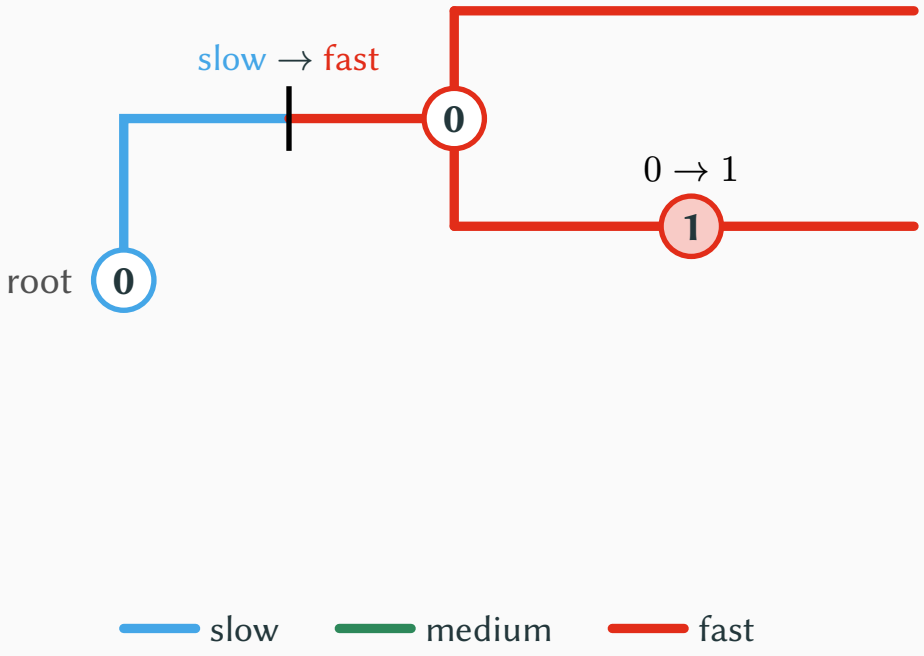
In state j and category i, the waiting time to the next event is $\mathrm{Exp}\big(\delta - r_i \mathcal{Q}_{jj}\big)$. An event can be a:

state change: $\dfrac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$

category switch: $\dfrac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph Generative Process



Fast category: state changes come quickly. Here: $0 \to 1$.

Two possible events along a branch

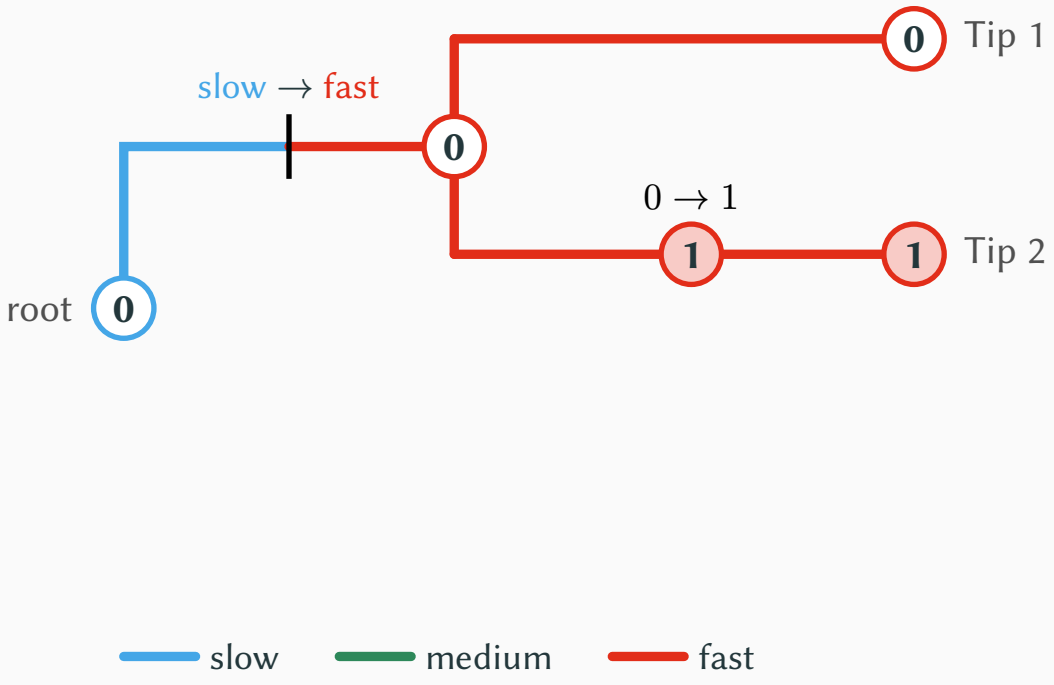
In state j and category i, the waiting time to the next event is $\mathrm{Exp}\big(\delta - r_i \mathcal{Q}_{jj}\big)$. An event can be a:

state change: $\dfrac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$

category switch: $\dfrac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph Generative Process



Tips keep whatever state the lineage ended in.

Two possible events along a branch

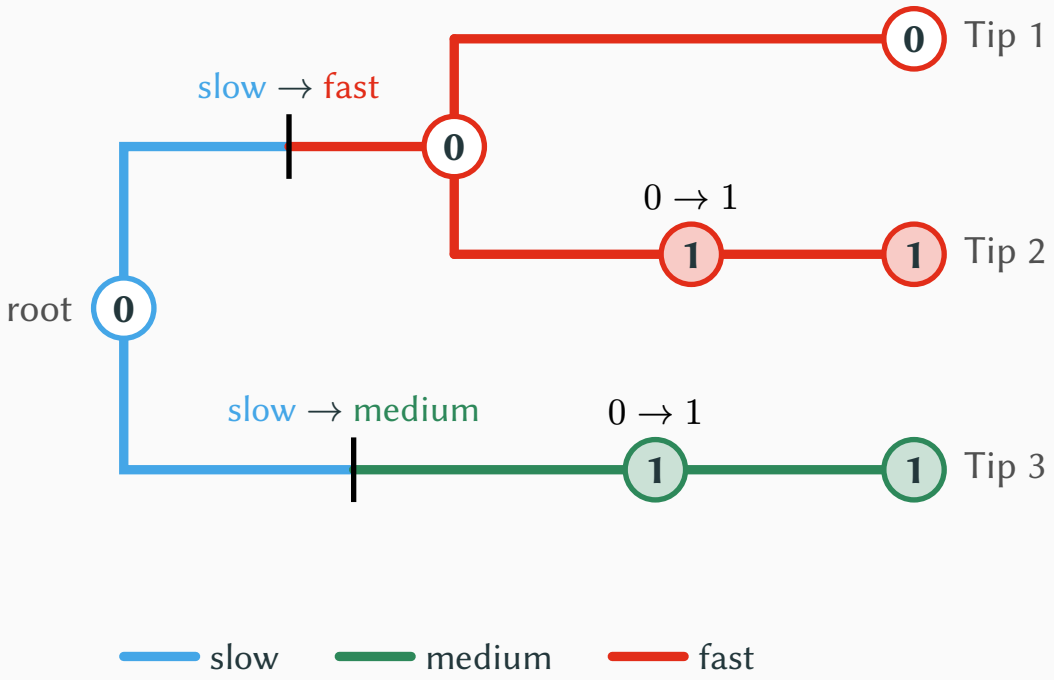
In state j and category i, the waiting time to the next event is $\mathrm{Exp}\left(\delta - r_i \mathcal{Q}_{jj}\right)$. An event can be a:

state change: $\dfrac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$

category switch: $\dfrac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph Generative Process



Lower lineage switches from **slow to medium**, then changes once.

Two possible events along a branch

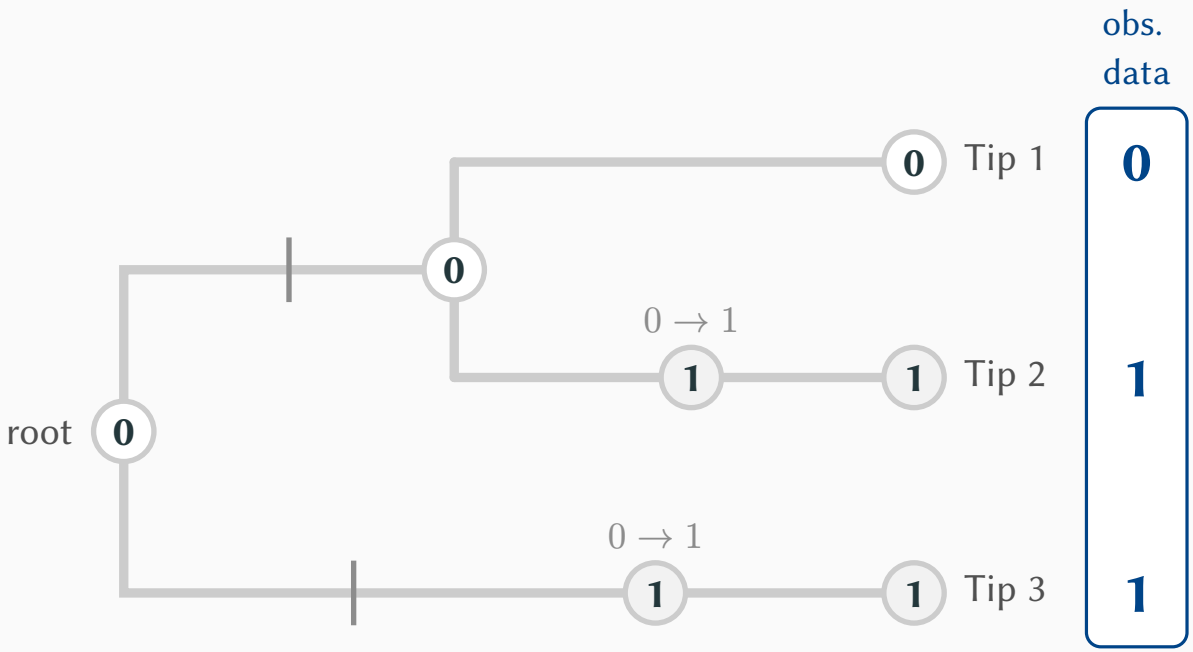
In state j and category i, the waiting time to the next event is $\mathrm{Exp}\left(\delta - r_i \mathcal{Q}_{jj}\right)$. An event can be a:

state change: $\dfrac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$

category switch: $\dfrac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph Generative Process



We only observe the tip states.
The rate categories are hidden.

