Identification and Phylogenetic Analysis of Basic Helix-Loop-Helix Genes in the Diamondback Moth

Zhen Zeng,1 Yong Wang,2,3 Nana Adwoa Nkuma Johnson,2 Guang-Dong Wang,1 Qin Yao,1 and Ke-Ping Chen1

1Institute of Life Sciences, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, China, 2School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, China, and 3Corresponding author, e-mail: ywang@ujs.edu.cn

Subject Editor: Jianxiu Yao

Received 10 January 2018; Editorial decision 19 May 2018

Abstract

Basic helix-loop-helix (bHLH) transcription factors play essential roles in regulating eukaryotic developmental and physiological processes such as neuron generation, myocyte formation, intestinal tissue development, and response to environmental stress. In this study, the diamondback moth, Plutella xylostella (L.) (Lepidoptera: Plutellidae), genome was found to encode 52 bHLH genes. All 52 P. xylostella bHLH (PxbHLH) genes were classified into correspondent bHLH families according to their orthology with bHLHs from fruit fly and other insect species. Among these 52 PxbHLH genes, 19 have been annotated consistently with our classification in GenBank database. The remaining 33 PxbHLH genes are either annotated as general bHLH genes or as hypothetical genes. Therefore, our data provide useful information for updating annotations to PxbHLH genes. P. xylostella has four stem cell leukemia (SCL) genes (one of them has three copies), two Dys genes, two copies of MyoR, Mitf, and Sima genes, and three copies of Sage genes. Further studies may be conducted to elucidate functions of these specific bHLH genes in regulating P. xylostella growth and development.

Key words: basic helix-loop-helix, Blastp search, orthology, phylogenetic analysis, Plutella xylostella

A basic helix-loop-helix (bHLH) motif is approximately 60 amino acids in length. It is composed of a basic (alkaline) region capable of binding DNA and a helix-loop-helix region capable of forming dimer with another HLH motif. Based on statistical analysis to amino acid composition in a large number of bHLH motifs, Atchley et al. (1999) discovered 19 highly conserved sites in bHLH motif at which specific amino acids are present. For example, either arginine or lysine is present at the first, second, and 10th site of the basic region. Therefore, a criterion was established to qualify a candidate bHLH protein sequence through examining whether specific amino acids are present at the 19 conserved sites. According to this criterion, a qualified bHLH protein sequence should have no less than 11 specific amino acids present at the 19 conserved sites.

bHLH proteins constitute a large superfamily of transcription factors. Various bHLH proteins play significant regulatory roles in a wide range of eukaryotic developmental and physiological processes such as neuron generation, myocyte formation, intestinal tissue development, and response to environmental stress (Massari and Murre 2000). Various eukaryotic species have a greatly varied number of bHLH genes. For example, yeast, nematode, fruit fly, mouse, and zebrafish genomes were found to encode 8, 45, 59, 114, and 139 bHLH genes, respectively, while genomes of thale cress and rice were found to encode 147 and 167 bHLH genes, respectively (Robinson and Lopes 2000, Ledent et al. 2002, Bailey and Weisshaar 2003, Li et al. 2006, Simionato et al. 2007, Wang et al. 2009, Zheng et al. 2009).

Animal bHLH proteins are currently classified into groups A, B, C, D, E, and F according to the nucleotide composition of target DNA elements they recognize and the common structural features they possess. Group A and B bHLH proteins recognize and bind DNA elements containing E box CANNTG (N means any nucleotide), which is CA/GC CTG for group A and CA/GTG/TG for group B. Group C bHLH proteins recognize and bind DNA element containing (A/G)CGTG. Most group C proteins also contain a Per-Arm-Sim (PAS) domain that facilitates dimerization with another PAS-containing protein (Jones 2004). Group D bHLH proteins have no basic region. They do not recognize any target DNA elements but form inactive heterodimers with group A bHLH proteins. Group E bHLH proteins recognize and bind DNA elements containing N box CACG(C/A)T. Their bHLH motifs are closely followed by a structural domain named ORANGE. Besides, a WRPW (tryptophan-arginine-proline-tryptophan) motif is present at their carboxyl terminus. Group F bHLH proteins do not have basic region. Instead, they have an IPT (immunoglobulin-like, plekens and transcription factor) structural domain to facilitate dimerization and target DNA binding (Ledent and Vervoort 2001).

Animal bHLH proteins are also divided into 45 families according to their specific functions in regulating eukaryotic growth and
development (Simionato et al. 2007). Insect bHLH genes are distributed into 42 families and each family has one or two members (Liu et al. 2015). Thus, an insect species generally has around 55 bHLH genes in its genome. Although total number of bHLH genes is close in different insects, number of bHLH genes in each family can be quite different. For example, Asian citrus psyllid has two to three bHLH genes in Net, Hand (heart and neural crest derivatives), and SRC (steroid receptor coactivator) families (Peng et al. 2017), while other insect species have only one bHLH gene in each of these families. Mosquitoes have three Ato genes in Atonal family (Zhang et al. 2013), while other insect species have only one Ato gene. Jewel wasp was found to lack Net, MyoR (myogenic repressor) and Fer1 (forty-eight related 1) genes (Liu et al. 2015), which are all present in other insect species. The presence or absence of specific bHLH genes may lead to physiological and behavioral difference among insect species, because each bHLH gene has its specific role(s) in controlling expression of genes related to organismal development. For example, Atonal family genes are involved in developmental regulation of Drosophila chordotonal organs and photoreceptors (Jarman et al. 1995). Hand and SRC family genes play roles in controlling Drosophila heart morphogenesis and larval metamorphosis (Han et al. 2006, Jang et al. 2009). Net and MyoR family genes are found to regulate Drosophila intervein and muscle development respectively (Georgias et al. 1997, Brentrup et al. 2000).

The diamondback moth, Plutella xylostella (L.) (Lepidoptera: Plutellidae), is one of the most aggressive pests of brassica vegetables and oilseed crops (Zalucki et al. 2012, Furlong et al. 2013). The name diamondback moth is based on such fact that a few light-colored diamond shapes are present on posterior margins of its forewings (Adashkevich 1972). P. xylostella larvae feed on leaves of host plants from seedling stage, which greatly affects yield and quality of the crop (Furlong et al. 2013). Diamondback moths have very few natural enemies and strong resistance to various insecticides, including insecticidal toxins. Therefore, they are very hard to be controlled efficiently in field (Talekar and Shelton 1993). The annual cost for pest management against diamondback moth has reached more than US$ 1 billion in the world (Zalucki et al. 2012, Tian et al. 2013). Its resistance to over 79 insecticides, including insecticidal toxins, has led to the inability of growing cruciferous crops in certain areas (Liang et al. 2001, Sun et al. 2012). In view of the importance of bHLH transcription factors in regulating insect tissue/organ development, knowledge of bHLH gene composition in P. xylostella would facilitate further studies on functions of specific bHLH proteins in regulating P. xylostella development and may aid in establishment of biological strategies to control its occurrence. Therefore, in the present study, we employed Blast searches and phylogenetic analyses to identify bHLH genes encoded in the genome of diamondback moth. A comparison with other insects displayed that P. xylostella has additional bHLH genes and/or gene copies in six bHLH families.

