BioOne COMPLETE

The Basic Helix-Loop-Helix Transcription Factor Family in the Honey Bee, Apis mellifera

Authors: Wang, Yong, Chen, Keping, Yao, Qin, Wang, Wenbing, and Zhu, Zhi

Source: Journal of Insect Science, 8(40) : 1-12

Published By: Entomological Society of America

URL: https://doi.org/10.1673/031.008.4001

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

The basic helix-loop-helix transcription factor family in the honey bee, *Apis mellifera*

Yong Wang[1,a], Keping Chen[2,b], Qin Yao[2,c], Wenbing Wang[2,d] and Zhi Zhu[1,e]

[1] Department of Biotechnology, Faculty of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, P. R. China

[2] Institute of Life Sciences, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, P. R. China

Abstract

The basic helix-loop-helix (bHLH) transcription factors play important roles in a wide range of developmental processes in higher organisms. bHLH family members have been identified in a dozen of organisms including fruit fly, mouse and human. In this study, we identified 51 bHLH sequences *in silico* in the honey bee, *Apis mellifera* L. (Hymenoptera: Apidae), genome. Phylogenetic analyses revealed that they belong to 38 bHLH families with 21, 11, 9, 1, 8 and 1 members in high-order groups A, B, C, D, E and F, respectively. Using phylogenetic analyses, all of the 51 bHLH sequences were assigned to their corresponding families. Genes that encode ASCb, NeuroD, Oligo, Delilah, MyoRb, Figα and Mad were not found in the honey bee genome. The present study provides useful background information for future studies using the honey bee as a model system for insect development.

Keywords: phylogeny, blast search
Abbreviations: ASC - Achaete-Scute Complex, Ngn - Neurogenin, H/E(spl) - Hairy/E(spl), Hs - *Homo sapiens*, Mm - *Mus musculus*, Bf - *Branchiostoma floridae* (the Florida lancelet)
Correspondence: [a]ywang@ujs.edu.cn, [b]kpchen@ujs.edu.cn, [c]yaoqin@ujs.edu.cn, [d]wenbingwang@ujs.edu.cn, [e]zouz@zjcan.com
Received: 14 August 2007 | **Accepted:** 29 September 2007 | **Published:** 20 May 2008
Copyright: This is an open access paper. We use the Creative Commons Attribution 3.0 license that permits unrestricted use, provided that the paper is properly attributed.
ISSN: 1536-2442 | Volume 8, Number 40

Cite this paper as:
Wang Y, Chen K, Yao Q, Wang W, Zhu Z. 2008. The basic helix-loop-helix transcription factor family in the honey bee, *Apis mellifera*. 12pp. *Journal of Insect Science* 8:40, available online: insectscience.org/8.40

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Introduction

The basic helix-loop-helix (bHLH) family of transcription factors plays important roles in a wide range of developmental processes including neurogenesis, myogenesis, hematopoiesis, sex determination and gut development (Massari and Murre 2000). Since the first characterization of the mouse bHLH transcription factors E12 and E47 (Murre et al. 1989), hundreds of bHLH proteins have been identified so far. In 1999, Atchley et al developed a predictive motif for the bHLH domains based on amino acid frequencies at all positions of 242 bHLH proteins (Atchley et al. 1999). 19 conserved sites were found within the bHLH domain. Atchley et al. (1999) showed that a sequence with less than 8 mismatches to the predictive motif was very possibly a bHLH protein. Later, researchers found that a sequence even with 9 mismatches could also be a potential bHLH protein (Toledo-Ortiz et al. 2003). In recent years, more bHLH genes have been identified in organisms whose genome sequences were available. These include 8 bHLH genes in yeast, 16 in *Amphimedon queenslandica*, 33 in *Hydra magnipapillata*, 39 in *Caenorhabditis elegans*, 39 in *Gallus gallus*, 39 in *Brachydanio rerio*, 46 in *Ciona intestinalis*, 47 in *Xenopus laevis*, 50 in *Strongylocentrotus purpuratus*, 57 in *Daphia pulex*, 59 in *Drosophila melanogaster*, 63 in *Lottia gigantea*, 64 in *Capitella* sp 1, 68 in *Nematodtella vectensis*, 78 in *Branchiostoma floridae*, 87 in pufferfish, 102 in mouse, 118 in human, 147 in *Arabidopsis* and 167 in rice (Ledent et al. 2002; Li et al. 2006; Satou et al. 2003; Simionato et al. 2007; Toledo-Ortiz et al. 2003). Based on phylogenetic analyses of over 400 bHLH proteins, Ledent et al defined 45 orthologous families and 6 higher-order groups for all the identified bHLH genes, where the 44 families were named according to their name of the first discovered or best-known member of the family, and the higher-order groups were named A to F based on the different properties of these groups (Atchley and Fitch 1997; Ledent et al. 2002; Ledent and Vervoort 2001; Simionato et al. 2007). Groups A and B bHLH proteins bind to core DNA sequences typical of E boxes (CANNTG) which is CACCTG or CAGCTG for group A and CACGTG or CATGTTG for group B. Group C comprises the family of bHLH proteins known as bHLH-PAS because a PAS domain follows the bHLH motif. The core sequences to which they bind are ACGTG or GCGTG, while recent studies have demonstrated that the *Drosophila* Dysfusion/Tango bHLH-PAS heterodimer has a binding preference as TCGTG > GCGTG > ACGTG > CCGTG (Jiang and Crews 2007). Group D proteins lack a basic domain. They are not able to bind DNA. They function as antagonists of group A bHLH proteins. Group E proteins are mainly related to the Drosophila Hairy and E(spl) bHLH proteins. These proteins bind preferentially to sequences typical of N boxes (CACGCG or CACGAG). They also contain two additional domains named 'Orange' and WRPW peptide in the carboxyl-terminal part. Group F proteins have the COE domain which is characterized by the presence of an additional domain involved both in dimerization and in DNA binding (Ledent and Vervoort 2001).

