



Article

# Proteomic Identification of *Bombyx mori* Organelles Using the Engineered Ascorbate Peroxidase APEX and Development of Silkworm Organelle Proteome Database (SilkOrganPDB)

Tian Li <sup>1,2,\*</sup>, Chen Xu <sup>1,2</sup>, Jinzhi Xu <sup>1,2</sup>, Jian Luo <sup>1,2</sup>, Bin Yu <sup>1,2</sup>, Xianzhi Meng <sup>1,2</sup>, Chunfeng Li <sup>1,2</sup>, Guoqing Pan <sup>1,2</sup> and Zeyang Zhou <sup>1,2,3,\*</sup>

- <sup>1</sup> State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China; xuchen\_swu@163.com (C.X.); xjz\_swu@126.com (J.X.); jianluo0214@163.com (J.L.); yubin5868@outlook.com (B.Y.); mxzmdswu@126.com (X.M.); cfli@163.com (C.L.); gqpan@swu.edu.cn (G.P.)  
<sup>2</sup> Chongqing Key Laboratory of Microsporidia Infection and Control, Southwest University, Chongqing 400715, China  
<sup>3</sup> College of Life Science, Chongqing Normal University, Chongqing 400047, China  
\* Correspondence: lit@swu.edu.cn (T.L.); zyzhou@swu.edu.cn (Z.Z.)  
† These authors contributed equally to the work.



**Citation:** Li, T.; Xu, C.; Xu, J.; Luo, J.; Yu, B.; Meng, X.; Li, C.; Pan, G.; Zhou, Z. Proteomic Identification of *Bombyx mori* Organelles Using the Engineered Ascorbate Peroxidase APEX and Development of Silkworm Organelle Proteome Database (SilkOrganPDB). *Int. J. Mol. Sci.* **2021**, *22*, 5051. <https://doi.org/10.3390/ijms22095051>

Academic Editor: Thomas G. Emyr Davies

Received: 5 April 2021  
Accepted: 7 May 2021  
Published: 10 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Silkworm *Bombyx mori* is an economically important insect and a lepidopteran model. Organelle proteome is vital to understanding gene functions; however, it remains to be identified in silkworm. Here, using the engineered ascorbate peroxidase APEX, we constructed transgenic *B. mori* embryo cells (BmE) expressing APEX-NLS, COX4-APEX, APEX-Rev, and APEX-KDEL in nucleus, mitochondrial matrix (MM), cytosol, and endoplasmic reticulum (ER), and isolated the biotin-labeled proteins using streptavidin-affinity purification, respectively. The isolated proteins were determined using LC-MS/MS and annotated by searching *B. mori* genomes downloaded from GenBank, SilkBase, SilkDB 2.0, and SilkDB 3.0, resulting in 842, 495, 311, and 445 organelle proteins identified, respectively. We mapped the 296 MM proteins annotated in the GenBank data to mitochondrial protein databases of the fly, human, and mouse, and found that 140 (47%) proteins are homologous to 80 fly proteins, and 65 (22%) proteins match to 31 and 29 human and mouse proteins, respectively. Protein orthology was predicted in multiple insects using OrthoMCL, producing 460 families containing 839 proteins we identified. Out of 460 families, 363 were highly conserved and found in all insects, leaving only three proteins without orthology in other insects, indicating that the identified proteins are highly conserved and probably play important roles in insects. A gene ontology enrichment analysis by clusterProfiler revealed that the nucleus proteins significantly enriched in cellular component terms of nucleus and nucleolus, the MM proteins markedly enriched in molecular function terms of nucleotide binding, and the cytosol proteins mainly enriched in biological process terms of small molecule metabolism. To facilitate the usage and analysis of our data, we developed an open-access database, Silkworm Organelle Proteome Database (SilkOrganPDB), which provides multiple modules for searching, browsing, downloading, and analyzing these proteins, including BLAST, HMMER, Organelle Proteins, Protein Locations, Sequences, Gene Ontology, Homologs, and Phylogeny. In summary, our work revealed the protein composition of silkworm BmE organelles and provided a database resource helpful for understanding the functions and evolution of these proteins.

