Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits

Sergei Tarasov

National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA

E-mail: sergxf@yandex.ru

The author declares no conflict of interest.

ABSTRACT

Modeling discrete phenotypic traits for either ancestral character state reconstruction or morphology-based phylogenetic inference suffers from ambiguities of character coding, homology assessment, dependencies, and selection of adequate models. These drawbacks occur because trait evolution is driven by two key processes – hierarchical and hidden – which are not accommodated simultaneously by the available phylogenetic methods. The hierarchical process refers to the dependencies between anatomical body parts, while the hidden process refers to the evolution of gene regulatory networks (GRNs) underlying trait development. Herein, I demonstrate that these processes can be efficiently modeled using structured Markov models equipped with hidden states, which resolves the majority of the problems associated with discrete traits. Integration of structured Markov models with anatomy ontologies can adequately incorporate the hierarchical dependencies, while the use of the hidden states accommodates hidden GRN evolution and mutation rate heterogeneity. This model is insensitive to alternative coding approaches, which is shown by using a previously unknown "two-scientist paradox", and solving the tail color problem. The practical considerations for implementing this model in phylogenetic inference and comparative methods are demonstrated and discussed.

KEY WORDS:

discrete trait, character, morphology, homology, anatomy ontology, structured Markov models, hidden Markov models, lumpability, gene regulatory networks

The processes driving trait evolution are crucial for understanding evolutionary radiations (Price et al. 2010; Van Bocxlaer et al. 2010; Tobias et al. 2014), origin of complexity and novelty (Moczek 2008; Ramirez and Michalik 2014), and inferring phylogenies. To make a phenotypic trait analyzable, we need to *(1)* discretize it (delimit the trait within a phenotype), *(2)* assess its primary homology (similarity), and finally *(3)* encode the trait (observations) into a character string (see the definitions in Box 1). This procedure, called *character construction* (Wiens 2001), exhibits a basic stage of any analysis, and has a profound influence on all downstream stages. Despite the plethora of phylogenetic methods - be it parsimony, likelihood or Bayesian framework [reviewed in O'Meara (2012)] - the lack of repeatable and agreed-upon approaches for character construction generates considerable ambiguity. As a result, different hypotheses of discretization and different ways of coding the same hypothesis into a character may be proposed for the same trait (Hawkins et al. 1997a; Strong and Lipscomb 1999a; Ramirez 2007; Agnarsson and Coddington 2008). This disagreement naturally leads to inconsistent phylogenetic results [reviewed in Brazeau (2011)].

The ambiguity of character construction stems out from two key processes – hidden and hierarchical. The hidden process refers to the evolution of gene regulatory networks (GRNs) which underlay trait development (Wagner 2007; Carroll 2008; Houle et al. 2010); it implies that the actual driver of trait evolution is hidden from the direct observation of morphology. The hierarchical process refers to the evolution of hierarchical relationships between traits that occur due to dependencies among anatomical body parts. For example, digits are located on limbs; loss of the limbs during evolution simultaneously triggers loss of the digits. Evolution of developmental programs in organisms causes interactions between the hierarchical and hidden processes, making them into simultaneous drivers of trait changes. Unfortunately, the available phylogenetic methods do not accommodate these processes simultaneously. Thus, development of methods capable of concurrently modeling the two processes would automatically resolve the ambiguity associated with character construction. To tackle this problem, I propose a new integrative framework that uses the theory of structured Markov models [SMM, (Nodelman et al. 2002)], hidden Markov models [HMM, (Beaulieu et al. 2013)] and knowledge of organismal anatomies from anatomy ontologies. I assess the new framework using simulations and theoretical synthesis, and show that to a large extent this framework resolves all problems of character construction while modeling hierarchical and hidden processes. Specifically, the new approach solves the long-standing tail color problem (Maddison 1993), and presents a previously unknown issue called the "two-scientist paradox". This issue refers to the confounding nature of the coding of a trait and the hidden processes driving the trait's evolution; failing to account for the hidden process may result in a bias, which can be avoided by using hidden state models. All this provides a clear guideline for coding traits into characters. This paper gives practical examples of using the new framework for phylogenetic inference and comparative analysis.

OVERVIEW OF DISCRETE-STATE MARKOV MODELS FOR MORPHOLOGICAL DATA

Character as a Markov model. – The *traditional Markov model (MM)* implies that character is a discrete-state and continuous-time Markov chain that moves sequentially from one state to another over the course of evolution. Markov chain is defined by a transition rate matrix containing infinitesimal rates of change between the states, and an initial vector of probabilities at the root of a phylogenetic tree [for details see e.g., Huelsenbeck et al. (2003)]. This paper treats "Markov model" and "character" interchangeably meaning that a rate matrix fully characterizes a character. The commonly used MM for a phylogenetic inference is *Mk* model (Lewis 2001).

Structured Markov Models (SMM). – This class of models (Nodelman et al. 2002), also known as continuous-time Bayesian Networks (Shelton and Ciardo 2014), arises from traditional Markov models. The only difference between the two is that SMMs are equipped with a specific parameterization of the rate matrix to model dependencies. The application of SMM to phylogenetics was pioneered by Pagel (1994) to infer correlation between two binary characters. This paper extends the application of SMM for modeling various types of correlations and hierarchical relationships between anatomical parts, which have not been previously reviewed.

Hidden Markov Models (HMM). – HMM elaborates traditional MM by splitting the model states into two layers - observable and hidden (Fig. 1a). The observable states represent the actual states of a phenotypic character, while the hidden states correspond to some unobserved factors influencing its evolution. The transitions between states are allowed only within the hidden layer; the observable states usually exhibit one-to-many mapping with the hidden states; if the mapping is one-to-one, then HMM collapses to MM. The use of HMM for the analysis of discrete traits has been recently pioneered by Beaulieu and O'Meara (2014) and Beaulieu et al. (2013).

SMM+HMM. – Equipping SMM with hidden states results in a more general class of structured Markov models that can simultaneously account for hierarchical and hidden processes. These models are the focus of the present paper.

MATERIALS AND METHODS

The tree inference was performed using the Bayesian framework in *RevBaes* (Höhna et al. 2016). The ancestral character state reconstruction was done using maximum likelihood approach in *R* (Team 2008) with package *corHMM* (Beaulieu et al. 2013); the scripts are given in the Supplementary Materials. Additionally, I provide a separate set of *R* functions (see Supplementary Materials) that construct all types of SMMs and produce input to use in *corHMM*, and *R* packages for ancestral state reconstruction.

Representation of HMM. – Although, *RevBayes* and *corrHMM* have inbuilt functions to invoke HMM, I find it convenient to construct HMM using ambiguous/polymorphic coding, which is equivalent to the inbuilt functions. For example, HMM with three observable states {*absent, blue present (2), red present (3)*}, where "absent" includes two hidden states {*absent blue (0), absent red (1)*} is represented as a three-state character {*0&1, 2, 3*}; the first state is coded as polymorphic.

Character and character state invariance. – The tree inference was run for the two datasets: (i) three two-state characters, and (ii) one four-state character. The datasets were constructed by replicating character (i) 30 times and character (ii) 10 times (the amalgamation of three characters gives one character). The replication was done to make data sufficiently informative for the inference. The dataset of character (i) was analyzed using *Mk2* model, while the dataset from the character (ii) using *Mk4* and structured *MK* model (*Mk-SMM-ind*); the analyses were run for 10^6 generations, using the exponential prior for branch lengths ~ *Exp(0.2).*

Inference for the tail problems. – The tree inference for the tail color problem and tail armor problem was performed for datasets with one, ten, and fifty characters; the characters in each dataset were identical. The different number of characters was chosen to evaluate the behavior of the models under varying amount of information.

The dataset for the tail color problem included three pairs of species, each pair had the same character pattern; the six species were chosen to assess topologies using unrooted trees (there is only one unrooted tree for three taxa). The dataset for the tail armor problem included four species. All datasets were analyzed using three models *Mk-SMM-ind* and *Mk-SMM-sw* (Supplementary Materials S6-7). The tree inference was performed using the same set of parameters as in the previous section. Model performance was evaluated using Bayes factor (BF) that compares the ratio of marginal likelihoods of two candidate models (Lavine and Schervish 1999). To interpret BF, namely threshold for which one model shows better fit than the other, I use the scale proposed by Jeffreys (1961). In this scale the BF value <3.2 indicates equal fit for two models; BF>10 suggest strong support in favor of the first model. The marginal likelihood was calculated using stepping-stone and path-sampling approaches implemented in *RevBayes.* Since both methods yielded similar values, only the former is used in the discussion.

Two-scientist paradox simulation. – One hundred trees of 2000 taxa were simulated using *phytools* package (Revell 2012). This number of taxa was used to demonstrate the asymptotic behavior of models under conditions when the amount of data is sufficient. The two-state color and shape characters were simulated as independent on the generated trees. To simulate different rates, the first rate of the shape character was drawn from a uniform distribution [0, 0.1], then this value was multiplied by varying factors to yield the second rate as well as the rates of the color character; so each character was represented by the two unique rates. To create the scoring for the first scientist, the characters were left intact. For the second scientist, the states of the simulated characters were recoded to produce a

two-state character: {*red triangular present, red triangular absent*}. Next, both HMM and MM were fitted for each character (except the first scientist's simultaneous character of shape + color since its original characters were simulated as independent given the initial conditions) using *rayDisc* function of *corHMM* package. The models were compared using ΔAIC – the difference between values of Akaike information criterion (Akaike 1974); the threshold for ΔAIC was interpreted according to Burnham and Anderson (2003).

MODELING HIERARCHICAL PROCESS USING ANATOMY ONTOLOGIES AND STRUCTURED MARKOV MODELS WITH HIDDEN STATES

Unsolved problems of encoding traits into the character

Review of the problem. – The evolving structure of organismal anatomy generates hierarchical trait dependencies that render trait encoding into character ambiguous even if trait discretization is straightforward. In practice, this means that the same trait can be encoded using two main approaches: (1) a set of binary or one multistate character, and (2) a separate state or missing entity for inapplicable observation (i.e., when inapplicable character state is coded with "–" which is interpreted as a missing entity by available methods). The superiority of one approach over the other lacks consensus in the published studies (Maddison 1993; Pleijel 1995; Hawkins et al. 1997b; Strong and Lipscomb 1999b; Forey and Kitching 2000; Brazeau 2011). Additionally, a missing consensus on what constitutes a character versus character state challenges the encoding step. One group of studies insist that the distinction between the character and character state is important (Pinna 1991; Sereno 2007; Wagner 2015), another suggests that both concepts are the same (Nelson and Platnick, 1981; Patterson, 1982). The use of different coding approaches and character/state concepts for the same trait will naturally produce different results (Brazeau 2011).

The tail color problem. – The ambiguity of trait coding is best exemplified by the tail color problem (Maddison 1993; Hawkins et al. 1997b) that seeks optimal scheme for scoring traits in species with no tails, blue tails and red tails. This problem exhibits a common case of coding anatomically dependent traits, and a solution to it can be extrapolated to all other similar cases.

Four main schemes have been traditionally proposed to encode the tail traits (Fig. 2). The *first scheme* (Fig. 2a) uses two characters: (i) tail presence, and (ii) tail color; the absence of the tail in the second character is coded as the inapplicable observation. The *second scheme* (Fig. 2b) is similar to the previous one but encodes tail absence as a separate state. The *third scheme* (Fig. 2c) employs three binary characters: (i) tail present, (ii) blue tail present, and (iii) red tail present. Finally, the *fourth scheme* (Fig. 2d) uses one three-state character.

The tale color problem has been widely discussed in parsimony literature but has never been evaluated in the "model-based" framework. In parsimony, the suggested approach is to use the scheme #1 with inapplicable coding that, nevertheless, was shown to suffer from pseudo-parsimonious optimization and spurious results (Maddison 1993; Hawkins et al. 1997a; Strong and Lipscomb 1999b; Brazeau 2011).

Below, I demonstrate that unlike parsimony, "model-based" framework offers two unambiguous solutions to the tail color problem: one is based on using the scheme #1, and another, which is new, uses SMM+HMM. I derive the solutions through showing (1) the equivalence of character and character state concepts, and (2) SMM techniques for modeling various types of anatomical dependencies. Finally, I will demonstrate how the proposed solutions can be used in phylogenetic inference with morphological data and discuss practical issues of trait modeling.

