www.sciencevision.org

Original Research

Sci Vis 11 (4), 189-197
October-December, 2011
ISSN (print) 0975-6175
ISSN (online) 2229-6026

In-silico comparison of distal-less protein variation in insects

H. Tejmala[1], S. Brindha[1], G. Gurusubramanian[2] and N. Senthil Kumar[1*]

[1] *Bioinformatics Infrastructure Facility, Department of Biotechnology,*
[2] *Department of Zoology, Mizoram University, Aizawl 796 004, India*

Received 23 July 2011 | Revised 25 October 2011 | Accepted 31 October 2011

ABSTRACT

Distal-less (Dll) protein is the first genetic signal for limb formation to occur in the developing zygote in insects. The function seems to be different across various organisms, like the Dll in butterflies are not only involved in limb formation but also take part in eyespot and wing pattern formation. Hence, the study of the sequence variation of the Dll protein of different insects might help us in better understanding of its evolutionary divergence and in turn its function in different insects. The sequence of Dll protein were retrieved from the NCBI database and was used to study its relationships among other insects species using MEGA 4.0 and analysis of the physicochemical properties was done using a computational tool called PROT-PROP. The Dll protein in the insects showed variations (31-94% identity) in their sequences when BLAST was performed, but the homeobox domain exhibiting helix-turn-helix (HTH) was found to be conserved. Presence of motifs with identical amino acid sequence and presence of regions with poly-amino acids might be the reason for the differences in the role of Dll in different insects. In the phylogenetic tree, insects belonging to the same order were found to cluster together and exhibit genetic relatedness.

Key words: Distal-less; homeobox domain; motif; poly-amino acids phylogeny; PROT-PROP; signature sequence.

INTRODUCTION

A family of transcription factors of *Hox* gene are major regulators of animal development.[1] The homeobox domain which is now known to be well-conserved in many other animals, including vertebrates was first identified in a number of *Drosophila* homeotic and segmentation proteins.[2] The homeodomain-containing transcriptional regulators are encoded by *Hox* genes that operate differential genetic programs along the anterior-posterior axis of animal bodies.[3] This domain binds DNA through a helix-turn-helix (HTH) structure which is characterized by two alpha-helices makes intimate contacts with the DNA and are joined by a short turn.[4] This family of homeobox transcription factors known as *Distal-less* (*Dll*/*Dlx*) are known for being involved in limb formation in many insects and vertebrates.[5] Antennae, la-

Corresponding author: N. Senthil Kumar
Phone: +91-0389-2330859
E-mail: nskmzu@gmail.com

Science Vision © 2011 MIPOGRASS. All rights reserved

bium, legs and wings originate from imaginal discs (the small sacs of epithelial cells that grow and develop during the larval stage).[6] Dlx homeobox genes which is mammalian derived are homologs of the *Drosophila Distal-less* (*Dll*) gene. The *Dlx*/*Dll* gene family appears to be of ancient origin which play essential role in appendage development in all species in which it has been identified.[7] When *Dll* gene switches on at a certain stage during the growth of insect larva, it causes some of its cells to organize into legs. Stumps formation was observed when *Dll* was switched off.[8] During embryogenesis, Hox proteins are expressed largely in the abdominal segments, where they can suppress thoracic leg development.[9] Gain and loss of transcriptional activation and repression functions in Hox proteins can be a plausible mechanism to diversify morphology during animal evolution.[10]

Proteins were the first molecular sequences to be used for phylogenetic inference.[11] Proteins constantly change shape and form to perform their biological roles. Since amino acid residues play an important role in stability and function, they are likely to be evolutionarily conserved in a protein family.[12] The hydrophobic core is found to contain residues important for stability.[13] Functional residues are found to be close together in protein-rotein interfaces.[14] In order to know the structure and function that are evolutionarily conserved, it is important to understand about the amino acids composition and physicochemical features of protein.[13]

Our present work aims to study the diversity of Dll protein of some insects and understand their genetic relatedness through phylogeny. Pattern of variation of the Dll protein among insects were studied in detail using PROT-PROP and various web-based tools and phylogenetic analysis was performed using MEGA 4.0.

