


## ORIGINAL RESEARCH

# Information in morphological characters

Congyu Yu<sup>1,2</sup>  | Qigao Jiangzuo<sup>1,3,4,5</sup> | Emanuel Tschopp<sup>1,6</sup> | Haibing Wang<sup>3,5</sup> | Mark Norell<sup>1</sup>

<sup>1</sup>Division of Paleontology, American Museum of Natural History, New York, NY, USA

<sup>2</sup>Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA

<sup>3</sup>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

<sup>4</sup>Key Laboratory of Orogenic Belts and Crustal Evolution, School of Earth and Space Sciences, Peking University, Beijing, China

<sup>5</sup>CAS Center for Excellence in Life and Paleoenvironment, Beijing, China

<sup>6</sup>Center of Natural History, University of Hamburg, Hamburg, Germany

**Correspondence**

Congyu Yu, Division of Paleontology, American Museum of Natural History, New York, NY, USA.

Email: 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 nor specifically defined. It is important, however, to establish a standard to measure paleontological information because fossils are hardly complete, rendering the recognition of homologous and homoplastic structures difficult. Here, based on information theory, we show the deep connections between paleontological systematic study and communication system engineering. Information is defined as the decrease of uncertainty and it is the information in morphological characters that allows distinguishing operational taxonomic units (OTUs) and reconstructing evolutionary history. We propose that concepts in communication system engineering such as source coding and channel coding, correspond to the construction of diagnostic features and the entire character matrices in paleontological studies. The two coding strategies should be distinguished following typical communication system engineering, because they 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 character saturation of all matrices in transmitting paleontological information, indicating that, due to the presence of noise, oversampling characters not only increases the burden in character scoring, but also may decrease quality of matrices. We further test the use of information entropy, which measures how informative a variable is, as a character weighting criterion in parsimony-based systematic studies. The results show high consistency with existing knowledge with both good resolution and interpretability.

**KEYWORDS**

character weighting, information theory, morphology, systematics

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Most extinct fossil organisms only preserved their morphology rather than macro-biomolecules including DNA and proteins. Therefore, researchers need to convert the morphology of fossils into sequences, a series of scored morphological characters, for example, and analyze such sequences to identify each OTU (operational taxonomic unit, 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 that can digitize the morphology of extinct organisms into sequences. A practical and probably the most common way to convert morphology into sequences is constructing morphological characters matrices, which contain various OTUs and characters. According to the morphology of different OTUs, they are scored different states, usually 0 and 1, for different characters. 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 have been ongoing on whether to use giant matrices (Laing et al., 2018) or not (Simões et al., 2017), which anatomical structures should be represented by characters (Brocklehurst & Benevento, 2020), whether to combine morphological characters with molecular data and shape data (Nylander et al., 2004; Catalano et al., 2010), etc. Moreover, due to the incompleteness and distortions 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, analyzing character matrices under the framework of information theory may help to better understand those arguments.

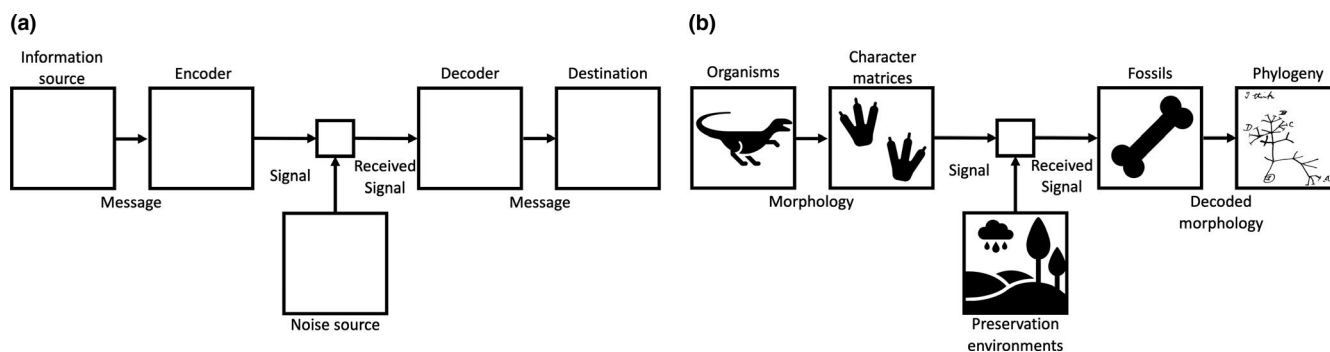
The word information is repeatedly used in systematic studies (Cracraft, 1974; Farris, 1979; Mickevich & Platnick, 1989; Wilkinson et al., 2004; Sereno, 2007; Simões et al., 2017; Laing et al., 2018) but often it seems to be confused with data, signal, or its embedded

semantic meaning. Few studies have connected information theory with systematic studies, especially for fossil-based analyses. Similarly, during the early development of tele-communication systems, even after the extensive applications of telegraph, telephone, and broadcast in 1940s, people did not formulate a complete theory of communication system engineering until information theory was proposed by Shannon (1948). Before constructing any communication system, it should be noticed that the signals themselves are irrelevant to their semantic meaning. Imagine a paleontologist and a local guide are working in a remote fossil locality, the guide stays in the camp and the paleontologist is looking for fossils. The paleontologist finds a dinosaur skeleton and needs tools from the camp to dig it out, but it takes too long to walk back. The paleontologist wants the local guide to bring with those tools. A smart way to do so may be to make an agreement with the local guide before leaving the camp as following: raising a red flag means the paleontologist needs fossil digging tools; raising a blue flag means he needs food. With such agreement, the paleontologist and local guide are communicating fairly efficiently, the only concern would be whether the local guide can see the color of the flag, but not the meaning of the color itself.