Invariance: character and character states are the same

Amalgamation of characters. – The properties of SMM (Shelton and Ciardo 2014) allow mathematically valid amalgamation of any number of characters thus removing distinction between character and character state. The

amalgamation is especially straightforward when characters are independent (the dependent cases are reviewed in the next section). Suppose, there are two characters: $\boldsymbol{T}$ - tail presence {with states: *absent (a), present (p)*}, and $\boldsymbol{C}$ - tail color {*red (r), blue (b)*}, which as we assume (for now) evolve independently according to the rate matrices:

$$\boldsymbol{T} = \begin{array}{c} \\ \boldsymbol{a} \\ \boldsymbol{p} \end{array} \begin{array}{c} \begin{array}{cc} \boldsymbol{a} & \boldsymbol{p} \end{array} \\ \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \end{array}, \boldsymbol{C} = \begin{array}{c} \\ \boldsymbol{r} \\ \boldsymbol{b} \end{array} \begin{array}{c} \begin{array}{cc} \boldsymbol{r} & \boldsymbol{b} \end{array} \\ \begin{pmatrix} -\gamma & \gamma \\ \theta & -\theta \end{pmatrix} \end{array}. \quad (1)$$

The rate matrix of the amalgamated character can be constructed via the following equation:

$$A_{mal.(ind)}(\boldsymbol{T},\boldsymbol{C}) = \boldsymbol{T} \otimes \boldsymbol{I_C} + \boldsymbol{I_T} \otimes \boldsymbol{C} = \begin{array}{c} \\ \boldsymbol{ar} \\ \boldsymbol{ab} \\ \boldsymbol{pr} \\ \boldsymbol{pb} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{ar} & \boldsymbol{ab} & \boldsymbol{pr} & \boldsymbol{pb} \end{array} \\ \begin{pmatrix} -\alpha-\gamma & \gamma & \alpha & 0 \\ \theta & -\alpha-\theta & 0 & \alpha \\ \beta & 0 & -\beta-\gamma & \gamma \\ 0 & \beta & \theta & -\beta-\theta \end{pmatrix} \end{array}, \quad (2)$$

where $\boldsymbol{I_T}$ and $\boldsymbol{I_C}$ are the identity matrices, and $\otimes$ denotes the Kronecker product; hereafter I refer to this model as SMM-ind (for derivation see Supplementary Materials S1). The four-state space of the amalgamated character is the state product (i.e. Cartesian product) of the initial characters that exhibits all combinations of $\boldsymbol{T}$ and $\boldsymbol{C}$ states, which are {*ar, ab, pr, pb*}. The zero elements in the matrix (2) penalize the initial states of $\boldsymbol{T}$ and $\boldsymbol{C}$ to change simultaneously over infinitesimal interval of time since simultaneous change would imply a correlation between the characters (see the section Synchronous state change). Also, these elements introduce a notion of state accessibility in phenotype that indicates which states are immediately accessible from present state through one-state change (Stadler et al. 2001). To combine *n* independently evolving characters the equation (2) has to be successively repeated *n-1* times. For example, in the case of the three initial characters $\boldsymbol{T}$, $\boldsymbol{C}$, and $\boldsymbol{Z}$ this means:

$$A_{mal.(ind)}(\boldsymbol{T},\boldsymbol{C},\boldsymbol{Z}) = (\boldsymbol{T} \otimes \boldsymbol{I_C} + \boldsymbol{I_T} \otimes \boldsymbol{C}) \otimes \boldsymbol{I_Z} + \boldsymbol{I_{TC}} \otimes \boldsymbol{Z}. \quad (3)$$

The matrices of amalgamated characters have peculiar symmetries - although matrix dimension grows fast, the vast majority of cells are zeros (Fig. 3c); the transition rates are located along the secondary diagonals. For a chain of *n* coevolving characters with the equal number of states ω the proportion of non-zero elements in the matrix is $[1 + n(\omega - 1)]/\omega^n$. The number and pattern of secondary diagonals populated with transition rates increases with the number of states. If initial characters have equal number of states, the total number of secondary diagonals is $n(\omega - 1)$.

Aggregation of states. – The opposite to amalgamation is the aggregation of states that allows decomposing one character into a set of several characters. The aggregation of states is mathematically valid only if Markov model is lumpable, which requires specific symmetries of rate matrix (see the section Strong lumpability) but always holds for independently evolving characters. The amalgamated character in the equation (2) can be decomposed with respect to the two partitions of its state space: $\mathcal{B}_1$={{*ar, ab*}, {*pr, pb*}} and $\mathcal{B}_2$={{*ar, pr*}, {*ab, pb*}}, that gives transition matrices for the characters $\boldsymbol{T}$ and $\boldsymbol{C}$ respectively. This operation can be expressed using the following equation:

$$A_{ggr.}(A_{mal.}(\boldsymbol{T},\boldsymbol{C})|\mathcal{B}) = \mathbf{NP}A_{mal.}(\boldsymbol{T},\boldsymbol{C})\mathbf{P^T}, \quad (4)$$

the meaning of the equation terms is provided in the section Strong lumpability.

The amalgamation technique indicates that separate characters can be combined into one single character; in turn, the aggregation technique allows the states of the same character to be represented as separate characters. Thus, regardless the initial way of discretizing organismal features into characters and states, the character and states remain conceptually equal and are thereby invariant with respect to each other if rate matrices are appropriately structured. The toy simulation in Figure 3 exemplifies this theoretical consideration: the tree topology inferred using *Mk* model for the three two-state characters is the same as the topology when those three characters are amalgamated into one four-state character (Fig. 3a) and analyzed using structured *Mk* model (i.e., SMM-ind) from the equation (3). If the four-state character is analyzed with traditional *Mk* model that is not properly structured, then the tree is unresolved since such amalgamation is invalid mathematically (Fig. 3b).

Modeling character dependencies

To the best of my knowledge, different coding schemes and trait dependencies generate three main types of correlation between characters: (i) general type of correlation, (ii) "switch-on" dependency, and (iii) synchronous state change. The techniques of modeling these correlations for the two-character case are discussed below; they can be extrapolated for an arbitrary number of characters using the equation (3).

General type of correlation. – Independent evolution of characters $\boldsymbol{T}$ and $\boldsymbol{C}$ is defined by the rate matrix with the specific rate symmetries and the maximum of four free rate parameters [equation (2)]. Those symmetries can be relaxed by making all matrix rates different (Pagel 1994), which produces a complex pattern of correlation between the states (see Supplementary Materials S2b). Such matrix corresponds to the general type of correlation between the characters $\boldsymbol{T}$ and $\boldsymbol{C}$ (hereafter SMM-gen) that is:

$$A_{mal.(gen)}(\boldsymbol{T},\boldsymbol{C}) = \begin{pmatrix} -\alpha-\gamma_1 & \gamma_1 & \alpha & 0 \\ \theta_1 & -\alpha_1-\theta_1 & 0 & \alpha_1 \\ \beta & 0 & -\beta-\gamma & \gamma \\ 0 & \beta_1 & \theta & -\theta-\beta_1 \end{pmatrix}. \quad (5)$$

"Switch-on" type of dependency. – This type arises due to anatomical dependencies between traits when hierarchically upstream trait switches-on and -off the downstream one. Consider two previous characters: $\boldsymbol{T}$ - tail presence $\{a, p\}$, and $\boldsymbol{C}$ - tail color $\{r, b\}$. Apparently, the tail color depends on the tail presence – both states of the character $\boldsymbol{C}$ are observable if and only if the character $\boldsymbol{T}$ is in the state present; if the tail is absent, then the color character is "switched-off" and does not evolve. One of the ways to model such dependency is to amalgamate $\boldsymbol{T}$ and $\boldsymbol{C}$ as independently evolving using SMM-ind. However, one may want to impose stricter relationships that reflect the anatomical hierarchy of the traits by prohibiting color changes when the tail is absent. It can be done using the modified version of the equation (2) that gives the following amalgamated rate matrix (hereafter SMM-sw):

$$A_{mal.(sw)}(\boldsymbol{T},\boldsymbol{C}) = \boldsymbol{T}\otimes\boldsymbol{I_C} + \boldsymbol{D_T}\otimes\boldsymbol{C} = \begin{array}{c} \\ \boldsymbol{ar} \\ \boldsymbol{ab} \\ \boldsymbol{pr} \\ \boldsymbol{pb} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{ar} & \boldsymbol{ab} & \boldsymbol{pr} & \boldsymbol{pb} \end{array} \\ \begin{pmatrix} -\alpha & 0 & \alpha & 0 \\ 0 & -\alpha & 0 & \alpha \\ \beta & 0 & -\beta-\gamma & \gamma \\ 0 & \beta & \theta & -\beta-\theta \end{pmatrix} \end{array}, \quad (6)$$

where $\boldsymbol{D_T}$ is a diagonal matrix (Supplementary Materials, S2c). The difference between this matrix and that of SMM-ind is that here the transitions $ar \rightarrow ab$ and $ab \rightarrow ar$ are set to zero.

Synchronous state change. – This type frequently occurs when traits are redundantly coded using binary (absent/present) approach. Suppose, there are two characters: (1) $\boldsymbol{R}$ – tail color red {*red absent* (a_r), *red present* (p_r)}, and (2) $\boldsymbol{B}$ - tail color blue {*blue absent* (a_b), *blue present* (p_b)}, defined by:

$$\boldsymbol{R} = \begin{array}{c} \\ \boldsymbol{a_r} \\ \boldsymbol{p_r} \end{array} \begin{array}{c} \begin{array}{cc} \boldsymbol{a_r} & \boldsymbol{p_r} \end{array} \\ \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \end{array}, \quad \boldsymbol{B} = \begin{array}{c} \\ \boldsymbol{a_b} \\ \boldsymbol{p_b} \end{array} \begin{array}{c} \begin{array}{cc} \boldsymbol{a_b} & \boldsymbol{p_b} \end{array} \\ \begin{pmatrix} -\beta & \beta \\ \alpha & -\alpha \end{pmatrix} \end{array}. \quad (7)$$

Obviously, at certain observation event, the tail can have only one color - either blue or red. This implies that if red is observed than blue is absent and vice versa. So, the states between $\boldsymbol{R}$ and $\boldsymbol{B}$ are mutually exclusive and hence change simultaneously over the course of evolution. In contrast to the previous amalgamation techniques, the synchronous evolution must allow only two-step transitions except for the transitions $p_r p_b \rightarrow a_r a_b$ and $a_r a_b \rightarrow p_r p_b$, which are biologically impossible. This yields the following amalgamated rate matrix (hereafter SMM-syn):

$$A_{mal.(syn)}(\boldsymbol{R},\boldsymbol{B}) = \begin{array}{c} \\ \boldsymbol{a_r a_b} \\ \boldsymbol{a_r p_b} \\ \boldsymbol{p_r a_b} \\ \boldsymbol{p_r p_b} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{a_r a_b} & \boldsymbol{a_r p_b} & \boldsymbol{p_r a_b} & \boldsymbol{p_r p_b} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\alpha & \alpha & 0 \\ 0 & \beta & -\beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}. \quad (8)$$

The two states $p_r p_b$ and $a_r a_b$ can be removed from this matrix as they are never visited by the Markov chain. Their remove renders the matrix (8) to be absolutely equal to the two-state matrix in the equation (1) that defines the tail color.

Solutions to the tail color problem

Assessment of the coding schemes. – As shown above, the tail traits can be modeled using two main approaches: (i) amalgamating the two characters as independently evolving (SMM-ind), or (ii) amalgamating the two characters through the "switch-on" dependency (SMM-sw). Let us assess these two techniques against the alternative coding schemes (Fig. 2). To make the schemes comparable, the variable number of characters between them has to be eliminated by amalgamating the characters of each scheme as SMM-ind into one single character; this renders the amalgamated character states to be the products of the initial characters (Supplementary Materials S3). The amalgamation as SMM-ind is equivalent to treating each initial character independently but is more convenient for the comparison.