MATERIALS AND METHODS

Sequence retrieval and local sequence alignment

The sequence of Dll protein was retrieved from the NCBI (http://www.ncbi.nlm.nih.gov) in FASTA format. BLAST[15] was performed for the *Bicyclus anynana* Dll protein sequence retrieved from NCBI to identify closely related sequences of Dll protein in different organisms. The Dll protein sequences for the 9 different insects included in the study are as follows: AF404825_1 [*B. anynana*]; ABD97849.1| [*Tetranychus urticae*]; AAB24059.1| [*Drosophila melanogaster*]; NP_001124509.1| [*Apis mellifera*]; AAG39634.1|AF317551_1 [*Tribolium castaneum*]; BAE78537.1| [*Harmonia axyridis*]; AF404110_1 homeotic transcription factor dll [*Junonia coenia*]; BAG06741.1| homeotic protein [*Athalia rosae*]; XP_002423032.1| Homeotic protein, putative [*Pediculus humanus corporis*]. Multiple sequence alignment of all these sequences was carried out using ClustalW from EMBL-EBI to find out the sequence similarity and conserved regions.

Phylogenetic analysis

Phylogenetic analysis of Dll protein sequence through neighbour-joining (NJ) with bootstrapping was carried out using Mega 4.0 software, prior to which the multiple sequence alignment of the sequences were also analyzed using ClustalX from MEGA 4.0.[16] Phylogenetic tree was constructed by the software showing the relationship among the sequences. The tree gives different clusters showing their relationship with each other. The sequences which lie in the same cluster are closely related. Bootstrap support was estimated using 1000 replicates.[17]

Analysis of physicochemical properties of the protein sequences

Dll sequences were analysed with PROT-

Science Vision © 2011 MIPOGRASS. All rights reserved

Figure 1. Arthropods used in the study. 1: *Bicyclus anynana* 2: *Junonia coenia* (Lepidoptera), 3: *Tetranychus urticae* (Arachnida), 4: *Athalia rosae* (Hymenoptera), 5: *Tribolium casteneum* 6: *Harmonia axyridis* (Coleoptera), 7: *Pediculus humanus corporis* (Phthiraptera), 8: *Drosophila melanogaster* (Diptera), 9: *Apis mellifera* (Hymenoptera).

PROP that characterizes physico-chemical properties of a protein in a single-window application. Other significant features of this software includes finding the subcellular location (intra or extra) of a protein, calculation of numerical values for hydrophobicity, hydrophilicity, composition of small and large amino acids, net hydrophobic content in terms of low/high, and Navie's algorithm to calculate theoretical pI.[18] Web-based tools like PSIPRED[19], Scansite[20] were used for secondary structure and motif analysis respectively.

