THOMPSON RIVERS UNIVERSITY

Reconstructing Ancestral Bird Plumage Colors Using the Maximum Likelihood Approach

By

Hu, Ning

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science in Data Science

KAMLOOPS, BRITISH COLUMBIA

April 26, 2024

SUPERVISOR

Shaikh, Mateen

©Ning Hu, 2024

ABSTRACT

This project employs the Maximum Likelihood (ML) approach to reconstruct the ancestral states of bird plumage colors, with a special focus on the belly colors of male birds due to their diversity and significance. Markov chain Monte Carlo simulations are used to assess the accuracy of the ML model in inferring the correct states, examining the influence of factors such as phylogenetic tree size, number of states, transition rates, node positions, and the root state's prior probability. The study finds orange to be the most likely color for the most ancient ancestor in the bird phylogenetic tree. However, the accuracy of these predictions is greatly influenced by the evolutionary rates, becoming less reliable for nodes further from the tips, especially for the root node. The research suggests that lower evolutionary rates, larger trees, and fewer states enhance the accuracy of the ML models.

Key Words: maximum likelihood; reconstruct the ancestral states; simulations; phylogenetic tree.

Contents

1	Intr	roduction	1
2	Dat	a Description	5
3	Lite	erature Review	8
	3.1	Notation of Phylogenetic Tree	8
	3.2	Maximum Likelihood	9
	3.3	Performance	13
4	Met	thodology	16
	4.1	Ancestral State Reconstruction Illustrative Example	17
	4.2	Simulation	22
5	Res	ult	29
	5.1	Simulations on Small-scale Scenarios	29
	5.2	Simulations for Real Data	37
	5.3	Plumage Color Reconstruction	39
6	Cor	clusion and Discussion	42
	6.1	Impact of Evolution on Accuracy	42
	6.2	Impact of Prior Root State on Accuracy	43
	6.3	Discussion on Reconstructing Ancestral Bird Colors	44

List of Figures

2.1	Distribution of the number of colors per bird and the distribution of the	
	number of production mechanisms per bird	6
2.2	Birds species color distribution break down by parts	7
3.1	A phylogenetic tree with 4 tips	9
4.1	An example phylogenetic tree of 30 tips with branch lengths $\ldots \ldots \ldots$	18
4.2	The process of modelling using <i>iterative</i> method	19
4.3	Estimated posterior probabilities of the ancestral nodes using iterative	
	approach on data Figure 4.1	21
4.4	Three patterns of transition matrix	23
4.5	Examples of transition rate matrix factors for state count of 5 \ldots .	24
4.6	Change of transition probability over time with different transition rates	
	among 5 states	25
4.7	Change of transition probability over time with different transition rates	
	among 13 states	28
5.1	States reconstruction accuracy for different scenarios	30
5.1	States reconstruction accuracy for different scenarios (cont.) $\ldots \ldots$	31
5.2	Success rate distribution for <i>iterative</i> and <i>likelihood weight</i> methods with	
	100 tips, 5 states and symmetric transition	33
5.3	Root state reconstruction success rate boxplot for <i>iterative</i> method and	
	likelihood weight method	34

5.4	Distribution of average success rate in each scenarios	35
5.5	Root state reconstruction success rate distribution for <i>iterative</i> method	
	and likelihood weight method with 5 states and the $sym1$ transition rate	36
5.6	90% bootstrap percentile confidence interval of the average success rate	
	and ESR of root state in each scenario	38
5.7	Success rate from simulations on a phylogenetic tree with 346 tips and 13 $$	
	states	39
5.8	Likelihood weight of each color	41
5.9	Transition rate matrix under <i>likelihood weight</i> model	41

List of Tables

1.1	Color codes and types of production	4
2.1	An instance of the original bird coloration data	6
4.1	Iterative approach illustration	19
4.2	Simulation scenarios with phylogenetic tree and number of states fixed $% \left({{{\bf{n}}_{{\rm{s}}}}} \right)$.	26
4.3	Simulation scenarios with number of states and transition rate matrix fixed	27
4.4	Simulation scenarios with phylogenetic tree and transition rate matrix fixed	27
5.1	P-value of hypothesis test on the mean value of average success rate and	
	ESR of root node	37
5.2	Maximum log-likelihoods of the birds color supposing the root state in	
	each color	40

Chapter 1

Introduction

The plumage color of birds plays a vital role in their phylogenetic history. Their vibrant hues and patterns serve as signals for crucial aspects for their survival and reproduction, such as attracting mates, deterring predators, and facilitating parent-offspring communication (Price-Waldman and Stoddard, 2021). Furthermore, the unique structure of birds' retinas allows them to perceive colors in a four-dimensional space, including the ultraviolet spectrum, making their colors even more impressive to themselves than to humans (Hill and McGraw, 2006). Therefore, understanding the ancestral coloration of birds and the most likely color transitions can offer valuable insights into their phylogenetic history. Given the limited evidence about the color of their ancestral species, statistical methods can be employed to estimate the likely plumage colors of extinct species by analyzing the plumage colors of existing bird species in conjunction with the phylogenetic tree.

The colorful plumage of birds is generated by the interaction between molecular structures and light. As outlined by Hill and McGraw (2006), bird coloration is influenced by various factors, including pigments and the arrangement of pigments within the feather microstructure, as well as environmental conditions.

According to Hill and McGraw (2006), the pigmentary color is determined by the

structure and arrangement of pigments within the tissue. There are two primary types of pigments: melanins and carotenoids. Melanins contribute to darker shades, with two distinct types: eumelanins, responsible for black, grey, or brown hues, and pheomelanins, which give rise to chestnut, rufous, and reddish. Carotenoids, a diverse group of chemical compounds, play a significant role in generating red, orange, and yellow feathers. However, the presence of melanins can also mask the colors yielded by carotenoids. Structural colors, produced by the physical interactions of light waves with nanostructures, are another important component of bird coloration. These colors can complement or interact with pigmentary colors to produce a wide range of hues like blue, green, and iridescence. Unpigmented feathers appear as white in color, while those with structural effects exhibit an even brighter white appearance.

The evolution of plumage color patterns is primarily influenced by pigments and nanometre scale structures changes in feathers. However, due to limited biological data, estimating colors through their production mechanisms is challenging. In a comprehensive overview, Vinther (2015) synthesizes findings from multiple studies to reveal that the presence of pigments has been identified in fossil material and some studies have established correlations between color categories and melanosome morphology. The accuracy of inferring the color categories such as brown, black, and grey from the melanin morphological variables can exceed 80%. However, Vinther also notes that carotenoids do not preserve well and are barely observed in fossil, and the production of structural colors is not well studied. Given the complexity of avian plumage coloration, including interactions between structural and pigmentary colors, limited evidence is available for extinct bird coloration.

In evolutionary biology, comparative methods are commonly used to reconstruct the ancestral states (Felsenstein, 1985), which involve mapping the states of characters or traits onto a phylogenetic tree. A phylogenetic tree is a graphical hypothesis that illustrates the lineage relationships among different organisms (Hall, 2013). The length of branches on the tree can be measured in units of time, representing things like genetic distance, possibly generation time, etc (Pagel, 1999). Comparative methods rely on comparing traits, behaviors, or genetic sequences across different species to infer their evolutionary history. The states reconstruction process involves analyzing similarities and differences among species to estimate the most possible states for those that are currently unknown.

In the realm of discrete character reconstruction in biology, two widely utilized comparative methods are Maximum Parsimony (MP) and Maximum Likelihood (ML) (Williams et al., 2006). The MP approach operates under the assumption that the probabilities of character state gains and losses are equal and seeks to minimize the number of state change events. Basically, the MP algorithm follows the rule that an ancestral node in the phylogenetic tree should share the same states with its immediate descendants. In cases where there are no common states shared by the descendants, the ancestral state is assigned the union of its descendants' states. The MP approach is usually satisfied, especially when the transition rates between states are relatively slow (Cunningham et al., 1998). On the other hand, ML assumes constant state transition rates, treating evolution as a time-continuous Markov process, which takes the evolution time into account. The ML approach considers every combination of ancestral states, searching for the optimal transition rate between states to maximize the likelihood of observed states of the extant species.

