Am J Transl Res 2016;8(11):4844-4856
www.ajtr.org /ISSN:1943-8141/AJTR0034984

Original Article

Phylogenetic analysis of basic helix-loop-helix transcription factors in the genome of a typical human-disease vector

Meng-Yun Chen[1,2,3*], Ying Dong[4*], Rui-Xue Chang[5*], Qian-Qian Ang[6], Ran Zhang[1,2,3], Yan-Yan Wu[1,2,3], Yi-Hui Xu[7], Wen-Sheng Lu[5], Xiao-Dong Zheng[1,2,3]

[1]Institute of Dermatology and Department of Dermatology at No.1 Hospital, Anhui Medical University, Hefei, Anhui, China; [2]Department of Dermatology and Venereology, Anhui Medical University, Hefei, Anhui, China; [3]State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology & Key laboratory of Dermatology, Ministry of Education, Hefei, Anhui, China; [4]Department of Dermatology, Anhui Provincial Children's Hospital, Hefei 230051, China; [5]Department of Dermatology, Affiliated Provincial Hospital, Anhui Medical University, Hefei 230001, China; [6]Allergy and Immunology Research Center, Anhui Medical University, Anhui, China; [7]School of Life Sciences, Anhui Medical University, Hefei, Anhui, China. [*]Equal contributors.

Received July 4, 2016; Accepted September 17, 2016; Epub November 15, 2016; Published November 30, 2016

Abstract: *Ixodes scapularis*, the black-legged tick, is one of the most common human-disease vectors and transmits *Borrelia* species, such as *B. burgdorferi*, as well as *Theileria microti*, *Anaplasma phagocytophilum*, etc. As basic helix-loop-helix (bHLH) transcription factors have been recognized for many years as important regulators of various developmental processes, we performed phylogenetic analysis of the black-legged tick genome in order to identify the number and family of bHLH transcription factors. Because bHLH family members have been identified in many organisms, including silkworm and fruit fly, we were able to conduct this survey and identify 58 putative bHLH transcription factors. Phylogenetic analysis revealed that the black-legged tick has 26, 10, 9, 1, 9, and 1 member in groups A, B, C, D, E, and F, respectively, whereas two were orphan genes. This analysis also revealed that unlike silkworm and fruit fly, the black-legged tick has no Mesp, Mlx, or TF4 family members, but has one more MyoRb family member. The present study provides useful background information for future studies of the black-legged tick as a disease vector with the goal of prevention and treatment.

Keywords: Disease vector, black-legged tick, bHLH transcription factor, phylogenetic analysis, gene ontology

Introduction

Ticks are one of the most significant vectors of human and animal pathogens, and they all belong to the Arthropoda. One species of this phylum, *Ixodes scapularis*, the black-legged tick or deer tick, can transmit the pathogens causing Lyme disease [1], human granulocytic anaplasmosis, babesiosis [2], and others to humans and other species. Because the ticks are usually infected with more than one pathogen, and co-infection makes diagnosis and treatment difficult and often elusive, they are becoming a huge threat to human health [3].

The basic helix-loop-helix (bHLH) transcription factors play essential roles in various developmental processes [4]. Two of these proteins pair to form homodimers or heterodimers. Generally, the bHLH motif has 60 amino acids, of which 19 are highly conserved in organisms ranging from yeast to mammals. Atchley WR et al. constructed a predicted motif of bHLH proteins based on statistical analysis of amino acid frequencies within the bHLH motif [5]. Owing to their important functions, bHLH transcription factors have been intensively studied in various organisms. Through examination of amino acids at the 19 highly conserved sites, more than 1000 bHLH sequences have been identified in organisms whose genome sequences are available. E12 and E47 were first reported by Murre C et al. in 1989 [6]. Since then, many bHLH proteins have been identified in numerous species. Among plants, rice and *Arabidopsis* have 167 and 147 bHLH members, respectively [7, 8]. In

mammals, the human and mouse genomes were found to encode 118 and 114 bHLH proteins, respectively [9, 10]. Additionally, other species, especially insects such as silkworm (*Bombyx mori*) and fruit fly (*Drosophila melanogaster*) were found to possess 52 and 59 bHLH members, respectively [11]. Based on phylogenetic analyses, Ledent et al. (2002) defined 44 orthologous families and 6 supergroups based on DNA-binding activity after large-scale phylogenetic analyses. Subsequently, Simionato, et al. (2007) revised this classification and defined 45 families in metazoans [5, 9, 12].

The black-legged tick is estimated to have diverged from other arthropods, such as the fruit fly and silkworm, more than 750 million years ago. The fruit fly and silkworm have both been surveyed regarding bHLH members [9, 11]. Consequently, genome sequencing and other genomic studies are anticipated to be highly informative for the analysis and control of tick-borne diseases. However, bHLH transcription factors have not been systematically characterized and reported in the black-legged tick. Questions such as how many bHLH genes exist and to what families they belong have not been addressed. Identification of bHLH members encoded in the black-legged tick genome will greatly facilitate studies of Arthropoda developmental biology and human granulocytic anaplasmosis and more. To provide an overview of the whole set of black-legged tick bHLH family members, we conducted a genome-wide survey of the latest version of the black-legged tick genome sequence and defined names and families for the identified bHLH members through phylogenetic analyses.

Materials and methods

BLAST searches and sequence retrieval

We gathered amino acid sequences of 45 representative bHLH motifs from the files of previous reports [9, 13]. Each sequence was used to perform tblastn and blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=OGP__6945__16233 and http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=OGP__6945__16233&LINK_LOC=blasttab, respectively) searches against genomic sequences of black-legged tick. Next, bHLH domains of silkworm (BmbHLH) and fruit fly (DmbHLH) were used to perform tblastn and blastp searches to confirm that no potential black-legged tick bHLH domains (IsbHLH) were missed. The sequences obtained were then manually examined to eliminate redundant sequences, add the missing amino acids on two ends of the bHLH motif, and find introns within the bHLH motifs. Intron analysis was done using the NetGene2 application online (http://www.cbs.dtu.dk/services/NetGene2/).

Phylogenetic analyses

We used PAUP 4.0 Beta 10 (http://paup.csit.fsu.edu/about.html) to perform phylogenetic analyses and then constructed neighbor-joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) trees of each amino acid sequence of the identified black-legged tick bHLH motifs with 45 representative bHLH motifs, 59 DmbHLH, and 52 BmbHLH motifs. NJ trees were bootstrapped with 1000 replicates; MP analysis was performed using heuristic searches and bootstrapped with 100 replicates; and ML trees were constructed using TreePuzzle 5.2 [14, 15].

