American Journal of Biochemistry and Biotechnology

Original Research Paper

Comparative Functional Analysis of the Basic Helix-Loop-Helix Proteins in the Clawed Frogs' Genomes With Common Essential Pathways and Enriched Gene Ontology Terms

[1, 2]Wuyi Liu

[1]*Biological and Food Engineering Faculty,*
Anhui Province Key Laboratory of Environmental Hormone and Reproduction, Fuyang Normal University, China
[2]*Department of Science and Technology, Fuyang Normal University, China*

Article history
Received: 31-12-2018
Revised: 13-02-2019
Accepted: 27-03-2019

Corresponding Author:
Wuyi Liu
Biological and Food Engineering Faculty, Anhui Province Key Laboratory of Environmental Hormone and Reproduction, Fuyang Normal University, Fuyang City, China
Email: lwycau@163.com

Abstract: The international Gene Ontology (GO) and pathway databases were used to functionally analyze the clawed frogs' Basic Helix-Loop-Helix (bHLH) transcription factors of *Xenopus tropicalis* and *Xenopus laevis* in a updated genome-wide survey. There were 41 GO terms and one pathway significantly enriched for *Xenopus tropicalis*, whereas there were 45 GO terms and 3 pathways significantly enriched for *Xenopus laevis*. Among those significantly enriched GO terms, the two clawed frogs share 31 common functional GO annotations of these bHLH genes, including DNA-dependent transcription and (negative) transcription regulation, DNA binding and bHLH binding, transcription factor complex and protein heterodimerization activity, (negative) regulation of RNA metabolic processes, nuclear translocator and repressor, myogenic basic muscle-specific protein, neurogenic differentiation factor and NeuroD. Furthermore, these frogs' bHLH genes were also found to play important roles in the regulation of gene expression in some important developmental or physiological processes, such as (skeletal) muscle cell differentiation, muscle organ development, biological rhythms and rhythmic process, hypoxia (adaption) and hypoxia-inducible factors, neurogenesis, neural tube development and neurogenic differentiation, whereas they were commonly significantly enriched in TGF-beta signaling pathway. These resulted data and information are very important for us to understand the functions, classification and evolution of frog bHLH genes.

Keywords: Basic Helix-Loop-Helix Transcription Factor, Functional Annotation, Pathway, Gene Ontology, Clawed Frog

Introduction

The Basic Helix-Loop-Helix (bHLH) proteins are currently recognized as the most important class of transcription factors. They can form specific interactions with the genetic cis-elements of eukaryotes, thereby activating or inhibiting the transcription and translation of the gene and they may also bind to the DNA binding proteins with activation or inhibition activities (Murre *et al*., 1989; Murre, 2019). They can combine with other transcription factors to form a complex genetic regulatory network too (Murre *et al*., 1989; Atchley and Fitch, 1997; Boggon *et al*., 1999; Luscombe *et al*., 2000; Riechmann *et al*., 2000; Stevens *et al*., 2008; Murre, 2019). At present, members of the bHLH transcription factor family are found to be crucial and they play many important roles in the cell proliferation and differentiation, body immunity, muscle tissue formation, neurons, resistance to stress, development of the eye and intestine, hematopoietic function and coagulation function, adaptation to hypoxic environment, sex determination and the process of genetic development of animals and plants (Murre *et al*., 1989; Murre, 2019; Atchley and Fitch, 1997; Boggon *et al*., 1999; Luscombe *et al*., 2000: Riechmann *et al*., 2000; Stevens *et al*., 2008; Wang *et al*., 2010). The earliest reported bHLH protein in animals is the mouse transcription factors E12 and E47 (Murre *et al*., 1989; Massari and Murre, 2000; Stevens *et al*., 2008; Murre, 2019). Later studies suggested the animal bHLH proteins

to be divided into six large categories of protein subclasses that were subdivided as 45 sub-families (Ledent and Vervoort, 2001; Ledent *et al*., 2002; Simionato *et al*., 2007; Stevens *et al*., 2008; Wang *et al*., 2010), whereas the plant bHLH proteins were divided into more than 20 subclasses with conserved motifs or domains identified (Carretero-Paulet *et al*., 2010; Yan *et al*., 2015; Liu *et al*., 2018; Wei and Chen, 2018; Gao *et al*., 2019), varying from 21 to 32 subfamilies (Carretero-Paulet *et al*., 2010; Song *et al*., 2014; Hudson and Hudson, 2014; Sun *et al*., 2015; Wang *et al*., 2015; Yan *et al*., 2015; Zhang *et al*., 2015; Kavas *et al*., 2016; Gao *et al*., 2017; Guo and Wang, 2017; Niu *et al*., 2017; Liu *et al*., 2018; Lu *et al*., 2018; Wei and Chen, 2018; Zhang *et al*., 2018; Gao *et al*., 2019).

Recently, as the whole genome genetic mapping of model animals and plants and the genome sequencing projects of numerous species have been completed, more and more transcription factors are identified and utilized, which is important and feasible for studies on the issues of functional characteristics and genetic evolution of specific transcription factors. On the one hand, these inherent information stored in various genomes may be explored to rapidly elucidate the genetics and developmental mechanisms regulating the processes of cell differentiation and organ development and organization growth of animal and plant and fungi species with the great development of modern bioinformatics and genomics. On the other hand, the bHLH transcription factor families have been identified and analyzed in the currently available genomes of many metazoan species, such as human, orangutan, mouse, rat, giant panda, chicken, sparrow, pig, cow, dog, zebrafish, lizard, silkworm, bee and other insects (Ledent and Vervoort, 2001; Ledent *et al*., 2002; Li *et al*., 2006; Simionato *et al*., 2007; Wang *et al*., 2007; Wang *et al*., 2008; Stevens *et al*., 2008; Wang *et al*., 2009; Zheng *et al*., 2009; Wang *et al*., 2010; Liu and Zhao, 2010; 2011; Dang *et al*., 2011; Liu *et al*., 2012; 2013; Liu and Chen, 2013; Wang *et al*., 2015; Liu and Li, 2015; Liu, 2015; Li and Liu, 2017; Zhang *et al*., 2017; Liu *et al*., 2018; Murre, 2019).

On the other hand, the bioinformatics databases of biological macromolecules and phenotypes and genotypes of specific genes and/or traits, such as the annotations of Gene Ontology (GO) forum and biological pathway databases mainly including the pathway databases of Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG), have shown a leap forward development. These large bioinformatics databases have become an extremely important method and main analytical tool for studying the functional characteristics of genes and genomics in the bioinformatics fields. Furthermore, their development has greatly accelerated the integration and utilization of modern genomes and biological data and "Omics" information, which is gradually changing the way we perceive and understand the genomic datasets of genomes and/or the organization and heredity of biological organisms. Among them, the GO forum has a database of dozens of animals, plants and microorganisms. The GO forum and KEGG pathway databases have constructed the relatively independent ontological vocabularies (Kanehisa and Goto, 2000). For instance, the GO forum developed the standard language "Ontology" with three levels of intrinsic structure, namely Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). All the common attributes of those genes, transcripts and their products can be used to organize the different functional concepts and/or annotations of GO and KEGG pathways into the organic systems of databases (Boggon *et al*., 1999; Luscombe *et al*., 2000; Riechmann *et al*., 2000, Dennis *et al*., 2003; Huang *et al*., 2009). Therefore, the genomic databases of GO and KEGG pathways are the most basic functional descriptions, structural composition descriptions, descriptions of synthesis and decomposition and metabolic maps of specific genes and their expression and transcripts.

Both *Xenopus tropicalis* and *Xenopus laevis* are the well-established biomedical model organisms for the genetics and developmental research. In practice, *Xenopus tropicalis* and *Xenopus laevis* are common and yet important clawed frogs that are generally used in the laboratories of biomedical and developmental biology (Bowes *et al*., 2008; Hellsten *et al*., 2010; Session *et al*., 2016; Elurbe *et al*., 2017; Kamran *et al*., 2018). Our group previously identified some bHLH transcription factors both the genomes of *Xenopus tropicalis* and *Xenopus laevis* (Liu and Chen, 2013; Liu and Li, 2015), in which we used the predefined bHLH gene/protein classification criteria (Atchley *et al*., 1999; Atchley and Fitch, 1997) and the verified 45 representative bHLH proteins and 118 human bHLH protein motifs in search of novel bHLH sequence hints (Ledent and Vervoort, 2001; Ledent *et al*., 2002; Simionato *et al*., 2007). In total, 105 bHLH proteins were identified from the genomic databases of *Xenopus tropicalis* and 106 bHLH proteins were found from the genomic databases of *Xenopus laevis* (Liu and Chen, 2013; Liu and Li, 2015). All these bHLH proteins retrieved from the those clawed frogs' genomes are used in the preliminary classification and comparison of the frog bHLH transcription factors performed in the present study. Furthermore, in view of the ongoing genome sequencing projects of the clawed frogs' genomes, the unearthed new annotations and functional information and structural features of many bHLH proteins should be identified and rediscovered and/or corrected. In this study, we further surveyed and identified the clawed frogs' bHLH proteins with the last updated genomic databases and totally 215 bHLH transcription factors were rediscovered or corrected and updated.