The mixture of signals and their semantic meaning can cause serious problems in communication, because exactly the same signal may have totally different meanings. 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.

## 2 | PALEONTOLOGICAL SYSTEMATIC STUDY AS A COMMUNICATION SYSTEM

Shannon (1948) indicated that information is the decrease of uncertainty and a typical communication system can be divided into 5 parts, the information source, encoder, channel (which usually introduces noise), decoder, and the destination (Figure 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 (Figure 1b)



**FIGURE 1** (a) Typical communication system modified from Shannon (1948); (b) Paleontological systematic studies in abstract

and focus on reconstructing the evolutionary history of extinct organisms. Most modern communication systems such as telephone, email, and instant messaging apps are for communication in spatial domains, whereas paleontological systematic studies represent communication in the temporal domain. The original organisms can be treated as the information source, fossils discovered as the received message, and the preservation environments as noisy channels. Although some signals are either lost or distorted during preservation, we are interested in how much information is preserved and whether or how we can reconstruct those lost or distorted signals based on known ones. The encoder in Figure 1a encodes the original messages into signals, for example encoding “I need fossil digging tools” into a red flag as the example before, and the decoder does the *vice versa*. In paleontology, a widely used encoder is the morphological character matrix that encodes each OTU as a sequence of character states. The fundamental problem of paleontological studies is reconstructing at present either exactly or approximately organisms living in another age. Two questions must be answered to do so: (a) that how much information was in an organism or taxon? and (b) how much information can be preserved?

For the first question, vast publications have described the morphology of fossils in fine detail, but such description cannot be directly applied in paleontological systematic studies. As the morphological character matrix is probably the most common encoder in paleontology, the number of characters in it determines how many signals can be encoded, which draws the upper limit of how much information can be transmitted.

For the second question, although most fossils themselves are not well-preserved, the preserved elements may help to reconstruct the missing parts. Many dinosaur species are named based on limited fossils, for example, *Deinocoelurus* (Dinosauria, Theropoda) was firstly collected in 1965 (Kielan-Jaworowska and Dovchin, 1966), when only forelimbs and several other fragments were discovered, then Ostrom (1972) recognized its affinity with ornithomimosaurs, which was later supported by Senter (2007), and finally the discovery of almost entire skeletons ended most arguments by Lee et al. (2014).

In communication system engineering, such processes are named source coding and channel coding; their differences are listed in Table 1. Source coding focuses on minimizing the cost at encoding all original messages. For example, Morse Code uses different lengths of codes to represent each letter in the alphabet, minimizing coding costs is attained by attributing the shortest code (a single dot) to the letter with the highest frequency in English (the letter E), whereas rarer letters such as X, Y, and Z have longer codes. 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 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 inefficient because in this example the encoding has tripled the cost but accuracy only improves 8.4%.

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, even in a noisy channel. 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 systematic studies. In various studies including Nelson (1972) and Cracraft (1974), researchers have shown the differences between classification (Linnaeus classification and its variants) and systematics (phylogenetic classification, evolutionary classification, evolutionary systematics, etc.). Harrison (1993) 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.

Morphological characters usually have two states which can be coded as 0 and 1. Although sometimes more states are available, multi-state characters can always be split into several binary ones. A morphological character can be treated as a variable with discrete distribution on a group of organisms and we are interested in how much information is in a morphological character. Information is defined as the decrease of uncertainty (Shannon, 1948). If a character is scored the same among OTUs in a group, the information given by this character should be 0 because it does not decrease any uncertainty. The information given by a variable is its information entropy and can be calculated as follows:

**TABLE 1** Comparison between source coding and channel coding

	Source coding	Channel coding
Approximation	Source information entropy	Channel capacity
Redundancy	Discard	Introduce
Purpose	Increasing efficiency	Increasing robustness
Examples	Morse Code	Repeated codes

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

where  $H$  is the information entropy (measures in bit if the base of logarithm function is 2) and  $p_i$  represents the probability of  $i$ -th possible value of the source variable, putatively possible states of morphological characters in paleontological studies. With the change of distribution of character states, the information given by a character also changes and reaches its maximum at equal distribution. For a binary variable  $P$  with its values in probability of  $p$  and  $(1 - p)$ , its information entropy is as follows:

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

and this relationship between information entropy and probability is illustrated in Figure 2a. Consider the premaxillary teeth in Ceratopsia (Dinosauria, Ornithischia) as a morphological character, which is scored as 0 if not present and 1 if present. There are about 80 species of ceratopsians reported and 12 of them have premaxillary teeth, the information entropy of the premaxillary teeth character in Ceratopsia is 0.61 bit, which means an observation of this character in a ceratopsian species provides 0.61 bit information in average. However, it should be noticed that the distribution of character states is subject to change with new fossils, which will change its information entropy. With the increase of OTUs, the influence from newly added OTU decreases. If every binary character has an information entropy of 1 bit, namely, character states have equal distribution,  $n$  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.