Materials and Methods

Data Collection

The amino acids of 45 representative bHLH motifs were prepared from previous report (Ledent and Vervoort 2001). Subsequently, they were used as query sequences to conduct Blastp searches for retrieving candidate bHLH protein sequences in diamondback moth at https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=OGP_.51655_.68127&LINK_LOC=blasttab. 'Annotated proteins' was selected as the target database and all other parameters were of default settings. As a result, a great number of P. xylostella protein sequences were obtained, which were then manually examined to remove the redundant ones. Amino acids of each P. xylostella bHLH (PxblH) motif were used to conduct tBlastn search against P. xylostella genome at https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BasicSearch&BLAST_SPEC=OGP_.51655_.68127&LINK_LOC=blasttab for obtaining contig number, reading frame and coding region(s) of each PxblH motif. 'RefSeq Genomic' was selected as target database, low complexity regions were not filtered, and other parameters were of default values.

Multiple Sequence Alignment

From the above Blastp and tBlast searches, amino acid sequences of candidate P. xylostella bHLH motifs were obtained. Each of the obtained motifs was manually examined to confirm whether it has sufficient conserved amino acids. If 11 or more conserved amino acids are found at the 19 conserved sites as indicated by Atchley et al. (1999), the bHLH motif is considered as a potential bHLH family member. Because bHLH motifs of both groups D and F have no basic region and the typical group D and F bHLH motifs have only 33 and 45 amino acids, the number of minimum conserved amino acids to qualify a group D or F bHLH protein is reduced to 5 and 8, respectively. Amino acid sequences of all eligible bHLH motifs were aligned with MUSCLE (Edgar 2004) program which is embedded in MEGA 5.2 (Tamura et al. 2011) using default settings. The aligned P. xylostella bHLH (PxblH) motifs were saved in FASTA format and subsequently exported to GeneDoc (Edgar 2004) for displaying degrees of amino acid conservatism. The multiple sequence alignment was copied and saved as a rich text file for further annotations.

Phylogenetic Analysis

The qualified PxblH motifs from above examination were subject to phylogenetic analysis for determining their orthology with known bHLH family members. Our previous reports indicated that in-group phylogenetic analysis was efficient in determining whether two genes are orthologous (Wang et al. 2007, Wang et al. 2008). Briefly, this method is divided into two steps. Step 1, all the obtained PxblH sequences were used to construct a maximum likelihood (ML) phylogenetic tree in MEGA 5.2 together with 59 DmbHLH (Drosophila melanogaster) (Meigen) (Diptera: Drosophilidae) bHLH motifs. This tree was referenced for determining to which group a specific PxblH belongs. Step 2, a single PxblH motif was used to construct NJ (neighbor-joining), MP (maximum parsimony), and ML trees with DmbHLH motifs from the group determined in step 1. For example, if step 1 showed that a PxblH motif was located in the phyletic clade formed by DmbHLH motifs from group A, then step 2 used this PxblH motif to construct phylogenetic trees with DmbHLH motifs only from group A. In step 2, if a PxblH motif formed monophyletic clade with a specific DmbHLH motif and all bootstrap values supporting this monophyletic clade were higher than 50, this PxblH was determined as an ortholog of that specific DmbHLH sequence. If a PxblH motif could not form monophyletic clade with any DmbHLH motif or the formed monophyletic clade was supported by bootstrap values below 50, known bHLH motifs from other insect species were used to determine its orthology. Detailed steps for conducting in-group phylogenetic analysis are available in Liu et al. (2015).
Protein Functional Domain Prediction

Full-length PxbHLH protein sequences were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/) using the correspondent protein accession numbers. The obtained protein sequences were then submitted to SMART (simple modular architecture research tool) (http://smart.embl-heidelberg.de/) for prediction of structural domain with default settings.

Results and Discussion

bHLH Family Members in P. xylostella

Through Blastp searches and manual examination, 52 PxbHLH (P. xylostella bHLH) family members were identified in P. xylostella protein databases (Fig. 1). Each identified PxbHLH motif contains more than 11 conserved amino acids, meaning that proteins containing these motifs are eligible bHLH proteins. Through in-group phylogenetic analyses, all identified PxbHLHs have been classified into correspondent bHLH families with bootstrap values higher than 50 (Table 1). Among them, 37 PxbHLH genes were classified and named according to their bHLH orthologs from fruit fly. The rest 15 PxbHLH genes were classified and named according to their bHLH orthologs from domestic silkworm (Bombyx mori (L.) (Lepidoptera: Bombycidae)), jewel wasp (Nasonia vitripennis (Walker) (Hymenoptera: Pteromalidae)), or Asian citrus psyllid (Diaphorina citri (Kuwanyama) (Hemiptera: Liviidae)). P. xylostella has two Ash2 and Dys genes and four SCL genes. These genes have been named as PxAsh2a and PxAsh2b, PxDys1 and PxDys2, and PxSCL1 to PxSCL4. Based on our classification, P. xylostella has 25, 9, 11, 1, 5, and 1 bHLH genes in groups A to F, respectively.