Two possible events along a branch

In state j and category i, the waiting time to the next event is $\mathrm{Exp}\big(\delta - r_i \mathcal{Q}_{jj}\big)$. An event can be a:

state change: $\dfrac{-r_i \mathcal{Q}_{jj}}{\delta - r_i \mathcal{Q}_{jj}}$

category switch: $\dfrac{\delta}{\delta - r_i \mathcal{Q}_{jj}}$

A switch picks one of the other $m - 1$ categories, each at rate $\delta/(m-1)$.

Covariomorph rate matrix

Each character evolves in one of m rate categories and can switch between categories as it evolves.

$$\tilde{\mathcal{Q}} = \begin{pmatrix} r_1\mathcal{Q} & \frac{\delta}{m-1}\mathcal{I} & \cdots & \frac{\delta}{m-1}\mathcal{I} \\ \frac{\delta}{m-1}\mathcal{I} & r_2\mathcal{Q} & \cdots & \frac{\delta}{m-1}\mathcal{I} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta}{m-1}\mathcal{I} & \frac{\delta}{m-1}\mathcal{I} & \cdots & r_m\mathcal{Q} \end{pmatrix}$$

$r_i\mathcal{Q}$ = scaled Mk matrix

δ = switching rate

$\mathcal{Q}$ = Mk rate matrix
r_i = rate of category i
m = number of rate categories
$\mathcal{I}$ = identity matrix



Simulation setup

Empirical trees

Short tree

21 taxa

$\mathrm{TL} = 3.18$

Intermediate tree

56 taxa

$\mathrm{TL} = 10.54$

Long tree

121 taxa

$\mathrm{TL} = 16.71$

Simulation setup

Empirical trees

Short tree

21 taxa

$\mathrm{TL} = 3.18$

Intermediate tree

56 taxa

$\mathrm{TL} = 10.54$

Long tree

121 taxa

$\mathrm{TL} = 16.71$

Data-generating models

Mk

no rate variation

Mk + ACRV ($m = 4$)

static rate variation

$\sigma = 1.17, \delta = 0$

Covariomorph ($m = 4$)

dynamic rate variation

$\sigma = 1.17$

$\delta \in \{0.005, 0.05, 0.5, 1, 2, 3, 4, 5\}$

Simulation setup

Empirical trees

Short tree
21 taxa
$\mathrm{TL} = 3.18$

Intermediate tree
56 taxa
$\mathrm{TL} = 10.54$

Long tree
121 taxa
$\mathrm{TL} = 16.71$

Data-generating models

Mk
no rate variation

Mk + ACRV $(m = 4)$
static rate variation
$\sigma = 1.17, \delta = 0$

Covariomorph $(m = 4)$
dynamic rate variation
$\sigma = 1.17$
$\delta \in \{0.005, 0.05, 0.5, 1, 2, 3, 4, 5\}$

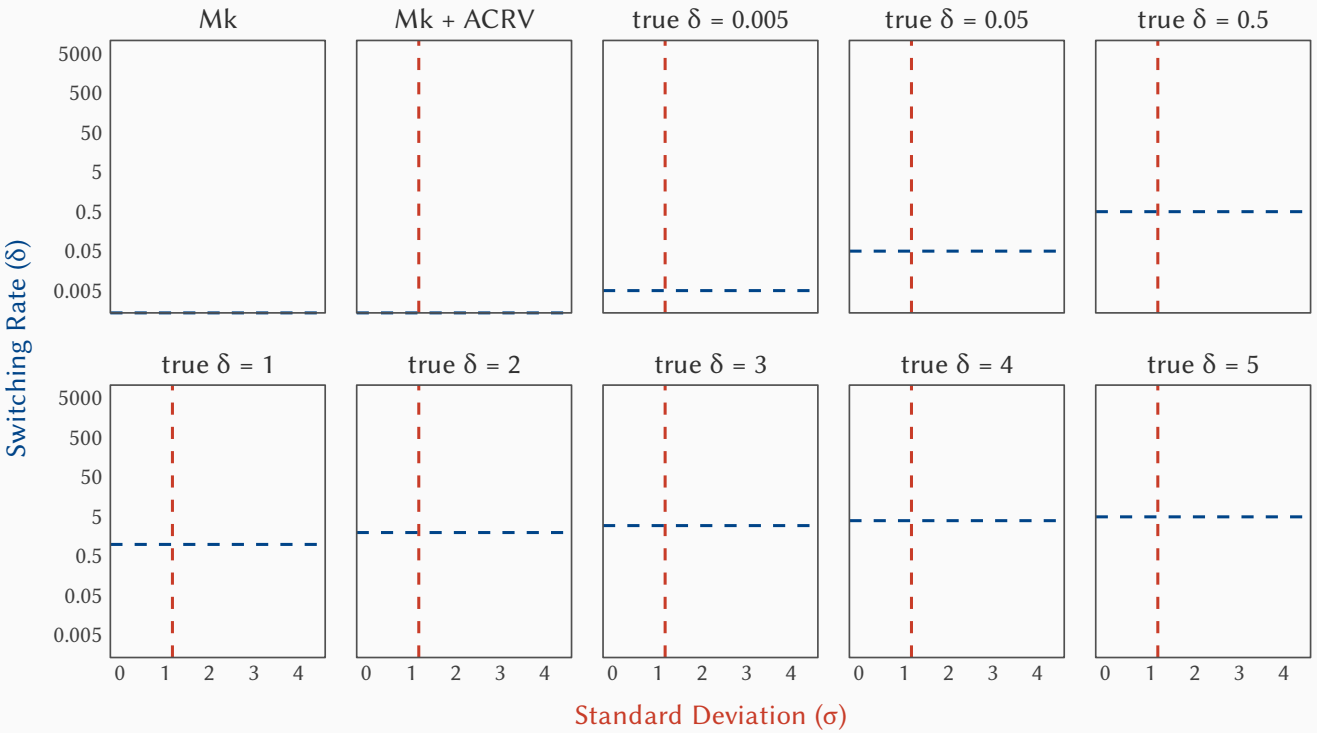
Inference

Covariomorph $(m = 4)$
focusing on two parameters: σ and δ

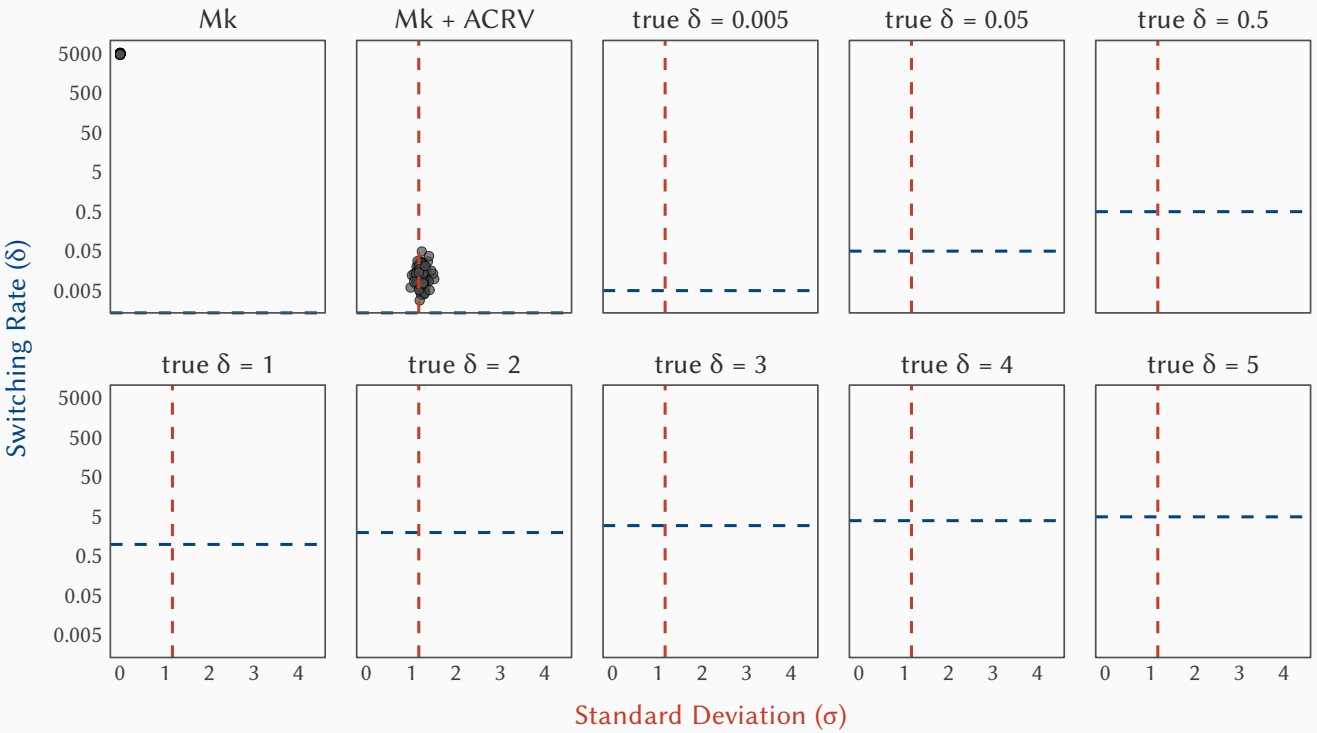
Prior on σ
$1/\sigma \sim \mathrm{Uniform}(0, 10^3)$

