

Information in morphological characters

Congyu Yu¹, Qigao Jiangzuo¹, Emanuel Tschopp², Haibing Wang³, and Mark Norell¹

¹American Museum of Natural History

²University of Hamburg

³Institute of Vertebrate Paleontology and Paleoanthropology Chinese Academy of Sciences

June 15, 2021

Abstract

The construction of morphological character matrices is central to paleontological systematic study, which extracts paleontological information from fossils. Although the word information has been repeatedly mentioned in a wide array of paleontological systematic studies, its meaning has rarely been clarified and there has not been a standard to measure paleontological information due to the incompleteness of fossils, difficulty of recognizing homologous and homoplastic structures, etc. Here, based on information theory, we show the deep connections between paleontological systematic study and communication system engineering. It is information, the decrease of uncertainty, in morphological characters that distinguishes operational taxonomic units (OTUs) and reconstructs evolutionary history. We propose that concepts in communication system engineering such as source coding and channel coding correspond in paleontological studies to the construction of diagnostic features and the entire character matrices, which should be distinguished as how typical communication systems are engineered because these two steps serve dual purposes. With character matrices from six different vertebrate groups, we analyzed their information properties including source entropy, mutual information, and channel capacity. Estimation of channel capacity shows upper limits of all matrices in transmitting paleontological information, indicating that, due to the presence of noise, too many characters not only increase the burden in character scoring, but also may decrease quality of matrices. Information entropy, which measure how informative a variable is, of each character is tested as a weighting criterion in parsimony-based systematic studies, the results show high consistence with existing knowledge with both good resolution and interpretability.

Information in morphological characters

Authors

Congyu Yu^{1,2*}, Qigao Jiangzuo^{1,3,4,5}, Emanuel Tschopp^{1,6}, Haibing Wang^{3,5}, Mark Norell¹

1. Division of Paleontology, American Museum of Natural History, New York, NY, USA
2. Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA
3. Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China
4. Key Laboratory of Orogenic Belts and Crustal Evolution, School of Earth and Space Sciences, Peking University, Beijing, China
5. CAS Center for Excellence in Life and Paleoenvironment, Beijing, China
6. University of Hamburg, Center of Natural History

Corresponding author: Congyu Yu, cyu@amnh.org

Abstract

The construction of morphological character matrices is central to paleontological systematic study, which extracts paleontological information from fossils. Although the word information has been repeatedly mentioned in a wide array of paleontological systematic studies, its meaning has rarely been clarified and there has not been a standard to measure paleontological information due to the incompleteness of fossils, difficulty of recognizing homologous and homoplastic structures, etc. Here, based on information theory, we show the deep connections between paleontological systematic study and communication system engineering. It is information, the decrease of uncertainty, in morphological characters that distinguishes operational taxonomic units (OTUs) and reconstructs evolutionary history. We propose that concepts in communication system engineering such as source coding and channel coding correspond in paleontological studies to the construction of diagnostic features and the entire character matrices, which should be distinguished as how typical communication systems are engineered because these two steps serve dual purposes. With character matrices from six different vertebrate groups, we analyzed their information properties including source entropy, mutual information, and channel capacity. Estimation of channel capacity shows upper limits of all matrices in transmitting paleontological information, indicating that, due to the presence of noise, too many characters not only increase the burden in character scoring, but also may decrease quality of matrices. Information entropy, which measure how informative a variable is, of each character is tested as a weighting criterion in parsimony-based systematic studies, the results show high consistence with existing knowledge with both good resolution and interpretability.

Key words

Systematics, information theory, morphology, character weighting

Introduction

Most extinct fossil organisms only preserved their morphology rather than macro biomolecules including DNA and proteins. Therefore, researchers need to transfer the morphology of fossils into sequences, a series of scored morphological characters for example, and analyse such sequences to identify each OTU (classification) and reconstruct their evolutionary history (systematics). However, unlike DNA or protein sequences coded by fixed alphabets (4 nucleotides and 20 amino acids), there is not a universal morphological alphabet of extinct organisms. A practical and the most common way to transfer morphology into sequences is constructing morphological characters matrices. The difficulties in constructing morphological characters have been realized early (Wilkinson 1995), and many early attempts to propose methods/guidance in character construction are far from satisfactory (Estabrook et al., 1975; Hawkins et al., 1997; Sereno 2007). The definition of “character” (in cladistics analysis) has also been discussed a lot (see review by Sereno 2007) but is far from being universally applied .