GenBank protein accession numbers and annotations for all 52 PxbHLH family members are listed (Table 1). A comparison between GenBank annotations and our classification shows that not all PxbHLH proteins have been annotated in agreement with our classification. Firstly, our classification of 19 PxbHLH proteins is consistent with GenBank annotations.
| bHLH family | Fruit fly gene | Px bHLH gene | Bootstrap values | Protein accession No. | Annotation in GenBank | Group |
|-------------|----------------|--------------|------------------|----------------------|------------------------|-------|
| ASCa        | Ase            | PxAseBm      | 75 85 96         | XP_011568337.1       | ASC protein T8         | A     |
|             | Ash2Bm         | PxAsh2Bm     | 85 56 87         | XP_011568336.1       | ASC protein T3-like    | A     |
|             | Ash3Bm         | PxAsh3Bm     | 85 53 88         | XP_011552936.1       | ASC protein T3-like    | A     |
| E12/E47     | da             | PxDa         | 99 100 99        | XP_011562492.1       | da                      | A     |
| MyoD        | nau            | PxNau        | 82 95 97         | XP_011562883.1       | MyoD1                  | A     |
| Ngn         | tap            | PxTap        | 80 91 93         | XP_011550236.1       | TAP-like               | A     |
| Mist        | dmm            | PxDmm        | 79 94 97         | XP_011567391.1       | Ngn2                   | A     |
| Beta3       | Oli            | PxOli        | 100 100 100      | XP_011552126.1       | class E bHLH protein   | A     |
| Atonal      | ato            | PxAto        | 72 96 93         | XP_011563274.1       | Atonal-like            | A     |
| Net         | net            | PxNet        | 82 95 94         | XP_011549879.1       | atonal8                | A     |
| MyoRa       | MyoR           | PxMyoR       | 97 76 88         | XP_011554754.1       | scleraxis-like         | A     |
| Delilah     | ix             | PxTx         | 61 96 91         | XP_011566491.1       | fer3-like protein      | A     |
| Mesp        | sage           | PxSage       | 79 99 98         | XP_011558845.1       | transcription factor 15| A     |
| Paraxis     | Pxs            | PxPxs        | 85 77 89         | XP_011547896.1       | twist-related protein 2| A     |
| Twist       | twi            | PxTwi        | 90 88 93         | XP_01154850.1        | PTFα                    | A     |
| PTFα        | Fer1           | PxFer1       | 92 88 90         | XP_011561521.1       | Tal protein 1          | A     |
| PTFb        | Fer2           | PxFer2       | 84 96 90         | XP_011545751.1       | protein Fer3           | A     |
| PTFb        | Fer3           | PxFer3       | 99 99 99         | XP_011549084.1       | atonal7-B-like         | A     |
| Hand        | Hand           | PxHand       | 94 94 86         | XP_011557047.1       | Hypothetical protein   | A     |
| SCL         | SCL            | PxSCL1       | 98 96 99         | XP_011551236.1       | Hypothetical protein   | A     |
|             | SCL2Bm         | PxSCL2Bm     | 81 71 90         | XP_011568653.1       | Hypothetical protein   | A     |
|             | SCL2Bm         | PxSCL2Bm     | 84 63 93         | XP_011551236.1       | Hypothetical protein   | A     |
| NSCL        | NSCL           | PxNSCL       | 92 98 97         | XP_011557149.1       | Hypothetical protein   | A     |
| Mnt         | Mnt            | PxMnt        | 61 53 78         | XP_011557464.1       | MNT-like               | B     |
| Max         | max            | PxMax        | 87 97 91         | XP_011550502.1       | protein max            | B     |
| Mad         | MadBc          | PxMadBc      | 88 97 96         | XP_011551922.1       | Mad 1-like             | B     |
| Myc         | dm             | PxDm         | 98 100 91        | XP_011554509.1       | Myc protein            | B     |
| USF         | USF            | PxUSF        | 90 84 96         | XP_011549434.1       | USF2                    | B     |
| MITF        | Mitf           | PxMitf       | 91 100 99        | XP_011566220.1       | Mitf-like              | B     |
| TF4         | tf4            | PxBmx        | 92 92 90         | XP_011552773.1       | max-like protein X     | B     |
| MLX         | Mio            | PxMio        | 74 84 95         | XP_011546485.1       | MLX-interacting protein| B     |
| SRC         | tae            | PxTaeBm      | 81 98 95         | XP_011554670.1       | Hypothetical protein   | B     |
| Clock       | clk            | PxClk        | 94 99 98         | XP_011531451.1       | Clock                   | C     |
| Met         | Met            | PxMetBm      | 85 93 94         | XP_011557479.1       | ARNT2                   | C     |
| JHRBm       | JHRBm          | PxJHRBm      | 96 97 97         | 1934544.4            | Hypothetical protein   | C     |
| HNR         | Dps1           | PxDps1       | 99 99 99         | XP_011559031.1       | NPA54                   | C     |
|             | Dps2           | PxDps2       | 99 99 99         | XP_011565531.1       | NPA54                   | C     |
|             | ss             | PxSs         | 80 100 97        | XP_011548350.1       | AHR                     | C     |
| Sm          | Sm             | PxSm         | 96 100 100       | XP_011561807.1       | Sim-like                | C     |
| Trh         | trh            | PxTrh        | 54 97 96         | XP_011556672.1       | protein Trh            | C     |
| HIF         | sima           | PxSima       | 93 85 96         | XP_011553430.1       | HIP1α                   | C     |
| ARNT        | tgo            | PxTgo        | 75 100 100       | XP_011558891.1       | ARNT                    | C     |
| Bmal        | cyc            | PxCycBm      | 65 54 60         | XP_011557457.1       | protein cycle           | C     |
| Emc         | Emc            | PxEmc        | 80 94 91         | XP_011568487.1       | protein Emc             | C     |
| Hey         | Hey            | PxHey        | 53 81 88         | XP_011560171.1       | Hey1                    | C     |
| E(spl)      | e(spl)         | PxGs         | 87 94 96         | XP_011562380.1       | hairy-like              | E     |
|             | e(spl)md       | PxE(spl)mdBm | 92 91 98         | XP_011563228.1       | E(spl)md                 | E     |
| COE         | kn             | PxKn         | 92 99 99         | XP_011563271.1       | COE                      | F     |

Each PxbHLH gene is named according to its ortholog of fruit fly (D. melanogaster) or other insects as indicated with superscript letters. Bootstrap values were from in-group phylogenetic analyses. For group B candidates, OsRa (the Oryza sativa bHLH motif sequence of R family) was used as outgroup. For group A and C–F candidates, DmMnt (a D. melanogaster bHLH motif sequence of B group) was used as outgroup. Superscript letters Bm, Dc, and Nv indicate gene orthology assignment using Bombyx mori (Bm), Diaphorina citri (Dc), and Nasonia vitripennis (Nv) bHLH motifs. In the last column, bold letters indicate consistent GenBank annotations with our classifications. Bold-italic letters indicate that GenBank annotations are based on family names which do not contain any information about its orthology with known insect bHLH gene. Italic letters indicate different GenBank annotations with our classification. Normal letters indicate hypothetical protein.
annotation and our classification to protein No. XP_01156249.1 are da (daughterless). Secondly, GenBank annotations to 14 PxbHLH proteins are based on bHLH family names which do not contain any information about its orthology with known insect bHLH gene. For example, GenBank annotation to protein No. XP_011568337.1 is ASC (achaete-scute complex) protein T8, which is based on the family name ASC. Our classification to this protein is Ase (asense), which is a specific gene name in ASC family. Thus, our classification provides useful information for improving annotations to these 14 PxbHLH proteins. Thirdly, GenBank annotations to 13 PxbHLH proteins are different with our classification. For example, GenBank annotation to protein No. XP_011566759.1 is Ngn2 (neurogenin 2). It is Dimm (dimmed) in our classification. Our classification to each PxbHLH protein is based on in-group phylogenetic analysis supported by bootstrap values higher than 50, while GenBank annotation is mainly based on its sequence identity with known proteins. Thus our classification is considered to be more accurate than GenBank annotation. Finally, six PxbHLH proteins are annotated as hypothetical proteins in GenBank. They have been classified as specific bHLH genes by us. Thus, six new bHLH proteins are found in *P. xylostella* protein databases.