Next, we compared and characterized their functional enrichments of GO annotations and KEGG pathways by systematically analyzing the common functions and characteristics of gene enrichment distribution with recently updated genetic ontology databases. Finally, the clawed frogs' common 31 functional GO annotations of the bHLH transcription factors were identified and analyzed.

Materials and Methods

Genome-Wide Survey and Data Acquisition of the Latest Nucleic Acid and Protein Sequences from the Genomic Databases of the Clawed Frogs

The genome-wide survey and data acquisition of those nucleic acid and protein sequences of the latest frog bHLH transcription factors were obtained primarily through the independent BLAST search tools from the genomic databases of NCBI (URL: https://www.ncbi.nlm.nih.gov/genome/) and Xenbase Version 4.10.1 (with the sub-databases version 9.1 on JBrowser and UCSC Genome Browser for *Xenopus tropicalis* and the sub-databases version 9.2 on JBrowser or sub-databases version 9.1 on UCSC Genome Browser for *Xenopus laevis*; URL: http://www.xenbase.org/, RRID: SCR_003280). Briefly, based on our previous findings (Liu and Chen, 2013; Liu and Li, 2015), the protein motifs of these frog bHLH proteins were first searched and retrieved with those representative bHLH protein subtypes and the classification criteria of bHLH transcription factors predefined by Atchley *et al*. (1999) utilizing the BLASTN and BLASTP and TBLASTN search algorithms in both the genomic databases of NCBI and Xenbase. Then, the putative bHLH protein motifs were mainly surveyed in the genomic databases of NCBI for accurate sequence hints using TBLASTN and BLASTP search algorithms with the 45 representative bHLH proteins and 118 human bHLH protein motifs reported by Ledent *et al*. (2002) and Simionato *et al*. (2007). Next, all the sequences of putative frog bHLH proteins suggested by BLAST searches with high scores hints were retrieved and compared and selected with strict phylogenetic analyses for final candidate transcription factors. Among these bHLH protein motifs obtained from the independent results of BLAST searches, each sequence was further put in the genomic databases of NCBI and Xenbase, in which the stringent value was set to E<10 to allow us search for multiple motif hints and retrieve all the possible putative bHLH proteins.

In practice, we performed the repeated BLAST searches with TBLASTN and BLASTP search algorithms in the frog genomic databases of NCBI. Meanwhile, we also searched many times the frog databases of Xenbase (Bowes *et al*., 2008; Hellsten *et al*., 2010; Session *et al*., 2016; James-Zorn *et al*., 2015; Elurbe *et al*., 2017; Kamran *et al*., 2018). For *Xenopus tropicalis*, we searched for the latest bHLH sequences with BLAST algorithms in the genomic databases of NCBI (URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Xenbase (URL: http://www.xenbase.org/genomes/blast.do?db=Nucleotide/Xentr_9_1_Scaffolds). For *Xenopus laevis*, we searched for the latest bHLH sequences with BLAST algorithms in the genomic databases of NCBI (URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Xenbase (URL: http://www.xenbase.org/genomes/blast.do?db=Nucleotide/Xenla_9_2_Scaffolds).

Finally, according to the latest released genomic information and sequenced datasets Xenbase (http://www.xenbase.org/, RRID: SCR_003280), the genetic clones (scaffolds or genomic clones) of these two clawed frogs, i.e., *Xenopus tropicalis* and *Xenopus laevis*, the careful alignment and selection were carried out with the compared results of gene coding regions, the sequence alignment of putative genes and proteins, genome acquisition numbers, the sequence characteristics of protein motifs. After the removal of redundant sequences, we obtained the final candidate bHLH transcription factors. From the point view of comparative genomics, we then performed the comprehensive analyses of functional annotations of GO forum and KEGG pathway databases to compare and analyze the enrichment and distribution characteristics of bHLH transcription factors in these two clawed frogs (Note: The lists of updated bHLH proteins were shown in supplementary Tables 1 and 2).

Enrichment Distribution Analysis of the GO Functional Annotations and Kegg Pathways

After the search and retrieval of frog bHLH gene and protein sequences and datasets from the genomic databases of NCBI (URL: https://www.ncbi.nlm.nih.gov/genome/) and Xenbase (URL: http://www.xenbase.org/, RRID:SCR_003280), we used the functional annotation tools of DAVID Bioinformatics Resources (Dennis *et al*., 2003; Huang *et al*., 2009; URL: https://david.ncifcrf.gov/) to execute the enrichment distribution analyses of the GO functional annotations (GO forum; URL: http://www.geneontology.org/) and KEGG pathways (KEGG: Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000; URL: https://www.kegg.jp/). All the significant thresholds of the enrichment distribution analyses of the GO functional annotations and KEGG pathways (Kanehisa and Goto, 2000) were set with both the Benjamini Corrected P-Values and False Positive Values (FDRs) significantly controlled below 0.05 (marked as P<0.05,

FDR<0.05). Actually, P-Values corrected by the other alternative methods, such as P-Values corrected by the Bonferroni method, were carried out and considered in the study too (data shown in Tables 1, 2 and 3).

Results

Genome-Wide Identification and Rediscovery of the Last Updated Frog bHLH Transcription Factors

With the predefined classification criteria and representative sequences of bHLH proteins and the human bHLH motifs (Atchley and Fitch, 1997; Atchley *et al*., 1999; Ledent and Vervoort, 2001; Ledent *et al*., 2002; Simionato *et al*., 2007), we analyzed and updated the initial results of our previous studies using TBLASTN and BLASTP search algorithms described above. We carefully searched and totally identified 107 bHLH protein sequences of *Xenopus tropicalis* and 108 bHLH protein sequences of *Xenopus laevis*. Among these putative bHLH proteins retrieved, 15 sequences were updated in the genome of *Xenopus tropicalis*, including the protein sequences of Xsash3 (XP_002940370.1 updated to XP_004913964.1), Oligo2 (XP_002938491.1 updated to XP_004912201.1), Hes5e (NP_001107462.1 updated to XP_004916212.1) and Tal2 (XP_002934026.1 updated to XP_017948432.1 and XP_004918959.1), whereas 6 sequences were updated in the genome of *Xenopus laevis*, including the protein sequences of ARNT2 and Baml1 (i.e., ARNT2b and ARNT2c, Baml1b and Baml1c; NP_001080540.1 updated to XP_018106068.1 and NP_001089031.1 updated to XP_018110819.1). Some predicted proteins of previously poorly described bHLH members (Liu and Chen, 2013; Liu and Li, 2015) are also reanalyzed and identified. In addition, we further validated and corrected the annotations of three frog bHLH protein sequences with error names or ambiguities.

Enrichment Distribution Analyses of the GO Functional Annotations and KEGG Pathways

In general, the major functional activities of bHLH transcription factors and bHLH-like proteins are the activities of DNA binding, protein heterodimerization and protein polymerization, transcriptional co-activation and transcription regulation or repression. However, in addition to the common functions and roles shared by these ordinary transcription factors, the frog bHLH transcription factors have their own specific functional activities too. To further explore the overall common molecular characteristics and specific functional activities of the frog bHLH transcription factor families, we collected all the functional datasets of GO annotations and pathways for the 108 bHLH proteins of Xenopus laevis and 107 bHLH proteins of Xenopus tropicalis. There are totally 41 classes of GO annotations in Xenopus tropicalis and 45 classes of GO annotations in Xenopus laevis found to be statistically significantly enriched (P<0.05, FDR<0.05; see the data and GO annotations shown in Tables 1 and 2, Fig. 1 and 2) in the hypergeometric distributions.

