



# Codon Usage Bias in Autophagy-Related Gene 13 in Eukaryotes: Uncovering the Genetic Divergence by the Interplay Between Nucleotides and Codon Usages

Yicong Li<sup>1†</sup>, Rui Wang<sup>2†</sup>, Huihui Wang<sup>1</sup>, Feiyang Pu<sup>1</sup>, Xili Feng<sup>1</sup>, Li Jin<sup>1</sup>, Zhongren Ma<sup>1</sup> and Xiao-xia Ma<sup>1\*</sup>

<sup>1</sup> Biomedical Research Center, Northwest Minzu University, Lanzhou, China, <sup>2</sup> Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

## OPEN ACCESS

### Edited by:

Yongsheng Liu,  
Lanzhou Veterinary Research Institute  
(CAAS), China

### Reviewed by:

Jinzhong Wang,  
Nankai University, China  
Xin Cao,  
Jilin Agriculture University, China

### \*Correspondence:

Xiao-xia Ma  
maxiaoxia@xbmu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Virus and Host,  
a section of the journal  
Frontiers in Cellular and Infection  
Microbiology

**Received:** 05 September 2021

**Accepted:** 12 October 2021

**Published:** 05 November 2021

### Citation:

Li Y, Wang R, Wang H, Pu F,  
Feng X, Jin L, Ma Z and Ma X-x (2021)  
Codon Usage Bias in Autophagy-  
Related Gene 13 in Eukaryotes:  
Uncovering the Genetic Divergence  
by the Interplay Between  
Nucleotides and Codon Usages.  
*Front. Cell. Infect. Microbiol.* 11:771010.  
doi: 10.3389/fcimb.2021.771010

Synonymous codon usage bias is a universal characteristic of genomes across various organisms. Autophagy-related gene 13 (*atg13*) is one essential gene for autophagy initiation, yet the evolutionary trends of the *atg13* gene at the usages of nucleotide and synonymous codon remains unexplored. According to phylogenetic analyses for the *atg13* gene of 226 eukaryotic organisms at the nucleotide and amino acid levels, it is clear that their nucleotide usages exhibit more genetic information than their amino acid usages. Specifically, the overall nucleotide usage bias quantified by information entropy reflected that the usage biases at the first and second codon positions were stronger than those at the third position of the *atg13* genes. Furthermore, the bias level of nucleotide 'G' usage is highest, while that of nucleotide 'C' usage is lowest in the *atg13* genes. On top of that, genetic features represented by synonymous codon usage exhibits a species-specific pattern on the evolution of the *atg13* genes to some extent. Interestingly, the codon usages of *atg13* genes in the ancestor animals (*Latimeria chalumnae*, *Petromyzon marinus*, and *Rhinatrema bivittatum*) are strongly influenced by mutation pressure from nucleotide composition constraint. However, the distributions of nucleotide composition at different codon positions in the *atg13* gene display that natural selection still dominates *atg13* codon usages during organisms' evolution.

**Keywords:** autophagy-related gene 13, phylogenetic analyses, nucleotide usage, synonymous codon usage, nucleotide composition distribution

## INTRODUCTION

Autophagy (or macroautophagy) is an intracellular protein-degradation process that is conserved from yeast to mammalian cells. This self-degradation system, where lysosomes/vacuoles degrade unnecessary cytoplasmic components from nutrient starvation, was first discovered in mammalian cells in the 1960s (Ashford and Porter, 1962). With the ongoing studies on autophagy in yeasts, a series of autophagy-related genes (*atg*) encoding proteins that are involved in autophagy

membranes biogenesis came into research attention (Tsukada and Ohsumi, 1993; Thumm et al., 1994; Harding et al., 1995). By investigating the molecular machinery of autophagy, 35 different *atg* genes have been identified in the morphologic and biochemical pathways from yeast genetic studies. Among these autophagy-related genes, *atg1-10*, *atg12-14*, *atg16-18*, *atg29*, and *atg31* play important roles in shaping canonical autophagosomes (Nakatogawa et al., 2009). The identified *atg* genes are highly conserved across various eukaryotes. The findings of these proteins take the studies of autophagosome formation to the molecular level where the genetic characteristics of autophagy are further investigated in mammalian cells. While it has been believed that ATG proteins are mainly involved in autophagy, some of them were actually identified to display non-autophagic functions, such as antiviral responses, cell proliferation, and development assistance (Kroemer and Levine, 2008; Radoshevich et al., 2010; Fan et al., 2017; Ma et al., 2020). Such diverse dynamics of the ATG proteins attract deeper investigations at a genetic level.

ATG13 is a serine-rich protein that contains several phosphorylation sites and other critical genetic regions (Fujioka et al., 2014). For example, ATG13-17BR and ATG13-17LR, two important functional regions, are able to interact with ATG17 (another autophagy-related protein) to construct a complex with ATG29 and ATG31 (Fujioka et al., 2014; Yamamoto et al., 2016). In addition, ATG13, which is the direct substrate of the target of rapamycin complex 1 (TORC1), is hyper-phosphorylated under nutrient-rich conditions (Funakoshi et al., 1997; Kamada et al., 2000; Kamada et al., 2010). In addition, another functional domain of ATG13, the MIM domain, can bind to the protein kinase ATG1 (Fujioka et al., 2014). The MIM domain under phosphorylation weakens the interaction between ATG1 and ATG13, therefore, as in the case of ATG17, the interplay between ATG1 and ATG13 can be facilitated by starvation in a TORC1-dependent regulation manner (Kamada et al., 2000; Fujioka et al., 2014; Noda, 2017).

The degeneracy of the genetic code results in the synonymous codon where different codons would translate into the same amino acid. The genetic codes that define the amino acid sequence of a protein are able to mediate the efficiency and rate of translation (Brule and Grayhack, 2017). Synonymous codon usages are commonly regarded as a bridge between nucleotide usages and amino acid compositions (Novoa and Ribas De Pouplana, 2012). Interestingly, the homogenous gene among different species may prefer different sets of synonymous codon usage, and the overall codon usage pattern appears to be similar among closely related species but displays divergence among distantly related species (De Mandal et al., 2020; Ge et al., 2020; Liu et al., 2020; Gupta et al., 2021). The reasons behind the uneven usage frequencies of synonymous codons in protein-coding have been puzzling molecular evolutionary researchers for decades (Pedrotti and Guerra, 1989). In animals, studies of the population genomics mark the codon usage bias by the effects of translational selection and GC-biased gene conversion on population transcriptomics data (Galtier et al., 2018). Moreover, the overall codon usage bias and the interplay between nucleotide usages and synonymous codon usage patterns

manifest the non-randomness in coding sequences, making it an important area of study for cellular genetic processes (Fedorov et al., 2002; Carlini, 2005; Nabiyouni et al., 2013). In addition, synonymous codon usage bias also has strong correlations with the formations of different folding structures of protein, resulting in altering biological functions of protein (Komar et al., 1999; Zhou et al., 2009; Saunders and Deane, 2010; Weygand-Durasevic and Ibba, 2010; Zhou et al., 2013b). Having a large intrinsically disordered region that can tether ATG17 molecules and lead to supramolecular ATG1 complex assembly, ATG13 works as an essential autophagy factor in mediating the supramolecular assembly for autophagy initiation (Yamamoto et al., 2016). While the existing sequencing technology has granted more accessibility to the studies on eukaryotic genome patterns, the knowledge about the conservation and evolution of *atg* genes across eukaryotes still remains unclear. Among the questions which have been raised is: do *atg13* genes derived from different species display similar patterns in their genetic characterizations on synonymous codon usage or not? Here, in this study, we wanted to uncover more insights into how nucleotide, synonymous codon usage, and amino acid usages act on the *atg13* gene evolutionary trend in eukaryotes by investigating the *atg13* genes derived from 226 different eukaryotic organisms.

## MATERIALS AND METHODS

### Phylogenetic Analyses for Different *atg13* Genes at the Nucleotide and Amino Acid Levels

The different *atg13* genes derived from 226 eukaryotic organisms were obtained from the National Center for Biotechnology (NCBI) Genome database (<https://www.ncbi.nlm.nih.gov/>) (Table S1). According to the whole coding sequences of 226 *atg13* genes (Table S1), the two phylogenetic trees were constructed by Maximum Composite Likelihood model with UPGMA statistical method at the nucleotide and amino acid levels, respectively.

### Nucleotide Usage Bias of the *atg13* Genes

To quantify the degree of the bias at the overall nucleotide usage of every nucleotide position in a code, the information entropy formula was used for calculating the nucleotide usage bias of the *atg13* gene per species.

$$Entropy = -\sum_i f_i \times \log_2(f_i)$$

$$f_i = \frac{F_i}{F(A) + F(T) + F(G) + F(C)}$$

where  $f_i$  stands for the probability of the specific nucleotide ( $F_i$ ),  $F_i$  means the number of occurrences of the specific nucleotide. The value for information entropy of nucleotide usage bias reflects how dispersed the contribution of the overall usage of the four nucleotide types is: the higher the value, the more uniform the usage between different nucleotides is; in contrast,

a lower value reflects a more biased usage between different nucleotides (Ge et al., 2020).

In addition, to assess the usage distribution for each nucleotide across the three positions in a code of *atg13* gene, we also implemented the strategy of information entropy to quantify the overall usage bias for each nucleotide.

$$\text{Entropy} = -\sum_i f_i \times \log_2(f_i)$$

$$f_n = \frac{F_n}{F(N_1) + F(N_2) + F(N_3)}$$

where  $f_n$  stands for the probability of the specific nucleotide type ('n' corresponding to A, T, C, or G),  $F_n$  means the number of occurrences of the specific nucleotide type.  $F(N_1)$ ,  $F(N_2)$ , and  $F(N_3)$  correspond to the nucleotide content at the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> codon positions, respectively.

## Relative Synonymous Codon Usage Calculation

The relative synonymous codon usage value (RSCU) was used to quantify the synonymous codon usage bias for each *atg13* gene in this study. In principle, three canonical stop codons (UAA, UAG, and UGA), AUG for Met, and UGG for Try are unavailable for RSCU analyses. According to the previous reports for estimating synonymous codon usage (Sharp and Li, 1986; Zhou et al., 2013a; Gun et al., 2018), the related standard was established that codons with RSCU values >1.6 would be regarded as overrepresented ones, while codons with RSCU values <0.6 were thought to be underrepresented. Therefore, a synonymous codon with an RSCU value <0.6 or >1.6 can be defined as a biased one.

To better quantify the dispersion of synonymous codon usages in *atg13* genes across different eukaryotic species, we introduced the formula ( $V_s = \sigma/\bar{x}$ ) into this study.