**Structural Domains in PxbHLH Protein Sequences**

Previous studies revealed that bHLH proteins of group C, E, and F usually possess typical conserved structural domains (Jones 2004). To further validate the reliability of our classification, we constructed an ML phylogenetic tree with the 52 PxbHLH motif sequences (Fig. 2, left panel) and predicted structural domains of PxbHLH proteins using SMART program (Fig. 2, right panel). Eleven PxbHLH proteins of group C have two PAS (Per-Arnt-Sim) domains and nine of them have a PAC (C-terminal to PAS motif) domain, while five members of group E have ORANGE domain, and PxKn protein of group F has three additional domains, viz. COE1 (collier/offactory-1/early B-cell factor), IPT (immunoglobulin plexin transcription factor) and MSF1 (major facilitator superfamily 1) (Fig. 2). In summary, typical structural domains are present in PxbHLH proteins of groups C, E, and F respectively. Therefore, our classification to PxbHLH proteins of these groups is not only supported by in-group phylogenetic analysis with bootstrap values higher than 50 but also supported by presence of specific structural domains in these proteins.

Among all PxbHLH proteins, PxHand has dual HLH motifs. Previously, four bHLH proteins, viz. Clk (clock), Sima (similar), Cyc (cycle), and Cwo (clockwork orange), were found to have dual HLH motifs in Asian citrus psyllid (Peng et al. 2017). No bHLH proteins have dual HLH motifs in jewel wasp, human body louse, and brown planthopper (Wang et al. 2014, Liu et al. 2015, Wan et al. 2016). In order to see whether other insect bHLH proteins have dual HLH motifs, full-length bHLH protein sequences of ten insect species, viz. fruit fly (*D. melanogaster*), yellow fever mosquito (*Aedes aegypti* (L.) (Diptera: Culicidae)), African malaria mosquito (*Anopheles gambiae* (Giles) (Diptera: Culicidae)), southern house mosquito (*Culex quinquefasciatus* (Jupp) (Diptera: Culicidae)), honey bee (*Apis mellifera* (L.) (Hymenoptera: Apidae)), Jerdon’s jumping ant (*Harpegnathos saltator* (Jerdon) (Hymenoptera: Formicidae)), domestic silkworm (*B. mori*), monarch butterfly (*Danaus plexippus* (L.) (Lepidoptera: Nymphalidae)), red flour beetle (*Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae))

![Fig. 2. Architecture of P. xylostella bHLH protein conserved domains. The left panel is a ML tree constructed using 52 PxbHLH motif amino acids with OsRa (the *Oryza sativa* bHLH motif sequence of R family) as outgroup. PxbHLH names of groups A to F are displayed in blue, red, green, purple, magenta, and aqua, respectively. The right panel is a schematic diagram showing HLH and other protein domains detected by SMART program online. Seven different protein domains, namely HLH, PAS, PAC, ORANGE, COE1, IPT, and MSF1, are found in *P. xylostella* bHLH proteins.](https://academic.oup.com/jinsectscience/article-abstract/18/3/17/5036101)
### Table 2. Coding regions, intron location and length of 52 *P. xylostella* bHLH motifs

| Family | PxBHLH name | Genomic coding sequence(s) | Intron location and length | Group |
|--------|--------------|----------------------------|----------------------------|-------|
|        |              | Contig No. | Frame | Coding region(s) |                   |
| ASCa   | PxAse        | NW_011952028.1 | +2    | 991757–991951   | A      |
|        | PxAsh2a      | NW_011952110.1 | +2    | 351530–351721   | A      |
|        | PxAsh2b      | NW_011952028.1 | -2    | 848610–848419   | A      |
|        | PxAsh3       | NW_011952110.1 | -2    | 372924–372733   | A      |
| E12/E47| PxDa         | NW_011952010.1 | -3    | 998879–998851   | Basic: 3026 bp |
| MyoD   | PxNau        | NW_011952428.1 | -1    | 92621–92577     | Helix 1: 20855 bp |
|        |              |             | -3    | 71721–71614     | A      |
| Ngn    | PxTap        | NW_011952067.1 | -1    | 273995–273837   | A      |
| Mist   | PxDimm       | NW_011952025.1 | -1    | 397556–397494   | Helix 1: 308 bp |
|        |              |             | -3    | 397185–397087   | A      |
| Beta3  | PxOli        | NW_011952096.1 | -1    | 760649–760602   | Helix 1: 400 bp |
|        |              |             | -2    | 760201–760085   | A      |
| Atonal | PxAto        | NW_011952451.1 | +2    | 100991–101149   | A      |
| MyoRa  | PxMyoR       | NW_011952029.1 | -2    | 1464417–1464353 | Helix 1: 3828 bp |
|        |              |             | -3    | 1460524–1460431 | A      |
| Delilah| PxTx         | NW_011952025.1 | -2    | 1659310–1659134 | A      |
| Mesp   | PxSage       | NW_011952173.1 | +3    | 428397–428527   | Helix 2: 662 bp |
|        |              |             | +3    | 429190–429220   | A      |
| Paraxis| PxPxs        | NW_011952256.1 | +3    | 333673–333718   | Helix 1: 387 bp |
| Twist  | PxTwI        | NW_011952038.1 | +3    | 1035549–1035704 | A      |
| PTFa   | PxFer1       | NW_011952151.1 | +2    | 98544–98386     | A      |
| PTFb   | PxFer2       | NW_011952355.1 | +3    | 145638–145681   | Helix 1: 1858 bp |
|        |              |             | +3    | 147540–147654   | A      |
| Hand   | PxHand       | NW_011952203.1 | +3    | 535854–535967   | Helix 2: 1283 bp |
|        |              |             | +2    | 537251–537295   | A      |
| SCL    | PxSCL1       | NW_011952051.1 | -2    | 276186–276047   | Helix 2: 868 bp |
|        |              |             | -2    | 275178–275160   | A      |
|        | PxSCL2       | NW_011952031.1 | +3    | 921273–921414   | Helix 2: 958 bp |
|        |              |             | +3    | 922373–922391   | A      |
|         |              |             | -1    | 894211–894354   | Helix 2: 6842 bp |
|         |              |             | +3    | 901197–901211   | A      |
|         |              |             | -3    | 901291–901311   | A      |
|         |              |             | -2    | 260131–259992   | Helix 2: 634 bp |
| SCL    | PxSCL3       | NW_011952080.1 | +3    | 20603–20623     | A      |
|         |              |             | +2    | 932273–932412   | Helix 2: 11849 bp |
|         |              |             | +1    | 944262–944280   | A      |
|         |              |             | -3    | 245480–245322   | A      |
|         |              |             | -2    | 651821–651717   | Loop: 761 bp |
|         |              |             | -2    | 650955–650902   | B      |
| Mad    | PxMAD        | NW_011952092.1 | -1    | 980031–979997   | Basic: 155422 bp |
|         |              |             | -2    | 824574–824460   | Helix 2: 12791 bp |
|         |              |             | -1    | 811676–811668   | A      |
| Myc    | PxDm         | NW_011952144.1 | +2    | 424292–424134   | B      |
| USF    | PxUsf        | NW_011952056.1 | +2    | 192728–192886   | B      |
| MITF   | PxMif        | NW_011952746.1 | -3    | 10517–10496     | Basic: 1104 bp |
|         |              |             | -3    | 9391–9234       | B      |
|         |              |             | -1    | 360653–360632   | Basic: 340 bp |
|         |              |             | -1    | 360291–360134   | A      |
and pea aphid (Acyrthosiphon pisum (Harris) (Hemiptera: Aphididae)), were retrieved and analyzed using SMART program online. As a result, Hand protein of D. plexippus, Fer2 (forty-eight related 2), and Gce (germ cell-expressed) proteins of C. quinquefasciatus were found to have dual HLH motifs. Therefore, dual HLH motifs have been found in four insect species, among which P. xylostella and D. plexippus belong to Lepidoptera, C. quinquefasciatus belongs to Diptera and D. citri belongs to Hemiptera. In summary, dual HLH motifs exist in Hand protein of two Lepidopteran insects, P. xylostella and D. plexippus, but not in B. mori; Dual motifs were also found in two bHLH proteins of one Dipteran insect, C. quinquefasciatus, but not existed in other Dipteran insects, like D. melanogaster, A. aegypti and A. gambiae; The HLH dual motifs were also identified in four bHLH proteins of one Hemipteran insect, D. citri, but not found from two other Hemipteran insects, A. pisum and Nilaparvata lugens (Stal) (Hemiptera: Delphacidae). Because only a few dual HLH motifs are found in all bHLH proteins of fifteen insect species and dual HLH motifs are only shared by Hand protein of two Lepidopteran insect species, it is considered that these dual HLH motifs were not inherited from the common ancestor of insects. Instead, they were resulted from independent duplication of HLH-coding DNA segment in individual species or in specific lineage of insects.

