American Journal of Life Sciences
2015; 3(4): 268-273
Published online June 29, 2015 (http://www.sciencepublishinggroup.com/j/ajls)
doi: 10.11648/j.ajls.20150304.13
ISSN: 2328-5702 (Print); ISSN: 2328-5737 (Online)

Bioinformatics Analysis of the Structure and Function of CG17196 Protein of *Drosophila Melanogaster*

Hongchao Liu*, Xianming Zou, Yiming Wang, Xuanlong Du, Qian Wang, Junbao Xie, Xinming Tu

Medical College, Henan University of Science and Technology, Luoyang, Henan, China

Email address:
lhongchao@hotmail.com (Hongchao Liu)

To cite this article:
Hongchao Liu, Xianming Zou, Yiming Wang, Xuanlong Du, Qian Wang, Junbao Xie, Xinming Tu. Bioinformatics Analysis of the Structure and Function of CG17196 Protein of Drosophila Melanogaster. *American Journal of Life Sciences*. Vol. 3, No. 4, 2015, pp. 268-273.
doi: 10.11648/j.ajls.20150304.13

Abstract: Recent studies have suggested that chimeric genes may account for the formation and evolution of new genes and functional divergence. However, the biological function of the new chimeric gene *CG17196* of *Drosophila melanogaster* remains unknown, therefore, this study aims to analyze the structure and function of CG17196 protein using bioinformatics methods. Based on the amino acid sequence of CG17196 protein from NCBI database, the bioinformatics analyses were performed, including protein physical and chemical properties, transmembrane region, signal peptide, subcellular localization, domain, tertiary structure, and the phylogenetic tree of CG17196 related proteins from different species. The results showed that CG17196 protein was an unstable hydrophobic protein, performing biological function in the endoplasmic reticulum. It contained DHHC-type zinc finger domain and three transmembrane regions, but without signal peptide. The prediction result of gene ontology showed that the chance that the CG17196 protein actually had palmitoyltransferase activity was 70%. CG17196 protein and its related proteins in *Schizosaccharomyces pombe*, *Ashbya gossypii*, *Dictyostelium discoideum* and *Arabidopsis thaliana* showed high homology. In conclusion, CG17196 protein belongs to DHHC protein family and contains palmitoyltransferase activity, which may participate in the protein palmitoylation in the endoplasmic reticulum of *Drosophila melanogaster*, providing theoretical references for further systematic research on the function and evolution of new chimera *CG17196*.

Keywords: Chimeric Gene, New Genes, Palmitoylation, Bioinformatics, *Drosophila Melanogaster*

1. Introduction

The new genes are defined as those that are newly evolved and found in one or a few closely related species. Origination of new genes is a fundamental process underlying evolution of biological diversity [1,2]. The molecular mechanisms involved in the generation of new genes include gene duplication, exon shuffling, retroposition, lateral gene transfer, gene fusion/fission and de novo origination, among which gene duplication was firstly recognized [3-6]. Redundant copies generate through duplication and then accumulate various mutations due to the absence of selection pressure, eventually one copy can maintain the ancestral function, leaving the other copy free to develop a new gene with novel function. Studies confirm that gene duplication has been regarded as the major source for genetic novelties [5,7]. However, recent studies have suggested that chimeric genes formed through the fusion of pieces of different genes may also account for the formation and evolution of new genes and functional divergence [8,9]. Approximately 30% of the new genes in *Drosophila melanogaster* recruited a variety of genomic sequences (such as coding regions of other genes, transposons, simple tandem repeats, intronic or intergenic sequences, etc.) and generated new chimeric genes by different mechanisms, consequently producing proteins with new functions [5]. Rogers et al [8] have identified 14 chimeric genes that formed through DNA-level mutations within the *Drosophila melanogaster* genome, among which chimera *CG17196* with age between 5.4 and 12.8 Myr, formed through the tandem duplication and combination of portions of coding sequences of parental genes *CG17197* and *CG17195*. The new genes occur adaptive changes under Darwinian positive selection after the time of origin, usually appearing three alternative fates in the evolution: nonfunctionallization, subfunctionalization and neofunctionalization. The neofunctionalized and subfunctionalized new genes retain in the genome, while the genes with nonfunctionallization

gradually decay into pseudogenes [3,10,11]. Studies in *Drosophila* have shown that new genes can quickly become integrated into genetic networks and become essential for survival or fertility and play an important role in the adaptation and diversification of species [12-15]. Currently, the biological function and evolutionary fate of the new chimera *CG17196* remains unknown, therefore, this study aims to analyze the structure and function of *CG17196* encoding protein using bioinformatics methods, providing theoretical references for further systematic research on the function and evolution of new gene *CG17196*.