There are some significant enriched groups of GO functional annotations and KEGG pathways for these two clawed frogs' bHLH transcription factors (P < 0.05, FDR < 0.05, Tables 1-4, Fig. 1 and 2). These GO functional annotations and KEGG pathways represent some important biological processes and information, molecular functions, cellular components and pathways, such as DNA-dependent (DNA-templated) transcription and (negative) transcription regulation, DNA binding and bHLH transcription factor binding, transcription factor complex and protein heterodimerization activity, orange and PAS domains/motifs, PAC motif and PAS fold, (negative) regulation of RNA metabolic processes, nuclear translocator and repressor, myogenic basic muscle-specific protein, muscle organ development, neurogenic differentiation factor and NeuroD, Notch signaling pathway and TGF-beta signaling pathway were highly frequent (P<0.05, FDR<0.05). In Tables 1-3 and Fig. 1 and 2, the observed data and phenomena indicated that these GO annotations are common and crucial functions of those frog bHLH genes. These common significant GO annotation terms show us the general genetic functions and most important roles of the frog bHLH transcription factors (Table 4, Fig. 1 and 2). In addition to these common functional annotation enrichment and cellular information, the significant GO annotation terms of these two frogs' bHLH proteins show that many crucial developmental or important physiological processes, such as (skeletal) muscle cell differentiation, muscle organ development, rhythmic process and biological rhythms, hypoxia (adaption) and hypoxia-inducible factors, neurogenesis, neural tube development and neurogenic differentiation are also highly significantly enriched in the hypergeometric statistical tests (P<0.05, FDR<0.05; Tables 1 and 2, Fig. 1 and 2). Furthermore, in Table 3, the predominantly enriched KEGG pathway of the bHLH genes in *Xenopus tropicalis* is the TGF-beta signaling pathway (Bonferroni corrected p-value 0.0150, Benjamini corrected p-value 0.0075), whereas those of the bHLH genes in *Xenopus laevis* are the Notch signaling pathway (Bonferroni Corrected p-value 1.77E-06, Benjamini corrected p-value 1.77E-06) and the TGF-beta signaling pathway (Bonferroni Corrected p-value 1.18E-05, Benjamini corrected p-value 5.92E-06) and the Fanconi anemia pathway (Bonferroni Corrected p-value 0.0062, Benjamini corrected p-value 0.0021). This is consistent with the functional enrichment test results of the in-group analysis of six high-order groups of the two clawed frogs' bHLH genes.

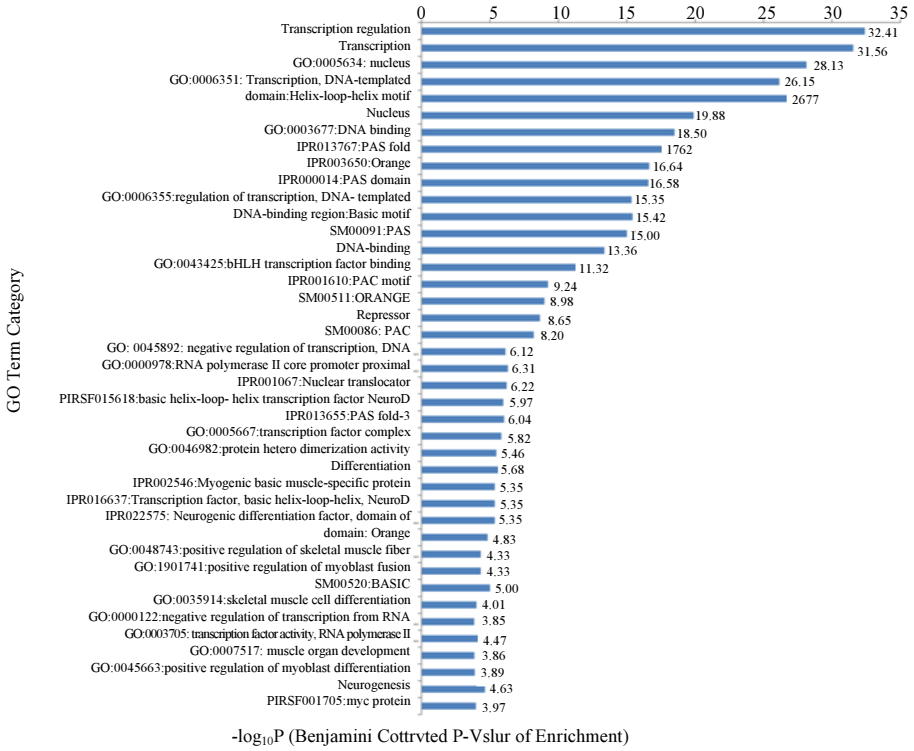

Fig. 1: Significantly enriched GO terms of *Xenopus tropicalis* bHLH transcription factors (P<0.05, FDR<0.05). **Note:** The figure shows significantly enriched GO terms of *Xenopus tropicalis* bHLH transcription factors identified with both the Benjamini corrected P values and FDR values (P<0.05, FDR<0.05)

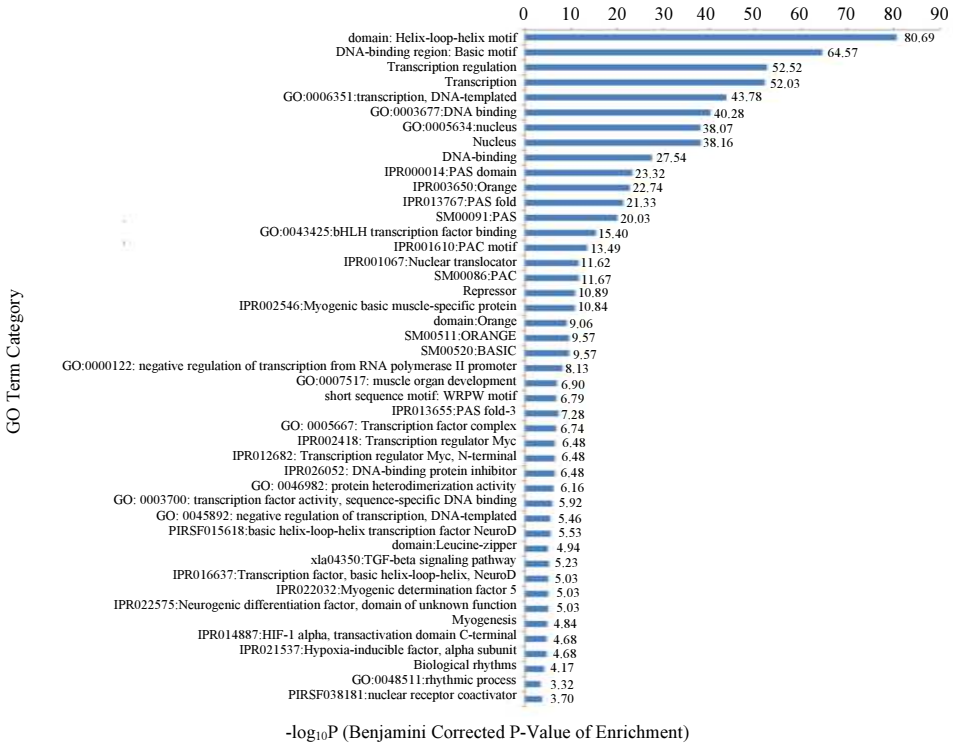

Fig. 2: Significantly enriched GO terms of *Xenopus laevis* bHLH transcription factors (P<0.05, FDR<0.05). **Note:** The figure shows significantly enriched GO terms of *Xenopus laevis* bHLH transcription factors identified with both the Benjamini corrected P values and FDR values (P<0.05, FDR<0.05)

Table 1: Significantly enriched GO terms of *Xenopus tropicalis* bHLH transcription factors