The honey bee, *Apis mellifera* L. (Hymenoptera: Apidae), is a key model for social behavior. Many studies have been conducted to elucidate the developmental processes that result in its particular social organization. However, not many bHLH transcription factors have been characterized. So far, seven honey bee bHLH sequences have been reported. They are AmCYC and AmCLK for which cDNA sequences were cloned (Rubin et al. 2006), two Achaete-Scute genes and three Enhancer of split genes that were identified in the honey bee genome (Schlatter and Maier 2005). The latest version of honey bee genome sequence has been available in the GenBank since October 2007. In this study, we used both the representative sequences of the 45 bHLH families (Ledent and Vervoort 2001) and the known 59 *Drosophila melanogaster* bHLH (DmbHLH) sequences (Ledent et al. 2002; Simionato et al. 2007) to conduct tblastn searches against database of the *Apis mellifera* genome sequences. After examining the amino acid residues at the 19 conserved sites, we found that 51 *Apis mellifera* bHLH (AmbHLH) sequences satisfied the screening criterion. Phylogenetic analyses with the 45 representative bHLH domains and with the 59 DmbHLH sequences defined the families to which the 51 AmbHLH sequences belong.

Materials and Methods

tblastn searches

The sets of 45 representative bHLH domains and 59 DmbHLH motifs were from the additional files of (Ledent and Vervoort 2001) and (Simionato et al. 2007), respectively. Each sequence of both sets was used to perform tblastn searches against the database of *Apis mellifera* genome draft sequences (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=7460).
Tblastn searches compare a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames of both strands. Stringency was set as $E < 10$ in order to obtain all bHLH related sequences for later examination.

Manual improvement to the obtained sequences

The obtained subject sequences from the tblastn searches were examined manually to keep only one sequence for those that have the same scaffold number, reading frame and coding regions. Manual improvement was also done to the sequences lacking a few amino acids on their two ends. This was realized by retrieving the whole subject sequence from GenBank and translating it with EditSeq program (version 5.01) of the DNAStar package to obtain the absent amino acid residues. To those subject

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

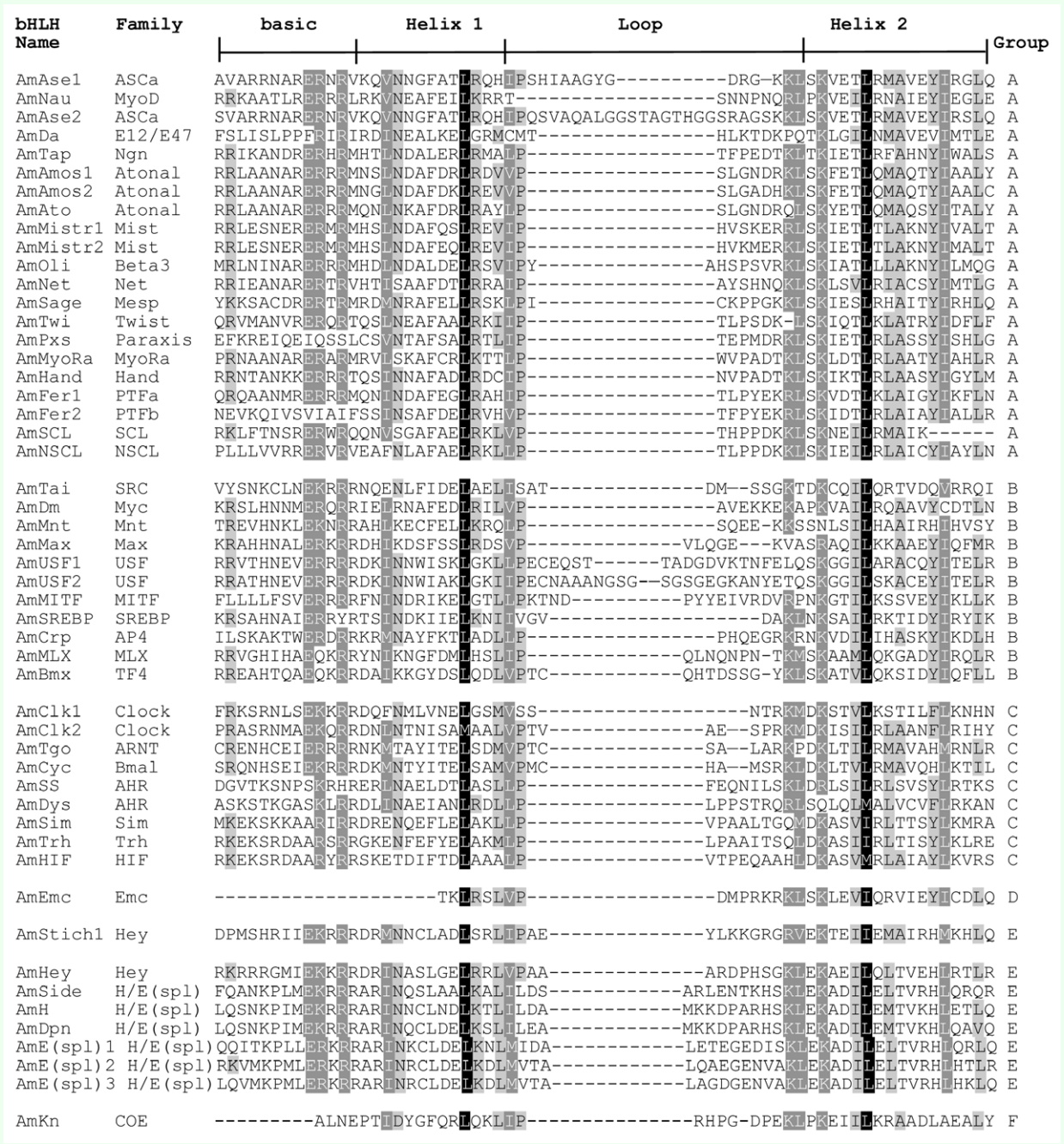

Figure 1. Alignment of 51 AmbHLH members. Designation of basic, helix 1, loop and helix 2 follows Ferre-D'Amare et al. (Ferre-D'Amare 1993). The family names and high-order groups have been organized according to Table 1 in Ledent et al (Ledent et al. 2002). AmbHLHs were named in accordance with fruit fly nomenclature.

sequences that had coding regions separated by dozens to thousands of nucleotides, the SpliceView application (http://www.itb.cnr.it/sun/webgene/) was used to analyze if the sequences had introns.

