SGID: a comprehensive and interactive database of the silkworm

Zhenglin Zhu[1*], Zhufen Guan[1], Gexin Liu[1], Yawang Wang[1,2], Ze Zhang[1*]

1. The School of Life Sciences, Chongqing University, Chongqing, China

2. Khoury College of Computer Sciences, Northeastern University, Seattle, WA, USA

*Corresponding authors

Zhenglin Zhu, The School of Life Sciences, Chongqing University, Chongqing, 400044 China

TEL: (86)23-6512-2686, FAX: (86)23-6512-2689, zhuzl@cqu.edu.cn

Ze Zhang, The School of Life Sciences, Chongqing University, Chongqing, 400044 China

TEL: (86)23-6512-2686, FAX: (86)23-6512-2689, zezhang@cqu.edu.cn

Keywords: silkworm, *bombyx mori*, informative, database

Abstract

Although the domestic silkworm (*Bombyx mori*) is an important model and economic animal, there is a lack of comprehensive database for this organism. Here, we developed the silkworm genome informatics database, SGID. It aims to bring together all silkworm related biological data and provide an interactive platform for gene inquiry and analysis. The function annotation in SGID is thorough and covers 98% of the silkworm genes. The annotation details include function description, gene ontology, KEGG, pathway, subcellular location, transmembrane topology, protein secondary/tertiary structure, homologous group and transcription factor. SGID provides genome scale visualization of population genetics test results based on high depth resequencing data of 158 silkworm samples. It also provides interactive analysis tools of transcriptomic and epigenomic data from 79 NCBI BioProjects. SGID is freely available at http://sgid.popgenetics.net. This database will be extremely useful to silkworm research in the future.

KEYWORDS: silkworm; *Bombyx mori*; informatics; database

Introduction

The silkworm, *Bombyx mori*, domesticated from its wild ancestry, *B. mandarina*, nearly 5000 years ago, contributes to silk industry, pest control (Gu et al., 2017; Li et al., 2015) and evolutionary biology. It is a promising model organism in life sciences (Meng et al., 2017). Because of its importance, the silkworm genome was sequenced and annotated in 2004 (Mita et al., 2004; Xia et al., 2004) and supplementarily annotated in 2012 (Shao et al., 2012). According to the records in NCBI PubMed, more than nine thousand silkworm related works have been published. The chromosome-level assembly of the silkworm genome is accomplished in 2017 and published in 2019 (Kawamoto et al.). Meanwhile, with the development of the sequencing technology, massive silkworm transcriptomic or epigenomic data were produced (Gu et al., 2019; Li et al., 2019a; Li et al., 2019b; Wu et al., 2019; Xiang et al., 2018). Up to now, there are already more than one thousand records of silkworm DNA-seq or RNA-seq data documented in NCBI SRA database. A comprehensive archiving and synthesis of these data is important to silkworm research.

Many model organisms have their own online bioinformatics analysis platform, such as TAIR for *Arabidopsis* (Poole, 2007), Flybase (Ashburner and Drysdale, 1994; Thurmond et al., 2018) for *Drosophila* and MGI (Bult et al., 2019; Smith et al., 2019) for mouse. Currently, SilkBase (Mita et al., 2003) is the only workable dedicated database for silkworm. It mainly focuses on the archive of sequences and is lack of analysis tools. Ensembl Silkworm (Kersey et al., 2018; Zerbino et al., 2018) and InsectBase (Yin et al., 2016) are absence of update and without workable whole genome browser. Until now, there is still not a comprehensive online analysis platform of the silkworm.

To assist silkworm research, we developed the silkworm genome informatics database, SGID, through collecting and cataloguing comprehensive genomics, transcriptomics, proteomics and epigenomics data. On the basis of previous works (Duan et al., 2010; Kawamoto et al., 2019; Mita et al., 2003; Wang et al., 2005), we thoroughly annotated silkworm genes in the contents of function, protein structure,

homolog and transcription factor. We also incorporated repeat elements, population statistic tests and epigenomic analysis results into the genome browser, which will help users to get a more comprehensive picture of genome segments. We developed interactive and click-one type analysis tools in SGID, letting users to obtain one or more genes' overall information swiftly.

Materials and methods

Data processing for basic gene annotation

We used the high quality assembly of the silkworm genome (Kawamoto et al., 2019) as the reference. Based on gene models (2017) in SilkBase, we re-annotated all silkworm genes by UniProt (Apweiler et al., 2004; Patient et al., 2008). Generally speaking, we BLAST the protein sequence of each gene against the UniProt protein database and took hits with significant similarity (*E-value* < 0.05, Coverage > 0.7). In this way, we made connections of the silkworm genes and UniProt proteins, and obtained UniProt annotations of the silkworm genes, including Pubmed ID, EMBL ID, Proteomes ID, Pfam (Bateman et al., 2000; Finn et al., 2016), Interpro (Mitchell et al., 2019; Mulder et al., 2002), Gene Ontologies (GO), KEGG (Kanehisa, 2002; Kanehisa et al., 2017) and so on.

To validate gene expression in protein level, we manually collected all sequenced silkworm peptides referred in published proteomics related works and did alignments of them and all silkworm genes. We filtered out results with cutoffs of *E-value* <0.05 and coverage > 0.8. If the predicted protein of one gene matches two or more peptides, we consider this gene has expression evidence.