GO Category	GO Term	Gene Count	P-Value	Fold Enrichment	Bonferroni Corrected P-value	Benjamini Corrected P- value	FDR
UP_KEYWORDS	Transcription regulation	37	2.30E-34	15.920113	3.91E-33	3.91E-33	1.66E-31
UP_KEYWORDS	Transcription	37	3.24E-33	14.798978	5.51E-32	2.75E-32	2.34E-30
GOTERM_CC_DIRECT	GO:0005634: Nucleus	51	5.65E-30	4.7932478	7.34E-29	7.34E-29	3.74E-27
GOTERM_BP_DIRECT	GO:0006351: Transcription, DNA-templated	35	3.85E-29	11.504812	7.04E-27	7.04E-27	4.75E-26
UP_SEQ_FEATURE	domain:Helix-loop-helix motif	13	6.24E-29	113.30769	1.69E-27	1.69E-27	5.17E-26
UP_KEYWORDS	Nucleus	41	2.33E-21	5.7074491	3.96E-20	1.32E-20	1.69E-18
GOTERM_MF_DIRECT	GO:0003677:DNA binding	31	1.10E-20	8.0497512	3.18E-19	3.18E-19	9.27E-18
INTERPRO	IPR013767:PAS fold	11	1.17E-19	119.04157	4.82E-18	2.41E-18	1.08E-16
INTERPRO	IPR003650:Orange	11	1.68E-18	99.201307	6.91E-17	2.30E-17	1.55E-15
INTERPRO	IPR000014:PAS domain	12	2.58E-18	72.146405	1.06E-16	2.65E-17	2.38E-15
GOTERM_BP_DIRECT	GO:0006355:regulation of transcription, DNA- templated	27	4.89E-18	8.3313687	8.95E-16	4.48E-16	6.04E-15
UP_SEQ_FEATURE	DNA-binding region:Basic motif	10	2.83E-17	70.817308	7.63E-16	3.82E-16	2.34E-14
SMART	SM00091:PAS	12	2.34E-16	45.596273	2.00E-15	9.99E-16	1.33E-13
UP_KEYWORDS	DNA-binding	27	1.03E-14	6.5533937	1.76E-13	4.39E-14	7.47E-12
GOTERM_MF_DIRECT	GO:0043425:bHLH transcription factor binding	7	3.33E-13	156.58065	9.66E-12	4.83E-12	2.81E-10
INTERPRO	IPR001610:PAC motif	7	7.04E-11	87.408145	2.89E-09	5.77E-10	6.48E-08
SMART	SM00511:ORANGE	7	3.52E-10	61.175	3.17E-09	1.06E-09	2.03E-07
UP_KEYWORDS	Repressor	10	7.95E-10	21.298529	1.35E-08	2.25E-09	5.75E-07
SMART	SM00086:PAC	7	2.80E-09	47.057692	2.52E-08	6.31E-09	1.61E-06
GOTERM_BP_DIRECT	GO:0045892:negative regulation of transcription, DNA-templated	9	1.65E-08	18.942465	3.03E-06	7.56E-07	2.04E-05
GOTERM_MF_DIRECT	GO:0000978:RNA polymerase II core promoter proximal region sequence-specific DNA binding	7	6.74E-08	31.316129	1.95E-06	4.89E-07	5.69E-05
INTERPRO	IPR001067:Nuclear translocator	5	8.78E-08	101.45588	3.60E-06	6.00E-07	8.08E-05
PIR_SUPERFAMILY	PIRSF015618:basic helix-loop- helix transcription factor NeuroD	4	3.59E-07	165.875	1.08E-06	1.08E-06	1.13E-04
INTERPRO	IPR013655:PAS fold-3	5	1.57E-07	90.183007	6.45E-06	9.22E-07	1.45E-04
GOTERM_CC_DIRECT	GO:0005667:transcription factor complex	8	2.31E-07	17.954943	3.00E-06	1.50E-06	1.53E-04
GOTERM_MF_DIRECT	GO:0046982:protein hetero dimerization activity	6	6.02E-07	34.795699	1.75E-05	3.49E-06	5.08E-04
UP_KEYWORDS	Differentiation	8	8.68E-07	15.145621	1.48E-05	2.11E-06	6.28E-04
INTERPRO	IPR002546:Myogenic basic muscle-specific protein	4	8.67E-07	162.32941	3.55E-05	4.44E-06	7.98E-04
INTERPRO	IPR016637:Transcription factor, basic helix-loop-helix, NeuroD	4	8.67E-07	162.32941	3.55E-05	4.44E-06	7.98E-04
INTERPRO	IPR022575: Neurogenic differentiation factor, domain of unknown function	4	8.67E-07	162.32941	3.55E-05	4.44E-06	7.98E-04
UP_SEQ_FEATURE	domain: Orange	4	1.65E-06	113.30769	4.45E-05	1.48E-05	0.0013652
GOTERM_BP_DIRECT	GO:0048743:positive regulation of skeletal muscle fiber development	4	1.29E-06	141.01613	2.35E-04	4.71E-05	0.0015889
GOTERM_BP_DIRECT	GO:1901741:positive regulation of myoblast fusion	4	1.29E-06	141.01613	2.35E-04	4.71E-05	0.0015889
SMART	SM00520:BASIC	4	5.53E-06	87.392857	4.98E-05	9.96E-06	0.0031829
GOTERM_BP_DIRECT	GO:0035914:skeletal muscle cell differentiation	4	3.20E-06	112.8129	5.85E-04	9.76E-05	0.0039526
GOTERM_BP_DIRECT	GO:0000122:negative regulation of transcription from RNA polymerase II promoter	8	5.45E-06	11.28129	9.98E-04	1.43E-04	0.0067371
GOTERM_MF_DIRECT	GO:0003705:transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding	4	8.12E-06	89.474654	2.35E-04	3.36E-05	0.0068525
GOTERM_BP_DIRECT	GO:0007517:muscle organ development	5	6.10E-06	39.171147	0.0011157	1.40E-04	0.0075346
GOTERM_BP_DIRECT	GO:0045663:positive regulation of myoblast differentiation	4	6.37E-06	94.010753	0.0011647	1.29E-04	0.0078657
UP_KEYWORDS	Neurogenesis	6	1.11E-05	20.045675	1.88E-04	2.36E-05	0.0080261
PIR_SUPERFAMILY	PIRSF001705:myc protein	3	7.14E-05	165.875	2.14E-04	1.07E-04	0.0224829

Note: The significant thresholds of enrichment distribution analyses of the above GO functional annotations were set with both the Benjamini Corrected P-Values and Bonferroni Corrected P-Values and FDR Values controlled below 0.05 ($P<0.05$, $FDR<0.05$)

Table 2: Significantly enriched GO terms of *Xenopus Laevis* bHLH transcription factors

GO Category	GO Term	Gene Count	P-Value	Fold Enrichment	Bonferroni Corrected P-Value	Benjamini Corrected P-Value	FDR
UP_SEQ_FEATURE	domain:Helix-loop-helix motif	40	2.45E-83	69.653999	2.03E-81	2.03E-81	2.62E-80
UP_SEQ_FEATURE	DNA-binding region:Basic motif	35	6.51E-67	64.072749	5.40E-65	2.70E-65	6.97E-64
UP_KEYWORDS	Transcription regulation	63	9.11E-55	9.8202072	3.01E-53	3.01E-53	7.95E-52
UP_KEYWORDS	Transcription	63	5.60E-54	9.5439707	1.85E-52	9.24E-53	4.89E-51
GOTERM_BP_DIRECT	GO:0006351:transcription, DNA-templated	62	1.75E-46	6.991864	1.66E-44	1.66E-44	1.92E-43
GOTERM_MF_DIRECT	GO:0003677:DNA binding	53	2.74E-42	9.1370621	5.21E-41	5.21E-41	2.06E-39
GOTERM_CC_DIRECT	GO:0005634:nucleus	68	8.51E-40	4.08	8.51E-39	8.51E-39	5.11E-37
UP_KEYWORDS	Nucleus	69	6.28E-40	4.5124629	2.07E-38	6.91E-39	5.49E-37
UP_KEYWORDS	DNA-binding	47	3.48E-29	6.6522927	1.15E-27	2.87E-28	3.04E-26
INTERPRO	IPR000014:PAS domain	15	2.79E-25	83.272959	9.49E-24	4.74E-24	2.45E-22
INTERPRO	IPR003650:Orange	13	1.61E-24	111.03061	5.47E-23	1.82E-23	1.41E-21
INTERPRO	IPR013767:PAS fold	13	5.54E-23	96.226531	1.88E-21	4.71E-22	4.87E-20
SMART	SM00091:PAS	15	2.67E-21	43.155612	1.87E-20	9.35E-21	1.38E-18
GOTERM_MF_DIRECT	GO:0043425:bHLH transcription factor binding	10	4.20E-17	86.982323	7.98E-16	3.99E-16	3.15E-14
INTERPRO	IPR001610:PAC motif	9	4.79E-15	90.843228	1.62E-13	3.24E-14	4.20E-12
INTERPRO	IPR001067:Nuclear translocator	8	4.22E-13	88.82449	1.43E-11	2.39E-12	3.71E-10
SMART	SM00086:PAC	9	9.06E-13	47.07885	6.34E-12	2.11E-12	4.67E-10
UP_KEYWORDS	Repressor	17	2.73E-12	10.463308	9.01E-11	1.29E-11	2.38E-09
INTERPRO	IPR002546:Myogenic basic muscle-specific protein	7	2.98E-12	111.03061	1.01E-10	1.45E-11	2.62E-09
UP_SEQ_FEATURE	domain:Orange	7	3.14E-11	71.395349	2.61E-09	8.69E-10	3.36E-08
SMART	SM00511:ORANGE	7	1.53E-10	57.540816	1.07E-09	2.68E-10	7.90E-08
SMART	SM00520:BASIC	7	1.53E-10	57.540816	1.07E-09	2.68E-10	7.90E-08
GOTERM_BP_DIRECT	GO:0000122:negative regulation of transcription from RNA polymerase II promoter	13	2.35E-10	12.512447	2.23E-08	7.44E-09	2.58E-07
GOTERM_BP_DIRECT	GO:0007517:muscle organ development	8	5.28E-09	28.791182	5.01E-07	1.25E-07	5.80E-06
UP_SEQ_FEATURE	short sequence motif:WRPW motif	6	7.73E-09	61.196013	6.41E-07	1.60E-07	8.28E-06
INTERPRO	IPR013655:PAS fold-3	6	1.23E-08	66.618367	4.20E-07	5.25E-08	1.09E-05
GOTERM_CC_DIRECT	GO:0005667:transcription factor complex	9	3.63E-08	16.978846	3.63E-07	1.82E-07	2.18E-05
INTERPRO	IPR002418:Transcription regulator Myc	5	8.78E-08	92.52551	2.99E-06	3.32E-07	7.72E-05
INTERPRO	IPR012682:Transcription regulator Myc, N-terminal	5	8.78E-08	92.52551	2.99E-06	3.32E-07	7.72E-05
INTERPRO	IPR026052:DNA-binding protein inhibitor	5	8.78E-08	92.52551	2.99E-06	3.32E-07	7.72E-05
GOTERM_MF_DIRECT	GO:0046982:protein heterodimerization activity	8	1.08E-07	19.626781	2.05E-06	6.85E-07	8.10E-05
GOTERM_MF_DIRECT	GO:0003700:transcription factor activity, sequence-specific DNA binding	18	2.54E-07	4.4387887	4.83E-06	1.21E-06	1.90E-04