Amongst all alternative schemes only the scheme #1 {amalgamated states: *ar, ab, pr, pb*} that uses inapplicable coding is biologically meaningful - it precisely exhibits SMM-ind that models the tail traits as independent initial characters. This solution has been known and widely applied in phylogenetics before. In parsimony, this solution was shown to suffer from pseudo-parsimonious optimization that results in spurious groupings of species and arises when distantly related taxa influence each other's parsimony scores (Maddison 1993). In the "model-based" methods, this problem should not be the case since likelihood calculation at distantly related parts of the tree is independent (Yang 2006). The rate matrix for the scheme #1 can be further elaborated using the "switch-on" dependency to yield the second solution via SMM-sw that has not been previously reported.

The scheme #4 (single three-state character), having at first glance similar state space, is quite different from the scheme #1. For comparison, let us expand the state space of the scheme #4 to make it identical to that of the scheme #1. This implies creating a four-state rate matrix that if lumped collapses to the original three-state one (Supplementary Materials S4a). In contrast to the scheme #1, the expanded matrix of the scheme #4 has non-zero rates for the two-step transitions, thus failing to account for the hierarchical dependencies.

The amalgamated state space of the schemes #2 and #3 differs significantly from the rest by having spurious states. For example, the state of scheme #2 (*tail present, no color*) and the state of scheme # 3 (*tail present, tail blue, tail red*) are logically contradictory since the tail is colored once it appears, and tail cannot be blue and red at the same time. During phylogenetic inference, these states will be reconstructed as ancestral ones which will make the inference biologically misleading as was previously noted for parsimony (Forey and Kitching 2000). Thus, the schemes #2 and #3 should not be used for coding.

Character coding invariance. – As shown above, SMM-ind and SMM-sw represent two solutions for the tail color problem, which can be coded with the scheme #1. The schemes #2, #3 and #4 fail to meaningfully accommodate hierarchical relationships because they contain synchronous changes, redundant states or improperly structured rate matrices (see the dependency graphs in Figure 2) that cannot be modeled directly through SMM-ind or SMM-sw. In fact, the amalgamation of the schemes #2, #3 and #4 has to be derived differently by considering their specific dependencies and using SMM-syn. Interestingly, this derivation makes the schemes #2, #3 and #4 to be identical to the scheme #1 (Supplementary Materials S4). Thus, the alternative coding approaches become invariant (i.e., equal) with respect to each other if anatomical dependencies between traits are appropriately incorporated using SMM.

Modeling SMM using HMM. – The amalgamation via SMM-ind and SMM-sw produces matrices with four states {*ar, ab, pr, pb*}. The states *ar* and *ab* correspond to the same observation specifying the absence of the tail since tail color cannot be observed when the tail is absent. Noteworthy, these matrices, as might be expected, cannot be reduced to a three-state matrix with states {*a, r, b*} under any values of their rate parameters because they are not lumpable under the partitioning scheme {{*ar, ab*}, {*pr*}, {*pb*}} (Supplementary Materials S5). Thus, modeling SMM-ind and SMM-sw requires using HMM that includes three observable states {*a, r, b*}, and the observable state a consists of two hidden states *ar* and *ab*. Alternatively, SMM-ind can be modeled using two separate characters with inapplicable coding (i.e., the scheme #1) because SMM-ind is lumpable with respect to the two initial characters [see the equation (4)]. This does not hold for SMM-sw whose rate matrix cannot be lumped with respect to the tail color character (Supplementary Materials S5); thus, SMM-sw can be modeled only using the hidden states.

I suggest that all dependent traits have to be scored as one multistate character despite that SMM-ind can be represented using two characters. First of all, one-character representation will help to avoid data recoding if both

models - SMM-ind and SMM-sw - are tested for the same dataset. Second, if SMM-ind is extended to have more free parameters as, for example, SMM-gen or a model similar to F81 (Felsenstein 1981), then its two-character representation will no longer be valid, and this extension will require only one multistate character.

Tree inference with morphology. – Both SMM-ind and SMM-sw can be straightforwardly adopted for phylogenetic inference by converting them into *Mk*-type models (Lewis 2001). This conversion is done by constraining their rate matrices to include one free parameter (that is usually interpreted as a branch length), and setting their initial vector to contain equilibrium distribution at the root. Hereafter, I refer to these models as *Mk-SMM-ind* and *Mk-SMM-sw* respectively (Supplementary Materials S6).

Despite having identical observable state space, *Mk-SMM-ind* and *Mk-SMM-sw* imply different assumption for trait evolution. In analyses, these models can be compared using statistical methods for model selection (e.g., Akaike information criterion, Bayes factor, etc.). The use of model selection methods becomes possible due to the identity of their observable state spaces that keep the data the same.

Ancestral reconstruction on a known tree. – The provided considerations for hierarchically dependent traits hold for both phylogenetic inference and ancestral character state reconstruction. Using inapplicable coding (i.e., the scheme #1), that exhibits one of the appropriate ways for coding hierarchical trait, has been common in tree inference. This is in contrast to ancestral character state reconstruction on a known tree where the use of one character with multiple states would be a common choice (i.e., the scheme #4). As was shown earlier, this is a misleading approach since SMM-ind and SMM-sw cannot be substituted by a three-state rate matrix. Thus, I suggest using hidden state models (i.e., SMM-ind and SMM-sw) for ancestral reconstruction whenever anatomical dependence in traits are detected. In ancestral reconstruction, the rate constraints in SMM-ind and SMM-sw can be relaxed by, for example, making all rates different as those in SMM-gen.

Demonstrative analyses

For the demonstrative purpose, I perform a phylogenetic inference for the tail color problem using three toy datasets that contain three pairs of identical species and a varying number of identical characters (Fig. 4d-f). For each dataset, I assess the performance of the two models: *Mk-SMM-ind* and *Mk-SMM-sw* (Supplementary Materials S6, Figure 4). Bayes factor (BF) and topologies of unrooted trees are used for model comparison (see also, Materials and Methods)

To provide a better insight into modeling hierarchical dependencies, I also use a modified version of the tail color problem to which I refer to as the “tail armor case” (Fig. 4a-c). This case considers four species that possess a tail with blue or red armor, tail without armor, and no tail; thus, it implies a two-level hierarchical dependency – armor color depends on armor presence that, in turn, depends on tail presence. The same set of alternative models *Mk-SMM-ind* and *Mk-SMM-sw* is applied to analyze it (Supplementary Materials S7); these models include four observable and eight hidden states (Fig. 4).

Discussion of the analyses. – The expected topology for the tail color problem must group together the pairs of species with the identical characters (Fig. 4d). All two models yield the majority consensus tree to be the same with the expected topology when datasets include more than one character (Fig. 4f). The inference with one character does not differentiate between the candidate models (all BF values <3.2); however, as the character number increases to fifty, *Mk-SMM-ind* yields a moderately better fit than *Mk-SMM-sw* [BF(*Mk-SMM-ind*, *Mk-SMM-sw*) = 5.4]. In the armor color case, the intuitive tree is expected to group the species with armor in one clade, and those without armor in another thus reflecting the putative evolutionary sequence of tail and armor emergence (Fig. 4a). All analyses with one and ten characters yield unresolved majority consensus, and similar fit for the models (Figure 4c). In the analyses with fifty characters, the two models produce the expected consensus but *Mk-SMM-ind* reveals a significantly better fit [BF(*Mk-SMM-ind*, *Mk-SMM-sw*) = 25.6].

The analyzed toy datasets demonstrate that dependent traits can be efficiently modeled by SMMs with hidden states, and the models can be differentiated using model selection criteria. The better fit of *Mk-SMM-ind*, in both sets

of the simulations, likely occurs due to the simplicity of the used datasets (*Mk-SMM-sw* describes more complex relationships). Both models reveal the expected tree when the amount of data is sufficient; however, the topological performance of the two models should not be expected to be the same in complex datasets. Thus, similar to DNA data, model selection for morphology-based inference is important.

Suggestions for trait coding

To sum up, structured Markov models equipped with hidden sates provide a natural way to model anatomical dependencies between traits for tree inference and ancestral character state reconstruction. These models give unambiguous solutions to the tail color problem and can be flexibly extended to model complex hierarchies as was shown using the tail armor case. The provided solutions via SMM-ind and SMM-sw can be further generalized using SMM-gen to account for various types of relationships which may exist between traits. The efficient use of SMM+HMM requires an identification of anatomical dependencies and their incorporation through the appropriate structuring of rate matrix. This means that trait coding and model selection are the parts of the same analytical procedure as their choice is affected by organismal anatomy. I suggest that hierarchically dependent traits have to be coded as one character that can be directly linked to an appropriate rate matrix. For phylogenetic inference, all characters of a dataset, which encode dependencies, can be placed in one data partition to which a selected model is assigned. This strategy will allow testing different candidate models using the same data.

The artificial recoding of characters (e.g., when a multistate character is recoded to a set of binary) should not be used - it introduces spurious model assumptions that are inconsistent with the properties of character amalgamation and state aggregation. At the same time, the invariance of character and character state can be a helpful tool as it, to a certain extent, eliminates the ambiguity of discretizing traits into character and character states. Trait scoring and model choice must be done thoughtfully and must be consistent with the underlying mathematics, organismal anatomies, and biological knowledge.

Ontology-informed structured Markov models

The hierarchical dependencies between traits are imposed by the structure of organismal anatomy and hence can be retrieved using anatomy ontologies; this is turning ontologies into a promising tool for arranging, querying and managing anatomical data (Deans et al. 2015). The need for integrating ontologies with morphology-based phylogenetic analysis has been recently emphasized and discussed (Vogt 2017a, 2017b, 2018). Unfortunately, by far, anatomy ontologies are available only for several taxa [e.g., Mungall et al. (2012) and Yoder et al. (2010)]. Nevertheless, the available software allows direct linking of character matrices with anatomy ontologies (Balhoff et al. 2010). This provides an opportunity for an automatic extraction of anatomical dependencies from ontologies linked to characters (Dececchi et al. 2016). Alternatively, the incorporation of the dependencies can be done by simply using a scientist's own knowledge of organismal anatomy. Integration of SMM with ontologies opens possibilities for reconstructing ancestral anatomy ontologies in a way similar to the parsimony-based method proposed by Ramirez and Michalik (2014). In this light, a character becomes not merely a vector of numbers (i.e., states) but a graph reflecting various relationships, each with its own anatomical meaning. Such graph, produced by linking the tail character with UBERON ontology (Mungall et al. 2012), is shown in Figure 5. Structured Markov models can be used to infer topology of such graphs as well. It can be anticipated that further developments in this field will lead to a creation of an automated pipeline for constructing ontology- informed SMM+HMM models. This pipeline will open up new perspectives for modeling evolution of entire organismal anatomies.

MODELING HIDDEN PROCESS DRIVING MORPHOLOGICAL EVOLUTION

Overview of correspondence between GRNs and morphology

The genotype-to-phenotype map is a function that describes complex interactions between genetic, cellular, and environmental variables resulting in phenotype (Cowperthwaite and Meyers 2007). Gene regulatory networks (GRNs) – sets of interacting genes – are one of the key components in the genotype-to-phenotype map that drive trait changes (Pigliucci 2010), especially at a macroevolutionary scale (Tsong et al. 2006). Thus, GRNs can be reasonably considered to be the elementary genetic units whose realizations, over spatiotemporal scales of embryo development, produce phenotypic traits. In this respect, it is essential to summarize general mechanisms of GRN evolution, and properties of GRN-to-phenotype (GRP) maps. Note that general genotype-to-phenotype maps have been intensively studied from the theoretical perspectives (Stadler et al. 2001; Ahnert 2017; Fortuna et al. 2017). Application of those interesting findings is currently challenging since no studies show how they can be translated into phylogenetic models. Thus, here, I work with a simple representation of GRP maps, that can be straightforwardly formalized using Markov models.