The ancestral character reconstruction problem is widely studied using binary states, such as the presence of phenotypes, DNA sequences or protein sequences. In the context of protein sequence reconstruction (Williams et al., 2006), both MP and ML are capable of accurately estimating ancestral states, with ML generally outperforming MP. In simulations regarding binary state reconstruction, Royer-Carenzi et al. (2013) suggest that the accuracy of both models relies on specific conditions, such as the topology of the tree and even the position of individual nodes within the evolutionary tree. While Holland et al. (2020) finds that MP outperforms ML in most cases, except the the transition rate between two state is highly asymmetrical. These simulations show that the performance is also influenced by the transition rate and the distance of the ancestral to the tips. However, usually, it is considered that when state changes are infrequent, MP can be

Code	Color	\mathbf{Type}
BK	black	melanin (eumelanin)
BR	brown	melanin (eumelanin)
BL	blue	structural (melanin components)
GR	green	structural (melanin components)
RE	red	carotenoid
WH	white	unpigmented (structural)
OR	orange	carotenoid
GE	grey	melanin (eumelanin)
IR	irridescent	structural (melanin components)
RU	rufous	melanin (phaeomelanin)
TU	turquoise	structural (carotenoids components)
YG	yellow-green	structural (carotenoids components)
YE	yellow	carotenoid

Table 1.1: Color codes and types of production.

viewed as a limiting case of ML and ML performs better when the evolution rate is high (Schluter et al., 1997).

The bird color dataset employed in this project was compiled by a student under the supervision of Dr. Reudink, drawing data from *Birds of the Worlds* (Billerman et al., 2022). The colors were identified by referencing online bird photographs, and were categorized into 13 distinct color categories, as detailed in Table 1.1 along with their corresponding production types. While most studies focus on binary state, both MP and ML approaches can be applied to the reconstruction for more than two states. For the phenotype of plumage color, each color category is considered as one possible state value, and the values can change from one to another. In this project, we assume that the process of bird plumage color change parallels the DNA mutation, without delving into the mechanisms behind color production. Given that plumage traits are believed to evolve rapidly (Omland and Lanyon, 2000) and the lack of certainty regarding equal probabilities for transitions between colors and their reversals, the ML method is used and assessed for reconstructing ancestral bird plumage colors. The phylogenetic tree of birds used in this project is one of the hypothesis trees on the website 'birdtree.org' with 9993 species (Jetz et al., 2012). All the species in the dataset are included in the tree.

Chapter 2

Data Description

The bird color dataset contains color records for 346 bird species. This dataset includes 20 variables representing colors found in 10 different parts of the birds. Each part can have up to two colors, and records for male and female birds are recorded separately. The first entry from the dataset is presented in Table 2.1, showcasing the coloration of the male species Melanogenys in the Anisognathus genus, belonging to the Thraupidae family. The auricular plumage is black, while the rump displays a combination of blue and black plumage.

Within the dataset, birds exhibit varying colors on different parts of their bodies. On average, each species displays an average of 3.32 distinct colors in their plumage. Male birds typically have a slightly higher average of 3.38 colors, while females average 3.25 colors. It is worth noting that only a minority of species have a single color across their entire body, and most species showcase a diversity of coloration attributed to different production mechanisms (Figure 2.1).

Figure 2.2 illustrates the distribution of feather colors across various parts of the birds. Notably, black, brown, and green colors are the most prevalent in each part, with the exception of crissum feathers, where rufous and white are the most common colors. It's worth highlighting that carotenoid-based coloration, encompassing red, yellow, and

Column	Value
Species	Anisognathus_melanogenys
Sex	Μ
Auricular	ВК
Auricular2	<na></na>
Rump	BL
Rump2	ВК
Back	BL
Back2	ВК
Crown	BL
Crown2	<na></na>
WingTip	ВК
WingTip2	<na></na>
DorTail	BL
DorTail2	ВК
Throat	YE
Throat2	<na></na>
Breast	YE
Breast2	<na></na>
Belly	YE
Belly2	<na></na>
Crissum	YE
Crissum2	<na></na>

Table 2.1: An instance of the original bird coloration data. Column names end with '2' is the second color presents on one bird.



(a) Color diversity of bird species.

(b) Color production mechanisms of Bird Species.

Figure 2.1: Distribution of the number of colors per bird and the distribution of the number of production mechanisms per bird.



Figure 2.2: Birds species color distribution break down by parts.

orange hues, is relatively uncommon among all body parts in the dataset.

In this project, we have chosen to focus on the belly color, given that it is a body part that does not exhibit extreme color preferences like the *DorTail* and *WingTip*. In cases where a species presents two colors at the belly, we consider a 50% probability to the appearance of each color.

Chapter 3

Literature Review

3.1 Notation of Phylogenetic Tree

Figure 3.1 is an illustrative example of an phylogenetic tree. The nodes located at the tips correspond to extant species, while the internal nodes are the common ancestors. For instance, node 1 and node 2 share a most recent common ancestor denoted as node 5. Similarly, node 1 and node 3 trace their most recent common ancestor to node 6, and so forth. A phylogenetic tree can be defined by a set of nodes, denoted as N, an immediate ancestor function, represented by f(i), where i is a node, and a branch length function e(f(i), i), which denotes the length of branch to node i and its parent (Farris, 1970). In Figure 3.1, we can observe a tree consisting of nodes 1, 2, 3, 4, 6, 7, along with their corresponding ancestor functions:

$$f(1) = 5, f(2) = 5, f(3) = 6,$$

 $f(4) = 7, f(5) = 6, f(6) = 7$

and branch length functions, which simply equal the prescribed constants for illustration:

$$e(5,1) = e_{5,1}, e(5,2) = e_{5,2}, e(6,3) = e_{6,3},$$

 $e(7,4) = e_{7,4}, e(6,5) = e_{6,5}, e(7,6) = e_{7,6}.$



Figure 3.1: A phylogenetic tree with 4 tips.

3.2 Maximum Likelihood

Felsenstein (1981) introduced the ML approach for constructing phylogenetic trees from DNA sequences. This approach paved the way for the development of computationally feasible methods to find the ML of a tree. Pagel (1994) applied these methods to reconstruct the states of discrete characters. The use of the ML method is also extended to test correlations between two binary characters in the context of phylogenetic analysis.

The likelihood of a phylogenetic tree in a particular state is calculated as the product of the probabilities of all the nodes achieving those states. In this context, let s_i represent the state of node i, and $p(s_i)$ denotes the probability of node i occupies state s_i . Assuming that in the evolutionary process, each branch's evolution depends solely on its immediate ancestor, The likelihood of Figure 3.1 in a state of $(s_1, s_2, ..., s_7)$ can be expressed as:

$$\mathcal{L}^{s_1,\dots,s_7} = p(s_7)p(s_6|s_7)p(s_5|s_6)p(s_1|s_5)p(s_2|s_5)p(s_3|s_6)p(s_4|s_7).$$

Here, $\mathcal{L}^{s_{j_1}, s_{j_2}, \dots, s_{j_N}}$ denotes the likelihood of the tree with node *i* as the root and its descendants j_1, j_2, \dots, j_N having states $s_{j_1}, s_{j_2}, \dots, s_{j_N}$, respectively. By grouping the expressions within brackets

$$\mathcal{L}^{s_1,\dots,s_7} = p(s_7)\{p(s_6|s_7)[p(s_5|s_6)p(s_1|s_5)][p(s_2|s_5)p(s_3|s_6)]\}\{p(s_4|s_7)\},\$$

we can observe that the likelihood can be defined in a recursive format, where it becomes a function of the likelihood of its two sub-trees.