Table 2: Countinue

GOTERM_BP_DIRECT	GO:0045892:negative regulation of transcription, DNA-templated	11	1.81E-07	9.2910319	1.72E-05	3.44E-06	1.99E-04
PIR_SUPERFAMILY	PIRSF015618:basic helix-loop-helix transcription factor NeuroD	4	9.76E-07	127.3	2.93E-06	2.93E-06	3.07E-04
UP_SEQ_FEATURE	domain:Leucine-zipper	6	6.86E-07	30.598007	5.70E-05	1.14E-05	7.35E-04
KEGG_PATHWAY	xla04350:TGF-beta signaling pathway	7	1.32E-06	16.560102	1.18E-05	5.92E-06	7.57E-04
INTERPRO	IPR016637:Transcription factor, basic helix-loop-helix, NeuroD	4	2.73E-06	111.03061	9.28E-05	9.28E-06	0.0024005
INTERPRO	IPR022032:Myogenic determination factor 5	4	2.73E-06	111.03061	9.28E-05	9.28E-06	0.0024005
INTERPRO	IPR022575:Neurogenic differentiation factor, domain of unknown function	4	2.73E-06	111.03061	9.28E-05	9.28E-06	0.0024005
UP_KEYWORDS	Myogenesis	5	4.37E-06	41.427126	1.44E-04	1.44E-05	0.0038108
INTERPRO	IPR014887:HIF-1 alpha, transactivation domain C-terminal	4	6.78E-06	88.82449	2.30E-04	2.10E-05	0.0059623
INTERPRO	IPR021537:Hypoxia-inducible factor, alpha subunit	4	6.78E-06	88.82449	2.30E-04	2.10E-05	0.0059623
UP_KEYWORDS	Biological rhythms	5	2.27E-05	28.344875	7.49E-04	6.81E-05	0.0198128
GOTERM_BP_DIRECT	GO:0048511:rhythmic process	5	3.02E-05	25.867077	0.0028611	4.77E-04	0.0331462
PIR_SUPERFAMILY	PIRSF038181:nuclear receptor coactivator	3	1.33E-04	127.3	3.99E-04	1.99E-04	0.0418314

Note: The significant thresholds of enrichment distribution analyses of the above GO functional annotations were set with both the Benjamini Corrected P-Values and Bonferroni Corrected P-Values and FDR Values controlled below 0.05 ($p<0.05$, FDR<0.05)

Table 3: Significantly enriched KEGG pathways of bHLH transcription factors identified in the clawed frog genomes

Frog Species	KEGG Pathway Term	Gene Count	P-Value	Fold Enrichment	Bonferroni Corrected P-Value	Benjamini Corrected P-Value
Xenopus tropicalis	xtr04350:TGF-beta signaling pathway	4	0.0012592	16.509946	0.0150065	0.0075316
Xenopus Laevis	xla04330:Notch signaling pathway	6	1.96E-07	38.408304	1.77E-06	1.77E-06
Xenopus Laevis	xla04350:TGF-beta signaling pathway	7	1.32E-06	16.560102	1.18E-05	5.92E-06
Xenopus Laevis	xla03460:Fanconi anemia pathway	4	6.87E-04	20.728291	0.00617045	0.0020611

However, the overall functional enrichment of pathways of the bHLH genes in *Xenopus laevis* is relatively more complicated than that of the bHLH genes in *Xenopus tropicalis*, as shown in Table 3. The KEGG pathway enrichment analysis of these bHLH genes in *Xenopus laevis* include two more significantly enriched signaling pathways (i.e., the Notch signaling pathway and the Fanconi anemia pathway) than that of those bHLH genes in *Xenopus tropicalis*. However, the two clawed frogs' bHLH gene were significantly enriched in a common KEGG pathway (i.e., TGF-beta signaling pathway, Table 3).

Discussion

Both *Xenopus tropicalis* and *Xenopus laevis* are important well-established model organisms for the genetics and developmental biology and biomedical studies worldwide. In recent years, many metazoan genomes have been sequenced, including those of *Xenopus tropicalis* and *Xenopus laevis* and the inherent information stored in various genomes can be explored to elucidate the genetics regulatory processes and developmental mechanisms of animals and plants and fungi. These data provide us with rich resources for comparative genomic analysis with modern advanced analyzing technologies and tools including many software packages and algorithms and databases of bioinformatics and genomics. In this study, we used the international annotations of GO and KEGG pathway databases to functionally analyze and characterize the enrichment distributions of these two clawed frogs' bHLH transcription factors from the point view of comparative genomics. A total of 215 bHLH transcription factors were identified and rediscovery and analyzed in the research. Firstly, we searched and updated the initial results of the previous studies using TBLASTN and BLASTP search algorithms described in materials and methods. We totally retrieved and identified 107 bHLH protein sequences of *Xenopus tropicalis* and 108 bHLH protein sequences of *Xenopus laevis*. Among these putative bHLH proteins retrieved, 15 sequences were updated in the genome of *Xenopus tropicalis* and 6 sequences were updated in the genome of *Xenopus laevis*. In this study, some predicted proteins of previously poorly described bHLH members are further identified with in-group analyses of phylogenetic analyses (data not shown). We also validated and corrected the annotations of three frog bHLH protein sequences BLAST searches and phylogenetic analyses. Next, we compared and characterized their functional enrichments of GO annotations and KEGG pathways by systematically analyzing the common functions and characteristics of gene enrichment distribution with recently updated genetic ontology databases. There were 41 GO terms and one pathway identified as significantly enriched for *Xenopus tropicalis*, whereas there were 45 GO terms and 3 pathways identified as significantly enriched for *Xenopus laevis*. Among those significantly enriched GO terms, the two clawed frogs share 31 common functional GO annotations of bHLH genes, including DNA-dependent transcription and (negative) transcription regulation, DNA binding and bHLH binding, transcription factor complex and protein heterodimerization activity, (negative) regulation of RNA metabolic processes, nuclear translocator and repressor, myogenic basic muscle-specific protein, neurogenic differentiation factor and NeuroD. Further more, these frogs' bHLH genes were also found to play important roles in the regulation of gene expression in some important developmental or physiological processes, such as (skeletal) muscle cell

differentiation, muscle organ development, biological rhythms and rhythmic process, hypoxia (adaption) and hypoxia-inducible factors, neurogenesis, neural tube development and neurogenic differentiation with high frequencies or scores of enrichment, whereas they were commonly significantly enriched in TGF-beta signaling pathway. These results were consistent with the previous observed functionally enriched data (Liu and Chen, 2013; Liu and Li, 2015) that indicated muscle organ development and (negative) regulation of muscle development, muscle fiber development and skeletal muscle (tissue) development, neural tube development, embryonic development, (nuclear) hormone receptor binding, circadian rhythm and circadian clock, TGF-beta signaling pathway and Notch signaling pathway have high frequent enrichments of GO and KEGG pathways (Liu and Chen, 2013; Liu and Li, 2015). However, the other GO categories and KEGG pathways including Fanconi anemia pathway were enriched in low frequencies as previously described (Liu and Chen, 2013; Liu and Li, 2015).

In the present study, the bHLH genes of the clawed frogs were naturally enriched in myogenesis, (skeletal) muscle cell differentiation, muscle organ development, neurogenesis, neural tube development and neurogenic differentiation and TGF-beta signaling pathway, since there are many research hotspots of bHLH genes, such as MyoD and Myogen and NeuroD (neuronal differentiation 1), regulating the cellular and developmental processes of myogenesis and neurogenesis and TGF-beta signaling pathway in the corresponding organs' development (Tazumi *et al*., 2008; Della *et al*., 2012; Curran *et al*., 2014; Hardwick *et al*., 2016). Actually, it is well known that the clawed frogs are a kind of muscular amphibians adapted to jumping and hunting insects (Ferenczi *et al*., 2004). Both the clawed frogs are well-established model organisms in physiological experiments. Many previous experiments have demonstrated that the changes of the frog muscle fibre's volumes agreed with the expected muscular movement and energy requirement from simple osmotic behaviour (Ferenczi *et al*., 2004). Furthermore, Ascl1 (achaete-scute family bHLH transcription factor 1), aliases ASCL1 and ASH1, is a multi-functional regulator of neural development in invertebrates and vertebrates, whereas the ectopic expression of Ascl1 can generate functional neurons from non-neural somatic cells. Functional studies have identified Ascl1 as a crucial maternal regulator of the germ layer pattern formation in clawed frogs, since the maternally expressed proteins can establish the major embryonic body axes and a pre-patterned transitional stage for later-acting zygotic signals (Gao *et al*., 2016; Min *et al*., 2016). Previous studies have revealed that the maternally supplied Ascl1 was capable to set a pre-patterning tendency for frog embryonic cells to adopt neural fates and represses the mesendoderm formation via the HDAC-dependent antagonism of frog VegT (Gao *et al*., 2016; Min *et al*., 2016). On the other hand, the hairy genes are key bHLH transcription factors required in the early neural crest development of clawed frogs (Vega-López *et al*., 2015), whereas neural crest formation is one of the fundamental processes in the early stages of the frog embryonic development to generate a variety of tissues and cell types.