2. Materials and Methods

The amino acid sequence of *CG17196* encoding protein was obtained from NCBI database (NCBI reference sequence: NP_001262975.1). The physical and chemical properties were predicted based on ProtParam tool [16] (http://www.expasy.org/tools/protparam.html); To analyze the transmembrane helices in this protein, TMHMM Server v. 2.0 was used [17] (http://www.cbs.dtu.dk/services/TMHMM-2.0/); Signal peptide was predicted by online SignalP 4.1 Server [18] (http://www.cbs.dtu.dk/services/SignalP/) and subcellular localization was predicted using PSORT program [19] (http://www.psort.org/). The secondary and three-dimensional structures were predicted according to PredictProtein website [20] (http://www.predictprotein.org/) and SWISS-MODEL homology-modelling server [21,22] (http://swissmodel.expasy.org/) , respectively. The conserved domain (CD) and gene ontology (GO) annotation were predicted by CD-Search tool [23] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and Metastudent method [24] (http://www.predictprotein.org/). Amino acid sequence homology analysis was based on BLASTp program (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

3. Results

3.1. Physical and Chemical Properties

The protein encoded by *CG17196* was 253 aa in length and its molecular weight was 29.65 kDa with formula $C_{1370}H_{2054}N_{360}O_{336}S_{22}$. The total number of negatively charged residues (Asp + Glu) was 11, while total number of positively charged residues (Arg + Lys) was 27. The theoretical isoelectric point (pI) was 9.37. The estimated half-life was 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), >10 hours (Escherichia coli, in vivo), respectively. The instability index was computed to be 52.19, thus this classified the protein as unstable. The aliphatic index was 85.97 and grand average of hydropathicity (GRAVY) was 0.117, showing that CG17196 protein was hydrophobic protein.

3.2. Transmembrane Region Prediction

The analysis result of CG17196 protein by TMHMM Server demonstrated that two possible transmembrane helices from outside to inside (from position 31 to 53 and 168 to 190) and one possible transmembrane helix from inside to outside were found (from position 126 to 148), with 1aa-30aa, 149aa-167aa, locating outside the cell and 54aa-125aa, 191aa-253aa locating inside the membrane (Fig 1). The expected number of amino acids in transmembrane helices was 66.53348, which was larger than 18, thus CG17196 protein was a membrane protein that contained three membrane-spanning helices.

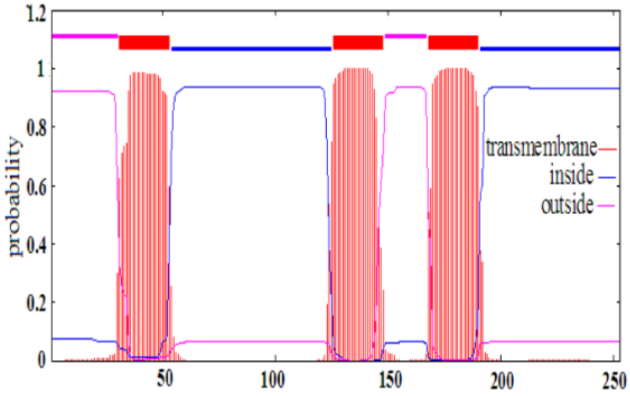

Fig. 1. *Transmembrane region prediction of CG17196 protein.*

3.3. Signal Peptide Prediction

SignalP 4.1 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. Based on a combination of several artificial neural networks, the SignalP results showed that the maximum S value (at position 55) was 0.115, mean S value (position 1-69) was 0.084. So the protein had no signal peptide (Fig 2).

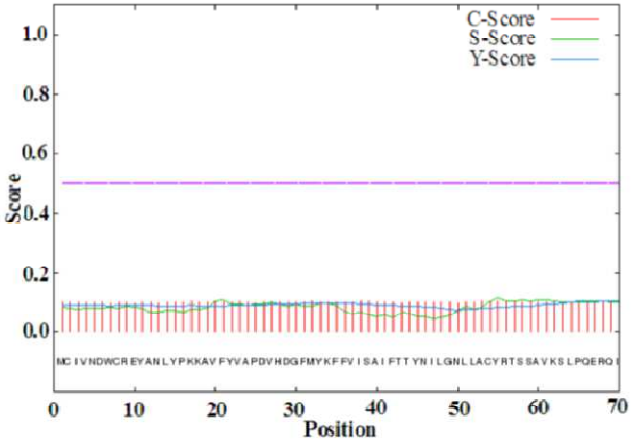

Fig. 2. *Signal peptide prediction of CG17196 protein.*

3.4. Subcellular Localization Prediction

PSORT II uses *k*-nearest neighbor (k-NN) algorithm for assessing the probability of localizing at each candidate sites. The peptide chain was predicted to be localized to the endoplasmic reticulum, cytoplasm and Golgi with the