Prior on δ
$\delta \sim \mathrm{Uniform}(0, 10^4)$

Simulation Results

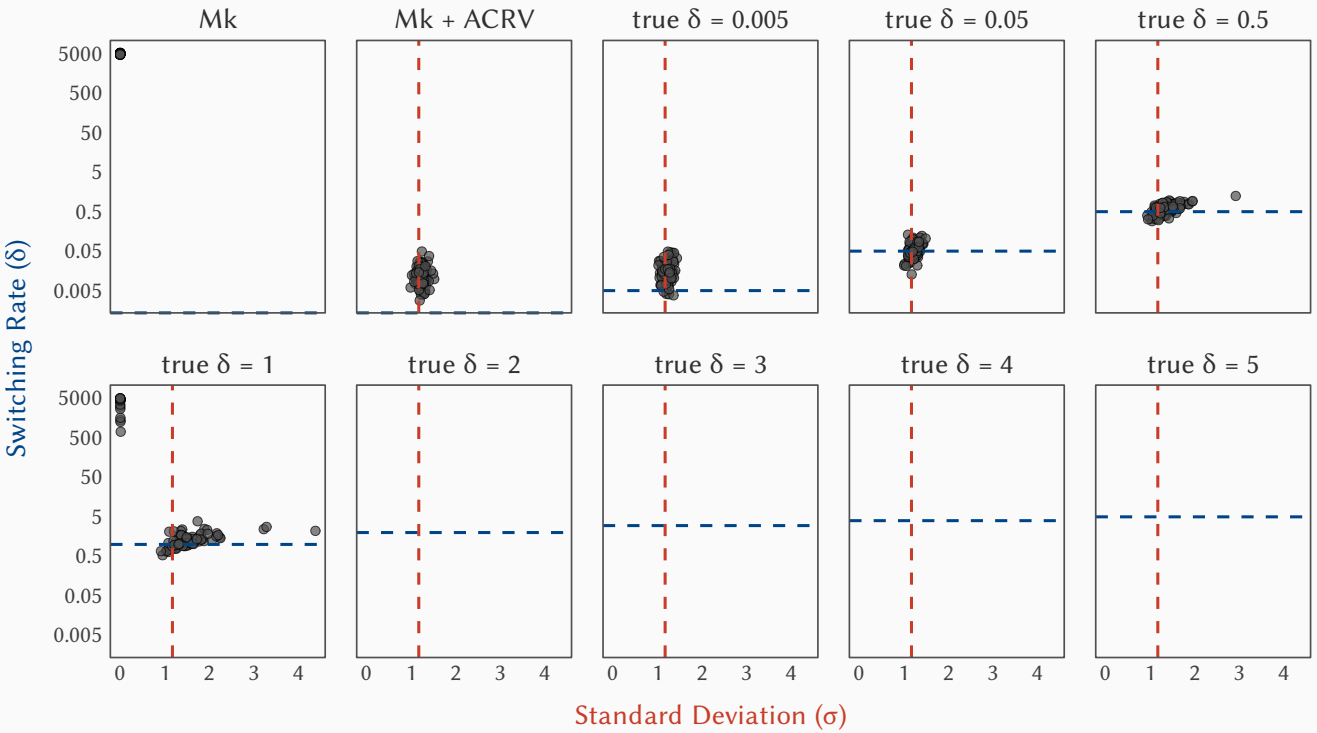


Simulation Results



Mk and ACRV data:
no false heterotachy.

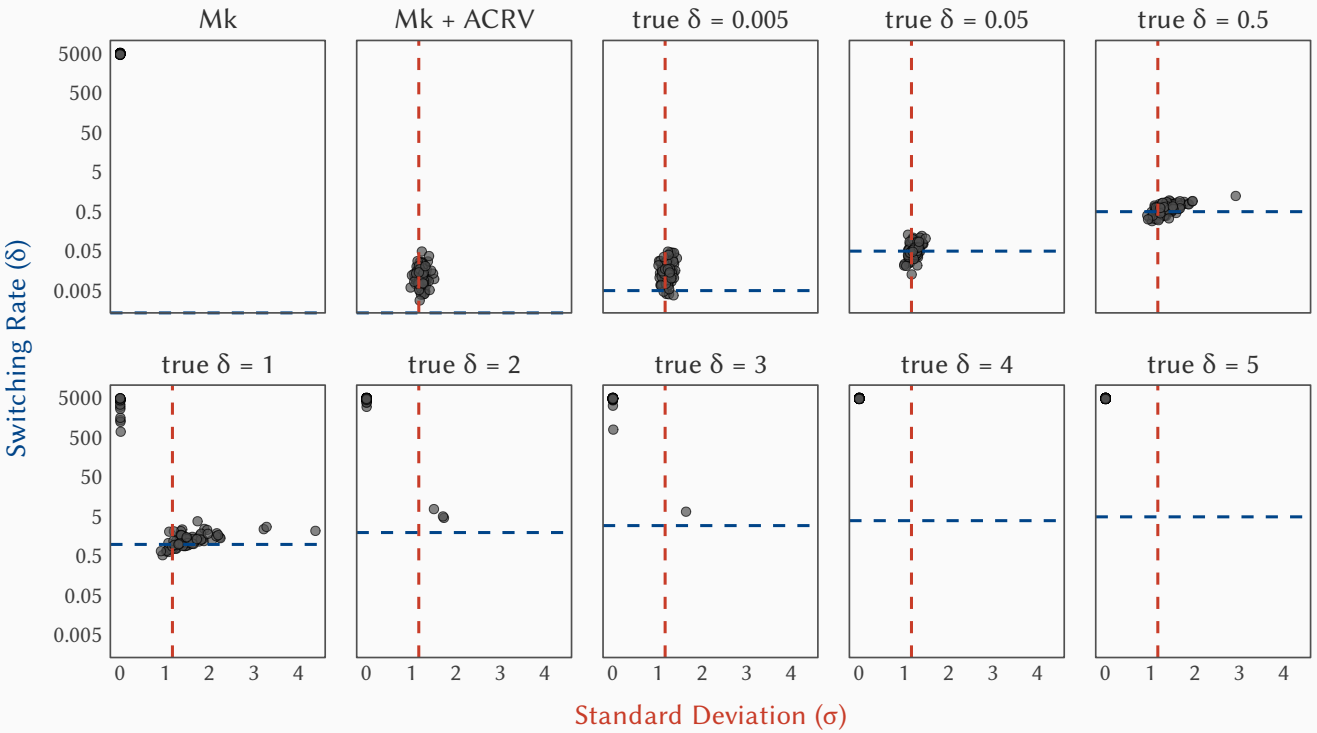
Simulation Results



Mk and ACRV data:
no false heterotachy.

$\delta \leq 1$: both σ and δ recovered.
Very slow switching (≈ 0.005) looks like ACRV.

Simulation Results



Mk and ACRV data:
no false heterotachy.

$\delta \leq 1$: both σ and δ recovered.
Very slow switching (≈ 0.005) looks like ACRV.

$\delta \geq 2$: too fast to detect.
$\sigma \to 0$, δ stays at its prior: looks like Mk.

σ: how different are the rate categories?



Category rates come from a discretized lognormal with mean 1; σ sets the **magnitude of rate variation**.

$\sigma \approx 0$: all categories share one rate, so there is nothing to switch between: the model is **Mk-like**.

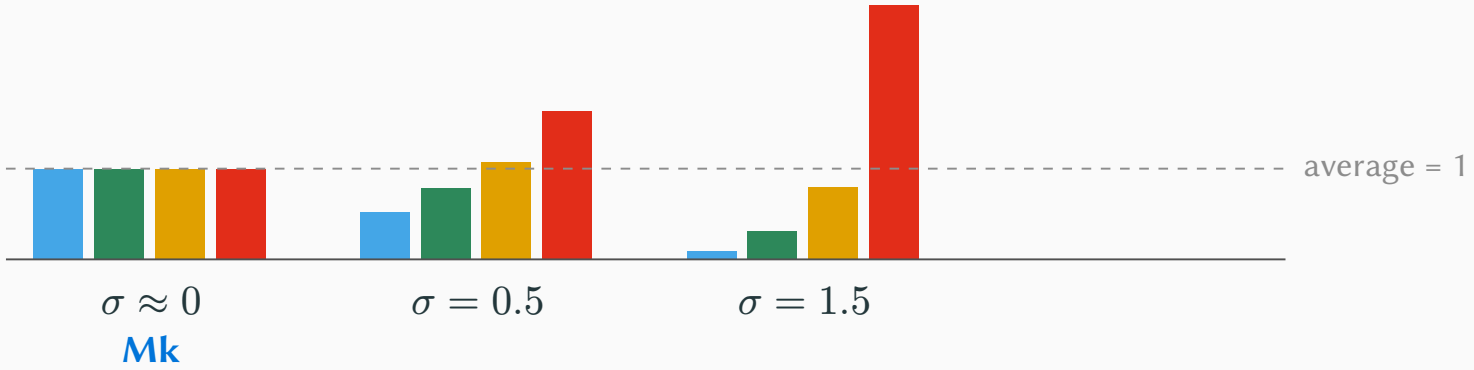
σ: how different are the rate categories?



Category rates come from a discretized lognormal with mean 1; σ sets the **magnitude of rate variation**.

$\sigma \approx 0$: all categories share one rate, so there is nothing to switch between: the model is **Mk-like**.

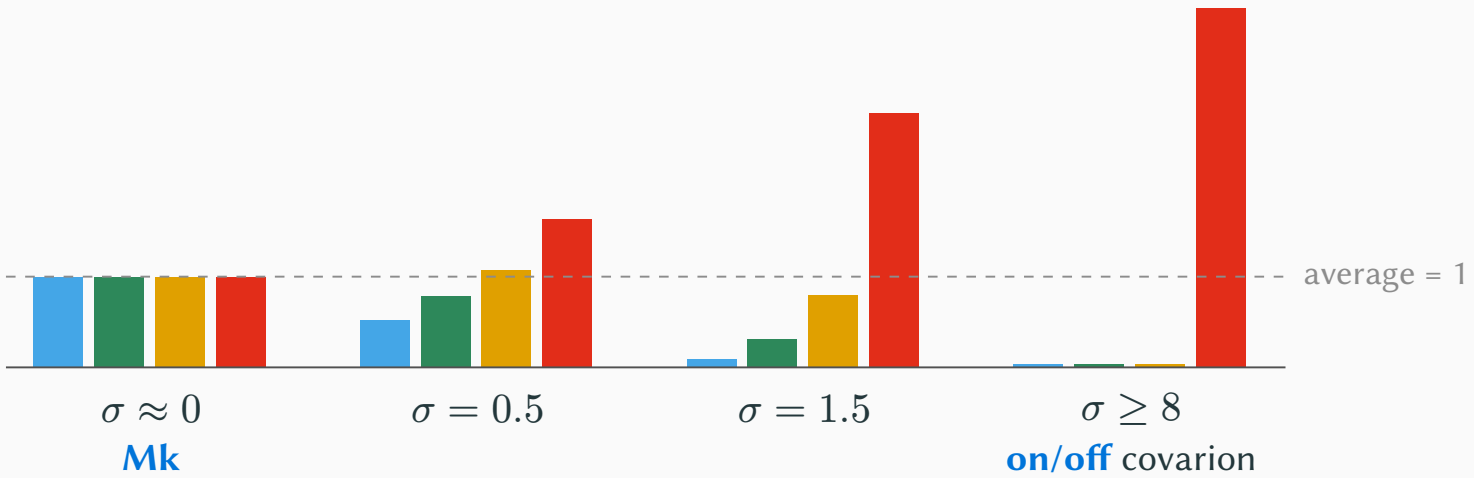
σ: how different are the rate categories?



