aJava

Asian Journal of Animal and Veterinary Advances

Academic
Journals Inc.

www.academicjournals.com

Asian Journal of Animal and Veterinary Advances 7 (5): 390-402, 2012
ISSN 1683-9919 / DOI: 10.3923/ajava.2012.390.402

Phylogenetic and Evolutionary Analyses of the Animal Hairy and Enhancer of Split Basic Helix-loop-helix Transcription Factors

[1,2]Wu-yi Liu and [3]Ke-jun Zhang

[1]Department of Scientific Research, Fuyang Normal College, Fuyang City 236041, China
[2]Department of Biology Sciences, Fuyang Normal College, Fuyang City 236041, China
[3]Qingdao Institute of Biomass Energy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao City 266101, China

Corresponding Author: Wu-yi Liu, Department of Biological Sciences, Fuyang Normal College, No. 100 West Qing He Road, Fuyang City 236037, People's Republic of China Tel: +86-558-2596562,+86-18298199628 Fax: +86-558-2596561

ABSTRACT

The basic helix-loop-helix (bHLH) transcription factors are the products of main regulatory genes in metazoans. Regulatory genes are usually believed to play a large role in morphological and genetic diversifications and are often characterized by elevatedrates of evolution. In previous studies, it was showed there was much genetic diversification among the members of Hairy and Enhancer of split (Hairy/E(spl) or H/E(spl)) bHLH family. This promoted us to do further research on the mechanism of genetic diversification within the H/E(spl) bHLH family. With all the data available, we carried out an evolutionary analysis among sequence sites of the enlarged H/E(spl) bHLH family and reported that all the Hes gene lineages were under purifying selection, but the Hes2, Hes3, Hes5 lineages evolved faster than the other ones with relatively higher omega values. We found nine sites of Hes5 genes under significantly positive selection and concluded that the diversity of H/E(spl) genes has been established through relaxed selective constraint.

Key words: Transcription factor, hairy and enhancer of split, bHLH, selective pressure

INTRODUCTION

Transcription Factor (TF) plays crucial roles in the regulation processes of gene expression of all living organisms at the transcriptional levels, such as development, growth and responses to environmental stimulus, etc. Some of these families are common to most eukaryotic organisms, while others are specific to particular taxonomic groups (Luscombe *et al*., 2000; Riechmann *et al*., 2000). The basic helix-loop-helix (bHLH) TFs constitute one of the largest families of functionally important proteins and were believed to play key roles in a wide range of developmental processes in metazoan and other organisms (Atchley and Fitch, 1997; Massari and Murre, 2000; Ledent and Vervoort, 2001; Jones, 2004). Murre *et al*. (1989) reported the first case of bHLH TFs E12 and E47 found in the murine organism. Many studies focused on the bHLH transcription factors and aimed at identification and characterization of their functions and classifications. Especially in recent years, the number of bHLH TFs characterized rapidly increased. More and more bHLH TF genes has been identified and bHLH TF families have been analyzed in many organisms whose genome drafts have been available (Morgenstern and Atchley, 1999; Atchley *et al*., 1999; Ledent *et al*., 2002; Toledo-Ortiz *et al*., 2003; Heim *et al*., 2003; Li *et al*., 2006a, b; Simionato *et al*., 2007, 2008;

Stevens *et al.*, 2008; Wang *et al.*, 2009; Zheng *et al.*, 2009; Pires and Dolan, 2010; Carretero-Paulet *et al.*, 2010; Liu and Zhao, 2011). With the genome resources of interested organisms being available, it would be desirable to have a more refined classification scheme of various types of bHLH motifs, as well as a better understanding of their functions and evolutionary implications within or among species.

Previous studies found that there was much genetic diversification among the members of H/E(spl) bHLH family (Zheng *et al.*, 2009; Liu and Zhao, 2011). However, the mechanism of genetic diversification within this gene family was still unclear. From the viewpoint of gene evolution, regulatory genes believed to play important roles in morphological and genetic diversifications are often characterized by the elevated rates of evolution. This promoted us to do further researches within the H/E(spl) bHLH family. In this study, with all of the data from previous studies and our unpublished data, we made phylogenetic analyses of the H/E(spl) bHLH transcription factors. We next carried out an evolutionary analysis among sequence sites of the enlarged H/E(spl) bHLH family and reported the results of selective pressure estimation among sites of H/E(spl) bHLH genes.