Sequence alignment and conserved-domain shading

After phylogenetic analyses, all sequences were aligned using ClustalW online (http://www.ebi.ac.uk/clustalw/) with default settings, followed by examination of the amino acid residues of the 19 conserved sites. Sequences with more than nine conserved sites, as well as one COE family member that has eight conserved sites, were regarded as potential black-legged tick bHLHs. The aligned IsbHLHs were shaded in the GeneDoc Multiple Sequence Alignment Editor and Shading Utility (ver. 2.6.02) [16] and copied to a rich text file (RTF) for further annotation.

Sequence logo and protein interaction network (PIN)

Multiple alignment of all identified IsbHLH domains was performed using MEGA software (http://www.megasoftware.net/), and the result of sequence alignment was used as input file to construct a sequence logo on the RTH website (http://rth.dk/resources/plogo/#form) [17, 18].

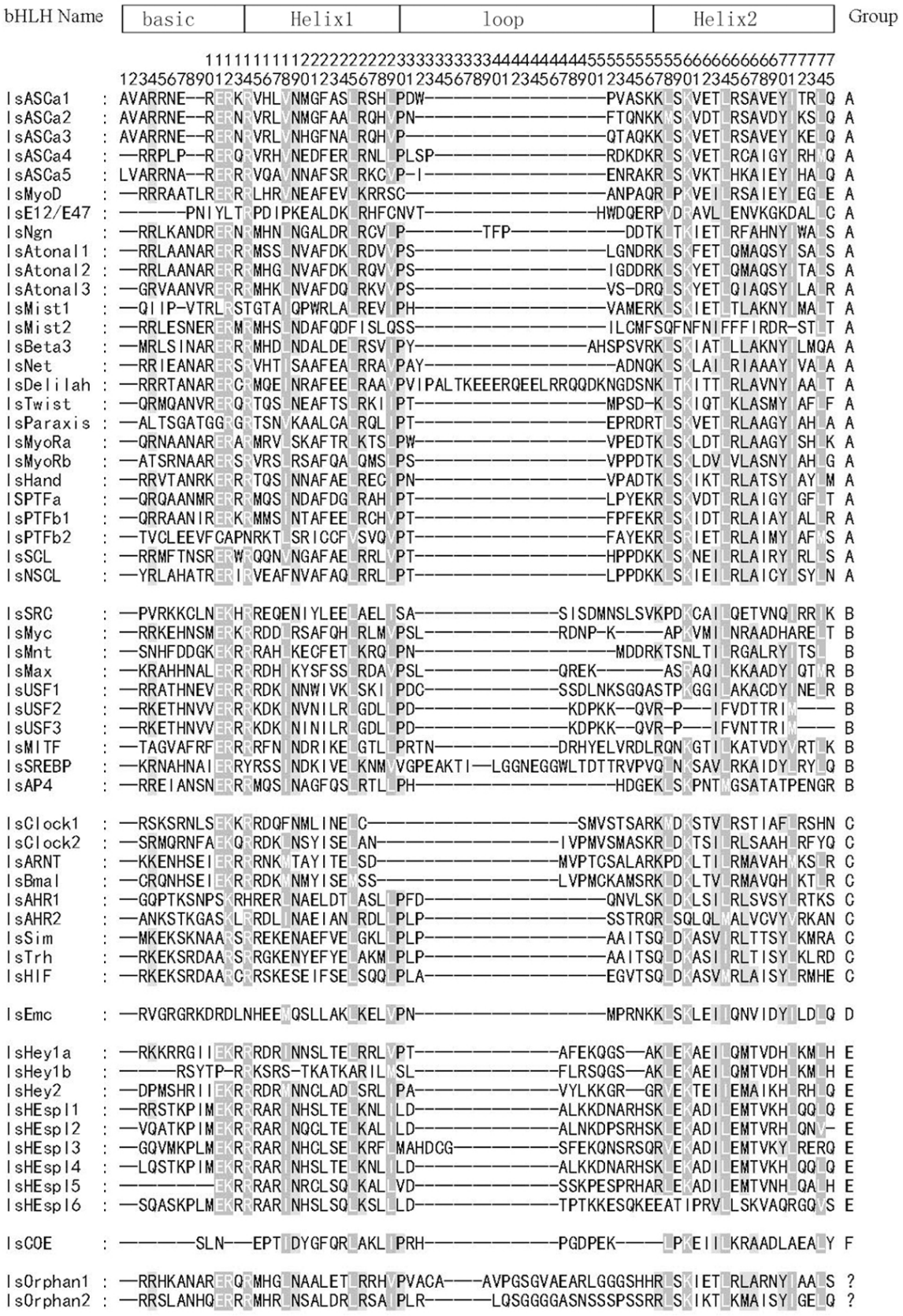

Figure 1. Multiple sequence alignment of 58 black-legged tick bHLH motifs. Designation of basic, helix 1, loop and helix 2 follows Ferre-D'Amare et al. [25]. The family names and high-order groups have been organized according to **Table 1** in Ledent et al. [12]. Highly conserved amino acid residues are shaded in black.