Sequence alignment

All sequences that had undergone the above improvement were aligned using ClustalW online (http://www.ebi.ac.uk/clustalw/) with default settings. The aligned sequences were transferred into a Microsoft Excel worksheet for examining the amino acid residues at the 19 conserved sites at specific sites. Sequences with less than 9 variations were regarded as potential AmbHLHs and were aligned again using ClustalW. The aligned AmbHLHs were shaded in GeneDoc Multiple Sequence Alignment Editor and Shading Utility (Version 2.6.02) (Nicholas et al. 1997) and copied to a Word RTF file for further annotation.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Phylogenetic analyses

Phylogenetic analyses were conducted using PAUP 4.0 Beta 10 (Swofford 1998) based on a stepmatrix constructed from Dayhoff PAM 250 distance matrix by R. K. Kuzoff (http://paup.csit.fsu.edu/nfiles.html). The obtained AmbHLH sequences were used to construct neighbor-joining distance trees with the 45 representative bHLH domains and with the 59 DmbHLH motifs, respectively. Sequences were first aligned in ClustalW and then copied into PAUP window to prepare nexus files. Neighbor joining trees were bootstrapped with 1,000 replicates to provide information about their statistical reliability. Maximum parsimony trees were constructed using PAUP 4.0 Beta 10 by executing command "bootstrap nreps = 100 search = heuristic/addseq = random". Other parameters were set to default values. Maximum likelihood trees were constructed using TreePuzzle 5.2 (Schmidt et al. 2002). The number of puzzling steps was set to 25,000. Model of substitution was set to Jones-Taylor-Thornton (Jones et al. 1992). Other parameters were default values. The trees were displayed using the TreeView program (version 1.6.6) (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html), saved as Phylip format, edited using MEGA3.1 (Kumar et al. 2004), copied to clipboard, and then annotated in Microsoft PowerPoint.

EST searches

In order to find existing expressed sequence tags (ESTs) matching the obtained AmbHLH sequences, tblastn searches were performed against honey bee EST database on NCBI tblastn website using each AmbHLH as

Table 1. Assignment of AmbHLH members into corresponding families.

bHLH name	Family	Bootstrap support from various phylogenetic analyses					
		NJ[a]	MP[a]	ML[a]	NJ[b]	MP[b]	ML[b]
AmAse1	ASCa	100	72	100	66	65	73
AmAse2	ASCa	100	91	83	36	n/m*	85
AmAmos1	Atonal	98	84	94	83	71	50
AmAmos2	Atonal	99	93	94	71	44	57
AmAto	Atonal	94	76	94	97	96	78
AmMistr1	Mist	98	88	92	100	100	71
AmMistr2	Mist	99	79	85	100	100	67
AmFer2	PTFb	73-PTFa	56-PTFa	65-PTFa	89-PTFb	98-PTFb	71-PTFb
AmDm	Myc	46	47	55	97	99	77
AmUSF1	USF	99	94	85	97	96	92
AmUSF2	USF	100	98	68	89	93	85
AmAP4	AP4	29-SRC	35-Figα	n/m	55-AP4	n/m	53-AP4
AmClk2	Clock	79	67	68	80	59	62
AmDys	AHR	34	40	88	100	100	82
AmSim	Sim	n/m	50	n/m	72	92	87
AmStich1	Hey	47	58	n/m	100	100	100
AmSide	H/E(spl)	42	50	100	99	99	97
AmH	H/E(spl)	98	63	66	93	79	64
AmDpn	H/E(spl)	92	n/m	n/m	68	n/m*	61
AmE(spl)1	H/E(spl)	n/m	73	n/m	39	n/m	66
AmE(spl)2	H/E(spl)	n/m	35	n/m	n/m*	n/m*	n/m*
AmE(spl)3	H/E(spl)	n/m	41	n/m	n/m*	34	n/m*

n/m = not monophyletic.
n/m* means that a AmbHLH does not form a monophyletic group with one specific bHLH protein but is in a monophyletic group with other bHLH proteins of the same family.
NJ[a], MP[a] and ML[a]: bootstrap support from constructing NJ, MP and ML trees of in-group phylogenetic analyses with the 45 domains (the constructed trees not shown).
NJ[b], MP[b] and ML[b]: bootstrap support from constructing NJ, MP and ML trees of in-group phylogenetic analyses with the 59 fruit fly bHLH proteins (the constructed trees not shown).
Each figure in the table is a bootstrap value supporting assignment of the AmbHLH sequence either into the family listed in the column of "Family" or into another family following the figure after a hyphen.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

the query sequence. The stringency was set as $E < 0.0001$. A 90% or higher identity was considered to be an EST corresponding to the specific AmbHLH sequence.

Results and Discussion

Identification of AmbHLH sequences in the *A. mellifera* genome database

Tblastn searches with the 45 bHLH domains and 59 DmbHLH motifs followed by manual improvement and examination led to the identification of 50 and 1 potential AmbHLH sequences, respectively. The alignment of all 51 AmbHLH domains is shown in Figure 1. Since there had been sufficient bootstrap support in the

Table 2. Coding regions of 51 AmbHLH domains.