MATERIALS AND METHODS

Data collection, sequence alignment and selection: A number of closely related genes, known as the Hairy and Enhancer of split (Hairy/E(spl) or H/E(spl)) bHLH family, including Hes, Her, or ESR genes and enhancer of split related factors have been isolated from many vertebrates. Like the *Drosophila* E(spl) genes, many of their vertebrate homologues are expressed in response to Notch activity (Campos-Ortega, 1993, 1994, 1995) and the products of these genes are essential to implement many of the cell fate decisions mediated by Notch signaling, such as the selection of cells to become neural precursors (Artavanis-Tsakonas *et al.*, 1995; Greenwald, 1998). More than 300 sequences were downloaded and finally a total number of 163 gene sequences from 15 different species were analyzed. The bHLH proteins alignment by ClustalX 2.0 (Thompson *et al.*, 1997) was transformed into an aligned cds (coding sequence) fasta file using PAL2NAL (Suyama *et al.*, 2006), which is a program to construct multiple codon alignments from matching amino acid sequences. Incomplete sequences and highly divergent regions or gaps resulting in uncertain alignments were excluded from the further analysis. The final data set included a total of 163 sequences from 15 different species (sequence data not shown), including Hey/Herp, Dec, Hes and Hes-like genes, i.e., *Homo sapiens* (17), *Pan troglodytes* (10), *Macaca mulatta* (1), *Canis lupus* (6), *Sus scrofa* (1), *Bos taurus* (16), *Mus musculus* (15), *Rattus norvegicus* (11), *Danio rerio* (21), *Gallus gallus* (6), *Taeniopygia guttata* (10), *Xenopus tropicalis* (13), *Xenopus laevis* (17), *Apis mellifera* (7) and *Drosophila melanogaster* (12).

Phylogenetic analyses of H/E(spl) bHLH factors: Phylogenetic analyses were conducted by MRBAYES 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and PHYML 2.4.4 (Guindon and Gascuel, 2003), with the JTT substitution frequency matrix (Jones *et al.*, 1992). The obtained bHLH sequences were used to construct phylogenetic trees of bayesian inference and maximum likelihood matching with the human bHLH domains. Maximum Likelihood (ML) analyses were performed using the frequencies of amino acids estimated from the data set and rate heterogeneity across sites modeled by one constant rate and eight γ-rates. Statistical support for the different internal branches was assessed by bootstrap resampling with 100 replicates in PHYML (Guindon and Gascuel, 2003). Phylogenetic analyses of Bayesian inference was performed

with two independent Markov chains, each containing from 80 to 150 million Monte Carlo steps until the standard deviation of split frequencies was below 0.01 with sample frequency saved every 1000 generations. Finally, the trees obtained in the two runs of Markov chains were meshed and the first 25% of the trees were discarded as 'burnin' and the 50% majority consensus trees were edited and displayed.

Evolutionary analysis and selective pressure estimation: To determine whether the H/E (spl) split bHLH family was subject to positive (diversifying) or negative (purifying) selection, the substitution rates at synonymous (dS) and nonsynonymous (dN) were calculated and compared on the basis of nucleotide coding regions of Hes and Hes-like genes. The dN should be significantly smaller than the dS (ω=dN/dS<1) for a protein-coding gene under functional constraint or purifying selection, while the dN would be significantly larger than the dS (i.e., ω=dN/dS>1) under positive selection, which drives the rapid evolution of genes (Yang and Nielsen, 2002; Yang *et al*., 2005; Yang, 2007; Anisimova and Yang, 2007). Different regions with different functions may experience different selection pressure and positive selection would likely affect a few amino acids at particular genes. Thus, variable selective pressure among sites should be detected carefully.

In total, a number of 163 gene sequences from 15 different species were analyzed. Numbers of dS and dN nucleotide substitutions per site between orthologous gene sequences and likelihood ratios were computed using the Maximum Likelihood approach with alternative models using the CODEML program implemented in the PAML 4.4 package (Yang, 2007). In the maximum likelihood approach, site-specific models M1 (neutral), M2 (selection), M7 (beta), M8 (beta and ω; Nielsen and Yang, 1998; Yang *et al*., 2000) and M8a (beta and ω=1; Swanson *et al*., 2003) are used with CODEML program in the PAML 4.4 package (Yang, 2007). Model M1 (neutral) allows two classes of sites with ω_0=0 and ω_1=1 in proportions p_0 and p_1=1-p_0, respectively. Model M2 (selection) has an additional class with ω_2, which takes on any nonnegative value and applies to a proportion p_2 of sites, now with the constraint $p_0+p_1+p_2=1$. Test for positive selection should be done by comparing twice the log-likelihood difference between M1 and M2 with a χ^2 distribution in the LRT (Yang *et al*., 2000). Model M7 (beta) assumes a β-distribution for $0\leq\omega\leq1$. Model M8 (beta and ω) adds to M7 an extra category, with proportion p_1 of sites with ω1, while the rest of sites (at frequency p_0=1-p_1) have ω from the β-distribution between 0 and 1. Here we compare twice the log-likelihood difference between M7 and M8 with a χ^2 distribution to test for positive selection (Yang *et al*., 2000; Anisimova *et al*., 2001). Model M8a is similar to model M8 except that the category ω_1 is fixed at ω_1=1.