Organismal GRN consists of nested modules, each module is a smaller GRN exhibiting a relatively autonomous cluster of interacting genes (Oakley; Babu et al. 2004; Longabaugh et al. 2005; Kuratani 2009; McDougall et al. 2011; Siegal 2013; Rebeiz et al. 2015; Voordeckers et al. 2015). The multifaceted evolution of GRN modules is driven by three general mechanisms. *(1) Module birth* refers to a birth of new GRN module that occurs by either co-option of a pre-existing module into a new body place (Babu et al. 2004; Wagner 2007; Erwin and Davidson 2009; Monteiro 2012; Siegal 2013; Hinman and Cheatle Jarvela 2014; McKeown et al. 2014; Glassford et al. 2015; Rebeiz et al. 2015) or by integration of several pre-existing modules into a new module (Clark-Hachtel et al. 2013; Arendt et al. 2016). *(2) Module transformation* refers to a transformation of a pre-existing module from one state to another that is caused by reorganizing regulatory linkage between genes (Abouheif 1999; Erkenbrack and Davidson 2015). *(3) Module death* refers to an inactivation of GRN module by a mutation in the upstream regulatory module that disables its realization (Shapiro et al. 2006; Shbailat and Abouheif 2013; Held 2014). The limits between those mechanisms are somewhat arbitrary and, at the global scale, GRN evolution occurs under a mixture of the interacting processes.

Evolution of GRNs can be conveniently formalized using Markov models - each step in the evolutionary path of a GRN module is a state of a Markov chain. The complete evolutionary path of the module involves: module birth, the transition between module states, and module death. In this respect, GRP map can be viewed as a mapping of the states of GRN Markov model to those of a discrete character. This mapping falls into three main types (Fig. 6).

Type 1: one-to-one. – One-to-one mapping between states of GRN modules and those of a discrete character indicates that primary homology hypothesis correctly identifies underlying genetic space (Fig. 6a). This case is ideal but rarely realistic since morphological traits are usually the products of complex gene interactions.

Type 2: many-to-one. – Many-to-one mapping is a scenario when several GRN states are mapped onto one trait state (Fig. 6b) indicating that the trait is controlled by multiple genetic factors which are usually unknown to a researcher (Rebeiz et al. 2015). Numerous evo-devo studies show that this scenario is a commonplace (Abouheif 1999; Hall 2003; Moczek 2008; McCune and Schimenti 2012; Wagner 2015). For instance, some males of *Drosophila* have a pigmented spot, located on the wing, that they use in a courtship display (Prud'Homme et al. 2006). This spot is controlled by one gene *yellow* and has similar shapes across different species. For a phylogenetic analysis, it would be reasonable to code this trait using a two-state character {*spot present, spot absent*}, which was done in the study of Prud'Homme et al. (2006). That study revealed multiple gains of the spot during the evolution of *Drosophila*. Interestingly, the independent gains were caused by a different mechanism - the co-option of different *cis*-regulatory elements associated with *yellow*. Evidently, even a seemingly simple trait such as the wing spot may, in fact, have a complex GRN state space.

Another situation when many-to-one correspondence can be a commonplace refers to the incompleteness of morphological examination that may arise when external structures are examined without reference to skeletal

structures (in vertebrates), or when external skeletal structures are examined without reference to underlying muscles (in invertebrates). For example, different lineages of salamanders have similarly elongated body shapes which may be hypothesized to be homologous assuming that the phylogeny is unknown. However, the mechanism of body elongation varies across the lineages by adding and extending different individual vertebrae (Wake et al. 2011). If body shape is studied without reference to the skeleton, then similarly elongated bodies would be scored with the same character state. However, this approach is misleading since the different mechanism of elongation may produce the same shapes, which reflects many-to-one mapping between the underlying state space and observation. Thus, in the model formalism, this and the *Drosophila* case are the same.

Type 3: partial mapping. – The partial mapping occurs when one trait is controlled by interdependent GRN modules but this interdependence in not reflected in the coding meaning that the GRN space producing the trait is only partially identified (Fig. 6c). For example, the aforementioned tail color problem can be roughly formulated in the GRN terms using the two modules: one controlling tail development, and another, nested module, controlling color; considering the color character as a separate entity (i.e., using the coding schemes #2-#4) results in this type of mapping because tail color and tail presence are the parts of the same developmental process; thus, they have to be treated as one character. The partial mapping may be a consequence of pleiotropy and epistasis when seemingly unrelated traits are correlated due to the shared genetic factors. Moreover, in the model formalism, partial mapping corresponds to any type of correlation that is missed due to a spurious trait discretization or model choice. For example, traits of insect mouthparts can undergo simultaneous evolutionary changes when species get adapted to new feeding conditions. In this situation, treating different elements (e.g., mandibles, labrum, etc.) of the mouthparts as separate characters is misleading; the dependencies have to be incorporated into the model to reflect the correlated evolution. The diversity of various types of partial mapping occurs because, in phylogenetics, they all match the same general model - SMM-gen. In some cases, the existence of correlation is obvious and can be retrieved from the structure of anatomy; however, cases with unobservable correlation - when the correlation is unknown *a priori* and can be detected only be means of a statistical analysis - seem to be widespread.

Modeling GRN-to-phenotype mapping

As shown above, the GRP map is characterized (except one-to-one mapping) by a significantly larger GRN state space than that of a morphological trait. In terms of Markov model, this means that the construction of a discrete character is a substitution of the original GRN Markov model with a large number of states (Fig. 1b) by a morphological character model with the reduced number of states (Fig. 1c). In this reduced model, at least, some states consist of an aggregated set of the original GRN states. A natural question which can be raised is - when such aggregation is mathematically valid for modeling observable character without referencing to the complex underlying processes? Below, I show that the aggregation of states is tightly connected to the property of Markov chain lumpability meaning that the aggregated states can be unbiasedly modeled if and only if the generating GRN process is lumpable. If the process is not lumpable then modeling traits using traditional MM may be biased. The bias can be avoided using HMM (Beaulieu et al. 2013; Beaulieu and O'Meara 2014) that is capable to explicitly model GRN states.

When traditional models are not biased: lumpable Markov models

If aggregation of states in an original Markov chain produces an aggregated chain that is still Markovian then the original chain is called lumpable (Kemeny and Snell 1960; Rubino and Sericola 2014). The lumpability guarantees that transition rates and state sequence can be unbiasedly modeled using the aggregated chain regardless the complexity of the original state space. There are three relevant types of lumpability which are discussed below.

Strong lumpability. – The aggregation of states is a partitioning of a transition rate matrix into partition blocks that correspond to the transition rates in the aggregated chain (Fig. 7a). The strong lumpability implies that the original chain can be aggregated under any possible values of initial vector. The sufficient and necessary condition (hereafter, the row-wise sum rule, Figure 7a) for Markov chain to be strongly lumpable in respect to a given partitioning scheme is that the row-wise sum of rates within one partition block of rate matrix must be the same for all rows within the given partition block, and this property must hold for all blocks in the rate matrix (Kemeny and Snell 1960; Rubino and Sericola 2014). Suppose, we want to lump an original rate matrix $\boldsymbol{M}$ using a partitioning scheme $\mathcal{B}$ into the aggregated matrix $A_{ggr.}(\boldsymbol{M}|\mathcal{B})$. If the model is lumpable the row-wise sum rule implies the following equality to hold (Kemeny and Snell 1960):

$$\mathbf{P}^{\mathbf{T}}\mathbf{N}\mathbf{P}\boldsymbol{M}\mathbf{P}^{\mathbf{T}} = \boldsymbol{M}\mathbf{P}^{\mathbf{T}},$$

where **N** and **P** are the matrices specifying state assignment between the initial and aggregated models (Supplementary Materials S2a). So, the rates in the aggregated matrix $A_{ggr.}(\boldsymbol{M}|\mathcal{B})$ can be expressed via the equation:

$$A_{ggr.}(\boldsymbol{M}|\mathcal{B}) = \mathbf{N}\mathbf{P}\boldsymbol{M}\mathbf{P}^{\mathbf{T}}.$$

For example, consider the four-state rate matrix $A_{mal.(ind)}(\boldsymbol{T},\boldsymbol{C})$ from the equation (2), its states can be aggregated using the partitioning schemes $\mathcal{B}_1$={{ar, ab}, {pr, pb}} that results in the character $\boldsymbol{T}$ from the equation (1). The aggregation is possible because the row-wise sum rule is satisfied that implies the following equalities:

$$(ar \to pr) + (ar \to pb) = (ab \to pr) + (ab \to pb) = \alpha$$
$$(pr \to ar) + (pr \to ab) = (pb \to ar) + (pb \to ab) = \beta$$

Here, the rates α and β on RHS form exactly the rates of the aggregated character $\boldsymbol{T}$. The partitioning scheme $\mathcal{B}_2$={{ar, pr}, {ab, pb}} also produces a lumpable model resulting in the character $\boldsymbol{C}$ from the equation (1). In contrast to that, SMM-gen from the equation (5) cannot be lumped since its rate matrix, with all rates different, violates the row-wise sum rule under any partitioning scheme. The performance of lumpability in simulations is discussed in the next section.

Weak lumpability. – The weak lumpability allows lumping an original chain only under particular values of the initial vector (Kemeny and Snell 1960; Rubino and Sericola 2014), which imposes stricter dependencies between the initial vector and rate matrix. Generally, these dependencies cannot be found analytically but a finite algorithm can be used to elucidate them (Rubino and Sericola 1993, 2014). An interesting case of weak lumpability arises when the initial vector of a Markov chain contains equilibrium distribution meaning that the probability of observing chain's states after some time has the same equilibrium distribution. Such chain is weakly lumpable with respect to any possible partitioning scheme (Supplementary Materials S8). Although Markov models that have the initial vector with equilibrium distribution are commonly used in phylogenetics, the weak lumpability should not be expected to hold due to the way Felsenstein's pruning algorithm (Felsenstein 1981) calculates likelihood (Supplementary Materials, section S8). Thus, weak lumpability is omitted from the further discussion in this paper.

Nearly lumpable chains. – This type of lumpability occurs in a large multistate character where the character states are correlated, thereby allowing to lump multistate character with the insignificant error. Such character can be constructed by amalgamating many elementary characters into one large rate matrix using SMM-gen (see the section General type of correlation). Aggregation of states in such matrix can be thought of as mapping evolutionary processes occurring at the level of DNA sites or numerous GRNs to its realization at the phenotypic level. For example, consider a DNA locus of 1000 sites with each site being a four-state character (four nucleotides). All sites can be amalgamated into one large character with 4^{1000} states; this character possess peculiar symmetries of the rate matrix(e.g., see Fig. 3c) in which over 99% cells are set to zero (see the section Amalgamation of characters). Suppose that those 4^{1000} states of the DNA character are mapped only to a few states of a phenotypic characters. Since the original state space is significantly larger, each state of the phenotypic character is an aggregation of many original DNA states.

In a trivial case of equal evolutionary rates across all sites, the strong lumpability condition can be satisfied for numerous partitioning schemes that can be applied to amalgamate the DNA character. In contrast to it, if the amalgamation is constructed using SMM-gen with all rates different to reflect the most complex scenario of correlated evolution, then, obviously, such character drastically violates the condition of strong lumpability. However, if rates in the rate matrix and values of the initial vector are identically and independently distributed (*i.i.d.*), which does not rule out for them the possibility to be different, then the amalgamated DNA character is nearly lumpable under any possible

partitioning scheme. This occurs because the *i.i.d.* condition produces an aggregated chain whose error, in approximating the original rates, is insignificant as the number of the original states increases, while the number of the aggregated states decreases (Supplementary Materials S6).

When traditional models are biased: hidden Markov models

The aforementioned conditions for strong lumpability and nearly lumpable chains are likely to be violated in real-life situations due to the rather strict constraints imposed on rate matrix and initial vector; one should not expect that evolution will specifically follow these conditions. The violation can be caused by the complexity of GRN space that will make the aggregation over the GRN states mathematically invalid; so, the aggregated chain will lose the Markovian property that will make the inference of a morphological character misleading using traditional MM.

Hidden Markov models are a perfect tool to avoid the aggregation - their structure, consisting of the unobservable and observable layers, can be flexibly adopted to account for any type of mapping between GRNs and traits (Fig. 1a). In this regard, hidden and observable states of HMM can be thought of as states of GRN modules, and states of a trait respectively. The notable case of lumpability violation is given below, to which I refer to as the "two-scientist paradox".