possibility of 77.8%, 11.1%, 11.1%, respectively. The maximum possible location for CG17196 protein was in the endoplasmic reticulum (k=9).

3.5. Secondary Structure and Three-Dimensional Structure Prediction

Secondary structure is predicted by a system of neural networks with an expected average accuracy >72% for the three states of secondary structure: helix, extended strand and loop. The predicted secondary structure composition of CG17196 protein included helix (54.55%), extended strand (7.91%) and loop (37.55%). A total of 36.76% amino acid residues were exposed with more than 16% of their surface, suggesting CG17196 protein was hydrophobic, which was in accordance with the results of Prot Param.

SWISS-MODEL is a fully automated protein structure homology-modelling server. The result showed that the amino acid sequence identity between CG17196 protein (79-104)and template 2dkt.1.A was 30.77%, suggesting CG17196 protein may have the same structure and function with those of RING finger and CHY zinc finger domain-containing protein 1 (Solution structure of the CHY zinc finger domain of the RING finger and CHY zinc finger domain-containing protein 1 from Mus musculus. To be Published) (Fig 3). The global and per-residue model quality had been assessed using the QMEAN scoring function [25], and the results showed that the GMQE score and QMEAN4 score was 0.04 and -2.04, respectively.

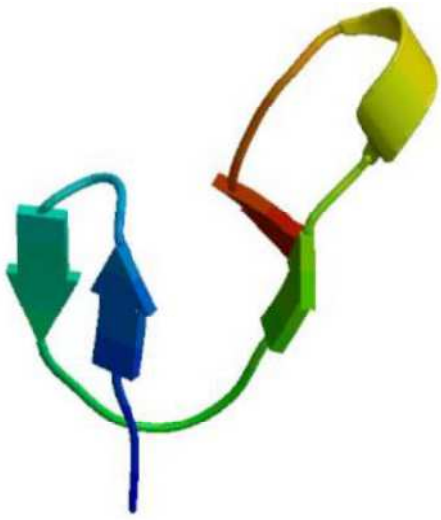

Fig. 3. *Three-dimensional structure prediction of CG17196 prorein.*

3.6. Conserved Domain Prediction and Gene Ontology Annotation

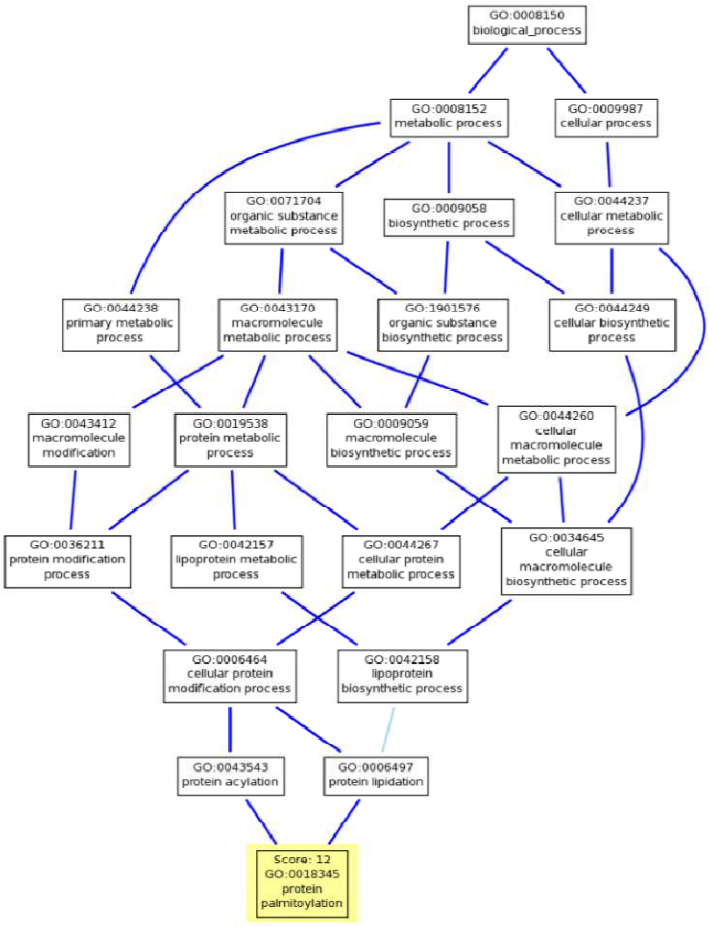

Fig. 4. *Biological process ontology prediction of CG17196 prorein.*

One conserved domain, DHHC-type zinc finger domain, was found from position 69 to 205 by CD-search with Bit score being 95.25 and E-value 5.29e-24. This domain is found in the DHHC proteins which are palmitoyltransferases [26]. Therefore, CG17196 protein may have palmitoyltransferase activity.

Metastudent predicts GO terms for protein sequences via homology to already annotated proteins. It first runs a PSI-BLAST query for a given target against a custom BLAST database containing sequences and GO terms of annotated proteins. If a similar sequence is found, the output is used by three base classifiers to calculate three separate sets of GO terms in different ways. The meta classifier finally takes all predictions from the base classifiers and combines them in an optimal way [24]. The tabular result of metastudent showed that the chance that the CG17196 protein actually had palmitoyltransferase activity (GO: 0016409) was 70% and that CG17196 protein participate in protein palmitoylation process (GO: 0018345) with a reliability of 12%. The subgraph shows where the predicted terms lie with respect to their inferred parent terms (Fig 4).

3.7. Amino Acid Sequence Homology Analysis

BLAST programs were the most powerful tools for nucleotide sequences and amino acid sequences similarity analysis. The amino acid sequence of CG17196 protein was submitted online to Blastp, then Swiss-Prot database was searched. The result showed that the amino acid sequence identity between CG17196 protein with those of palmitoyltransferase encoded by *Schizosaccharomyces pombe*, *Ashbya gossypii*, *Dictyostelium discoideum*, *Arabidopsis thaliana* were 52%, 52%, 50%, and 45%, respectively. The phylogenetic tree based on "neighbor Joining" method was showed in Fig. 5, illustrating they were most close in the evolutionary process, which showed high homology.