We estimated the selective pressures acting on the coding regions by taking the phylogenetic-based Maximum Likelihood (ML) analyses. These analyses were conducted under different competing evolutionary hypotheses. We first investigated whether the distribution of selective constraints acting on the each gene fluctuated across lineages by comparing the fit to data of the one ratio model (M0), which assumes a constant selective pressure across branches, with the free ratios model (FR). We next examined other evolutionary scenarios to detect sites under positive selection in Hes gene lineages, using three codon-based ML substitution models that are site-specific (i.e., models with different ω ratios among sites) (Yang *et al*., 2000, 2005; Yang and Nielsen, 2002) but assume the same selection pattern for a site in all lineages. The likelihood Ratio Test (LRT) was used to compare the fit to data of two nested models, assuming that twice the log likelihood

difference between the two models (2ΔL) follows a χ^2 distribution with a number of degrees of freedom equal to the different numbers of free parameters.

RESULTS AND DISCUSSION

Phylogenetic analyses of H/E(spl) bHLH factors in different animal species: Determining the phylogenetic relationships of the bHLH proteins is an important step for elucidating the evolutionary and functional divergence of this gene family as well. Meanwile, with genome sequence data for more and more species becoming available, it is now feasible to compare the bHLH gene family among different animal species at the genome-wide level. Herein, phylogenetic analyses of Bayesian Inference (BI) and maximum likelihood estimate (ML) were used to identify putative orthologous relationships of the H/E (spl) split bHLH members in different phylogenetic trees. The H/E (spl) split bHLH family, including diverse hairy/enhancer of split related genes (mainly Hes proteins and Hes-like factors). Previously, we showed most invertebrate species have an invariable number of either 11 or 12, while the vertebrate species have 6-15 members in the H/E(spl) bHLH family (Liu and Zhao, 2011). The phylogenetic tree of hairy/enhancer of split like orthologous genes from human, mouse, rat, zebrafish, *Xenopus tropicalis* and chicken was explored by a maximum likelihood method with bHLH protein sequences and the zebrafish HEYL being used as out-group (Fig. 1). It was found that Hes genes from human, mouse, rat, zebrafish, *Xenopus tropicalis* and chicken form clear monophyletic groups (except Hes4), indicating that each Hes lineage has its own ancestral sequence (Fig. 1).

Detecting variable selective pressure among sites of the H/E(spl) bHLH family: The analysis of nonsynonymous to synonymous substitution rate ratio was used to detect selective pressure on the enlarged H/E(spl) gene sequences. We estimated ω as an average over all sites and branches from the Hes and Hes-like paralogus sequences and the ratio was substantially smaller than 1 (one ratio model, ω=0.2742) that indicated that purifying selection had been the predominant force acting on the evolution of these genes. However, selective constrains, are unevenly distributed across the phylogeny (data not shown). Models of variable ω ratios among sequence sites were used to test for the presence of sites under positive selection (with ω=1) and to identify them. Positive selection would likely affect a few amino acids at specific lineages on the phylogeny and models estimating ω ratios averaged by codons or amid acids are highly conservative. Variations in ω among sites could occur along with particular lineage heterogeneity, which can be tested by fitting the data to a model comprising different site classes. The test results of site models for each lineage were shown in Table 1.

The M3 and M0 LRT was significant, indicating that one category of ω was not fit data well to describe the variability in selection pressure across amino acid sites. The estimates of ω for Hes1, Hes2, Hes3, Hes4, Hes5, Hes6, Hes7, Hes-like, Hey/Herp and Dec gene phylogenies were 0.0439, 0.2149, 0.2150, 0.0985, 0.2464, 0.1297, 0.1303, 0.1280, 0.0850 and 0.1300, respectively (Table 1). The tests contrasting the models M1a against M2a resulted in the p values less than 1 for all the lineages or groups suggesting a lack of power and the amino acid changes within each cluster were assumed neutral or under negative selection. M1a, the parameter estimates for the least parameter rich model describes that most sites with low ω estimates (indicative of selective constraints). The estimates of ω for Hes1, Hes2, Hes3, Hes4, Hes5, Hes6, Hes7, Hes-like, Hey/Herp and Dec gene