No.	bHLH name	Fruit fly ortholog* (Gene symbol | ID)	Family	Scaffold number	Frame	Coding region(s)	Remark
1	AmAse1	*ase* | CG3258	ASCa	NW_001252982.1	3	178410–178586	
2	AmAse2	? - ortholog of AmAse1	ASCa	NW_001252982.1	3	134250–134459	
3	AmNau	*nau* | CG10250	MyoD	NW_001253207.1	2	81488–81532	Intron (114 bp)
					2	81647–81757	
4	AmDa	*da* | CG5102	E12/E47	NW_001253102.1	–2	472361–472200	
5	AmTap	*tap* | CG7659	Ngn	NW_001253047.1	–2	1837563–1837405	
6	AmAmos1	*amos* | CG10393	Atonal	NW_001253389.1	–2	37521–37363	
7	AmAmos2	? - ortholog of AmAmos1	Atonal	NW_001253389.1	–2	61944–61786	
8	AmAto	*ato* | CG7508	Atonal	NW_001253032.1	–3	69265–69107	
9	AmMistr1	*Mistr* | CG8667	Mist	NW_001259485.1	2	44171–44233	Intron (72 bp)
					2	44306–44401	
10	AmMistr2	? - ortholog of AmMistr1	Mist	NW_001253188.1	–1	187093–187031	Intron (2441 bp)
					–3	184589–184494	
11	AmOli	*Oli* | CG5545	Beta3	NW_001253016.1	–1	1095766–1095602	
12	AmNet	*net* | CG11450	Net	NW_001262732.1	2	40334–40492	
13	AmSage	*sage* | CG12952	Mesp	NW_001253168.1	3	462963–463124	
14	AmTwi	*twi* | CG2956	Twist	NW_001253504.1	2	21713–21868	
15	AmPxs	*Pxs* | CG12648	Paraxis	NW_001260530.1	–3	9268–9110	
16	AmMyoRa	*MyoRa* | CG5005	MyoRa	NW_001253177.1	–1	869042–868884	
17	AmHand	*Hand* | CG18144	Hand	NW_001253428.1	2	39326–39484	
18	AmFer1	*Fer1* | CG10066	PTFa	NW_001253524.1 NW_001253522.1	1	7003–7161	2 copies
					3	372942–373100	
19	AmFer2	*Fer2* | CG5952	PTFb	NW_001253506.1	3	114579–114737	
20	AmSCL	*SCL* | CG2655	SCL	NW_001253369.1	–2	673350–673210	
21	AmNSCL	*NSCL* | CG3052	NSCL	NW_001255426.1 NW_001253565.1	–3	770–612	2 copies
					1	17407–17565	
22	AmTai	*tai* | CG13109	SRC	NW_001253165.1	–1	328401–328240	
23	AmDm	*dm* | CG10798	Myc	NW_001253009.1	–2	81742–81584	
24	AmMnt	*Mnt* | CG2856	Mnt	NW_001253181.1	–3	21235–21080	
25	AmMax	*max* | CG9648	Max	NW_001253535.1	1	794938–795096	
26	AmUSF1	*USF* | CG17592	USF	NW_001253047.1	–2	1011771–1011631	Intron (103 bp)
					–3	1011527–1011477	

*Gene symbols and ID numbers are from Table 3 of Ledent et al.(2002). To those whose orthologs were not found in fruit fly, a question mark and a description of its orthologous AmbHLH is in place.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

following phylogenetic analyses, the AmbHLHs were named according to their orthologs in *D. melanogaster*. Data supporting this nomenclature are provided in Figures 2, 3 and Table 1. The *D. melanogaster* orthologs are listed in Table 2 for reference. All of the phylogenetic analyses revealed that the 51 AmbHLHs belong to 38 families with 21, 11, 9, 1, 8 and 1 members in groups A, B, C, D, E and F, respectively (Figure 1).

Identification of orthologous families

Identification of orthologous genes has been uncertain since there is no absolute criterion that can be used to decide if two genes are orthologous (Ledent and Vervoort 2001). Based on the criterion used by Ledent et al (Ledent et al. 2002; Ledent and Vervoort 2001), a more stringent criterion was used: a single AmbHLH must form a monophyletic group with another bHLH of a known family in phylogenetic trees constructed with different methods, and all the bootstrap values must exceed 50.

Table 2. [continued from previous page]

No.	bHLH name	Fruit fly ortholog* (Gene symbol | ID)	Family	Scaffold number	Frame	Coding region(s)	Remark
27	AmUSF2	? - ortholog of AmUSF1	USF	NW_001253565.1	–2	69204–69052	Intron (80 bp)
					–1	68971–68921	
28	AmMITF	*MITF* | CG17469	MITF	NW_001253276.1	3	486417–486515	Intron (765 bp)
					3	487281–487361	
29	AmSREBP	*SREBP* | CG8522	SREBP	NW_001254263.1	–2	3763–3611	
30	AmCrp	*crp* | CG7664	AP4	NW_001253018.1	–3	593295–593137	
31	AmMlx	*Mlx* | CG18362	MLX	NW_001253310.1	–1	2346–2182	
32	AmBmx	*bmx* | CG3350	TF4	NW_001260288.1	1	19873–20043	
33	AmClk1	*clk* | CG7391	Clock	NW_001253243.1	1	192466–192618	
34	AmClk2	? - ortholog of AmClk1	Clock	NW_001253823.1 NW_001253487.1	–3	2138–1977	2 copies
					–3	78214–78053	
35	AmTgo	*tgo* | CG11987	ARNT	NW_001253455.1	–1	85074–84913	
36	AmCyc	*cyc* | CG8727	Bmal	NW_001252991.1	2	150056–150217	
37	AmSS	*ss* | CG6993	AHR	NW_001253513.1	–1	948637–948476	
38	AmDys	*dys* | CG12561	AHR	NW_001253007.1	–3	9070–8909	
39	AmSim	*sim* | CG7771	Sim	NW_001253228.1	–2	1077732–1077571	
40	AmTrh	*trh* | CG6883	Trh	NW_001253054.1	–2	198756–198595	
41	AmHIF	*sima* | CG7951	HIF	NW_001253383.1	1	276958–277119	
42	AmEmc	*emc* | CG1007	Emc	NW_001253177.1	–3	826635–826537	
43	AmStich1	*Stich 1* | CG17100	Hey	NW_001253057.1	–1	1286192–1286025	
44	AmHey	*Hey* | CG11194	Hey	NW_001252977.1	1	120133–120300	
45	AmSide	*side* | CG10446	H/E(spl)	NW_001253507.1	–3	39984–39811	
46	AmH	*h* | CG6494	H/E(spl)	NW_001253070.1	3	722031–722132	Intron (253 bp)
					1	722386–722457	
47	AmDpn	*dpn* | CG8704	H/E(spl)	NW_001253070.1	–2	387108–387007	Intron (3642 bp)
					–2	383364–383293	
48	AmE(spl)1	*E(spl) mC(d)* | CG8328	H/E(spl)	NW_001253088.1	–3	137141–137046	Intron (4460 bp)
					2	132585–132508	
49	AmE(spl)2	? - ortholog of AmE(spl)3	H/E(spl)	NW_001253088.1	–1	177463–177290	
50	AmE(spl)3	? - ortholog of AmE(spl)2	H/E(spl)	NW_001253088.1	3	333567–333740	
51	AmKn	*kn* | CG10197	COE	NW_001253246.1	3	746790–746924	