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$

where  $\{x_1, x_2, x_3, \dots, x_N\}$  are the RSCU values of the given synonymous codon of *atg13* gene derived from the specific eukaryotic organisms in this study;  $\bar{x}$  means the average value of these RSCU values; and 'N' is the number of eukaryotic species in this study. Based on  $V_s$  values, the lower the  $V_s$  value is, the more stable the specific synonymous codon usage is.

Genetic divergence at synonymous codon and amino acid usages for *atg13* gene:

To visualize the overall genetic divergence at synonymous codon or amino acid usages among the 226 *atg13* genes, we implemented principal component analysis (PCA), a multivariate statistical pathway. Specifically, the 59 canonical synonymous codons were regarded as a multi-dimensional vector (each dimension corresponds to the RSCU value for each sense codon). Similarly, to estimate ATG13 protein evolution across these eukaryotic organisms, the 20 canonical

amino acids were regarded as a 20-dimensional vector in regard to the different frequencies of the amino acid.

## Analyses for the Overall Codon Usage Patterns of *atg13* Gene Influenced by Selective Forces:

To quantify the absolute magnitude of bias at the overall codon usage for each *atg13* gene, the effective number of codons (ENC) calculation was carried out, ranking the codon usage bias for *atg13* gene per eukaryotic organism in this study. While ENC values can range from 20 to 61 (Wright, 1990), the lower the value is, the more biased the absolute codon usage is. It has been pointed out that one gene would have a significant codon bias when the ENC value is less than 35 (Comeron and Aguadé, 1998). To further evaluate the interplay between mutation pressure and natural selection forcing evolutionary trends of *atg13* gene at codon usages, the parity rule 2 (PR2) plot analysis was performed for each sample in this study. In detail, the PR2 plot is constructed by the ordinate (AU-bias [ $A_3/(A_3 + T_3)$ ]) and the abscissa (the GC-bias [ $G_3/(G_3 + C_3)$ ]) (Butt et al., 2016). Here,  $A_3$ ,  $T_3$ ,  $G_3$ , and  $C_3$  stand for frequencies of occurrence of the corresponding base at the third codon position of the four-codon amino acids of gene population). As for the plot represented by AU-bias vs. GC-bias, the center of the plot, where both coordinate are 0.5, is where base A equal to base U and base G equal to base C (PR2), with no bias between the influence of the mutation and selection rates (substitution rates) (Sueoka, 1995; Sueoka, 1999).

## Estimating Effects of Nucleotide Compositional Distribution of *Atg13* Gene on Its Protein Property

For a better understanding of the compositional distribution influenced by nucleotide usages in these *atg13* genes, we implemented nucleotide skew variants. From the variants, we would be able to explore the previously hidden relationships between the overall nucleotide skew across coding sequence and nucleotide composition distribution in the specific position in a code. AT skew and GC skew were commonly used for displaying the compositional distribution of nucleotide usages (Tillier and Collins, 2000). Positive AT and GC skew values represent a higher proportion of A base over T and similarly, higher G over C and vice versa. The deviation of skew value from zero marks the disproportional usage of two bases in the gene. Based on the analysis strategy of GC and AT skews, we also calculated purine skew [ $(A-G)/(A+G)$ ], pyrimidine skew [ $(T-C)/(T+C)$ ], keto skew [ $(T-G)/(T+G)$ ] and amino skew [ $(A-C)/(A+C)$ ]. The bivariate correlation analyses were conducted between nucleotide skew data for the specific codon position and the corresponding overall nucleotide skew data from using the Pearson method with a two-tailed test.

## RESULTS

### The Obvious Specific-Species Genetic Diversity for *atg13* Genes

Serving as one conserved gene across eukaryotic species, the phylogenetic tree of the *atg13* gene displays the obvious species-

specific genetic divergences at the nucleotide level (**Figure 1**). As shown in **Figure 1**, the three *Brachionus* spp. (*B. calyciflorus*, *B. plicatilis*, and *B. rotundiformis*) have a significantly divergent evolutionary trend relative to those of other animals. Likewise, *Locusta migratoria* (a kind of insect) retains its specific evolutionary trend at the nucleotide level. As for the fish family in this study, it can be observed that there are three significantly different genetic branches. Specifically, the three divergent genetic populations do not correlate to geographic factors or living habits on the *atg13* evolutionary trends of fishes. However, the species-specific nature still plays an important role in *atg13* genes among different kinds of fishes. Turning to *atg13* genes in the amphibian family in this study, Anura (*Nanorana parkeri* and *Xenopus tropicalis*) does not serve as a sister group with Apoda (*Microcaecilia unicolor*, *Geotrypetes seraphini*, or *Rhinatrema bivittatum*). Interestingly, *Petromyzon marinus* and *Erpetoichthys calabaricus* (two ancient animals) exhibit similar genetic niches of *atg13* gene at nucleotide organization. Likewise, the species-specific evolutionary dynamic also acts on genetic divergence of *atg13* genes between Mammals and *Sauropsida* animals (**Figure 1**).

Next, from the phylogenetic tree, we found that at the amino acid level, *atg13* genes display a species-specific dynamic in their evolutionary trends (**Figure 2**). Compared with the genetic diversity of *atg13* genes with other eukaryotes (**Figure 1**), ATG13 proteins of both *Brachionus* spp. and *L. migratoria* still display distinct evolutionary trends in regard to their species, while the majority of the fish in this study has been classified into one genetic group. As for the genetic diversity of ATG13 proteins, we found that Anura and Apoda compose of a sister group. In addition, the four marsupials (*Monodelphis domestica*, *Phascolarctos cinereus*, *Sarcophilus harrisi*, and *Vombatus ursinus*) and *Ornithorhynchus anatinus* are found to have different genetic niches in comparison to the rest mammals in this study (**Figure 2**). Taken together, the evolutionary trend of the *atg13* gene is strongly influenced by the species-specific selection at either nucleotide or amino acid level to different degrees.

## Amino Acid Composition Constraint Influences Nucleotide Usages in the *atg13* Gene

To quantify the magnitudes of nucleotide usage bias in different codon positions of these sequences, we used information entropy to estimate the differences of nucleotide usage bias in each position in comparison to the overall nucleotide usage bias for the whole coding sequence. Generally, the nucleotide usage biases in the first and second codon positions are significantly stronger than the overall usage bias ( $p$  value < 0.001) (**Figure 3**), implying that nucleotide usage patterns for *atg13* genes are subject to selective forces derived from nucleotide composition constraint and translation selection. Of note, despite significantly different biases between the overall nucleotide usage bias and that of the third codon position, the observation of the highly variant ranks of nucleotide usage biases are carried out at the third codon position of *atg13* genes in some eukaryotic

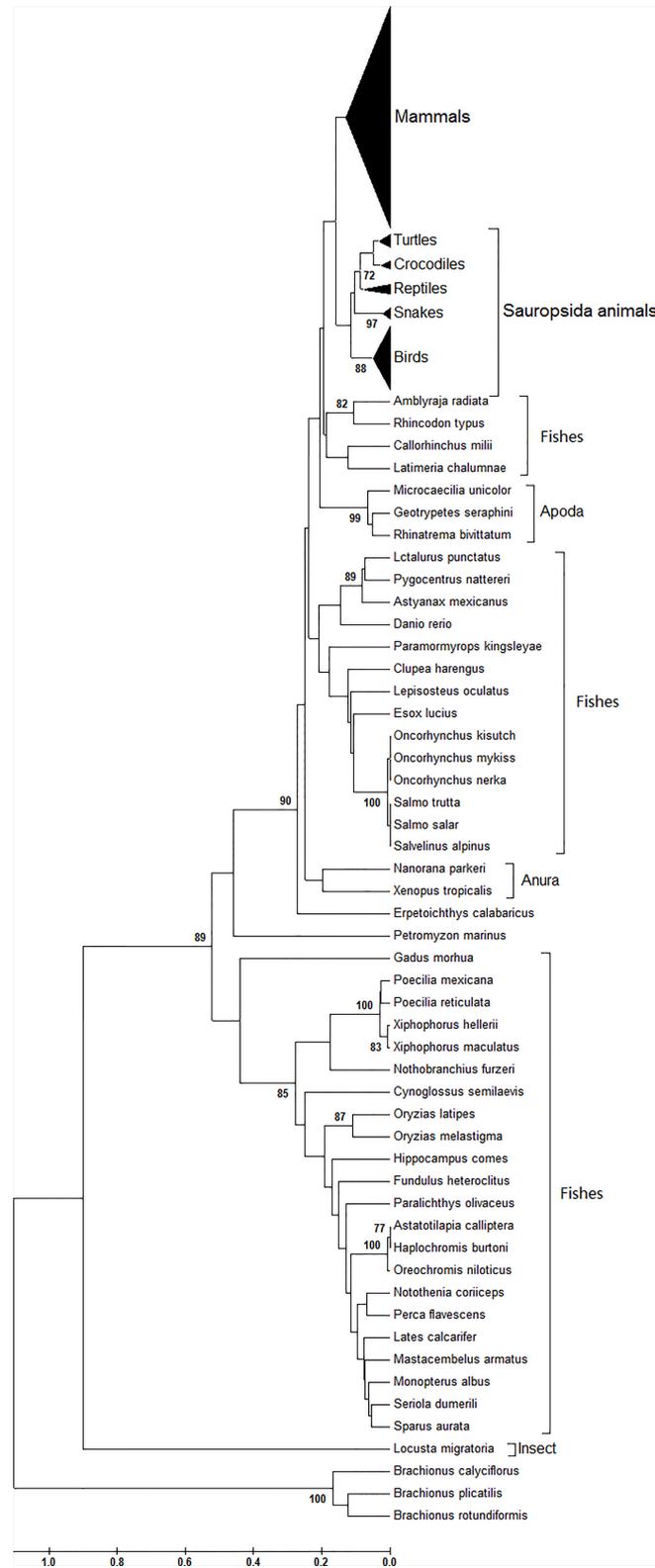
organisms (**Figure 3**). According to the data of nucleotide usage bias, the outliers are observed in columns 'N', 'N2', and 'N3'. It can be found in **Figure 3** that *Brachionus calyciflorus* is included in column 'N', *Brachionus rotundiformis*, *Brachionus plicatilis*, and *Brachionus calyciflorus* in column 'N2', and *Gadus morhua*, *Oryzias latipes*, *Oryzias melastigma*, *Brachionus calyciflorus*, *Petromyzon marinus*, *Fukomys damarensis*, *Nothobranchius furzeri*, *Poecilia reticulata*, *Poecilia Mexicana*, *Sarcophilus harrisi*, and *Fundulus heteroclitus* in column 'N3'. These outliers suggest the existence of stronger selective forces acting on nucleotide usage patterns at the specific codon position.