We used CD-HIT (Li and Godzik, 2006) to search for homologous genes, requiring identify > 50% and coverage > 70%, and identified 1064 gene clusters. We did multiple alignments of the sequences within each cluster by MUSCLE (Edgar, 2004a; Edgar, 2004b), and built the phylogenetic tree by FastTree 2.1 (Price et al., 2010), with the parameter '-boot 5000' to test trees' likelihoods. We called repeat elements by RepeatMasker (Tarailo-Graovac and Chen, 2009; Tempel, 2012) and predicted transcription factors by the pipeline referred in AnimalTFDB 3.0 (Hu et al., 2015; Zhang et al., 2012; Zhang et al., 2019).

Subcellular localization and structure prediction

Annotations from UniProt only cover part of the silkworm proteins. Thus, we re-do protein function and structure prediction for all genes. We predicted subcellular locations of ASFV proteins through CELLO v2.5 (Yu et al., 2006; Yu et al., 2004) and transmembrane helixes within proteins' sequences using TMHMM 2.0 (Krogh et al., 2001). The output images were converted into PNG format for display in websites by Magick (www.imagemagick.org).

We BLAST all protein sequences against the PDB database (Berman et al., 2000; Burley et al., 2019) and extracted significant alignment results as we did for UniProt. We also did protein structure prediction by InterproScan (Jones et al., 2014; Zdobnov and Apweiler, 2001) and put significant results into SGID. We used SignalP (Almagro Armenteros et al., 2019; Nielsen, 2017) to predict signal peptides for silkworm proteins.

Gene Ontolgy

We put the GO (Gene Ontology) from InterproScan, UniProt and SilkBase together as SGID GO data set. We made alignments of the silkworm genes and KEGG silkworm proteins, and extracted the hits with *E-value* <0.05 and identity >0.9. In this way, we made connections between the silkworm genes and KEGG proteins. We also extracted KEGG pathway information (Aoki-Kinoshita and Kanehisa, 2007) from KEGG and made connections between pathways and silkworm genes using KEGG protein ID as a bridge. We obtain Entrez IDs of silkworm genes by KOBAS (Wu et al., 2006; Xie et al., 2011).

To enable a gene search using old gene models (Wang et al., 2005; Xia et al., 2004), we made connections between old gene models and SilkBase gene models 2017. Like we did for KEGG proteins, we made alignments of predicted proteins between old gene models and gene models 2017 and selected the best hit of each alignment.

Pre-processing of transcriptomic and epigenomic data

We collected transcriptomic and epigenomic data of silkworm related projects from NCBI. For transcriptomes, we classified them into three categories, 'DEG' (differentially expressed genes), 'Stage' and 'Tissue'. 'DEG' means the project is to identify differentially expressed genes in different experimental conditions. 'Stage' means the project is to observe gene expression at stages of different time points. Tissue means the project is to obtain expression profiling in different tissues. Following the standard RNA-seq analysis protocol (Ghosh and Chan, 2016; Pollier et al., 2013), we mapped transcriptomic reads onto the reference genome by bowtie (Langdon, 2015; Langmead and Salzberg, 2012) and called FKPM (Fragments per Kilobase Million) by cuffnorm (Wang et al., 2017).

For epigenomic data, according to experimental methodologies, we classified them into ChIP-Seq, Bisulfite-Seq and miRNA. ChIP-Seq stands for combining chromatin immunoprecipitation (ChIP) assays with next-generation sequencing. Bisulfite-Seq is the use of bisulfite treatment of DNA before routine sequencing to determine the pattern of methylation (Chatterjee et al., 2012). Small RNA means small RNA sequencing. For ChIP-Seq data, we used Bowtie to align reads onto the reference genome and inspected signatures by MAC2 (Liu, 2014; Zhang et al., 2008). We used Bismark (Krueger and Andrews, 2011) to pre-process Bisulfite-Seq data. We aligned small RNA reads onto the reference genome by Bowtie. Epigenomic analysis results are converted into bigWig format by the UCSC Genome tool bedGraphToBigWig for display in terminals.

Identification of domestication genes

We used Bowtie to map the genome resequencing data of 142 domesticated and 16 wild silkworm samples, including PRJDB4743, PRJNA402108 and an unpublished resequencing data (depth = 30x) of 15 silkworm samples produced by our lab, onto the reference genome and made bam files by SAMtools (Li et al., 2009). To illustrate the evolution pressure at a genome wide scale, we slid along the silkworm genome with a window size of 2000 bp and a step size of 200 bp. In each sliding window, we

calculated Pi, Theta (Watterson, 1975), Tajima's D (Tajima, 1989) and the composite likelihood ratio (CLR) (Nielsen et al., 2005) by ANGSD (Durvasula et al., 2016; Korneliussen et al., 2014) and SweapFineder2 (DeGiorgio et al., 2016). We also did the four population genetics test for each gene and made coalescent simulation (Pavlidis et al., 2010) ranking test (CSRT) (Zhu et al., 2007) according the silkworm domestication mode (Yang et al., 2014). We called domestication genes in a strict method, requiring a Tajima's $D_{domesticated} < -1$, a CSRT <0.05, a Tajima's $D_{domesticated} <$ Tajima's D_{wild}, a Tajima's $D_{min, domesticated}$ > top 5% point value in ascending order, a Fst_{max} > the top 5% point value in descending order and a $CLR_{max, domesticated}$ > the top 5% threshold in the whole genome. 'Min' or 'max' indicates the minimal or max value in the genic region extended by 10% of gene length, to accommodate the situation that the 5' or 3' terminals of a gene is under evolutionary forces. A subscript of 'domesticated' or 'wild' means the population genetics test is done on domesticated or wild silkworms. We also appended identified domestication genes in (Xia et al., 2009) and (Xiang et al., 2018) to our domestication genes dataset. We searched for genes possibly under balancing selection in the criteria that a Tajima's $D_{domesticated}$ > the top 5% point value in ascending ranking, a Tajima's $D_{wild} < 0.5$, a Tajima's $D_{domesticated} > 1$ and a CSRT < 0.95.