**Keywords:** silkworm; *Bombyx mori*; organelle; proteome; database

## 1. Introduction

Silk moth is a large group of lepidopteran insects that includes the domestic silkworm *Bombyx mori* and some wild animals such as *Actias selene*, *Antheraea assamensis*, *Antheraea pernyi*, *Bombyx huttoni*, *Bombyx mandarina*, *Caligula japonica*, *Eriogyna pyretoum*, *Philosamia cythiaricini*, and *Samia cynthia*. The *B. mori* is an economically important insect and has

become a model organism for lepidoptera. To date, three silkworm genomes from the *B. huttoni*, *B. mandarina*, and *B. mori* have been sequenced and deposited in the GenBank database under the accession numbers GCA\_002197625.1, GCA\_003987935.1, and GCA\_000151625.1, respectively. The genomic and genetic maps greatly promoted studies of the functional and comparative genomics of silkworms [1–3], especially with a high-quality genome assembly of *B. mori* obtained and published recently [4]. This improved assembly and annotation provided a more accurate reference for transcriptomics and proteomics studies. Referring to the genome sequence, large-scale transcriptome analysis of silkworms was performed with microarray and RNA-seq technologies, respectively [5,6]. Transcriptional profiles in different tissues and at different development stages partially indicate the functional characteristics of silkworm protein-coding genes; however, these need to be further revealed at the proteomics level. Silkworm proteomes have been further dissected at the organ and tissue levels. Proteins from the head, silk gland, colleterial gland, hemolymph, midgut, embryo, and spinneret have been isolated and identified with mass spectrometry [7–16]. In addition, silkworm proteomic responses against infections by virus, bacteria, and fungi pathogens have also been widely investigated using high-throughput methods [17–22]. To date, little is known about the protein composition of silkworm subcellular organelles like the nucleus, mitochondria, cytosol, and endoplasmic reticulum (ER), except that Wang et al. isolated proteins from the cytosol, mitochondria and microsomes of silkworm infected by nucleopolyhedrovirus [23]. However, organelle proteomes that identified with targeted and accurate isolation methods have not been reported in silkworms.

Studies of organelle proteome have been limited by the technology for specific purifying of objective organelles from cells and tissues. Recently, a proteomic mapping technology via spatially restricted enzymatic tagging using the engineered ascorbate peroxidase APEX was developed and successfully used to identify organelle proteins from living cells and tissues [24–27]. Proteins from human mitochondrial matrix (MM), intermembrane space, membranes of ER and outer mitochondria, and lipid droplets, fly tissues, yeast, and worm organelles have been identified using the engineered APEX [28–32]. The APEX-based technology has become an effective tool for identifying proteomes in cells and tissues.

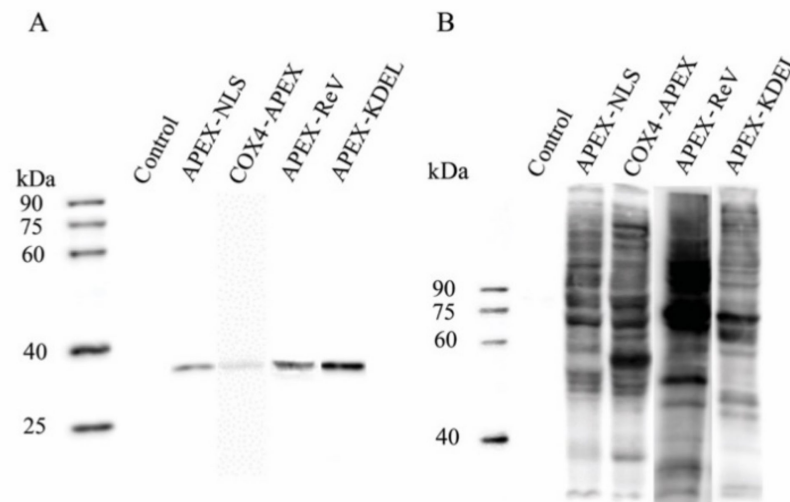
Here, for the first time, we isolated and identified protein components from the nucleus, MM, ER, and cytosol of silkworm embryo cells (BmE), and developed an open-access database resource for silkworm organelle proteomics.