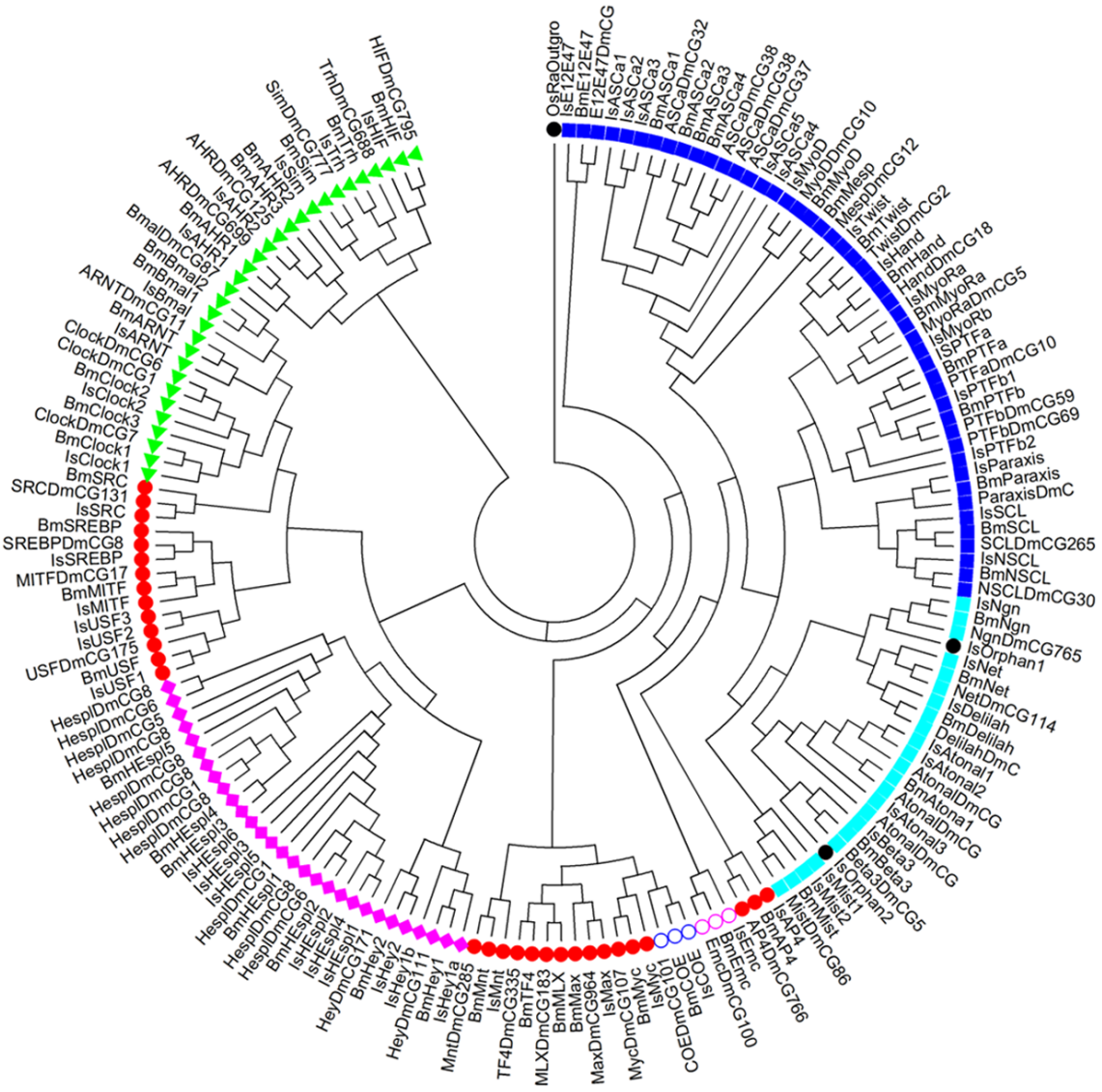

Figure 2. The neighbor-joining (NJ) phylogenetic tree in circle format. The phylogenetic relationship constructed by 58 IsbHLH members, 52 BmbHLH domains and 59 DmbHLHs domains. The higher-order group labels are in accordance with Ledent et al. [12].

Using the protein-protein interaction online search tool STRING (version 10.0) (http://string-db.org/) [19, 20], we identified direct or indirect interactions between those black tick proteins (*Ixodes scapularis* organism, STRING protein database) that had 100% identity with 58 IsbHLH members.

Results

Phylogenetic analyses and identification of IsbHLH transcription factors

Tblastn and blastp searches with 45 bHLH domains, 59 DmbHLH motifs, and 52 BmbHLH motifs of the black-legged tick genome were carried out, followed by removal of the redundant sequences and examination of the 19 conserved amino acid sites. We finally identified 58 putative bHLH transcription factors encoded in the black-legged tick genome (**Figure 1**).

In this study, using 58 putative IsbHLH domains, 59 DmbHLH motifs, and 52 BmbHLH motifs, we constructed an overall NJ tree to investigate the evolutionary relationships among three insects (**Figure 2**). After that, each identified black-legged tick bHLH sequence was used to conduct in-group phylogenetic analyses with

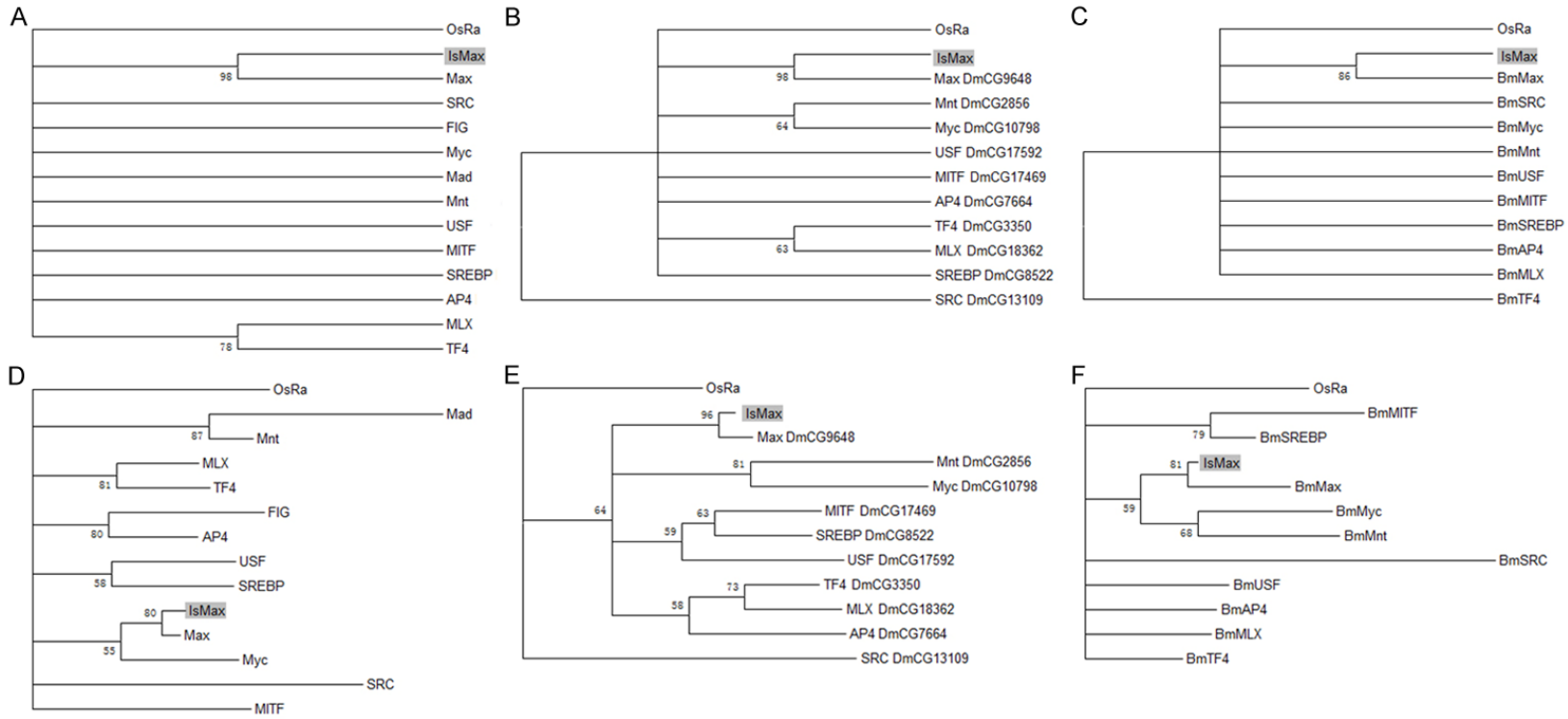

Figure 3. In-group phylogenetic analyses of IsMax. A-C: Are MP trees constructed by one black-legged tick bHLH member (IsMax) and all the group B bHLH members of bHLH family, Fruit fly and Silkworm respectively, besides; D-F: Are constructed ML trees. In all trees, OsRa was used as the out group.