RESULTS AND DISCUSSION

Sequence retrieval

B. anynana DII protein sequence was retrieved from the NCBI in FASTA format.

The sequence of the protein is the following:

```
B.anynana           -----------MTTQELDH---QHHHLGGSQTPHDISNSTNSTPTNVSSKSAFIELQQH 45
J.coenia            -----------MTTQELDH---QHHHLGGSQTPHDISNSTNSTPTNVSSKSAFIELQQH 45
D.melanogaster      -----------MDAPDAPH---TPKYMDGGNTAASVTPGIN-----IPGKSAFVELQQH 40
T.castaneum         -----------MSG---------EAHIG-PPTPHESNTSTP------VSKSAFIELQQH 32
H.axyridis          -----------MSS---------EGHLG-PPTPHESNNSTP------VSKSAFIELQQH 32
A.mellifera         -----------MEQHLHHT---GLGSVTPGPPGNGGDSTTSSTPVSQSGKSAFIELQHN 45
A.rosae             -----------MEQHLHHA---GLGSVTPGPAGNGGDS-ASSTPVSQSGKSAFIELQHN 44
P.humanus           MWSNKEFVVVVVVTGPQIDMSLLKGSPFGSGTGNNNGGSNTPTPTPIPLNSNDNLDHHHS 60
T.urticae           ------------------------------------------------------------

B.anynana           ------------GYGPFKGGYQHPHHFGSPGGQQNP-------HEASGFPSPR-SLG-- 82
J.coenia            ------------GYGPFKGGYQHPHHFGSPGGQQNP-------HEASGFPSPR-SLG-- 82
D.melanogaster      ------------AAAGYGGIRSTYQHFGPQGGQD-----------SGFPSPRSALG-- 73
T.castaneum         ------------GYGPLR--TSYQHHFNSPAGNAHTGPTG---THDAGFPSPRGALGA- 73
H.axyridis          ------------GYGPLR--TSYQHHFNSPAANAHGGPGG---GHDGGFPSPRSALSA- 73
A.mellifera         NLYNPASLRGGYPGGHNVPGAHQFSHQVGALSHQTSGPGSGNQQHADAGFPSPR-SLAG- 103
A.rosae             NLYNPASLRGGYPGGHNVPGSHQFPHQVGSLGHQGSGPGSGNQQHADSGFPSPRSALAG- 103
P.humanus           THGHHLNPPPTPHNNESLSNPNTPVPPSNTPGGLNSLNKSAFIELQQQYNPAIRAAYGHF 120
T.urticae           ------------------------------------------------------------

B.anynana           -----------YPFPPMHQNTY-GYHIGSYAPQCASPPKDEK-----------CGLSDD 118
J.coenia            -----------YPFPPMHQNTY-GYHLGSYAPQCASPPKDEK-----------CGLSDD 118
D.melanogaster      -----------YPFPPMHQNSYSGYHLGSYAPPCASPPKDDFSI----------SDKCED 112
T.castaneum         -----------YPFRPMHQNSYTGYHLGSYAANCPPSPKDDEKC----------LSLERP 112
H.axyridis          -----------YPFPPMHQNSYTGYHLGSYAANCPPSPKDDDKC----------LSLERP 112
A.mellifera         -----------YPFAPMHQHNAYGYHLGSYAPQCPSPPKDEKCG----------GSLVDE 142
A.rosae             -----------YPFAPMHQHNTYGYHLGSYAPQCPSPPKDGEYPNYLSPLGDKSGSLVDE 152
P.humanus           NHHQTSPGSGGYPFPPMHQNSYSSYHLGGYGPQS—PPK ESELDK---------IGVEDG 169
T.urticae           ------------------HEASPHFQPGPTPPRDEKPTLEEISR---------------- 26
                                     :.    :: *   .    .. :
B.anynana           PGLR-----VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 173
J.coenia            PGLR-----VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 173
D.melanogaster      SGLR-----VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 167
T.castaneum         SG--------GGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 164
H.axyridis          SG--------GGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 164
A.mellifera         LGGGGGGSLRNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 202
A.rosae             LGGGGGGSLRNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 212
P.humanus           PLR------VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 223
T.urticae           ---------VNGKNKKNWKPRTIYSSIQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 77
                          .**.**  ********:**********************************

B.anynana           VKIWFQNRRSKYKKMMKAAQVG----AVPPGLGLPPGSPPNNNQLLHGGGGSSSGSQHSP 229
J.coenia            VKIWFQNRRSKYKKMMKAAQVG----GAPPGLG-PPASPPNNNQLLHGGGGSSSGSQHSP 228
D.melanogaster      VKIWFQNRRSKYKKMMKAAQGPGTNSGMPLGGGGPNPGQHSPNQMHSGGNNGGGSNSGSP 227
T.castaneum         VKIWFQNRRSKYKKMMKAAQVS---------GGNNNNGTPGGHSGFWGQLGAAVDESRTP 215
H.axyridis          VKIWFQNRRSKYKKMMKAAQVP---------GGTNNNGTPGGHAGLLGSN--ANLPSSSP 213
A.mellifera         VKIWFQNRRSKYKKMMKAAQQG----------GGGGGGQHG--SLLAGGTALPGGPSPQP 250
A.rosae             VKIWFQNRRSKYKKMMKAAQQG----------GGGGGGQHG--SLLAGGTALPGGPSPQP 260
P.humanus           VKIWFQNRRSKYKKMMKAAQNTGPGGQNNNGGGNGSPAPPGGATVGILGGNNSTGSQNSP 283
T.urticae           VKIWLQNKRSKNKKMQKAQEAVN--------GGGQVNGSQGVACGTGGGRRGRGGNQGQG 129
                    ****:**:*** *** ** :            *    .  .             .

B.anynana           SGYAGGPAQAHSPTPSSTPVSELSPGLSPTATPWDVKQPPQPSWDVKVGYPTAGRSPDGT 289
J.coenia            SAYQSGPTQAHSPTPSSTPVSELSPGLSPTGTPWDVKQPPQPAWDVKVGYPTAGRSPDGT 288
D.melanogaster      SHYLP---PGHSPTPSSTPVSELSPEFPPTG----LSPPTQAPWDQKP----------H 269
T.castaneum         RPKHDGRGRIVERFTSDRVHAALVT--SPRA-----RPLARTCRDNRLLVLAG------- 261
H.axyridis          GPQGQG-----------------------------------MTQGKRL------------ 226
A.mellifera         GQPGSLMQGGGGSSVSGSPTTGYLGGMGGGA---GGHTPGSSSPGSEMSPQHN------- 300
A.rosae             GQPGSLMQGGGGSSVSGSPTTGYLGGMGGGG---GGHTPGSSSPGSEMSPQHN------- 310
P.humanus           -NYGHHNQNQTSPSPSSTPVSDMSPHGLSGS-------PPTMNWDMKPNINNLG------ 329
T.urticae           QNQSQQQQQAQQQAQHQQQQQQQQQHQQQQH-----QQHQQQQQQAQVQQTLSN------ 178
```