According to the present analysis, some bHLH genes of the clawed frogs were also enriched in the specific processes of biological rhythms and circadian rhythm and hypoxia (adaption), but there are presently reported only a few candidate genes of hypoxia-inducible factors and circadian genes, including HIF1alpha (hypoxia-inducible factor 1 alpha) HIF2alpha (hypoxia-inducible factor 2 alpha), Clock, Bmal1 (xBmal1), cryptochromes 1 and 2, periods 1 and 2 (xPeriod1 and xPeriod2) and xNocturnin (Green, 2003; Van *et al*., 2007; Beaucourt and Coumailleau, 2007; Li *et al*., 2008; Curran *et al*., 2008; Kriegmair *et al*., 2013; Curran *et al*., 2014). In fact, vertebrate retinas contain the endogenous circadian clocks that control many aspects of retinal physiology and circadian rhythms control the temporal arrangement of molecular, physiological and behavioral processes within an organism and synchronize these processes with the external environment too (Green, 2003; Van *et al*., 2007; Curran *et al*., 2008; Kriegmair *et al*., 2013; Curran *et al*., 2014). Previous studies indicated that virtually all organisms ranging from metazoan to humans exhibit circadian rhythms in their organic and cellular processes, from genetic phenotypes and physiology to cell hormone levels and gene expression (Green, 2003; Van *et al*., 2007; Curran *et al*., 2008; 2014). The circadian rhythms of clawed frogs are typical intracellular mechanisms composed of interlocking transcriptional and/or translational feedback loops and/or internal circadian timekeeping mechanisms of clocks (Van *et al*., 2007; Curran *et al*., 2008; 2014). This internal timekeeping mechanism allows the frog organisms internally rearrange their physiological processes to temporally anticipate in or adapt to the external changes (Van *et al*., 2007; Curran *et al*., 2008; 2014). Actually, frog circadian rhythms are approximately a full day of 24 hours in duration and persist in the constant conditions within a physiological range of body temperatures and internal environment (Green, 2003; Van *et al*., 2007; Curran *et al*., 2008; 2014). Meanwhile, the hypoxia-inducible factors (HIF1alpha and HIF2alpha) are members of the bHLH transcription factor family. It was shown that frog xHIF1alpha heterodimerized with the Arnt1 protein

(xArnt1) with the protein complex being mediated by the HLH and PAS domains (Beaucourt and Coumailleau, 2007). These results indicated that the endogenous circadian rhythms and hypoxia adaptive transcription factors may help clawed frogs to adapt to the terrestrial environments since frogs are the transitional amphibians from fish to land animals that may be tolerant of hypoxia and fast circadian clocks (Del, 2018). However, there is presently no more report of hypoxia-inducible factors in frog experiments.

According to the present study, the two clawed frogs' bHLH genes share a common TGF-beta signaling pathway which was significantly enriched in both the frogs genomes. Previous reports have shown that the specific signal transduction by TGF-beta family in vertebrates involves many sets of polypeptide growth factors, receptor complexes of serine/threonine kinases and Smad proteins acting as receptor substrates with Smad-associated transcription factors (Liu *et al*., 1996; Lagna *et al*., 1996; Derynck *et al*., 1996; Chen *et al*., 1998; Upadhyay *et al*., 2017; Kim and Baek, 2019). Members of TGF-beta family include TGF-beta factors, activins, Bone Morphogenetic Proteins (BMPs) and other related growth factors that regulate cell division, differentiation, motility, adhesion and death in the metazoan organs and tissues (Liu *et al*., 1996; Lagna *et al*., 1996; Derynck *et al*., 1996; Chen *et al*., 1998; Upadhyay *et al*., 2017; Kim and Baek, 2019). Because of the diverse processes controlled by different TGF-beta family members and their various roles in pathogenesis and/or tumorigenesis and development biology, there is an intense interest in studies on the compositional basis and relevant genes for the tissue-specific signal transduction pathways in vertebrates (Liu *et al*., 1996; Lagna *et al*., 1996; Derynck *et al*., 1996; Chen *et al*., 1998; Upadhyay *et al*., 2017; Kim and Baek, 2019). In our analysis, it has been reported that the two clawed frogs' bHLH proteins as important transcription factors were commonly involved in TGF-beta signaling pathway with significant P Values of functional enrichment. These results indicated that 4 bHLH genes in *Xenopus tropicalis* and 7 bHLH genes in *Xenopus Laevis* are involved in the TGF-beta signaling pathway. However, more functional details about the TGF-beta signaling pathway should be further analyzed in the frog experiments of genetics and development biology.

In addition, according to the classification and grouping criteria of six high-order groups of animal bHLH transcription factors by Ledent *et al*. (2002) and Simionato *et al*. (2007), we also tested the functional enrichment analyses of GO annotations and KEGG pathways on each high-order group members of the frog bHLH genes, respectively. Actually, the functional enrichment of some group specific GO annotations of bHLH genes in the six high-order groups was extremely significant ($P<0.001$, $FDR<0.05$). In consistent with the previous studies (Liu and Chen, 2013; Liu and Li, 2015), the bHLH transcription factors in the high-order groups have some significant differences in the functional enrichment and in-group distribution of GO annotations due to their own independent sequence and functional characteristics. These results provide a foundation for understanding the functional roles of bHLH transcription factors in the genetics and development and evolution of clawed frogs.

Conclusion

Comparative genomic research of bHLH genes in different animal species can show us specific biological data and information of common and/or differentiated phenomena observed between these two clawed frogs. In the study, we found 41 (for *Xenopus tropicalis*) and 45 (for *Xenopus laevis*) statistically significant enrichment GO terms of bHLH transcription factors in the two clawed frogs' genomes. Among those significantly enriched GO terms, the two clawed frogs share 31 common functional GO annotations of these bHLH genes, including DNA-dependent (DNA-templated) transcription and (negative) transcription regulation, DNA binding and bHLH transcription factor binding, transcription factor complex and protein heterodimerization activity, orange and PAS domains/motifs, PAC motif and PAS fold, (negative) regulation of RNA metabolic processes, nuclear translocator and repressor, myogenic basic muscle-specific protein, muscle organ development, neurogenic differentiation factor and NeuroD. These data and phenomena indicated that these GO annotations are common functions of those frog bHLH genes. Furthermore, the bHLH genes of the two clawed frogs were also found to play important roles in the regulation of gene expression in some crucial developmental or important physiological processes, such as (skeletal) muscle cell differentiation, muscle organ development, rhythmic process and biological rhythms, hypoxia (adaption) and hypoxia-inducible factors, neurogenesis, neural tube development and neurogenic differentiation, whereas they were significantly enriched in a common KEGG pathway (TGF-beta signaling pathway). These resulted data and observed information analyzed are very important for us to understand the functions, classification and evolution of frog bHLH genes.

Acknowledgement

The author is grateful to the anonymous reviewers for their constructive comments and suggestions.

Funding Information

It is jointly funded by the Major Project of Municipal Horizontal Cooperation between Fuyang Government and Fuyang Normal University (No. XDHX201725) and the Major Project of Anhui Provincial Educational Commission Natural Science Foundation (No. KJ2016SD47) and the Project of National Natural Science Foundation of China (No.31301965)..

Author's Contributions

Wuyi liu designed, collected, checked and analyzed all the data and prepare the full draft manuscript.

Ethics

This is original and unpublished article. There are no animals dealt in the study and the author declared that no competing interests and no ethical issues involved. The corresponding author confirms that the author has read and approved the manuscript.