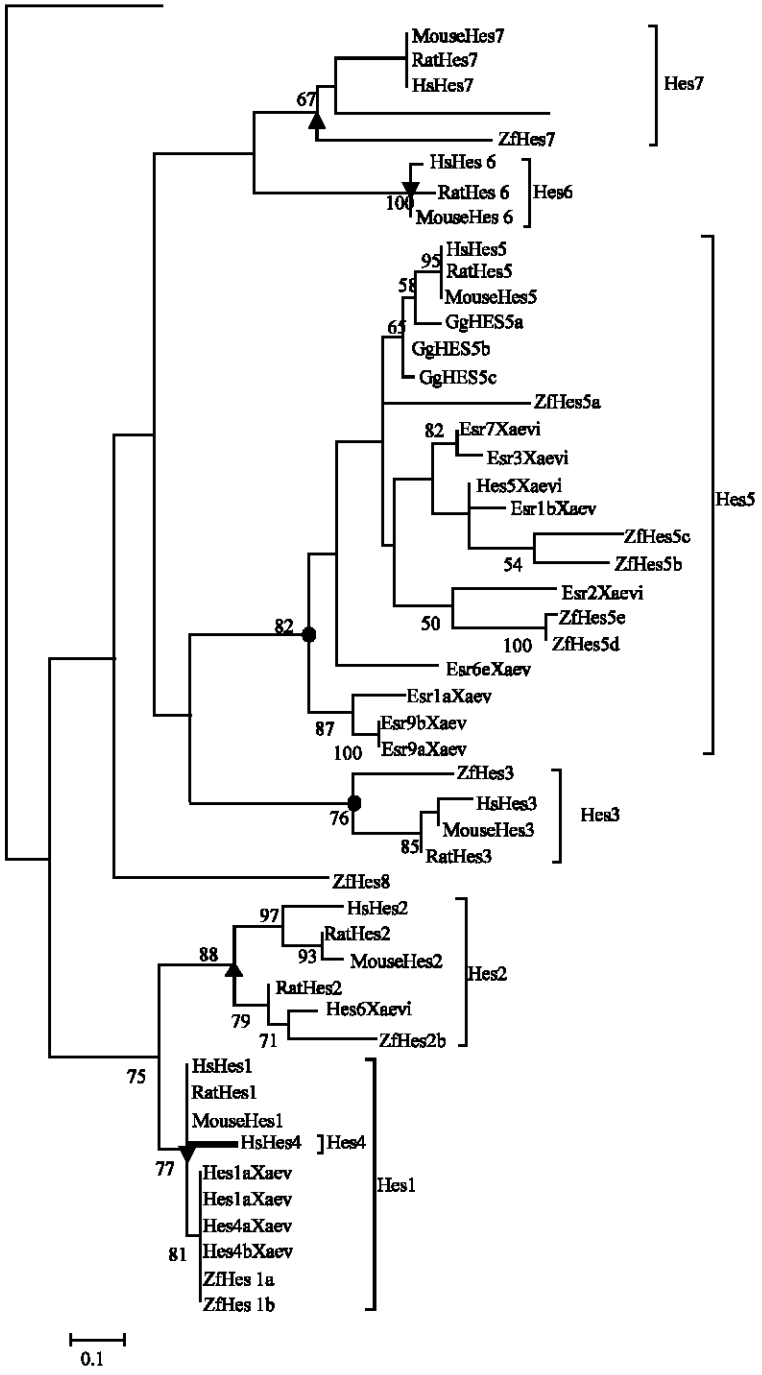

Fig. 1: Phylogenetic tree of the H/E(spl) bHLH family members from human, mouse, rat, zebrafish, *Xenopus laevis*, and chicken, with the zebrafish HEYL as outgroup. Figures around the node are the transferred bootstrap values of maximum likelihood estimates (MLEs) for corresponding branches. The phylogenetic tree of Hes factor motifs revealed that Hes1, Hes2, Hes3, Hes5, Hes6, and Hes7 had their own common ancestor sequences, respectively

Table 1: Selection pressures detection and model comparison among sequence sites for the H/E(spl) gene lineages

	Models comparison and parameter estimates										
	M3 vs M0			M2a vs M1a				M8 vs M7			
Gene lineages	2ΔL = (L1-L0)	p-value	ω-value	2ΔL = (L1-L0)	p-value	ω-value	Positively selected sites (ω>1)	2ΔL = (L1-L0)	p-value	ω-value	Positively selected sites (ω>1)
Hes1	241.8714	<0.001	0.0439	0	>0.05	0.1518	159G,237A	1.7828	>0.05	0.0467	237A
Hes2	115.6373	<0.001	0.2149	0	>0.05	0.4312	N/A	0.3592	>0.05	0.2250	115D
Hes3	85.3758	<0.001	0.2150	0	>0.05	0.4392	N/A	0.00041	>0.05	0.2378	N/A
Hes4	37.4375	<0.001	0.0985	0	>0.05	0.1226	188I	0.0451	>0.05	0.0952	166A,188I,190S
Hes5	513.4400	<0.001	0.2464	0	>0.05	0.3011	54S,64K	735.0716	<0.001	0.2422	30L, 76T**, 158S**, 169S*, 171K**, 170P, 172G**, 174P*,176A*,177T**,179K**
Hes6	343.4217	<0.001	0.1297	0	>0.05	0.3780	7V	1.5919	>0.05	0.1319	7V,11G,14T
Hes7	10.8885	<0.05	0.1303	0	>0.05	0.1420	136G,213P	0.000004	>0.05	0.1325	82A,136G,153A,173C, 213P
Hes-like	358.6621	<0.001	0.1280	0.1572	>0.05	0.3691	68M	4.2481	>0.05	0.1325	68M
Dec1,2	311.0111	<0.001	0.1300	0	>0.05	0.2801	N/A	0.00076	>0.05	0.1319	N/A
Hey/Herp	497.3909	<0.001	0.0850	0	>0.05	0.1264	N/A	0.00351	>0.05	0.0834	N/A