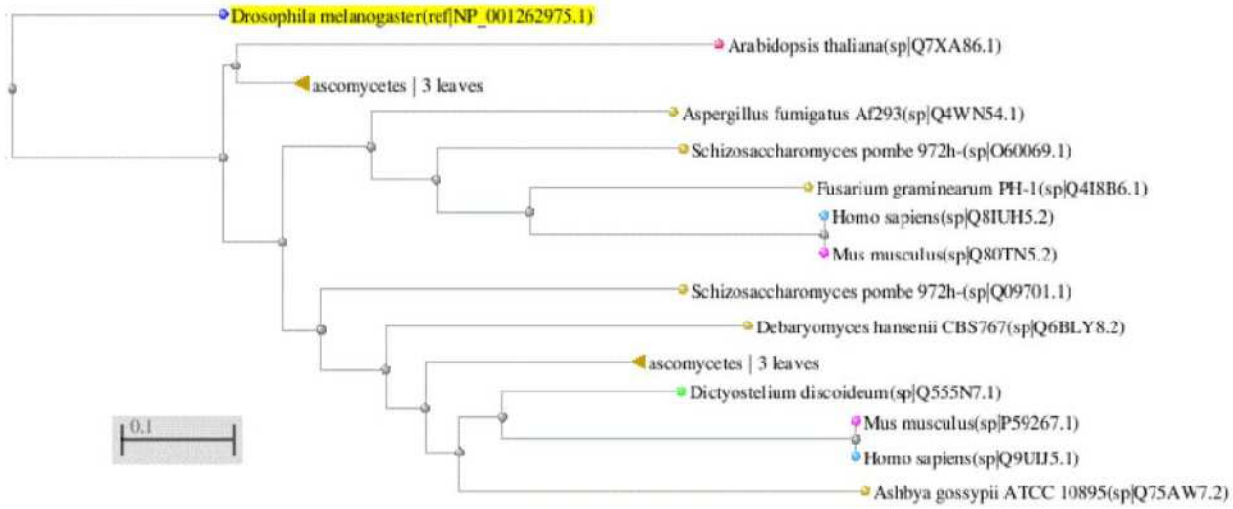

Fig. 5. *The phylogenetic tree based on amino acid sequence of CG17196 protein.*

4. Discussion

The new chimeric gene of *Drosophila melanogaster* appears to have formed through tandem duplication that did not respect gene boundaries. By DNA-based exon shuffling, the parts of two parental genes (*CG17197* and *CG17195*) were duplicated and transcribed together, forming a young chimeric gene (*CG17197*) [8]. It locates on chromosome 3R and its cytogenetic location is 96F2, bearing two exon and showing 950 bp region. Since all regulatory material is inherited from either the 5 or 3 parental gene, chimera expression patterns should reflect those of the parental genes, and the expression profile of *CG17196* appears in parallel with that of the 3 parental gene-*CG17195* [9]. The analysis result of high-throughput expression pattern data showed that *CG17196* had no expression or extremely low expression in embryo, early larva and adult female, while had moderately high expression in later larva, pupae and adult male, suggesting that *CG17196* may be expressed in a sex-specific manner and involve in the testes development [27,28]. However, the expressed tissues of this chimeric gene and its parental genes were unknown. Whether or not this new chimeric gene is functional remains uncertain.

Based on the amino acid sequence of *CG17196* encoding protein of *Drosophila melanogaster* from NCBI database, this study applied a variety of bioinformatics methods, including ProtParam, SignalP, CD-Search, SWISS-MODEL, ect., to analyze the structure and function of CG17196 protein. The results showed that the CG17196 protein was 253 aa in length with molecular weight 29.65 kDa. The theoretical pI was 9.37. The instability index was 52.19, which was larger than 40, classifying the protein as unstable. The grand average of hydropathicity was 0.117, showing that CG17196 protein was hydrophobic. CG17196 protein performed biological function in the endoplasmic reticulum, which contained three transmembrane regions, but without signal peptide. The main secondary structure of CG17196 protein included helix and

loop. The result of homology-modelling showed that CG17196 protein may have the same function with RING finger and CHY zinc finger domain-containing protein 1, from which template 2dkt.1.A derivated. One DHHC-type zinc finger domain, which is also known as NEW1 [29], was found in CG17196 protein. The DHHC zinc finger was first isolated in the *Drosophila* putative transcription factor DNZ1 and was named