According to nucleotide usage bias at the three codon positions (**Figure 3**), we further estimated the usage bias extent of each type of nucleotide across all positions in a code for the *atg13* gene. The usage bias extent of nucleotide 'G' usage is highest, while that of nucleotide 'C' usage is lowest in *atg13* gene (**Figure 4**). Moreover, the outliers are also observed at the four nucleotide types, namely *Fundulus heteroclitus*, *Oryzias melastigma*, *Oryzias latipes*, and *Gadus morhua* in the column 'T'; *Locusta migratoria* and *Brachionus calyciflorus* in the column 'C'; *Oryzias latipes*, *Oryzias melastigma*, and *Fundulus heteroclitus* in the column 'A'; *Brachionus plicatilis*, *Brachionus rotundiformis*, and *Brachionus calyciflorus* in the column 'G'. These outliers further prove that nucleotide usage bias for the *atg13* gene is more influenced by natural selection than by mutation pressure from the nucleotide composition constraint.

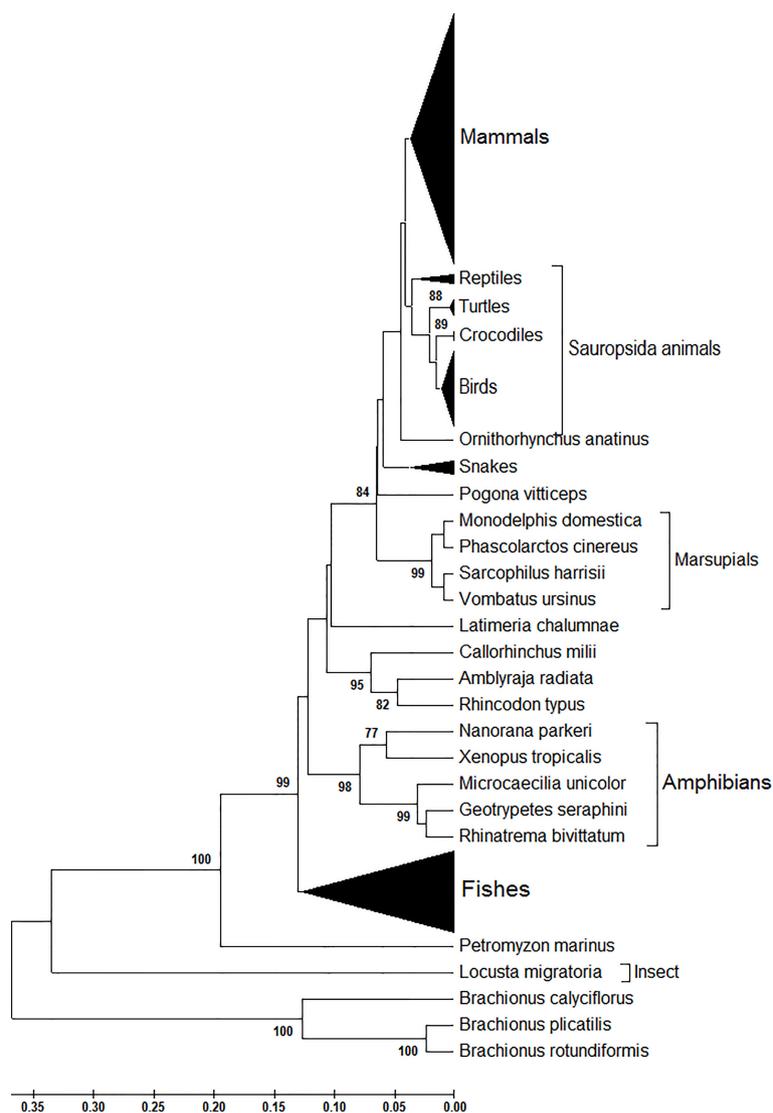
## Strong Selective Forces Acting On Synonymous Codon Usage Patterns of *atg13* Gene

Given the strong bias at nucleotide usages of the *atg13* gene (**Figures 3 and 4**), we subsequently calculated RSCU data for the *atg13* gene per species (**Table S2**). Generally, different eukaryotic organisms exhibit their specific codon usage patterns. Moreover, despite the higher composition of GC than AT in most of the genomes in this study, synonymous codons with neither G/C ends nor A/T end display a uniform pattern. To estimate the roles of nucleotide composition constraint and natural selection in synonymous codon usages, we estimated the dispersion extent ( $V_s$  value) for each synonymous codon usage in *atg13* genes across these eukaryotic species. Among the synonymous codon usages (**Table 1**), some display highly variant usage patterns, including CTA for Leu ( $V_s=0.718$ ), TCG for Ser ( $V_s=0.966$ ), CCG for Pro ( $V_s=0.903$ ), GCG for Ala ( $V_s=1.093$ ), and CGC for Arg ( $V_s=0.883$ ). In the meantime, however, some exhibit strongly stable usage patterns, including TTT for Phe ( $V_s=0.158$ ), GTG for Val ( $V_s=0.196$ ), TCT for Ser ( $V_s=0.188$ ), CCT for Pro ( $V_s=0.180$ ), CAG for Gln ( $V_s=0.159$ ), AAG for Lys ( $V_s=0.183$ ), GAC for Asp ( $V_s=0.162$ ), and GAG for Glu ( $V_s=0.197$ ). These results prove the existence of the strong selective forces acting on synonymous codon usage patterns in the *atg13* gene with the specific species.

From the variant RSCU data for these *atg13* genes (**Table S2**), 57 out of 59 synonymous codon usage patterns display a wide spectrum of RSCU values among different species. However, the



**FIGURE 1** | Phylogenetic analysis of *atg13* genes of 226 eukaryotic organisms. This phylogenetic tree was constructed by the Maximum Composite Likelihood model with UPGMA statistical method at the nucleotide level. Branches leading to these groups of sequences were supported by 100% of bootstrap replications in all cases.



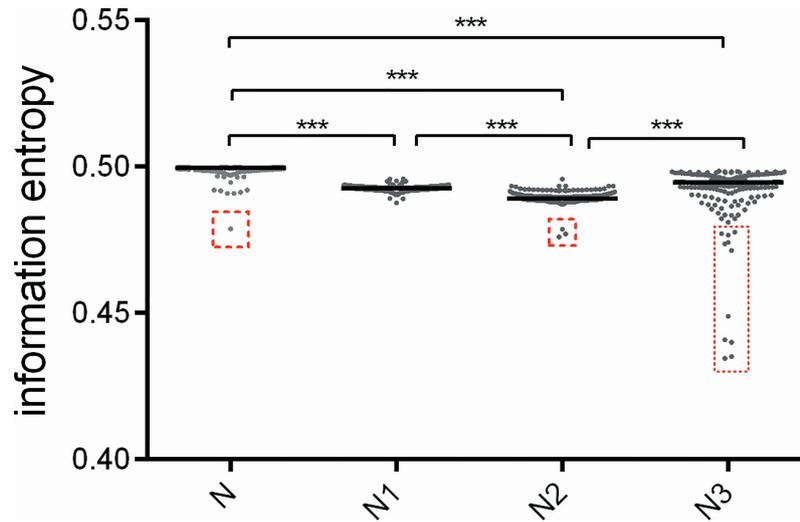
**FIGURE 2** | Phylogenetic analysis of ATG13 protein sequences of 226 eukaryotic organisms. This phylogenetic tree was constructed by the Maximum Composite Likelihood model with UPGMA statistical method at the nucleotide level. Branches leading to these groups of sequences were supported by 100% of bootstrap replications in all cases.

synonymous codons (GCG for Ala and CTA for Leu) are suppressive selections ( $RSCU < 1.0$ ), except *atg13* genes of *Petromyzon marinus* ( $RSCU_{GCG}=1.04$ ), *Meleagris gallopavo* ( $RSCU_{CTA}=1.02$ ), *Numida meleagris* ( $RSCU_{CTA}=1.07$ ), and *Cygnus atratus* ( $RSCU_{CTA}=1.26$ ). Interestingly, some synonymous codons are absent from *atg13* genes in some eukaryotic organisms, such as TTA, CTA & CTG for Leu; ATT & ATA for Ile; GTC & GTA for Val; TCG for Ser, CCG for Pro; ACG for Thr, GCG for Ala; TAT for Tyr; TGT & TGC for Cys, and CGC, CGA, CGG & AGA for Arg (Table 2), suggesting that not only do invertebrates with low rank (such as wheel animalcule) but also vertebrates with a high rank (such as *Homo sapiens*, *Pan troglodytes* and *Macaca mulatta*) would skip the synonymous codon selection. However, such a pattern

has only been found in *atg13* genes. Taken together, it can be proven that the synonymous codon usage bias of *atg13* genes in different species is more prone to the natural selection power than to the nucleotide composition constraint.

### Codon Usages of *atg13* Gene Strongly Influenced by Natural Selection

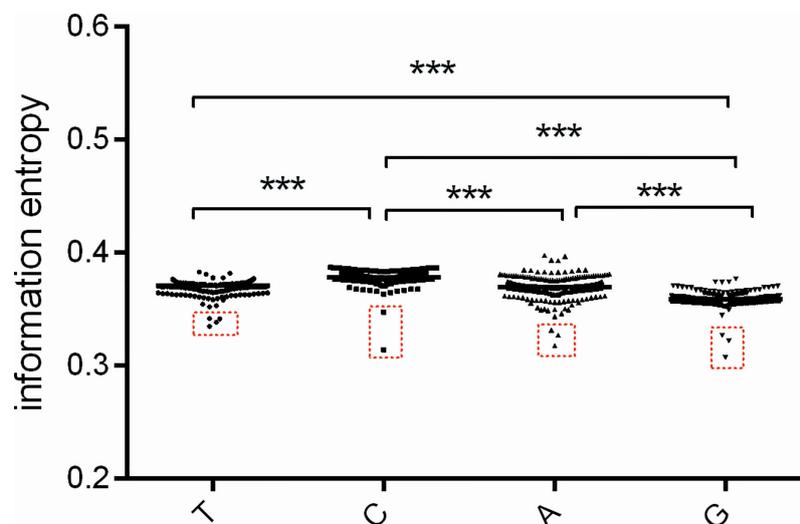
Based on the synonymous codon usage patterns of *atg13* genes in different eukaryotic organisms (Table S2), we next estimated the magnitudes of the overall codon usage for transcript variants via ENC value vs. GC3 content. The plot of ENC value vs. GC3 content reveals that no eukaryotic organism owns significantly strong bias (ENC value < 35) at the overall codon usage for *atg13* gene, and most eukaryotic organisms own *atg13* genes with ENC



**FIGURE 3** | The patterns of nucleotide usage bias for *atg13* genes of 226 eukaryotic organisms are represented by information entropy. The 'N' means the overall nucleotide usage bias across the complete gene, The 'N1', 'N2', and 'N3' stand for the patterns of nucleotide usage bias at the first, second, and third codon positions, respectively. The red box with broken lines highlights the outliers which correspond to *atg13* genes of the specific organisms. \*\*\* means  $p$ -value < 0.001.

values ranging from 50 to 55 (**Figure 5**). Specifically, these below-curve scattering dots prove that natural selection is stronger than the mutation pressure derived from nucleotide component constraint, especially *atg13* genes in *Brachionus calyciflorus*, *Gadus morhua*, *Oryzias melastigma*, *Oryzias latipes*, and *Fundulus heteroclitus*. Interestingly, with one exception for *Petromyzon marinus*, the nucleotide composition constraint dominates the mutation pressure in the overall codon usage formation of the *atg13* gene (**Figure 5**).