Besides the most basic question of what a character is, discussions on whether to use giant matrices (Laing et al., 2017) or not (Simoes et al., 2016), which anatomical structures should characters come from (Brocklehurst & Benevento 2020), whether to combine morphological characters with molecular data and shape data (Nylander et al., 2004, Catalano et al., 2010), etc., have been ongoing accessed. Moreover, due to the incompleteness and distortion from preservation environments, most morphological character matrices can only be partially scored. If morphological characters are the most basic units in morphology-based systematic studies, which resemble the nucleotides in DNA sequences and amino acids in proteins, analysing character matrices under the framework of information theory may help to better understand those arguments. .

The word information is repeatedly used in systematic studies (Cracraft 1974, Farris 1979; Mickevich & Platnick 1989; Wilkinson et al., 2004; Sereno 2007; Simoes et al., 2016; Laing et al., 2017) but often it

seems to be confused with data, signal, or its embedded semantic meaning, and few studies have connected information theory with systematic studies, especially for fossil-based ones. Similarly, during the early development of tele-communication system, even after the extensive applications of telegraph, telephone, and broadcast in 1940s, people didn't formulized theories of communication system engineering until information theory was proposed by Shannon (1948). By that time, the transmitted signals, for example the “.” and “-” in Morse Code, and their semantic meaning, for example “we found a dinosaur skull”, are not separated properly. And this ignorance had brought difficulties in improving the quality of communication because no guidance existed to maximize the efficiency of coding information source or to minimize the influence of noises in communication channels.

Paleontological systematic study as a communication system

Shannon (1948) indicated that information is the decrease of uncertainty and the semantic meaning of information is not related to its communication. A typical communication system can be divided into 5 parts, the information source, encoder, channel (with noise), decoder, and the destination (Fig. 1a). Shannon (1948, pp. 379) stated that “*The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point*”. Paleontological systematic studies share abundant similarities with a communication system (Fig. 1b) and focus on reconstructing the evolutionary history of extinct organisms either exactly or approximately. While most communication systems such as telephone, email, and instant messaging apps are for communication in spatial domains, whereas paleontological systematic studies are communication in the temporal domain. The original organisms can be treated as the information source, fossils discovered as the received message, and preservation environments as noisy channels. The encoder in Fig. 1a encodes the original message into signals, for example encoding “we found a dinosaur skull” into Morse Codes, and decoder does the *vice versa*. In paleontology, a widely used encoder is morphological character matrix that encodes each OTU as a series of character states. Most organisms cannot be preserved as fossils, namely, transmitting information through the preservation environment channel, and those fossils discovered are more or less incomplete and distorted. The fundamental problem of paleontological studies is reconstructing at present either exactly or approximately organisms living in another age. Therefore, two questions need to asked that how much information was in an organism or taxon and how much information can be preserved.

To efficiently transmit information, both source coding and channel coding are essential in communication engineering and their differences are listed in table 1. Source coding focuses on minimizing the cost at encoding original messages. For example, Morse Code uses different length of codes to represent different letters, while *E* with the highest frequency in alphabet has the shortest code, a single dot, rarer letters such *X*, *Y*, and *Z* have longer codes, therefore maximize the information entropy, which measures how informative a variable is, of each code. On the other hand, channel coding is designed to resist noises in the preservation environments. The simplest but inefficient example of channel coding is repeated codes. If an information source is randomly sending 0 and 1 via a noisy channel that has a 30% chance to reverse the original message, thus any 0 or 1 received has a 70% chance to be correct. To resist such noise, the encoder repeats each message for three times, which turns “0” into “000” and “1” into “111”, thus under maximum likelihood decoding principle that “000”, “100”, “010”, and “001” are decoded as “0” and others as “1”, the received message has a 78.4% chance to be correct ($0.7^3 + 3 \times 0.7^2 \times 0.3 = 0.784$), which is better than the original encoding method. However, repeated code is usually seen as inefficient because in this example the encoding has tripled the cost but only improves 7.84% accuracy.

Table 1. Comparison between source coding and channel coding

	Source coding	Channel coding
Approximation	Source information entropy	Channel capacity
Redundancy	Discard	Introduce

	Source coding	Channel coding
Purpose	Increasing efficiency	Increasing robustness
Examples	Morse Code	Repeated codes

The joint source-channel coding theorem (Shannon 1948), also known as source-channel separation theorem, shows that source coding and channel coding can be separated without influencing the other. If the channel capacity is strictly greater than source information entropy, noiseless communication can be achieved via sophisticated engineering. In practice, the information encoder is often engineered into decoupled source and channel encoders to serve different purposes as in Table 1.