Codon sites with ambiguity data were removed from the alignment before analysis. Marker ** Means sites inferred under selection at the 99% level (p>99%) and marker * means those at the 95% level (p>95%) by Bayes Empirical Bayes (BEB) analysis (Yang *et al*., 2005)

phylogenies were 0.1518, 0.4312, 0.4392, 0.1226, 0.3011, 0.378, 0.142, 0.3691, 0.1264 and 0.2801, respectively (Table 1). The test using M7 against M8, which allows for beta-distributed site-specific ω ratio, detected no gene lineage or group under positive selection. The estimates of ω for Hes1, Hes2, Hes3, Hes4, Hes5, Hes6, Hes7, Hes-like, Hey/Herp and Dec gene phylogenies were 0.0467, 0.2250, 0.2378, 0.0952, 0.2422, 0.1319, 0.1325, 0.1325, 0.0834 and 0.1319, respectively (Table 1). But for Hes5 genes, Model M8 found nine sites were significantly under possible positive selection with Bayes Empirical Bayes (BEB) analysis (six at the 0.05 level and three at the 0.001 level).

The Naïve Empirical Bayes (NEB) (Nielsen and Yang, 1998; Yang *et al*., 2000) and the Bayes Empirical Bayes (BEB) (Yang *et al*., 2005) methods are implemented since PAML version 3.14 for calculating the posterior probabilities for site classes and identifying sites under positive selection if the likelihood ratio test is significant. NEB uses the MLEs (Maximum likelihood Estimates) of parameters but do not account for their sampling errors, while BEB deals with the sampling errors by applying a Bayesian prior. BEB is available and used under models M2a and M8 only. From these three pair-wise nested comparisons, we found all the 163 Hes gene lineages were under purifying selection and/or strong selective constraints, but there were different evolutionary rates among these lineages. From the selection and neutral model (M2a and M1a) analyses, Hes2, Hes3, Hes5 lineages obviously evolved faster than the other ones with relatively higher ω values (i.e., 0.4312, 0.4392, 0.3011, Table 1). From the site model (M8 and M7) analyses, all the Hes gene sequences, except Hes5 genes, have no site under significant or strong positive selective pressure. No positively selected sites were found in the Hes3 gene lineages. Similar to Hes5 genes, some sites from Hes1, Hes2, Hes4, Hes6 and Hes7 lineages appeared to be under putative positive selection, but the p values detected with BEB analysis were not significant ($p>0.05$).

Hairy and Enhancer of split (H/E(spl) in brief), including several Hes or Her or ESR genes in vertebrates, is one of the *Drosophila melanogaster* neurogenic loci required for normal segregation of neural and epidermal cell progenitors (Campos-Ortega, 1993, 1994, 1995). The E(spl) locus was found to be composed of a complex of several genes clustered together, called E(sp1)-C (enhancer of split complex), required to allow epidermal development of neuroectodermal cells (Campos-Ortega, 1993, 1994, 1995). These genes and others formed the so-called H/E(spl) gene family. Apart from their presence in epidermoblasts, H/E(spl) transcripts have occasionally been detected in the neuroblasts after separation of the lineages. Molecular data suggested that at least seven transcription units are members of the gene complexes, they were found to be expressed in nearly identical patterns (Knust *et al*., 1987, 1992). These seven genes of enhancer of split complex in *Drosophila melanogaster* encode basic-helix-loop-helix transcription factors (Md, Mb, Mg, M3, M5, M7 and M8) which are components of the Notch signaling pathway. Their expression are in response to Notch activation and mediate some effects of the Notch signaling pathway by regulating the expression of target genes (Artavanis-Tsakonas *et al*., 1995; Greenwald, 1998).

Notch function is essential for many cell fate decisions in *Drosophila* that involve the singling out of one cell from a group of competent cells (Muskavitch, 1994). In these processes, the Notch protein appears to function as a receptor in a cell-cell signaling pathway whose other components include two ligands 'Delta' and 'Serrate', the intracellular transducer suppressor of hairless and the nuclear proteins encoded by the enhancer of split complex (Artavanis-Tsakonas *et al*., 1995). The most immediate transcriptional target genes of Notch activation in *Drosophila melanogaster* encode seven bHLH proteins (Md, Mb, Mg, M3, M5, M7 and M8) which are clustered in the E(spl)-C

(Jennings *et al.*, 1999). The expression of E(spl) genes are required downstream of Notch since the accumulation of E(spl) proteins depends on Notch activation and they mediate the effects of constitutively active forms of Notch during neurogenesis (Artavanis-Tsakonas *et al.*, 1995; Greenwald, 1998). Therefore, mutations of E(spl) genes are deleterious during neurogenesis of embryo. Detailed genetic analyses suggested that there is some overlap in the functions of these proteins (Celis de *et al.*, 1996). Consistent with the proposed functional redundancy, the patterns of expression of E(spl) genes are also very similar during embryo neurogenesis (Knust *et al.*, 1987, 1992; Celis de *et al.*, 1996). This could explain why there are many E(spl) genes in different vertebrates and whether there is possible positive selection driving the genetic diversity of the E(spl) proteins.