## 2. Results

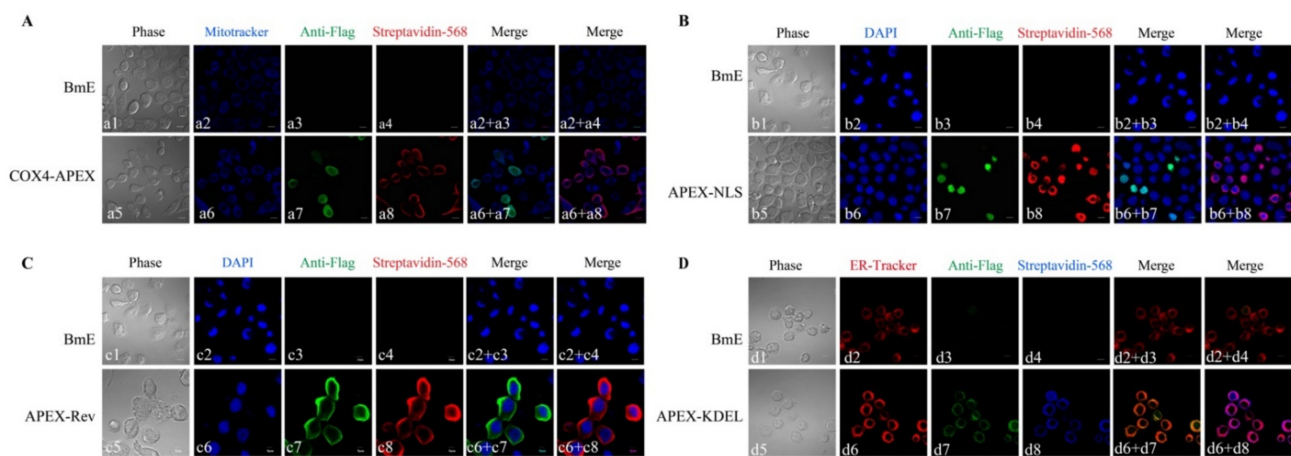
### 2.1. Expression of APEX in Transgenic BmE cells

Transgenic BmE cells expressing APEX fused with a leading peptide including COX4, NLS, Rev, and KDEL were made for translocating the APEX in MM, nucleus, cytosol, and ER, respectively. For verifying the expression and subcellular localization of the APEX, a FLAG tag was inserted between the APEX and leading peptide. The expression of COX4-FLAG-APEX, APEX-FLAG-NLS, APEX-FLAG-Rev, and APEX-FLAG-KDEL was determined using Western blotting with an antibody against the FLAG (anti-FLAG). As a result, the fusion proteins were expressed explicitly in transgenic cells (Figure 1A).

The subcellular localization of APEX in transgenic BmE cells was analyzed using an indirect immunofluorescence assay (IFA) with anti-FLAG. The COX4-APEX was colocalized with mitochondria stained by MitoTracker Deep Red (Figure 2A), indicating that the COX4 specifically led the APEX into the mitochondria. The APEX-NLS colocalized with the nuclear dye DAPI (sigma) (Figure 2B), APEX-Rev localized in the cytosol (Figure 2C), and the APEX-KDEL colocalized with the ER dye (Thermo Fisher) (Figure 2D). These results manifested that the APEX was successfully expressed in BmE cells, and the leading sequences could specifically translocate the APEX into target organelles.



**Figure 1.** Western blotting analysis of APEX expression and APEX-mediated protein biotinylation. (A) Western blotting analysis of the expression of APEX with anti-FLAG. (B) Western blotting analysis of APEX-mediated biotinylation in transfected BmE cells using streptavidin. The control was the BmE cells without APEX-transfection.



**Figure 2.** IFA of the subcellular localization of COX4-APEX (A), APEX-NLS (B), APEX-Rev (C), APEX-KDEL (D), and biotinylated proteins in BmE organelles, respectively. The BmE cells in a1-a4, b1-b4, c1-c4, d1-d4 were transfected without APEX (control). The BmE cells in a5-a8, b5-b8, c5-c8 and d5-d8 were transfected with COX4-APEX, APEX-NLS, APEX-Rev, and APEX-KDEL, respectively. The bar in A, B, and D indicates 10  $\mu$ m, and the bar in C represents 5  $\mu$ m.