Category rates come from a discretized lognormal with mean 1; σ sets the **magnitude of rate variation**.

$\sigma \approx 0$: all categories share one rate, so there is nothing to switch between: the model is **Mk-like**.

σ: how different are the rate categories?

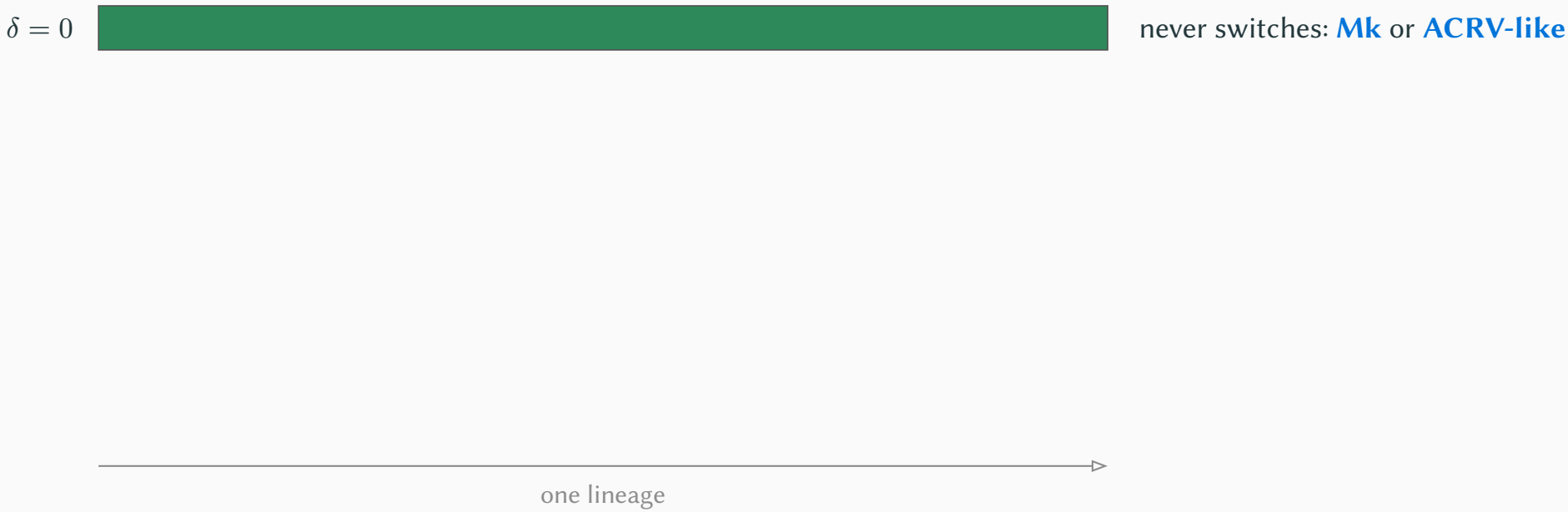


Category rates come from a discretized lognormal with mean 1; σ sets the **magnitude of rate variation**.

$\sigma \approx 0$: all categories share one rate, so there is nothing to switch between: the model is **Mk-like**.

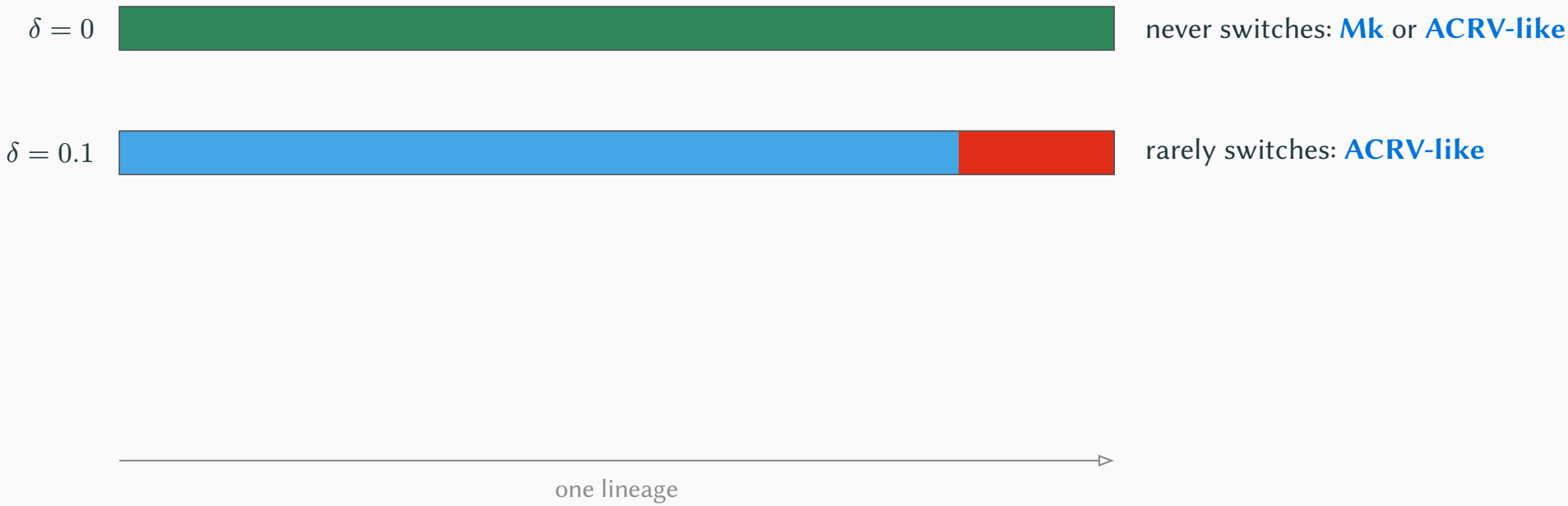
Large σ: most categories are nearly *frozen* and one is fast: similar to the "on/off" covarion.

δ: how often does a character switch rate category?



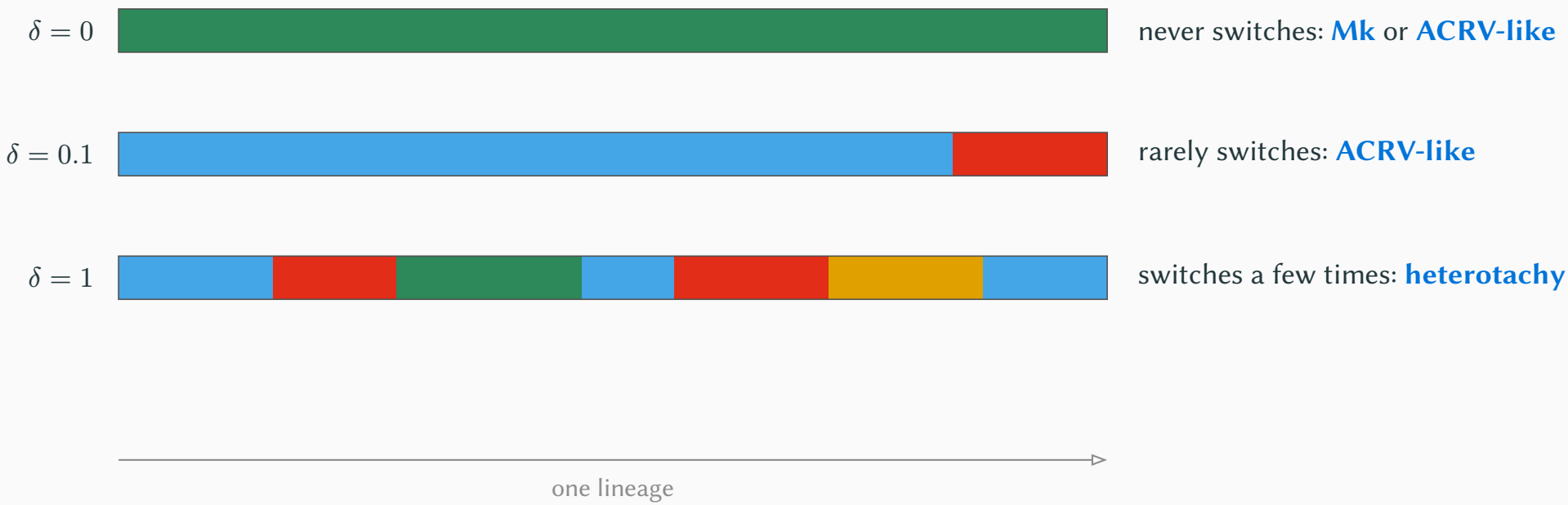
δ is scaled *per change of state*: $\delta = 1$ means about **one category switch for every change** in the character.

δ: how often does a character switch rate category?



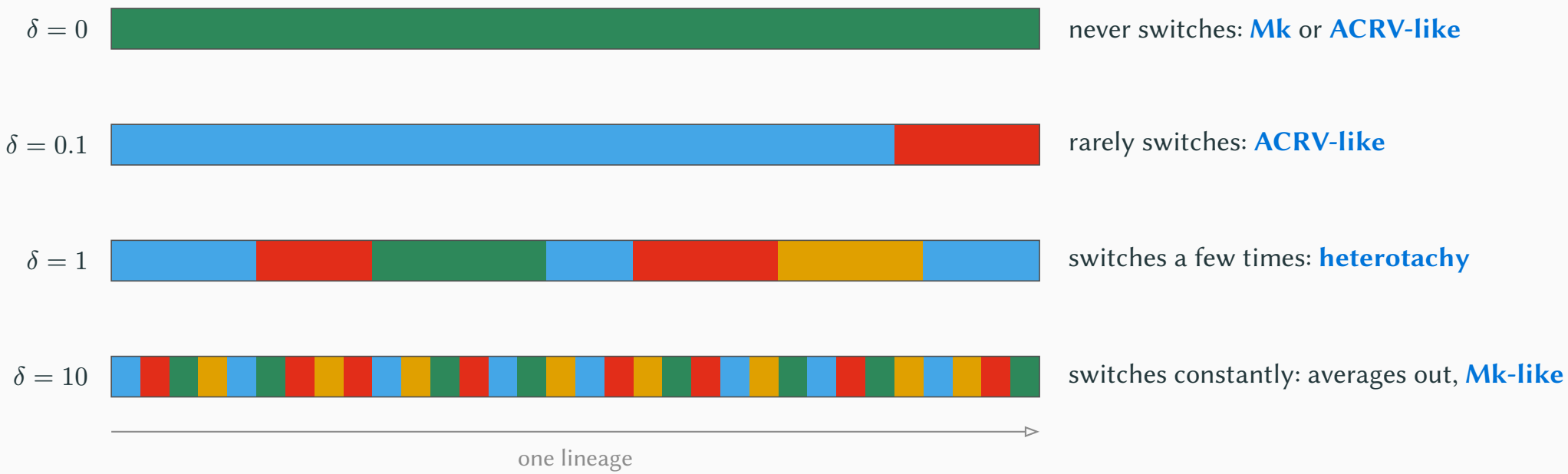
δ is scaled *per change of state*: $\delta = 1$ means about **one category switch for every change** in the character.