bHLH domains of fruit fly and silkworm, which enabled us to allocate all the identified IsbHLHs to defined evolutionarily conserved orthologous groups. **Figure 3**, as examples here, shows two groups, six in-group trees that were constructed with IsbHLH member (IsMax) and group B bHLH members from the fruit fly and silkworm bHLH motif, respectively. The bootstrap values among these IsbHLH, DmbHLH, and BmbHLH domains are listed in **Table 1**. In all these phylogenetic trees, we used OsRa (the rice bHLH sequence of R family) sequence as outgroup.

These phylogenetic analysis results allow us to define the name of each IsbHLH. One criterion was used; i.e., a single IsbHLH must form a monophyletic group with another bHLH of a known family in phylogenetic trees constructed by at least three methods and having bootstrap value exceeding 50. The results show that, among all 58 black-legged tick bHLH members, 23 have all NJ, MP, and ML bootstrap values >50 (ranging from 54 to 100), enabling us to confidently assign corresponding fruit fly and silkworm homologues for them. Second, 30 black-legged tick bHLH members have two or more bootstrap values >50 and some n/m* or n/m (see explanation in **Table 1**) in the constructed phylogenetic trees, whereas three other members have constructed conserved topologies with exactly one family in three species. Although these three black-legged tick bHLH members were not supported by adequate bootstrap values, we assigned the corresponding homologues for them by considering that most of the values had supported the formation of a monophyletic clade with the same fruit fly or silkworm counterpart. In particular, the IsbHLH member that we defined as IsASCa5 always formed one conserved topology with the ASCa family and had no bootstrap value exceeding 50. However, these assignments can be regarded as arbitrary and are subject to modification upon acquisition of new data.

Each black-legged tick bHLH gene was named according to its phylogenetic relationship (explained below) with the corresponding bHLH domain, fruit fly or silkworm homologue. In the case where one known bHLH sequence has two or more black-legged tick homologues, we used "1,", "2,", "3," etc., to number them. For instance, two homologues of the MmMIST1 domain and the MmUSF1 domain were found in black-legged tick. Thus, the black-legged tick genes were named IsMist1 and IsMist2; and IsUSF1, IsUSF2, and IsUSF3, respectively.

It was found that the black-legged tick has 26, 10, 9, 1, 9, and 1 bHLH genes in groups A, B, C, D, E, and F, respectively. An additional two members could not be assigned to any known families and were thus regarded as "orphan". By cross validation, we found no Mesp, Mlx, or TF4 family members in the black-legged tick genome with blastp or tblastn searches using members from the silkworm or fruit fly genome. Further, there was one additional bHLH family member identified as BmMyoRb in the silkworm genome by blastp search using IsMyoRb, which was found in the black-legged tick genome.

Sequence alignment and domain conservation analyses

Protein sequence accession numbers and their genomic contig numbers for all 58 black-legged tick bHLH motifs are also listed in **Table 1**. It should be noted that when the amino acid sequence of an individual *I. scapularis* bHLH motif was used to conduct blastp searches against various black-legged tick protein databases, generally a considerable number of "hits" with 100% identity with the bHLH motif could be obtained. These protein sequences often varied slightly in length. Yet most did not represent different protein sequences encoded in the black-legged tick genome, because most tblastn searches using the amino acid sequence of each black-legged tick bHLH motif against the black-legged tick genome yielded only one coding region in the genome.

After identification of all these domains, we performed multiple sequence alignment of 58 bHLH domain sequences to analyze the sequence features of IsbHLH members. Most of the IsbHLH domains had more than 10 conserved amino acid residues, and highly conserved amino acid residues are shaded in black. The results of alignment of all 58 IsbHLH motifs showed that the basic region (13 amino acids) and two helix regions (16 and 19 amino acids) are more highly conserved than the loop region (27 amino acids). Further, 18 conserved residues were identified in the IsbHLH domains (**Figure 1**).

Table 1. The complete list of 58 bHLH transcription factors from the black-legged tick