Similarly, the differences between source coding and channel coding have been realized and practiced in many paleontological studies. From various studies including Nelson (1972) and Cracraft (1974), researchers have shown that the differences between classification (Linnaeus classification and its variants) and systematics (phylogenetic classification, evolutionary classification, evolutionary systematics, etc.). Harrison (1997) emphasized the necessity of separating classification, corresponding to source coding, and systematics, corresponding to channel coding, in paleontological systematic studies. This separation is actually automatically applied in paleontological systematic studies, especially studies reporting new taxa, in which the characterization of the new taxon needs only few diagnostic features, whereas subsequent systematic analysis requires many.

If every character has information entropy of 1 bit, n binary characters can classify 2^n taxa in the ideal situation. Table 2 is an example character matrix including 9 taxa and 3 scored binary characters (0 means absence and 1 means presence of a structure) to illustrate the differences between source coding and channel coding in paleontological systematic studies.

Table 2. Example character matrix showing source coding

Character Taxa	Tail	Feather	Five digits	Scored Sequence	Other characters
1	0	0	0	000	...
2	0	0	1	001	...
3	0	1	0	010	...
4	1	0	0	100	...
5	0	1	1	011	...
6	1	0	1	101	...
7	1	1	0	110	...
7.1	1	1	0	110	...
8	1	1	1	111	...

In the character matrix shown in Table 2, there are 9 taxa and only 3 scored binary characters. The sequences of taxon 7 and 7.1 are the same based the given three characters, hence they cannot be distinguished without other characters being observed. If we would combine taxa 7 and 7.1 as a single OTU, all three characters have information entropy of 1 bit. Although such 3 binary characters are sufficient to distinguish 8 OTUs, they are far from enough to produce a resolved evolutionary cladogram. Usually in practice, the number of characters is much larger than the number of taxa in a character matrix and larger character matrices seems to be a trend in paleontological systematic studies (O’leary et al., 2013; Laing et al., 2017; Baron et al., 2017a). In table 2, the construction of characters is not only insufficient to represent source information entropy, 7 and 7.1 are indistinguishable, but also vulnerable in systematic analysis. There are no redundant scored characters to resist noise or loss of data as the mutual information between each pair of characters is 0 bit if 7 and 7.1 are combined as a single OTU. For example, if the digit parts of the fossils are not preserved, taxa 1&2, taxa 3&5, taxa 4&6, and taxa 7&7.1&8 are indistinguishable because they are equally scored in

the other two characters in the same states. Or for some reason, some fungus fossils in taxon 4 are identified as feather, so it can be confused with taxa 7 and 7.1. From this simplified example we can conclude that, to construct a comprehensive and robust character matrix, the sequences of character states should represent the source information entropy completely, and enough redundancy based on mutual information should be incorporated to minimize the influence of incomplete fossils and misidentification of character states.

A study by Baron et al. (2017a) proposed a significantly different dinosaur phylogeny, in which Theropoda and Ornithischia are sister groups, forming the Ornithoscelida, and Sauropodomorpha and Herrerasauridae form Saurischia as sister group to Ornithoscelida. In a comment to Baron et al. (2017a), Langer et al. (2017) recovered the “traditional” topology of dinosaur phylogeny with dichotomy of Ornithischia and Saurischia including Theropoda, Sauropodomorpha, and “Herrerasauridae”. The subsequent reply by Baron et al. (2017b) mentioned that, “*Langer et al., identify numerous disagreements in terms of character scoring and suggest changing approximately 2,500 scorings, around 10% of the character data*”. Given that there are only tiny differences between methods (Langer et al., 2017 supplementary information), it is clear that the incongruence of original data had led to the contrasting results, but not the algorithm used to reconstruct the phylogeny. Both sides of authors tried to score the vast number of morphological characters in the matrix (“457 anatomical features scored for 74 early dinosaurs and close relatives”) as accurately as possible, but rescoring a single character of a single taxon, *Pisanosaurus mertii*, has led to a considerably different result (Baron et al., 2017b Fig. 1). This vulnerability reflects the fact that this morphological character matrix cannot provide robust results, although the taxon and character numbers in these studies are larger than many previous studies. Comparably, even before Shannon proposed the information theory, communication engineers have designed codes, for example Morse Code, and found factors influencing transmission quality in noisy channel (Nyquist 1924, 1928). But a general problem had been realized that blindly increasing the power of signal cannot improve communication quality after certain threshold in noisy channels.