Science Vision © 2011 MIPOGRASS. All rights reserved

```
B.anynana           SCDVKPPHQQSWDPRVGYGAPPGPWDMKGAHAHALHHQ--GAPQPHPPYVPQYSWYQADA 347
J.coenia            S-DVKPPHQQSWDPRVGYGAPPGPWDMKGAD-HALHHQ--GAPQPHPPYVPQYSWYQADA 344
D.melanogaster      WIDHKPPPQMTPQP---------PHPAATLHPQTHHHN--PPPQ-MGGYVPQY-WYLPET 316
T.castaneum         --TSSPSNRFTRHLIQ-------RRTHTRHTTTSHNTR--GTPRTTPASSRCGQLFNTEE 310
H.axyridis          ------------------------------------------------------------
A.mellifera         --ESPPAPSWPGEMKHHPHPHAPPHTHHHPHTPGHHPPPPPPPPPHHAGYMPQYSWYQADP 358
A.rosae             --ESPPAPVWPGEMKHHPHPHGPPHTHHHPHTPGHHPPPPPPPPPHHAGYMPQYSWYQADP 368
P.humanus           ---VTPTHHTTGHP--------------HHTPTHHHP------THHSYMPQYSWYNADT 365
T.urticae           --QQQSLDTQPGVGLSSGSPFIKGEGYIPQHSPEVPSESHTPLHSSLGPNPGSNGVNSNG 236

B.anynana           -NPGLLTVWPAV---------------------- 358
J.coenia            -NPGLLTENGLALFHRMGLRSQFLGSNEQNNYDS 377
D.melanogaster      -NPSLVTVWPAV---------------------- 327
T.castaneum         -SD------------------------------- 312
H.axyridis          ---------------------------------
A.mellifera         -NPGLLT--------------------------- 364
A.rosae             -NPGLLTVWPAV---------------------- 379
P.humanus           ANQPLLTVWPAV---------------------- 377
T.urticae           PGAPVNSIGNAT---------------------- 248
```

Figure 2. CLUSTAL 2.1 multiple sequence alignment of Dll protein sequences of nine insects and Poly-G regions.

```
(1) KPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRRSKYKKM
                HHHHHHHHHHHHHHCCCCCHHHHHHHHHH
(2) KPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQVKIWLQNKRSKNKKM
```

Figure 3. The amino acid differences in the homeobox region of other insects (1) and *Tetranychus urticae* (2). The helix-turn-helix region is also specified.

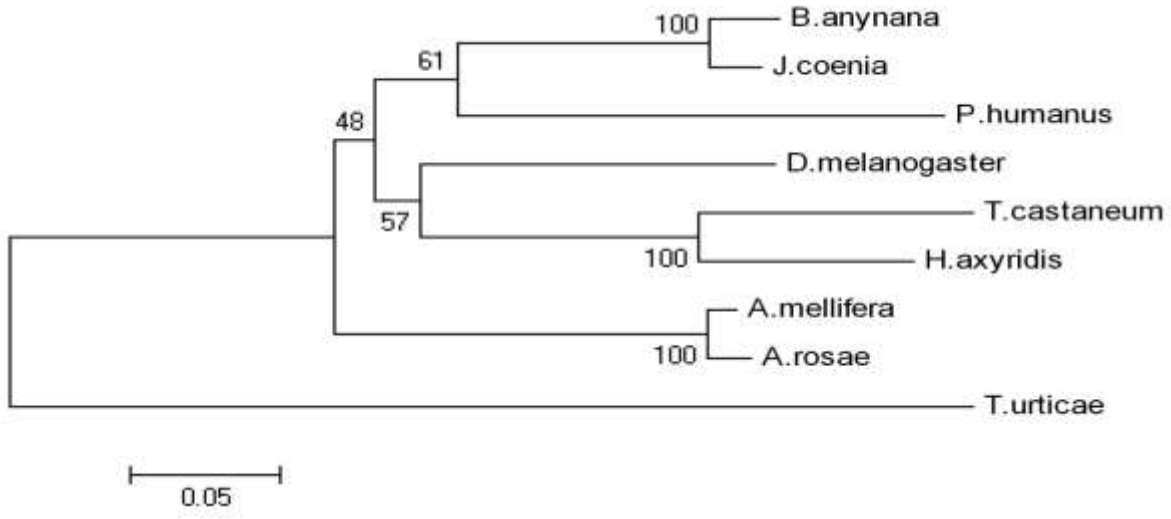

Figure 4. Phylogenetic tree built using NJ method of MEGA 4.0.