δ: how often does a character switch rate category?



δ is scaled *per change of state*: $\delta = 1$ means about **one category switch for every change** in the character.

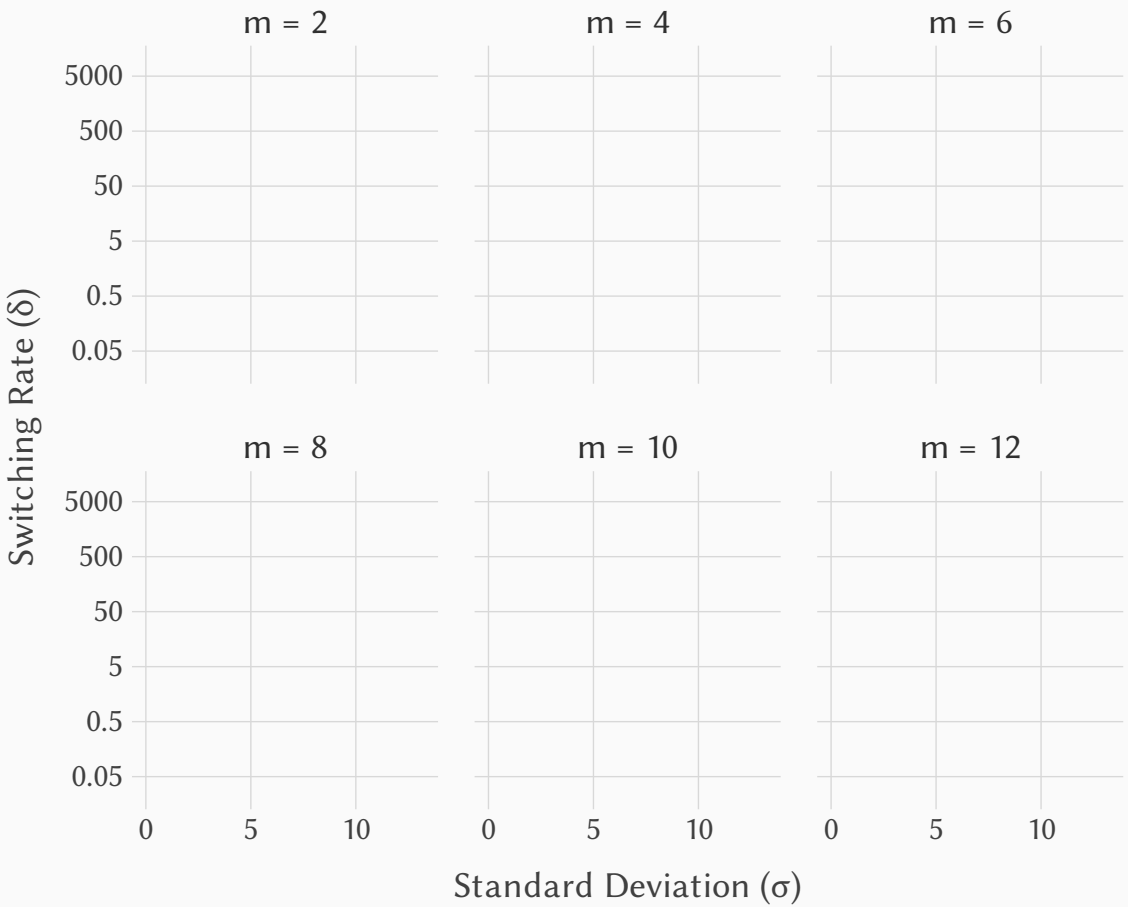
δ: how often does a character switch rate category?



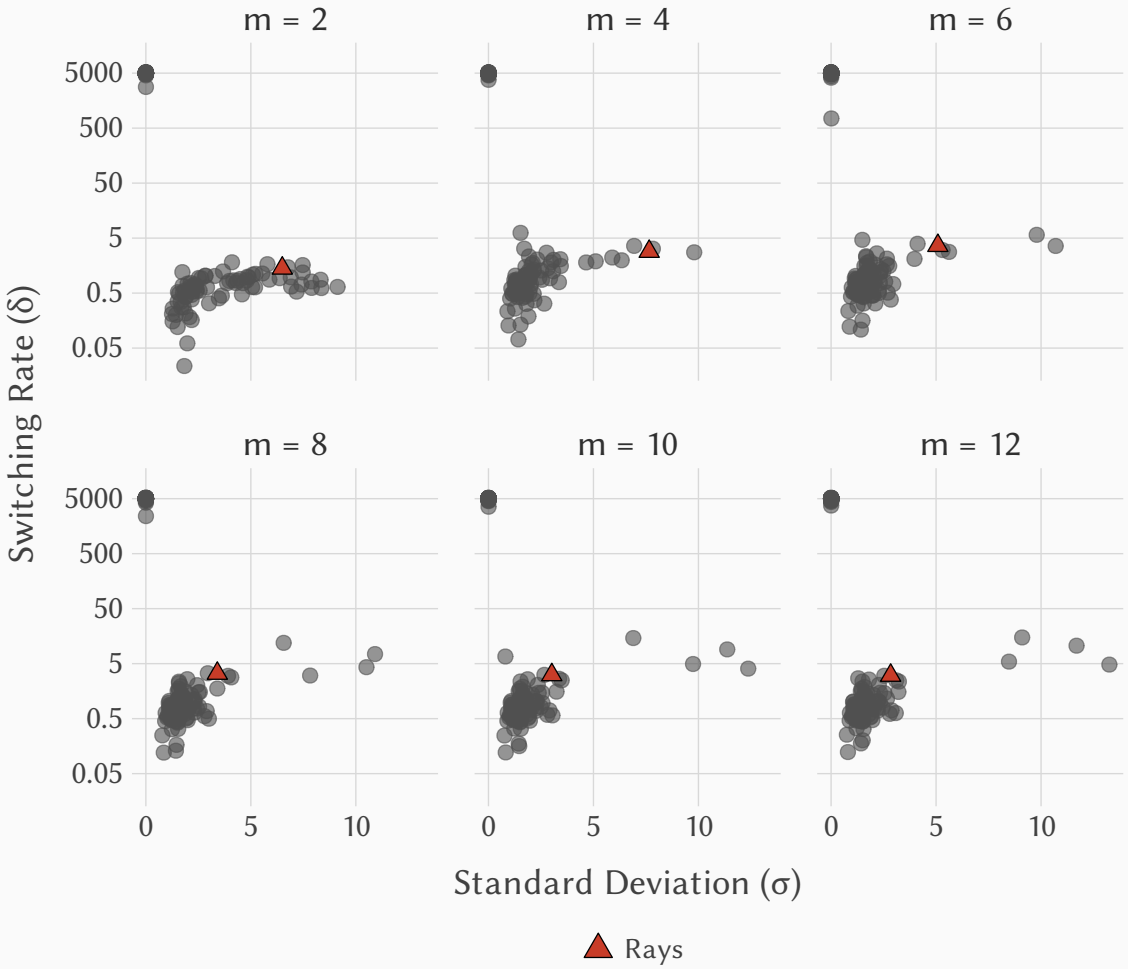
δ is scaled *per change of state*: $\delta = 1$ means about **one category switch for every change** in the character.

Mk and ACRV are **special cases** of the covariomorph: without signal for switching, the model reduces to the simpler one.

Is heterotachy present in empirical matrices?