after a conserved sequence motif. This domain has palmitoyltransferase activity and this post-translational modification reversibly attach 16-carbon saturated fatty acids to specific cysteine residues in protein substrates through thioester linkages [30]. Palmitoylation will enhance the surface hydrophobicity and membrane affinity of protein substrates, and play important roles in modulating proteins trafficking, stability, and sorting, etc.. Protein palmitoylation has also been involved in numerous cellular processes, such as apoptosis, signaling and neuronal transmission [31-32]. The DHHC motif is found within a cysteine-rich domain which is thought to contain the catalytic site that acts as an enzyme, which adds a palmitoyl chemical group to proteins in order to anchor them to cell membranes. DHHC palmitoyltransferase family includes the well known DHHC zinc binding domain as well as three of the four conserved transmembrane regions found in this family of palmitoyltransferase enzymes. The result of metastudent showed that the chance that the CG17196 protein actually had palmitoyltransferase activity was 70%. Therefore, CG17196 protein may have palmitoyltransferase activity. The amino acid sequence of CG17196 protein showed high homology with those of palmitoyltransferase encoded by *Schizosaccharomyces pombe*, *Ashbya gossypii*, *Dictyostelium discoideum* and *Arabidopsis thaliana*, suggesting the conservation of function.

In summary, CG17196 protein of *Drosophila melanogaster* had DHHC-type zinc finger domain and showed high homology with palmitoyltransferase. It may have palmitoyltransferase activity and participate in the post-translational modification, which provided references for further research on the function and evolution of new chimera *CG17196*. Future studies of new genes, especially chimeric genes, and their functions will help to determine the role of genetic novelty in the adaptation and diversification of species.

References

[1] Zhang J, Dean AM, Brunet F, Long M. Evolving protein functional diversity in new genes of Drosophila. Proc Natl Acad Sci U S A, 2004, 101, 16246-16250.

[2] Long M, VanKuren NW, Chen S, Vibranovski MD. New gene evolution: little did we know. Annu Rev Genet, 2013, 47, 307-333.

[3] Assis R, Bachtrog D. Neofunctionalization of young duplicate genes in Drosophila. Proc Natl Acad Sci U S A, 2013, 110, 17409-17414.

[4] Arguello JR, Chen Y, Yang S, Wang W, Long M. Origination of an X-linked testes chimeric gene by illegitimate recombination in Drosophila. PLoS Genet, 2006, 2, e77.