IsbHLH	bHLH family homolog	Fruit fly homolog	Silkworm homolog	NJ	MP	ML	Accession Number	Groups
IsASCa1	MmMash1	ASCa?	BmASCa?	n/m*,n/m*,n/m*	89,n/m*,n/m*	91,n/m*,n/m*	XP_002414095.1	A
IsASCa2	MmMash1	ASCa?	BmASCa?	n/m*,n/m*,n/m*	98,n/m*,n/m*	90,n/m*,n/m*	XP_002409128.1	A
IsASCa3	MmMash1	ASCa?	BmASCa?	n/m*,n/m*,n/m*	98,n/m*,n/m*	87,n/m*,n/m*	XP_002407277.1	A
IsASCa4	MmMash3	Mesp(DmCG12952)	BmASCa?	54,18,n/m*	n/m*,n/m*,n/m	n/m*,n/m, n/m	XP_002406630.1	A
IsASCa5	MmMash1/MmMash3	ASCa?	BmASCa?	n/m,n/m*,n/m*	n/m*,n/m*,n/m*	n/m*,n/m*,n/m	XP_002400075.1	A
IsHand	MmdHand	Hand(DmCG18144)	BmHand	100,98,99	100,77,97	82,n/m,67	XP_002410508.1	A
IsAtonal01	MmMath1	Atonal(DmCG10393)	BmAtonal1	72,n/m*,n/m*	97,80,78	94,n/m*,84	XP_002433778.1	A
IsAtonal02	MmMath1	Atonal?	BmAtonal1	72,n/m*,n/m*	95,n/m,71	94,n/m,79	XP_002433775.1	A
IsAtona103	MmMath1	Atonal(DmCG7760)	BmAtonal1	94,n/m*,n/m*	59,n/m,54	91,68,83	XP_002433776.1	A
IsNgn	MmMath4A	Ngn(DmCG7659)	BmNgn	99,98,100	92,98,97	71,84,73	XP_002409093.1	A
IsDelilah	BfDelilah	Delilah(DmCG5441)	BmDelilah	43,79,83	n/m,73,87	79,58,58	XP_002400731.1	A
IsMist1	MmMIST1	Mist(DmCG8667)	BmParaxis	49,n/m,n/m*	n/m,76,n/m	88,63,n/m	?	A
IsMist2	MmMIST1	Mist(DmCG8667)	BmMist	n/m,54,50	68,74,62	82,75,58	XP_003745166.1	A
IsBeta3	MmBeta3	Beta3(DmCG5545)	BmBeta3	100,100,100	88,100,100	84,58,64	XP_002410334.1	A
IsNet	MmMATH6	Net(DmCG11450)	BmNet	99,99,98	95,96,100	73,58,67	XP_002412927.1	A
IsMyoD	MmMyoD	MyoD(DmCG10250)	BmMyoD	100,99,99	81,100,77	64,60,62	XP_002403013.1	A
IsTwist	MmTwist	Twist(DmCG2956)	BmTwist	98,100,99	93,97,99	83,84,67	XP_002405105.1	A
IsParaxis	MmParaxis	Paraxis(DmCG12648)	BmParaxis	47,44,85	61,59,87	77,85,n/m	?	A
IsSCL	MmLyl1	SCL(DmCG2655)	BmSCL	100,100,100	100,95,100	85,55,65	XP_002410238.1	A
IsMyoRa	HsMyoRa	MyoRa(DmCG5005)	BmMyoRa	100,99,100	100,96,98	54,58,91	XP_002408270.1	A
IsMyoRb	HsMyoRb	MyoRa(DmCG5005)	BmMyoRa	99,n/m*,n/m*	93,n/m,61	71,n/m,90	XP_002435930.1	A
ISPTFa	HsPTFa	PTFa(DmCG10066)	BmPTFa	100,99,100	98,76,94	69,54,66	XP_002434241.1	A
IsPTFb1	HsPTFb	PTFb(DmCG5952)	BmPTFb	n/m*,93,90	50,69,89	61,51,69	XP_002408569.1	A
IsPTFb2	HsPTFb	PTFb(DmCG6913)	BmPTFb	48,40,n/m*	73,86,94	64,76,n/m	?	A
IsNSCL	MmHen1	NSCL(DmCG3052)	BmNSCL	100,100,100	100,98,100	78,63,66	XP_002414911.1	A
IsE12/E47	MmE2a	E12/E47(DmCG5102)	BmE12E47	n/m,n/m*,62	n/m,58,86	72,90,58	XP_002407150.1	A
IsMnt	MmMNT	Mnt(DmCG2856)	BmMnt	99,99,99	92,85,97	60,71,73	XP_002404397.1	B
IsAP4	MmAP4	AP4(DmCG7664)	BmAP4	98,96,98	97,84,100	89,93,95	XP_002400892.1	B
IsSRC	MmSRC1	SRC(DmCG13109)	BmSRC	100,98,82	96,95,91	82,74,70	XP_002402360.1	B
IsMyc	MmNMyc	Myc(DmCG10798)	BmMyc	53,60,64	79,62,n/m	n/m,66,n/m	XP_002415391.1	B
IsMax	MmMax	Max(DmCG9648)	BmMax	100,100,99	98,98,86	80,96,81	XP_002415337.1	B
IsUSF1	MmUSF1	USF(DmCG17592)	BmUSF	100,99,99	100,89,100	78,89,76	XP_002416465.1	B
IsUSF2	MmUSF1	USF(DmCG17592)	BmUSF1	n/m*,n/m*,n/m*	n/m,58,n/m	n/m,80,74	?	B
IsUSF3	MmUSF1	USF(DmCG17592)	BmUSF2	n/m*,n/m*,n/m*	n/m,54,n/m	n/m,82,80	?	B
IsMITF	MmMITF	MITF(DmCG17469)	BmMITF	100,99,99	96,94,100	77,78,80	XP_002412123.1	B
IsSREBP	MmSREBP1	SREBP(DmCG8522)	BmSREBP	100,100,100	100,98,99	77,73,82	XP_002408259.1	B

IsARNT	MmARNT2	ARNT(DmCG11987)	BmARNT	100,99,100	99,100,100	100,99,100	XP_002416629.1	C
IsClock1	MmNPAS2	Clock(DmCG7391)	BmClock1	99,100,100	100,92,98	97,96,70	XP_002406740.1	C
IsClock2	MmNPAS2	Clock?	BmClock2	n/m*,n/m*,83	n/m,n/m*,86	n/m,n/m,74	XP_002410323.1	C
IsBmal	MmBmal1	Bmal(DmCG8727)	?	88,97,n/m*	58,64,n/m*	n/m,69,n/m*	?	C
IsAHR1	MmAHR	AHR(DmCG6993)	BmAHR1	100,100,100	99,100,100	100,81,94	XP_002401283.1	C
IsAHR2	?	AHR(DmCG12561)	BmAHR3	n/m,100,75	n/m,100,n/m*	n/m,84,n/m*	XP_002412571.1	C
IsSim	MmSim1	Sim(DmCG7771)	BmSim	77,98,98	57,99,93	56,84,84	?	C
IsTrh	MmNPAS3	Trh(DmCG6883)	BmTrh	94,100,100	84,94,100	66,98,100	XP_002434754.1	C
IsHIF	MmEPAS1	HIF(DmCG7951)	BmHIF	89,92,88	86,78,91	81,75,98	XP_002414889.1	C
IsEmc	MmId2	Emc(DmCG1007)	EmcDmCG	100,99,99	88,85,72	88,85,n/m	XP_002414917.1	D
IsHey1a	MmHey1	Hey(DmCG11194)	BmHey1	100,97,99	100,89,96	100,94,98	XP_002412333.1	E
IsHey1b	MmHey1	Hey(DmCG11194)	BmHey1	n/m*,n/m*,n/m*	73,n/m,76	100,66,90	?	E
IsHey2	MmHey1	Hey(DmCG17100)	BmHey2	n/m*,100,100	n/m,100,100	n/m,100,100	XP_002400926.1	E
IsH/E(spl)1	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	100,n/m*,64	100,n/m*,62	XP_002408713.1	E
IsH/E(spl)2	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	100,n/m*,55	100,n/m*,76	XP_002408712.1	E
IsH/E(spl)3	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	n/m,n/m*,n/m*	100,n/m*,n/m*	?	E
IsH/E(spl)4	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	100,n/m*,82	100,n/m*,72	?	E
IsH/E(spl)5	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	91,n/m*,n/m*	91,n/m*,n/m*	XP_002415605.1	E
IsH/E(spl)6	MmHES1	H/E(spl)?	BmH/E(spl)2	n/m*,n/m*,n/m*	63,n/m,n/m	63,n/m*,n/m*	XP_002403347.1	E
IsCOE	COE_MmCoe1	COEDmCG	BmCOE	100,100,100	n/m,n/m*,n/m*	n/m,n/m*,n/m*	XP_002412012.1	F
IsOrphan1	MmMath3	Ngn(DmCG7659)	BmNgn	64,n/m*,n/m*	n/m,n/m,n/m	n/m,n/m,n/m	XP_002411626.1	O
IsOrphan2	MmAP4	AP4(DmCG7664)	BmAP4	n/m,n/m*,n/m*	n/m,61,n/m	n/m,n/m,n/m	?	O