Genome browser and analysis tools

The genome browser of SGID is developed based on an open source population genetics visualization and analysis package SWAV (swav.popgenetics.org). We used MSAViewer (Yachdav et al., 2016) to show multiple alignments of homologous proteins, and phylotree (Shank et al., 2018) to display phylogenetic trees of gene clusters. The fuzzy text search in the home page is compatible with gene ID, gene name and gene function annotations. The alignment search tools in SGID are Perl codes to parse BLAT (Kent, 2002) or NCBI BLAST results (Johnson et al., 2008). The interface to exhibit gene expression is built upon D3 and JQuery. The overall web structure is Mysql + PHP + CodeIgniter (www.codeigniter.com) + JQuery (jquery.com).

Result and discussions

The biological data in SGID

SilkBase annotated 16880 genes in the high quality assembly of the silkworm genome (Kawamoto et al., 2019), but left 3329 without function descriptions. For these, SGID incorporated protein information from UniProt (Apweiler et al., 2004; Patient et al., 2008) and re-annotated the functions of 15594 genes, within which 2962 got function annotations for the first time. For a lot of genes, SGID gives not only simple descriptions, but also information on function details, chemical properties, related publications, protein structure, topologies, pathways and gene ontologies. In addition to the available gene ontology (GO) annotations of 9147 genes in SilkBase, SGID newly labeled GO IDs for 5521 genes. Besides, SGID made KEGG annotations for 16028 genes and Entrez IDs for 16320 genes. These are important for research, especially for gene set function enrichment analysis (Fig. 1).

Using peptide sequences from published experiments, we validated 2999 protein coding genes. They are of proteomics evidence. To depict one gene's function in a cell, SGID provides information on gene's subcellular localization and topology prediction. More than half (9592, 56.8%) of the silkworm genes are located in the nuclear, and 2878 genes (17.0%) have transmembrane regions (Fig. 2). Furthermore, 1960 silkworm genes are predicted to have signal peptides. Encouragingly, 9844 silkworm proteins are of PDB matches with *E-value* <0.05, which infers that more than half (58.3%) silkworm expressed proteins have structural information. External links to UniProt Proteomes, PRIDE (Perez-Riverol et al., 2019; Reisinger et al., 2015), Pfam (Bateman et al., 2000; Finn et al., 2016), Interpro (Jones et al., 2014; Zdobnov and Apweiler, 2001), SUPFAM (Pandit et al., 2004), Gene 3D (Dawson et al., 2017; Pearl et al., 2003; Pearl et al., 2002), Protein Modal Potal (Arnold et al., 2009; Haas et al., 2013) and PANTHER (Mi et al., 2017; Mi and Thomas, 2009) are also provided and they are helpful to understand the protein structure and related functions of one gene.

As a domesticated insect, the silkworm is important in evolution research. Totally, we

identified 569 domestication gene candidates. Users can view and inspect theses domestication genes by a SGID tool named "Population Genetics". Population genetics test results are also displayed in the genome bowser, where users can do sliding window analysis of interested genomic or genic segments. We also identified 81 genes possibly under balancing selection.

SGID includes transcriptomic data of 41 projects and epigenomic data of 38 projects. For transcriptomes, 28 are 'DEG', 9 are 'Stage' and 4 are 'Tissue' as we described in Materials and Methods. SGID includes 704 transcription factors (TF) belonging to 68 TF families. It also has 571401 repeat segments covering 27.5% of the silkworm genome, which is generally in accordance with previous records (Osanai-Futahashi et al., 2008). There are more retrotransposons (93%) than DNA transposons (7%). For retrotranposons, most are LINE (46%) and SINE (44%).

The genome browser in SGID

In SGID's genome browser page (Fig. 3), users can view the silkworm genes, repeat elements and population genetics test tracks subsequently. An input box and a list of buttons above the browser allow users to move, zoom in, zoom out, setting focus bar, generating figures or downloading the data of one track. A click onto a gene figure will take users to the gene detail page. Clicking on one point of some track will raise a dialog displaying the value at the point. Except for a genome browser, SGID also provides a browser to view epigenomic data. In the browser, users could view gene regulation signals at some specific genome position.

Retrieve genes' information from SGID

As a one-click type platform, SGID offers to search genes by a gene ID, a gene name, a gene function or even a brief description. In the page displaying search results, there are a list of gene information buttons within each result list. With the buttons, users can jump to to view gene details, a gene in genome browser, gene ontology and pathway, gene expression, regulation elements, gene structure and population genetics analysis results (Fig. S1A). In the detail page of each gene, aside from basic annotations (such as gene name, description, subcellular location and sequences, Fig.

S1B), six information groups are listed subsequently, including "Summary", "Ontologies", "Topology", "Population Genetics", "Multiple Alignment" and "Gene Tree". "Summary" mainly includes information resulted from protein sequence analysis (Fig. S1C). "Ontologies" displays a gene's annotation on GO, KEGG Function, KEGG Pathway, and PANTHER (Fig. S1C). In the part of "Topology", transmembrane regions are listed and marked in a diagram (Fig. S1F). If one gene's protein product is of signal peptide, the region of the signal peptide will also be listed and marked. "Population Genetics" listed 5 population genetic test results (Pi, Theta, Tajima's D, CLR and CSRT) and will give an interpretation about evolutionary forces. "Multiple Alignment" and "Gene Tree" displayed the multiple alignment of homologous genes at protein level and the phylogenetic tree produced based on the alignment.