References

Atchley, W.R., W. Terhalle and A. Dress, 1999. Positional dependence, cliques and predictive motifs in the bHLH Protein Domain. J. Mol. Evol., 5: 501-516. DOI: 10.1007/pl00006494

Atchley, W.R. and W.M. Fitch, 1997. A natural classification of the basic helix-loop-helix class of transcription factors. Proc. Nat. Acad. Sci., 9: 5172-5176. DOI: 10.1073/pnas.94.10.5172

Boggon, T.J., W.S. Shan, S. Santagata, S.C. Myers and S. Lawrence, 1999. Implication of tubby proteins as transcription factors by structure-based functional analysis. Science, 286: 2119-2125. DOI: 10.1126/science.286.5447.2119

Bowes, J.B., K.A. Snyder, E. Segerdell, R. Gibb and Jarabek C. *et al*., 2008. Xenbase: A Xenopus biology and genomics resource. Nucleic Acids Res., 36: D761- D767. DOI: 10.1093/nar/gkm826

Carretero-Paulet, L., A. Galstyan, I. Roig-Villanova, J.F. Martinez-Garcia and J.R. Bilbao-Castro *et al*., 2010. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss and algae. Plant. Physiol., 153: 1398-1412. DOI: 10.1104/pp.110.153593

Chen, Y.G., A. Hata, R.S. Lo, D. Wotton and Y. Shi *et al*., 1998. Determinants of specificity in TGF-beta signal transduction. Genes Dev., 12: 2144-2152. DOI: 10.1101/gad.12.14.2144

Curran, K.L., L. Allen, B.B. Porter, J. Dodge and C. Lope *et al*., 2014. Circadian genes, xBmal1 and xNocturnin, modulate the timing and differentiation of somites in Xenopus laevis. PLoS One, 9: e108266. DOI: 10.1371/journal.pone.0108266

Curran, K.L., S. Larue, B. Bronson, J. Solis and A. Trow *et al*., 2008. Circadian genes are expressed during early development in Xenopus laevis. PLoS One, 3: e2749. DOI: 10.1371/journal.pone.0002749

Dang, C., Y. Wang, D. Zhang, Q. Yao and K. Chen, 2011. A genome-wide survey on basic helix-loop-helix transcription factors in giant panda. PLoS One, 6: e26878. DOI: 10.1007/s00335-009-9196-3

Beaucourt, B.A. and P. Coumailleau, 2007. Molecular cloning and characterization of the Xenopus hypoxia-inducible factor 1alpha (xHIF1alpha). J. Cell Biochem., 102: 1542-1552. DOI: 10.1002/jcb.21376

Del, P.E.M., 2018. The extraordinary biology and development of marsupial frogs (*hemiphractidae*) in comparison with fish, mammals, birds, amphibians and other animals. Mechanisms Developm., 154: 2-11. DOI: 10.1016/j.mod.2017.12.002

Della, G.B., A.S. Armand, S. Lecolle, F. Charbonnier and C. Chanoine, 2012. Mef2d acts upstream of muscle identity genes and couples lateral myogenesis to dermomyotome formation in Xenopus laevis. PLoS One, 7: e52359. DOI: 10.1371/journal.pone.0052359

Dennis, G.J., B.T. Sherman, D.A. Hosack and J. Yang *et al*., 2003. DAVID: Database for annotation, visualization and integrated discovery. Genome Biol., 4: R60. DOI: 10.1186/gb-2003-4-5-p3

Derynck, R., W.M. Gelbart, R.M. Harland, C.H. Heldin and S.E. Kern *et al*., 1996. Nomenclature: Vertebrate mediators of TGFbeta family signals. Cell, 87: 173. DOI: 10.1016/S0092-8674(00)81335-5

Elurbe, D.M., S.S. Paranjpe, G. Georgiou, K.I. Van and O. Bogdanovic *et al*., 2017. Regulatory remodeling in the allo-tetraploid frog *Xenopus laevis*. Genome. Biol., 18: 198. DOI: 10.1186/s13059-017-1335-7

Ferenczi, E.A., J.A. Fraser, S. Chawla, J.N. Skepper and L.H. Huang, 2004. Membrane potential stabilization in amphibian skeletal muscle fibres in hypertonic solutions. J. Physiol., 555: 423-438. DOI: 10.1113/jphysiol.2003.058545

Gao, C., J. Sun, C. Wang, Y. Dong and S. Xiao *et al*., 2017. Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. PLoS One, 12: e0181843. DOI: 10.1371/journal.pone.0181843

Gao, F., K. Robe, F. Gaymard, E. Izquierdo and C. Dubos, 2019. The transcriptional control of iron homeostasis in plants: A tale of bHLH transcription factors? Front. Plant. Sci., 10: 6. DOI: 10.3389/fpls.2019.00006

Gao, L., X. Zhu, G. Chen, X. Ma and Y. Zhang *et al*., 2016. A novel role for Ascl1 in the regulation of mesendoderm formation via HDAC-dependent antagonism of VegT. Development, 143: 492-503. DOI: 10.1242/dev.126292

Green, C.B., 2003. Molecular control of Xenopus retinal circadian rhythms. J. Neuroendocrinol., 15: 350-4. DOI: 10.1046/j.1365-2826.2003.00999.x

Guo, X.J. and J.R. Wang, 2017. Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat. BMC Plant. Biol., 17: 90. DOI: 10.1186/s12870-017-1038-y

Hardwick, L.J., J.D. Davies and A. Philpott, 2016. MyoD phosphorylation on multiple C terminal sites regulates myogenic conversion activity. Biochem. Biophys. Res. Commun., 481: 97-103. DOI: 10.1016/j.bbrc.2016.11.009

Hellsten, U., R.M. Harland, M.J. Gilchrist, D. Hendrix and J. Jurka *et al*., 2010. The genome of the Western clawed frog Xenopus tropicalis. Science, 328: 633-636. DOI: 10.1126/science.1183670

Huang, D.W., B.T. Sherman and R.A. Lempicki, 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature. Protocol., 4: 44-57. DOI: 10.1038/nprot.2008.211

Hudson, K.A. and M.E. Hudson, 2014. The basic Helix-loop-Helix transcription factor family in the sacred Lotus, Nelumbo Nucifera. Trop. Plant Biol., 7: 65-70. DOI: 10.1007/s12042-014-9138-4

James-Zorn, C., V.G. Ponferrada, K.A. Burns, J.D. Fortriede and V.S. Lotay *et al*., 2015. Xenbase: Core features, data acquisition and data processing. Genesis, 53: 486-97. DOI: 10.1002/dvg.22873

Kamran, K., J.D. Fortriede, V.S. Lotay, K.A. Burns and D.Z. Wang *et al*., 2018. Xenbase: A genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Res., 46: D861-D868. DOI: 10.1093/nar/gkx936

Kanehisa, M. and S. Goto, 2000. KEGG: Kyoto encyclopaedia of genes and genomes. Nucleic Acids Res., 28: 27-30. DOI: 10.1093/nar/28.1.27

Kavas, M., M.C. Baloglu, E.S. Atabay, U.T. Ziplar and H.Y. Dasgan *et al*., 2016. Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration. Mol. Genet. Genomics., 291: 129-143. DOI: 10.1007/s00438-015-1095-6.

Kim, S.Y. and K.H. Baek, 2019. TGF-β signaling pathway mediated by deubiquitinating enzymes. Cell Mol. Life Sci., 76: 653-665. DOI: 10.1007/s00018-018-2949-y

Kriegmair, M.C., S. Frenz, M. Dusl, W.M. Franz and R. David *et al*., 2013. Cardiac differentiation in Xenopus is initiated by mespa. Cardiovasc. Res., 97: 454-63. DOI: 10.1093/cvr/cvs354

Lagna, G., A. Hata, A. Hemmati-Brivanlou and J. Massagué, 1996. Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. Nature, 383: 832-6. DOI: 10.1038/383832a0

Ledent, V. and M. Vervoort, 2001. The basic Helix-Loop-Helix protein family: Comparative genomics and phylogenetic analysis. Genome Res., 5: 754-770. DOI: 10.1101/gr.177001

Ledent, V., O. Paquet and M. Vervoort, 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Bio., 3: H30. DOI: 10.1186/gb-2002-3-6-research0030

Li, F. and W. Liu, 2017. Genome-wide identification, classification and functional analysis of the basic helix-loop-helix transcription factors in the cattle, Bos Taurus. Mamm. Genome, 28: 176-197. DOI: 10.1007/s00335-017-9683-x

Li, J., Q. Liu, M. Qiu, Y. Pan and Y. Li, 2006. Identification and analysis of the mouse basic/Helix-Loop-Helix transcription factor family. Biochem. Bioph. Res. Co., 3: 648-656. DOI: 10.1016/j.bbrc.2006.09.114

Li, R., J. Yue, Y. Zhang, L. Zhou and W. Hao *et al*., 2008. CLOCK/BMAL1 regulates human nocturnin transcription through binding to the E-box of nocturnin promoter. Mol. Cell Biochem., 317: 169-77. DOI: 10.1007/s11010-008-9846-x

Liu, W.Y. and C.J. Zhao, 2010. Genome-wide identification and analysis of the chicken basic helix-loop-helix factors. Comp. Funct. Genomics., 2010: 682095. DOI: 10.1155/2010/682095

Liu, A., Y. Wang, C. Dang, D. Zhang and H. Song *et al*., 2012. A genome-wide identification and analysis of the basic helix-loop-helix transcription factors in the ponerine ant, Harpegnathos saltator. BMC Evolutionary Biol., 12: 165. DOI: 10.1186/1471-2148-12-165

Liu, A., Y. Wang, D. Zhang, X. Wang and H. Song *et al*., 2013. Classification and evolutionary analysis of the basic helix-loop-helix gene family in the green anole lizard, Anolis carolinensis. Mol. Genetics Genomics., 288: 365-380. DOI: 10.1007/s00438-013-0755-7

Liu, B., X. Guan, W. Liang, J. Chen and L. Fang *et al*., 2018. Divergence and evolution of cotton bHLH proteins from diploid to allotetraploid. BMC Genomics., 19: 162. DOI: 10.1186/s12864-018-4543-y.