*Gene symbols and ID numbers are from Table 3 of Ledent et al.(2002). To those whose orthologs were not found in fruit fly, a question mark and a description of its orthologous AmbHLH is in place.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

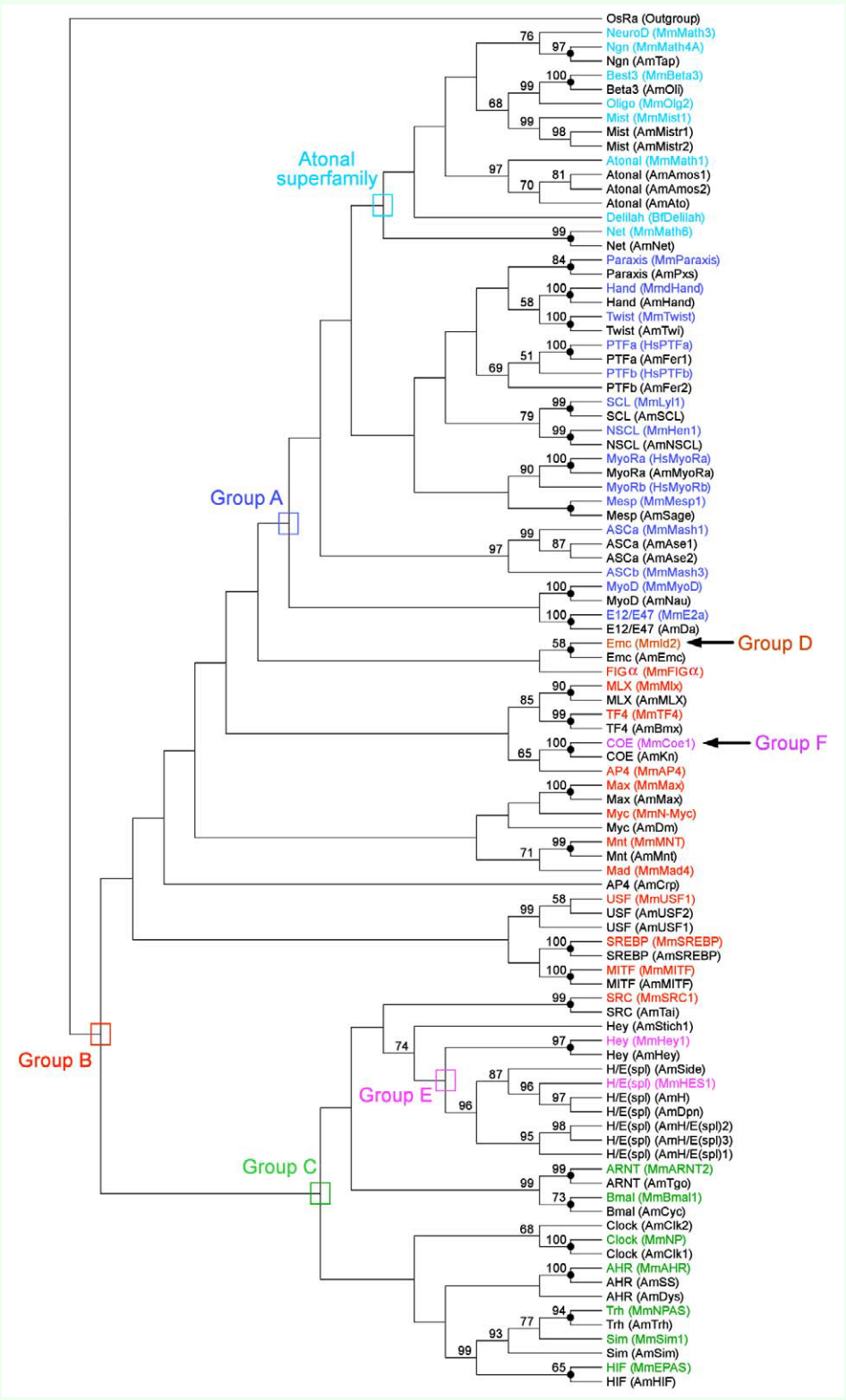

Figure 2. Phylogenetic relationship of 51 AmbHLH members with 45 bHLH domains. A neighbor-joining (NJ) tree is shown. For simplicity, branch lengths of the tree are not proportional to distances between sequences, and bootstrap values less than 50 are not shown. The higher-order group labels are in accordance with (Ledent et al. 2002).