Parameters	*B. anynana*	*J. coenia*	*D. melano gaster*	*T. casteneum*	*H. axyridis*	*A. mellifera*	*A. rosae*	*P. humanus*	*T. urticae*
SULPHUR amino acids	2% Cystine: 3 Methionine: 6	2% Cystine: 2 Methionine: 7	3% Cystine: 2 Methionine: 10	2% Cystine: 4 Methionine: 5	3% Cystine: 2 Methionine: 6	3% Cystine: 2 Methionine: 10	2% Cystine: 1 Methionine: 10	2% Cystine: 0 Methionine: 10	0% Cystine: 1 Methionine: 1
BASIC Amino Acids	13% Lysine: 17 Arginine:12 Histidine:21	13% Lysine: 17 Arginine:14 Histidine:21	12% Lysine:15 Arginine:11 Histidine:16	19% Lysine: 13 Arginine:31 Histidine:16	15% Lysine: 13 Arginine:12 Histidine:10	14% Lysine: 13 Arginine:11 Histidine:28	13% Lysine: 13 Arginine:11 Histidine:28	13% Lysine: 15 Arginine10H istidine:25	12% Lysine: 12 Arginine11 Histidine: 9
ACIDIC Amino Acids	17% Aspartic Acid:12 Glutamic Acid: 7 Asparagine: 12 Glutamine: 30	19% Aspartic Acid: 14 Glutamic Acid: 9 Asparagine: 16 Glutamine: 33	16% Aspartic Acid: 10 Glutamic Acid: 7 Asparagine: 14 Glutamine: 24	16% Aspartic Acid: 8 Glutamic Acid: 11 Asparagine: 15 Glutamine: 16	17% Aspartic Acid: 4 Glutamic Acid: 6 Asparagine: 13 Glutamine: 16	14% Aspartic Acid: 5 Glutamic Acid: 9 Asparagine: 12 Glutamine: 27	14% Aspartic Acid: 6 Glutamic Acid: 9 Asparagine: 13 Glutamine: 27	20% Aspartic Acid: 8 Glutamic Acid: 8 Asparagine: 39 Glutamine: 22	33% Aspartic Acid: 2 Glutamic Acid: 10 Asparagine: 17 Glutamine: 55
Amino acid residues	A 24, R 12, N 12, D 12, C 3, E 7, Q 30, G 40, H 9, I 21, L 5, K 23, M 17, F 6, P 7, S 49, T 36, W 18, Y 7, V 15, 14	A 23, R 14, N 16, D 14, C 2, E 9, Q 33, G 42, H 21, I 4, L 27, K 17, M 7, F 9, P 48, S 39, T 19, W 6, Y 16, V 11	A 21, R 11, N 14, D 10, C 2, E 7, Q 24, G 36, H 16, I 7, L 19, K 15, M 10, F 8, P 50, S 31, T 18, W 5, Y 14, V 9	A 23, R 31, N 15, D 8, C 4, E 11, Q 16, G 30, H 16, I 6, L 22, K 13, M 5, F 10, P 25, S 29, T 29, W 2, Y 9, V 8	A 16, R 12, N 13, D 4, C 2, E 6, Q 16, G 29, H 10, I 3, L 19, K 13, M 6, F 6, P 23, S 24, T 11, W 1, Y 9, V 3	A 23, R 11, N 12, D 5, C 2, E 9, Q 27, G 64, H 28, I 3, L 26, K 13, M 10, F 6, P 48, S 37, T 17, W 3, Y 13, V 7	A 22, R 11, N 13, D 6, C 1, E 9, Q 27, G 69, H 28, I 3, L 28, K 13, M 10, F 6, P 51, S 39, T 14, W 4, Y 15, V 10	A 15, R 10, N 39, D 8, C 0, E 8, Q 22, G 40, H 25, I 9, L 25, K 15, M 10, F 7, P 45, S 37, T 29, W 5, Y 13, V 15	A 13, R 11, N 17, D 2, C 1, E 10, Q 55, G 28, H 9, I 7, L 14, K 12, M 1, F 3, P 18, S 21, T 11, W 2, Y 3, V 10