In the character matrix shown in Table 2, there are 9 taxa and only 3 scored binary characters. The sequences of taxa 7 and 7.1 are the same based on 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 these 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., 2018; 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 associated with taxon 4 are erroneously identified as feathers, 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.

Since there are many characters in a character matrix, we are interested in their mutuality that strongly influences the quality of character matrix. If two characters are strongly dependent, we can infer the state of a missing character according to observed state of the other, which may provide insight in dealing with incomplete specimens and dividing modules in mosaic evolution studies. In previous content, we show that the information of a variable is defined as the uncertainty it decreases, and thus, the uncertainty of a variable  $A$  given by another variable  $B$  is the mutuality between them, the mutual information.

$$I(A, B) = H(A) - H_B(A) = H(B) - H_A(B) = H(A) + H(B) - H(A, B) \quad (3)$$

where  $I(A, B)$  is the mutual information,  $H(A)$ ,  $H(B)$ , and  $H(A, B)$  are the information entropy of variable  $A$ ,  $B$ , and their joint distribution, respectively,  $H_B(A)$  and  $H_A(B)$  are the conditional entropy of these two variables. If the marginal and joint distributions of variable  $A$  and  $B$  are known:

$$I(A, B) = \sum_a \sum_b P(a, b) \log_2 \frac{P(a, b)}{P(a)P(b)} \quad (4)$$

In Table 2, we can calculate the mutual information between each pair of characters and the results are 0 for all pairs (taxa 7 and 7.1 are treated as a single OTU). The zero mutual information in this designed character matrix indicates that the tail, feather, and five digits characters are independent from one another, namely, the knowledge of a character in a taxon does not decrease the uncertainty from another character. The lack of dependence also explains the vulnerability of diagnosis when a character cannot be observed.

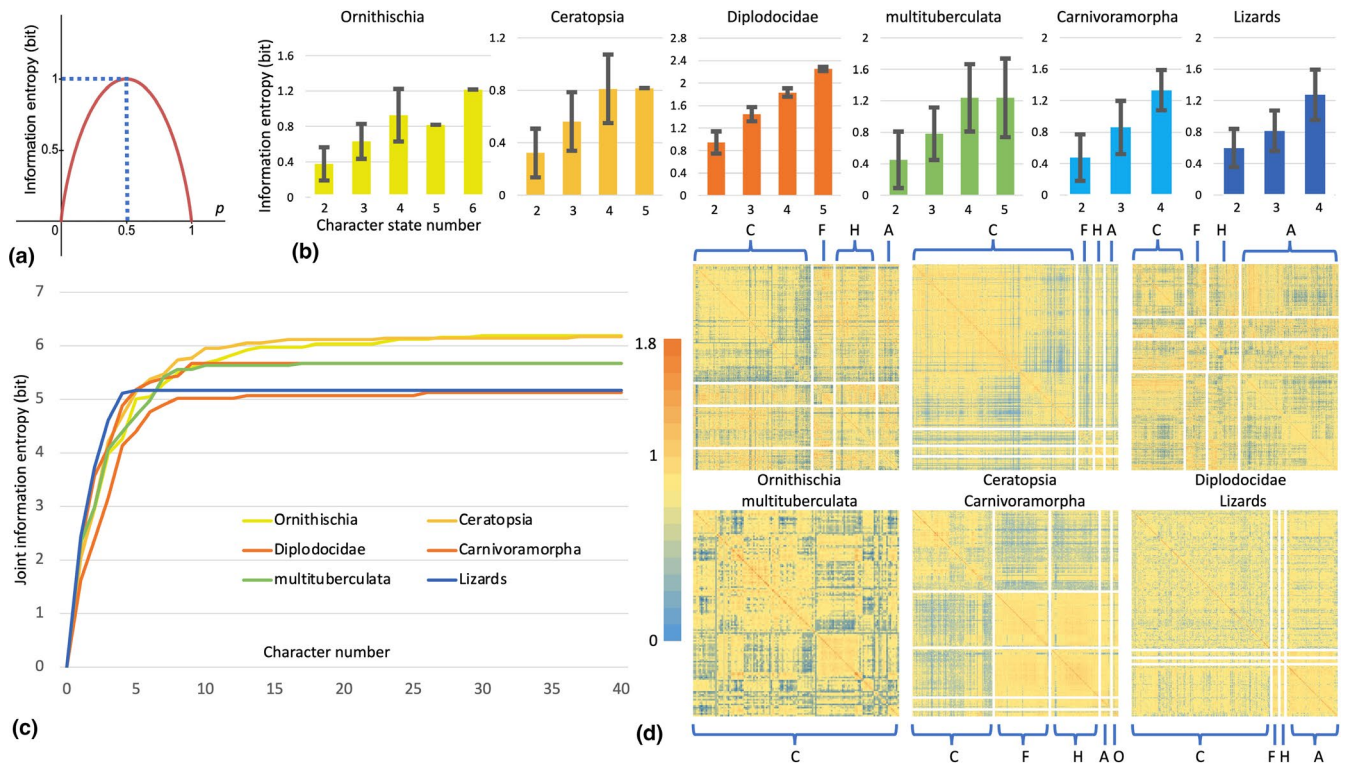
Mutual information and joint information entropy can be further generalized to multiple variables. We can calculate the joint information entropy of the entire character matrix according to the joint distribution of character states as following:

$$H(A_1, A_2, \dots, A_n) = - \sum_{a_1} \dots \sum_{a_n} P(a_1, a_2, \dots, a_n) \log_2 P(a_1, a_2, \dots, a_n) \quad (5)$$

where  $A_j$  represents one of the multiple variables. To simplify expression (5), we use  $s_{ij}$  to represent the  $j$ -th unique sequence of first  $i$  characters in the matrix. For example, in the character matrix given by Table 2,  $s_{21} = 00$  and  $s_{33} = 010$ .

$$H(\text{first } i \text{ characters in matrix}) = - \sum_j P(s_j) \log_2 P(s_j) \quad (6)$$

A study by Baron et al. (2017a) proposed a significantly different dinosaur phylogeny, in which Theropoda and Ornithischia are sister



**FIGURE 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, only first 40 characters are shown to save space; (d) mutual information distribution heat map 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

**TABLE 2** Example character matrix showing source coding

Taxa	Character				Other characters
	Tail	Feather	Five digits	Scored sequence	
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	...

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 a dichotomy of Ornithischia and Saurischia, the latter 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 character scoring 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, Figure 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 channels (Nyquist 1924, 1928). A general problem had been realized that blindly increasing the power of signal cannot improve communication quality after certain thresholds 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 s 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 ideal situation. According to the similarity between communication system and paleontological systematic studies discussed before, the concept of frequency in communication systems in the spatial domain can be transcribed in paleontological systematic studies in the temporal domain as the number of characters, namely, bandwidth. Intuitively, if every fossil specimen is complete and undeformed (noiseless channel), increasing the number of morphological characters can describe their morphology in finer detail, which correlates the trend of using giant matrices currently. However, such positive correlation only partially holds under noisy circumstances, as noises also increase with the increase of bandwidth. Channel capacity, the maximum rate of reliable communication, is limited by the presence of noises even with arbitrarily large bandwidth. Shannon (1948) shows the relation between channel capacity and bandwidth of Additive White Gaussian Noise (AWGN) channel:

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (7)$$

where  $C$  is the channel capacity (bit),  $B$  is the bandwidth (range of available frequency in wireless communication and number of characters in this study),  $S$  and  $N$  are the power of signal (scored characters) and noise (unscored characters), respectively. AWGN channel assumes that noises are (a) additive to the system but not intrinsic; (b) white, uniform power across frequency; (c) Gaussian, in normal distribution in the time domain. AWGN is a basic model simplifying many natural random processes as a whole. Because the unscored character states in a matrix are controlled by many complicated factors and the extreme difficulty of modeling most non-Gaussian systems, AWGN channel offers convincing results at this stage. It should be noticed that  $N = n_0 B$ , where  $n_0$  is the noise density per bandwidth unit, and thus, the maximum channel capacity is approximate  $1.44B \frac{S}{N} = 1.44 \frac{S}{n_0}$  bits in AWGN channel.

### 3 | MATERIAL AND METHODS

In this study, we calculated the information properties and run parsimony-based phylogenetic analyses on character matrices from 6 different vertebrate groups: Ornithischia (Han et al., 2017), Ceratopsia (Yu et al., 2020), Diplodocidae (Tschopp & Mateus, 2017), multituberculata (Wang et al., 2019), Carnivoramorpha (Spaulding &

Flynn, 2012), and lizards (Tschopp et al., 2018). We first quantified the information entropy of each character in six matrices. To investigate the mutuality among characters, the mutual information in each character matrix is calculated. 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. Last, we use the model of AWGN discrete channel to estimate the channel capacity of fossil preservation environments.

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% OTUs, 1 in 40% OTUs, and missing in 40% OTUs, the estimated distribution would be 0 in  $20\% + 0.5 \times 40\% = 40\%$  and 1 in  $40\% + 0.5 \times 40\% = 60\%$ . We also calculate those values without the estimation of missing data as a reference. Calculation is done by custom Python 3.7 scripts.

Phylogenetic analysis was done in TNT 1.5 using traditional search with default settings (Goloboff & Catalano, 2016), in which implied weighting used  $k = 3$  and 12. The strict consensus tree was appended to the last of tree list in each group. CI and RI are discussed for the strict consensus trees.