The general model for the likelihood of a sub-tree with root k at state S_k (a collection of states for nodes in tree k) can be expressed as:

$$\mathcal{L}^{S_k} = p(s_k | s_{f(k)}) \mathcal{L}^{S_i} \mathcal{L}^{S_j}, \tag{3.1}$$

where *i* and *j* represent the immediate descendants of *k*, and $p(s_k|s_{f(k)})$ represents the probability of node *k* being in state s_k given the state of its parent node, f(k). \mathcal{L}^{S_k} is the likelihood for one possible state assignment.

In Equation (3.1), when considering a node k without an ancestor, $p(s_k|s_f(k))$ is replaced with its prior probability $p(s_k)$. If the state of the node k is known, such as in the case of tips, then \mathcal{L}^{S_k} will be zero for all S_k except when S_k is observed, in which case $\mathcal{L}^{S_k} = 1$.

To take into account the uncertainty of internal nodes, the overall likelihood is calculated as the sum over all possible assignments of states to all the nodes. Let S_k represent all possible state combination of the tree with root k. The likelihood of all the states combination can be expressed as

$$\sum_{S_k \in \mathbb{S}_k} \mathcal{L}^{S_k} = \sum_{S_k \in \mathbb{S}_k} p(s_k | s_{f(k)}) \mathcal{L}^{S_i} \mathcal{L}^{S_j}$$
$$= \sum_{s_k} p(s_k | s_{f(k)}) (\sum_{S_i \in \mathbb{S}_i} \mathcal{L}^{S_i}) (\sum_{S_j \in \mathbb{S}_j} \mathcal{L}^{S_j})$$

The summation operators on S_i and S_j could move to right because the change of probability on one sub-tree does no influence that of its sibling sub-tree. We use \mathcal{L}_k denotes the likelihood of a sub-tree with root k with all possible state combinations. Then,

$$\mathcal{L}_k = \sum_{s_k} p(s_k | s_{f(k)}) \mathcal{L}_i \mathcal{L}_j, \qquad (3.2)$$

For a ML model M on a tree with a root node r, the likelihood

$$\mathcal{L}(M) = \mathcal{L}_r = \sum_{s_r} p(s_r) \mathcal{L}_i \mathcal{L}_j.$$
(3.3)

To estimate the value of $p(s_k|s_{f(k)})$, let $q(s_{f(k)}, s_k)$ be the probability of transition from node f(k) at state $s_{f(k)}$ to node k at state s_k , then $p(s_{f(k)}, s_k) = p(s_{f(k)})q(s_{f(k)}, s_k)$. The posterior probability

$$p(s_k|s_{f(k)}) = \frac{p(s_{f(k)}, s_k)}{p(s_{f(k)})} = q(s_{f(k)}, s_k).$$
(3.4)

The transition probability is commonly assumed to adhere to a time-continuous Markov process. In this process, state change rates depend solely on the previous state, and they remain constant over a specific period of evolutionary time. Felsenstein (1981) initially assumed the existence of a base change rate shared by all states, while Pagel (1994) extended this idea by developing simultaneous transition rate estimation. Following Pagel's approach, all the transition probability from one state to another across different time can be represented by the transition matrix $\mathbf{Q}(t)$, with each entry $q_{xy}(t)$ denoting the probability of transition from state x to state y after a time interval t. If r_{xy} is the transition rate from x to y, then $r_{xy} = \frac{dq_{xy}(t)}{dt}$ when x and y are different. To build the transition rate matrix \mathbf{R} , each entry $r_{x,y}$ reflects the transition rate from state x to state y if $x \neq y$. Otherwise, r_{xy} is defined as $-\sum_{x\neq y} r_{xy}$. For instance, in the case of 3×3 transition rate matrix \mathbf{R} , it can be illustrated as follows:

$$\begin{pmatrix} -(r_{12}+r_{13}) & r_{12} & r_{13} \\ r_{21} & -(r_{21}+r_{23}) & r_{23} \\ r_{31} & r_{32} & -(r_{31}+r_{32}) \end{pmatrix}$$

The transition probability after a short time dt from time t can be approximated as:

$$\boldsymbol{Q}(t+dt) = \boldsymbol{Q}(t)(\boldsymbol{I} + \boldsymbol{R}dt). \tag{3.5}$$

To find the solution, the equation can be rearranged and expressed differentially as

$$\frac{\boldsymbol{Q}(t)}{dt} = \frac{\boldsymbol{Q}(t+dt) - \boldsymbol{Q}(t)}{dt} = \boldsymbol{Q}(t)\boldsymbol{R}.$$

From this, the solution can be derived as

$$\boldsymbol{Q}(t) = \exp(\boldsymbol{R}\boldsymbol{t}), \tag{3.6}$$

in which the exp function is the matrix exponential. Let X be a square matrix, the exponential of X is defined as ¹

$$\exp(\boldsymbol{X}) = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{X}^k$$

Combining with Equation (3.4) and (3.6), the objective of ML model is to maximize Equation (3.3) by optimizing all non-diagonal entries of the transition rate matrix \mathbf{R} . However, the prior probability of the root is unknown, which significantly influences the inference process and also poses a challenge in the analysis (FitzJohn et al., 2009). To assign priors to the root node, one can carefully choose appropriate values based on prior knowledge or beliefs. However, in cases where there is no prior belief, Pagel (1999) suggests an iterative approach in finding a proper prior root state.

In Equation (3.3), each element in the summation can be considered as the contribution of the root in each separate state to the likelihood. The proportion of root in state a

$$P(s_r = a) = \frac{p(s_r = a)\mathcal{L}_i\mathcal{L}_j}{\mathcal{L}(M)}$$
(3.7)

is used as the posterior weight as the root in preference of state a. In Pagel's method, model M is achieved by maximizing the likelihood of each combination of ancestral states, which uses different transition rate parameter, and is therefore considered as the "quasiposterior weight".

¹ "Matrix exponential", Wikipedia, last modified Nov 13, 2023, https://en.wikipedia.org/w/index.php?title=Matrix_exponential&oldid=1184902135

The iterative approach initially builds an ML model with an equally distributed root state. The prior probability of the root state is then replaced with the quasi-posterior weight and recalculates the ML. Repeat this procedure until there is no improvement in the likelihood. During the this process, the prior probability of root node could shift the weights towards the most likely state. Pagel suggests more simulation work to assess its behaviour.

3.3 Performance

The likelihood ratio (LR) test is a practical tool for comparing any two nested models, where one model is a special case of the other (Pagel, 1994). In this context, nested models mean that one model partially constrains the parameters of the other, making it a more restricted or specific version of the larger model. The LR test used to assess whether the constrained model (I) significantly degenerates the general model (D) is defined as:

$$LR = -2\ln[\mathcal{L}^I/\mathcal{L}^D]. \tag{3.8}$$

When model I can be expressed as a special case of D, the test statistic is asymptotically distributed as χ^2 with degree of freedom equal to the difference between the number of free parameters of two model D and I.

When dealing with non-nested models, the LR test may not follow a χ^2 distribution. In such cases, a Monte Carlo procedure (Goldman, 1993), can be applied to estimate the distribution of the LR test statistic. By using this Monte Carlo procedure and combining it with nonparametric techniques like bootstrapping, an empirical estimate of the distribution of the LR test statistic can be obtained. In this approach, LR is often interpreted as a measure of "support". If there is evidence of an LR value greater than 2, it is typically considered an indication that the two models are significantly different (Pagel, 1999). This can help in assessing whether a model performs similarly with the other model.

Both Royer-Carenzi et al. (2013) and Holland et al. (2020) conduct simulations to

assess the performance of ML models with binary states. Performance is measured by the models' accuracy in correctly inferring unknown node states, with a match between the inferred and true states considered as a success.