The two-scientist paradox. – Suppose, two independently evolving genes produce organisms with two traits: (i) red or blue color, and (ii) triangular or round shapes (Fig. 8a-b). Two scientists are willing to reconstruct ancestral states of those traits on a known phylogeny but are unfamiliar with the gene-to-trait mapping and each other's work. The first scientist codes the traits using three different characters: (1) color, (2) shape, and (3) color and shape simultaneously (Fig. 8c). The second scientist scores only one binary character - presence or absence of being triangular and red simultaneously - because she or he considers this trait to be a "key innovation" that is worth special attention with respect to all other traits (Fig. 8d). Even though the scientists have different opinions of how to code the traits, they test similar models – traditional MM and HMM - to identify the best one. To demonstrate the consequences of the scientists' approaches, I generated 100 random trees with characters corresponding to the respective scoring used; the fit of the models was compared using ΔAIC, calculated as $AIC_{(MM)}$-$AIC_{(HMM)}$ (see Materials and Methods).

The first scientist finds that traditional MMs have a significantly better fit for all characters, which is expected given the initial conditions for how the traits evolve (Fig. 8e-g). In contrast to that, the second scientist finds that HMM is favored – it yields a substantially better fit than MM (ΔAIC >10) in 80% of the simulations (Fig. 8h-j). At first glance, this result may seem odd since characters are independent; however the underperformance of MM occurs because the coding scheme of the second scientist corresponds to the partial mapping between the genes and trait meaning that the generating process is incompletely identified with MM; the amalgamated rate matrix that defines simultaneous evolution of the shape and color is not lumpable with respect to that coding scheme, thus the hidden factors have to be included in the model that yields better fit for HMM. The lack of lumpability holds even if the transition rates between and within the genes in the generating process are all equal (Supplementary Materials S10). Note that in the cases where MM performed similar or better (20%), the state aggregation was very close to form a lumpable chain due to the stochastic nature of the simulations, thus, favoring MM.

Obviously, different discretization schemes may imply different assumptions of trait evolution; thus, phylogenetic analysis is strongly conditional on a discretization scheme used. Does it mean that the discretization scheme of the first scientist is better than that of the second one? I would argue that it does not since discretization of traits is always a subjective process. An optimal discretization scheme cannot be selected in advance without prior knowledge of a generating process. In fact, this procedure has to be reversed – an optimal model has to be found that can adequately model the chosen discretization scheme. In this regard, the subjectivity associated with the choice of the discretization becomes minimal. For example, MM of the four-state character of the first scientist, and HMM of the second scientist are different but have similar state spaces, and, in fact, they model similar events – the hidden

transitions in HMM tend to approximate the observable transitions of MM (Fig. 8i,g). Thus, both scientists are doing a good job on modeling their characters since they both perform model comparison and use the best-selected models.

Inferring hidden states. – The topology of hidden space in HMM (i.e., the number of hidden states and transitions between them) can be anything that does not produce a lumpable model. Since the topology is unknown *a priori*, one needs testing different models by manually varying parameters of the hidden topology and selecting the best models using model selection criteria. This approach can be implemented in *corHMM* or *RevBayes*. The method similar to that of Pagel and Meade (2006) that uses reversible jump Markov Chain Monte Carlo to sample different models from their posterior distribution can be implemented in *RevBayes* but this direction requires further research. When testing different HMMs, one should keep in mind that the data must be identical, which can be achieved as described in the tail color problem. Note that HMM used in the example above has some similarity to SMM – its rate matrix is structured to reflect independent evolution between the genes. The techniques reviewed for constructing SMM can be used to build hidden topologies of HMM.

Performance of HMM vs. MM. – If lumpability is violated but traditional MM is used instead of HMM it may lead to a significant error in rate estimation and hence biased inference of ancestral states. The error magnitude depends on the rate values and rate ratios in the original transition matrix. For example, if the original three-state chain is aggregated into the two-state chain under scenario when lumpability is fulfilled (*rate.diff.=1*) and not (*rate.diff.≠1*), then the increase in rate values and rate ratios increases the estimation error (Fig. 7b). The substantial bias in rate estimation between MM and variable types of HMM were reported for ancestral character state reconstruction with the empirical and simulated datasets (Beaulieu et al. 2013). If lumpability is violated, the phylogenetic inference may be also biased, which was shown for DNA data when four nucleotides were recoded into fewer groups (Vera-Ruiz et al. 2014). At the same time, if lumpability is valid, state aggregation can be used in algorithms to reduce the size of rate matrix which improves computational performance (Davydov et al. 2017). The simulations for the two-scientist paradox, despite showing a substantially better fit for HMM, detected insignificant differences between HMM and MM rate estimates, which converged on the generating rates (mean squared error was 7×10^{-4} and 9×10^{-4} respectively). This indicates that rate underestimation by MM is context-dependent and if rates are the only focus of inference then, in certain cases, MM can be used as an appropriate estimator. However, if a study aims at inferring both rates and hidden factors then HMM should be preferred.

It has to be born in mind that the performance of HMM is strongly data dependent since, on average, HMM uses more free parameters than MM. The study of Beaulieu et al. (2013) sets the lower limit when HMM can be inferred in ancestral character state reconstruction to 60 - 120 taxa. If fewer taxa are used, then MM provides a sufficient approximation. The behavior of HMM under varying simulation conditions is reported in that study as well. To sum up, HMM provides an insight into the underlying hidden process that cannot be achieved through MM. The application of HMM minimizes subjectivity in trait coding. Hidden Markov model will constantly overperform MM if the generating process is not lumpable and the amount of data is sufficient. The use of HMM is especially essential for analyzing large datasets which naturally contain heterogeneity. The model selection between MM and HMM must be an obligatory component of ancestral character state reconstruction since even a complex MM, because of lumpability violation, might be inappropriate for inference. The type of HMM discussed in the present section, can be also used in tree inference with morphological data, however, HMM application will be limited due to stricter model parametrization; it will likely require the assignment of the same HMM with *a priori* specified hidden states to all characters of the same partition.

Interpreting hidden sates

Rate heterogeneity and GRN states. – Initially, HMMs were primarily proposed to model heterogeneity of evolutionary rates across time (Tuffley and Steel 1998; Beaulieu et al. 2013). As was shown above, hidden states can be interpreted as states of GRN modules; also they are conditional on a discretization scheme. Therefore, all three components – time-heterogeneity, GRN transitions, and subjectivity in trait discretization – are confounded in HMM.

In this light, interpreting hidden states must be taken with caution. For example, in the two-scientist paradox, despite the same process that generated the traits, the two scientists may come to drastically different conclusions. While the first scientist may not report something unexpected, the second scientist may claim that the "key innovation" of being triangular and red was subjected to complex shifts during evolution. Such conclusion, depending on the scientist's imagination, can be connected with some other biological observations to result in a "spectacular" finding, which, in fact, is misleading. In general, I suggest using a moderate interpretation of hidden states – as a simultaneous realization of all confounding processes mentioned above.

Trait homology and HMM. – In phylogenetics, every character statement is a hypothesis of primary homology (Hawkins et al. 1997b). The "real" assessment of whether an observation in one species is homologous or homoplasious in respect to that in the other can be done only through a phylogenetic inference or reconstruction of character evolution (secondary homology). It has been considered that a thoughtful identification of primary homology is a prerequisite for a successful phylogenetic analysis. Obviously, in traditional MM and parsimony, the secondary homology is conditional on primary homology (Agnarsson and Coddington 2008; Brazeau 2011). The discordance between primary and secondary homology, to a large extent, occurs due to the discordance between the morphological traits and underlying GRN states. Hidden Markov models are capable to directly account for this discordance. Theoretically, this means that the quality of primary homology statement should not matter if HMM is used since it can automatically adjust the underlying hidden space. The practical aspect of this consideration requires further research though.

CONCLUSIONS

The present paper lays out a framework for modeling hierarchical and hidden processes whose realization produce discrete phenotypic traits. Both processes, at first glance, being dissimilar are, in fact, interacting - modeling hierarchical process often needs equipping SMM with HMM; at the same modeling hidden processes requires structuring HMM using SMM techniques. The main suggestion of the paper is to use a joint structured Markov model with hidden states (SMM+HMM) that, as was shown, can accommodate manifold of process driving trait evolution. Additionally, the paper summarizes all main techniques which can be used to amalgamate, decompose and structure morphological characters in a mathematically consistent way, which gives a rigorous ground for coding traits into character. The included simulation studies demonstrate how the proposed framework can be implemented in practice. The proposed approaches can be used for developing new methods aiming at reconstructing correlated trait evolution and ancestral ontologies.

SUPPLEMENTARY MATERIALS

The supplementary files are available in the section Supplementary material at *bioRxiv*.

ACKNOWLEDGEMENTS

This work was conducted while a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Thanks are due to Brian O'Meara (University of Tennessee), Sergey Gavrilets (University of Tennessee), Sarah Flanagan (National Institute for Mathematical and Biological Synthesis, University of Tennessee), and Dimitar Dimitrov (University of Copenhagen, Natural History Museum of Denmark) for useful suggestions on the text of the manuscript.

REFERENCES

Abouheif E. 1999. Establishing homology criteria for regulatory gene networks: prospects and challenges. Novartis Foundation symposium. 222:207–21; discussion 222–5.

Agnarsson I., Coddington J.A. 2008. Quantitative tests of primary homology. Cladistics. 24:51–61.

Ahnert S.E. 2017. Structural properties of genotype–phenotype maps. Journal of The Royal Society Interface. 14:20170275.

Akaike H. 1974. A new look at the statistical model identification. IEEE transactions on automatic control. 19:716–723.

Arendt D., Musser J.M., H Baker C.V., Bergman A., Cepko C., Erwin D.H., Pavlicev M., Schlosser G., Widder S., Laubichler M.D., Wagner G.P. 2016. The origin and evolution of cell types. Nature Publishing Group. 17.

Babu M.M., Luscombe N.M., Aravind L., Gerstein M., Teichmann S.A. 2004. Structure and evolution of transcriptional regulatory networks. Current Opinion in Structural Biology. 14:283–291.

Balhoff J.P., Dahdul W.M., Kothari C.R., Lapp H., Lundberg J.G., Mabee P., Midford P.E., Westerfield M., Vision T.J. 2010. Phenex: ontological annotation of phenotypic diversity. PLoS One. 5:e10500.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic biology. 62:725–37.

Beaulieu J.M., O'Meara B.C. 2014. Hidden Markov models for studying the evolution of binary morphological characters. Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer. p. 395–408.

Van Bocxlaer I., Loader S.P., Roelants K., Biju S., Menegon M., Bossuyt F. 2010. Gradual adaptation toward a range-expansion phenotype initiated the global radiation of toads. Science. 327:679–682.

Brazeau M.D. 2011. Problematic character coding methods in morphology and their effects. Biological Journal of the Linnean Society. 104:489–498.

Burnham K.P., Anderson D.R. 2003. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.

Carroll S.B. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell. 134:25–36.

Clark-Hachtel C.M., Linz D.M., Tomoyasu Y. 2013. Insights into insect wing origin provided by functional analysis of vestigial in the red flour beetle, Tribolium castaneum. Proceedings of the National Academy of Sciences. 110:16951–16956.

Cowperthwaite M.C., Meyers L.A. 2007. How mutational networks shape evolution: lessons from RNA models. Annu. Rev. Ecol. Evol. Syst. 38:203–230.

Davydov I.I., Robinson-Rechavi M., Salamin N. 2017. State aggregation for fast likelihood computations in molecular evolution. Bioinformatics. 33:354–362.

Deans A.R., Lewis S.E., Huala E., Anzaldo S.S., Ashburner M., Balhoff J.P., Blackburn D.C., Blake J.A., Burleigh J.G., Chanet B., others. 2015. Finding our way through phenotypes. PLoS Biol. 13:e1002033.