Figure 5. Types of amino acids and residues present in the Dll protein of arthropods using PROT-PROP software.

```
A.mellifera     LGGGGGGSLRNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 202
A.rosae         LGGGGGGSLRNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 212

A.mellifera     VKIWFQNRRSKYKKMMKAAQQG----------GGGGGGQHG--SLLAGGTALPGGPSPQP 250
A.rosae         VKIWFQNRRSKYKKMMKAAQQG----------GGGGGGQHG--SLLAGGTALPGGPSPQP 260

B.anynana       PGLR-----VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 173
J.coenia        PGLR-----VNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQ 173

B.anynana       VKIWFQNRRSKYKKMMKAAQVG----AVPPGLGLPPGSPPNNNQLLHGGGGSSSGSQHSP 229
J.coenia        VKIWFQNRRSKYKKMMKAAQVG----GAPPGLG-PPASPPNNNQLLHGGGGSSSGSQHSP 228
```

Figure 6. HTH conserved region is highlighted in black and occurrence of long stretches of poly- amino acids are highlighted in grey (*Apis mellifera* and *Athalia rosae*; *Bicyclus anynana* and *Junonia coenia*).

Science Vision © 2011 MIPOGRASS. All rights reserved

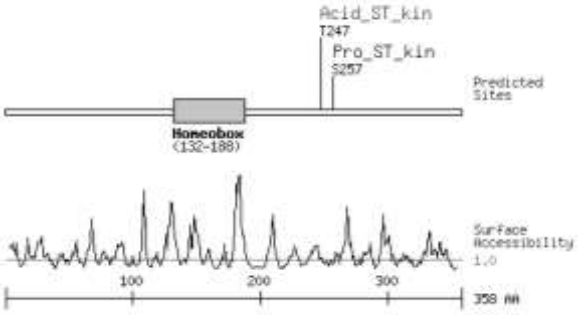

Figure 7. Motif scan graphic results of *B. anynana*.

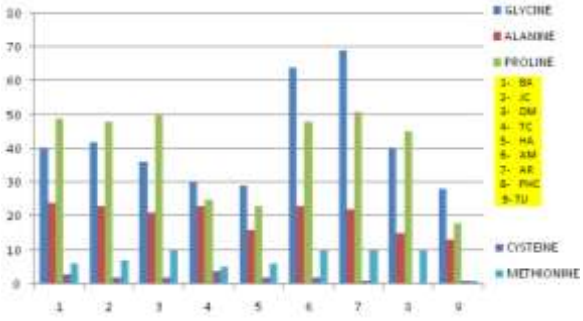

Figure 8. Physicochemical analysis of Distal-less protein using PROT-PROP software. 1: *B. anynana* & 2: *J. coenia* (Lepidoptera), 3: *D. melanogaster* (Diptera), 4: *T. urticae* (Arachnida), 5: *H. axyridis* (Coleoptera), 6: *A. mellifera* (Hymenoptera), 7: *A. rosae* (Hymenoptera) 8: *P. humanus corporis* (Phthiraptera), 9: *T. casteneum*.

>AF404825_1 distal-less [*B. anynana*]

MTTQELDHQHHHLGGSQTPHDISNSTNSTPTNVSSK-
SAFIELQQHGYGPFKGGYQHPHHFGSPGGQQNPHEASGFPSP
RSLGYPFPPMHQNTYGYHIGSYAPQCASPPKDEKCGLSDDPG
LRVNGKGKKMRKPRTIYSSLQLQQLNRRFQRTQYLALPERAEL
AASLGLTQTQVKIWFQNRRSKYKKMMKAAQVGAVPPGLGLP
PGSPPNNNQLLHGGGGSSSGSQHSPSGYAGGPAQAHSPTPS
STPVSELSPGLSPTATPWDVKQPPQPSWDVKVGYPTAGRSPD
GTSCDVKPPHQQSWDPRVGYGAPPGPWDMKGAHAHALHH
QGAPQPHPPYVPQYSWYQADANPGLLTVWPAV

BLAST was performed using the above sequence and after observing the result, Dll protein sequences of nine insects were selected for further analysis (Fig. 1). BLAST result shows ~31-94 % identity among the Dll sequences of various insects. A multiple sequence alignment using ClustalW was performed for all the 9 selected sequences. The conserved signature sequences which include the homeobox region have been highlighted in grey (Fig. 2). The region highlighted in dark grey is the HTH motif observed using PSIPRED. It is through this motif where Hox genes bind to DNA. However, slight differences were observed in the homeobox regions of *T. urticae* as compared to the sequences of the other insects (Fig. 3). From the motif scanner results obtained from Scansite, it was found that the homeobox domain of Dll consists of 57 amino acids with a HTH region in all the insects included in our study.