In a ML model M, the posterior probability of node k in state i given the observed data d, which is the combination of leaf states, is given by

$$p_M(s_k = i | \boldsymbol{d}) = \frac{p_M(s_k = i, \boldsymbol{d})}{p_M(\boldsymbol{d})}.$$

By the definition of likelihood, $p_M(\mathbf{d}) = \mathcal{L}(M|\mathbf{d})$, in which $\mathcal{L}(M|\mathbf{d})$ is the same with the former notation $\mathcal{L}(M)$. Then the posterior probability can be expressed by likelihood of the data:

$$p_M(s_k = i | \boldsymbol{d}) = \frac{\mathcal{L}(M | s_k = i)}{\mathcal{L}(M)},$$
(3.9)

where $\mathcal{L}(M|s_k = i)$ is the likelihood of the leaf states with the state of node k in i. In the recursive process of calculating likelihood (Equation 3.2), \mathcal{L}_k is replaced with $p(s_k = s_i|s_{f(k)})\mathcal{L}_i\mathcal{L}_j$.

Royer-Carenzi uses expectation of success assessing the performance. The simulation returns the most likely state of each unknown node, which is the state with largest posterior probability. In case the two states are equally weighted, a random state is returned. The probability of success is then set to 1 if the returned state matches the true state, which is recorded during simulation, otherwise the probability is set to 0.

Holland introduces two metrics to assess the error rate: quantised score and raw score. To calculate the quantised score, the probability of a node state is converted to state as Royer-Carenzi's method, except that it uses 0.7 as the threshold. A state of a node is returned when probability of being in the state is greater than 0.7, otherwise the state is considered as ambiguous. Only an unambiguous state that does not match the true state counts as an error. This error rate is referred as quantised score. The raw score directly uses the the posterior probability of incorrectly identifying the state of the unknown node. If a node is estimated to have a 0.10 probability of being in state A, 0.90

probability of being in state B, and the true state is A, the error rate is 0.90.

Both expectation of success and quantised score loses some certain of information compared to raw score, because they convert the continuous probability to discrete numbers. They are useful while evaluating the performance between ML and MP models, because MP only yields the state of the node other than probability.

Chapter 4

Methodology

The ML problem-solving process of Equation (3.3) is referred to Holland et al. (2020). It is executed through the functions 'make.mkn' and 'find.mle' from 'diversitree' package in R. The method that is used for optimization is *nlminb*. The 'make.mkn' function is slightly modified to make it support tips with more than one possible states. The *nlminb* optimization method necessitates initial parameter values for the optimization process. Depending on these initial values and the nature of the objective function, the outcomes may either reach global optimization or become entrapped in local optima. To avoid the local convergence problem, five different random starting conditions are set to perform optimizations and the best solution is taken. For the first condition, all non-diagonal entries of the transition rate matrix—our initial parameters—are set to a value of 10^{-6} . For subsequent conditions, we multiply each initial parameter by a random factor ranging from 1.2 to 2.5. This results in initial transition rates varying from 10^{-6} to 10^{-4} . Such a strategy ensures that the initial parameters are not excessively close to the true transition rates in any scenario within this project, thus maintaining the comparability of our findings. The 'keep.tip' function from the same package is used to prune a subset from the entire birds phylogenetic tree, keeping only the 346 species in the dataset.

In our model, all combination of ancestral states shares the same transition rate

matrix. In this case, to employ the iterative approach for inferring the root state Equation (3.7) is equivalent to the posterior probability of the root state. The approach provides a way to shift the root state towards a preferred distribution that maximizes the likelihood. It, however, does not guarantee that the root state would move toward the true state value, in which case the iteration could make the performance even worse.

In this project, we come up with a likelihood weight approach to infer the prior of the root state. In this approach, we first build ML models for the root in each separate state. Let M_i be the model when root state is *i*, the prior probability of the root state at state *i* is given by

$$p(s_r = s_i) = \frac{\mathcal{L}(M_i)}{\sum_{s_r} \mathcal{L}(M_{s_r})}.$$
(4.1)

The proportion of the likelihood in each state over the that of the summation of all the states does not work as a true weight, as the likelihood in the denominator comes from different models. This method, however, provides a sensible estimator on how much the root is preferred in each state.

This chapter begins with a demonstration to illustrate the ML process of reconstructing the ancestral states using both methods to infer the prior probability of root node. Following this, simulations are designed to compare the accuracy of both models in inferring the correct states of unknown nodes.

4.1 Ancestral State Reconstruction Illustrative Example

The ML method is applied to the data shown in Figure 4.1 to demonstrate the process of reconstructing the color of the ancestral species. In this phylogenetic tree, there are 4 colors for the species - black, yellow, blue and red. The number of parameters to the optimize problem is 12, which are all the entries of the transition rate matrix except for the diagonals.

In the iterative approach to infer the root state, the prior probability of the root is



Figure 4.1: An example phylogenetic tree of 30 tips with branch lengths. The states of tips are represented by the color of nodes.

equally set to 25% in initial. The maximum log-likelihood of the model with this prior is -23.94, with contributions of -30.86, -30.40, -24.77, and -24.51 from the root in each state (Table 4.1). The posterior distribution of the root state, calculated based on the likelihood proportion of the root in each state, is (0.10%, 0.16%, 43.46%, 56.28%). This posterior is subsequently assigned as the new prior to root node. After nine iterations, the ML no longer improves beyond four decimal places, and the root node demonstrates a preference for the *red* color in this process (Figure 4.2).

The root state in the final model is predominantly *red*, with nearly 100% probability.

Table 4.1: Applying *iterative* approach to the tree of Figure 4.1. The table only show first three loops of the iteration. The proportion of root in each state contribute to the likelihood is used as the prior of the root state.

Iteration	Root Prior			Log	Contribution of Each State				
	black	yellow	blue	red	Likelihood	black	yellow	blue	\mathbf{red}
1	0.25	0.25	0.25	0.25	-23.94	-30.86	-30.40	-24.77	-24.51
2	9.8e-4	1.6e-3	0.43	0.56	-23.22	-36.31	-35.78	-24.47	-23.56
3	2.1e-6	3.5e-6	0.29	0.71	-23.07	-42.27	-43.06	-25.79	-23.14



(a) Change of ML during iteration.



(b) Change of root state prior distribution across iterations. Beginning with the second iteration, the prior probability for the root is derived from the posterior distribution of the previous iteration. The colors of the lines represent corresponding states.

Figure 4.2: The process of modelling using *iterative* method.

The transition rate matrix \boldsymbol{R} for the model is

	black	yellow	blue	red
black	(-0.0113)	0.0113	0	0)
yellow	0	-0.0070	0.0070	0
blue	0	0	-0.0097	0.0097
red	0.0039	0	0.0219	-0.0258

The entry in the row corresponding to *black* and the column corresponding to *yellow* indicates that the transition from *black* to *yellow* occurs at a rate of 0.0113. Under this transition rate, the state transition matrix between node \boldsymbol{r} and \boldsymbol{a} is

$$exp(\mathbf{R} * 8) = \begin{array}{c} black \\ yellow \\ black \\ blue \\ red \end{array} \begin{pmatrix} 0.9137 & 0.0839 & 0.0024 & 0.0001 \\ 0 & 0.9453 & 0.0527 & 0.0020 \\ 0.0011 & 0 & 0.9314 & 0.0675 \\ 0.0270 & 0.0013 & 0.1521 & 0.8196 \end{pmatrix}$$

With the transition rate matrix, the posterior probability of each node state can be computed following Equation (3.9). The estimated distribution of each node is illustrated in Figure 4.3.