[5] Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. On the origin of new genes in Drosophila. Genome Res, 2008, 18, 1446-1455.

[6] Reinhardt JA, Wanjiru BM, Brant AT, Saelao P, Begun DJ, Jones CD. De novo ORFs in Drosophila are important to organismal fitness and evolved rapidly from previously non-coding sequences. PLoS Genet, 2013, 9, e1003860.

[7] Zhou Q, Wang W. On the origin and evolution of new genes--a genomic and experimental perspective. J Genet Genomics, 2008, 35, 639-648.

[8] Rogers RL, Bedford T, Hartl DL. Formation and longevity of chimeric and duplicate genes in Drosophila melanogaster. Genetics, 2009, 181, 313-322.

[9] Rogers RL, Hartl DL. Chimeric genes as a source of rapid evolution in Drosophila melanogaster. Mol Biol Evol, 2012, 29, 517-529.

[10] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science, 2000, 290, 1151-1155.

[11] Santos ME, Athanasiadis A, Leitão AB, DuPasquier L, Sucena E. Alternative splicing and gene duplication in the evolution of the FoxP gene subfamily. Mol Biol Evol, 2011, 28, 237-247.

[12] Ding Y, Zhao L, Yang S, Jiang Y, Chen Y, Zhao R, Zhang Y, Zhang G, Dong Y, Yu H, Zhou Q, Wang W. A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. PLoS Genet, 2010, 6, e1001255.

[13] Chen S, Zhang YE, Long M. New genes in Drosophila quickly become essential. Science, 2010, 330, 1682-1685.

[14] Kemkemer C, Long M. New genes important for development. EMBO Reports, 2014, 15, 460-461.

[15] Ross BD, Rosin L, Thomae AW, Hiatt MA, Vermaak D, de la Cruz AF, Imhof A, Mellone BG, Malik HS. Stepwise evolution of essential centromere function in a Drosophila neogene. Science, 2013, 340, 1211-1214.

[16] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server, In: John M. Walker, ed., The Proteomics Protocols Handbook, Humana Press 2005, pp. 571-607.

[17] Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics, 2001, 17, 646-653.

[18] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods, 2011, 8, 785-786.

[19] Horton P, Nakai K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol, 1997, 5, 147-152.

[20] Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B. PredictProtein--an open resource for online prediction of protein structural and functional features. Nucleic Acids Res, 2014, 42, W337-343.

[21] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 2006, 22, 195-201.

[22] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res, 2014, 42, W252-258.

[23] Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res, 2011, 39, D225-229.

[24] Hamp T, Kassner R, Seemayer S, Vicedo E, Schaefer C, Achten D, Auer F, Boehm A, Braun T, Hecht M, Heron M, Hönigschmid P, Hopf TA, Kaufmann S, Kiening M, Krompass D, Landerer C, Mahlich Y, Roos M, Rost B. Homology-based inference sets the bar high for protein function prediction. BMC Bioinformatics, 2013, 14, S7.

[25] Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 2011, 27, 343-350.

[26] Fukata M, Fukata Y, Adesnik H, Nicoll RA, Bredt DS. Identification of PSD-95 palmitoylating enzymes. Neuron, 2004, 44, 987-996.

[27] Zhan ZB, Zhang Y, Zhao RP, Wang W. Evolutionary fate and expression patterns of chimeric new genes in Drosophila melanogaster. Dongwuxue Yanjiu, 2011, 32, 585-595. (in Chinese).

[28] Bannan BA, Van Etten J, Kohler JA, Tsoi Y, Hansen NM, Sigmon S, Fowler E, Buff H, Williams TS, Ault JG, Glaser RL, Korey CA. The Drosophila protein palmitoylome: characterizing palmitoyl-thioesterases and DHHC palmitoyl-transferases. Fly (Austin), 2008, 2, 198-214.

[29] Mesilaty-Gross S, Reich A, Motro B, Wides R. The Drosophila STAM gene homolog is in a tight gene cluster, and its expression correlates to that of the adjacent gene ial. Gene, 1999, 231, 173-186.

[30] Resh MD. Trafficking and signaling by fatty-acylated and prenylated proteins. Nat Chem Biol, 2006, 2, 584-590.

[31] Linder ME, Deschenes RJ. Palmitoylation: policing protein stability and traffic. Nat Rev Mol Cell Biol, 2007, 8, 74-84.

[32] Greaves J, Chamberlain LH. Palmitoylation-dependent protein sorting. J Cell Biol, 2007,176, 249-254.