## 4 | RESULTS

The distribution of character information entropy (Figure 2b) shows 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 (Figure 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 show 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 within the 6 matrices is calculated (Figure 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 do not strictly correspond to the character's information entropy but are still generally higher than other areas of the heat maps. The distribution of mutuality seems to have no pattern in most matrices. After reordering and partitioning 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 intramutuality than other anatomical structures.

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

## 5 | DISCUSSION

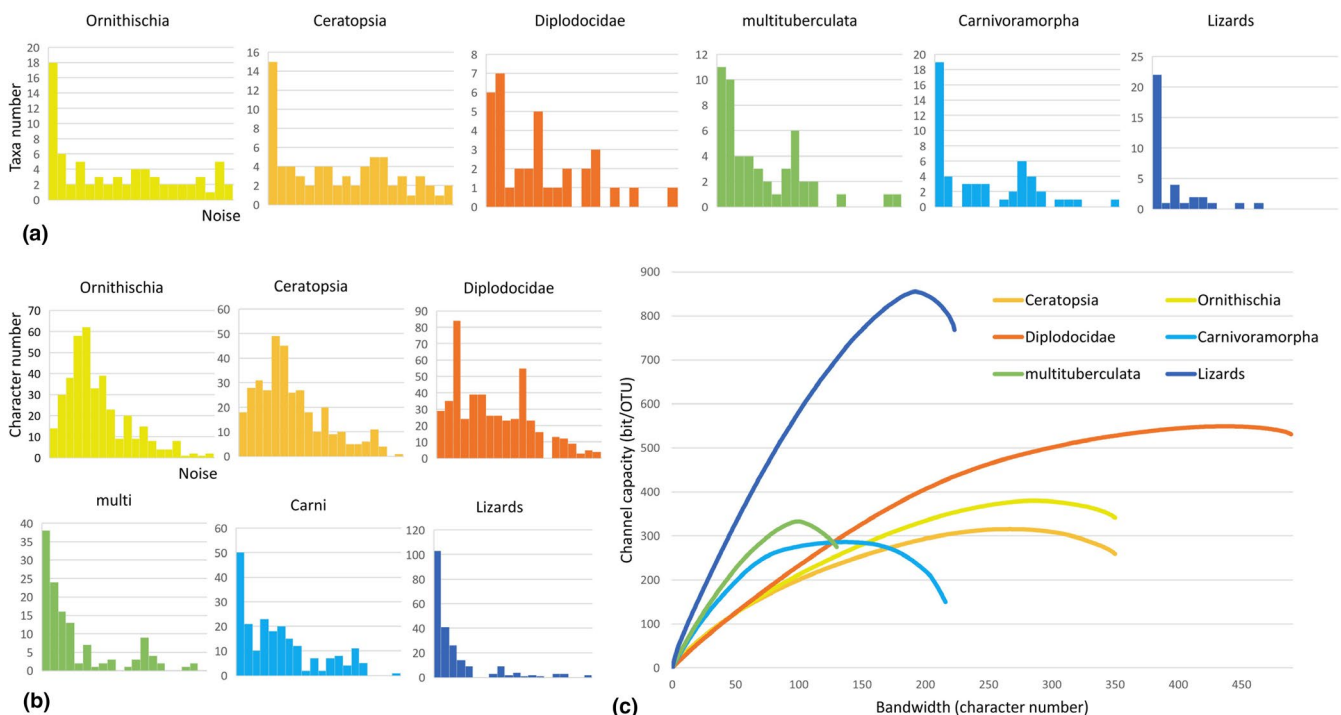
### 5.1 | Information source

No matter what algorithm is being used in systematics studies, the common aspect is using sequences (DNA, amino acid, and morphological characters) to characterize organisms and to interpret their evolutionary history. With fixed alphabets, DNA and protein sequences resemble digital 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 digital signals from analog signals, and meanwhile, the probably infinite original information entropy of fossil morphology is converted into finite entropy, represented by hundreds to thousands of morphological characters, that can be more easily compared. More morphological characters usually describe organisms more completely, but it is extremely difficult to measure how completely 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 seem to reach the saturation of characters already. Shannon (1949) proposed the Sampling Theorem (also known as Nyquist-Shannon Sampling Theorem because early work was done by Nyquist 1924, 1928), 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 (Figure 3c) does not necessarily mean those morphological character matrices fully represent the entire morphology of fossil specimens. Such saturation only shows these matrices cannot sufficiently transmit the sampled morphological information in themselves while some other information may be left as the sampling of characters are strongly biased. The morphological matrix of multituberculata (Wang et al., 2019) comprises only characters from the cranial region, but the postcrania of those organisms also have information.

With the wide applications 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



**FIGURE 3** (a) noise power distribution in taxa domain; (b) noise power distribution in character domain; (c) channel capacity and bandwidth in character matrices

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.

## 5.2 | The properties of the 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. Treating the character number as bandwidth, then the characters probably correspond to the frequency domain in a typical communication system and the OTUs to the temporal domain. This model sounds natural based on the model in Figure 1 as every organism ever lived on earth was a message sent, and fossils are a small fraction received. However, in the 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 Figure 3a. For the temporal 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 (Figure 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.