Furthermore, to clarify whether the overall codon usage pattern of the *atg13* gene was shaped solely by natural selection, mutation pressure, or both, PR2 plot analysis was performed to evaluate the relationships between the nucleotides A-T contents and the G-C contents in the fourfold degenerate codon families (Ala, Gly, Pro, Thr, and Val). Specifically for *atg13* genes of these eukaryotic organisms, nucleotide C is selected more frequently than nucleotide G at the third codon position, except *Latimeria chalumnae* (GC bias=0.505) and *Brachionus calyciflorus* (GC bias=0.565) (**Figure 6**). Though there is a



**FIGURE 4** | The usage bias for the specific nucleotide (A, T, G, or C) in the *atg13* genes of 226 eukaryotic organisms represented by information entropy. The red box with broken lines highlights the outliers which correspond to *atg13* genes of the specific organisms. \*\*\* means  $p$ -value < 0.001.

**TABLE 1 |** The dispersion magnitude of synonymous codon usages for *atg13* genes across all species in this study.

	Vs value		Vs value		Vs value
TTT(F)	0.158	CCT(P)	0.180	AAA(K)	0.237
TTC(F)	0.430	CCC(P)	0.234	AAG(K)	0.183
TTA(L)	0.480	CCA(P)	0.290	GAT(D)	0.227
TTG(L)	0.339	CCG(P)	0.903	GAC(D)	0.162
CTT(L)	0.265	ACT(T)	0.250	GAA(E)	0.220
CTC(L)	0.267	ACC(T)	0.223	GAG(E)	0.197
CTA(L)	0.718	ACA(T)	0.339	TGT(C)	0.249
CTG(L)	0.242	ACG(T)	0.574	TGC(C)	0.544
ATT(I)	0.275	GCT(A)	0.203	CGT(R)	0.300
ATC(I)	0.304	GCC(A)	0.245	CGC(R)	0.883
ATA(I)	0.368	GCA(A)	0.337	CGA(R)	0.496
GTT(V)	0.244	GCG(A)	1.093	CGG(R)	0.571
GTC(V)	0.274	TAT(Y)	0.306	AGA(R)	0.351
GTA(V)	0.496	TAC(Y)	0.222	AGG(R)	0.264
GTG(V)	0.196	CAT(H)	0.296	GGT(G)	0.370
TCT(S)	0.188	CAC(H)	0.287	GGC(G)	0.206
TCC(S)	0.215	CAA(Q)	0.473	GGA(G)	0.303
TCA(S)	0.293	CAG(Q)	0.159	GGG(G)	0.286
TCG(S)	0.966	AAT(N)	0.266		
AGT(S)	0.428	AAC(N)	0.244		
AGC(S)	0.206				

relatively big spectrum of AT bias for *atg13* genes of different species in this study, 37 out of 226 species have strong trends to select nucleotide A at the third codon position rather than nucleotide T (Figure 6). This unequal selection of different kinds of nucleotide at the third codon position exhibits the interplay between mutation pressure and natural selection on the overall codon usage in *atg13* genes for many eukaryotic organisms. In addition, the *atg13* genes of *Latimeria chalumnae*, *Petromyzon marinus*, and *Rhinatrema bivittatum* displayed almost even nucleotide usage (AT bias=0.5 and GC bias=0.5), and *Locusta migratoria*, *Brachionus plicatilis*, *Latalurus punctatus*, *Danio rerio*, and *Lepisosteus oculatus* are the next closest to the center than the rest of species (Figure 6). These interesting genetic phenomena indicate that the *atg13* gene could store its evolutionary trace at altering synonymous codon usages during species evolutions.

### The Resembling Evolutionary Trends for *atg13* Genes in Mammals at Synonymous Codon Usages

According to genetic characterizations reflected by synonymous codon usages (Table S2) and the overall codon usage bias (Figure 5), we next investigated the genetic diversity illustrated by 59 synonymous codon usages for these species in this study. Based on the classifications of mammals, *Sauropsida* animals, fishes, amphibians, and other organisms (the rest of the organisms in this study), the *atg13* genes of mammals and *Sauropsida* animals generally form two evolutionary groups, while *atg13* genes of amphibians appear to not form the resembling evolutionary trends (Figure 7). It can be clearly found that fish exhibit strong genetic dispersions at synonymous codon usages of the *atg13* gene. Moreover, *Brachionus calyciflorus*, *Brachionus plicatilis*, and *Brachionus rotundiformis* display the analogous evolutionary trends at synonymous codon usages of *atg13* gene, while *Locusta migratoria* exhibit its unique evolutionary trend relative to other species. The graphical representation of codon usage space for *atg13*

genes of different species further indicates that selective forces involved with specific-species evolutionary dynamic play important roles in the genetic development of the *atg13* gene at the synonymous codon usage patterns.

### Nucleotide Composition Distributions at Different Positions in Codon Acting in the *atg13* Gene Context

Based on the selective forces acting on evolutionary trends of different *atg13* genes at codon usages (Figures 5 and 6), we wanted to investigate the roles of nucleotide composition distribution at the specific codon position in the *atg13* gene. It is unsurprising that conventional nucleotide composition distributions involved with GC3 and AT3 skews play more important roles than those of GC and AT skews at the first and second codon positions in influencing the selection of the GC and AT compositions in the *atg13* gene (Table 3). However, variations of GC and AT skews do not completely uncover the genetic variations around the *atg13* gene. Therefore, we evaluated the genetic effects on the distributions of purine (A & G) and pyrimidine (T & C) compositions at different codon positions on the *atg13* gene. We found similar effects from the purine and pyrimidine composition distributions on nucleotide usage patterns at the first and second codon positions of the *atg13* gene. However, distributions of purine composition and pyrimidine composition do not exhibit the same roles in nucleotide usages at the third codon position (Table 3). Meanwhile, we estimated the roles of keto (T & G) and amino (A & C) skews at different codon positions in the *atg13* gene. Interestingly, both usage variations of nucleotides (T & G) and nucleotides (A & C) play more important roles at the third codon position than at the first and second codon positions of the *atg13* gene. These results strongly indicate that while the transition caused by nucleotide substitutions mainly acts on the nucleotide usages at the first and second codon positions, transversion