In typical digital communication systems, all messages are coded in 0 and 1 for transmission. The frequency of the transmitter is defined as how many changes can be made during 1 second with unit Hz. With the increase of frequency, more signals can be sent within a given time span thus more information can be transmitted in an ideal situation. According to the similarity between communication system and paleontological systematic studies discussed before, the concept of frequency in communication systems in spatial domain can be transplanted in paleontological systematic studies in temporal domain as the number of characters, namely bandwidth. Intuitively, if every fossil specimen is complete and undeformed, increasing the number of morphological characters can better describe their morphology, which correlates the trend of using giant matrices currently. However, such positive correlation is challenged under noisy situation, as noises also increase as the increase of bandwidth. Channel capacity, the maximum rate of reliable communication, can be limited by the presence of noises even with arbitrarily large bandwidth.

In this study, we run phylogenetic analyses based on parsimony on character matrices from 6 different vertebrate groups: Ornithischia (Han et al., 2017), Ceratopsia (Yu et al., 2020), Diplodocidae (Tschoop & Mateus 2017), multituberculata (Wang et al., 2019), Carnivoramorpha (Spaulding & Flynn 2012), and lizards (Tschoop et al., 2018). We first quantified the information entropy of each character in six matrices. To access the differences between source coding and channel coding, we then calculated the joint information entropy of first n ($n \leq$ total character number) characters. To investigate the mutuality among characters, the mutual information in each character matrix is calculated. Last, we use the model of additive white gaussian noise (AWGN) discrete channel to estimate the channel capacity of fossil preservation environments.

Material and methods

Information entropy

$$H = - \sum P_i \log_2 (P_i) \quad (1)$$

where P_i represents the possibility of i -th possible value of the source variable, putatively possible states of morphological characters in paleontological studies. For characters with missing data in the character matrices, we estimate the missing parts to have equal distribution among different states. For example,

a binary character is scored 0 in 20% taxa and 1 in 40% taxa, the estimated distribution would be 0 in 20%+20%=40% and 1 in 40%+20%=60%. We also calculate those values without the estimation of missing data as a reference.

For a binary variable P with probability of p and $(1 - p)$, its information entropy is:

$$H = -p \log_2(p) - (1 - p) \log_2(1 - p) \quad (2)$$

, and this relationship between information entropy and probability is illustrated in Fig. 2a

Joint information entropy of the first n characters:

$$H_{\text{accu}}(n) = - \sum S_i \log_2(S_i) \quad (3)$$

, where S_i is the probability of i -th distinct character sequence of the first n characters.

Channel capacity of AWGN discrete channel:

$$C = B \log_2 \left(1 + \frac{S}{N} \right) \quad (4)$$

, where C is the channel capacity, B is the bandwidth (number of characters), S and N are the power of signal (scored characters) and noise (unscored characters), respectively. Character matrices are from published studies (Spaulding & Flynn 2012; Tschopp & Mateus 2017; Han et al., 2018; Tschopp et al. 2018; Wang et al., 2019; Yu et al., 2020). Calculation is done by custom Python 3.7 scripts.

Phylogenetic analysis was done in TNT 1.5 using traditional search (Goloboff & Catalano 2016). The strict consensus tree was appended to the last of tree list in each method. CI and RI are only calculated for the strict consensus trees.

Results

The results of character information entropy (Fig. 2b) show that characters with more states tend to have higher information entropy, indicating those multi-state characters generally introduce more information in systematic studies. Among characters with the same number of states, the information entropy still varies a lot in most datasets.

Six matrices show consistent pattern in their joint information entropy (Fig. 2c, only first 40 characters are shown). For the first few characters, the joint information entropy increases fast to approximate the source information entropy, which is the upper limit of joint information entropy, and the majority of characters serve in channel coding as they do not contribute to the source coding much. The curves of joint information entropy shows that only a few characters are required to distinguish each OTU (classification) and the majority of characters in the matrices are for channel coding (systematics).

The mutual information across 6 matrices is also calculated (Fig. 2d) to test the mutuality between characters. Due to the existence of missing data, the diagonal line numbers showing mutuality between any character and itself are not strictly its information entropy but are still generally higher than other areas of the heatmaps. The distribution of mutuality seems to have no pattern in most matrices. After reorder and partition characters by anatomical structures (crania, pectoral girdle and forelimb, pelvic girdle and hindlimb, axial bones, and others), some parts exhibit relatively high mutuality, for example the forelimbs and hindlimbs of Carnivoramorpha (Spaulding & Flynn 2012) show both higher inter- and intra-mutuality than other anatomical structures.