Liu, F., A. Hata, J.C. Baker, J. Doody and J. Cárcamo *et al*., 1996. A human Mad protein acting as a BMP-regulated transcriptional activator. Nature, 381: 620-3. DOI: 10.1038/381620a0

Liu, W.Y. and C.J. Zhao, 2011. Molecular phylogenetic analysis of zebra finch basic helix-loop-helix transcription factors. Biochem. Genetics., 49: 226-241. DOI: 10.1007/s10528-010-9401-9

Liu, W.Y. and D.Y. Chen, 2013. Phylogeny, functional annotation and protein interaction network analyses of the *Xenopus tropicalis* basic helix-loop-helix transcription factors. Biomed Res. Int., 2013: 145037. DOI: 10.1155/2013/145037

Liu, W.Y. and F.M. Li, 2015. Identification and bioinformatics analyses of the basic Helix-loop-helix transcription factors in *Xenopus laevis*. Pak. J. Bio. Sci., 18: 149-165. DOI: 10.3923/pjbs.2015.149.165

Liu, W.Y., 2015. Genome-wide identification, classification and functional analyses of the bHLH transcription factor family in the pig, *Sus scrofa*. Mol. Genetics Genomics, 290: 1415-1433. DOI: 10.1007/s00438-015-1007-9

Lu, R., J. Zhang, D. Liu, Y.L. Wei and Y. Wang *et al*., 2018. Characterization of bHLH/HLH genes that are involved in Brassinosteroid (BR) signaling in fiber development of cotton (*Gossypium hirsutum*). BMC Plant. Biol., 18: 304. DOI: 10.1186/s12870-018-1523-y

Luscombe, N.M., S.E. Austin, H.M. Berman and J.M. Thornton, 2000. An overview of the structures of protein-DNA complexes. Genome Biol., 1: 1-37. DOI: 10.1186/gb-2000-1-1-reviews001

Massari, M.E. and C. Murre, 2000. Helix-Loop-Helix proteins: Regulators of transcription in eucaryotic organisms. Mol. Cell Biol., 2: 429-440. DOI: 10.1128/MCB.20.2.429-440.2000

Min, Z., H. Lin, X. Zhu, L. Gao and A.A. Khand *et al*., 2016. Ascl1 represses the mesendoderm induction in Xenopus. Acta Biochim. Biophys., 48: 1006-1015. DOI: 10.1093/abbs/gmw092

Murre, C., 2019. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. Genes. Dev., 33: 6-25. DOI: 10.1101/gad.320663.118

Murre, C., P.S. Mccaw, H. Vaessin, M. Caudy and L.Y. Jan *et al*., 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell, 58: 537-544. DOI: 10.1016/0092-8674(89)90434-0

Niu, X., Y. Guan, S. Chen and H. Li, 2017. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in Brachypodium distachyon. BMC Genomics., 18: 619. DOI: 10.1186/s12864-017-4044-4

Riechmann, J.L., J. Heard, G. Martin, L. Reuber and C. Jiang *et al*., 2000. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science, 290: 2105-10. DOI: 10.1126/science.290.5499.2105

Session, A.M., Y. Uno, T. Kwon, J.A. Chapman. and A. Toyoda *et al*., 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature, 538: 336-343. DOI: 10.1038/nature19840

Simionato, E., V. Ledent, G. Richards, M. Thomas-Chollier and P. Kerner *et al*., 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: Insights from comparative genomics. BMC Evol. Biol., 7: 33. DOI: 10.1186/1471-2148-7-33

Song, X.M., Z.N. Huang, W.K. Duan, J. Ren and T.K. Liu *et al*., 2014. Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (Brassica rapa ssp. pekinensis). Mol. Genet. Genomics., 289: 77-91. DOI: 10.1007/s00438-013-0791-3

Stevens, J.D., E.H. Roalson and M.K. Skinner, 2008. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: Genomic approach to cellular differentiation. Differentiation, 76: 1006-1022. DOI: 10.1111/j.1432-0436.2008.00285.x

Sun, H., H.J. Fan and H.Q. Ling, 2015. Genome-wide identification and characterization of the bHLH gene family in tomato. BMC Genomics., 16: 9. DOI: 10.1186/s12864-014-1209-2

Tazumi, S., S. Yabe, J. Yokoyama, Y. Aihara and H. Uchiyama, 2008. PMesogenin 1 and 2 function directly downstream of Xtbx6 in Xenopus somitogenesis and myogenesis. Dev. Dyn., 237: 3749-3761. DOI: 10.1002/dvdy.21791

Upadhyay, A., L. Moss-Taylor, M.J. Kim, A.C. Ghosh and M.B. O'Connor, 2017. TGF-β Family Signaling in Drosophila. Cold Spring Harb. Perspect. Biol., 9: a022152. DOI: 10.1101/cshperspect.a022152

Van, D.S.E.A., F.E. Conte, K.E. Marz and C.B. Green, 2007. Structure/function analysis of Xenopus cryptochromes 1 and 2 reveals differential nuclear localization mechanisms and functional domains important for interaction with and repression of CLOCK-BMAL1. Mol. Cell Biol., 27: 2120-2129. DOI: 10.1128/MCB.01638-06

Vega-López, G.A., M. Bonano, C. Tríbulo, J.P. Fernández and T.H. Agüero *et al*., 2015. Functional analysis of Hairy genes in Xenopus neural crest initial specification and cell migration. Dev. Dyn., 244: 988-1013. DOI: 10.1002/dvdy.24295

Wang, J., Z. Hu, T. Zhao, Y. Yang and T. Chen *et al*., 2015. Genome-wide analysis of bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato (Solanum lycopersicum). BMC Genomics., 16: 39. DOI: 10.1186/s12864-015-1249-2

Wang, X.H., Y. Wang, A. Liu, X.T. Liu and Y. Zhou *et al*., 2015. Genome-wide identification and analysis of basic helix-loop-helix domains in dog, *Canis lupus familiaris*. Mol. Genetics Genomics., 290: 633-648. DOI: 10.1007/s00438-014-0950-1

Wang, Y., K. Chen, Q. Yao, W. Wang and Z. Zhi, 2007. The basic helix-loop-helix transcription factor family in *Bombyx mori*. Develop. Genes Evolution., 217: 715-723. DOI: 10.1007/s00427-007-0184-x

Wang, Y., K. Chen, Q. Yao, W. Wang and Z. Zhu, 2008. The basic helix-loop-helix transcription factor family in the honey bee, *Apis mellifera*. J. Insect Sci., 8: 1-12. DOI: 10.1673/031.008.4001

Wang, Y., K. Chen, Q. Yao, X. Zheng and Z. Yang, 2009. Phylogenetic analysis of zebrafish basic helix-loop-helix transcription factors. J. Mol. Evolut., 68: 629-640. DOI: 10.1007/s00239-009-9232-7

Wang, Y., Q. Yao and K.P. Chen, 2010. Progress of studies on family members and functions of animal bHLH transcription factors. Yi Chuan, 32: 307-330. DOI: 10.3724/SP.J.1005.2010.00307

Yan, Q., H.S. Liu, D. Yao, X. Li and H. Chen *et al*., 2015. The basic/Helix-Loop-Helix protein family in Gossypium: Reference genes and their evolution during tetraploidization. PLoS One, 10: e0126558. DOI: 10.1371/journal.pone.0126558

Wei, K. and H. Chen, 2018. Comparative functional genomics analysis of bHLH gene family in rice, maize and wheat. BMC Plant Biol., 18: 309. DOI: 10.1186/s12870-018-1529-5

Zhang, C., R. Feng, R. Ma, Z. Shen and Z. Cai *et al*., 2018. Genome-wide analysis of basic helix-loop-helix superfamily members in peach. PLoS One, 13: e0195974. DOI: 10.1371/journal.pone.0195974

Zhang, D., G. Li and Y. Wang, 2017. A genome-wide identification and analysis of basic helix-loop-helix transcription factors in cattle. Gene, 626: 241. DOI: 10.1016/j.gene.2017.05.036

Zhang, X., H. Luo, Z. Xu, Y. Zhu and A. Ji *et al*., 2015. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in Salvia miltiorrhiza. Sci. Rep., 5: 11244. DOI: 10.1038/srep11244

Zheng, X., Y. Wang, Q. Yao, Z. Yang and K. Chen, 2009. A genome-wide survey on basic helix-loop-helix transcription factors in rat and mouse. Mamm. Genome, 20: 236-246. DOI: 10.1007/s00335-009-9176-7