In NJ column, the number or symbol indicates the bootstrap value from NJ phylogenetic analysis using 58 IsbHLH members and 45 bHLH family motifs, 59 DmbHLH members, 52 BmbHLH members respectively. While MP, ML column indicates the bootstrap value were in-group phylogenetic analysis using the same higher-order group of bHLH family motifs, DmbHLH members and BmbHLH members respectively. n/m = none monophyletic; n/m* = an individual IsbHLH sequence did not form a monophyletic group with another bHLH sequence of known family, but formed a monophyletic group with other bHLH sequences of the same family; IsbHLH genes were named according to the nomenclature used in silkworm.

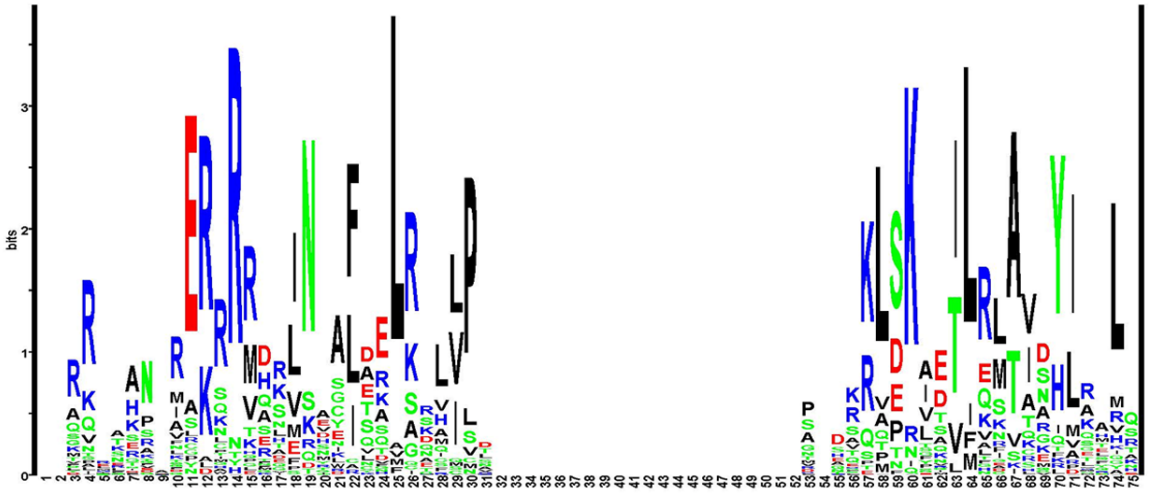

Figure 4. Sequence logo of all 58 IsbHLH members. The height of each symbol is in proportion to the fraction of the observed frequency and the expected frequency, besides, the symbol will be displayed up-side-down when it appears less than expected. The total height of the sequence information part is computed as the relative entropy between the observed fractions of a given symbol and the respective a priori probability, with the constraint that the a priori "probability" of the gap has always been one.

Table 2. 58 IsbHLH domains annotated by STRING black-legged tick protein database

Query sequence	STRING protein	Annotation	Identity	Bitscore
ISPTFa	ISCW003737-PA	Pancreas-specific transcription factor 1A, ptfa, putative	100%	112
IsAHR1	ISCW019892-PA	Aryl hydrocarbon receptor, putative	100%	107
IsAHR2	ISCW012136-PA	Hypothetical protein	100%	108
IsAP4	ISCW013477-PA	Activator protein, putative	100%	108
IsARNT	ISCW023999-PA	Aryl hydrocarbon receptor nuclear translocator, putative	100%	110
IsASCa1	ISCW013371-PA	Helix-loop-helix protein, putative	100%	113
IsASCa2	ISCW009120-PA	Conserved hypothetical protein	100%	108
IsASCa3	ISCW001920-PA	Conserved hypothetical protein	100%	108
IsASCa4	ISCW008705-PA	Conserved hypothetical protein	100%	104
IsASCa5	ISCW018149-PA	Conserved hypothetical protein	100%	95.1
IsAtonal1	ISCW003677-PA	Transcription factor, putative	100%	107
IsAtonal2	ISCW003667-PA	Transcription factor, putative	100%	109
IsAtonal3	ISCW003670-PA	Conserved hypothetical protein	100%	107
IsBeta3	ISCW009607-PA	Basic helix-loop-helix protein, putative	100%	110
IsBmal	ISCW023999-PA	Aryl hydrocarbon receptor nuclear translocator, putative	64%	79.7
IsCOE	ISCW021817-PA	IPT/TIG domaincontaining protein	100%	89.7
IsClock1	ISCW007040-PA	Circadian locomoter output cycles kaput protein, putative	100%	103
IsClock2	ISCW008805-PA	Helixloop-helix DNA-binding domain-containing protein	100%	111
IsDelilah	ISCW016573-PA	Conserved hypothetical protein	100%	147
IsE12/E47	ISCW019963-PA	Glyoxylate/hydroxypyruvate reductase, putative	100%	109
IsEmc	ISCW014470-PA	DNA-binding protein inhibitor, putative	100%	106
IsHEspl1	ISCW016540-PA	Transcription factor hes-1, putative	100%	118
IsHEspl2	ISCW016537-PA	Transcription factor hes-1, putative	100%	117
IsHEspl3	ISCW016540-PA	Transcription factor hes-1, putative	57%	67.8
IsHEspl4	ISCW016540-PA	Transcription factor hes-1, putative	98%	114
IsHEspl5	ISCW013710-PA	Transcription factor hes-1, putative	100%	102