Phylogenetic analysis

The NJ method is frequently used because of its demonstrated accuracy for smaller data sets and its computational speed.[16] Phylogenetic analysis of Dll protein sequence through NJ with bootstrapping was carried out using Mega 4.0 software. It was found that species belonging to same order were clustered together (Fig. 4). The two lepidopteran species of butterflies *B. anynana* and *J. coenia* are clustered together. The dipteran, *D. melanogaster*, is placed at a different node. *T. castaneum* and *H. axyridis*, coleopterans are clustered together. The hymenopteran *A. mellifera* and *A. rosae* are also clustered together. *P. humanus corporis* a phthirapteran is found to lie at a different node. *T. urticae* an arachnid, order trombidiformes which is an outgroup is found to lie at an isolated node.

Physicochemical analysis of Distal-less protein using PROT-PROP

The percentage occurrence of each amino acid of the Dll sequence of each insect as well as the individual sequence details are clearly

generated by the PROT-PROP tool. Transcription factors are found to contain numerous repeats of serine, glycine, proline and alanine.[22] In our analysis, the same kind of amino acids are found to be quite prevalent (Fig. 5). Amino acid residues are likely to be evolutionarily conserved in a protein family because they play an important role in stability and function. Understanding about the amino acids composition and physicochemical features of protein are essential for knowing the structure and function that are evolutionarily conserved.[12]

Among insects falling in the same clade, sequences seems to exhibit replication slippage in which a codon becomes repeated numerous times and causes the occurrence of "coding tandem repeats" creating long stretches of the same amino acid. An example of such a repeat is the polyalanine sequence of the (ultrabithorax) Ubx protein of insects.[23] Occurrence of Poly-G is observed in all the Dll sequences of the insects studied (Fig. 2 & 6). Poly-N and Poly-S are also observed in some sequences. These repeats are capable of changing the function of a protein.[23]

B. anynana and *J. coenia* (lepidopteran) are observed to have glycine(G) and serine(S) poly-amino acids regions. Hence the similarity in the poly-amino acids regions might contribute to the similarity of the functions of *Dll* in these two Nymphalid butterflies and also in their morphological characters as *Dll* is involved in limb as well as eye-spot formation in both the butterflies. Since the poly-amino acids regions vary in different insects, it might imply the difference in the function of *Dll* seen in different insects. *Drosophila Dll* takes part in appendage and sense organ development[24] and in butterflies *Dll* reflect the potential for pattern formation in the wing.[25] However, it is also found that the poly-amino acids sometimes belong to region of a motif and is found to be conserved between two closely related species. This may also be the reason for the exhibition of similar function of a protein between two closely related species. From the Scansite result obtained, these motif regions were found to belong mostly to different kinase groups which are believed to be involved in signal transducing events (Fig. 7).

Binding is an important way the activity of transcription factors are regulated and the interaction and binding of transcription factors with one another can be affected by chains of amino acids. Polyglutamine can increase the rate of transcription in the genes that it regulates, while polyalanine can reduce it thereby displaying a situation of modulating activity between the two competing components.[26] In our analysis, poly-Q and poly-A regions were observed in the all the sequences, more importantly in the HTH region which might be playing an important role in the modulation of activity during transcription. By observation, comparatively total of poly-Q and poly-A is highest in *T. urticae* which clearly shows the fact that it is an outgroup, belonging to the class arachnida. This variation in total poly-Q + poly-A in the different Dll sequences might affect its interaction and binding with other transcription factors and causes functional and morphological differences. According to Fondon and Garner[27] repeat length variations in transcription factors correlate with the evolution of specific morphologies in dogs. This can be similar in insects as well.

Cuticle phenotypic variation is observed when *Dll* expression is altered.[28] Hence, *Dll* might be related with the variability of cuticle phenotypes. Cysteine and methionine are found to be rare in insect cuticular proteins[29] and alanine, glycine and proline are found to be in high content.[30] This was also observed in the insects under study from PROT-PROP analysis (Fig. 8). Intensive study of the physicochemical properties of proteins and their interactions with other homeodomain might be helpful for future protein-protein interactive studies.