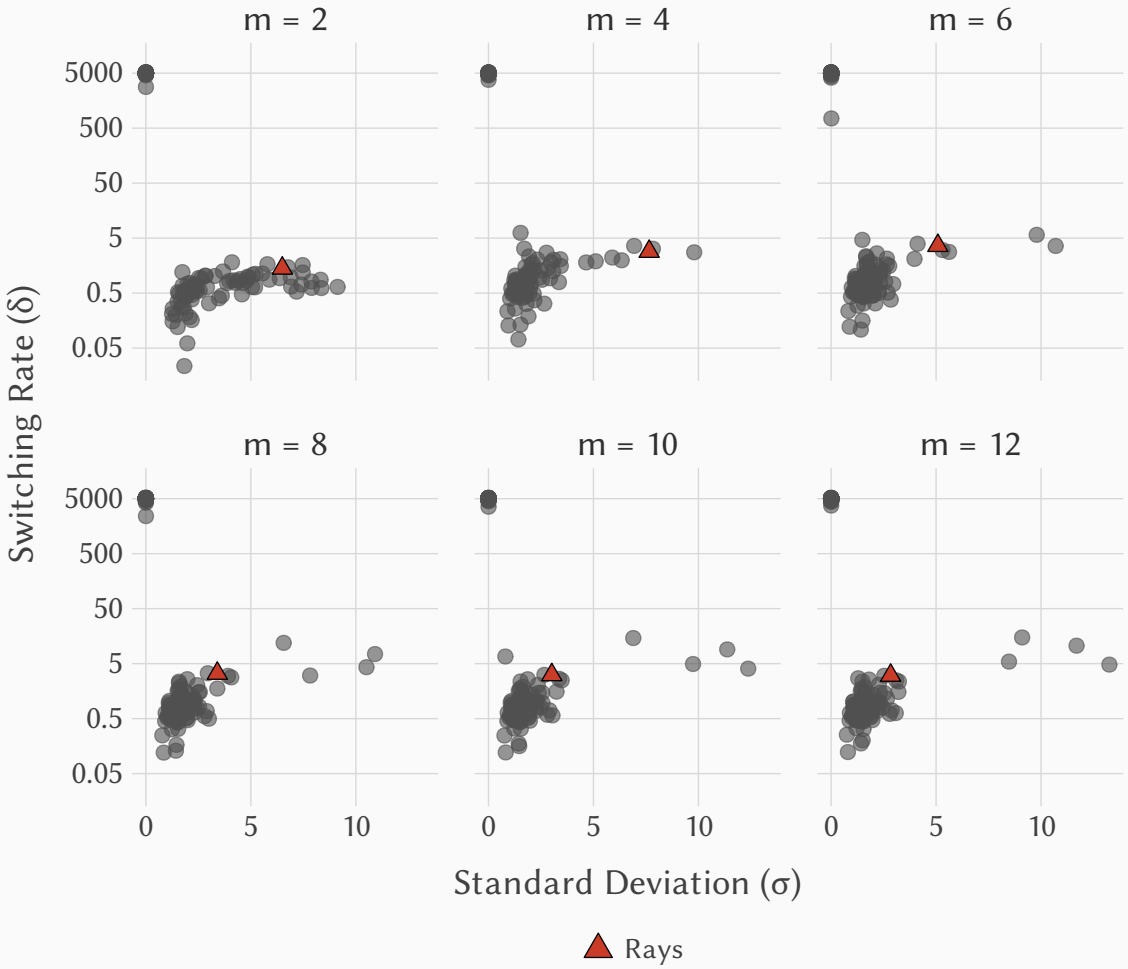


Is heterotachy present in empirical matrices?



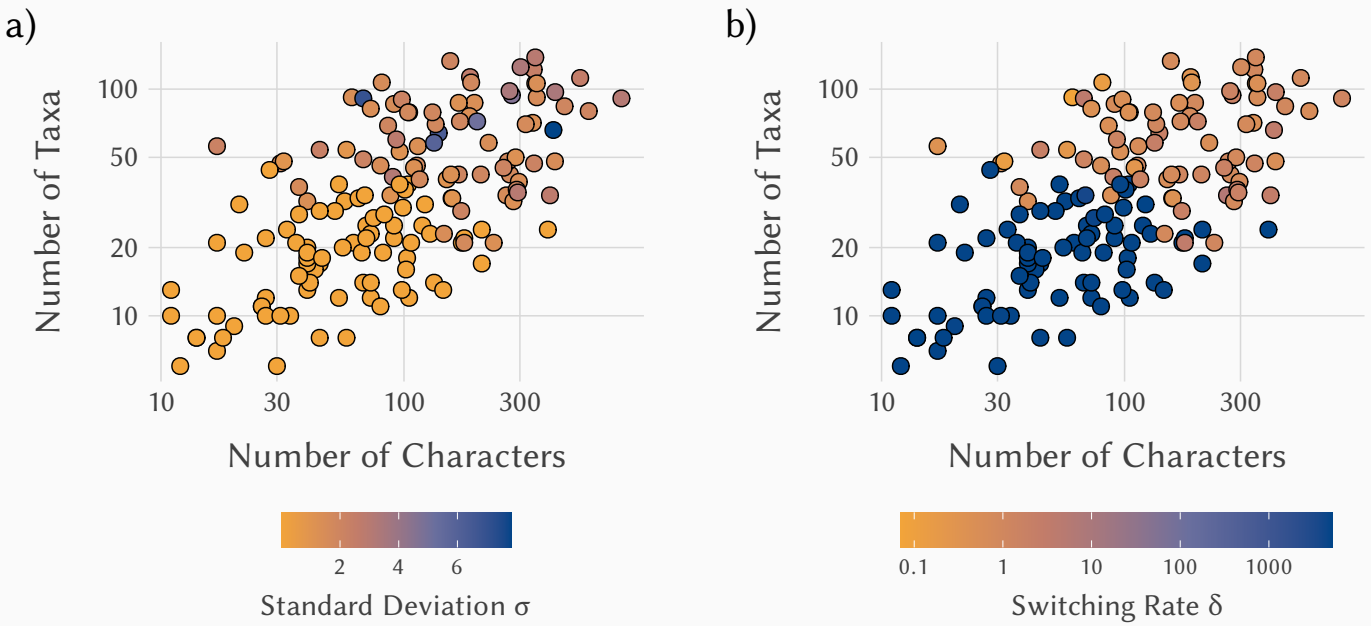
- 50% of datasets cluster near the prior mean, showing little deviation from homogenous rates.

Is heterotachy present in empirical matrices?



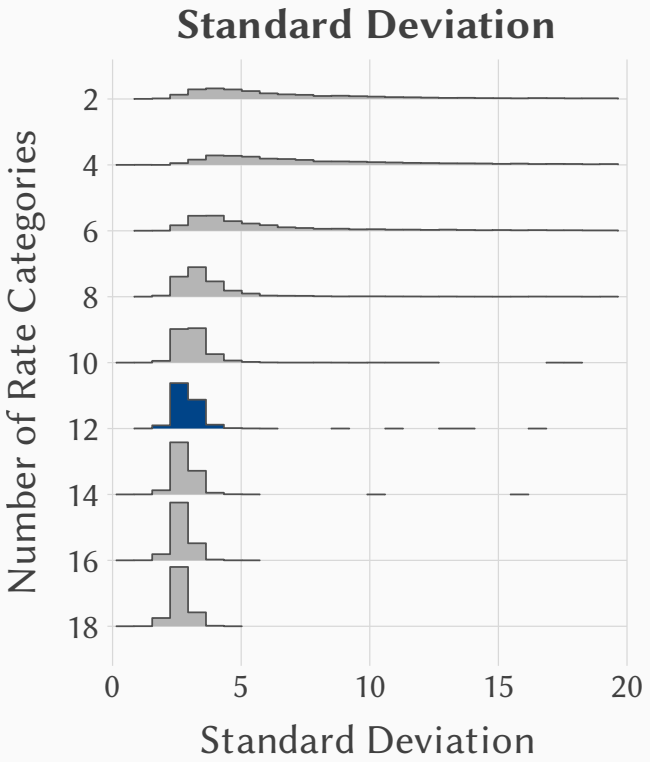
- 50% of datasets cluster near the prior mean, showing little deviation from homogenous rates.
- Remaining 50% display signs for rate heterogeneity.
  - a mix of ACRV and Covariomorph dynamics

Is my matrix big enough?



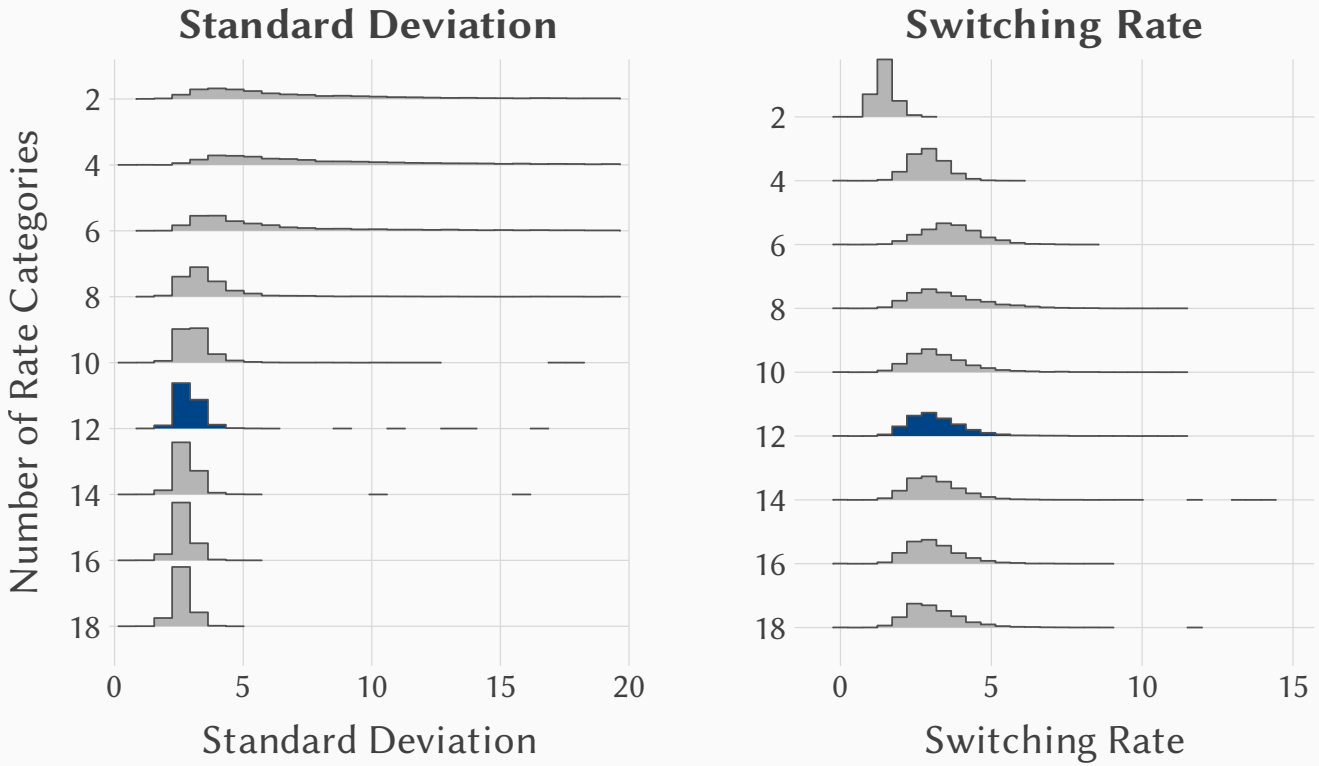
- Small matrices mostly show $\sigma \approx 0$ and a δ that *stays at its prior mean.*
- Switching becomes detectable once there are **enough taxa and characters** to record many changes.