To facilitate users to analyze a list of genes, SGID also offers to generate a list of gene information buttons through inputting a list of gene IDs. With the buttons, users can jump to some information view page directly like they do in search result page as referred above. Analogously, users can input a list of chromosome positions and obtain a list of genomic infomration links, with which users can view the genome browser or the epigenomics browser swiftly.

SGID analysis tools

To help users to visit data more quickly, we developed a list of analysis tools in SGID. As shown in the home page, "Gene Ontology" is a tool to retrieve GO, KEGG or Entrez numbers using a list of gene IDs. "Transcriptome" is a tool to view the expression of several genes in different experiment conditions, tissue or development stages. The results will be displayed in a heatmap figure. Stopping the mouse cursor at one cell of the heatmap will display the FKPM value of one gene at an experiment condition. The project's name is listed at the top right and users can click it to view the project's description. "Protein Structure", "TF", "Population Genetics", "Repeat Elements" and "Subcellular localization" are interactive search tools, with which users can obtain a group of genes or items with some similar biological properties. "Cluster" listed the 1064 gene clusters we identified. A summary of SGID search engines and analysis

tools is shown in Fig. 4.

KWMTBOMO08141, a case study

KWMTBOMO08141, BGIBMGA001085 in the previous annotation (Mita et al., 2004; Xia et al., 2004), is a domestication gene referred in (8), Bmor_03834. Through searching in the home page, we found this gene is of other 17 functional related members as transient receptor (Fig. S1A). Using the buttons listed below genes in the search result page, we can obtain genes' information one by one. In the detail page, we found this gene is of a full name "transient receptor potential-gamma protein" (Sanyal et al., 2004; Selbie et al., 1997; Woodard et al., 2007) and an alternative name "Transient receptor potential cation channel gamma". The gene is annotated to be located in plasma membrane (Fig. S1B) and function in interacting preferentially with trpl and to a lower extent with trp (Fig. S1C). Encouragingly, the protein product of this gene is validated by experiment peptides (Fig. S1B) and of significant similarity to a real protein structure 5Z96 recorded in PDB (Fig. S1D). In gene ontology, we obtained the GO and KEGG IDs of this gene and found KWMTBOMO08141 plays roles in the pathway of phototransduction in cell membrane (Fig. S1E). In topology, this gene has 6 transmembrane regions (Fig. S1F), which is in accordance to its subcellular localization prediction. As a domestication gene, KWMTBOMO08141 has low Tajima's D (-1.949945) and high CLR (629.851816). In the genome browser, we observed a CLR peak at the gene's region (Fig. 2) and a higher CLR peak at the right, indicating there may be genetic hitchhiking effects in this case (Fig. S1G). In transcription analysis, we found the expression of this gene is higher in brain than other tissues (Fig. S1H) and affected by ectopic expression of ecdysone oxidase (Fig. S1I). Through scanning this gene in the SGID epigenomics browser, we observed that epigenomic signals within the genic region disappear in some cell lines (Fig. S1J).

Conclusion

SGID is informative and user friendly. Under the idea of 'Click-one', SGID integrated different biological data and made them connective. SGID allows to search genes in fuzzy mode and to do analysis of more than one genes simultaneously. SGID

pre-analyzed available transcriptomic data and developed a search tool to view the expression of genes in different conditions. Similar such SGID tools made the initial bioinformatics analysis of silkworm projects more efficiently. With the advance of sequencing and experiments of the silkworm, more and more data will be incorporated into SGID, making the platform to be more and more powerful.

Data availability

All SGID data are publicly and freely accessible at http://sgid.popgenetics.net. Feedback on any aspect of the SGID database and discussions of silkworm gene annotations are welcome by email to zhuzl@cqu.edu.cn.

Author contributions

Z.L.Z. developed the web interface of the database. Z.Z.L., Z.G., G.L. and Y.W. collected and compiled the data and performed the analysis. Z.L.Z. and Z.Z. wrote the manuscript, conceived the idea and coordinated the project.

Funding

This work was supported by grants from the National Natural Science Foundation of China (31772524 to Z.Z., 31200941 to Z.L.Z.) and the Fundamental Research Funds for the Central Universities (106112016CDJXY290002).

Acknowledgements

We thank Dr. Yong Zhang for insightful suggestions and Dr. Anyuan Guo for help on transcription factor prediction. We also thank Dr. Quanyou Yu, Dr. Wei Sun and Mr. Yun Wang for useful discussions.

Figure Legends

Figure 1. An overview of the data in SGID.

Figure 2. The distribution of silkworm genes in different cellular subunits.

Figure 3. A snapshot of the genome browser of SGID with KWMTBOMO08141 in the center. Below are tracks of population genetics test results.

Figure 4. Search engines and analysis tools in SGID. Lines with arrows pointed out a general analysis flow in SGID.

References

Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology 37, 420-423.

Aoki-Kinoshita, K.F. and Kanehisa, M., 2007. Gene annotation and pathway mapping in KEGG. Methods in Molecular Biology 396, 71-91.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. and Yeh, L.S., 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Research 32, D115-9.

Arnold, K., Kiefer, F., Kopp, J., Battey, J.N., Podvinec, M., Westbrook, J.D., Berman, H.M., Bordoli, L. and Schwede, T., 2009. The Protein Model Portal. Journal of Structural and Functional Genomics 10, 1-8.

Ashburner, M. and Drysdale, R., 1994. FlyBase--the Drosophila genetic database. Development 120, 2077-9.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L., 2000. The Pfam protein families database. Nucleic Acids Research 28, 263-6.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Research 28, 235-42.

Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A. and Richardson, J.E., 2019. Mouse Genome Database (MGD) 2019. Nucleic Acids Research 47, D801-D806.

Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D.S., Green, R.K., Guranovic, V., Guzenko, D., Hudson, B.P., Kalro, T., Liang, Y., Lowe, R., Namkoong, H., Peisach, E., Periskova, I., Prlic, A., Randle, C., Rose, A., Rose, P., Sala, R., Sekharan, M., Shao, C., Tan, L., Tao, Y.P., Valasatava, Y., Voigt, M., Westbrook, J., Woo, J., Yang, H., Young, J., Zhuravleva, M. and Zardecki, C., 2019. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Research 47, D464-D474.

Chatterjee, A., Stockwell, P.A., Rodger, E.J. and Morison, I.M., 2012. Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Research 40, e79.

Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A. and Sillitoe, I., 2017. CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Research 45, D289-D295.

DeGiorgio, M., Huber, C.D., Hubisz, M.J., Hellmann, I. and Nielsen, R., 2016. SweepFinder2: increased sensitivity, robustness and flexibility. Bioinformatics 32, 1895-7.

Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z. and Xia, Q., 2010. SilkDB v2.0: a platform for silkworm (Bombyx mori) genome biology. Nucleic Acids Research 38, D453-6.

Durvasula, A., Hoffman, P.J., Kent, T.V., Liu, C., Kono, T.J., Morrell, P.L. and Ross-Ibarra, J., 2016. angsd-wrapper: utilities for analysing next-generation sequencing data. Molecular Ecology Resources 16, 1449-1454.

Edgar, R.C., 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space

complexity. BMC Bioinformatics 5, 113.

Edgar, R.C., 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-7.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 44, D279-85.

Ghosh, S. and Chan, C.K., 2016. Analysis of RNA-Seq Data Using TopHat and Cufflinks. Methods in Molecular Biology 1374, 339-61.

Gu, J., Li, Q., Chen, B., Xu, C., Zheng, H., Zhou, Y., Peng, Z., Hu, Z. and Wang, B., 2019. Species identification of Bombyx mori and Antheraea pernyi silk via immunology and proteomics. Scientific Reports 9, 9381.

Gu, Z., Li, F., Hu, J., Ding, C., Wang, C., Tian, J., Xue, B., Xu, K., Shen, W. and Li, B., 2017. Sublethal dose of phoxim and Bombyx mori nucleopolyhedrovirus interact to elevate silkworm mortality. Pest Management Science 73, 554-561.

Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L. and Schwede, T., 2013. The Protein Model Portal--a comprehensive resource for protein structure and model information. Database (Oxford) 2013, bat031.

Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q. and Guo, A.Y., 2015. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Research 47, D33-D38.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Research 36, W5-9.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236-40.

Kanehisa, M., 2002. The KEGG database. Novartis Foundation Symposia 247, 91-101; discussion 101-3, 119-28, 244-52.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45, D353-D361.

Kawamoto, M., Jouraku, A., Toyoda, A., Yokoi, K., Minakuchi, Y., Katsuma, S., Fujiyama, A., Kiuchi, T., Yamamoto, K. and Shimada, T., 2019. High-quality genome assembly of the silkworm, Bombyx mori. Insect Biochemistry and Molecular Biology 107, 53-62.

Kent, W.J., 2002. BLAT--the BLAST-like alignment tool. Genome Research 12, 656-64.

Kersey, P.J., Allen, J.E., Allot, A., Barba, M., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Grabmueller, C., Kumar, N., Liu, Z., Maurel, T., Moore, B., McDowall, M.D., Maheswari, U., Naamati, G., Newman, V., Ong, C.K., Paulini, M., Pedro, H., Perry, E., Russell, M., Sparrow, H., Tapanari, E., Taylor, K., Vullo, A., Williams, G., Zadissia, A., Olson, A., Stein, J., Wei, S., Tello-Ruiz, M., Ware, D., Luciani, A., Potter, S., Finn, R.D., Urban, M., Hammond-Kosack, K.E., Bolser, D.M., De Silva, N., Howe, K.L., Langridge, N., Maslen, G., Staines, D.M. and Yates, A., 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Research 46, D802-D808.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R., 2014. ANGSD: Analysis of Next Generation

Sequencing Data. BMC Bioinformatics 15, 356.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305, 567-80.

Krueger, F. and Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571-2.

Langdon, W.B., 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Mining 8, 1.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357-9.

Li, B., Wang, X., Li, Z., Lu, C., Zhang, Q., Chang, L., Li, W., Cheng, T., Xia, Q. and Zhao, P., 2019a. Transcriptome-wide analysis of N6-methyladenosine uncovers its regulatory role in gene expression in the lepidopteran Bombyx mori. Insect Molecular Biology.

Li, F., Ni, M., Zhang, H., Wang, B., Xu, K., Tian, J., Hu, J., Shen, W. and Li, B., 2015. Expression profile analysis of silkworm P450 family genes after phoxim induction. Pesticide Biochemistry Physiology 122, 103-9.

Li, G., Zhou, K., Zhao, G., Qian, H. and Xu, A., 2019b. Transcriptome-wide analysis of the difference of alternative splicing in susceptible and resistant silkworm strains after BmNPV infection. 3 Biotech 9, 152.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-9.

Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-9.

Liu, T., 2014. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Methods in Molecular Biology 1150, 81-95.

Meng, X., Zhu, F. and Chen, K., 2017. Silkworm: A Promising Model Organism in Life Science. Journal of Insect Science 17.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research 45, D183-D189.