ACKNOWLEDGEMENT

The authors thank the DBT, New Delhi

Science Vision © 2011 MIPOGRASS. All rights reserved

for the Bioinformatics Infrastructure Facility and Traineeship for the first author, HT.

REFERENCES

1. Myers P (2008). Hox genes in development: The Hox code. Nature Education 1(1), http://scienceblogs.com/pharyngula/2007/09/the_hox_code.php.
2. McGinnis W, Levine MS, Hafen E, Kuroiwa A & Gehring WJ (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila antennapedia* and *bithorax* complexes. *Nature,* 308, 428-33.
3. Alonso CR (2002). Hox proteins: sculpting body parts by activating localized cell death. *Curr Biol,* 12, 22.
4. Schofield PN (1987). Patterns, puzzles and paradigms - The riddle of the homeobox. *Trends Neurosci,* 10, 3-6.
5. Casci T (2002). Developmental biology: *Distal-less* makes the leap. *Nat Rev Genet*, 218, 554-562.
6. Cohen B, Simcox AA & Cohen SM (1993). Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development,* 117, 597-608.
7. Robledo RF et al. (2002). The Dl*x5* and *Dlx6* homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev,* 16, 1089-1101.
8. Zimmer C (1998). Hidden Unity. Discover Magazine 19(1), http://www.discover. com/cover_story/9801-3.html.
9. Carroll SB, Grenier JK & Weatherbee SD (2001). *From DNA to Diversity*. Blackwell Science, London.
10. Li X & McGinnis W (1999). Activity regulation of Hox proteins, a mechanism for altering functional specificity in development and evolution. *Proc Natl Acad Sci USA,* 96, 6802-6807.
11. Fitch & Margoliash (1967). Construction of phylogenetic trees. *Science,* 155, 279-284.
12. Schueler-Furman O & Baker D (2003). Conserved residue clustering and protein structure prediction. *Proteins*, 52, 225-235.
13. Shortle D (1992) Mutational studies of protein structures and their stabilities. *Q Rev Biophys*, 25, 205-250.
14. DeLano WL (2002). Unraveling hot spots in binding interfaces: Progress and challenges. *Curr Opin Struct Biol,* 12, 14-20.
15. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol*, 215, 403-410.
16. Tamura K, Dudley J, Nei M & Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol,* 24, 1596-1599.
17. Pairot P, Komgrit W & Peter HA (2011). Cryptic biodiversity and phylogenetic relationships revealed by DNA barcoding of Oriental black flies in the subgenus *Gomphostilbia*. *Genome,* 54, 1-9.
18. Brindha S, Senthil Kumar N & Gurusubramanian G (2011). PROT-PROP: J-Tool to classify and predict the subcellular location of proteins based on physiochemical characterization. *Comput Biol Chem* (Communicated).
19. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS & Jones DT (2005). Protein structure prediction servers at University College London. *Nucleic Acids Res*, 33, 36-38.
20. Obenauer JC, Cantley LC & Yaffe MB (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res,* 31, 3635-3641.
21. Tamura K, Nei M & Kumar S (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA,* 101, 11030-11035.
22. Caburet S, Cocquet J, Vaiman D & Veitia RA (2005). Coding repeats and evolutionary "agility". *BioEssays,* 27, 581-558.
23. Gilbert Scott F (2006). *Developmental Biology*, 8th edition, Chapter 23, pp. 739-741.
24. Panganiban G (2000). Distal-less function during *Drosophila* appendage and sense organ development. *Dev Dyn,* 218, 554-562.
25. Brakefield PM & French V (1999). Butterfly wings: the evolution of development of colour patterns. *BioEssays,* 21, 391-401.
26. Myers P (2007). Tandem repeats and morphological variation. *Nature Education* 1(1), http://scienceblogs.com/pharyngula/2007/10/tandem_repeats_and_morphologic.php.
27. Fondon JW & Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci USA,* 101, 18058-18063.
28. Shiga Y & Yasumoto R (2002). Evolving role of Antennapedia protein in arthropod limb patterning. *Development,* 129, 3555-3561.
29. Willis JH (1999). Cuticular proteins in insects and crustaceans. *Amer Zool*, 39, 600-609.
30. Andersen SO, Hojrup P & Roepstorff P (1995). Insect cuticular protein. *Insect Biochem Mol Biol*, 25, 153-176.