In the likelihood weight approach for selecting the prior probability of the root node, we initially hypothesize the root state to be *black*, *yellow*, *blue* and *red*, respectively. The corresponding maximum log-likelihoods of the leaf states are found to be -24.07, -23.90, -22.97 and -22.79. Then the likelihood weights of the root are calculated to be 11.39%, 13.52%, 34.28%, 40.80% according to Equation (4.1). We employ this probability distribution as the prior probability of the root, and recalculate the ML. In this model,



Figure 4.3: Estimated posterior probabilities of the ancestral nodes using iterative approach on data Figure 4.1.

The transition rate matrix R is

$$\begin{array}{cccccccccccccc} black & yellow & blue & red \\ black \\ yellow \\ blue \\ red \end{array} \begin{pmatrix} -0.0113 & 0.0113 & 0 & 0 \\ 0 & -0.0069 & 0.0069 & 0 \\ 0.0011 & 0 & -0.0130 & 0.0119 \\ 0.0029 & 0 & 0.0226 & -0.0255 \end{pmatrix}$$

which is similar with that from the iterative approach, with a notable difference in the transition from *blue* to *black*. In the *likelihood weight* method, *blue* has a chance to transit to *black*, while in the iterative method, this is 0, because of the difference of the prior of the root state.

4.2 Simulation

The simulation is to evaluate the performance of ML method in estimating the average probability of correctly inferring the states of all the unknown nodes. Markov chain Monte Carlo (MCMC) method is used to generate the samples. Then based on the states of the tips, we build the ML models to infer the probability of true state of unknown nodes. Multiple scenarios are established to evaluate the impact of variables such as the number of tips, total states, and evolutionary rate on performance. The scenario of bird ancestral plumage color reconstruction is also simulated to estimate the accuracy of ML model for the bird plumage color dataset. The sample size is 100 in each scenario.

Input

The simulation input includes phylogenetic trees with branch length, number of states, a root state and transition rates between all states.

Three subsets of phylogenetic trees are created from phylogenetic tree we use to reconstruct the ancestral birds plumage color. These subsets keep the first 50, 100 and 150 tips, respectively. The length of each tree from root to the tips is normalized to 10, which is not necessarily varied because it only affects the transition probability on a given branch, while this can be achieved by scaling transition rate.

The number of states is varied from 3 to 9 step by 2. The root state is randomly drawn from all states.

There are three patterns that we consider for the transition rate matrix. The first pattern is symmetric (Figure 4.4a). That is the transition rates between two states in two directions are almost the same. The second pattern is grouped states matrix (Figure 4.4b), in which the state transition shows high rate within the same group, while the transition rate is low for states in different groups. The third pattern is asymmetric (Figure 4.4c), which is opposite to the symmetric pattern. The the transition rates



Figure 4.4: Three patterns of transition matrix in the example where there are 5 states. The color depth represents different levels of transition rate.

between two states in two directions are greatly different.

Four sets of transition rate matrices are designated for the symmetric pattern. Each set includes free parameters with mean values of 0.02, 0.2, 0.1, and 1, respectively. These parameters follow uniform distributions: Uniform(0.01, 0.03), Uniform(0.07, 0.13), Uniform(0.17, 0.23), and Uniform(0.7, 1.3). For grouped and asymmetric pattern, there is one transition rate matrix for each. The larger values are sampled from aforementioned distributions, while the smaller values are drawn from another distribution with a lower mean. The choice of the distributions depends on the performance of the symmetric pattern, as explained in the scenario section.

In summary, there are six sets of transition rate matrices, with examples of 5x5 matrices illustrated in Figure 4.5. We refer to these sets as *sym1*, *sym2*, *sym3*, *sym4*, *grp* and *asy*, respectively.

The mean of lowest transition rate is 0.02, because when the number of states is 5, the probability of the state remains unchanged is about 50% after 10 units of length during evolution. Lower transition rate than that may result in insufficient diversity in exhibition all the states in tips. For the transition rate being 0.1, the transition probability does not change too much after 10 units length. When the transition rate is 0.2, the transition probability achieves almost stable after evolving 5 units length, and that is 1 unit length when the transition rate is 1. Higher transition rate will not change the transition matrix too much (Figure 4.6).

-0.0770	0.0220	0.0221	0.0197	0.0132
0.0160	-0.0820	0.0300	0.0129	0.0231
0.0288	0.0298	-0.1033	0.0152	0.0295
0.0174	0.0297	0.0151	-0.0761	0.0139
0.0174	0.0150	0.0108	0.0260	-0.0692

-0.4152	0.1133	0.0893	0.0863	0.1264
0.0784	-0.3400	0.0792	0.0735	0.1090
0.1256	0.1160	-0.3979	0.0863	0.0700
0.0923	0.0929	0.0942	-0.3527	0.0732
0.1082	0.0842	0.0910	0.0937	-0.3772

(a) Symmetric matrix: mean transition rate (b) Symmetric matrix: mean transition rate 0.02.



0.1.



(c) Symmetric matrix: mean transition rate (d) Symmetric matrix: mean transition rate 1. 0.2.





(e) Grouped matrix: High transition rates within first three and last two states (mean: 0.1). Mean transition rate between states in each group is 0.02.

(f) Asymmetric matrix: If transition rate from state i to j is sampled from uniform distribution with mean 0.02, the reverse rate has a mean of 0.1.

Figure 4.5: Examples of transition rate matrix factors for state count of 5.



Figure 4.6: Change of transition probability over time with different transition rates among 5 states.

Generate Samples

MCMC sampling process explains as following. Assuming the transition rate matrix to be \mathbf{R} , the transition probability matrix $\mathbf{Q}(t)$ on a single branch with length t can be obtained following Equation (3.6). If the parent state is i, the probability of its child node's state to be j is $q_{ij}(t)$, which is the i, j entry of transition matrix $\mathbf{Q}(t)$. By sampling from the transition probability distribution, starting from the root node and traversing the tree in preorder, we can simulate an evolutionary history with full knowledge of all node states. The ML models are constructed with the leaf states known, while the states of the other node remain unknown.

To compare the effect of the prior probability of the root state on the performance of the ML models, four choices of prior distributions of the root state are used to build the ML models on each sample. Two of the prior distributions are determined by using iterative method and likelihood weight method. The other two prior are set to the true root state and a random false root state, respectively, which are used to build benchmark models.

Number of Tips	Number of States	Transition Rate Matrix
100	5	sym1
100	5	sym2
100	5	sym3
100	5	sym4
100	5	grp
100	5	asy

Table 4.2: Simulation scenarios with phylogenetic tree and number of states fixed.

The output of all four models is the posterior probability of inferring the correct state of each node, which is used in this project to evaluate the performance of the models. The probability is similar to the raw score to evaluate error rate in Holland et al. (2020), except that we use "success rate" here.

Scenarios

To reduce the scenarios of the simulation, we initially conduct simulations by maintaining the phylogenetic tree featuring 100 tips and 5 states, while varying all 4 symmetric transition rate matrices. Subsequently, we identify two transition rate matrices that exhibit the best performance in inferring the correct states. The distributions utilized to generate these two transition matrices are then employed to generate grouped and asymmetric matrices. Finally, simulations are conducted using both grouped and asymmetric matrices, each with 100 tips and 5 states (Table 4.2).

Next, we maintain a constant transition rate matrix type and vary the phylogenetic tree by adjusting the number of tips to one of three specific values: 50, 100, or 150. The type of transition rate matrix which yields the highest average success rate in previous scenarios are chosen. Table 4.3 uses the symmetric matrix with the mean of transition rate 0.02 as example.

Finally, we choose the optimal phylogenetic tree which performs best in above scenarios, and vary the states. Suppose the number of tree tips is 150, the scenarios are presented as Table 4.4.

Number of Tips	Number of States	Transition Rate Matrix
50	5	sym1
100	5	sym1
150	5	sym1

Table 4.3: Simulation scenarios with number of states and transition rate matrix fixed.

Table 4.4: Simulation scenarios with phylogenetic tree and transition rate matrix fixed.