The distributions of noise power in taxa domain and character domain are shown in Figure 3a and 3b, respectively. The results show saturation in channel capacity when increasing bandwidth, the number of characters (Fig. 3c). Different character matrices reach the maximum channel capacity when having 62.5% (multituberculata) to 89.7% (Diplodocidae) of total characters.

Discussion

Information source

No matter what algorithm is being used in systematics studies, the common aspect is using sequences (DNA, amino acids, and morphological characters) to characterize organisms and to interpret their evolutionary history. With fixed alphabets, DNA and protein sequences resemble digit signals in modern communication systems, while morphology of fossils is more like analog signals. Therefore, the process of character construction is the same as sampling digit signals from analog signals, meanwhile, the probably infinite original information entropy of fossil morphology is transferred into finite entropy, represented by hundreds to thousands of morphological characters, that can be compared. More morphological characters usually describe organisms more completely, but it is extremely difficult to measure how complete the character matrix characterizes the overall morphology of a group of organisms. There is not a standard guidance on character selection and many characters in matrices are selected because researchers believe they carry morphological information. The interrelationship among morphological characters and how they connect to the overall morphology remains uncertain. At least from the results of mutual information and channel capacity against bandwidth, the number of characters, we show that the dependence between characters and different anatomical structures is complex, and current morphological character matrices seems to encounter the saturation of characters already. Shannon (1949) proposed the Sampling Theorem (also known as Nyquist-Shannon Sampling Theorem), which bridges the continuous signals and discrete signals. With a continuous signal source of a finite bandwidth, Sampling Theorem shows the lowest sample rate to capture all information, which is twice the rate of highest rate of original signals. As the connection between bandwidth in typical communication systems and character number of paleontological systematic studies is discussed before, Sampling Theorem may be a bridge between raw morphology and morphological characters.

However, the saturation of channel capacity (Fig. 3c) does not necessarily mean those morphological character matrices fully represent the entire morphology of fossil specimens, but cannot sufficiently convey the sampled morphological information in matrices while some other information may be left as the sampling of characters are strongly biased. The morphological matrix of multituberculata comprise only characters from the cranial region, but the postcrania of those organisms also have information.

With the wide application of advanced imaging techniques such as CT (computed tomography) scan, it is feasible to capture the complete morphology of fossil specimens without destruction. The unprecedented amount of data may be the stepstone to establish the connection between analog morphological data and digital character data. A standard workflow may be possible to morphological studies under the facilitation from information theory and high-resolution imaging.

The properties of channel (bandwidth, channel capacity, noise)

In this study we use one of the most basic models, AWGN channel, to mimic preservation environments with limited explanation. AWGN channel requires that noises have uniform power in frequency domain and gaussian distribution in time domain. In this study, we treat the character number as bandwidth, then the character probably corresponds to the frequency domain in typical communication system and OTUs to the time domain. This model sounds natural based on the model in Fig. 1 as every organism ever lived on earth was a signal sent, and fossils are a small fraction received. However, in character matrices analyzed here, many OTUs are scored based on multiple specimens, therefore result in the aggregation of scored characters in the first few columns in Fig. 3a. For the time domain/OTUs, the noises derived from natural preservation and are controlled by many factors, so it is probably fair to use AWGN channel model for both simplification and convenience.

From the estimation based on AWGN channel model (Fig. 3c), all character matrices show saturation of characters. The basic explanation of saturation is that with the increase of bandwidth, the noise also increases. Incompleteness, deformation, and misidentification are common among the fossil specimens. If the nature of the paleontological information channel is noisy, we cannot expect to efficiently transmit

paleontological information without channel coding. Moreover, the time costs in both encoding and decoding have to be considered when facing extremely giant character matrices.

Character matrix construction and weighting

The construction of (morphological) character matrices is central to systematic studies and has been discussed extensively. In this study, we make the initial attempt to quantify the information in existing morphological character matrices for the first time. Many results show consistence to common understanding of morphological characters, including different characters having different amount of information, mutuality existing among characters, more characters usually carrying more information, etc. Besides, we also propose that the information entropy of each character can be used as their weights in phylogenetic analysis.

As the information entropy represents how informative a character is, it may be a candidate of character weighting in phylogenetic analysis. Most researchers agree that some kind of weighting should be applied in systematic analysis and equal weighting is one of the weighting methods (Farris 1969, Sereno 2007). Based on the successive weighting proposed by Farris (1969), Goloboff (1993) proposed implied weighting and extended implied weighting (Goloboff 2014). These weighting methods refine the weights of different characters to reduce homoplasy. However, Congreve & Lamsdell (2016) indicated that implied weighting is not consistent with the idea of parsimony and increase both correctly and incorrectly resolved nodes with simulated datasets. The wide use suggest that implied weighting and its variants probably provide a direction in reconstructing better resolved trees, but neither the theoretical basis nor its utilization answer the core question of how much information is in each character and may fail when working with character matrices with too many homoplastic characters.