## 5.3 | 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 kinds 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 and 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 suggests that implied weighting and its variants probably provide a direction in reconstructing better resolved trees, but neither the theoretical basis nor its utilization answers the core question of how much information is in each character and may fail when working with character matrices with too many homoplastic characters.

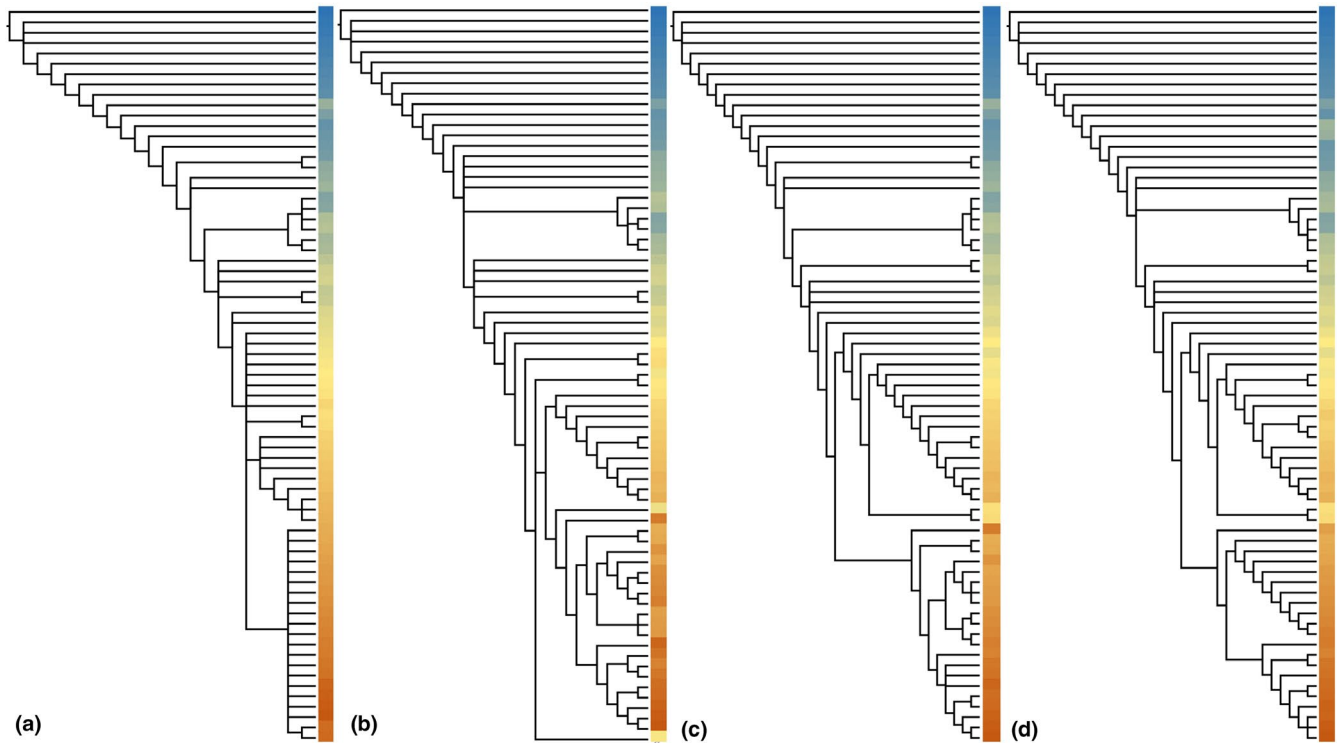
Birds and modern mammals are both endothermic, covered with filaments rather than scales, having four-chamber hearts, etc. If we would deliberately sample too many characters describing these features, the conclusion could easily be forced into that birds are mammals, and many synapomorphies between 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 characters can be hard to recognize in 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 matrix construction, the choice of characters is often extremely biased toward cranial characters in vertebrate paleontology studies (Figure 2d). In the six datasets we analyzed here, the proportion of cranial characters range from 40.7% to 100% with an average of 63.2%, which immediately shows that some parts are considered to have more morphological information (or to be "more important") than others in systematic studies.

Kälersjö et al. (1999) studied plant nucleotides data and their results showed that fast evolving and highly homoplastic third codon positions, contrary to traditional thought, have the 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, that is, 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





**FIGURE 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

**TABLE 3** CI and RI of different morphological character matrices

Index	Weighting	Ornithischia	Ceratopsia	Diplodocidae	multituberculata	Carnivoramorpha	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

analyzed 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. Detailed phylogenetic results are provided online at <https://doi.org/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 most parsimonious 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 fit better for entropy weighted characters.

## 6 | CONCLUSION

Under the framework of information theory and communication system engineering, we show that the information entropy, which measure how informative a character is, varies a lot in different characters. Characters with more states have significant higher information entropy than binary characters. Mutuality between characters does not show clear patterns except for limited body parts. All six character matrices analyzed in this study have oversampled characters that lead to saturation in channel capacity. Last, information entropy, without any prerequisites, can be used as a criterion for character weighting in systematic studies.

## ACKNOWLEDGEMENTS

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.

## CONFLICT OF INTEREST

All authors declare that they have no conflicts of interest.