Mi, H. and Thomas, P., 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods in Molecular Biology 563, 123-40.

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin, I.T., Abe, H., Shimada, T., Morishita, S. and Sasaki, T., 2004. The genome sequence of silkworm, Bombyx mori. DNA Research 11, 27-35.

Mita, K., Morimyo, M., Okano, K., Koike, Y., Nohata, J., Kawasaki, H., Kadono-Okuda, K., Yamamoto, K., Suzuki, M.G., Shimada, T., Goldsmith, M.R. and Maeda, S., 2003. The construction of an EST database for Bombyx mori and its application. Proceedings of the National Academy of Science of the United States of America 100, 14121-6.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani,

A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y. and Finn, R.D., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Research 47, D351-D360.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R., Courcelle, E., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Griffith-Jones, S., Haft, D., Hermjakob, H., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Orchard, S., Pagni, M., Peyruc, D., Ponting, C.P., Servant, F. and Sigrist, C.J., 2002. InterPro: an integrated documentation resource for protein families, domains and functional sites. Briefings in Bioinformatics 3, 225-35.

Nielsen, H., 2017. Predicting Secretory Proteins with SignalP. Methods in Molecular Biology 1611, 59-73.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. and Bustamante, C., 2005. Genomic scans for selective sweeps using SNP data. Genome Research 15, 1566-75.

Osanai-Futahashi, M., Suetsugu, Y., Mita, K. and Fujiwara, H., 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, Bombyx mori. Insect Biochemistry and Molecular Biology 38, 1046-57.

Pandit, S.B., Bhadra, R., Gowri, V.S., Balaji, S., Anand, B. and Srinivasan, N., 2004. SUPFAM: a database of sequence superfamilies of protein domains. BMC Bioinformatics 5, 28.

Patient, S., Wieser, D., Kleen, M., Kretschmann, E., Jesus Martin, M. and Apweiler, R., 2008. UniProtJAPI: a remote API for accessing UniProt data. Bioinformatics 24, 1321-2.

Pavlidis, P., Laurent, S. and Stephan, W., 2010. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. Molecular Ecology Resources 10, 723-7.

Pearl, F.M., Bennett, C.F., Bray, J.E., Harrison, A.P., Martin, N., Shepherd, A., Sillitoe, I., Thornton, J. and Orengo, C.A., 2003. The CATH database: an extended protein family resource for structural and functional genomics. Nucleic Acids Research 31, 452-5.

Pearl, F.M., Lee, D., Bray, J.E., Buchan, D.W., Shepherd, A.J. and Orengo, C.A., 2002. The CATH extended protein-family database: providing structural annotations for genome sequences. Protein Science 11, 233-44.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Perez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A. and Vizcaino, J.A., 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Research 47, D442-D450.

Pollier, J., Rombauts, S. and Goossens, A., 2013. Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. Methods in Molecular Biology 1011, 305-15.

Poole, R.L., 2007. The TAIR database.Methods in Molecular Biology 406, 179-212.

Price, M.N., Dehal, P.S. and Arkin, A.P., 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Reisinger, F., del-Toro, N., Ternent, T., Hermjakob, H. and Vizcaino, J.A., 2015. Introducing the PRIDE Archive RESTful web services. Nucleic Acids Research 43, W599-604.

Sanyal, S., Matthews, J., Bouton, D., Kim, H.J., Choi, H.S., Treuter, E. and Gustafsson, J.A., 2004. Deoxyribonucleic acid response element-dependent regulation of transcription by orphan nuclear receptor estrogen receptor-related receptor gamma. Molecular Endocrinology 18, 312-25.

Selbie, L.A., King, N.V., Dickenson, J.M. and Hill, S.J., 1997. Role of G-protein beta gamma subunits in the augmentation of P2Y2 (P2U)receptor-stimulated responses by neuropeptide Y Y1 Gi/o-coupled receptors. Biochemical Journal 328 (Pt 1), 153-8.

Shank, S.D., Weaver, S. and Kosakovsky Pond, S.L., 2018. phylotree.js - a JavaScript library for application development and interactive data visualization in phylogenetics. BMC Bioinformatics 19, 276.

Shao, W., Zhao, Q.Y., Wang, X.Y., Xu, X.Y., Tang, Q., Li, M., Li, X. and Xu, Y.Z., 2012. Alternative splicing and trans-splicing events revealed by analysis of the Bombyx mori transcriptome. RNA 18, 1395-407.

Smith, C.M., Hayamizu, T.F., Finger, J.H., Bello, S.M., McCright, I.J., Xu, J., Baldarelli, R.M., Beal, J.S., Campbell, J., Corbani, L.E., Frost, P.J., Lewis, J.R., Giannatto, S.C., Miers, D., Shaw, D.R., Kadin, J.A., Richardson, J.E., Smith, C.L. and Ringwald, M., 2019. The mouse Gene Expression Database (GXD): 2019 update. Nucleic Acids Research 47, D774-D779.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-95.

Tarailo-Graovac, M. and Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics Chapter 4, Unit 4 10.

Tempel, S., 2012. Using and understanding RepeatMasker. Methods in Molecular Biology 859, 29-51.

Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., Kaufman, T.C. and Calvi, B.R., 2018. FlyBase 2.0: the next generation. Nucleic Acids Research 47, D759-D765.

Wang, J., Xia, Q., He, X., Dai, M., Ruan, J., Chen, J., Yu, G., Yuan, H., Hu, Y., Li, R., Feng, T., Ye, C., Lu, C., Li, S., Wong, G.K., Yang, H., Xiang, Z., Zhou, Z. and Yu, J., 2005. SilkDB: a knowledgebase for silkworm biology and genomics. Nucleic Acids Research 33, D399-402.