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

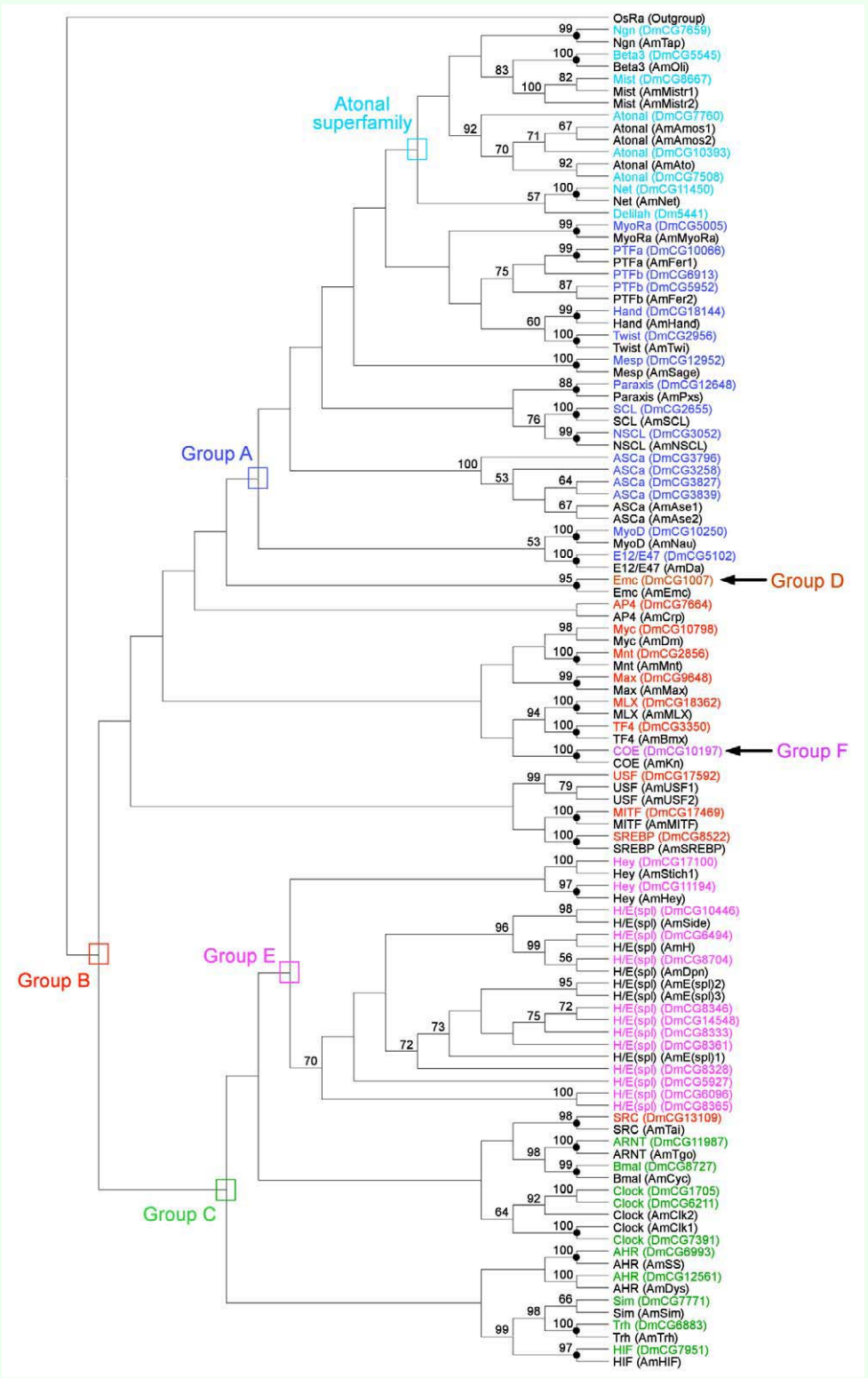

Figure 3. Phylogenetic relationship of 51 AmbHLH members with 59 *Drosophila* bHLHs. A neighbor-joining (NJ) tree is shown. For simplicity, branch lengths of the tree are not proportional to distances between sequences, and bootstrap values less than 50 are not shown. The higher-order group labels are in accordance with (Ledent et al. 2002).

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

The obtained 51 AmbHLH sequences had been used to construct neighbor joining trees with 45 bHLH domains (Figure 2) and with 59 DmbHLH motifs (Figure 3), respectively. In both trees, OsRa (the rice bHLH sequence of R family) sequence was used as outgroup. In both Figures 2 and 3, it can be seen that 29 out of 51 AmbHLH sequences had formed monophylogenetic groups with 29 other bHLH sequences, respectively. Their bootstrap values ranged from 56 to 100. These 29 AmbHLHs are AmTap, AmOli, AmNet, AmPxs, AmHand, AmTwi, AmMyoRa, AmSCL, AmNSCL, AmSage, AmFer1 and AmNau of group A, AmSREBP, AmMLX, AmBmx, AmDm, AmMax, AmMITF, AmMnt and AmTai of group B, AmClk1, AmTgo, AmCyc, AmSS, AmTrh and AmHIF of group C, AmEmc of group D, AmHey of group E, and AmKn of group F, all of which nodes are indicated with black dots in Figures 2 and 3).

In order to define families for the rest 22 AmbHLHs, each of them was used to construct neighbor joining, maximum parsimony and maximum likelihood phylogenetic trees within the members of a particular higher-order group. The results are summarized in Table 1 (the constructed trees are not shown). Table 1 shows that, by constructing phylogenetic trees using a single AmbHLH sequence with other bHLH proteins belonging to the same higher-order group (termed "in-group" analysis), all of the 22 AmbHLH sequences can be assigned to specific bHLH families. It should be noted that not all of the bootstrap values are over 50 for each assignment. However, with the bootstrap values from the construction of six in-group phylogenetic trees, there was sufficient support to make the assignments shown in Table 1.

The above phylogenetic analyses also enabled us to identify *D. melanogaster* orthologs for 44 AmbHLHs (Table 2). The remaining 7 AmbHLHs, namely AmAse2, AmAmos2, AmMistr2, AmUSF2, AmClk2, AmE(spl)2 and AmE(spl)3, did not form monophyletic groups with any *D. melanogaster* bHLHs. Instead, they formed monophyletic groups with other AmbHLHs as indicated with a

Table 3. The insect bHLH members

Group	Family name	*A.m.*	*B.m.*	*T.c.*	*D.m.*
A	ASCa	2	4	3	4
	ASCb	0	0	0	0
	MyoD	1	1	1	1
	E12/E47	1	1	1	1
	Ngn	1	1	1	1
	NeuroD	0	0	1	0
	Atonal	3	1	3	3
	Mist	2	1	1	1
	Beta3	1	1	1	1
	Oligo	0	0	0	0
	Net	1	1	1	1
	Delilah	0	1	2	1
	Mesp	1	1	0	1
	Twist	1	1	1	1
	Paraxis	1	1	1	1
	MyoRa	1	1	1	1
	MyoRb	0	0	0	0
	Hand	1	1	1	1
	PTFa	1	1	1	1
	PTFb	1	1	2	2
	SCL	1	1	1	1
	NSCL	1	1	1	1