Birds and modern mammals are both endothermic, covered with filaments rather than scales, have four-chamber hearts, etc. If we would deliberately sample too many characters describing these features, the conclusion could easily be forced into that birds are mammals, and many synapomorphies between birds and other reptiles, for example the presence of sclerotic rings, can be recovered as homoplasy. Fortunately, there are many other lines of evidence, which mean more information, showing that birds are more closely related to modern reptiles than modern mammals. The morphology and physiology of birds, the genetic data, and the fossil records all indicate that these similar features between birds and mammal are results of convergent evolution. It is not reasonable to refute that birds are dinosaurs with considerable fewer features against the overwhelming evidence from fossils, molecular biology, anatomy, and many other aspects. However, such biased sampling of character can be hard to be realized for extinct groups with only limited fossil materials and implied weighting may even strengthen such bias. But information theory may discover those biased sampling. If such a character matrix exists, since its biased sampling, the mutual information among characters would be high and the channel capacity may not be saturated by the number of characters, because there is only little information represented by biased sampled characters.

Successive weighting, implied weighting and their variants require an initial weight or an existing tree topology, whereas information entropy weighting only depends on the information entropy in each character. In character construction, the fact is that the selections of characters are extremely biased as most morphological character come from cranial area in paleontological vertebrate studies (Fig. 2d). In the six datasets we analyzed here, the proportion of cranial characters are from 40.7% to 100% with an average of 63.2%, which immediately shows that some parts have more morphological information (or “more important”) than others in systematic studies. Practically, any multi-state characters can be split into several binary characters, and character matrices examined in this study all have multi-state characters. If using equal weighting, the weights are different for multi-state characters and binary characters. But based on information theory, since $H(A + B) = H(A) + H(B) - I(A, B)$, multi-state characters can be accurately split into independent binary characters (with 0 mutual information) without losing or adding any information.

Intuitively, using information entropy as a measure of character weight also conforms with our understanding of what a character is. A character can be treated as a random variable and its states are possible values. For simplification, we only discuss discrete memoryless information source here, which has discrete signals

and previous signals do not influence later ones. If a character is coded the same across the entire group of organisms, $P(\text{Character } A = 0) = 1$, then it has an information entropy of 0, therefore should be weighted as 0, namely excluded. The character with the highest information entropy has the most equal distribution of character states (Fig. 2a). Those characters with very unequal distributions, for example only a few OTUs are scored as 0 and the majority as 1, should be down weighted because they only contribute little to the source coding by distinguishing very few OTUs, and easily influenced by environmental noises. Information entropy weighting gives characters that contribute more to the source information entropy higher weight. Kälersjö et al. (1999) studied plant nucleotides data and their results showed that fast evolving and highly homoplastic third codon positions, opposite to traditional thought, have the unexpectedly strongest phylogenetic information, and they also suggest that the frequency of change should be used as in character weighting and selection. Although these authors tried to quantify the information in different nucleotide sites, i.e., molecular characters, they did not provide an explanation on how they define information/informative sites.

We tested the results from equal weighting, implied weighting ($k = 3 \& 12$), and information entropy weighting of six matrices analysed before. Ceratopsia are illustrated in Figure 4. To save space and show the differences among trees, colored columns replace the OTU names on the right side of trees and color gradients correspond to the taxa order in character matrix. Detail phylogenetic results of six groups are provided in Dryad doi:10.5061/dryad.8sf7m0cnc. Generally, they show unexpected consistence between both equal weighting and implied weighting, but slight differences are common. The CI (consistence index) and RI (retention index) are also calculated for the strict consensus tree of each group in table 3. The CI of entropy weighting is generally slightly lower than other methods and RI is slightly higher, suggesting that more homologous characters are suggested and the trees fits better for entropy weighted characters.