Wang, Q., Zhang, Y., Guo, W., Liu, Y., Wei, H. and Yang, S., 2017. Transcription analysis of cochlear development in minipigs. Acta Otolaryngologica 137, 1166-1173.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7, 256-76.

Woodard, G.E., Sage, S.O. and Rosado, J.A., 2007. Transient receptor potential channels and intracellular signaling. International Review of Cytology 256, 35-67.

Wu, J., Mao, X., Cai, T., Luo, J. and Wei, L., 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Research 34, W720-4.

Wu, P., Shang, Q., Huang, H., Zhang, S., Zhong, J., Hou, Q. and Guo, X., 2019. Quantitative proteomics analysis provides insight into the biological role of Hsp90 in BmNPV infection in Bombyx mori. Jouenal of Proteomics 203, 103379.

Xia, Q., Guo, Y., Zhang, Z., Li, D., Xuan, Z., Li, Z., Dai, F., Li, Y., Cheng, D., Li, R., Cheng, T., Jiang, T., Becquet, C., Xu, X., Liu, C., Zha, X., Fan, W., Lin, Y., Shen, Y., Jiang, L., Jensen, J., Hellmann, I., Tang, S., Zhao, P., Xu, H., Yu, C., Zhang, G., Li, J., Cao, J., Liu, S., He, N., Zhou, Y., Liu, H., Zhao, J., Ye, C., Du, Z., Pan, G., Zhao, A., Shao, H., Zeng, W., Wu, P., Li, C., Pan, M., Yin, X., Wang, J., Zheng, H., Wang, W., Zhang, X., Li, S., Yang, H., Lu, C.,

Nielsen, R., Zhou, Z. and Xiang, Z., 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombyx). Science 326, 433-6.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., Shen, Y., Lan, X., Yuan, L., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen, W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R., Shi, J., Li, H., Su, J., Wang, X., Zhang, Z., Wu, Q., Li, J., Zhang, Q., Wei, N., Sun, H., Dong, L., Liu, D., Zhao, S., Zhao, X., Meng, Q., Lan, F., Huang, X., Li, Y., Fang, L., Li, D., Sun, Y., Yang, Z., Huang, Y., Xi, Y., Qi, Q., He, D., Huang, H., Zhang, X., Wang, Z., Li, W., Cao, Y., Yu, Y., Yu, H., Ye, J., Chen, H., Zhou, Y., Liu, B., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Wang, W., Ye, C., Wong, G.K. and Yang, H., 2004. A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science 306, 1937-40.

Xiang, H., Liu, X., Li, M., Zhu, Y., Wang, L., Cui, Y., Liu, L., Fang, G., Qian, H., Xu, A., Wang, W. and Zhan, S., 2018. The evolutionary road from wild moth to domestic silkworm. Nature Ecology & Evolution 2, 1268-1279.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.Y. and Wei, L., 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Research 39, W316-22.

Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., Lewis, S.E., Rost, B. and Goldberg, T., 2016. MSAViewer: interactive JavaScript visualization of multiple sequence alignments. Bioinformatics 32, 3501-3503.

Yang, S.Y., Han, M.J., Kang, L.F., Li, Z.W., Shen, Y.H. and Zhang, Z., 2014. Demographic history and gene flow during silkworm domestication. BMC Evolutionary Biology 14, 185.

Yin, C., Shen, G., Guo, D., Wang, S., Ma, X., Xiao, H., Liu, J., Zhang, Z., Liu, Y., Zhang, Y., Yu, K., Huang, S. and Li, F., 2016. InsectBase: a resource for insect genomes and transcriptomes. Nucleic Acids Research 44, D801-7.

Yu, C.S., Chen, Y.C., Lu, C.H. and Hwang, J.K., 2006. Prediction of protein subcellular localization. Proteins 64, 643-51.

Yu, C.S., Lin, C.J. and Hwang, J.K., 2004. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science 13, 1402-6.

Zdobnov, E.M. and Apweiler, R., 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847-8.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A. and Flicek, P., 2018. Ensembl 2018. Nucleic Acids Research 46, D754-D761.

Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H. and Guo, A.Y., 2012. AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Research 40, D144-9.

Zhang, H.M., Liu, T., Liu, C.J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y. and Guo, A.Y., 2019. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Research 43, D76-81.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biology 9, R137.

Zhu, Q., Zheng, X., Luo, J., Gaut, B.S. and Ge, S., 2007. Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice. Molecular Biology and Evolution 24, 875-88.

12752 lists of correllations between old and new gene models

RNA-seq, ChIP-seq, BS-seq, miRNA-seq and ncRNA-seq from 79 BioProjects

Sliding window analysis results (window=2000, step=200) based on 158 silkworm genomes, including Theta, Pi, Fst, Tajima's D and CLR