## 2.2. Verification of APEX-Mediated Protein Biotinylation

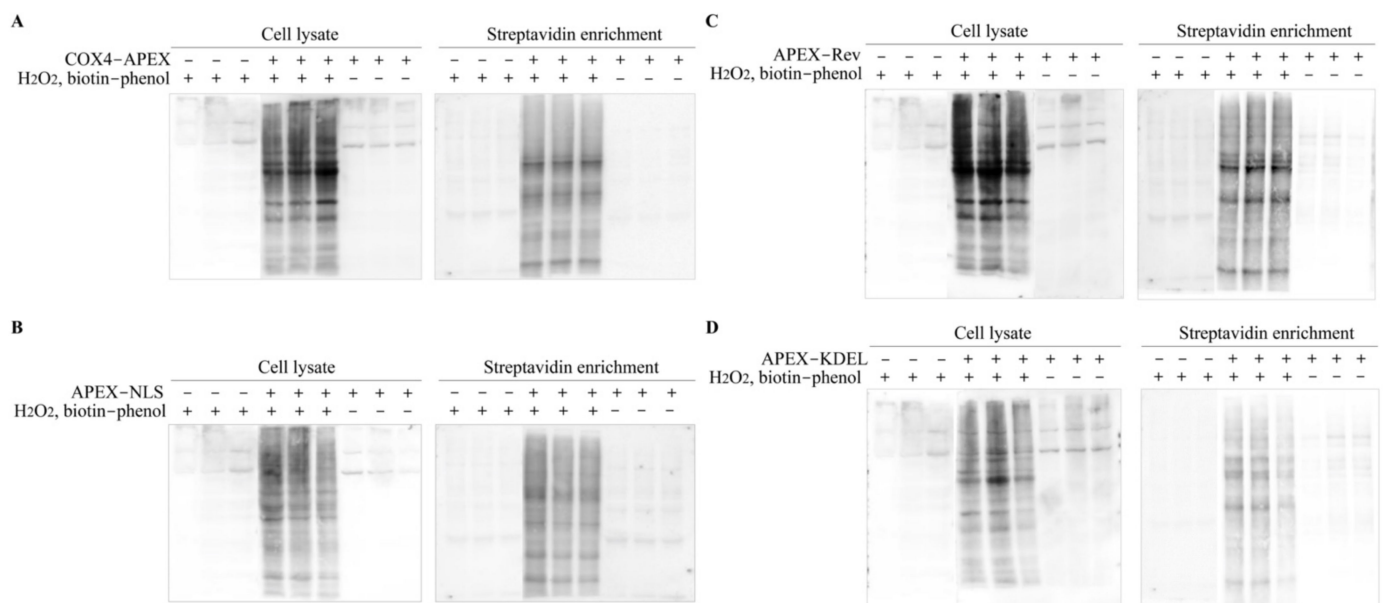
The catalytic activity of the APEX in the BmE organelles was verified by adding reaction substrates, biotin-phenol, and  $H_2O_2$  into the culture media. The biotinylation was demonstrated using Western blotting with an antibody against streptavidin-HRP (Figure 1B). In the result, protein biotinylation was found in APEX-transgenic BmE cells, but not in that without APEX transfection. To examine the subcellular localization of the biotinylated proteins, Alexa Flour 568 conjugated streptavidin was used to perform IFA, which results showed that the biotinylated proteins well overlapped with COX4-APEX in MM (Figure 2A), co-localized with APEX-NLS in the nucleus (Figure 2B), and merged with APEX-Rev and APEX-KEDL in cytosol and ER, respectively (Figure 2C, D). These results

indicated that the APEX successfully transferred biotin-phenol into biotin-radicals, leading to the specific biotinylation of proteins in targeted organelles.

### 2.3. Enrichment of the Biotinylated Proteins

To improve the accuracy of protein identification, we prepared two negative control groups and one experimental group, with three independent replicates for each group. One negative control group was BmE cell only expressing EGFP with the addition of substrates biotin-phenol and H<sub>2</sub>O<sub>2</sub>, and the other was BmE cell expressing APEX without the substrates. Thus, there should be no biotinylated proteins found in the two negative control groups. The experimental group was BmE cell expressing APEX and with the addition of reaction substrates, which should produce biotinylated proteins around where there was APEX protein.

Before the enrichment of the biotinylated proteins, we first determined the biotinylation in whole-cell lysate using Western blotting with streptavidin-HRP (Figure 3A–D). As a result, no biotin-labeled protein was detected in the negative control groups, while abundant biotinylated proteins were found in the experimental groups. Moreover, a significant difference in biotinylation was observed among the samples from the nucleus, MM, cytosol, and ER, suggesting that proteins in targeted organelles were successfully labeled by biotin. The biotinylated proteins were enriched using streptavidin-coated magnetic beads and then verified by Western blotting, the results of which manifested the successful enrichment (Figure 3A–D).