Table 3. CI and RI of different morphological character matrices

Index	Weighting	Ornithischia	Ceratopsia	Diplodocidae	multituberculata	Carnivoramorph	lizards
CI	Equal	0.369	0.518	0.345	0.327	0.261	0.246
	Implied ($k=3$)	0.352	0.502	0.344	0.44	0.267	0.241
	Implied ($k=12$)	0.352	0.511	0.345	0.446	0.270	0.246
	Entropy	0.343	0.498	0.331	0.433	0.262	0.236
RI	Equal	0.71	0.844	0.526	0.795	0.578	0.468
	Implied ($k=3$)	0.688	0.833	0.525	0.748	0.590	0.452
	Implied ($k=12$)	0.688	0.839	0.525	0.754	0.596	0.467
	Entropy	0.703	0.846	0.530	0.761	0.61	0.478

Figures

1. **a** . typical communication system modified from Shannon (1948); **b** . paleontological systematic studies in abstract.
2. **a** . information entropy distribution of a binary variable; **b** . information entropy of characters in different morphological character matrices, x-axis: number of character states, y-axis: average information entropy; **c** . joint information entropy in different morphological character matrices; **d** . mutual information distribution heatmap in different morphological character matrices, orange: high mutual information, blue: low mutual information, anatomical parts abbreviations, C: crania, F: forelimb and pectoral girdle, H: hindlimb and pelvic girdle, A: axial elements, O: others.
3. **a** . noise power distribution in taxa domain; **b** . noise power distribution in character domain; **c** . channel capacity and bandwidth in character matrices.
4. Tree results of Ceratopsia. **a** . Equal weighting; **b** . implied weighting ($k=3$); **c** . implied weighting ($k=12$); **d** . information entropy weighting. Colored columns on the right side of trees represent

OTUs and their color gradients correspond to the taxa order in the original character matrix. Data Accessibility -Phylogenetic results of Ornithischia, Ceratopsia, Diplodocidae, multituberculate, Carnivoramorphia, and lizards, by equal weighting, implied weighting ($k = 3&12$), and information entropy weighting are presented in .nex files. The strict consensus tree is appended in the end of trees in each file. Dryad doi:10.5061/dryad.8sf7m0cnc

Conflict of interest

All authors declare that they have no conflicts of interest.

Acknowledgments

We thank Jin Meng at the American Museum of Natural History, New York, USA and Xing Xu at the Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China for their comments on this study.

References

1. Baron, M. G., Norman, D. B. and Barrett, P. M. (2017a) ‘A new hypothesis of dinosaur relationships and early dinosaur evolution’, *Nature* , 543(7646), pp. 501–506. doi: 10.1038/nature21700.
2. Baron, M. G., Norman, D. B. and Barrett, P. M. (2017b) ‘Baron et al. reply’, *Nature* , 551(7678), pp. E4–E5. doi: 10.1038/nature24012.
3. Catalano, S. A., Goloboff, P. A. and Giannini, N. P. (2010) ‘Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework’, *Cladistics* , 26(5), pp. 539–549. doi: <https://doi.org/10.1111/j.1096-0031.2010.00302.x>.
4. Congreve, C. R. and Lamsdell, J. C. (2016) ‘Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices’, *Palaeontology* , 59(3), pp. 447–462. doi: <https://doi.org/10.1111/pala.12236>.
5. Cracraft, J. (1974) ‘Phylogenetic Models and Classification’, *Systematic Biology* , 23(1), pp. 71–90. doi: 10.1093/sysbio/23.1.71.
6. Crawford, N. G. *et al.* (2012) ‘More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs’, *Biology Letters* , 8(5), pp. 783–786. doi: 10.1098/rsbl.2012.0331.
7. Estabrook, G. F., Johnson, C. S. and Mc Morris, F. R. (1975) ‘An idealized concept of the true cladistic character’, *Mathematical Biosciences* , 23(3), pp. 263–272. doi: 10.1016/0025-5564(75)90040-1.
8. Farris, J. S. (1969) ‘A Successive Approximations Approach to Character Weighting’, *Systematic Biology* , 18(4), pp. 374–385. doi: 10.2307/2412182.
9. Farris, J. S. (1979) ‘The Information Content of the Phylogenetic System’, *Systematic Biology* , 28(4), pp. 483–519. doi: 10.2307/sysbio/28.4.483.
10. Goloboff, P. A. (1993) ‘Estimating Character Weights During Tree Search’, *Cladistics* , 9(1), pp. 83–91. doi: <https://doi.org/10.1111/j.1096-0031.1993.tb00209.x>.
11. Goloboff, P. A. (2014) ‘Extended implied weighting’, *Cladistics* , 30(3), pp. 260–272. doi: <https://doi.org/10.1111/cla.12047>.
12. Goloboff, P. A. and Catalano, S. A. (2016) ‘TNT version 1.5, including a full implementation of phylogenetic morphometrics’, *Cladistics* , 32(3), pp. 221–238. doi: <https://doi.org/10.1111/cla.12160>.
13. Han, F. *et al.* (2017) ‘Postcranial anatomy of Yinlong downsi (Dinosauria: Ceratopsia) from the Upper Jurassic Shishugou Formation of China and the phylogeny of basal ornithischians’, *Journal of Systematic Palaeontology* , 0(0), pp. 1–29. doi: 10.1080/14772019.2017.1369185.
14. Harrison, T. (1993) ‘Cladistic Concepts and the Species Problem in Hominoid Evolution’, in Kimbel, W. H. and Martin, L. B. (eds) *Species, Species Concepts and Primate Evolution* . Boston, MA: Springer US, pp. 345–371. doi: 10.1007/978-1-4899-3745-2_14.