Group	Family name	*A.m.*	*B.m.*	*T.c.*	*D.m.*
B	SRC	1	1	1	1
	Figα	0	0	0	0
	Myc	1	1	1	1
	Mad	0	0	1	0
	Mnt	1	1	1	1
	Max	1	1	1	1
	USF	2	1	1	1
	MITF	1	1	1	1
	SREBP	1	1	1	1
	AP4	1	1	1	1
	MLX	1	1	0	1
	TF4	1	1	1	1
C	Clock	2	3	2	3
	ARNT	1	1	1	1
	Bmal	1	2	1	1
	AHR	2	3	1	2
	Sim	1	1	0	1
	Trh	1	1	1	1
	HIF	1	1	1	1
D	Emc	1	1	1	1
E	Hey	2	2	1(2?)	1(2?)
	H/E(spl)	6	5	5(6?)	11(12?)
F	COE	1	1	1	1
	TOTAL	**51**	**52**	**50**	**59**

Data of A.m. (*Apis mellifera*) from this study. Those of B.m. (*Bombyx mori*) were from our unpublished data. Those of T.c. (*Tribolium castaneum*) and D.m. (*D. melanogaster)* were from Simionato et al. 2007.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

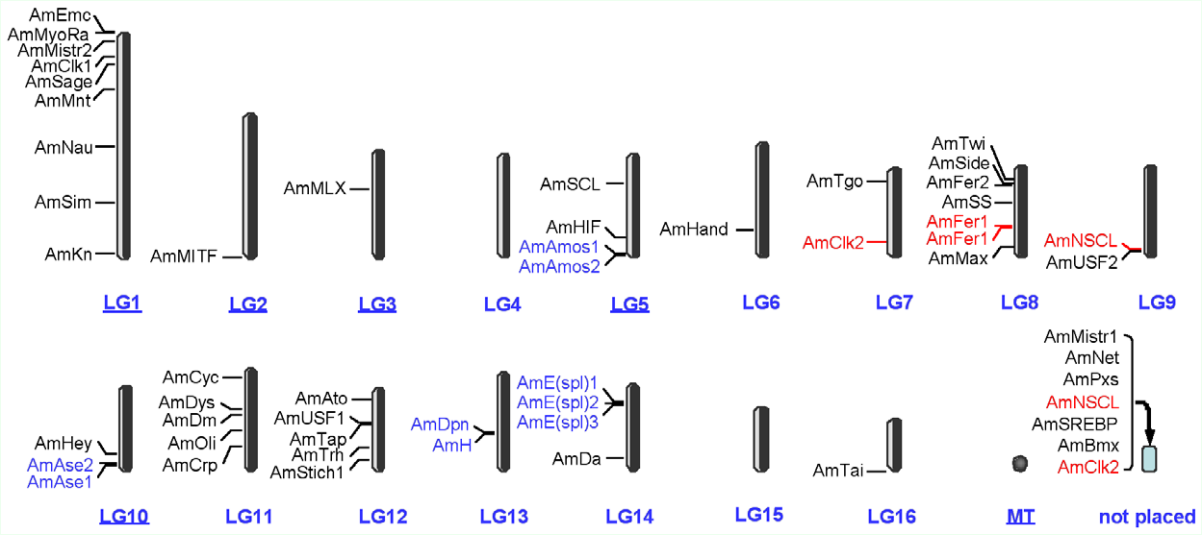

Figure 4. Localization of the AmbHLH coding regions. The AmbHLH names in red are those having two copies in the genome. The AmbHLH names in blue are those of the same family cluster together.

question mark followed by the description of its orthologous AmbHLH. This result strongly suggests that these 7 AmbHLHs arose after *A. mellifera* diverged from the other insects.

Coding regions and the localization of AmbHLH motifs

The coding regions for all the identified AmbHLH motifs are listed in Table 2. The data indicate that 9 AmbHLHs have introns in their bHLH motifs, among which AmNau, AmMistr1 and AmMistr2 have introns in helix 1 region, AmMITF, AmH, AmDpn and AmE(spl)1 have introns in the loop region, and AmUSF1 and AmUSF2 have introns in helix 2 region. The length of the introns ranged from 72 to 4460 base pairs. It was also found that three AmbHLHs had 2 copies in the genome. They are AmFer1, AmNSCL and AmClk2.

Searches with the scaffold numbers listed in Table 3 in the honey bee map view (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=9555) located the positions of sequences coding for all of the AmbHLHs (Figure 4),which shows that the distribution of AmbHLH genes is fairly uneven. Chromosomes 1 and 8 have 9 and 7 AmbHLH genes, respectively. Chromosomes 11 and 12 have 5 AmbHLH genes each. Those on chromosomes 2, 3, 5, 6, 7, 9, 10, 13, 14 and 16 vary between 1 and 4. Chromosomes 4 and 15 and mitochondrial DNA do not code for any bHLH proteins. It should be noted that 7 AmbHLHs have not been placed in the map. It can also be seen in Figure 4 that two or three members of the same bHLH family often cluster together. For example, AmAmos1 and AmAmos2, AmAse1 and AmAse2, AmH and AmDpn, and AmE(spl)1, 2 and 3 all locate on the same scaffold, respectively (indicated in blue). This suggests a possible origination for the other member by gene duplication.

The AmbHLH repertoire

The above searches and analyses allowed definition of families for the 51 obtained AmbHLHs. This figure is comparable with 52, 50 and 59 bHLH members in the domestic silkworm, red flour beetle and fruit fly, respectively (Table 3). It can be seen that all of these four insects lack genes of families ASCb, NeuroD, Oligo, MyoRb, Figα and Mad, and many of the families have the same number of genes. Major differences were seen in the number of genes of the H/E(spl) family. *D. melanogaster* have 11 to 12 H/E(spl) genes while other insects have 5 to 6. *A. mellifera* has fewer genes in families ASCa, PTFb and Clock than *D. melanogaster*. It is noteworthy that *A. mellifera* has one more gene in families Mist and USF than other insects. Another feature to be noted is that only 2 members of the family ASCa were found in *A. mellifera* and none were found of the family Delilah. Whether *A. mellifera* does have fewer members of these families, or if it was due to incompleteness of the genome sequences remains for further exploitation.