## AUTHOR CONTRIBUTIONS

**Congyu Yu:** Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Project administration (lead); Resources (lead); Software (lead); Supervision (lead); Validation (lead); Visualization (lead); Writing-original draft (lead); Writing-review & editing (lead). **Qigao Jiangzuo:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Emanuel Tschopp:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Haibing Wang:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Mark Norell:** Conceptualization (equal); Project administration (equal); Supervision (equal); Writing-original draft (equal).

## DATA AVAILABILITY STATEMENT

Phylogenetic results of Ornithischia, Ceratopsia, Diplodocidae, multituberculata, Carnivoramorpha, and lizards, by equal weighting, implied weighting ( $k = 3$  and  $12$ ), and information entropy weighting are presented in.nex files. The strict consensus tree is appended in the end of trees in each file. The phylogenetic files can be found online at <https://doi.org/10.5061/dryad.8sf7m0cnc>.

## ORCID

Congyu Yu  <https://orcid.org/0000-0002-8686-3362>

## REFERENCES

- Baron, M. G., Norman, D. B., & Barrett, P. M. (2017a). A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature*, *543*(7646), 501–506. <https://doi.org/10.1038/nature21700>
- Baron, M. G., Norman, D. B., Barrett, P. (2017b). Baron et al. reply. *Nature*, *551*(7678), E4–E5. <https://doi.org/10.1038/nature24012>
- Brocklehurst, N., & Benevento, G. L. (2020). Dental characters used in phylogenetic analyses of mammals show higher rates of evolution, but not reduced independence. *PeerJ*, *8*, e8744. <https://doi.org/10.7717/peerj.8744>
- Catalano, S. A., Goloboff, P. A., & Giannini, N. P. (2010). Phylogenetic morphometrics (I): The use of landmark data in a phylogenetic framework. *Cladistics*, *26*(5), 539–549. <https://doi.org/10.1111/j.1096-0031.2010.00302.x>
- Congreve, C. R., & Lamsdell, J. C. (2016). Implied weighting and its utility in palaeontological datasets: A study using modelled phylogenetic matrices. *Palaeontology*, *59*(3), 447–462. <https://doi.org/10.1111/pala.12236>
- Cracraft, J. (1974). Phylogenetic models and classification. *Systematic Biology*, *23*(1), 71–90. <https://doi.org/10.1093/sysbio/23.1.71>
- Estabrook, G. F., Johnson, C. S., & Mc Morris, F. R. (1975). An idealized concept of the true cladistic character. *Mathematical Biosciences*, *23*(3), 263–272. [https://doi.org/10.1016/0025-5564\(75\)90040-1](https://doi.org/10.1016/0025-5564(75)90040-1)
- Farris, J. S. (1969). A successive approximations approach to character weighting. *Systematic Biology*, *18*(4), 374–385. <https://doi.org/10.2307/2412182>
- Farris, J. S. (1979). The information content of the phylogenetic system. *Systematic Biology*, *28*(4), 483–519. <https://doi.org/10.2307/sysbio/28.4.483>
- Goloboff, P. A. (1993). Estimating character weights during tree search. *Cladistics*, *9*(1), 83–91. <https://doi.org/10.1111/j.1096-0031.1993.tb00209.x>
- Goloboff, P. A. (2014). Extended implied weighting. *Cladistics*, *30*(3), 260–272. <https://doi.org/10.1111/cla.12047>
- Goloboff, P. A., & Catalano, S. A. (2016). TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*, *32*(3), 221–238. <https://doi.org/10.1111/cla.12160>
- Han, F. et al (2017). Postcranial anatomy of *Yinlong downsii* (Dinosauria: Ceratopsia) from the Upper Jurassic Shishugou Formation of China and the phylogeny of basal ornithischians. *Journal of Systematic Palaeontology*, 1–29. <https://doi.org/10.1080/14772019.2017.1369185>
- Harrison, T. (1993) Cladistic concepts and the species problem in hominoid evolution. In W. H. Kimbel, & L. B. Martin (eds) *Species, Species Concepts and Primate Evolution*. Springer US, pp. 345–371. [https://doi.org/10.1007/978-1-4899-3745-2\\_14](https://doi.org/10.1007/978-1-4899-3745-2_14)
- Hawkins, J. A., Hughes, C. E., & Scotland, R. W. (1997). Primary homology assessment, characters and character states. *Cladistics*, *13*(3), 275–283. <https://doi.org/10.1111/j.1096-0031.1997.tb00320.x>
- Källersjö, M., Albert, V. A., & Farris, J. S. (1999). Homoplasy increases phylogenetic structure. *Cladistics*, *15*(1), 91–93. <https://doi.org/10.1111/j.1096-0031.1999.tb00400.x>
- Kielan-Jaworowska, Z., & Dovchin, N. (1966) 'Narrative of the Polish-Mongolian Palaeontological expeditions 1963-1965', p. 34.
- Laing, A. M., Doyle, S., Gold, M. E. L., Nesbitt, S. J., O'Leary, M. A., Turner, A. H., Wilberg, E. W., & Poole, K. E. (2018). Giant taxon-character matrices: The future of morphological systematics. *Cladistics*, *34*(3), 333–335. <https://doi.org/10.1111/cla.12197>
- Langer, M. C., Ezcurra, M. D., Rauhut, O. W. M., Benton, M. J., Knoll, F., McPhee, B. W., Novas, F. E., Pol, D., & Brusatte, S. L. (2017). Untangling the dinosaur family tree. *Nature*, *551*(7678), E1–E3. <https://doi.org/10.1038/nature24011>
- Lee, Y.-N., Barsbold, R., Currie, P. J., Kobayashi, Y., Lee, H.-J., Godefroit, P., Escuillié, F., & Chinzorig, T. (2014). Resolving the long-standing enigmas of a giant ornithomimosaur *Deinocheirus mirificus*. *Nature*, *515*(7526), 257–260. <https://doi.org/10.1038/nature13874>
- Mickey, M. F., & Platnick, N. I. (1989). On the information content of classifications. *Cladistics*, *5*(1), 33–47. <https://doi.org/10.1111/j.1096-0031.1989.tb00481.x>
- Nelson, G. J. (1972). Phylogenetic relationship and classification. *Systematic Zoology*, *21*(2), 227–231. <https://doi.org/10.2307/2412293>
- Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., & Nieves-Aldrey, J. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*(1), 47–67. <https://doi.org/10.1080/10635150490264699>

