

The Invariant Nature of a Morphological Character and Character State: Insights from Gene Regulatory Networks

SERGEI TARASOV

The tree inference was performed using the Bayesian framework in *RevBaes* (Höhna *et al.*, 2016) and for the imaginary species from Fig. 2F. The data were coded using (a) binary coding corresponding to four binary characters from Fig. 2C, (b) multistate coding with penalty corresponding to one multistate character from Fig. 2C [i.e, using structured Markov models *sensu* Tarasov (2019)], (c) multistate coding without penalty corresponding to one multistate character where all transitions between states are possible.

To demonstrate the consequences of the different approaches, I used two datasets (see the scripts in Supplementary Materials): (i) dataset of four two-state characters for the coding approach (a); and (ii) dataset of one five-state character for the coding approach (b) and (c). The datasets were constructed by replicating characters in datasets (i) and (ii) 10 times. The replication was done to make data sufficiently informative for the inference. The dataset of character (i) was analyzed using *Mk2* model (Lewis, 2001), while the dataset from the character (ii) using *Mk5* model that does not incorporate penalty and *Mk5*-like model (Matrices 1-2) that incorporates the penalty. The *Mk5* model with penalty was constructed as shown on Fig. 2E; this construction combines elementary characters from Fig. 2C as independently evolving (Pagel, 1994; Shelton & Ciardo, 2014; Tarasov, 2019). This *Mk5* model with penalty consists of 16 states (Matrix 1) that is larger in comparison to the nine states shown in Fig. 2E. This difference occurs because some states in Matrix 1 correspond to hidden states (Tarasov, 2019); however, for the demonstrative purpose of current paper those hidden states can be omitted without loss of generality. Thus, the states of Matrix 1 which were not observed in the data (i.e., combinations of states which are absent in species in Fig. 2F) were removed that rendered the models to include only five states (Matrix 2). The analyses were run for 10^6 generations, using the exponential prior for branch lengths $\sim \text{Exp}(0.2)$.

References

- Höhna S, Landis MJ, Heath TA, et al. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology*: syw021.
- Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic biology* 50: 913–925.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London B: Biological Sciences* 255: 37–45.
- Shelton CR & Ciardo G. 2014. Tutorial on Structured Continuous-Time Markov Processes. *J. Artif. Intell. Res.(JAIR)* 51: 725–778.

Matrix 1. The rate matrix describing evolution of GRN modues in Fig 2E. B1, B2, G1, G2 correspond to beta1,beta3, gamma1, gamma2 modules respectively. B1a (and others) reads as beta1 module absent; B1p (and others) reads as beta1 module present.

	B1aB2aG1aG2a	B1aB2aG1aG2p	B1aB2aG1pG2a	B1aB2aG1pG2p	B1aB2pG1aG2a	B1aB2pG1aG2p	B1aB2pG1pG2a	B1aB2pG1pG2p	B1pB2aG1aG2a	B1pB2aG1aG2p	B1pB2aG1pG2a	B1pB2aG1pG2p	B1pB2pG1aG2a	B1pB2pG1aG2p	B1pB2pG1pG2a	B1pB2pG1pG2p
B1aB2aG1aG2a	NA	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0
B1aB2aG1aG2p	1	NA	0	1	0	1	0	0	0	1	0	0	0	0	0	0
B1aB2aG1pG2a	1	0	NA	1	0	0	1	0	0	0	1	0	0	0	0	0
B1aB2aG1pG2p	0	1	1	NA	0	0	0	1	0	0	0	1	0	0	0	0
B1aB2pG1aG2a	1	0	0	0	NA	1	1	0	0	0	0	0	1	0	0	0
B1aB2pG1aG2p	0	1	0	0	1	NA	0	1	0	0	0	0	0	1	0	0
B1aB2pG1pG2a	0	0	1	0	1	0	NA	1	0	0	0	0	0	0	1	0
B1aB2pG1pG2p	0	0	0	1	0	1	1	NA	0	0	0	0	0	0	0	1
B1pB2aG1aG2a	1	0	0	0	0	0	0	0	NA	1	1	0	1	0	0	0
B1pB2aG1aG2p	0	1	0	0	0	0	0	0	1	NA	0	1	0	1	0	0
B1pB2aG1pG2a	0	0	1	0	0	0	0	0	1	0	NA	1	0	0	1	0
B1pB2aG1pG2p	0	0	0	1	0	0	0	0	0	1	1	NA	0	0	0	1
B1pB2pG1aG2a	0	0	0	0	1	0	0	0	1	0	0	0	NA	1	1	0
B1pB2pG1aG2p	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1
B1pB2pG1pG2a	0	0	0	0	0	0	1	0	0	0	1	0	1	0	NA	1
B1pB2pG1pG2p	0	0	0	0	0	0	0	1	0	0	0	1	0	1	1	NA

Matrix 2. The rate matrix describing evolution of GRN modues in Fig 2E. This matrix was derived from Matrix 1 by removing the unobserved states in Fig. 2F.

	B1aB2aG1aG2a	B1pB2aG1aG2a	B1pB2aG1pG2a	B1pB2pG1pG2a	B1pB2pG1pG2p
B1aB2aG1aG2a	NA	1	0	0	0
B1pB2aG1aG2a	1	NA	1	0	0
B1pB2aG1pG2a	0	1	NA	1	0
B1pB2pG1pG2a	0	0	1	NA	1
B1pB2pG1pG2p	0	0	0	1	NA