Number of Tips	Number of States	Transition Rate Matrix
150	3	sym1
150	5	sym1
150	7	sym1
150	9	sym1

An additional scenario is established to assess the accuracy of inferring bird plumage color for this project. In this scenario, the number of tips and states matches the dataset. However, the ML algorithm is time-consuming under these conditions. On average, the execution time for a single solution ranges from 5 minutes to 2 hours, with an average duration of 35 minutes, on a CPU clocked at 4.3 GHz. For the likelihood weight approach, 14 priors are set and each prior is executed five times with varying initial parameters to mitigate the risk of local convergence issues. Producing a single sample for a single task would require approximately 40 hours. To speed up the process, multitasking parallel processing is employed, reducing the time required to 8 days for 100 samples distributed across 20 tasks.

Therefore, the choice of transition rate is limited to two types for assessing the accuracy of plumage color inference and its sensitivity to different transition rates. Following insights gained from earlier simulation outcomes, the accuracy primarily depends on the average transition rate, with minimal influence from the transition pattern. Consequently, both transition rate matrices adopt a symmetric pattern.

To ensure that the simulated data is similar to the bird color dataset, the proportion of tips in each state ranges from 1% to 30%. Transition rates, with means of 0.02 and 0.05, are drawn from Uniform(0.01, 0.03) and Uniform(0.04, 0.06) distributions for



Figure 4.7: Change of transition probability over time with different transition rates among 13 states.

the two matrices, respectively. The phylogenetic tree's height is approximately 18.6. Throughout the evolutionary process, the transition probability stabilizes with a rate of 0.02, whereas it stabilizes midway with a rate of 0.05 (Figure 4.7).

Evaluation

The performance of the ML models is assessed by calculating the average success rate across all ancestral nodes, based on the success rates of each node derived from simulation outputs. Notably, the root node is specifically examined as a unique individual node to offer additional insights into the model's performance. The accuracy in reconstructing the bird plumage color is assessed through simulations, along with the examination of various factors such as the distance of the ancestral node to the tips, the transition pattern and rate, and the number of tips and states.

Chapter 5

Result

5.1 Simulations on Small-scale Scenarios

The simulations present a comprehensive examination of the accuracy of ancestral state reconstruction across different scenarios with varying node heights (Figure 5.1). In general, across all methods tested, the posterior probability of inferring the correct state, which defines our accuracy measure, decreases as the distance from the inferred nodes to the tips increases. The exception is that the accuracy of the models built with the true root state increases when the nodes close to root node. The *true root* method consistently shows the highest success rate across almost all scenarios, especially when the nodes are far from the tips. The *false root* method generally has the lowest success rate, significantly deviating from the performance of the *true root* method with the increase of nodes' height. The *iterative* method and *likelihood weight* method perform similarly between the two benchmark methods across most scenarios, with the *likelihood weight* method slightly outperforming the *iterative* method in some cases.

As seen in Figure 5.1a, the accuracy of all methods decreases with the increase transition rate when the transition follows the symmetric pattern. For a symmetric transition matrix with rate sym1, all the methods perform better than any other transition matrices. The accuracy of all methods is almost the same when the nodes are close to



(a) Accuracy for scenarios with a tree of 100 tips and 5 states. Each non-diagonal entry in symmetric transition rate matrices sym1, sym2, sym3 and sym4 are drawn from uniform distribution (0.01, 0.03), (0.07, 0.13), (0.17, 0.23), and (0.7, 1.3), respectively. The higher values in grouped and asymmetric matrices are from uniform(0.01, 0.03), while the lower values are from uniform(0.07, 0.13).



(b) Accuracy for scenarios with 5 states and the sym1 transition rate.

Figure 5.1: States reconstruction accuracy for different scenarios. The height of the nodes is defined as the distance from the inferred nodes to the tips. The root node and tips are not included to avoid their influence on the smoothing fitted line.



(c) Accuracy for scenarios with a tree of 100 tips and the *sym1* transition rate. Figure 5.1: States reconstruction accuracy for different scenarios (cont.).

the tips, while with the increase of the heights, that of the *false root* model decays the quickest. The performance of the *true root* model decreases slowest and even goes up when the height is greater than 4. As the transition rate increases, the performance of all four models converges, particularly at the higher-level nodes. For the sym2, sym3 and 4 transition rate matrices, the accuracy declines rapidly with the increase of transition rate and eventually stabilizing at a constant value for all models. For the sym2, the accuracy stops decreasing when the height of the node is 6, while that of the *true root* model slightly increases after that. The accuracy stabilizes at around 0.2 for nodes at a height of 4 for sym3, and at roughly 0.5 for sym4. The influence of the transition rate is 1.

The accuracy patterns observed in models using the grp and asy transition rate matrices exhibit similarities. In both scenarios, the models' performances are ranked between the symmetric transition sym1 and sym2 in terms of accuracy. The *true root* model consistently outperforms others, while the *false root* model ranks lowest in effectiveness. The *iterative* and *likelihood weight* models show comparable outcomes. This indicates that the specific pattern of the transition rate matrix, whether grouped or asymmetric, does not significantly affect model performance.

Figure 5.1b compares the success rates of four different methods across three different phylogenetic tree sizes. The success rate for all methods tends to increase as the tree size increases. The *false root* method shows the most sensitivity to node height, especially when the number of tree tips is 150. There is a deep decline in accuracy when the height of tree exceeds 4. The *likelihood weight* method consistently outperforms the *iterative* method, with both achieving intermediate success rates. While accuracy for these methods also decreases as the increase of node height, the decline is not as marked as seen with the *false root* method.

The difficulty of reconstructing ancestral states increases with the number of states increases as shown in Figure 5.1c. Same as the former scenarios, the *true root* method is consistently the most effective, while the *false root* method is the least effective and shows significant sensitivity to both the number of states and the height of the node. Uniquely in this context, the *iterative* method shows the poorest performance when the state count reaches 9.

Figure 5.2 presents scatterplots comparing success rates against node height for a phylogenetic tree characterized by 100 tips and 5 states using a symmetric transition pattern. For both the *iterative* and *likelihood weight* methods, the patterns of distribution across each scenario appear similar. At lower transition rates, the success rates predominantly cluster near 1, indicating a high likelihood of accurate ancestral state inference. As the transition rates and node heights increase, the success rates tend to converge around a value, which is 0.2 in this context. The distinction between the two methods lies in the *iterative* method's tendency to produce a higher frequency of extreme success rates (0 or 1) compared to the *likelihood weight* method when the nodes are close to the root.

The comparison of the root state reconstruction success rates between the *iterative*



Figure 5.2: Success rate distribution for *iterative* and *likelihood weight* methods with 100 tips, 5 states and symmetric transition.

method and *likelihood weight* method indicates that the performance of the two methods varies significantly across most scenarios (Figure 5.3). As seen in Figure 5.3a, the interquartile range (IQR) for *iterative* method is broader than that of the *likelihood weight* method. For the *iterative* method, the lower quartiles are near zero, suggesting a tendency towards lower success rates, while the *likelihood weight* method consistently shows higher values for these quartiles. As the transition rate increases, the success rate distributions for both methods appear to converge near a success rate of 0.2. It is noteworthy that with grp and asy transition matrices, both methods exhibit lower performance compared to the sym1 matrix. This observation contrasts with the patterns seen when evaluating average success rates.

As the number of tips in the phylogenetic trees increases, there is a noticeable improvement in the performance of both the *iterative* and *likelihood weight* methods. This improvement is marked by higher median success rates, as depicted in Figure 5.3b. By contrast, as depicted in Figure 5.3c), there is a decline in success rates with an increased number of states. These trends suggest that the complexity added by more



(a) Accuracy for scenarios with a tree of 100 tips and 5 states.



(b) Accuracy for scenarios with 5 states and (c) Accuracy for scenarios with a tree of 150 the sym1 transition rate. tips and the sym1 transition rate.