One gene was found to code for Bmal, 2 for ASCa and 3 for E(spl). This is consistent with previous reports (Rubin et al. 2006; Schlatter and Maier 2005). In our survey, 2 genes coding for the Clock family of transcription factors were identified. It is not known if the AmCLK gene cloned by Rubin et al. (2006) is one of them, since no sequence information was available for AmCLK.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Expression of AmbHLH genes

A tblastn search with the identified AmbHLH sequences against *A. mellifera* EST databases in GenBank indicated that 10 of them met the searching criterion (Table 4). They are AmOli, AmHand, AmSCL, AmMax, AmUSF2, AmCrp, AmClk1, AmTgo, AmE(spl)2 and AmE(spl)3. Table 4 indicates that the expression of these 10 AmbHLH genes was mainly seen in head tissue. The reason why only 10 AmbHLHs were found to have corresponding EST sequences was probably due to a relatively small deposit of the honey bee EST database which had 78,085 EST sequences as compared to 541,595 for *D. melanogaster* and 4,850,243 for the mouse.

Conclusions

By using the 45 representative bHLH domains and 59 identified DmbHLHs as query sequences, we identified 51 bHLHs from *A. mellifera* genome sequences. It was found necessary to use *D. melanogaster* bHLH sequences as

Table 4. EST sequences of 10 identified AmbHLHs.

bHLH name	Identity	EST accession number	Tissue type
AmOli	100%	BI503635	Brain
AmHand	100%	CK629732	Whole body
AmSCL	100%	BI517069	Brain
	100%	BI514458	Brain
	100%	BI514430	Brain
AmMax	100%	DB731368	Head
	100%	DB745739	Head
	100%	BI516438	Brain
	100%	BI513903	Brain
	100%	BI513837	Brain
AmUSF2	100%	BQ103821	Brain
AmCrp	98%	BI504679	Brain
AmClk1	100%	DB738928	Head
	100%	DB743239	Head
	100%	DB741916	Head
	100%	DB734425	Head
	93%	DB735483	Head
	100%	DB730331	Head
	96%	DB752332	Head
	93%	DB750344	Head
AmTgo	94%	DB754019	Head
AmE(spl)2	100%	DB736462	Head
	100%	BI510545	Brain
	100%	DB733661	Head
	91%	BI516208	Brain
	100%	DB747076	Head
	100%	DB753629	Head
	98%	DB742806	Head
AmE(spl)3	98%	BI516208	Brain
	89%	DB736462	Head
	89%	BI510545	Brain
	91%	DB733661	Head
	92%	DB747076	Head
	92%	DB753629	Head
	90%	DB742806	Head

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

query sequences. This helped us to identify 1 additional bHLHs in *Apis mellifera*. It was also advantageous to use *D. melanogaster* bHLHs of known families to help determine the orthologous genes for *A. mellifera*. Since the 45 representative bHLH domains were mainly from the mouse (Ledent et al. 2002; Simionato et al. 2007), it was reasonable to assign relationships depending on results from *D. melanogaster* when the results from phylogenetic analyses with both representative bHLH domains and DmbHLH motifs did not accord with each other. For instance, in-group analyses with 22 representative bHLH domains suggested orthologous families of PTFa for AmPTFb with bootstrap support of 56 to 73. But in-group analyses with 24 *D. melanogaster* bHLHs suggested PTFb with much higher bootstrap support (71 to 98) (Table 1). Therefore, PTFb was considered to be the orthologous family for that sequence.

Acknowledgments

We are grateful to Professor Bin Chen and two anonymous reviewers for constructive comments on the manuscript. This work was supported by grants from National Natural Science Foundation of China (No. 30370773) and National Basic Research Program of China (No. 2005CB121000).

References

Atchley WR, Fitch WM. 1997. A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Science U S A* 94: 5172-5176.

Atchley WR, Terhalle W, Dress A. 1999. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. *Journal of Molecular Evolution* 48: 501-516.

Jiang L, Crews ST. 2007. Transcriptional specificity of Drosophila dysfusion and the control of tracheal fusion cell gene expression. *Journal of Biological Chemistry* 282: 28659-28668.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8: 275-282.

Kumar CG, LeDuc R, Gong G, Roinishivili L, Lewin HA, Liu L. 2004. ESTIMA, a tool for EST management in a multi-project environment. *BMC Bioinformatics* 5: 176

Ledent V, Paquet O, Vervoort M. 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. *Genome Biology* 3: R30

Ledent V, Vervoort M. 2001. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Research* 11: 754-770.

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J, Zhang D. 2006. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiology* 141: 1167-1184.

Massari ME, Murre C. 2000. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology* 20: 429-440.

Murre C, McCaw PS, Baltimore D. 1989. A new DNA binding and dimerizing motif in Immunoglobulin enhancer binding, Daugtherless, MyoD, and Myc proteins. *Cell* 56: 777-783.

Nicholas KB, Nicholas-Jr HB, Deerfield-II DW. 1997. GeneDoc: Analysis and Visualization of Genetic Variation. *Embnet News* 4: 14

Rubin EB, Shemesh Y, Cohen M, Elgavish S, Robertson HM, Bloch G. 2006. Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (Apis mellifera) and shed new light on the molecular evolution of the circadian clock. *Genome Research* 16: 1352-1365.

Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N. 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis. I. Genes for bHLH transcription factors. *Development Genes and Evolution* 213 : 5–60213-221.

Schlatter R, Maier D. 2005. The Enhancer of split and Achaete-Scute complexes of Drosophilids derived from simple ur-complexes preserved in mosquito and honeybee. *BMC Evolutionary Biology* 5: 67-86.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18: 502-504.

Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evolutionary Biology* 7: 33

Swofford DL. 1998. PAUP*. *Phylogenetic Analysis Using Parsimony, Version 4*. Sinauer Associates.

Toledo-Ortiz G, Huq E, Quail PH. 2003. The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* 15: 1749-1770.

Downloaded From: https://bioone.org/journals/Journal-of-Insect-Science on 24 Apr 2024
Terms of Use: https://bioone.org/terms-of-use