**TABLE 2** | Some synonymous codons are absent from *atg13* genes from different species.

Synonymous codon/Amino acid	Species
UUA(L)	<i>Callorhinchus milii</i> and <i>Chelonia mydas</i>
CUA(L)	<i>Astatotilapia calliptera</i> , <i>Callithrix jacchus</i> , <i>Castor Canadensis</i> , <i>Cercocebus atys</i> , <i>Chlorocebus sabaeus</i> , <i>Colobus angolensis palliatus</i> , <i>Cricetulus griseus</i> , <i>Elephantulus edwardii</i> , <i>Fundulus heteroclitus</i> , <i>Gadus morhua</i> , <i>Gorilla gorilla</i> , <i>Grammomys surdaster</i> , <i>Homo sapiens</i> , <i>Hylobates moloch</i> , <i>Macaca fascicularis</i> , <i>Macaca mulatta</i> , <i>Macaca nemestrina</i> , <i>Mandrillus leucophaeus</i> , <i>Mesocricetus auratus</i> , <i>Mus caroli</i> , <i>Mus musculus</i> , <i>Nomascus leucogenys</i> , <i>Oreochromis niloticus</i> , <i>Pan paniscus</i> , <i>Pan troglodytes</i> , <i>Papio Anubis</i> , <i>Peromyscus leucopus</i> , <i>Ptilocolobus tephrosceles</i> , <i>Pongo abelii</i> , <i>Rattus norvegicus</i> , <i>Rattus rattus</i> , <i>Rhinopithecus bieti</i> , <i>Rhinopithecus roxellana</i> and <i>Theropithecus gelada</i>
CUG(L)	<i>Brachionus calyciflorus</i>
AUU(I)	<i>Nothobranchius furzeri</i>
AUA(I)	<i>Petromyzon marinus</i>
GUC(V)	<i>Locusta migratoria</i>
GUA(V)	<i>Oryzias latipes</i>
UCG(S)	<i>Alligator mississippiensis</i> , <i>Alligator sinensis</i> , <i>Callithrix jacchus</i> , <i>Calypte anna</i> , <i>Castor canadensis</i> , <i>Cercocebus atys</i> , <i>Chlorocebus sabaeus</i> , <i>Colobus angolensis palliatus</i> , <i>Crocodylus porosus</i> , <i>Dipodomys ordii</i> , <i>Elephantulus edwardii</i> , <i>Felis catus</i> , <i>Galeopterus variegates</i> , <i>Gavialis gangeticus</i> , <i>Gorilla gorilla</i> , <i>Homo sapiens</i> , <i>Hylobates moloch</i> , <i>Lctalurus punctatus</i> , <i>Macaca fascicularis</i> , <i>Macaca mulatta</i> , <i>Macaca nemestrina</i> , <i>Mandrillus leucophaeus</i> , <i>Monopterus albus</i> , <i>Mus musculus</i> , <i>Nomascus leucogenys</i> , <i>Notechis scutatus</i> , <i>Pan paniscus</i> , <i>Pan troglodytes</i> , <i>Pantherophis guttatus</i> , <i>Papio Anubis</i> , <i>Ptilocolobus tephrosceles</i> , <i>Pongo abelii</i> , <i>Pseudonaja textilis</i> , <i>Python bivittatus</i> , <i>Rattus norvegicus</i> , <i>Rhincodon typus</i> , <i>Rhinopithecus bieti</i> , <i>Rhinopithecus roxellana</i> , <i>Thamnophis elegans</i> and <i>Theropithecus gelada</i>
CCG(P)	<i>Alligator mississippiensis</i> , <i>Alligator sinensis</i> , <i>Amblyraja radiata</i> , <i>Anolis carolinensis</i> , <i>Aquila chrysaetos chrysaetos</i> , <i>Calypte anna</i> , <i>Carlito syricta</i> , <i>Castor Canadensis</i> , <i>Columba livia</i> , <i>Corvus brachyrhynchos</i> , <i>Corvus cornix cornix</i> , <i>Crocodylus porosus</i> , <i>Dipodomys ordii</i> , <i>Dromaius</i> , <i>Egretta garzetta</i> , <i>Falco peregrinus</i> , <i>Gavialis gangeticus</i> , <i>Geospiza fortis</i> , <i>Haliaeetus albicilla</i> , <i>Meleagris gallopavo</i> , <i>Melopsittacus undulates</i> , <i>Myotis lucifugus</i> , <i>Numida meleagris</i> , <i>Opisthocomus hoazin</i> , <i>Parus major</i> , <i>Pelodiscus sinensis</i> , <i>Phasianus colchicus</i> , <i>Pogona vitticeps</i> , <i>Python bivittatus</i> , <i>Rhincodon typus</i> , <i>Serinus canaria</i> , <i>Strigops habroptila</i> , <i>Sturnus vulgaris</i> , <i>Trichechus manatus latirostris</i> , <i>Tyto alba</i> and <i>Zootoca vivipara</i>
ACG(T)	<i>Danio rerio</i> , <i>Microcaecilia unicolor</i> , <i>Monopterus albus</i> , <i>Pantherophis guttatus</i> , <i>Protobothrops mucrosquamatus</i> , <i>Pseudonaja textilis</i> , <i>Python bivittatus</i> and <i>Xenopus tropicalis</i>
GCG(A)	<i>Ailuropoda melanoleuca</i> , <i>Aptenodytes forsteri</i> , <i>Calypte anna</i> , <i>Canis lupus dingo</i> , <i>Canis lupus familiaris</i> , <i>Capra hircus</i> , <i>Carlito syricta</i> , <i>Castor canadensis</i> , <i>Chaetura pelagica</i> , <i>Chelonia mydas</i> , <i>Chelonoidis abingdonii</i> , <i>Chrysemys picta bellii</i> , <i>Colobus angolensis palliatus</i> , <i>Corvus brachyrhynchos</i> , <i>Corvus cornix cornix</i> , <i>Dasyopus novemcinctus</i> , <i>Delphinapterus leucas</i> , <i>Dromaius</i> , <i>Eumetopias jubatus</i> , <i>Falco peregrinus</i> , <i>Galeopterus variegates</i> , <i>Gallus gallus</i> , <i>Geospiza fortis</i> , <i>Gopherus evgoodei</i> , <i>Gorilla gorilla</i> , <i>Jaculus jaculus</i> , <i>Miniopterus natalensis</i> , <i>Monodelphis domestica</i> , <i>Myotis brandtii</i> , <i>Myotis davidii</i> , <i>Myotis lucifugus</i> , <i>Nannospalax gallii</i> , <i>Nipponia nippon</i> , <i>Notechis scutatus</i> , <i>Odobenus rosmarus</i> , <i>Opisthocomus hoazin</i> , <i>Pantherophis guttatus</i> , <i>Parus major</i> , <i>Pelecanus colchicus</i> , <i>Podarcis muralis</i> , <i>Pogona vitticeps</i> , <i>Pseudonaja textilis</i> , <i>Pteropus alecto</i> , <i>Pteropus vampyrus</i> , <i>Python bivittatus</i> , <i>Rhincodon typus</i> , <i>Rousettus aegyptiacus</i> , <i>Serinus canaria</i> , <i>Sturnus vulgaris</i> , <i>Sus scrofa</i> , <i>Terrapene carolina triunguis</i> , <i>Thamnophis elegans</i> , <i>Trachemys scripta elegans</i> , <i>Trichechus manatus latirostris</i> , <i>Tupaia chinensis</i> , <i>Ursus arctos horribilis</i> , <i>Ursus maritimus</i> , <i>Zalophus californianus</i> and <i>Zootoca vivipara</i>
UAU(Y)	<i>Chelonia mydas</i> and <i>Oryzias latipes</i>
UGU(C)	<i>Protobothrops mucrosquamatus</i>
UGC(C)	<i>Halichoerus grypus</i> , <i>Hipposideros armiger</i> , <i>Mirounga leonina</i> , <i>Rhincodon typus</i> and <i>Rhinolophus ferrumequinum</i>
CGC(R)	<i>Alligator mississippiensis</i> , <i>Alligator sinensis</i> , <i>Amblyraja radiata</i> , <i>Anas platyrhynchos</i> , <i>Anolis carolinensis</i> , <i>Aptenodytes forsteri</i> , <i>Aythya fuligula</i> , <i>Bos Taurus</i> , <i>Calypte anna</i> , <i>Chaetura pelagica</i> , <i>Columba livia</i> , <i>Cygnus atratus</i> , <i>Dromaius</i> , <i>Egretta garzetta</i> , <i>Fulmarus glacialis</i> , <i>Gavia stellata</i> , <i>Geotrypetes seraphini</i> , <i>Haliaeetus albicilla</i> , <i>Latimeria chalumnae</i> , <i>Locusta migratoria</i> , <i>Manis javanica</i> , <i>Melopsittacus undulates</i> , <i>Microcaecilia unicolor</i> , <i>Nipponia nippon</i> , <i>Oxyura jamaicensis</i> , <i>Paramormyrops kingsleyae</i> , <i>Pelecanus crispus</i> , <i>Podarcis muralis</i> , <i>Pogona vitticeps</i> , <i>Protobothrops mucrosquamatus</i> , <i>Pygoscelis adeliae</i> , <i>Rattus norvegicus</i> , <i>Rhinatrema bivittatum</i> , <i>Rhincodon typus</i> , <i>Strigops habroptila</i> , <i>Struthio camelus australis</i> , <i>Sturnus vulgaris</i> , <i>Tyto alba</i> and <i>Zootoca vivipara</i>
CGA(R)	<i>Fundulus heteroclitus</i> , <i>Gorilla gorilla</i> , <i>Loxodonta africana</i> , <i>Notothenia coriiceps</i> , <i>Ornithorhynchus anatinus</i> , <i>Pogona vitticeps</i> and <i>Sparus aurata</i>
CGG(R)	<i>Gekko japonicus</i> , <i>Lates calcarifer</i> , <i>Lepisosteus oculatus</i> , <i>Oncorhynchus kisutch</i> , <i>Oncorhynchus mykiss</i> , <i>Oncorhynchus nerka</i> , <i>Paralichthys olivaceus</i> , <i>Protobothrops mucrosquamatus</i> , <i>Rhinatrema bivittatum</i> , <i>Salmo salar</i> , <i>Salmo trutta</i> and <i>Salvelinus alpinus</i>
AGA(R)	<i>Protobothrops mucrosquamatus</i>

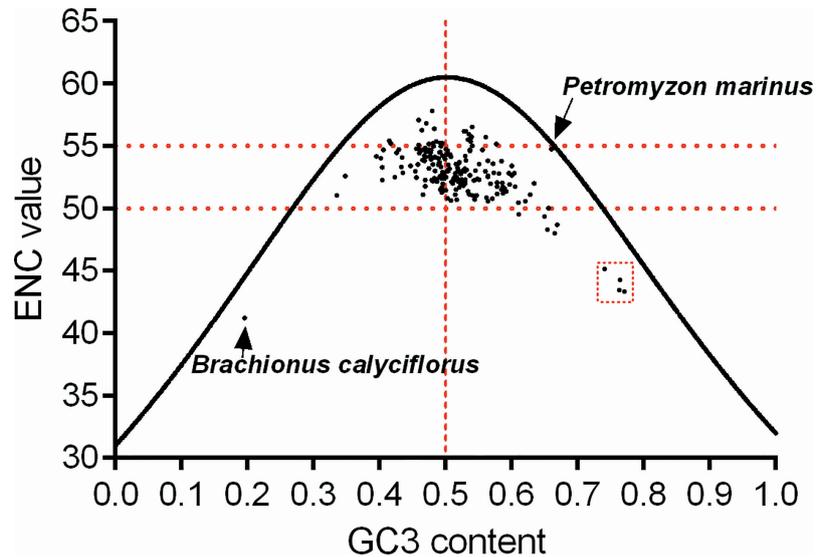
derived from nucleotide substitutions plays more important roles in the nucleotide usages at the third codon position of the *atg13* gene.

## DISCUSSION

Synonymous codons are great representations of non-randomness in coding sequences, playing an important role in many different cellular processes. One of the most studied synonymous codon usages investigates the mRNA decay process and protein folding past translation. (Precup and Parker, 1987; Pop et al., 2014; Presnyak et al., 2015; Hia et al., 2019; Guimaraes et al., 2020).

Autophagy, on the other hand, is one fundamental cellular process across eukaryotes, induced by various stimuli including environmental stressors which can be regarded as one selective force for cellular remodeling during organism development (Penalzo et al., 2006; Di Bartolomeo et al., 2010).

With the *atg13* gene being one essential factor for the formation of autophagy, its species-specific nucleotide usage in eukaryotic species has attracted attention from researchers around the world. In this study, we selected 226 eukaryotic species from invertebrates to vertebrates to uncover evolutionary traces involved in the interplay between nucleotide and synonymous codon usages. To our knowledge, this is the first

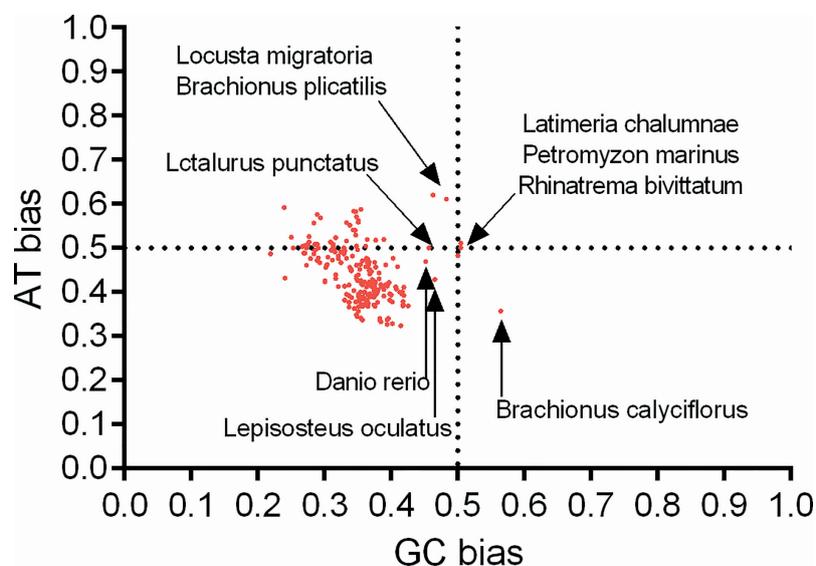


**FIGURE 5** | The relationship between ENC and GC content at the third synonymous codon position (GC3) in *atg13* genes of 226 eukaryotic organisms. The continuous curve line displays the expected codon usage if GC compositional constraints alone account for codon usage bias. The red broken lines highlight the distribution of *agt13* genes in this plot of ENC vs. GC3 content. In addition, the red box with broken lines highlights the outliers which correspond to *atg13* genes of the specific organisms.