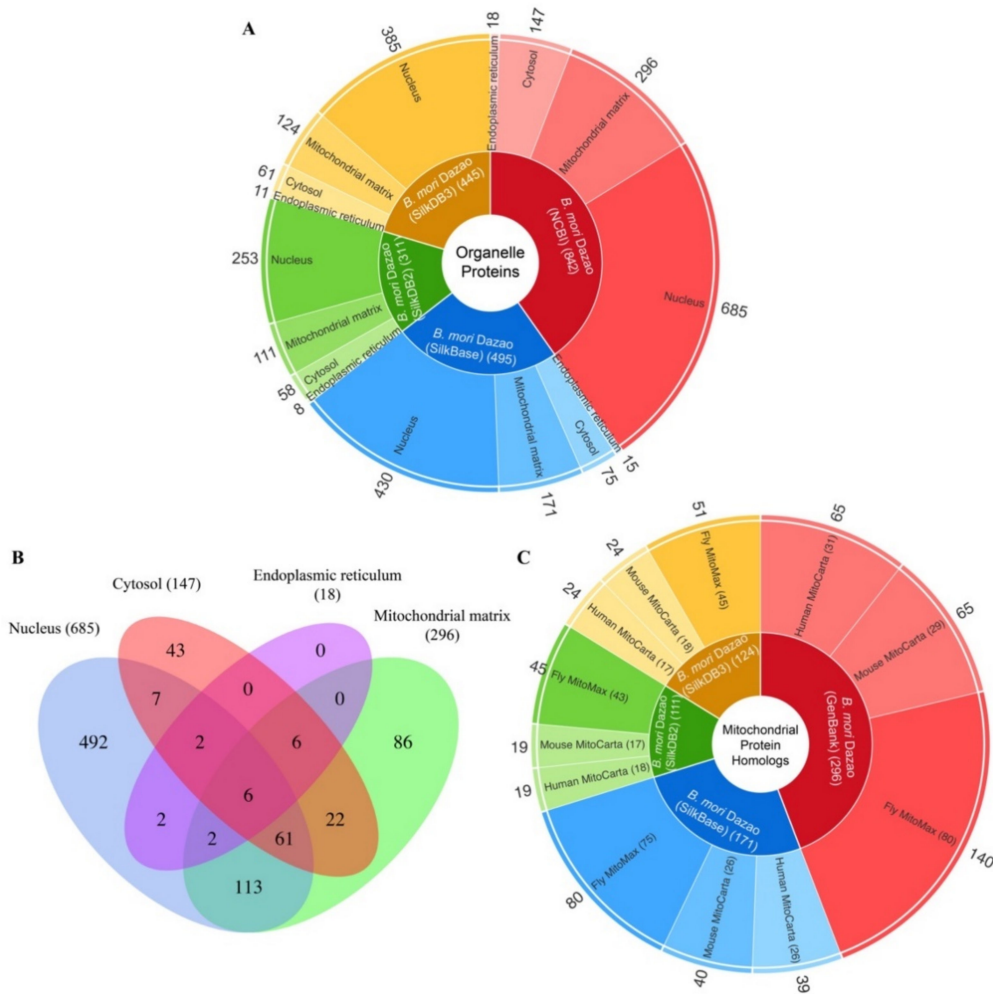


**Figure 3.** Streptavidin blotting analysis of biotinylated proteins before (cell lysate) and after enrichment (streptavidin enrichment) from MM (A), nucleus (B), cytosol (C) and ER (D).

### 2.4. Identification and Annotation of the Biotinylated Proteins

The enriched proteins from the nucleus, MM, cytosol, and ER were determined using mass spectrometry and searched against silkworm genomes, respectively. In general, we identified 842, 495, 311, and 445 proteins in genomes from GenBank, SilkBase, SilkDB 2.0, and SilkDB 3.0, respectively (Figure 4A, Table S1–S4). The largest number of proteins was obtained from the GenBank data, resulting in 685, 296, 147, and 18 proteins of the nucleus, MM, cytosol, and ER, respectively. Some proteins identified in the four organelles showed multiple subcellular localizations. For example, 193 out of 685 (28.2%) nucleus proteins and 210 out of 296 (70.9%) MM proteins were found in multiple organelles (Figure 4B).