IsHEspl6	ISCW017801-PA	Transcription factor hes-1, putative	100%	115
IsHIF	ISCW023657-PA	Hypoxia-inducible factor 1 alpha, putative	100%	108
IsHand	dhand	dhand, putative	100%	109
IsHey1a	ISCW011251-PA	Transcription factor hey, putative	100%	114
IsHey1b	ISCW011251-PA	Transcription factor hey, putative	63%	53.9
IsHey2	ISCW004855-PA	Conserved hypothetical protein	100%	115
IsMITF	ISCW011796-PA	Microphthalmia-associated transcription factor, putative	100%	128
IsMax	ISCW013570-PA	Upstream transcription factor 2/L-myc-2 protein, putative	100%	106
IsMist1	STRING found no matching protein in it's database			
IsMist2	STRING found no matching protein in it's database			
IsMnt	ISCW007968-PA	Max binding protein mnt, putative	100%	115
IsMyc	ISCW023000-PA	REF-cotransformation assay oncogene, putative	100%	109
IsMyoD	ISCW009440-PA	Myogenic factor 6, putative	100%	102
IsMyoRa	ISCW019706-PA	Musculin, putative	100%	108
IsMyoRb	ISCW024416-PA	Transcription factor basichelix-loop-helix protein	100%	107
IsNSCL	ISCW014797-PA	Helix-loop-helix protein, putative	100%	108
IsNet	ISCW011969-PA	Predicted protein	100%	103
IsNgn	ISCW020498-PA	Neurogenin-2, putative	100%	117
IsOrphan1	ISCW021517-PA	Neurogenic differentiation factor, putative	100%	139
IsOrphan2	ISCW009607-PA	Basic helix-loop-helix protein, putative	53%	52
IsPTFb1	ISCW024442-PA	N-twist protein, putative	100%	109
IsPTFb2	STRING found no matching protein in it's database			
IsParaxis	STRING found no matching protein in it's database			
IsSCL	ISCW010324-PA	Helix-loop-helix protein hen, putative	100%	112
IsSRC	ISCW017872-PA	Nuclear receptor coactivator, putative	100%	120
IsSREBP	ISCW018168-PA	Sterol regulatory element-binding protein, putative	100%	152
IsSim	ISCW018625-PA	Neuronal pas domain protein, putative	81%	86.7
IsTrh	ISCW018625-PA	Neuronal pas domain protein, putative	100%	110
IsTwist	musculin	Musculin, putative	100%	107
IsUSF1	ISCW015344-PA	Upstream stimulatory factor 2, usf2, putative	100%	123
IsUSF2	STRING found no matching protein in it's database			
IsUSF3	STRING found no matching protein in it's database			

The identified IsbHLH domains are annotated by Ixodes scapularis protein database, and only the STRING protein, which has 100 percent identical with corresponding domain will be selected to proceed to the association network.

Sequence logo and PIN results

We used the results of multiple alignment of these IsbHLH domains to construct a protein sequence logo. The outcome shows that multiple amino acid sides are conserved, which is as same as conserved domain shad analysis (**Figure 4**).

Using the STRING online annotation tool, we found that 46 IsbHLH motifs have 100% identity with *I. scapularis* proteins in the STRING database (**Table 2**), and we constructed the interaction network with these proteins (**Figure 5**). We found that about 10 IsbHLH proteins dissociate from the core-connected functional modules, which are composed of more than 20 other IsbHLH proteins.

Discussion

I. scapularis is commonly known as the deer tick or black-legged tick and is estimated to have diverged from other arthropods, such as the fruit fly, over 750 million years ago. It is a vector for several diseases of animals, including humans, mice [21], lizards [22], migratory birds [23], etc., especially while the tick is in the

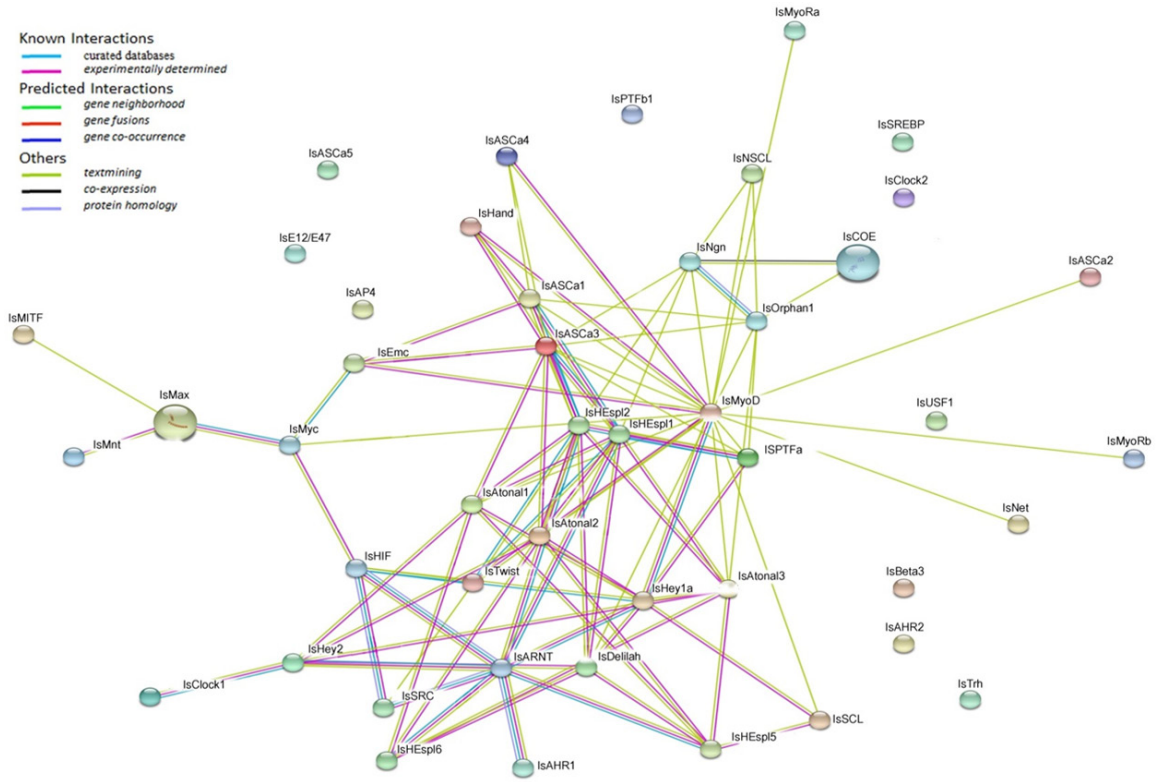

Figure 5. The STRING mapping profile of protein-protein interaction network (PIN) showing the functional interaction of IsbHLH transcription factors. There are 46 black-legged tick proteins which have 100 percent score of identity with 58 IsbHLH members have been inputted in the network. Lines of different colors represent the types of evidence for the protein functional association. The small panel in the bottom left magnified the implication of different connecting lines with different data sources in the main figure (color figure online).

larval or nymphal stage. Consequently, genome sequencing and other genomic studies are anticipated to be highly informative for the analysis and control of tick-borne diseases. Because the latest version of the black-legged tick genome sequence has been available in GenBank since August 2009, we had the opportunity to identify 58 IsbHLH proteins by tblastn and blastp searches in black-legged tick genome databases. Based on phylogenetic analyses of IsbHLH domains with orthologous bHLH family sequences, DmbHLH domains and BmbHLH domains, we assigned these 56 IsbHLH genes to the supergroups A, B, C, D, E, and F, and two members were categorized as orphan.