Genomic Coding Regions of PxbHLH Motifs
The coding information of 52 PxbHLH motifs is listed in Table 2. Five PxbHLH genes were found to have multiple copies in P. xylostella genome. Among them, PxSage and PxSCL2 have three copies, while PxMyoR, PxMitf, and PxSima have two copies. Thirty-two PxbHLH motifs were found to have coding regions interrupted by introns. Among them, coding regions of PxMad, PxBmx, PxH, and PxDpn motifs are interrupted by two introns respectively, and each of the rest 28 PxbHLH motifs is interrupted by one intron respectively. A comparison with other insect species (Table 3) reveals that P. xylostella has the highest number of bHLH motifs having introns and the highest number of total introns. Besides, it occupies the first, second, and fourth place in length of the shortest intron, length of the longest intron and average length of introns, respectively. These data indicate that coding regions of PxbHLH motifs are interrupted by more and longer introns than most other insects. These data could have important implications for future studies concerning intron gain or loss events during bHLH gene evolution.

Special bHLH Genes in P. xylostella
Up to now, bHLH repertories have been established for 15 insect species. Their gene numbers in each bHLH family are listed in Table 4. A comparison with other insects reveals the existence of special bHLH genes in P. xylostella.
Firstly, *P. xylostella* has four stem cell leukemia (*SCL*) genes among which *PxSCL2* has three copies. The multiple genes were defined because they have the different amino acid sequences, while gene copies were named because these gene copies have the identical amino acid sequences. There is only one copy of *SCL* gene in all other insects whose bHLH repertoire has been established. Phylogenetic tree constructed using SCL bHLH motif amino acids of 15 insect species displays that the four *PxSCL* genes cluster in a separate clade, indicating that they are originated from species-specific gene duplication in *P. xylostella* (Fig. 3a). *SCL* gene was first discovered in a human leukemic stem-cell line (Begley et al. 1989). It is expressed in a number of cells including hematopoietic stem cells, megakaryocytes, cells, progenitor cells, and committed erythroblasts. It plays a significant role in regulating the proliferation and differentiation of various hematopoietic cells (Begley et al. 1991, Green and Begley 1992, Curtis et al. 2012, Real et al. 2012). In *D. melanogaster*, restricted expression of *SCL* was observed in a subset of cells in the developing central nervous system (Varterasian et al. 1993). It would be interesting to study where and when the four *PxSCL* genes are expressed and what mechanisms are employed by *PxSCL* proteins to regulate growth and development in *P. xylostella*.

Secondly, *P. xylostella* has two *Dys* genes. Among the 15 insect species, all three Lepidopteran species (i.e., *B. mori*, *Danaus plexippus* and *P. xylostella*) have two *Dys* genes, while other insects have only one *Dys* gene. A phylogenetic tree constructed using *Dys* bHLH motif amino acids of 15 insect species (Fig. 3b) shows that *Dys1* and *Dys2* genes of *B. mori*, *D. plexippus*, and *P. xylostella* are located in separate clades, respectively. Such phylogenetic pattern demonstrates that the double *Dys* genes are originated from lineage-specific gene expansion in Lepidoptera. In fruit fly, *Dys* (dysfusion) is responsible for regulating gene expression in tracheal fusion (Jiang and Crews 2007). It is also involved in the regulation of pro-apoptosis and head involution defective in tarsal joints (Iordanou et al. 2011). Because the basic regions of *PxDys1* and *PxDys2* have three different amino acids (Fig. 1), it is possible that *PxDys1* and *PxDys2* proteins recognize different target DNA elements and play different regulatory roles in trachea development of *P. xylostella*.

Thirdly, *P. xylostella* has two copies of *MyoR*, *Mitf*, and *Sima* genes, respectively and three copies of *Sage* gene. Among them, the coding regions of *PxMyoR*, *PxSage*, and *PxMitf* bHLH motifs are interrupted by one intron of different length, respectively, suggesting that each gene copy has diverged slightly after it was duplicated. *MyoR* (myogenic repressor) gene is expressed in undifferentiated myoblasts and down-regulated in myoblast differentiation (Lu et al. 1999). *Sage* (salivary gland-expressed bHLH) protein can form dimer with *Da* (daughterless) protein, which is necessary to maintain expression of *sens* gene in embryonic salivary gland. The expression of *sens* gene can prevent apoptosis of salivary gland cells in embryos (Chandrasekaran and Beckendorf 2003). *Mitf* (microphthalmia transcription factor) gene is expressed during *Drosophila* embryonic development and in *Drosophila* eye-buds/antennae-buds (Hallsson et al. 2004). It is also involved in regulating lysosomal biogenesis and expression of multiple V-ATPase in *D. melanogaster* (Tognon et al. 2016). *Sima* (similar) and *Tgo* (tango) form a complex that activates the corresponding