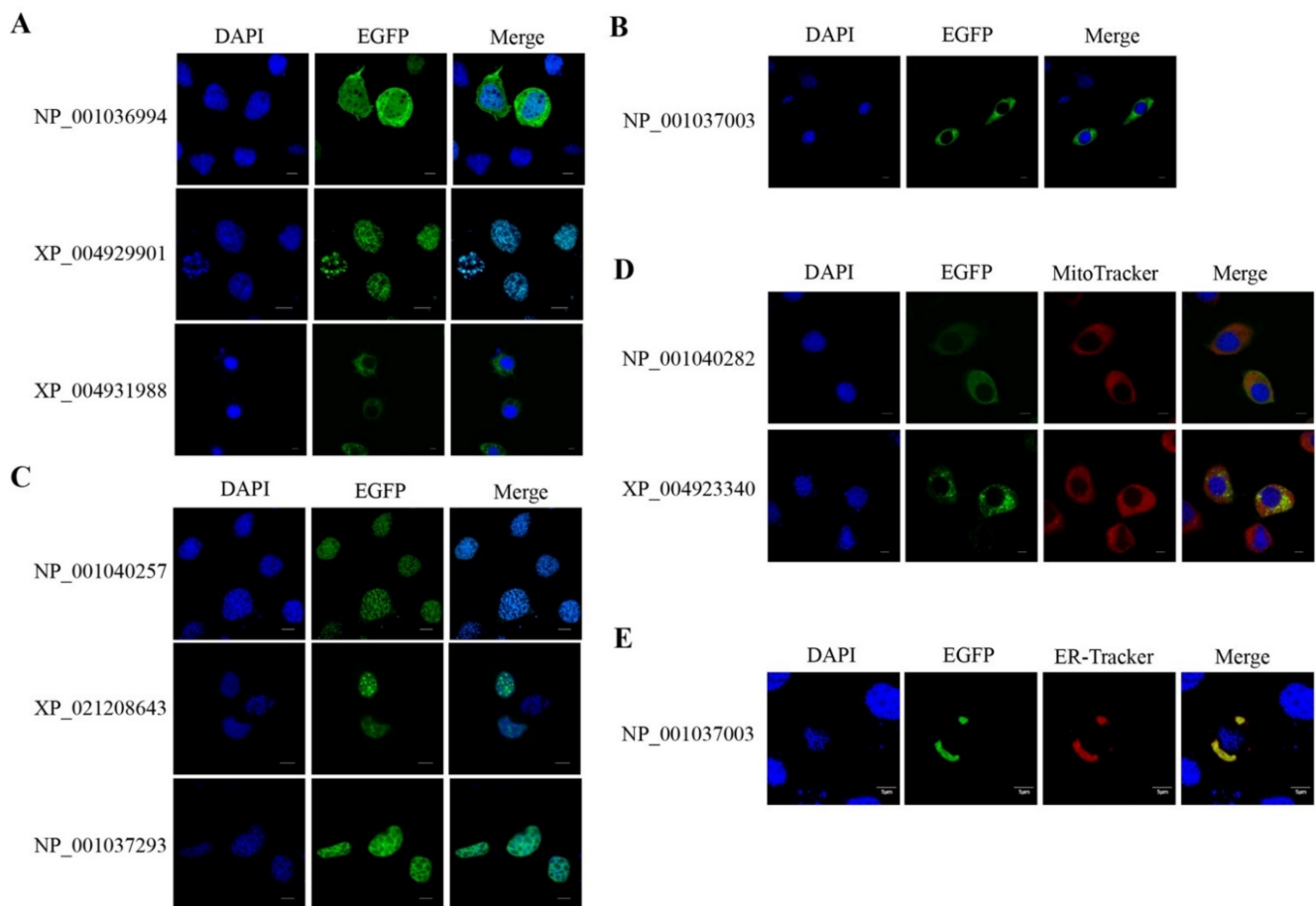




**Figure 4.** Summary of the identified silkworm organelle proteins. (A) The nucleus, MM, cytosol, and ER proteins are annotated in silkworm genomes from GenBank (red), SilkBase (blue), SilkDB 2.0 (green), and SilkDB 3.0 (yellow), respectively. (B) The distribution of the 842 organelle proteins annotated in GenBank data in the nucleus, cytosol, MM, and ER. (C) The identified silkworm MM proteins were mapped to databases of the fly MitoMax, mouse and human MitoCarta, in which proteins were mainly identified using the APEX method.

The identified MM proteins were aligned against the fly MitoMax, human and mouse MitoCarta databases, in which mitochondrial proteins were mainly identified by the APEX strategy. For the MM proteins annotated in the genome of the GenBank version, 140 out of 296 (47%) proteins were aligned to 80 fly mitochondrial proteins, and 65 out of 296 (22%) proteins were matched to 31 and 29 human and mouse mitochondrial proteins, respectively (Figure 4C), indicating that the identified MM proteins are more conserved between silkworm and fly.