Genome assembly in chromosomal level

16880 gene models

31074 gene descriptions

2487 EC numbers

4937 gene feature annotations

201309 annotations by InterproScan

20378 subcellular localization annotations

329 biophysicochemical properties

1725442 correlated EMBL IDs

8730 distinctive gene ontology lists

138 KEGG pathways

17762 correlated Gene3D IDs

2445 cactivity information lists

11712 correlated KEGG Gene IDs

1743 lists of cofactor information

9844 significantly correlated PDB protein structures

704 transcription factors

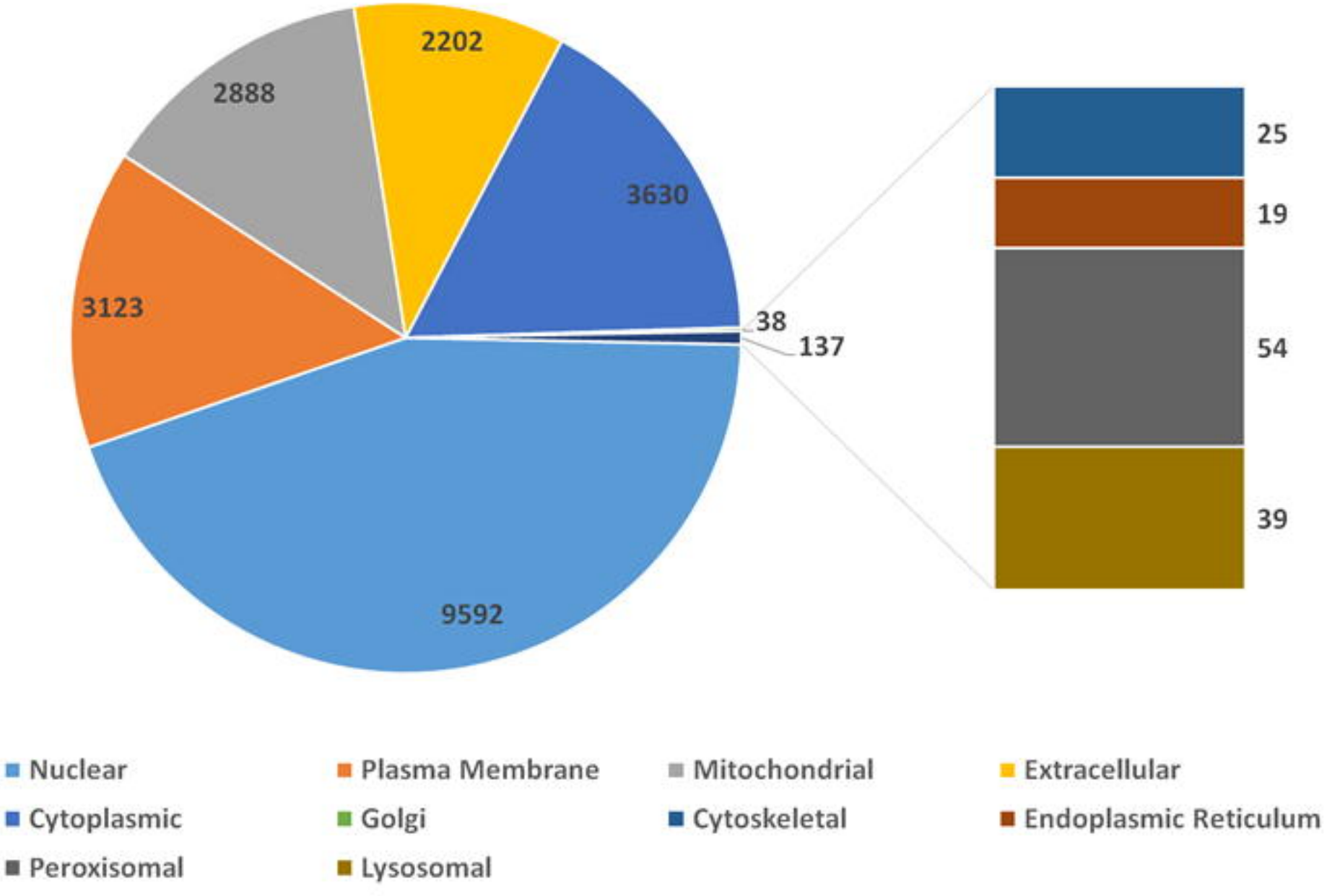
6257 CDD annotation lists

129394 real peptides from experiments

99183 correlated Interpro IDs

1960 signal peptides

10866 transmembrane predcitions


2202
2888
3630
3123
38
137
9592
25
19
54
39
Nuclear
Plasma Membrane
Mitochondrial
Extracellular
Cytoplasmic
Golgi
Cytoskeletal
Endoplasmic Reticulum
Peroxisomal
Lysosomal

Bomo_Chr13:17475200-17540002 GO

Location: 17475200-17540002

17,480,000 17,485,000 17,490,000 17,495,000 17,500,000 17,505,000 17,510,000 17,515,000 17,520,000 17,525,000 17,530,000 17,535,000 17,54

Gene Model

KWMTBOMO08138> KWMTBOMO08140> <KWMTBOMO08141

<KWMTBOMO08139

Repeat Elements

Theta 1,Domesticated 2,Wild

60 50 40 30 20 10 0

Bottom 5% for domesticated silkworms

Fst

0.6 0.5 0.4 0.3 0.2

Top-5% in the whole genome

Tajima's D 1,Domesticated 2,Wild

1 0 -2

Bottom 5% for domesticated silkworms

CLR

800 600 400 200

Top-5% for domesticated silkworms

Search Engines

Gene name, or gene ID,
or gene description

A list of gene IDs
or previous gene IDs

Chromosome segements

DNA/protein sequences

Fuzzy search

Search by gene IDs

Search by genomic regions

BLAT/Blast

Search Reaults

Information

Basic information

Gene function

Protein 2D/3D

GO, KEGG, pathways

Subcellular location

Protein topology

Homologous genes

Population genetics tests

Expression of genes
in different samples

Epigenomic signal
in different samples

Analysis Tools

Gene details

Genome browser

Subcellular localization

Protein Structure analysis

Transcriptome analysis

Epigenomic signals visualization

Gene ontology and pathway analysis

Population genetics analysis

Transcription factors

Repeat elements