Figure 5.4: Distribution of average success rate in each scenarios. 'L' and 'S' in the scenario label represents number of tips and number of states, respectively.

states negatively impacts the success rate, while a greater number of tips provides more data that may enhance the accuracy of both methods. As seen in both figures, the *iterative* method displays greater variability in response to changes in the number of tips and states, indicating that the *likelihood weight* method may be more robust under a range of scenarios.

Average success rate of all nodes is used to evaluate the overall performance of both *iterative* and *likelihood weight*. The distribution of average success rate of both method is all most identical in each scenario (Figure 5.4). As the success rates of both methods from each sample are calculated from the same object, the paired two sample test is used to compare the two-sample mean. The Shapiro-Wilk normality test on the difference of the average success rate shows that the difference is not normally distributed. Therefore, a percentile test combined with bootstrap sampling method is applied.

Bootstrap is also employed for the root node. However, the distributions of the success rate for the two methods are distinct. In Figure 5.5, the success rate of the *iterative* method demonstrates a bimodal distribution, indicating two peaks in performance at 0



Figure 5.5: Root state reconstruction success rate distribution for *iterative* method and *likelihood weight* method with 5 states and the sym1 transition rate.

and 1, whereas the *likelihood method* method presents a generally unimodal distribution with slight fluctuations. Consequently, using the sample mean as a comparative statistic is inappropriate. Instead, the comparison is based on the proportion of success rates exceeding the threshold associated with random chance. Specifically, this threshold is 0.20 for a state count of five, and 0.33 for three states. This chosen metric effectively quantifies the frequency at which each method successfully predicts the correct state with a notably high rate of success. We refer to this metric as Exceedance Success Rate (ESR).

The null hypothesis of the test is that the parameter (p) of two methods are equal. One-sided tests in both directions are conducted to determine which method performs better. All the tests use the significant level of 5%. The test results show that the *likelihood weight* method performs better than the *iterative* method in most scenarios (Table 5.1). In instances where this is not the case, the accuracy levels of the two methods are comparable. For the average success rate, the performance of the two methods aligns closely, except that there is significant evidence to suggest that mean value of *likelihood weight* method is greater than that of *iterative* method in scenarios 7, 8, and 11. In these scenarios, the success rates are moderate for both methods (Figure 5.1 and 5.4). In the majority of cases, the *likelihood weight* method outperforms the *iterative* method in terms of inferring the root state. The exception arises under extreme conditions where both methods face difficulties in accurately inferring the root state, leading to no significant difference in their performance. Overall, the *iterative* method

Table 5.1: P-value of hypothesis test on the mean value of average success rate and ESR of root node, comparing between the *iterative* (iter.) and *likelihood weight* (l.w.) methods. The p-value is the proportion of the difference less than 0 for right-sided test and more than 0 for left-sided test.

scenario		$H_1: p(ite)$	r.) > p(l.w.)	$H_1: p(iter.) < p(l.w.)$		
no.	scenario	average	root	average	root	
1	L:100, S:5, sym1	0.7678	1	0.2322	0	
2	L:100, S:5, sym2	0.3238	1	0.6762	0	
3	L:100, S:5, sym3	0.6694	0.9986	0.3306	0.0014	
4	L:100, S:5, sym4	0.7940	0.5617	0.2060	0.4383	
5	L:100, S:5, grp	0.4196	1	0.5804	0	
6	L:100, S:5, asy	0.3909	1	0.6091	0	
7	L:50, S:5, sym1	0.9976	1	0.0024	0	
8	L:150, S:5, sym1	0.9859	0.9998	0.0141	0.0002	
9	L:150, S:3, sym1	0.4896	1	0.5104	0	
10	L:150, S:7, sym1	0.3900	0.9951	0.6100	0.0049	
11	L:150, S:9, sym1	0.9967	0.9317	0.0033	0.0683	

does not surpass the *likelihood weight* method in any of the scenarios.

Figure 5.6 shows how the average success rate and ESR of root state changes across different scenarios. In both methods, The average success rate decreases as the transition rate and number of states increase, or as the size of the phylogenetic tree decreases (Figure 5.6a). This trend also holds for the ESR of the root state, with a notable exception: the accuracy is lower at a transition rate of 0.2 (sym3) compared to a rate of 1.0 (sym4). For *iterative method*, the accuracy of inferring the root state further declines under grouped or asymmetric transition patterns. Overall, the effectiveness of both ML models is significantly tied to factors such as transition rate, the scale of the phylogenetic tree, and the diversity of states.

5.2 Simulations for Real Data

Given its consistent performance, the *likelihood weight* method is exclusively utilized in simulations involving real data scenarios. Specifically, this approach is applied to a phylogenetic tree with 346 tips, where 13 distinct states have been identified at



(b) ESR of root state.

Figure 5.6: 90% bootstrap percentile confidence interval of the average success rate and ESR of root state in each scenario.



Figure 5.7: Success rate from simulations on a phylogenetic tree with 346 tips and 13 states. *Likelihood weight* method is employed. Root node is included.

the tips, to simulate conditions closely mirroring observed data. Across both transition rate scenarios, success rates diminish as node height increases, and it deteriorates faster with higher transition rate (Figure 5.7). Notably, the ESR for the root state at a mean transition rate of 0.02 is lower compared to when it's at 0.05. The 90% percentile confidence interval for the mean of average success rate ranges between 0.5070 and 0.5274 at a 0.02 transition rate, and between 0.2461 and 0.2533 at a 0.05 transition rate. The 90% confidence interval of the proportion of the success rate exceeding 1/13 is (0.21, 0.36) for the 0.02 transition rate, and is (0.43, 0.59) for the 0.05 transition rate.

5.3 Plumage Color Reconstruction

The *likelihood weight* method is employed to set the prior probability for the root state to reconstruct the ancestral plumage color. The log-likelihood of the observed data under the assumption of the root in each color is shown in Table 5.2, and the weight of likelihoods shows *orange* (OR) color weighs most among all the colors, exceeding 40% (Figure 5.8). Based on the new root prior, the posterior probability of root shows that the root has 99% chance to be *orange*. However, the accuracy is dependent on the transition

BK	\mathbf{BR}	BL	\mathbf{GR}	\mathbf{RE}		
-613.6069 WH	-611.9785 OR	-613.2790 GE	-614.7886 IR	-612.1578 RU		
-615.8971	-610.7384 TU	-612.4108 YG	-611.9778 YE	-616.0901		
	-613.2344	-613.5475	-616.6292			

Table 5.2: Maximum log-likelihoods of the birds color supposing the root state in each color.

rate among different colors. If the transition probability between colors has just reached the stable state over the evolution (transition rate is 0.02), the average success rate falls within 0.7007 and 0.7227. But the accuracy in inferring the root state is equivalent to making a random guess. An increase in the transition rate diminishes the accuracy in identifying the correct states. Specifically, for nodes situated midway between the root and the tips, the model loses its predictive efficacy when the transition rate is 0.05 (Figure 5.7).

The estimated transition rate matrix is shown as Figure 5.9. It indicates that orange (OR), red (RE), grey (GE) and rufous (RU) are the most unstable colors, as the rate that they stay unchanged are the lowest, whereas green (GR) and yellow (YE) are most unlikely to change to other colors. The transition probabilities from red (RE) to black (BK) and from rufous (RU) to green (GR) are higher than other transitions. As the given transition rate matrix is not verified for its accuracy, it simply indicates that, under these transition rates, the dataset aligns with a time-continuous Markov process evolution as suggested by Maximum Likelihood analysis.



Figure 5.8: Likelihood weight of each color.