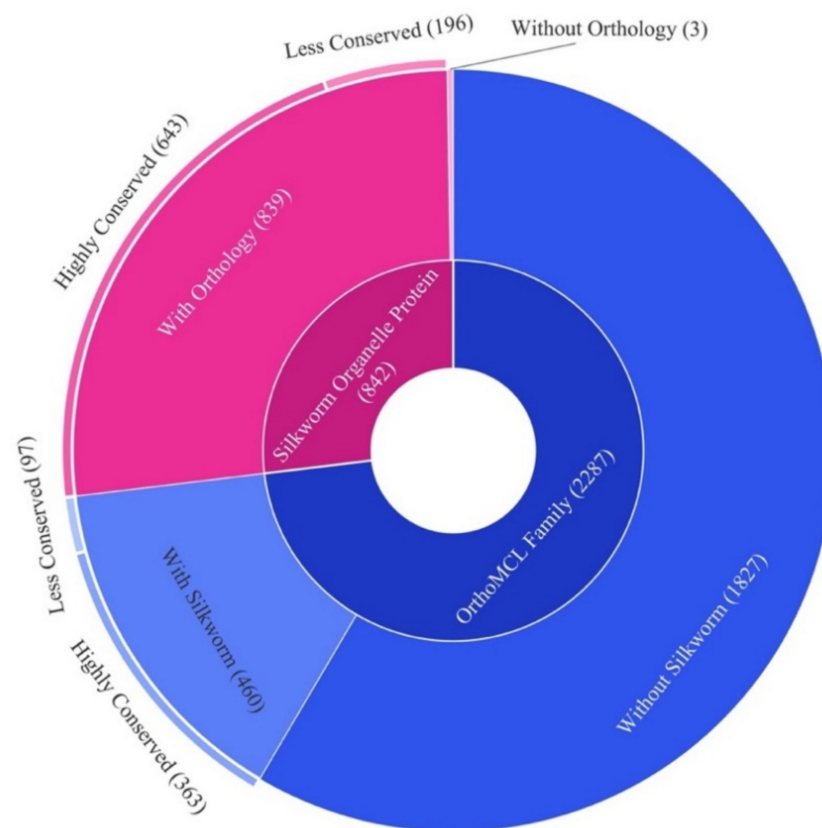
To verify the subcellular localization of the identified proteins, we performed transgenic analysis on some candidates in the BmE cells, including one protein from the cytosol and nucleus, two proteins from the cytosol, four proteins in the nucleus, and two proteins located in the mitochondrion, respectively. As shown in Figure 5, protein candidates fused with EGFP were shown in organelles where they were isolated from. These results indicated the high accuracy of our isolation and identification.



**Figure 5.** Verification of protein locations in BmE cells. The identified proteins were fused with EGFP and transfected BmE cells. The nucleus, MM, and ER were stained with DAPI (blue), MitoTracker (red) and ER-Tracker (red), respectively. (A) Subcellular localization of proteins identified in nucleus and cytosol. (B–D) Subcellular localization of proteins identified in the cytosol, nucleus, and MM, respectively. (E) Subcellular localization of protein identified in ER. The IDs indicate the GenBank accession number for each protein.