### Table 3. Intron number and length in coding regions of insect bHLH motifs

| Insect species | No. of bHLH motifs having introns | Total no. of introns | Length of the shortest intron (bp) | Length of the longest intron (bp) | Average length of introns (bp) |
|---------------|----------------------------------|---------------------|-----------------------------------|-----------------------------------|------------------------------|
| **Holometabola** |                                  |                     |                                   |                                   |                              |
| Diptera       |                                  |                     |                                   |                                   |                              |
| *Aedes aegypti* (Aa) | 25                                 | 28                  | 36                                | 315 344                           | 16 707                       |
| *Anopheles gambiae* (Ag) | 21                                 | 23                  | 57                                | 37 485                            | 2 279                        |
| *Culex quinquefasciatus* (Cq) | 22                                 | 24                  | 56                                | 14 434                            | 2 464                        |
| *Drosophila melanogaster* (Dm) | 18                                 | 20                  | 57                                | 11 845                            | 1 027                        |
| **Hymenoptera** |                                  |                     |                                   |                                   |                              |
| *Apis mellifera* (Am) | 24                                 | 29                  | 72                                | 129 558                           | 11 020                       |
| *Nasonia vitripennis* (Nv) | 22                                 | 27                  | 77                                | 174 325                           | 11 715                       |
| *Harpegnathos saltator* (Hs) | 23                                 | 27                  | 82                                | 127 364                           | 6 326                        |
| **Lepidoptera** |                                  |                     |                                   |                                   |                              |
| *Bombbyx mori* (Bm) | 24                                 | 28                  | 78                                | 11 651                            | 1 749                        |
| *Danaus plexippus* (Dp) | 25                                 | 30                  | 74                                | 4 539                             | 607                          |
| *Plutella xylostella* (Px) | 32                                 | 36                  | 222                               | 155 422                           | 9 663                        |
| **Coleoptera** |                                  |                     |                                   |                                   |                              |
| *Tribolium castaneum* (Tc) | 24                                 | 29                  | 44                                | 100 326                           | 4 841                        |
| **Paraneoptera** |                                  |                     |                                   |                                   |                              |
| **Hemiptera** |                                  |                     |                                   |                                   |                              |
| *Diaphorina citri* (Dc) | 23                                 | 28                  | 82                                | 68 654                            | 6 759                        |
| *Acyrthosiphon pisum* (Ap) | 28                                 | 36                  | 62                                | 30 718                            | 4 003                        |
| *Nilaparvata lugens* (Nl) | 23                                 | 29                  | 58                                | 14 128                            | 2 736                        |
| **Phthiraptera** |                                  |                     |                                   |                                   |                              |
| *Pediculus humanus corporis* (Phc) | 22                                 | 27                  | 66                                | 6 723                             | 695                          |
| Average       | 24                                 | 28                  | 75                                | 80 168                            | 5 326                        |

Insect species have been organized into two groups (i.e., Holometabola and Paraneoptera) under infraclass Neoptera of class Insecta. Data of *P. xylostella* are from this study. Data of *Danaus plexippus* are from our unpublished work. Data of *Apis mellifera*, *Pediculus humanus corporis* (Light) (Phthiraptera: Pediculidae), *Diaphorina citri*, *Acyrthosiphon pisum*, *Harpegnathos saltator*, *Bombbyx mori*, *Aedes aegypti*, *Anopheles gambiae*, *Nasonia vitripennis* and *Culex quinquefasciatus* are from previous reports (Wang et al. 2007, Wang et al. 2008, Deng et al. 2011, Liu et al. 2012, Zhang et al. 2013, Wang et al. 2014, Liu et al. 2015, Peng et al. 2017). Data of *Drosophila melanogaster*, *Nilaparvata lugens*, and *Tribolium castaneum* are from our own survey based on reports of Simionato et al. (2007), Bitra et al (2009), and Wan et al (2016). The same sources of data are used in Table 4.
gene expression under hypoxic condition (Lavista-Llanos et al. 2002). Under hypoxic condition, Sima protein accumulates in Drosophila SL2 cells (Bacon et al. 1998). Taken together, MyoR, Sage, Mitf, and Sima genes are mainly involved in regulation of myoblast differentiation, seng gene expression, eye development, and gene expression under hypoxic condition. Further studies may be conducted to understand functions of these multiple copy genes in regulating growth and development of specific cells/tissues such as myoblasts and eye-buds in P. xylostella.

Finally, it is to be noted that we have not found AP4 (activating element-binding protein 4) and SREBP (sterol regulatory element-binding protein) genes in P. xylostella, while all other 14 insects have one to three such genes. AP4 is a protein that binds to viral SV40 enhancer elements and activates viral late transcription (Mermod et al. 1988). In addition, AP4 can form a complex with geminin and negatively regulate its target gene in non-neuronal cells (Kim et al. 2006). SREBP is crucial to survival of Drosophila larvae. If this gene was deleted, Drosophila