Moreover, all the seven E(spl) proteins have two additional helix motifs and terminate with the same tetrapeptide 'WRPW', besides the HLH domain. And the degree of conservation is variable in the two additional regions (helix III/IV, or helix III and IV). Alignment showed a region of 41 amino acids present in seven proteins of the E(spl)-C. Besides the great similarity in the basic domain of the E(sp1) proteins, including the proline residue (Fig. 1), the other domains are also conserved, such as the HLH domain and the C-terminal tetrapeptide 'WRPW'. Although there is some amino acid sequence conservation in the hypothetical helix III/IV region of the E(sp1) proteins, the ability to form an a helix is only apparent in helix IV and less obvious in helix III. It is obvious that the two helix motifs are more variable than the other regions and drive the diversification of E(sp1) proteins. This was verified by our analysis of selective pressure on the nine sites of Hes5 genes. Among Hes5 genes, the nine sites under putative positive selection were 76T**, 158S**, 169S*, 171K**, 172G**, 174P*, 176A*, 177T** and 179K** (corresponding sites before deleting alignment gaps; the markers, ** and *, denote sites under selection at the 99 and 95% level, respectively). It is found that only one site (76T) lies in the HLH domain, while the other eight sites locate in the rest part following the HLH domain including the additional Hairy Orange domain.

Determining selective pressure among sites of the H/E(spl) bHLH domains: The analysis of nonsynonymous to synonymous substitution rate ratio was further used to detect possible selective pressure on the H/E(spl) bHLH and/or HLH domains. To determine whether there has been putative positive selection, we also did three analyses (M3 and M0, M2a and M1a, M8 and M7) and estimated ω values over all sites and branches from the Hes and Hes-like paralogus sequences. From these pairwise nested comparisons, we found the 163 Hes bHLH and/or HLH domains were under purifying selection or strong function constraints, although there were different evolutionary rates among different gene lineages. From the selection and neutral model (M2a and M1a) analyses, Hes3, Hes5, Hes7 lineages obviously evolved faster than the other ones with relatively higher ω values (i.e., 0.2605, 0.1941, 0.1202, Table 2). From the site model (M8 and M7) analyses, all the Hes gene sequences have no site, except one site (37N, corresponding site before deleting alignment gaps) in Hes2 genes, under putative positive selective pressure ($p>0.05$) and Hes2, Hes3, Hes5 lineages obviously evolved faster than the other ones with relatively higher ω values (i.e., 0.1697, 0.1156, 0.1175, Table 2). However, all the estimated ratios were all substantially smaller than 1 (Table 2), indicating that there might be predominant purifying selection affecting on the evolution of bHLH or HLH domains.

Table 2: Selection pressures detection and model comparison among sequence sites for the H/E(spl) bHLH domains

	Models comparison and parameter estimates										
	M3 vs M0			M2a vs M1a				M8 vs M7			
Gene lineages	2ΔL = (L1-L0)	p-value	ω-value	2ΔL = (L1-L0)	p-value	ω-value	Positively selected sites (ω>1)	2ΔL = (L1-L0)	p-value	ω-value	Positively selected sites (ω>1)
Hes1	0	>0.05	0.0332	0.0020	>0.05	0.0332	N/A	0.0020	>0.05	0.0343	N/A
Hes2	80.0450	<0.001	0.0623	5.4629	>0.05	0.0716	37N	9.4283	>0.05	0.1697	37N
Hes3	74.6365	<0.001	0.1156	0	>0.05	0.2605	N/A	0.0002	>0.05	0.1156	N/A
Hes4	14.8906	<0.01	0.0056	0.0012	>0.05	0.0042	N/A	0.0014	>0.05	0.0061	N/A
Hes5	175.6435	<0.001	0.1239	0	>0.05	0.1941	N/A	2.1354	>0.05	0.1175	N/A
Hes6	23.0424	<0.001	0.0256	0	>0.05	0.0781	N/A	0.0012	>0.05	0.0255	N/A
Hes7	30.8204	<0.001	0.0232	0	>0.05	0.1202	N/A	0.0018	>0.05	0.0361	N/A
Dec1,2	10.1245	<0.05	0.0192	0.0014	>0.05	0.0163	N/A	0.0016	>0.05	0.0188	N/A
Hey/Herp	144.1811	<0.001	0.0173	0	>0.05	0.0351	N/A	3.3841	>0.05	0.0143	N/A