Dececchi T.A., Mabee P.M., Blackburn D.C., Strier K., Altmann J., Brockman D., Bronikowski A., Cords M., Fedigan L., Tacutu R., Craig T., Budovsky A., Wuttke D., Lehmann G., Taranukha D., Smith C., Finger J., Hayamizu T., McCright I., Xu J., Berghout J., Sprague J., Clements D., Conlin T., Edwards P., Frazer K., Schaper K., Parr C., Schulz K., Hammock J., Leary P., Hammock J., Rice J., O'Leary M., Kaufman S., Mabee P., Balhoff J., Dahdul W., Lapp H., Midford P., Vision T., Deans A., Lewis S., Huala E., Anzaldo S., Ashburner M., Balhoff J., Mabee P., Ashburner M., Cronk Q., Gkoutos G., Haendel M., Segerdell E., Dahdul W., Balhoff J., Engeman J., Grande T., Hilton E., Kothari C., Vogt L., Nickel M., Jenner R., Deans A., Dececchi T., Balhoff J., Lapp H., Mabee P., Deans A., Yoder M., Balhoff J., Vogt L., Bartolomaeus T., Giribet G., Göpel T., Richter S., Richter S., Wirkner C., Poe S., Wiens J., Wiens J., O'Keefe F., Wagner P., Wake D., Wake D., Wake D., Wake M., Specht C., West-Eberhard M., Daeschler E., Shubin N., Thomson K., Amaral W., Shubin N., Daeschler E., Coates M., Mabee P., Edmunds R., Su B., Balhoff J., Eames B., Dahdul W., Lapp H., Boisvert C., Boisvert C., Garvey J., Johanson Z., Warren A., Shubin N., Daeschler E., Jr F.J., Shubin N., Daeschler E., Jr F.J., Boisvert C., Mark-Kurik E., Ahlberg P., Coates M., Daeschler E., Shubin N., Jr F.J., Mungall C., Gkoutos G., Washington N., Lewis S., Mungall C., Gkoutos G., Smith C., Haendel M., Lewis S., Ashburner M., Balhoff J., Dahdul W., Kothari C., Haendel M., Lewis S., Ashburner M., Mungall C., Torniai C., Gkoutos G., Lewis S., Haendel M., Haendel M., Balhoff J., Bastian F., Blackburn D., Blake J., Bradford Y., Dahdul W., Lundberg J., Midford P., Balhoff J., Lapp H., Vision T., Dahdul W., Balhoff J., Blackburn D., Diehl A., Haendel M., Hall B., Bastian F., Parmentier G., Roux J., Moretti S., Laudet V., Robinson-Rechavi M., Bairoch A., Cohen-Boulakia S., Froidevaux C., Niknejad A., Comte A., Parmentier G., Roux J., Bastian F., Robinson-Rechavi M., Gkoutos G., Green E., Mallon A.-M., Hancock J., Davidson D., Midford P., Dececchi T., Balhoff J., Dahdul W., Ibrahim N., Lapp H., Jarvik E., Jarvik E., Lebedev O., Coates M., Dahdul W., Dececchi T., Ibrahim N., Lapp H., Mabee P., Swartz B., Sereno P., Kazakov Y., Krötzsch M., Simancík F., Clack J., Ahlberg P., Blom H., Finney S., Ruta M. 2016. Data Sources for Trait Databases: Comparing the Phenomic Content of Monographs and Evolutionary Matrices. PLOS ONE. 11:e0155680.

Erkenbrack E.M., Davidson E.H. 2015. Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. Proceedings of the National Academy of Sciences. 112:E4075–E4084.

Erwin D.H., Davidson E.H. 2009. The evolution of hierarchical gene regulatory networks. Nature Reviews Genetics. 10:141–148.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of molecular evolution. 17:368–376.

Forey P.L., Kitching I. 2000. Experiments in coding multistate characters. Systematics Association Special Volume. 58:54–80.

Fortuna M.A., Zaman L., Ofria C., Wagner A. 2017. The genotype-phenotype map of an evolving digital organism. PLoS computational biology. 13:e1005414.

Glassford W.J., Johnson W.C., Dall N.R., Smith S.J., Liu Y., Boll W., Noll M., Rebeiz M. 2015. Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty. Developmental Cell. 34:520–531.

Hall B.K. 2003. Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. Biological Reviews. 78:S1464793102006097.

Hawkins J.A., Hughes C.E., Scotland R.W. 1997a. Primary Homology Assessment, Characters and Character States. Cladistics. 13:275–283.

Hawkins J.A., Hughes C.E., Scotland R.W. 1997b. Primary Homology Assessment, Characters and Character States. Cladistics. 13:275–283.

Held L.I. 2014. How the snake lost its legs: curious tales from the frontier of evo-devo. Cambridge University Press.

Hinman V.F., Cheatle Jarvela A.M. 2014. Developmental gene regulatory network evolution: Insights from comparative studies in echinoderms. genesis. 52:193–207.

Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic biology.syw021.

Houle D., Govindaraju D.R., Omholt S. 2010. Phenomics: the next challenge. Nature Reviews Genetics. 11:855–866.

Huelsenbeck J.P., Nielsen R., Bollback J.P. 2003. Stochastic mapping of morphological characters. Systematic Biology. 52:131–158.

Jeffreys H. 1961. Theory of probability,(Oxford: Oxford University Press). .

Kemeny J.G., Snell J.L. 1960. Finite markov chains. van Nostrand Princeton, NJ.

Kuratani S. 2009. Modularity, comparative embryology and evo-devo: Developmental dissection of evolving body plans. Developmental Biology. 332:61–69.

Lavine M., Schervish M.J. 1999. Bayes factors: what they are and what they are not. The American Statistician. 53:119–122.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic biology. 50:913–925.

Longabaugh W.J.R., Davidson E.H., Bolouri H. 2005. Computational representation of developmental genetic regulatory networks. Developmental Biology. 283:1–16.

Maddison W.P. 1993. Missing data versus missing characters in phylogenetic analysis. Systematic Biology. 42:576–581.

McCune A.R., Schimenti J.C. 2012. Using genetic networks and homology to understand the evolution of phenotypic traits. Current genomics. 13:74–84.

McDougall C., Degnan B.M., Lowe C., Wu M., Salic A., Evans L., Lander E., Stange-Thomann N., Gruber C., Gerhart J., Kirschner M., Yankura K., Martik M., Jennings C., Hinman V., Denes A., Jékely G., Steinmetz P., Raible F., Snyman H., Prud B., Wawersik S., Maas R., Nielsen C., Degnan S., Degnan B., Jackson D., Meyer N., Seaver E., Pang K., McDougall C., Moy V., Gordon K., Degnan B., Martindale M., Burke R., Peterson K., Dunn E., Moy V., Angerer L., Angerer R., Morris R., Peterson K., Collin R., Raff R., Byrne M. 2011. Modularity of gene-regulatory networks revealed in sea-star development. BMC Biology. 9:6.

McKeown A.N., Bridgham J.T., Anderson D.W., Murphy M.N., Ortlund E.A., Thornton J.W. 2014. Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. Cell. 159:58–68.

Moczek A.P. 2008. On the origins of novelty in development and evolution. BioEssays. 30:432–447.

Monteiro A. 2012. Gene regulatory networks reused to build novel traits. BioEssays. 34:181–186.

Mungall C.J., Torniai C., Gkoutos G.V., Lewis S.E., Haendel M.A. 2012. Uberon, an integrative multi-species anatomy ontology. Genome biology. 13:R5.

Nodelman U., Shelton C.R., Koller D. 2002. Continuous time Bayesian networks. Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence.378–387.

O'Meara B.C. 2012. Evolutionary Inferences from Phylogenies: A Review of Methods. http://dx.doi.org/10.1146/annurev-ecolsys-110411-160331.

Oakley T.H. The eye as a replicating and diverging, modular developmental unit. .

Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. The American Naturalist. 167:808–825.

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.

Pigliucci M. 2010. Genotype–phenotype mapping and the end of the "genes as blueprint"metaphor. Philosophical Transactions of the Royal Society B: Biological Sciences. 365:557–566.

Pinna M.C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics. 7:367–394.

Pleijel F. 1995. On character coding for phylogeny reconstruction. Cladistics. 11:309–315.

Price S.A., Wainwright P.C., Bellwood D.R., Kazancioglu E., Collar D.C., Near T.J. 2010. Functional innovations and morphological diversification in parrotfish. Evolution. 64:3057–3068.

Prud'Homme B., Gompel N., Rokas A., Kassner V.A., Williams T.M., Yeh S.-D., True J.R., Carroll S.B. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature. 440:1050–1053.

Ramirez M.J., Michalik P. 2014. Calculating structural complexity in phylogenies using ancestral ontologies. Cladistics. 30:635–649.

Ramirez M.J. 2007. Homology as a parsimony problem: a dynamic homology approach for morphological data. Cladistics. 23:588–612.

Rebeiz M., Patel N.H., Hinman V.F. 2015. GG16CH05-Rebeiz Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development. Annu. Rev. Genomics Hum. Genet. 16:103–31.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 3:217–223.

Rubino G., Sericola B. 1993. A finite characterization of weak lumpable Markov processes. Part II: The continuous time case. Stochastic processes and their applications. 45:115–125.

Rubino G., Sericola B. 2014. Markov chains and dependability theory. Cambridge University Press.

Sereno P.C. 2007. Logical basis for morphological characters in phylogenetics. Cladistics. 0:070907095847001–???

Shapiro M.D., Bell M.A., Kingsley D.M. 2006. Parallel genetic origins of pelvic reduction in vertebrates. Proceedings of the National Academy of Sciences. 103:13753–13758.

Shbailat S.J., Abouheif E. 2013. The Wing-Patterning Network in the Wingless Castes of Myrmicine and Formicine Ant Species Is a Mix of Evolutionarily Labile and Non-Labile Genes. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 320:74–83.

Shelton C.R., Ciardo G. 2014. Tutorial on Structured Continuous-Time Markov Processes. J. Artif. Intell. Res.(JAIR). 51:725–778.

Siegal M.L. 2013. Evolution of molecular networks. The Princeton guide to evolution. Princeton University Press, Princeton.428–436.

Stadler B.M., Stadler P.F., Wagner G.P., Fontana W. 2001. The topology of the possible: Formal spaces underlying patterns of evolutionary change. Journal of Theoretical Biology. 213:241–274.

Strong E.E., Lipscomb D. 1999a. Character Coding and Inapplicable Data. Cladistics. 15:363–371.

Strong E.E., Lipscomb D. 1999b. Character Coding and Inapplicable Data. Cladistics. 15:363–371.

Team R.D.C. 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Tobias J.A., Cornwallis C.K., Derryberry E.P., Claramunt S., Brumfield R.T., Seddon N. 2014. Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. Nature. 506:359–363.

Tsong A.E., Tuch B.B., Li H., Johnson A.D. 2006. Evolution of alternative transcriptional circuits with identical logic. Nature. 443:415.

Tuffley C., Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. Mathematical biosciences. 147:63–91.

Vera-Ruiz V.A., Lau K.W., Robinson J., Jermiin L.S. 2014. Statistical tests to identify appropriate types of nucleotide sequence recoding in molecular phylogenetics. BMC bioinformatics. 15:S8.

Vogt L. 2017a. Assessing similarity: on homology, characters and the need for a semantic approach to non-evolutionary comparative homology. Cladistics. 33:513–539.

Vogt L. 2017b. The logical basis for coding ontologically dependent characters. Cladistics.1–21.

Vogt L. 2018. Towards a semantic approach to numerical tree inference in phylogenetics. Cladistics. 34:200–224.

Voordeckers K., Pougach K., Verstrepen K.J. 2015. How do regulatory networks evolve and expand throughout evolution? Current Opinion in Biotechnology. 34:180–188.

Wagner G.P. 2007. The developmental genetics of homology. Nature Reviews Genetics. 8:473–479.

Wagner G.P. 2015. Homology in the Age of Developmental Genomics. Evolutionary Developmental Biology of Invertebrates 1. Vienna: Springer Vienna. p. 25–43.