- Nyquist, H. (1924). Certain factors affecting telegraph speed. *Transactions of the American Institute of Electrical Engineers*, XLIII, 412–422. <https://doi.org/10.1109/T-AIEE.1924.5060996>
- Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2), 617–644. <https://doi.org/10.1109/T-AIEE.1928.5055024>
- O'Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M. J., Perini, F. A., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., ... Cirranello, A. L. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science*, 339(6120), 662–667. <https://doi.org/10.1126/science.1229237>
- Ostrow, J. H. (1972). *McGraw-Hill Yearbook of Science and Technology for 1971*, 176–179. McGraw-Hill.
- Senter, P. (2007). A new look at the phylogeny of coelurosauria (Dinosauria: Theropoda). *Journal of Systematic Palaeontology*, 5(4), 429–463. <https://doi.org/10.1017/S1477201907002143>
- Sereno, P. C. (2007). Logical basis for morphological characters in phylogenetics. *Cladistics*, 23(6), 565–587. <https://doi.org/10.1111/j.1096-0031.2007.00161.x>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10–21. <https://doi.org/10.1109/JRPROC.1949.232969>
- Simões, T. R., Caldwell, M. W., Palci, A., & Nydam, R. L. (2017). Giant taxon-character matrices: Quality of character constructions remains critical regardless of size. *Cladistics*, 33(2), 198–219. <https://doi.org/10.1111/cla.12163>
- Spaulding, M., & Flynn, J. J. (2012). Phylogeny of the Carnivoromorpha: The impact of postcranial characters. *Journal of Systematic Palaeontology*, 10(4), 653–677. <https://doi.org/10.1080/14772019.2011.630681>
- Tschopp, E., & Mateus, O. (2017). Osteology of *Galeamopus pabsti* sp. nov. (Sauropoda: Diplodocidae), with implications for neurocentral closure timing, and the cervico-dorsal transition in diplodocids. *PeerJ*, 5, e3179. <https://doi.org/10.7717/peerj.3179>
- Tschopp, E., Villa, A., Camaiti, M., Ferro, L., Tuveri, C., Rook, L., Arca, M., & Delfino, M. (2018). The first fossils of Timon (Squamata: Lacertinae) from Sardinia (Italy) and potential causes for its local extinction in the Pleistocene. *Zoological Journal of the Linnean Society*, 184(3), 825–856. <https://doi.org/10.1093/zoolinnean/zly003>
- Wang, H., Meng, J., & Wang, Y. (2019). Cretaceous fossil reveals a new pattern in mammalian middle ear evolution. *Nature*, 576(7785), 102–105. <https://doi.org/10.1038/s41586-019-1792-0>
- Wilkinson, M. (1995). A comparison of two methods of character construction. *Cladistics*, 11(3), 297–308. [https://doi.org/10.1016/0748-3007\(95\)90017-9](https://doi.org/10.1016/0748-3007(95)90017-9)
- Wilkinson, M., Cotton, J. A., & Thorley, J. L. (2004). The information content of trees and their matrix representations. *Systematic Biology*, 53(6), 989–1001. <https://doi.org/10.1080/10635150490522737>
- Yu, C., Prieto-Marquez, A., Chinzorig, T., Badamkhatan, Z., & Norell, M. (2020). A neoceratopsian dinosaur from the early Cretaceous of Mongolia and the early evolution of ceratopsia. *Communications Biology*, 3(1), 1–8. <https://doi.org/10.1038/s42003-020-01222-7>

**How to cite this article:** Yu, C., Jiangzuo, Q., Tschopp, E., Wang, H., & Norell, M. (2021). Information in morphological characters. *Ecology and Evolution*, 11, 11689–11699. <https://doi.org/10.1002/ece3.7874>