15. Hawkins, J. A., Hughes, C. E. and Scotland, R. W. (1997) 'Primary Homology Assessment, Characters and Character States', *Cladistics* , 13(3), pp. 275–283. doi: 10.1111/j.1096-0031.1997.tb00320.x.
16. Källersjö, M., Albert, V. A. and Farris, J. S. (1999) 'Homoplasy Increases Phylogenetic Structure', *Cladistics* , 15(1), pp. 91–93. doi: <https://doi.org/10.1111/j.1096-0031.1999.tb00400.x>.
17. Kluge, A. G. and Farris, J. S. (1969) 'Quantitative Phyletics and the Evolution of Anurans', *Systematic Biology* , 18(1), pp. 1–32. doi: 10.1093/sysbio/18.1.1.
18. Laing, A. M. *et al.* (2018) 'Giant taxon-character matrices: the future of morphological systematics', *Cladistics* , 34(3), pp. 333–335. doi: 10.1111/cla.12197.
19. Langer, M. C. *et al.* (2017) 'Untangling the dinosaur family tree', *Nature* , 551(7678), pp. E1–E3. doi: 10.1038/nature24011.
20. Luo, Z.-X. *et al.* (2015) 'Mandibular and dental characteristics of Late Triassic mammaliaform Haramiyavia and their ramifications for basal mammal evolution', *Proceedings of the National Academy of Sciences* , 112(51), pp. E7101–E7109. doi: 10.1073/pnas.1519387112.
21. Nelson, G. J. (1972) 'Phylogenetic Relationship and Classification', *Systematic Zoology* , 21(2), pp. 227–231. doi: 10.2307/2412293.
22. O'Leary, M. A. *et al.* (2013) 'The Placental Mammal Ancestor and the Post-K-Pg Radiation of Placentals', *Science* , 339(6120), pp. 662–667. doi: 10.1126/science.1229237.
23. Sereno, P. C. (2007) 'Logical basis for morphological characters in phylogenetics', *Cladistics* , 23(6), pp. 565–587. doi: <https://doi.org/10.1111/j.1096-0031.2007.00161.x>.
24. Shannon, C. E. (1948) 'A Mathematical Theory of Communication', *The Bell System Technical Journal* , 27, pp. 379–423.
25. Shannon, C. E. (1949) 'Communication in the Presence of Noise', *Proceedings of the IRE* , 37(1), pp. 10–21. doi: 10.1109/JRPROC.1949.232969.
26. Simões, T. R. *et al.* (2017) 'Giant taxon-character matrices: quality of character constructions remains critical regardless of size', *Cladistics* , 33(2), pp. 198–219. doi: 10.1111/cla.12163.
27. Sokal, R. R. (1966) 'Numerical Taxonomy', *Scientific American* , 215(6), pp. 106–117.
28. Spaulding, M. and Flynn, J. J. (2012) 'Phylogeny of the Carnivoramorpha: The impact of postcranial characters', *Journal of Systematic Palaeontology* , 10(4), pp. 653–677. doi: 10.1080/14772019.2011.630681.
29. Wang, H., Meng, J. and Wang, Y. (2019) 'Cretaceous fossil reveals a new pattern in mammalian middle ear evolution', *Nature* , pp. 1–4. doi: 10.1038/s41586-019-1792-0.
30. Wilkinson, M. (1995) 'A comparison of two methods of character construction', *Cladistics* , 11(3), pp. 297–308. doi: 10.1016/0748-3007(95)90017-9.
31. Wilkinson, M., Cotton, J. A. and Thorley, J. L. (2004) 'The Information Content of Trees and Their Matrix Representations', *Systematic Biology* , 53(6), pp. 989–1001. doi: 10.1080/10635150490522737.
32. Yu, C. *et al.* (2020) 'A neoceratopsian dinosaur from the early Cretaceous of Mongolia and the early evolution of ceratopsia', *Communications Biology* , 3(1), pp. 1–8. doi: 10.1038/s42003-020-01222-7.