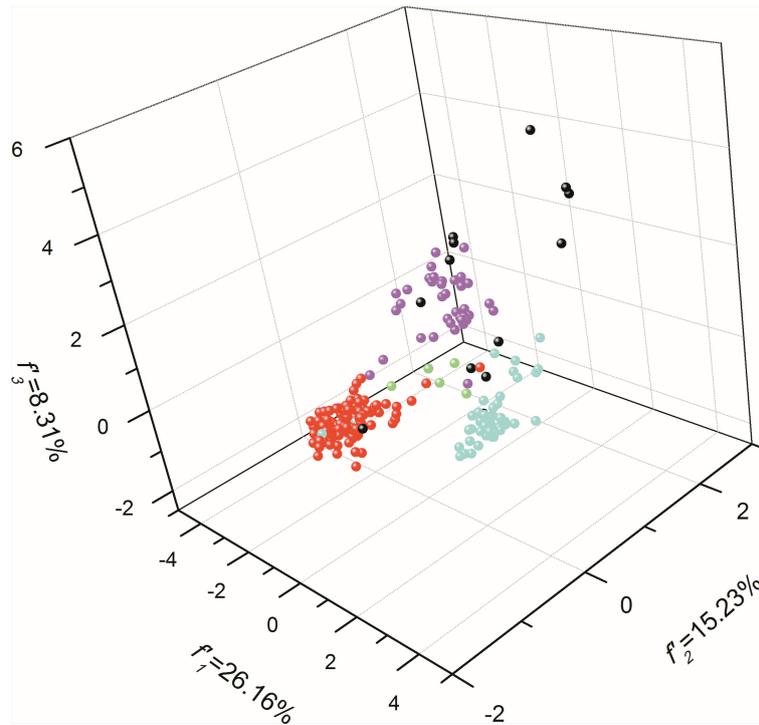
genome-wide investigation of *atg13* genes across eukaryotic organisms at the nucleotide, synonymous codon, and amino acid levels. All the target genes were separated into groups and were observed to have explicit orthologous relationships at the nucleotide and amino acid levels (Figures 1 and 2), implying that the *atg13* genes in either invertebrates or vertebrates share conserved functions from their common ancestors. Generally, nucleotide usages in the *atg13* gene exhibit more genetic

information on its evolutionary trend than amino acid usages, strongly suggesting that synonymous codon usages serve as a genetic bridge between nucleotide usage and amino acid usages store additional evolutionary traces for the *atg13* gene.

After all, all proteins primarily derived from nucleotide organization are encoded by the corresponding transcript sequences. The previous studies reported that the *atg* genes owned different intron/exon organizations from lower



**FIGURE 6** | Parity rule 2 bias plot for *atg13* genes of 226 eukaryotic organisms.



**FIGURE 7** | The plot for the overall codon usage visualized by PCA method for *atg13* genes of 226 eukaryotic organisms. The red dots mean mammals, the blue dots stand for *Sauropsida* animals, the purple dots mean fishes, the green dots mean amphibians, and the black dots correspond to the rest eukaryotic organisms in this study.

organisms to higher ones (Ohsumi, 2014; Kang et al., 2018). Based on different strengths of compact genomic structures of *atg13* genes of different organisms, synonymous codon usage patterns should store additional evolutionary scenarios for different eukaryotic species. The interplay between nucleotide usage pattern and amino acid composition is considered a

central platform for biological processes and probably plays an important role in the evolutionary development of the protein-based life cycle (Nicholson et al., 2018).

Specific identification of autophagy-related structures participates in clarifying autophagic activity, and clarification of autophagy-related genes plays an important role in identifying

**TABLE 3** | The correlations between the overall nucleotide skew and the nucleotide skew at the specific codon position at the gene level.

	<b>GC1 skew</b>	<b>GC2 skew</b>	<b>GC3 skew</b>
<b>GC skew</b>	r=0.398***	r=0.448***	r=0.606***
	<b>AT1 skew</b>	<b>AT2 skew</b>	<b>AT3 skew</b>
<b>AT skew</b>	r=0.071 <sup>NS</sup>	r=0.587***	r=0.888***
	<b>Purine1 skew</b>	<b>Purine2 skew</b>	<b>Purine3 skew</b>
<b>Purine skew</b>	r=0.862***	r=0.638***	r=0.939***
	<b>Pyrimidine1 skew</b>	<b>Pyrimidine2 skew</b>	<b>Pyrimidine3 skew</b>
<b>Pyrimidine skew</b>	r=0.859***	r=0.681***	r=-0.293***
	<b>Keto1 skew</b>	<b>Keto2 skew</b>	<b>Keto3 skew</b>
<b>Keto skew</b>	r=0.669***	r=0.618***	r=0.956***
	<b>Amino1 skew</b>	<b>Amino2 skew</b>	<b>Amino3 skew</b>
<b>Amino skew</b>	r=0.834***	r=0.559***	r=0.971***

\*\*\* $p < 0.001$ , <sup>NS</sup> $p > 0.05$ . NS, non-significance.

the requirements of autophagy in different animals (Kuma et al., 2017). From what we have found, the bias extents of nucleotide usages at different codon positions show that the degrees of bias at the third codon position of the *atg13* genes are more variant than those at other positions (Figure 3). While nucleotide usages at the third codon position reflect the existence of broad distributions of the fitness effect of synonymous mutations in the *atg13* genes during their evolution, the first and second codon positions are influenced by the conserved functions of ATG13 protein which directly dominate amino acid usages. A similar pattern can also be found during mRNA splicing where the nucleotide usages at the third codon position can be influenced by translation-independent selective pressures (Wu and Hurst, 2015). Interestingly, nuclear STAT3 fine-tunes autophagy via the transcriptional regulation of several autophagy-related genes (You et al., 2015), strongly suggesting that some regulators involved in various cellular signaling pathways possibly interact with *atg* genes at the transcriptional level. Among the usage bias extents for each nucleotide in the *atg13* genes (Figure 4), the nucleotide 'C' displays a relatively convergent evolution for most species, while nucleotide 'A' usage exhibits an obviously dispersed pattern. Although adenine is not the active component of protein-nucleotide binding, it serves as a 'molecular handle' which enhances binding specificity and affinity (Cai et al., 2009; Hong et al., 2009; Rogne et al., 2018). Based on the studies mentioned above, adenine is a potentially promising candidate for analyzing the evolutionary development of molecular recognition in proteins (Narunsky et al., 2020).

The strong selective forces acting on usage biases of different kinds of nucleotide contribute to synonymous codon usage bias for the *atg13* genes. Despite the divergent evolutionary pathways of the *atg13* genes across eukaryotes at synonymous codon usages (Table S2), some synonymous codons with conserved usage patterns retain in the *atg13* genes (Table 1), strongly suggesting that these synonymous codons with the conserved usage patterns benefit the *atg13* gene for sustaining the normal biological functions. Of note, the *atg13* gene and the unc-51 like autophagy activating kinase 1 (Ulk1) are the crucial substrates of the target of rapamycin complex 1 (TORC1) in autophagy (Noda, 2017), suggesting that the *atg13* gene requires one conserved sequence structure for sustaining its correct protein properties. These results also imply that synonymous codon usage patterns for the *atg13* gene are most likely to fit in the species-specific evolutionary pathway with conserved biological functions and fine-tune translation selection. Moreover, synonymous codons are not uniformly represented in the transcriptome; the observed codon bias has coevolved with tRNA abundances under selection for translation accuracy and efficiency (Ikemura, 1981; Bulmer, 1991; Drummond and Wilke, 2008; Guimaraes et al., 2020). As a result, codon usage impacts gene expression on both translation efficiency and mRNA decay process (Presnyak et al., 2015; Hanson and Collier, 2018).

To provide a reference for the *atg13* gene of *Petromyzon marinus* at the overall codon usage vs. GC3 content, we clarified the evolutionary dynamics derived from natural selection for driving the evolution of the *atg13* genes (Figure 5). Furthermore,

RP2 plot analyses prove that natural selection plays a key role in the evolutionary pathway of the *atg13* gene, namely for the *atg13* genes of *Latimeria chalumnae*, *Petromyzon marinus*, and *Rhinatrema bivittatum* at synonymous codon usage patterns for Ala, Gly, Pro, Thr, and Val (Figure 6). This is, to our knowledge, the first report of general codon usage for *atg13* genes of different species affected by natural selection. We assume the overall codon usage pattern to be quite stable over the course of evolution since switching to a new codon usage pattern would impose a high genetic load by simultaneously modifying the selection coefficient at many synonymous positions. Since ATG13 protein plays an important role in the maintenance of cell viability under starvation conditions, *atg13* mutations, along with other transcript variants of the *atg13* gene, were defective in autophagy (Funakoshi et al., 1997; Stephan et al., 2009). Moreover, variations in synonymous codon usage are abundant across multiple levels of organization: between codons for one amino acid, between genes in a genome, and between genomes of different species (Labella et al., 2019). Here, the evolution of the *atg13* genes in different eukaryotic species represents the species-specific model at synonymous codon usages (Figure 7), strongly suggesting that genetic changes of synonymous codon usage in the *atg13* gene benefit the corresponding protein products with functional fit to perform the proper biological functions in specific species. Given that genetic changes in genes can alter phenotypes and even result in diseases, the synonymous codon usage bias has a strong correlation with protein biological functions (Chakraborty et al., 2017; Chakraborty et al., 2019; De Mandal et al., 2020; Malakar et al., 2020). With the ongoing analyses for the genetic signatures of *atg* genes, some members have been regarded as markers for predicting some diseases (such as cancers) (Baradaran Ghavami et al., 2019; Chen et al., 2020; Yang et al., 2020; Lai et al., 2020; Wang et al., 2020). Despite a weak pressure caused by translation selection, the selective force from natural selection plays an obvious role in GC-conservative codons in different organisms, leading to alterations of the gene structures (Carlini et al., 2001; Carlini and Stephan, 2003; Galtier et al., 2018). Based on the distributions of different nucleotide compositions with respect to the nucleotide context of the *atg13* gene (Table 3), it can be found that selective forces from natural selection are more significant than mutation pressure from nucleotide composition in terms of affecting the synonymous codon usage. It is worth noting that adaptive evolution caused by nucleotide compositional distribution and protein property also plays an important role in genetic diversity among genomes (Foerster et al., 2005; Reichenberger et al., 2015).