Wake D.B., Wake M.H., Specht C.D., Jablonka E., Raz G., Wake M.H., Wake D.B., Larson A., Wake D.B., Hall B.K., Hall B.K., Diogo R., Arendt J., Reznick D.N., Arendt J., Reznick D.N., Leander B.S., Davidson E.H., Donoghue M.J., Ree R.H., Shapiro M.D., Albert A.Y., Chan Y.F., Schluter D., Marchinko K.B., Barrett R.D.H., Rogers S.M., King M.-C., Wilson A.C., Nachman M.W., Hoekstra H.E., D'Agostino S.L., Manceau M., Domingues V.S., Linnen D.R., Rosenblum E.B., Hoekstra H.E., Zanis M.J., Soltis P.S., Qiu Y.L., Zimmer E., Soltis D.E., Craene L.P.R.D., Irish V.F., Gregory T.R., Alberch P., Gale E., Roth G., Nishikawa K.C., Wake D.B., Shubin N., Tabin C., Carroll S., Shubin N., Tabin C., Carroll S., Carroll S.B., Piatigorsky J., Kirchoff B.K., Lagomarsino L.P., Newman W.H., Bartlett M.E., Specht C.P., Kramer E.M., Abouheif E., Preston J.C., Hileman L.C., Collin R., Miglietta M.P., Kohlsdorf T., Wagner G.P., Goldberg E.E., Igic B., Galis F., Arntzen J.W., Lande R., Marshall C.R., Raff E.C., Raff R.A., Kohlsdorf T., Lynch V.J., Rodrigues M.T., Brandley M.C., Wagner G.P., Zufall R.A., Rausher M.D., Igic B., Lande R., Kohn J.R., Hoekstra H.E., Coyne J.A., Mueller R.L., Macey J.R., Jaekel M., Wake D.B., Boore J.L., Parra-Olea G., Wake D.B., Wake M.H., Gower D.J., Giri V., Dharne M.S., Shouche Y.S., Coen E.S., Meyerowitz E.M., Ditta G., Pinyopich A., Robles P., Pelaz S., Yanofsky M.F. 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. Science (New York, N.Y.). 331:1032–5.

Wiens J.J. 2001. Character Analysis in Morphological Phylogenetics: Problems and Solutions. Syst. Biol. 50:689–699.

Yang Z. 2006. Computational molecular evolution. Oxford University Press.

Yoder M.J., Miko I., Seltmann K.C., Bertone M.A., Deans A.R. 2010. A gross anatomy ontology for Hymenoptera. PloS one. 5:e15991.

BOXES

Trait	An observation of some feature(s) of a phenotype.
Character	A formalized coding of a trait (observation) into a character string (i.e., character) that consists of two or more entities called "character states."
Phenotype	A set of all traits of an organism

Box 1. Definitions of the terms used in this paper.

LEGENDS TO FIGURES

Figure 1. Hidden Markov model and lumpability.

This original three-state Markov model **(a)** can be reduced to a two-state model by either directly aggregating the states (e.g., the states *B* and *C*) if the model is lumpable **(b)**, or by using HMM with two observable and three hidden states if the model is not lumpable **(c).**

Figure 2. The coding schemes for the tail color problem.

(a-d) The four alternative schemes for coding the tail traits; a graph next to each scheme shows dependencies between the characters, which arise due to the anatomical dependencies between the traits. If each of the four schemes is amalgamated into one character with the appropriate incorporation of those dependencies, then all schemes collapse to the same Markov model **(e)**.

Figure 3. Character amalgamation.

(a) Amalgamation of three-two state characters evolving under *Mk2* model results in one four-state character evolving under *Mk-SMM-ind*. The phylogenetic inference using these characters yields the identical and resolved topology. **(b)** If amalgamation of the three two-state characters is done without appropriately structuring the rate matrix (i.e. using *Mk4* model) then the topology is unresolved. **(c)** The amalgamated rate matrix (1024 states) of ten two-state characters; non-zero cells in the matrix are shown in color.

Figure 4. Tree inference for the tail problems.

(a-c) Tail armor case. **(d-f)** Tail color problem. **(a, d)** Expected trees**. (b, e)** Graphs depicting anatomical dependencies between the traits. **(c, f)** Results of inference; *Mk-ind* and *Mk-sw* correspond to *Mk-SMM-ind*, *Mk-SMM-sw* respectively; thick graph bars indicate BF values calculated as shown on the left-side legend; thin orange bars are the posterior probabilities of the expected trees; note that the higher BF values indicate worse model fit, the threshold of BF<3.2 means that the two models yield similar fit. The graphs on the bottom show the topologies of SMM models used for the inference.

Figure 5. Ontology-informed character.

The tail characters linked with ontology. The green links show various types of ontological relationships between the characters and between entities of UBERON anatomy ontology.

Figure 6. GRN-to-phenotype mapping.

The tail color trait is divided into the two states: blue and red (shown with rectangles). The states of hypothetical GRN modules that produce the trait are shown with balls. **(a)** One-to-one mapping. **(b)** Many-to-one mapping. **(c)** Partial mapping.

Figure 7. Lumpable Markov chain and estimation error.

(a) Symmetries in the rate matrix that are necessary for strong lumpability (row-wise sum rule); the original four-state $\{s_1, s_2, s_3, s_4\}$ matrix with rates q_{ij} (shown in purple) is aggregated into the two-state $\{f_1, f_2\}$ matrix with rates $\hat{q}_{ij}$ (shown in grey); the aggregation is lumpable iff $\hat{q}_{12} = q_{13} + q_{14} = q_{23} + q_{24}$ and $\hat{q}_{21} = q_{31} + q_{32} = q_{41} + q_{42}$. **(b)** Rate estimation when aggregated chain is lumpable (*rate.diff.=1*) and not (*rate.diff.≠1*); error (*y* axis) of approximating the original three-state Markov chain with the aggregated two-state Markov chain; the rates in the original chain were identical except one rate in the aggregated state that was divided by a scaling parameter (*rate diff.*) specifying the rate ratio (*x* axis) between this and the remaining rates; the scaling parameter was used to violate strong lumpability property if *rate.diff.≠1*; the results are shown for a range of the transition rates q ={0.1, 0.5, 1, 2} in the original chain (Supplementary Materials, section S7).

Figure 8. Two-scientist paradox.

(a) Two independent genes, that produce color and shape traits, evolve on the phylogeny **(b)**. **(c-g)** The analyses of the first scientist; neither of them supports HMM. **(d-j)** The analyses of the second scientist; HMM used in **(i)** outperforms MM used in **(h)**. The dashed line from **(i)** to **(g)** indicates that HMM in **(i)** approximates MM of **(g).**

FIGURES

Figure 1.

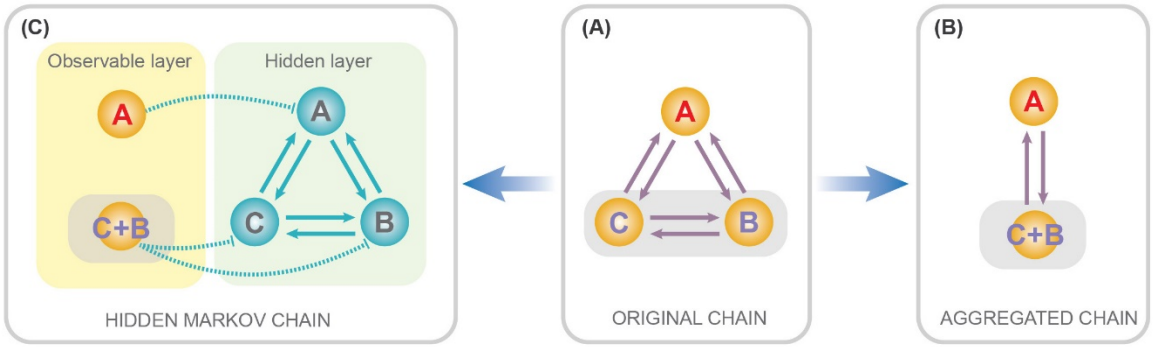


Figure 2.

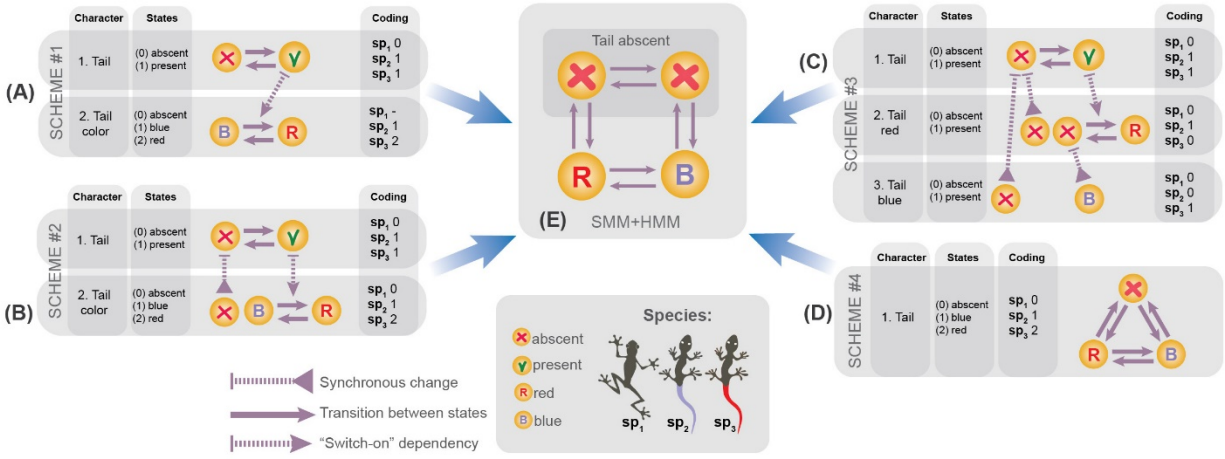


Figure 3.

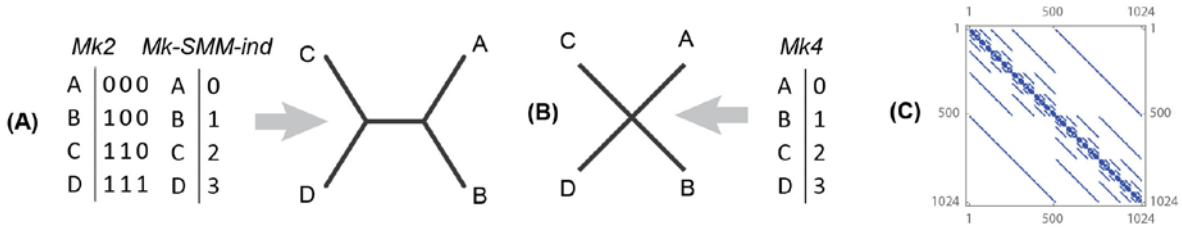


Figure 4.