Parameter estimates



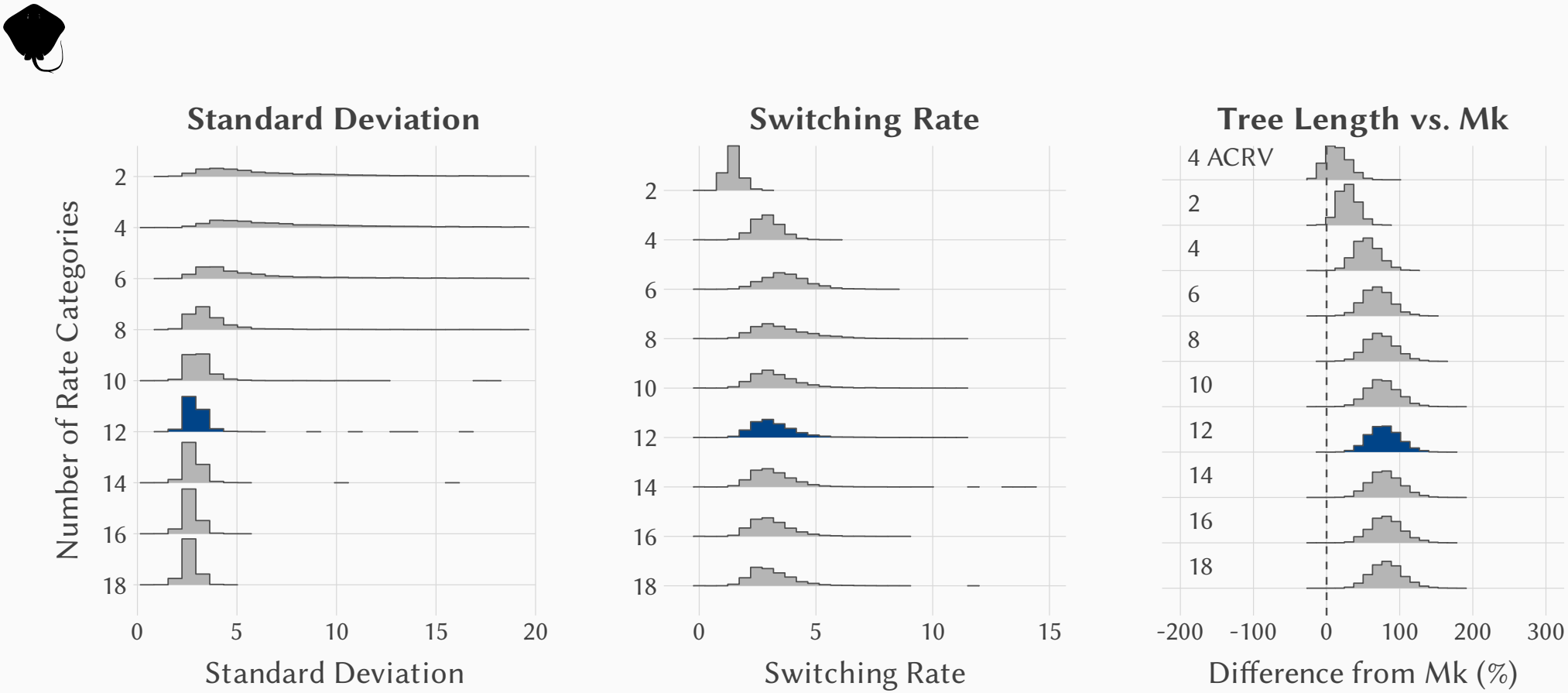
Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Parameter estimates



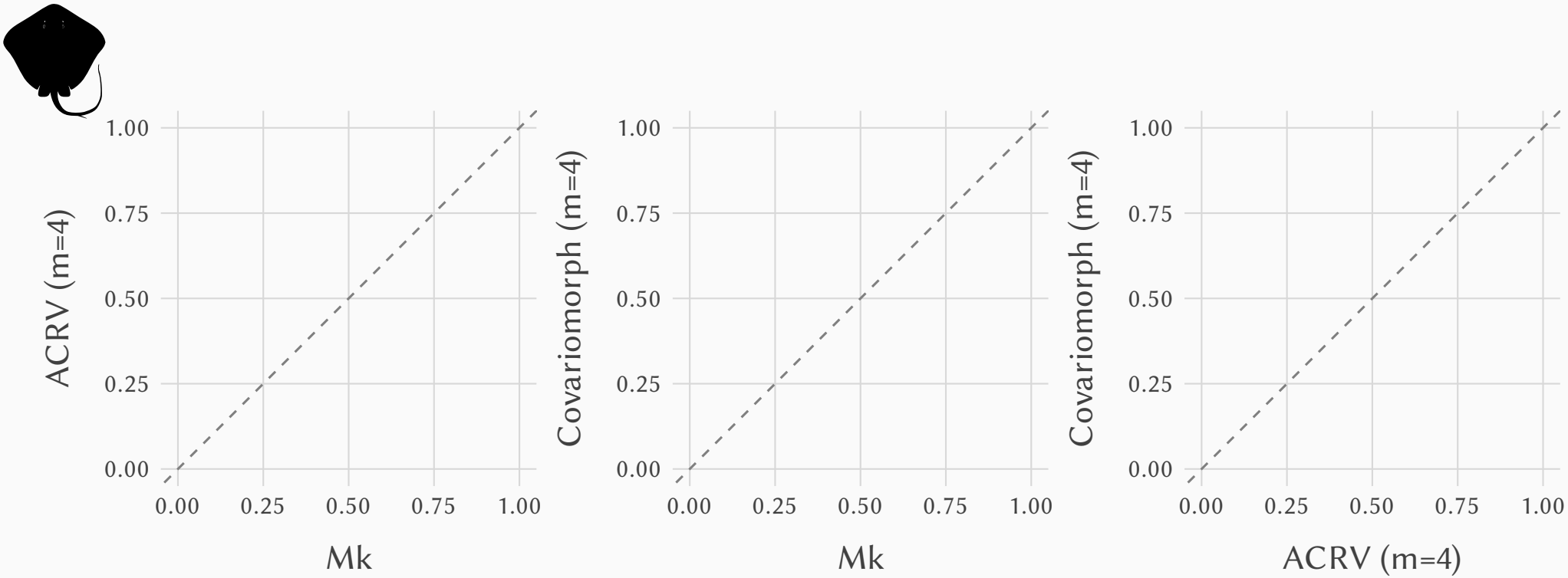
Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Parameter estimates



Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Topological impact

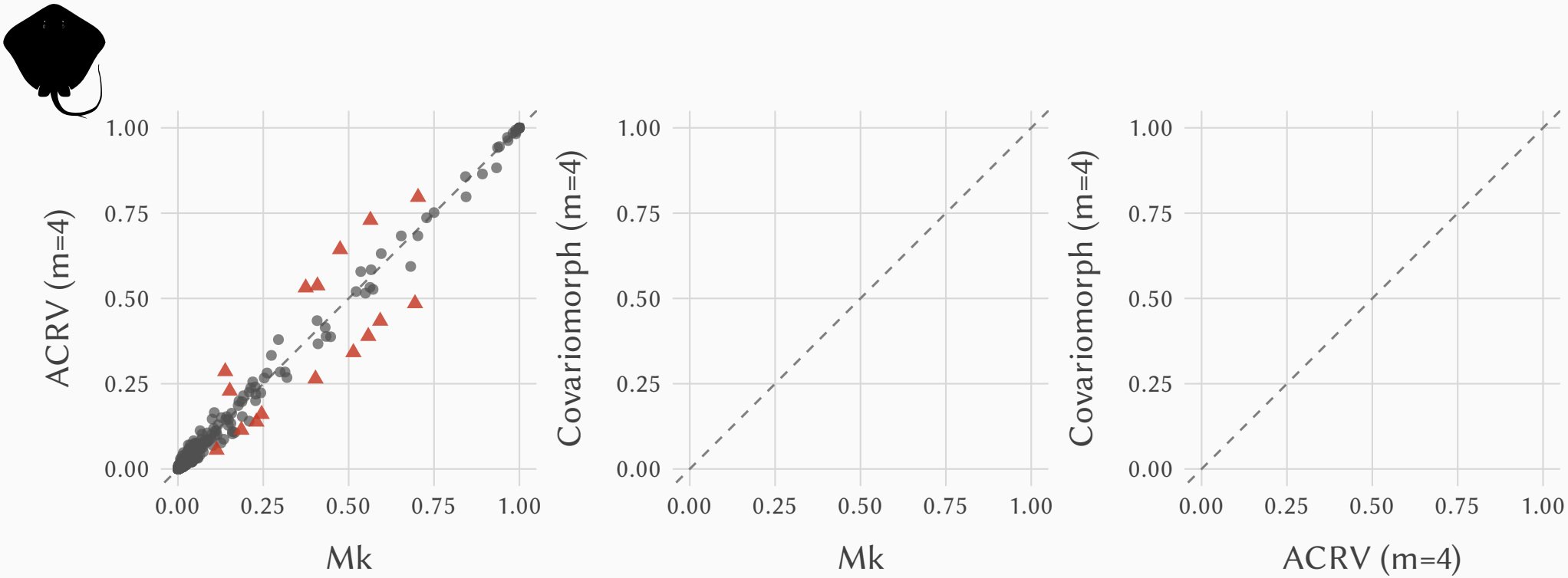


Each point = one clade

▲ = support differs by more than MCMC noise

Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Topological impact

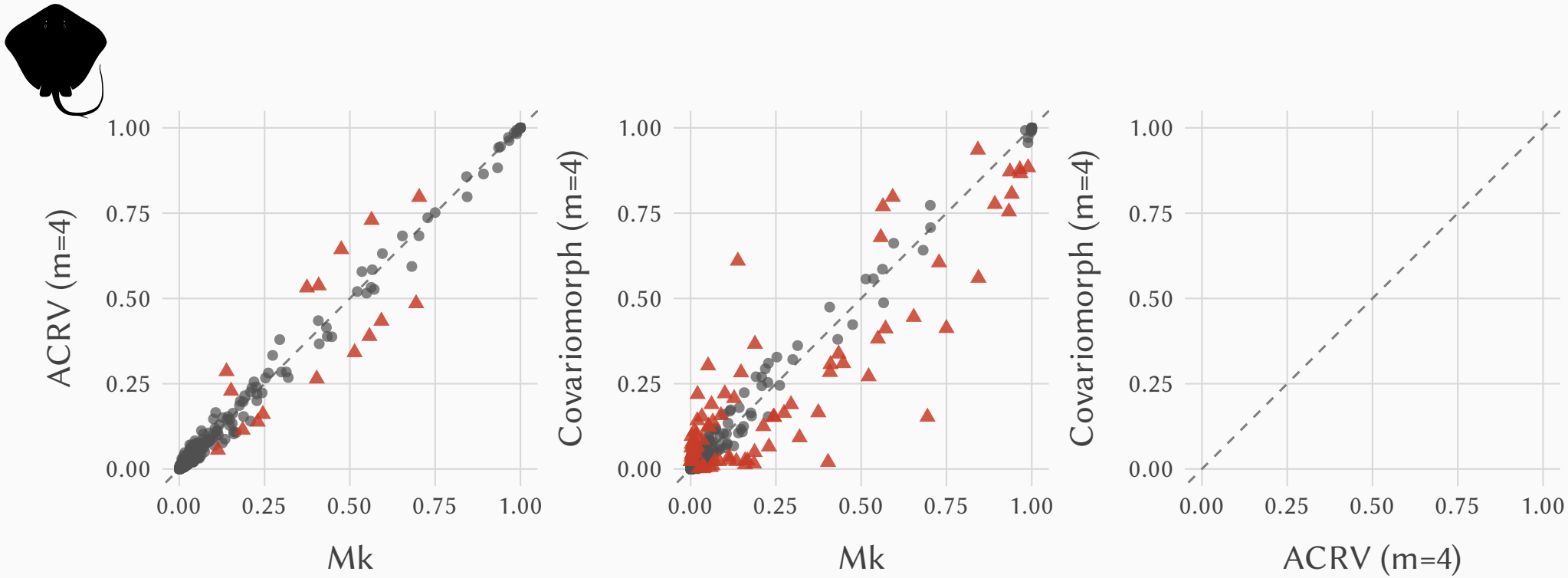


Each point = one clade

▲ = support differs by more than MCMC noise

Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Topological impact

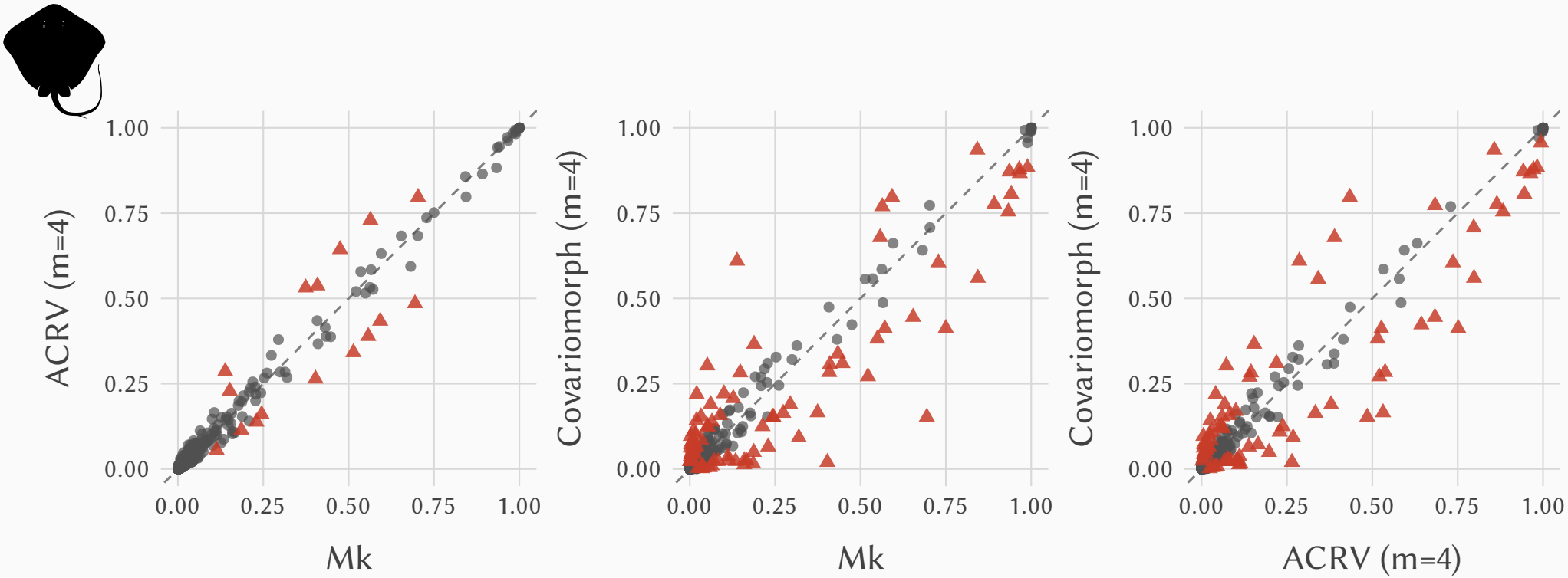


Each point = one clade

▲ = support differs by more than MCMC noise

Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Topological impact



Each point = one clade

▲ = support differs by more than MCMC noise

Rays, 52 taxa × 124 characters (Marramà et al. 2023). Khakurel & Höhna (2026)

Take-home messages

Heterotachy is common

About **half** of 164 published matrices show characters switching rate.

It changes inference

Branch lengths and clade support shift: in rays, tree length **5.3 → 8.2**.

The covariomorph model explicitly models heterotachy and thus offers a more flexible and potentially more realistic representation of morphological evolution by accounting for temporal variation.

· ascertainment-bias correction · decoupling switching from state change · per-character switching

Thank you!

b.khakurel@lmu.de basantakhakurel.github.io basantakhakurel

Khakurel & Höhna (2026) *Systematic Biology*, 10.1093/sysbio/syag014



Tutorial: revbayes.github.io/tutorials/morph_tree/covariomorph.html

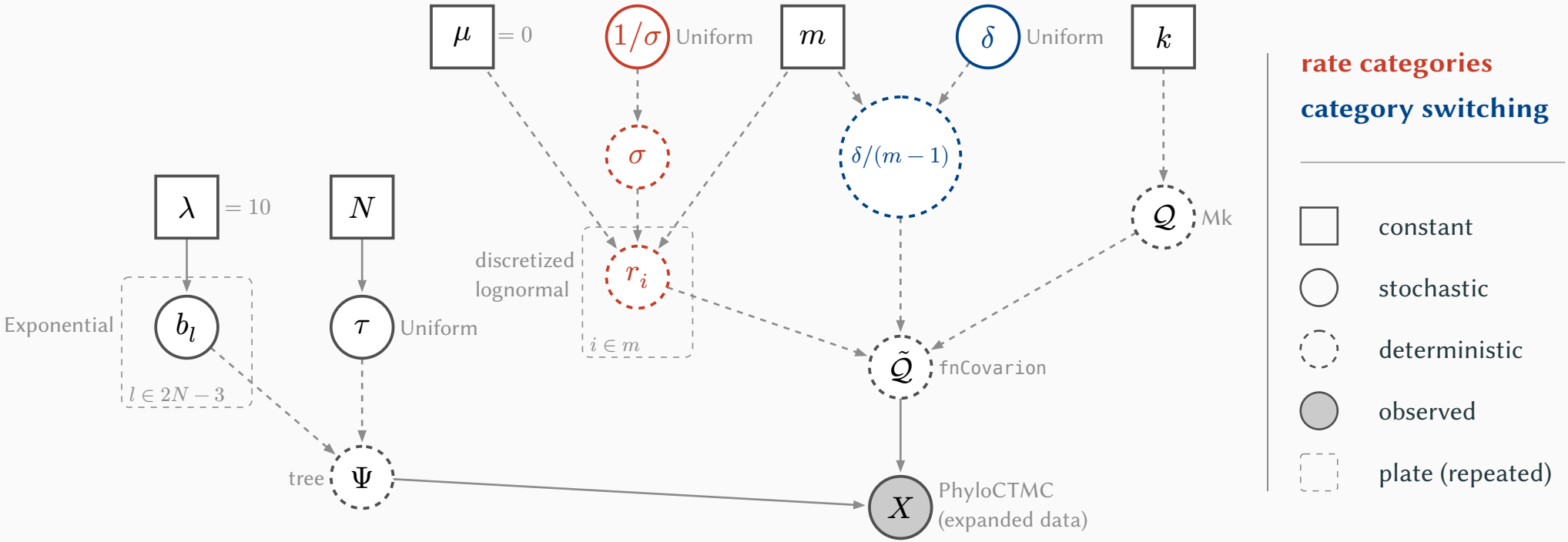
Slides: basantakhakurel.github.io/presentations







The covariomorph model as a graph



References

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *Plos Biology*, *4*(5), e88.

Fitch, W. M., & Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics*, *4*, 579–593.

Galtier, N. (2001). Maximum-likelihood phylogenetic analysis under a covarion-like model. *Molecular Biology and Evolution*, *18*(5), 866–873.

Harmon, L. J. (2022). *The Mk Model*. University of Idaho.

Harrison, L. B., & Larsson, H. C. E. (2015). Among-Character Rate Variation Distributions in Phylogenetic Analysis of Discrete Morphological Characters. *Systematic Biology*, *64*(2), 307–324. https://doi.org/10.1093/sysbio/syu098

Huelsenbeck, J. P. (2002). Testing a covariotide model of DNA substitution. *Molecular Biology and Evolution*, *19*(5), 698–707.

Khakurel, B., & Höhna, S. (2026). A covarion model for phylogenetic estimation using discrete morphological datasets. *Systematic Biology*, syag014. https://doi.org/10.1093/sysbio/syag014

Lee, M. S., Cau, A., Naish, D., & Dyke, G. J. (2014). Morphological clocks in paleontology, and a mid-Cretaceous origin of crown Aves. *Systematic Biology*, *63*(3), 442–449.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*(6), 913–925.

Lockhart, P., Steel, M., Barbrook, A., Huson, D., Charleston, M., & Howe, C. (1998). A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Molecular Biology and Evolution*, *15*(9), 1183–1188.

Marramà, G., Villalobos-Segura, E., Zorzin, R., Kriwet, J., & Carnevale, G. (2023). The evolutionary origin of the durophagous pelagic stingray ecomorph. *Palaeontology*, *66*(4), e12669.

Tuffley, C., & Steel, M. (1998). Modeling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences*, *147*(1), 63–91.

Wang, M., & Lloyd, G. T. (2016). Rates of morphological evolution are heterogeneous in Early Cretaceous birds. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1828), 20160214.

Warnock, R. C., & Wright, A. M. (2020). *Understanding the tripartite approach to Bayesian divergence time estimation*. Cambridge University Press.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, *39*, 306–314.