After obtaining the 58 black-legged tick bHLH members, we found MyoRb only in the black-legged genome, so we investigated the presence of a MyoRb ortholog in the silkworm and fruit fly genomes. We conducted the same tblastn and blastp searches of the latest version of the silkworm and the fruit fly genomes with the tick MyoRb member. As expected, we found a new MyoRb member in the silkworm genome, which was named BmMyoRb, and the identified sequence is EPHRNAANARERARMRVLSKAFCRLKTTLPWVPADTKLSKLDTLRLAASYIAHLR (Accession Number: NP_001124342.1). But a similar result was not found in the fruit fly genome. By gene ontology (GO) annotation [24], we considered that the MyoRb family is associated with multiple developmental processes, such as glomerular capillary formation and muscle organ development, etc. (data not shown). On the other hand, genes that encode Mesp, Mlx, or TF4 were not found in the black-legged tick genome which may be the result of the incomplete genome sequence.

In conclusion, the present study deepens our understanding of, and provides insight into, the evolution and functional aspects of insect bHLH proteins and should serve as a solid foundation for further analyses of specific bHLH

transcription factors in other insects. Further, the identification of black-legged tick bHLH family members and investigation of their significance in genetic evolutionary events provide useful information for studies on Arthropoda development and for related studies in Ixodida, silkworm, fruit fly, and other fly species. Finally, this study provides a framework for applying a whole-genome phylogenetic analysis of IsbHLH transcription factors and will aid future investigations of the black-legged tick, which is one of the most common vectors of human disease.

Acknowledgements

The research was supported by the incubation program of the national youth fund, the first affiliated hospital of Anhui medical university (2560). This work was also funded by Anhui Provincial Natural Science Foundation (1608085MH175).

Disclosure of conflict of interest

None.

Address correspondence to: Wen-Sheng Lu, Department of Dermatology, Affiliated Provincial Hospital, Anhui Medical University, No. 17, Lu Jiang Road, Hefei 230001, Anhui, PR China. Tel: +8655162283151; Fax: +8655162283151; E-mail: luws2004@sina.com; Xiao-Dong Zheng, Institute of Dermatology and Department of Dermatology at No. 1 Hospital, Anhui Medical University, Hefei 230032, Anhui, China. Tel: +8655165161053; Fax: +8655165161053; E-mail: zhengxiaodong.1234@163.com

References

[1] Brownstein JS, Holford TR and Fish D. Effect of Climate Change on Lyme Disease Risk in North America. EcoHealth 2005; 2: 38-46.
[2] Steere AC. Lyme disease. New Engl J Med 2001; 345: 115-125.
[3] Varde S, Beckley J and Schwartz I. Prevalence of tick-borne pathogens in Ixodes scapularis in a rural New Jersey County. Emerg Infect Dis 1998; 4: 97-99.
[4] Massari ME and Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol 2000; 20: 429-440.
[5] Atchley WR, Terhalle W and Dress A. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J Mol Evol 1999; 48: 501-516.
[6] Murre C, McCaw PS and Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 1989; 56: 777-783.
[7] Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J and Zhang D. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. Plant Physiol 2006; 141: 1167-1184.
[8] Toledo-Ortiz G, Huq E and Quail PH. The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell 2003; 15: 1749-1770.
[9] Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM and Vervoort M. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol Biol 2007; 7: 33.
[10] Zheng X, Wang Y, Yao Q, Yang Z and Chen K. A genome-wide survey on basic helix-loop-helix transcription factors in rat and mouse. Mamm Genome 2009; 20: 236-246.
[11] Wang Y, Chen K, Yao Q, Wang W and Zhi Z. The basic helix-loop-helix transcription factor family in Bombyx mori. Dev Genes Evol 2007; 217: 715-723.
[12] Ledent V, Paquet O and Vervoort M. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol 2002; 3: RESEARCH0030.
[13] Ledent V and Vervoort M. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res 2001; 11: 754-770.
[14] Schmidt HA, Strimmer K, Vingron M and von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 2002; 18: 502-504.
[15] Jones DT, Taylor WR and Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 1992; 8: 275-282.
[16] Nicholas KB, Nicholas HB Jr and Deerfield DW 2nd. GeneDoc: analysis and visualization of genetic variation. Embnet News 1997; 4: 14.
[17] Gorodkin J, Heyer LJ, Brunak S and Stormo GD. Displaying the information contents of structural RNA alignments: the structure logos. Comput Appl Biosci 1997; 13: 583-586.
[18] Schneider TD and Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 1990; 18: 6097-6100.
[19] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P and von Mering C. STRING

8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 2009; 37: D412-416.
[20] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-452.
[21] Mannelli A, Kitron U, Jones CJ and Slajchert TL. Influence of season and habitat on Ixodes scapularis infestation on white-footed mice in northwestern Illinois. J Parasitol 1994; 80: 1038-1042.
[22] Levine JF, Apperson CS, Howard P, Washburn M and Braswell AL. Lizards as hosts for immature Ixodes scapularis (Acari: Ixodidae) in North Carolina. J Med Entomol 1997; 34: 594-598.
[23] Ogden NH, Lindsay LR, Hanincova K, Barker IK, Bigras-Poulin M, Charron DF, Heagy A, Francis CM, O'Callaghan CJ, Schwartz I and Thompson RA. Role of migratory birds in introduction and range expansion of Ixodes scapularis ticks and of Borrelia burgdorferi and Anaplasma phagocytophilum in Canada. Appl Environ Microbiol 2008; 74: 1780-1790.
[24] Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY and Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 2011; 39: W316-322.
[25] Ferre-D'Amare AR, Prendergast GC, Ziff EB and Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature 1993; 363: 38-45.