In conclusion, based on the strong evidence of nucleotide and synonymous codon usage biases, we find the imbalance between selective forces from natural selection and mutation pressure from nucleotide composition constraint on the evolution of *atg13* genes of different species. During the formation of synonymous codon usages for the *atg13* genes, nucleotide usage patterns at the first and second codon positions are influenced by amino acid composition constraint, while nucleotide usage patterns at the third codon position are

affected by translation selection. Of note, natural selection is a significant determinant of codon usage patterns in *atg13* genes across eukaryotic organisms, but mutation pressure still plays an important role in the formation of codon usage of this gene in ancestry animals (such as *Petromyzon marinus*).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

YL, RW, and FP collected the related sequence data from NCBI website. HW, XF, and LJ analyzed and calculated the related data. ZM and XM designed and wrote the manuscript. All

authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Natural Science Foundation of Gansu Province (No. 20JR5RA505), the Innovative Research Team in University (No. IRT\_17R88), and the Center Universities Foundation supported by Northwest Minzu University (No. 31920200073).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2021.771010/full#supplementary-material>

## REFERENCES

- Ashford, T. P., and Porter, K. R. (1962). Cytoplasmic Components in Hepatic Cell Lysosomes. *J. Cell Biol.* 12, 198–202. doi: 10.1083/jcb.12.1.198
- Baradaran Ghavami, S., Kabiri, F., Nourian, M., Balaii, H., Shahrokh, S., Chaleshi, V., et al. (2019). Association Between Variants of the Autophagy Related Gene ATG16L1 in Inflammatory Bowel Diseases and Clinical Statuses. *Gastroenterol. Hepatol. Bed Bench* 12, S94–s100.
- Brule, C. E., and Grayhack, E. J. (2017). Synonymous Codons: Choose Wisely for Expression. *Trends Genet.* 33, 283–297. doi: 10.1016/j.tig.2017.02.001
- Bulmer, M. (1991). The Selection-Mutation-Drift Theory of Synonymous Codon Usage. *Genetics* 129, 897–907. doi: 10.1093/genetics/129.3.897
- Butt, A. M., Nasrullah, I., Qamar, R., and Tong, Y. (2016). Evolution of Codon Usage in Zika Virus Genomes Is Host and Vector Specific. *Emerg. Microbes Infect.* 5, e107. doi: 10.1038/emi.2016.106
- Cai, S., Li, Q. S., Fang, J., Borchardt, R. T., Kuczera, K., Middaugh, C. R., et al. (2009). The Rationale for Targeting the NAD/NADH Cofactor Binding Site of Parasitic S-Adenosyl-L-Homocysteine Hydrolase for the Design of Anti-Parasitic Drugs. *Nucleosides Nucleotides Nucleic Acids* 28, 485–503. doi: 10.1080/15257770903051031
- Carlini, D. B. (2005). Context-Dependent Codon Bias and Messenger RNA Longevity in the Yeast Transcriptome. *Mol. Biol. Evol.* 22, 1403–1411. doi: 10.1093/molbev/msi135
- Carlini, D. B., Chen, Y., and Stephan, W. (2001). The Relationship Between Third-Codon Position Nucleotide Content, Codon Bias, mRNA Secondary Structure and Gene Expression in the Drosophilid Alcohol Dehydrogenase Genes *Adh* and *Adhr*. *Genetics* 159, 623–633. doi: 10.1093/genetics/159.2.623
- Carlini, D. B., and Stephan, W. (2003). In Vivo Introduction of Unpreferred Synonymous Codons Into the Drosophila *Adh* Gene Results in Reduced Levels of ADH Protein. *Genetics* 163, 239–243. doi: 10.1093/genetics/163.1.239
- Chakraborty, S., Mazumder, T. H., and Uddin, A. (2019). Compositional Dynamics and Codon Usage Pattern of BRCA1 Gene Across Nine Mammalian Species. *Genomics* 111, 167–176. doi: 10.1016/j.ygeno.2018.01.013
- Chakraborty, S., Uddin, A., and Choudhury, M. N. (2017). Factors Affecting the Codon Usage Bias of SRY Gene Across Mammals. *Gene* 630, 13–20. doi: 10.1016/j.gene.2017.08.003
- Chen, H., Deng, Q., Wang, W., Tao, H., and Gao, Y. (2020). Identification of an Autophagy-Related Gene Signature for Survival Prediction in Patients With Cervical Cancer. *J. Ovarian Res.* 13, 131. doi: 10.1186/s13048-020-00730-8
- Comeron, J. M., and Agudé, M. (1998). An Evaluation of Measures of Synonymous Codon Usage Bias. *J. Mol. Evol.* 47, 268–274. doi: 10.1007/pl00006384
- De Mandal, S., Mazumder, T. H., Panda, A. K., Kumar, N. S., and Jin, F. (2020). Analysis of Synonymous Codon Usage Patterns of HPRT1 Gene Across Twelve Mammalian Species. *Genomics* 112, 304–311. doi: 10.1016/j.ygeno.2019.02.010
- Di Bartolomeo, S., Nazio, F., and Cecconi, F. (2010). The Role of Autophagy During Development in Higher Eukaryotes. *Traffic* 11, 1280–1289. doi: 10.1111/j.1600-0854.2010.01103.x
- Drummond, D. A., and Wilke, C. O. (2008). Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* 134, 341–352. doi: 10.1016/j.cell.2008.05.042
- Fan, X., Han, S., Yan, D., Gao, Y., Wei, Y., Liu, X., et al. (2017). Foot-And-Mouth Disease Virus Infection Suppresses Autophagy and NF- $\kappa$ B Antiviral Responses via Degradation of ATG5-ATG12 by 3C(Pro). *Cell Death Dis.* 8, e2561. doi: 10.1038/cddis.2016.489
- Fedorov, A., Saxonov, S., and Gilbert, W. (2002). Regularities of Context-Dependent Codon Bias in Eukaryotic Genes. *Nucleic Acids Res.* 30, 1192–1197. doi: 10.1093/nar/30.5.1192
- Foerster, K. U., Von Mering, C., Hooper, S. D., and Bork, P. (2005). Environments Shape the Nucleotide Composition of Genomes. *EMBO Rep.* 6, 1208–1213. doi: 10.1038/sj.embor.7400538
- Fujioka, Y., Suzuki, S. W., Yamamoto, H., Kondo-Kakuta, C., Kimura, Y., Hirano, H., et al. (2014). Structural Basis of Starvation-Induced Assembly of the Autophagy Initiation Complex. *Nat. Struct. Mol. Biol.* 21, 513–521. doi: 10.1038/nsmb.2822
- Funakoshi, T., Matsuura, A., Noda, T., and Ohsumi, Y. (1997). Analyses of APG13 Gene Involved in Autophagy in Yeast, *Saccharomyces Cerevisiae*. *Gene* 192, 207–213. doi: 10.1016/s0378-1119(97)00031-0
- Galtier, N., Roux, C., Rousselle, M., Romiguier, J., Figueat, E., Glémin, S., et al. (2018). Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion. *Mol. Biol. Evol.* 35, 1092–1103. doi: 10.1093/molbev/msy015
- Ge, Z., Li, X., Cao, X., Wang, R., Hu, W., Gen, L., et al. (2020). Viral Adaptation of Staphylococcal Phage: A Genome-Based Analysis of the Selective Preference Based on Codon Usage Bias. *Genomics* 112, 4657–4665. doi: 10.1016/j.ygeno.2020.08.012
- Guimaraes, J. C., Mittal, N., Gnann, A., Jedlinski, D., Riba, A., Buczak, K., et al. (2020). A Rare Codon-Based Translational Program of Cell Proliferation. *Genome Biol.* 21, 44. doi: 10.1186/s13059-020-1943-5
- Gun, L., Haixian, P., Yumiao, R., Han, T., Jingqi, L., and Liguang, Z. (2018). Codon Usage Characteristics of PB2 Gene in Influenza A H7N9 Virus From Different Host Species. *Infect. Genet. Evol.* 65, 430–435. doi: 10.1016/j.meegid.2018.08.028