-0.16	0.03	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	вк	0.4
0.00	-0.17	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	BR	0.2
0.00	0.00	-0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.01	BL	0
0.02	0.00	0.00	-0.06	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	GR	-0.2
0.32	0.00	0.00	0.00	-0.35	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	RE	-0.4
0.00	0.01	0.01	0.00	0.00	-0.13	0.00	0.01	0.00	0.07	0.00	0.01	0.03	wн	
0.00	0.05	0.08	0.00	0.00	0.14	-0.42	0.00	0.03	0.00	0.12	0.00	0.00	OR	
0.00	0.00	0.17	0.00	0.00	0.00	0.00	-0.33	0.00	0.00	0.00	0.17	0.00	GE	
0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	-0.23	0.00	0.00	0.00	0.07	IR	
0.06	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.00	-0.30	0.00	0.00	0.00	RU	
0.05	0.00	0.03	0.00	0.00	0.07	0.00	0.03	0.00	0.00	-0.21	0.00	0.03	тυ	
0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	-0.25	0.14	YG	
0.00	0.00	0.00	0.01	0.00	0.02	0.01	0.00	0.00	0.04	0.00	0.00	-0.08	YE	
BK	BR	BL	GR	RE	МM	OR	GE	R	RU	J	YG	ΎΕ		

Figure 5.9: Transition rate matrix under *likelihood weight* model.

Chapter 6

Conclusion and Discussion

The study evaluates the efficacy of the ML approach for reconstructing ancestral plumage colors of birds. It is found that the accuracy of inferring the ancestral state by ML models is influenced by several factors, including the size of the phylogenetic tree, the number of states, the transition rates, and the positions of nodes within the tree.

6.1 Impact of Evolution on Accuracy

Generally, the transition rate between states was a crucial factor, with lower rates leading to more accurate predictions as they allowed for a slower and thus more traceable evolution. The pattern of transition, whether grouped or asymmetric, seems to have a negligible impact, showing similar performance to the symmetric pattern with an average rate. High accuracy is achievable across all nodes when when the transition rate is relatively low, that is the transition probability has not reached the stable status during the evolutionary history. Conversely, higher transition rates reduce the reliability of predictions, especially for nodes further from the tips.

Large phylogenetic trees tend to provide more accurate predictions due to richer evidence supporting the hypothesis. On the other hand, an increase in the number of states, which introduces new variables at a exponential pace, results in greater uncertainty and subsequently leads to a decrease in accuracy.

The accuracy of inferring ancestral states typically diminishes with the increase of the height of the nodes within a tree. This decrease in accuracy can be accelerated by factors such as an increased number of states or a reduction in tree size, potentially reducing the model accuracy to near-random outcomes. However, in scenarios where the ancestral root state is known, accuracy tends to improve for the nodes that are close to the root.

The simulation scenarios presented offer a preliminary insight into the general trends of the accuracy of ancestral state inference. However, due to the computational demands, only a limited number of levels are assigned to each input factor. To comprehensively assess the quantitative effects of various factors on the accuracy of ancestral state inference, further in-depth studies and extensive simulations are required.

6.2 Impact of Prior Root State on Accuracy

The study also reveals the significance of the root state's prior probability, which has a remarkable influence on model accuracy when transition rates are not excessively high. Two methods for estimating the root's prior probability are evaluated against benchmarks of two fixed root states. Accuracy changes among the four models are similar across various scenarios. The model based on the true root state is consistently the highest reliability in ancestral states reconstruction. In contrast, the model using an incorrect root state is the least reliable, and its performance is particularly affected by evolution factors. The *iterative* and *likelihood weight* methods perform better than the *false root* method but do not reach the high success rates of the *true root* method. The *iterative* method exhibits extreme low or high accuracy, especially notable at the nodes close to root. In most scenarios, the *likelihood weight* method performs better in inferring the node states.

As the node height increases, the success rate declines across all methods, yet the *true root* method is notably less impacted by node height. For the nodes close to the tips,

all methods achieve very high success rates. However, under conditions of high transition rates, where transition probabilities reach a stable state throughout evolutionary history, the efficacy of all four methods aligns, demonstrating a uniform level of performance.

6.3 Discussion on Reconstructing Ancestral Bird Colors

The *likelihood weight* method is applied to determine the prior probability of the root states for reconstructing the belly color of the bird species provided in the dataset. The result shows that the root state of the bird color is most probably to be *orange*. However, the accuracy is notably affected by the rate of evolutionary changes in bird plumage colors. It is observed that the closer the nodes are to the tips, the more reliable the results become. Conversely, the accuracy diminishes as the height of the node increases. To gain initial insights into the expected accuracy, this study conducted simulations on a phylogenetic tree with 346 tips, distributed across 13 distinct states, and evaluated them at two separate transition rates.

Lacking evidence about the speed of evolutionary changes makes it challenging to ascertain the accuracy level confidently. Enhancing accuracy could benefit from more observed data and fewer parameters. One straightforward strategy could involve expanding the dataset to include a broader range of species.

This project has limited its focus to reconstructing the belly color of male birds only. If color changes across different body parts are independent and follow identical transition rates, employing more complex models that consider plumage colors across different parts of the bird's body can be applied to improve accuracy. This could be a potential direction for future research.

Bibliography

- Billerman, S. M., B. K. Keeney, P. G. Rodewald, and T. S. Schulenberg (2022). Birds of the world. https://birdsoftheworld.org/bow/home.
- Cunningham, C. W., K. E. Omland, and T. H. Oakley (1998). Reconstructing ancestral character states: a critical reappraisal. *Trends in ecology & evolution* 13(9), 361 366.
- Farris, J. S. (1970). Methods for computing wagner trees. Systematic Zoology 19(1), 83 - 92.
- Felsenstein, J. (1981). Evolutionary trees from dna sequences: A maximum likelihood approach. Journal of Molecular Evolution 17(6), 368 – 376.
- Felsenstein, J. (1985). Phylogenies and the comparative method. The American Naturalist 125(1), 1-15.
- FitzJohn, R. G., W. P. Maddison, and S. P. Otto (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58(6), 595 – 611.
- Goldman, N. (1993). Statistical tests of models of dna substitution. Journal of molecular evolution 36(2), 182 – 198.
- Hall, B. G. (2013, 03). Building Phylogenetic Trees from Molecular Data with MEGA. Molecular Biology and Evolution 30(5), 1229–1235.
- Hill, G. E. and K. J. McGraw (2006). Bird coloration. Harvard University Press.

- Holland, B. R., S. Ketelaar-Jones, A. R. O'Mara, M. D. Woodhams, and G. J. Jordan (2020). Accuracy of ancestral state reconstruction for non-neutral traits. *Scientific Reports* 10(1), 1 – 10.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers (2012). The global diversity of birds in space and time. *Nature* 491(7424), 444 – 448.
- Omland, K. E. and S. M. Lanyon (2000). Reconstructing plumage evolution in orioles (icterus): Repeated convergence and reversal in patterns. *Evolution* 54(6), 2119–2133.
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings: Biological Sciences* 255(1342), 37 – 45.
- Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology 48(3), 612 – 622.
- Price-Waldman, R. and M. C. Stoddard (2021). Avian coloration genetics: Recent advances and emerging questions. *Journal of Heredity* 112(5), 395 – 416.
- Royer-Carenzi, M., P. Pontarotti, and G. Didier (2013). Choosing the best ancestral character state reconstruction method. *Mathematical Biosciences* 242(1), 95 109.
- Schluter, D., T. Price, A. Mooers, and D. Ludwig (1997). Likelihood of ancestor states in adaptive radiation. *Evolution* 51(6), 1699 – 1711.
- Vinther, J. (2015). A guide to the field of palaeo colour: Melanin and other pigments can fossilise: Reconstructing colour patterns from ancient organisms can give new insights to ecology and behaviour. *BioEssays* 37(6), 643 – 656.
- Williams, P. D., D. D. Pollock, B. P. Blackburne, and R. A. Goldstein (2006). Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Computational Biol*ogy 2(6), e69 – 0605.