Marker ** means sites inferred under selection at the 99% level (p>99%) and marker * Means those at the 95% level (p>95%) by Bayes Empirical Bayes (BEB) analysis (Yang *et al*., 2005)

CONCLUSIONS

The basic helix-loop-helix (bHLH) transcription factors are key regulators in cell and organ and organism, especially animal's muscle development and the adaptation to hypoxia (Massari and Murre, 2000; Velleman *et al*., 2010; Liu *et al*., 2010; Fakheri *et al*., 2010; Gonchar and Mankovska, 2010; Song *et al*., 2010). Present study showed that a considerable number of bHLH genes were found to have a multi-member distribution pattern in human, mouse, rat, zebrafish, chicken and Zebra Finch bHLH gene families (Liu and Zhao, 2011). This case suggests that they should arise through gene duplication at least before the divergence of vertebrates from invertebrates. In the research, phylogenetic analysis of the H/E (spl) family revealed that more than four gene duplication events should have occurred at an early date. The primary objective of this study was also to determine whether accelerated rates of evolution in the H/E (spl) members or bHLH domains are due to increased positive selection or decreased constraint. Present results provide little support for the positive selection hypothesis. By modeling the variable selective pressure among sequence sites of the enlarged H/E(spl) bHLH family, we found that all the 163 Hes gene lineages were under purifying selection and/or strong selective constraints, but the Hes2, Hes3, Hes5 lineages evolved faster than the other ones with relatively ω values. We found that nine sites of the Hes5 gene lineage had been under significantly positive selection. We further explored and described the evolutionary analysis for bHLH and/or HLH domains and found no site under positive selection. More generally, the failure to detect positive selection or adaptive substitution in E(spl) proteins and other transcription factors suggests that selective function constraint is relaxed and most substitutions are neutral in variable domains of these genes (Streisfeld and Rausher, 2007). This provides us an evolutionary scenario in which the diversity of E(spl) genes has been established through relaxed selective constraint. The present study deepens our knowledge of the animal Hairy and Enhancer of split bHLH transcription factors.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers for their constructive comments and suggestions. This study was supported by Research Funds to LWY (2011PTFY04ZD, KJ2012A216).

REFERENCES

Anisimova, M. and Z. Yang, 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol. Biol. Evol., 25: 1219-1228.

Anisimova, M., J.P. Bielawski and Z. Yang, 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol., 18: 1585-1592.

Artavanis-Tsakonas, S., K. Matsuno and M.E. Fortini, 1995. Notch signaling. Science, 268: 225-232.

Atchley, W.R. and W.M. Fitch, 1997. A natural classification of the basic helix-loop-helix class of transcription?factors. PNAS., 94: 5172-5176.

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.

Campos-Ortega, J.A., 1993. Mechanisms of early neurogenesis in *Drosophila melanogaster*. J. Neurobiol., 24: 1305-1327.

Campos-Ortega, J.A., 1994. Genetic mechanisms of early neurogenesis in *Drosophila melanogasteru*. J. Physiol. Paris, 88: 111-122.

Campos-Ortega, J.A., 1995. Genetic mechanisms of early neurogenesis in *Drosophila melanogaster*. Mol. Neurobiol., 10: 75-89.

Carretero-Paulet, L., A. Galstyan, I. Roig-Villanova, J.F. Martinez-Garcia, J.R. Bilbao-Castro and D.L. Robertson, 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.

Celis de J.F., J. de Celis, P. Ligoxygakis, A. Preiss, C. Delidakis and S. Bray, 1996. Functional relationships between Notch, Su(H) and the bHLH genes of the E(spl) complex: the E(spl) genes mediate only a subset of Notch activities during imaginal development. Development, 122: 2719-2728.

Fakheri, H.T., A.Z. Kiasari, T. Taghvaii, V. Hosseini, R.A. Mohammadpour *et al.*, 2010. Assessment the effect of midazolam sedation on hypoxia during upper gastrointestinal endoscopy. Pak. J. Biol. Sci., 13: 152-157.

Gonchar, O. and I. Mankovska, 2010. Antioxidant system in adaptation to intermittent hypoxia. J. Biol. Sci., 10: 545-554.

Greenwald, I., 1998. LIN-12/Notch signaling: Lessons from worms and flies. Genes Dev., 12: 1751-1762.

Guindon, S. and O. Gascuel, 2003. A simple, fast and accurate algorithm to estimate larges phylogenies by maximum likelihood. Syst. Biol., 52: 696-704.

Heim, M.A., M. Jakoby, M. Werber, C. Martin, B. Weisshaar and P.C. Bailey, 2003. The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol., 20: 735-747.

Huelsenbeck, J.P. and F. Ronquist, 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17: 754-755.

Jennings, B.H., D.M. Tyler and S.J. Bray, 1999. Target specificities of *Drosophila enhancer* of split basic helix-loop-helix proteins. Mol. Cell Biol., 19: 4600-4610.