### Table 4. bHLH family members in 15 insect species

| Group | bHLH family | Holometabola | Lepidoptera | Colb | Hemiptera | Phlb |
|-------|-------------|--------------|-------------|------|------------|------|
| A     | ASCa        | 3 2 3 4      | 2 2 2 2     | 4 4 4 2 | 1 0 2 2   | 1    |
|       | ASCb        | 1 0 1 0      | 0 0 0 0     | 0 0 0 1 | 2 1 0 1   | 1    |
| E12/E47 | MyoD       | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 0 0 1 1   | 1    |
|       | Ngn         | 1 1 2 1      | 1 1 1 1     | 1 1 1 1 | 0 1 2 1   | 1    |
|       | NeuroD      | 1 1 0 0      | 0 0 0 0     | 0 0 0 1 | 0 0 0 0   | 0    |
|       | Mist        | 1 1 1 1      | 2 2 2 2     | 1 1 1 1 | 2 2 2 2   | 2    |
|       | Beta3       | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | Atonal      | 5 4 5 3      | 3 3 3 3     | 1 1 1 3 | 3 3 3 3   | 3    |
|       | Olig        | 0 0 0 0      | 0 0 0 0     | 0 0 0 0 | 0 0 0 0   | 0    |
|       | Net         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 2 1 1 1   | 1    |
|       | MyoRa       | 1 1 1 1      | 1 0 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | MyoRb       | 0 0 0 0      | 0 0 0 0     | 0 0 0 0 | 0 0 0 0   | 0    |
|       | Delilah     | 1 1 1 1      | 0 0 0 0     | 1 1 1 2 | 1 1 3 1   | 1    |
|       | Mesp        | 1 1 1 1      | 1 1 1 1     | 1 1 1 0 | 1 1 1 1   | 1    |
|       | Paraxis     | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 0 1 1 0   | 1    |
|       | Twist       | 1 1 1 1      | 1 1 2 1     | 2 1 2 1 | 1 1 1 1   | 1    |
|       | PTFa        | 2 2 2 2      | 2 2 2 2     | 2 2 2 2 | 2 2 2 2   | 2    |
|       | PTFb        | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 0 1 1 2   | 0    |
|       | Hand        | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 3 1 1 1   | 1    |
|       | SLCL        | 1 1 1 1      | 1 1 1 1     | 1 1 1 4 | 1 1 1 1   | 1    |
| B     | Mnt         | 0 0 0 0      | 0 1 0 1     | 0 1 1 1 | 0 1 1 0   | 0    |
|       | Max         | 1 1 1 1      | 1 1 2 1     | 1 1 1 1 | 1 3 1 1   | 1    |
|       | Myc         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | USF         | 1 1 1 1      | 2 1 2 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | MITF        | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 2   | 1    |
|       | AP4         | 1 1 1 1      | 1 1 1 2     | 2 1 2 1 | 1 1 1 2   | 1    |
|       | TF4         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 2 1 1   | 1    |
|       | MLX         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 0 1 1 1   | 0    |
|       | SREBP       | 1 1 2 2      | 1 1 1 1     | 1 1 1 0 | 1 1 1 3   | 1    |
|       | Figa        | 0 0 0 0      | 0 0 0 0     | 0 0 0 0 | 0 0 0 0   | 0    |
|       | SRC         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
| C     | Clock       | 2 2 2 2      | 3 3 3 3     | 3 3 3 3 | 2 2 2 2   | 2    |
|       | AHR         | 2 2 2 2      | 2 2 2 2     | 3 3 3 3 | 2 2 2 2   | 2    |
|       | Sim         | 1 2 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 2 1   | 1    |
|       | Thr         | 1 1 1 2      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | HIF         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | ARNT        | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | BMAL        | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
| D     | Enc         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 1 1   | 1    |
|       | Hey         | 3 3 3 2      | 2 2 2 2     | 2 2 2 2 | 2 2 2 2   | 2    |
|       | H/E(spl)    | 4 4 4 11     | 6 4 6 5     | 5 3 6 6 | 6 7 6 8   | 8    |
| F     | COE         | 1 1 1 1      | 1 1 1 1     | 1 1 1 1 | 1 1 2 1   | 1    |
|       | Total       | 55 55 57 59  | 55 55 48 56 | 52 53 52 54 | 52 55 60 55 | 5    |

Uncertainty of classification that previously existed in families ASCb, Hey, and H/E(spl) has been eliminated through our in-depth phylogenetic analysis. Please refer to Table 3 for full names of individual insect species.

Col: Coleoptera.

Ph: Phthiraptera.

Finally, it is to be noted that we have not found AP4 (activating element-binding protein 4) and SREBP (sterol regulatory element-binding protein) genes in P. xylostella, while all other 14 insects have one to three such genes. AP4 is a protein that binds to viral SV40 enhancer elements and activates viral late transcription (Mermod et al. 1988). In addition, AP4 can form a complex with geminin and negatively regulate its target gene in non-neuronal cells (Kim et al. 2006). SREBP is crucial to survival of Drosophila larvae. If this gene was deleted, Drosophila
larva growth was severely blocked, larval growth was severely blocked and larvae died before 3rd instar molting (Kunte et al. 2006). In view of the importance of these two genes in regulating animal growth and development, P. xylostella seems unlikely to lack these two genes. It is probably because the genome database of P. xylostella is incomplete. Therefore, when the genome sequences of diamondback moth are further refined in the future, we would come to check these data again.

Acknowledgments

We are thankful to NCBI (National Center for Biotechnology Information) for free access to various sequence databases. This work was supported by the National Basic Research Program (973) of China (2012CB114604) and the Project Funded by Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

References Cited

Adashevich, B. P. 1972. Biological features of Platella maculipennis Curt. in conditions of southwest of the USSR. Protection of Vegetable Plants, Collection of Papers. V. 12: 72–78.

Atchley, W. R., W. Terhalle, and A. Dress. 1999. Positional dependence, clukes, and predictive motifs in the bHLH protein domain. J. Mol. Evol. 48: 501–516.

Bacon, N. C., P. Wappner, J. F. O’Rourke, S. M. Bartlett, B. Shilo, C. W. Pugh, and P. J. Ratcliffe. 1998. Regulation of the Drosophila bHLH-PAS protein Sma by hypoxia: functional evidence for homology with mammalian HIF-1 alpha. Biochem. Biophys. Res. Commun. 249: 811–816.

Bailey, P. C., C. Martin, G. Toledo-Oritz, P. H. Quail, E. Huq, M. A. Heim, M. Jakoby, M. Werber, and B. Weinshaar. 2003. Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. Plant Cell. 15: 2497–2502.

Begley, C. G., P. D. Aplan, M. P. Davey, K. Nakahara, K. Tchorz, J. Kurtzberg, M. S. Hershfield, B. F. Haynes, D. I. Cohen, and T. A. Waldmann. 1989. Chromosomal translocation in a human leukemia stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc. Natl. Acad. Sci. USA 86: 2031–2035.

Begley, C. G., J. Visvader, A. R. Green, P. D. Aplan, D. Metcalf, I. R. Kirsch, and N. M. Gough. 1991. Molecular cloning and chromosomal localization of the murine homolog of the human helix-loop-helix gene SCL. Proc. Natl. Acad. Sci. USA 88: 869–873.

Bita, K., A. Tan, A. Dowling, and S. R. Palli. 2009. Functional characterization of PAS and HES family bHLH transcription factors during the metamorphosis of the red flour beetle, Tribolium castaneum. Gene. 448: 74–87.

Brentrup, D., H. Lerch, H. Jäckle, and M. Noll. 2000. Regulation of Drosophila wing vein patterning: net encodes a bHLH protein repressing rhomboid and is repressed by rhomboid-dependent Egfr signalling. Development. 127: 4729–4741.

Chandrasekaran, V., and S. K. Beckendorf. 2003. Senseless is necessary for the survival of embryonic salivary glands in Drosophila. Development. 130: 4719–4728.

Curtis, D. J., J. M. Salmon, and J. E. Pimanda. 2012. Concise review: blood relatives: formation and regulation of hematopoietic stem cells by the basic helix-loop-helix transcription factors stem cell leukemia and lymphoblastic leukemia-derived sequence 1. Stem Cells. 30: 1053–1058.

Dang, C. W., Y. Wang, K. P. Chen, Q. Yao, D. B. Zhang, and M. Guo. 2011. The basic helix-loop-helix transcription factor family in the pea aphid, Acyrthosiphon pisum. Journal of Insect Science 11: 1–15.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Ferré-D’Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley. 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature. 363: 38–45.

Furlong, M. J., D. J. Wright, and L. M. Dosdall. 2013. Diamondback moth ecology and management: problems, progress, and prospects. Annu. Rev. Entomol. 58: 517–541.