- Gupta, S., Paul, K., and Roy, A. (2021). Codon Usage Signatures in the Genus *Cryptococcus*: A Complex Interplay of Gene Expression, Translational Selection and Compositional Bias. *Genomics* 113, 821–830. doi: 10.1016/j.ygeno.2020.10.013
- Hanson, G., and Coller, J. (2018). Codon Optimality, Bias and Usage in Translation and mRNA Decay. *Nat. Rev. Mol. Cell Biol.* 19, 20–30. doi: 10.1038/nrm.2017.91
- Harding, T. M., Morano, K. A., Scott, S. V., and Klionsky, D. J. (1995). Isolation and Characterization of Yeast Mutants in the Cytoplasm to Vacuole Protein Targeting Pathway. *J. Cell Biol.* 131, 591–602. doi: 10.1083/jcb.131.3.591
- Hia, F., Yang, S. F., Shichino, Y., Yoshinaga, M., Murakawa, Y., Vandenbon, A., et al. (2019). Codon Bias Confers Stability to Human mRNAs. *EMBO Rep.* 20, e48220. doi: 10.15252/embr.201948220
- Hong, J. A., Bhave, D. P., and Carroll, K. S. (2009). Identification of Critical Ligand Binding Determinants in Mycobacterium Tuberculosis Adenosine-5'-Phosphosulfate Reductase. *J. Med. Chem.* 52, 5485–5495. doi: 10.1021/jm900728u
- Ikemura, T. (1981). Correlation Between the Abundance of Escherichia Coli Transfer RNAs and the Occurrence of the Respective Codons in Its Protein Genes: A Proposal for a Synonymous Codon Choice That Is Optimal for the E. Coli Translational System. *J. Mol. Biol.* 151, 389–409. doi: 10.1016/0022-2836(81)90003-6
- Kamada, Y., Funakoshi, T., Shintani, T., Nagano, K., Ohsumi, M., and Ohsumi, Y. (2000). Tor-Mediated Induction of Autophagy via an Apg1 Protein Kinase Complex. *J. Cell Biol.* 150, 1507–1513. doi: 10.1083/jcb.150.6.1507
- Kamada, Y., Yoshino, K., Kondo, C., Kawamata, T., Oshiro, N., Yonezawa, K., et al. (2010). Tor Directly Controls the Atg1 Kinase Complex to Regulate Autophagy. *Mol. Cell Biol.* 30, 1049–1058. doi: 10.1128/mcb.01344-09
- Kang, H. M., Lee, J. S., Kim, M. S., Lee, Y. H., Jung, J. H., Hagiwara, A., et al. (2018). Genome-Wide Identification of 99 Autophagy-Related (Atg) Genes in the Monogonont Rotifer Brachionus Spp. And Transcriptional Modulation in Response to Cadmium. *Aquat Toxicol.* 201, 73–82. doi: 10.1016/j.aquatox.2018.05.021
- Komar, A. A., Lesnik, T., and Reiss, C. (1999). Synonymous Codon Substitutions Affect Ribosome Traffic and Protein Folding During In Vitro Translation. *FEBS Lett.* 462, 387–391. doi: 10.1016/s0014-5793(99)01566-5
- Kroemer, G., and Levine, B. (2008). Autophagic Cell Death: The Story of a Misnomer. *Nat. Rev. Mol. Cell Biol.* 9, 1004–1010. doi: 10.1038/nrm2529
- Kuma, A., Komatsu, M., and Mizushima, N. (2017). Autophagy-Monitoring and Autophagy-Deficient Mice. *Autophagy* 13, 1619–1628. doi: 10.1080/15548627.2017.1343770
- Labella, A. L., Oputente, D. A., Steenwyk, J. L., Hittinger, C. T., and Rokas, A. (2019). Variation and Selection on Codon Usage Bias Across an Entire Subphylum. *PLoS Genet.* 15, e1008304. doi: 10.1371/journal.pgen.1008304
- Lai, J., Chen, B., Mok, H., Zhang, G., Ren, C., and Liao, N. (2020). Comprehensive Analysis of Autophagy-Related Prognostic Genes in Breast Cancer. *J. Cell Mol. Med.* 24, 9145–9153. doi: 10.1111/jcmm.15551
- Liu, X. Y., Li, Y., Ji, K. K., Zhu, J., Ling, P., Zhou, T., et al. (2020). Genome-Wide Codon Usage Pattern Analysis Reveals the Correlation Between Codon Usage Bias and Gene Expression in *Cuscuta Australis*. *Genomics* 112, 2695–2702. doi: 10.1016/j.ygeno.2020.03.002
- Malakar, A. K., Halder, B., Paul, P., Deka, H., and Chakraborty, S. (2020). Genetic Evolution and Codon Usage Analysis of NKX-2.5 Gene Governing Heart Development in Some Mammals. *Genomics* 112, 1319–1329. doi: 10.1016/j.ygeno.2019.07.023
- Ma, P., Li, L., Jin, L., Zhang, D., Cao, X., Guo, F., et al. (2020). Antiviral Responses of ATG13 to the Infection of Peste Des Petits Ruminants Virus Through Activation of Interferon Response. *Gene* 754, 144858. doi: 10.1016/j.gene.2020.144858
- Nabiyouni, M., Prakash, A., and Fedorov, A. (2013). Vertebrate Codon Bias Indicates a Highly GC-Rich Ancestral Genome. *Gene* 519, 113–119. doi: 10.1016/j.gene.2013.01.033
- Nakatogawa, H., Suzuki, K., Kamada, Y., and Ohsumi, Y. (2009). Dynamics and Diversity in Autophagy Mechanisms: Lessons From Yeast. *Nat. Rev. Mol. Cell Biol.* 10, 458–467. doi: 10.1038/nrm2708
- Narunsky, A., Kessel, A., Solan, R., Alva, V., Kolodny, R., and Ben-Tal, N. (2020). On the Evolution of Protein-Adenine Binding. *Proc. Natl. Acad. Sci. U. S. A.* 117, 4701–4709. doi: 10.1073/pnas.1911349117
- Nicholson, D. A., Sengupta, A., Sung, H. L., and Nesbitt, D. J. (2018). Amino Acid Stabilization of Nucleic Acid Secondary Structure: Kinetic Insights From Single-Molecule Studies. *J. Phys. Chem. B* 122, 9869–9876. doi: 10.1021/acs.jpcc.8b06872
- Noda, T. (2017). Regulation of Autophagy Through TORC1 and Mtorc1. *Biomolecules* 7 (3), 52. doi: 10.3390/biom7030052
- Novoa, E. M., and Ribas De Pouplana, L. (2012). Speeding With Control: Codon Usage, tRNAs, and Ribosomes. *Trends Genet.* 28, 574–581. doi: 10.1016/j.tig.2012.07.006
- Ohsumi, Y. (2014). Historical Landmarks of Autophagy Research. *Cell Res.* 24, 9–23. doi: 10.1038/cr.2013.169
- Pedrotti, M., and Guerra, G. F. (1989). Semi-Rigid Permeable Contact Lenses. Our Experience With Extended Wear. *Ophthalmologie* 3, 260–261.
- Penaloza, C., Lin, L., Lockshin, R. A., and Zakeri, Z. (2006). Cell Death in Development: Shaping the Embryo. *Histochem. Cell Biol.* 126, 149–158. doi: 10.1007/s00418-006-0214-1
- Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., et al. (2014). Causal Signals Between Codon Bias, mRNA Structure, and the Efficiency of Translation and Elongation. *Mol. Syst. Biol.* 10, 770. doi: 10.15252/msb.20145524
- Precup, J., and Parker, J. (1987). Missense Misreading of Asparagine Codons as a Function of Codon Identity and Context. *J. Biol. Chem.* 262, 11351–11355. doi: 10.1016/S0021-9258(18)60966-4
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* 160, 1111–1124. doi: 10.1016/j.cell.2015.02.029
- Radoshevich, L., Murrow, L., Chen, N., Fernandez, E., Roy, S., Fung, C., et al. (2010). ATG12 Conjugation to ATG3 Regulates Mitochondrial Homeostasis and Cell Death. *Cell* 142, 590–600. doi: 10.1016/j.cell.2010.07.018
- Reichenberger, E. R., Rosen, G., Hershberg, U., and Hershberg, R. (2015). Prokaryotic Nucleotide Composition Is Shaped by Both Phylogeny and the Environment. *Genome Biol. Evol.* 7, 1380–1389. doi: 10.1093/gbe/evv063
- Rogne, P., Rosselin, M., Grundström, C., Hedberg, C., Sauer, U. H., and Wolf-Watz, M. (2018). Molecular Mechanism of ATP Versus GTP Selectivity of Adenylate Kinase. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3012–3017. doi: 10.1073/pnas.1721508115
- Saunders, R., and Deane, C. M. (2010). Synonymous Codon Usage Influences the Local Protein Structure Observed. *Nucleic Acids Res.* 38, 6719–6728. doi: 10.1093/nar/gkq495
- Sharp, P. M., and Li, W. H. (1986). Codon Usage in Regulatory Genes in *Escherichia Coli* Does Not Reflect Selection for 'Rare' Codons. *Nucleic Acids Res.* 14, 7737–7749. doi: 10.1093/nar/14.19.7737
- Stephan, J. S., Yeh, Y. Y., Ramachandran, V., Deminoff, S. J., and Herman, P. K. (2009). The Tor and PKA Signaling Pathways Independently Target the Atg1/Atg13 Protein Kinase Complex to Control Autophagy. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17049–17054. doi: 10.1073/pnas.0903316106
- Sueoka, N. (1995). Intrastrand Parity Rules of DNA Base Composition and Usage Biases of Synonymous Codons. *J. Mol. Evol.* 40, 318–325. doi: 10.1007/bf00163236
- Sueoka, N. (1999). Translation-Coupled Violation of Parity Rule 2 in Human Genes Is Not the Cause of Heterogeneity of the DNA G+C Content of Third Codon Position. *Gene* 238, 53–58. doi: 10.1016/s0378-1119(99)00320-0
- Thumm, M., Egner, R., Koch, B., Schlumpberger, M., Straub, M., Veenhuis, M., et al. (1994). Isolation of Autophagocytosis Mutants of *Saccharomyces Cerevisiae*. *FEBS Lett.* 349, 275–280. doi: 10.1016/0014-5793(94)00672-5
- Tillier, E. R., and Collins, R. A. (2000). The Contributions of Replication Orientation, Gene Direction, and Signal Sequences to Base-Composition Asymmetries in Bacterial Genomes. *J. Mol. Evol.* 50, 249–257. doi: 10.1007/s002399910029
- Tsukada, M., and Ohsumi, Y. (1993). Isolation and Characterization of Autophagy-Defective Mutants of *Saccharomyces Cerevisiae*. *FEBS Lett.* 333, 169–174. doi: 10.1016/0014-5793(93)80398-e
- Wang, X., Xu, Y., Li, T., Chen, B., and Yang, W. (2020). Development of Prognosis Model for Colon Cancer Based on Autophagy-Related Genes. *World J. Surg. Oncol.* 18, 285. doi: 10.1186/s12957-020-02061-w
- Weygand-Durasevic, I., and Ibbá, M. (2010). Cell Biology. New Roles for Codon Usage. *Science* 329, 1473–1474. doi: 10.1126/science.1195567
- Wright, F. (1990). The 'Effective Number of Codons' Used in a Gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

- Wu, X., and Hurst, L. D. (2015). Why Selection Might Be Stronger When Populations Are Small: Intron Size and Density Predict Within and Between-Species Usage of Exonic Splice Associated Cis-Motifs. *Mol. Biol. Evol.* 32, 1847–1861. doi: 10.1093/molbev/msv069
- Yamamoto, H., Fujioka, Y., Suzuki, S. W., Noshiro, D., Suzuki, H., Kondo-Kakuta, C., et al. (2016). The Intrinsically Disordered Protein Atg13 Mediates Supramolecular Assembly of Autophagy Initiation Complexes. *Dev. Cell* 38, 86–99. doi: 10.1016/j.devcel.2016.06.015
- Yang, C., Wang, Y., Xue, W., Xie, Y., Dong, Q., and Zhu, C. (2020). Competing Endogenous RNA (ceRNA) Network Analysis of Autophagy-Related Genes in Hepatocellular Carcinoma. *Pharmacogenomics Pers. Med.* 13, 445–462. doi: 10.2147/pgpm.S267563
- You, L., Wang, Z., Li, H., Shou, J., Jing, Z., Xie, J., et al. (2015). The Role of STAT3 in Autophagy. *Autophagy* 11, 729–739. doi: 10.1080/15548627.2015.1017192
- Zhou, J. H., Gao, Z. L., Zhang, J., Ding, Y. Z., Stipkovits, L., Szathmary, S., et al. (2013a). The Analysis of Codon Bias of Foot-and-Mouth Disease Virus and the Adaptation of This Virus to the Hosts. *Infect. Genet. Evol.* 14, 105–110. doi: 10.1016/j.meegid.2012.09.020
- Zhou, T., Weems, M., and Wilke, C. O. (2009). Translationally Optimal Codons Associate With Structurally Sensitive Sites in Proteins. *Mol. Biol. Evol.* 26, 1571–1580. doi: 10.1093/molbev/msp070
- Zhou, J. H., You, Y. N., Chen, H. T., Zhang, J., Ma, L. N., Ding, Y. Z., et al. (2013b). The Effects of the Synonymous Codon Usage and tRNA Abundance on Protein Folding of the 3C Protease of Foot-and-Mouth Disease Virus. *Infect. Genet. Evol.* 16, 270–274. doi: 10.1016/j.meegid.2013.02.017

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Wang, Wang, Pu, Feng, Jin, Ma and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.