Jones, D.T., W.R. Taylor and J.M. Thornton, 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci., 3: 275-282.

Jones, S., 2004. An overview of the basic helix-loop-helix proteins. Genome Biol., Vol. 5, 10.1186/gb-2004-5-6-226

Knust, E., H. Schrons, F. Grawe and J.A. Campos-Ortega, 1992. Seven genes of the Enhancer of split complex of *Drosophila melanogaster* encode helix-loop-helix proteins. Genetics, 132: 505-518.

Knust, E., K. Tietze and J.A. Campos-Ortega, 1987. Molecular analysis of the neurogenic locus Enhancer of split of *Drosophila melanogaster*. EMBO J., 6: 4113-4123.

Ledent, V. and M. Vervoort, 2001. The basic helix-loop-helix protein family: Comparative genomics and phylogenetic analysis. Genome Res., 5: 754-770.

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

Li, J., Q. Liu, M. Qiu, Y. Pan, Y. Li and T. Shi, 2006a. Identification and analysis of the mouse basic/Helix-Loop-Helix transcription factor family. Biochem. Biophys. Res. Commun., 350: 648-656.

Li, X., X. Duan, H. Jiang, Y. Sun and Y. Tang *et al.*, 2006b. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and arabidopsis. Plant Physiol., 141: 1167-1184.

Liu, F., J. Fei, N. Li and Q. Meng, 2010. The muscle transcription factor MyoD regulates the expression of hepatocyte growth factor during the process of muscle development. Asian J. Anim. Vet. Adv., 5: 557-565.

Liu, W. and C. Zhao, 2011. Molecular phylogenetic analysis of zebra finch basic Helix-Loop-Helix transcription factors. Biochem. Genetics, 49: 226-241.

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.

Massari, M.E. and C. Murre, 2000. Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. Mol. Cell. Biol., 20: 429-440.

Morgenstern, B. and W.R. Atchley, 1999. Evolution of bHLH transcription factors: modular evolution by domain shuffling?. Mol. Biol. Evol., 16: 1654-1663.

Murre, C., P.S. McCaw and D. Baltimore, 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins. Cell, 56: 777-783.

Muskavitch, M.A.T., 1994. Delta-notch signaling and Drosophila cell fate choice. Dev. Biol., 166: 415-430.

Nielsen, R. and Z. Yang, 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148: 929-936.

Pires, N. and L. Dolan, 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. Mol. Biol. Evol., 27: 862-874.

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: 2105-2110.

Ronquist, F. and J.P. Huelsenbeck, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 12: 1572-1574.

Simionato, E., P. Kerner, N. Dray, M. Le Gouar, V. Ledent, D. Arendt and M. Vervoort, 2008. Atonal-and achaete-scute-related genes in the annelid Platynereis dumerilii: Insights into the evolution of neural basic-Helix-Loop-Helix genes. BMC Evol. Biol., 10.1186/1471-2148-8-170

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., Vol. 7, 10.1186/1471-2148-7-33

Song, M., W. Han, H. Bao, C. Liu, C. Wu and C. Zhao, 2010. Physiological adaptability of tibet chicken embryo liver to hypoxia and its protein profile analysis. Asian J. Anim. Vet. Adv., 5: 547-556.

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.

Streisfeld, M.A. and M.D. Rausher, 2007. Relaxed constraint and evolutionary rate variation between basic helix-loop-helix floral anthocyanin regulators in Ipomoea. Mol. Biol. Evol., 24: 2816-2826.

Suyama, M., D. Torrents and P. Bork, 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res., 34: W609-W612.

Swanson, W.J., R. Nielsen and Q. Yang, 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol., 20: 18-20.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougi and D.G. Higgins, 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., 25: 4876-4882.

Toledo-Ortiz, G., E. Huq and P.H. Quail, 2003. The arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell, 15: 1749-1770.

Velleman, S.G., K.E. Nestor, C.S. Coy, I. Harford and N.B. Anthony, 2010. Effect of posthatch feed restriction on broiler breast muscle development and muscle transcriptional regulatory factor gene and heparan sulfate proteoglycan expression. Int. J. Poult. Sci., 9: 417-425.

Wang, Y., K. Chen, Q. Yao, X. Zheng and Z. Yang, 2009. Phylogenetic analysis of zebrafish basic helix-loop-helix transcription factors. J. Mol. Evol., 6: 629-640.

Yang, Z. and R. Nielsen, 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specic lineages. Mol. Biol. Evol., 19: 908-917.

Yang, Z., 2007. PAML 4: phylogenetic a nalysis by maximum likelihood. Mol. Biol. Evol., 24: 1586-1591.

Yang, Z., R. Nielsen, N. Goldman and A.M.K. Pedersend, 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155: 431-449.

Yang, Z.H., S.W.W. Wendy and N. Rasmus, 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol., 22: 1107-1118.

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.