Georgias, C., M. Wasser, and U. Hinz. 1997. A basic-helix-loop-helix protein expressed in precursors of Drosophila longitudinal visceral muscles. Mech. Dev. 69: 115–124.

Green, A. R., and C. G. Begley. 1992. SCL and related hematopoietic helix-loop-helix transcription factors. Int. J. Cell Cloning. 10: 269–276.

Green, A. R., E. Salvaris, and C. G. Begley. 1991. Erythroid expression of the ‘helix-loop-helix’ gene. SCL. Oncogene. 6: 475–479.

Hallsson, J. H., B. S. Hafildadóttir, C. Stivers, W. Odendahl, H. Arnold, F. Pignoni, and E. Steingrímsson. 2004. The basic helix-loop-helix leucine zipper transcription factor Mitf is conserved in Drosophila and functions in eye development. Genetics. 167: 233–241.

Han, Z., P. Yi, X. Li, and E. N. Olson. 2006. Hand, an evolutionarily conserved bHLH transcription factor required for Drosophila cardiogenesis and hematopoiesis. Development. 133: 1175–1182.

Fig. 3. Evolutionary relationship among insect SCL and Dys genes. (a) A maximum-likelihood phylogenetic tree based on bHLH motif amino acids encoded by SCL genes of 15 insect species. Phylogenetic clades shown in thick lines indicate species-specific gene duplication in P. xylostella. Both trees have been rooted using the DmMnt (D. melanogaster Mnt) motif amino acids. Sequence names are indicated using a two-letter abbreviation of species name plus gene name. Please refer to Table 3 for full names of individual insect species.
Iordanou E., R. R. Chandran, N. Blackstone, and L. Jiang. 2011. RNAi interference by dsRNA injection into Drosophila embryos. J. Vis. Exp. 50: 2477.

Jang, A. C., Y. C. Chang, J. Bai, and D. Montell. 2009. Border-cell migration requires integration of spatial and temporal signals by the RTB protein Abrupt. Nat. Cell Biol. 11: 569–579.

Jarman, A. P., Y. Sun, L. Y. Jan, and Y. N. Jan. 1995. Role of the proneural gene, aonal, in formation of Drosophila chordotonal organs and photoreceptors. Development. 121: 2019–2030.

Jiang, L., and S. T. Crews. 2007. Transcriptional specificity of Drosophila dysfunction and the control of tracheal fusion cell gene expression. J. Biol. Chem. 282: 28659–28668.

Jones, S. 2004. An overview of the basic helix-loop-helix proteins. Genome Biol. 5: 226.

Kim, M. Y., B. C. Jeong, J. H. Lee, H. J. Kee, H. Kook, N. S. Kim, Y. H. Kim, J. K. Kim, K. Y. Ahn, and K. K. Kim. 2006. A repressor complex, AP4 transcription factor and geminin, negatively regulates expression of target genes in nonneuronal cells. Proc. Natl. Acad. Sci. USA 103: 13074–13079.

Kunte, A. S., K. A. Matthews, and R. B. Rawson. 2006. Fatty acid auxotrophy in Drosophila larvae lacking SREBP. Cell Metab. 3: 439–448.

Lavista-Llano, S., L. Centanin, M. Irisarri, D. M. Russo, J. M. Gleadle, S. N. Bocca, M. Muzzopappa, P. J. Ratcliffe, and P. Wappner. 2002. Control of the hypoxic response in Drosophila melanogaster by the basic helix-loop-helix PAS protein similar. Mol. Cell. Biol. 22: 6842–6853.

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

Ledent, V., O. Paquet, and M. Vervoort. 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3: 1–18.

Li, X., X. Duan, H. Jiang, Y. Sun, Y. Tang, Z. Yuan, J. Guo, W. Liang, L. Chen, J. Yin, et al. 2006. Genome-wide analysis of basic/helix-loop-helix transcription factors in brown planthopper, Plutella xylostella (Lepidoptera: Plutellidae). Proc. Natl. Acad. Sci. USA 103: 557–562.

Liang, P., X. Gao, B. Zheng, and H. Dai. 2001. Study on resistance mechanism and cross-resistance of abamectin in diamondback moth Plutella xylostella (L.). Chin. J. Pestic. Sci. 3: 41–45.

Liu A. K., Y. Wang, C. W. Dang, D. B. Zhang, H. F. Song, Q. Yao, and K. P. Chen. 2012. A genome-wide identification and analysis of the basic helix-loop-helix transcription factors in rat and mouse. BMC Evol. Biol. 7: 33.

Liu X. T., Y. Wang, X. H. Wang, X. F. Tao, Q. Yao, and K. P. Chen. 2015. Phylogenetic analyses of vector mosquito basic helix-loop-helix transcription factors. Insect Mol. Biol. 22: 249–257.

Nasonia vitripennis (Hymenoptera: Pteromalidae). Genome 57: 715–723.

Varterasian, M., S. Lipkowitz, I. Karsch-Mizrachi, B. Paterson, and I. Kirsch. 1993. Two new Drosophila genes related to human hematopoietic and neurogenetic transcription factors. Cell Growth Diff. 4: 885–889.

Wan P. J., S. Y. Yuan, W. X. Wang, X. Chen, F. X. Lai, and Q. Fu. 2016. A genome-wide identification and analysis of the basic helix-loop-helix transcription factors in brown planthopper, Plutella xylostella (Lepidoptera: Plutellidae). J. Biol. Chem. 282: 28659–28668.

Wang, Y., K. Chen, Q. Yao, W. Wang, and Z. Zhi. 2007. The basic helix-loop-helix transcription factor family in Bombyx mori. Dev. Genes Evol. 217: 715–723.

Wang, Y., K. Chen, Q. Yao, W. B. Wang, and Z. Zhi. 2008. The basic helix-loop-helix transcription factor family in the honey bee, Apis mellifera. J. Insect Sci. 8: 1–12.

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. 69: 629–640.

Wang, W. X., Y. Wang, D. B. Zhang, A. K. Liu, Q. Yao, and K. P. Chen. 2014. A genome-wide identification of basic helix-loop-helix motifs in Pediculus humanus corporis (Phthiraptera: Pediculidae). J. Insect Sci. 14: 195.

Zalucki, M. P., A. Shabbir, R. Silva, D. Adamson, L. Shu-Sheng, and M. J. Furlong. 2012. Estimating the economic cost of one of the world’s major insect pests, Plutella xylostella (Lepidoptera: Plutellidae): just how long is a piece of string? J. Econ. Entomol. 105: 1115–1129.

Zheng X. D., Y. Wang, Q. Yao, Z. Yang, and K. P. Chen. 2009. A genomewide survey on basic helix-loop-helix transcription factors in rat and mouse. Mamm. Genome 20: 236–246.