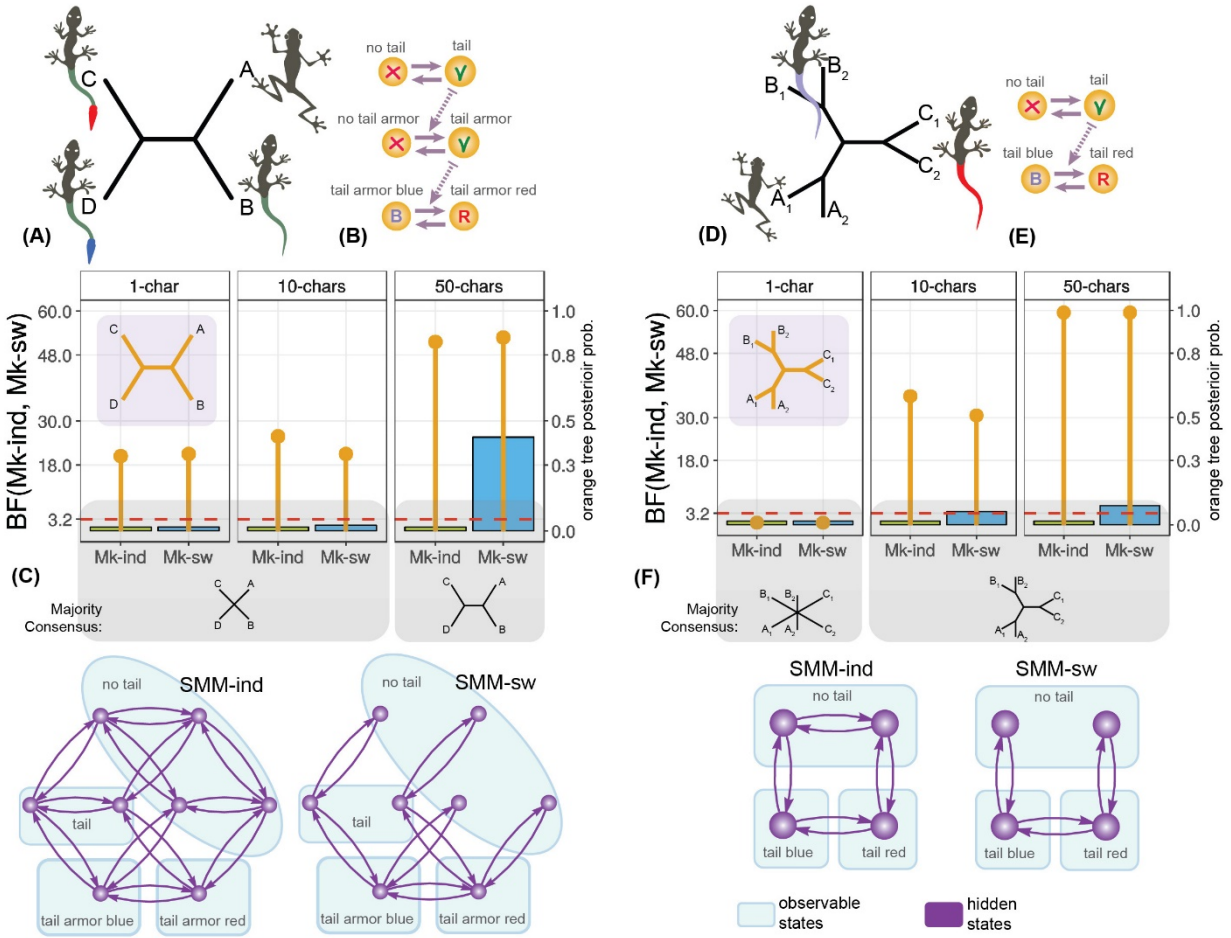

(A)
(B)
no tail
tail
no tail armor
tail armor
tail armor blue
tail armor red
(D)
(E)
tail blue
tail red
1-char
10-chars
50-chars
BF(Mk-ind, Mk-sw)
orange tree posterioir prob.
Mk-ind
Mk-sw
(C)
(F)
Majority Consensus:
SMM-ind
SMM-sw
observable states
hidden states

Figure 5.

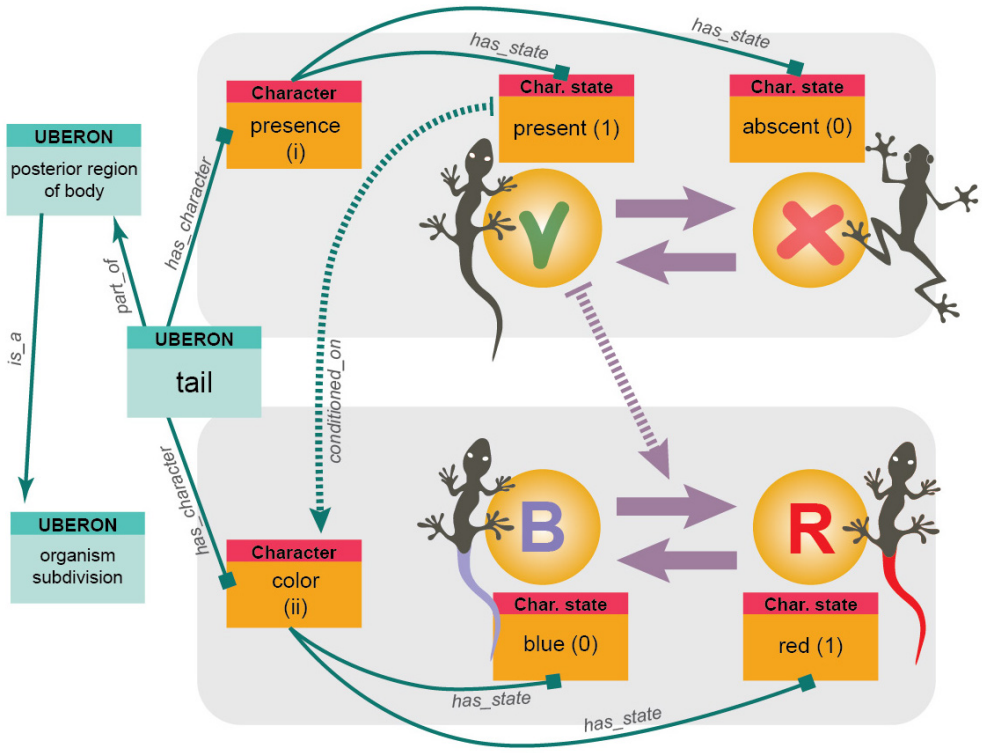


Figure 6.

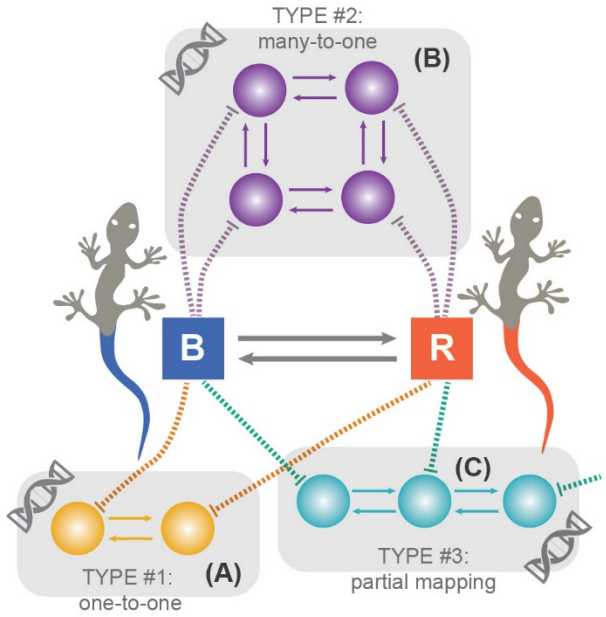


Figure 7.

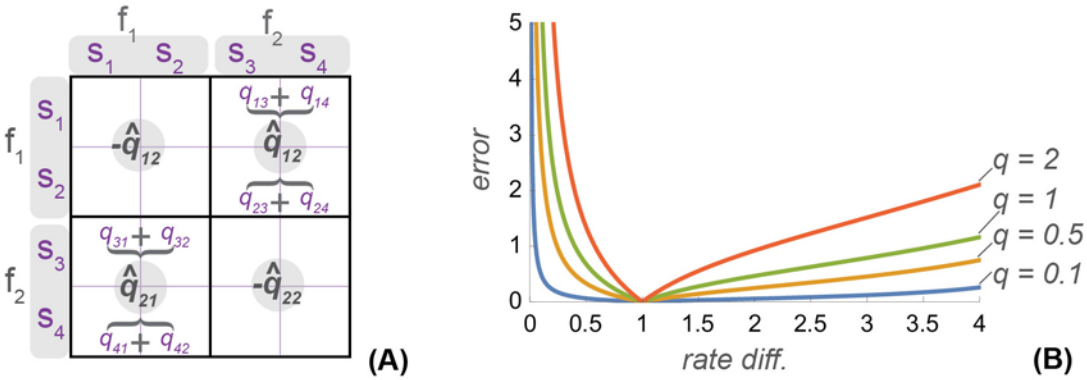
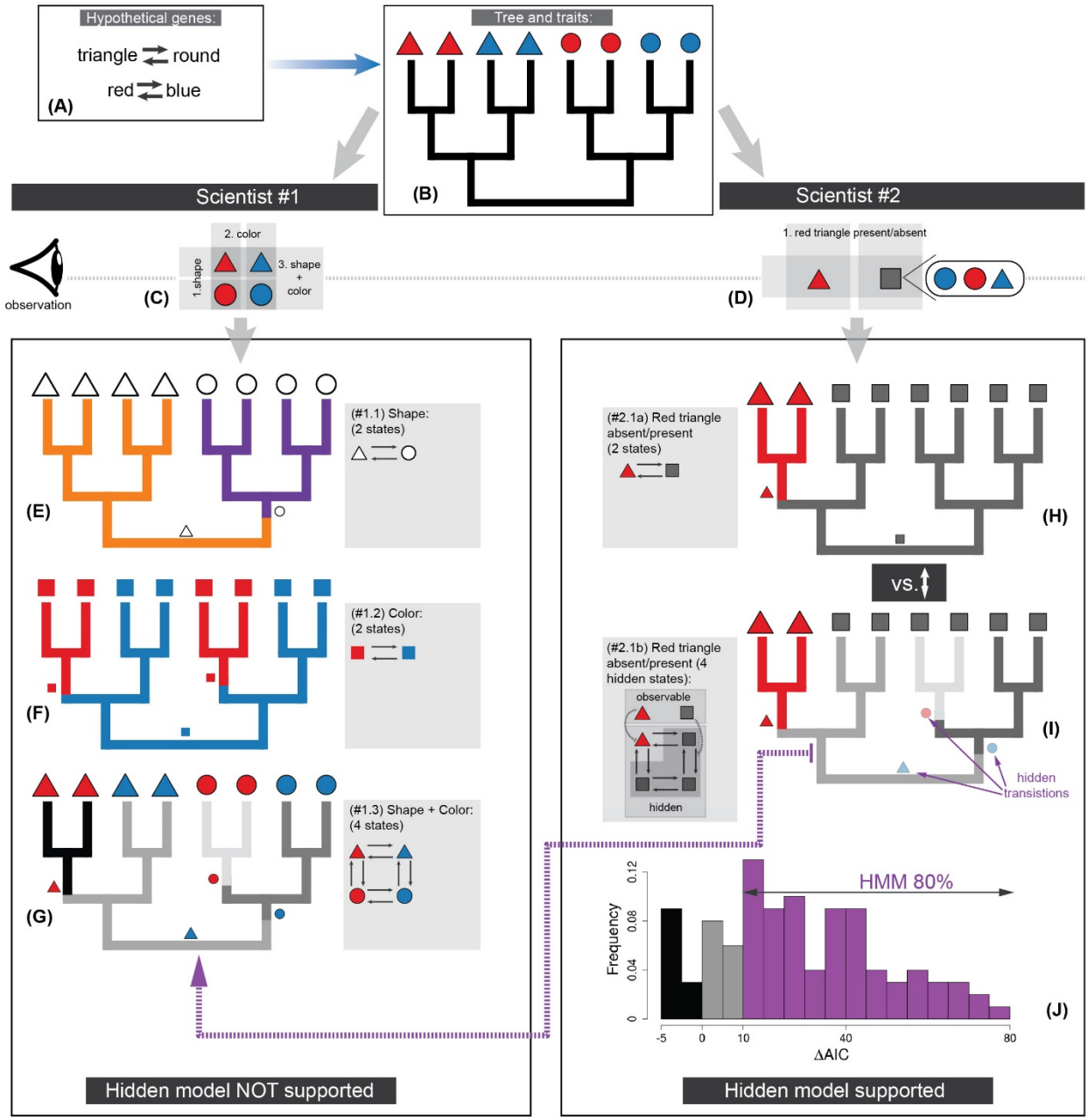

f1
f2
s1
s2
s3
s4
-q̂12
q13 + q14
q̂12
q23 + q24
q31 + q32
q̂21
q41 + q42
-q̂22
(A)
error
rate diff.
q = 2
q = 1
q = 0.5
q = 0.1
(B)

Figure 8.


Hypothetical genes:
triangle ⇄ round
red ⇄ blue
(A)
Tree and traits:
(B)
Scientist #1
Scientist #2
observation
(C)
2. color
1. shape
3. shape + color
(D)
1. red triangle present/absent
(E)
(#1.1) Shape: (2 states)
(F)
(#1.2) Color: (2 states)
(G)
(#1.3) Shape + Color: (4 states)
Hidden model NOT supported
(#2.1a) Red triangle absent/present (2 states)
(H)
vs.
(#2.1b) Red triangle absent/present (4 hidden states):
observable
hidden
(I)
hidden transistions
HMM 80%
Frequency
0
0.04
0.08
0.12
-5
0
10
40
80
ΔAIC
(J)
Hidden model supported