### 2.5. Orthology of the Organelle Proteins

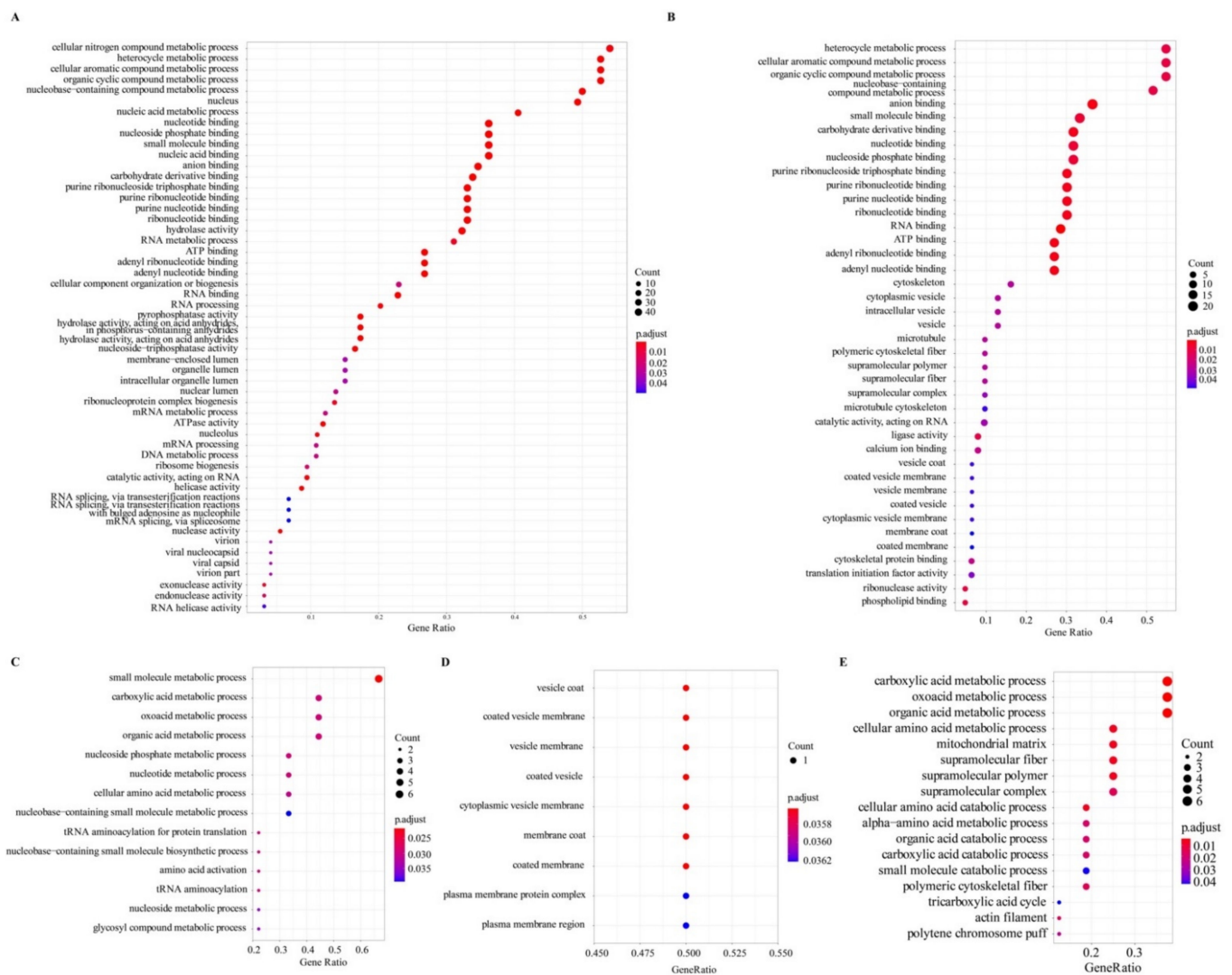
The protein orthology among *B. mori*, *Bombyx mandarina*, *Manduca sexta*, *Papilio xuthus*, *Spodoptera litura*, *Trichoplusia ni*, *Tribolium castaneum*, *Aedes aegypti*, *Drosophila melanogaster*, *Apis cerana*, and *Apis mellifera* was determined using the OrthoMCL. In the result, all proteins were grouped into 22,877 families, in which 460 families contained 839 out of 842 (99.6%) organelle proteins annotated in the GenBank silkworm data, leaving only 3 proteins without homolog in other insects. Moreover, 363 out of 460 (78.9%) families were found in all insects (Figure 6, Tables S5 and S6). These results suggest that the identified organelles proteins are highly conserved and probably play important roles in insects.



**Figure 6.** Summary of orthology for the identified silkworm organelle proteins annotated in the GenBank data. Pies in blue show the summary of protein families. Pies in magenta indicate the summary of silkworm organelle proteins included in families.

### 2.6. Gene Ontology Enrichment of the Identified Proteins

The organelle proteins annotated from the silkworm genome of the GenBank version were used to perform gene ontology (GO) enrichment analysis using clusterProfiler software. In the result, we obtained 53, 41, 20, and 9 GOs for 117, 54, 7, and 2 proteins of the nucleus, MM, cytosol, and ER, respectively. After enrichment, the nucleus proteins were mainly enriched in biological process terms of cellular nitrogen compound metabolism, heterocycle metabolism, nucleobase-containing compound metabolism, cellular aromatic compound metabolism and organic cyclic compound metabolism, molecular function terms of nucleotide binding, and cellular component terms of nucleus and nucleolus (Figure 7A); the MM proteins were significantly enriched in molecular function terms of nucleotide binding (Figure 7B); the cytosol proteins were markedly enriched in biological process terms of small molecule metabolism (Figure 7C); and the ER proteins were mainly involved in vesicle and plasma membrane (Figure 7D). Furthermore, homologs of the silkworm MM proteins searched in the fly MitoMax (Figure 4C) were retrieved to validate GO enrichment, which results showed that the fly homologs were enriched in mitochondrial component (Figure 7E).

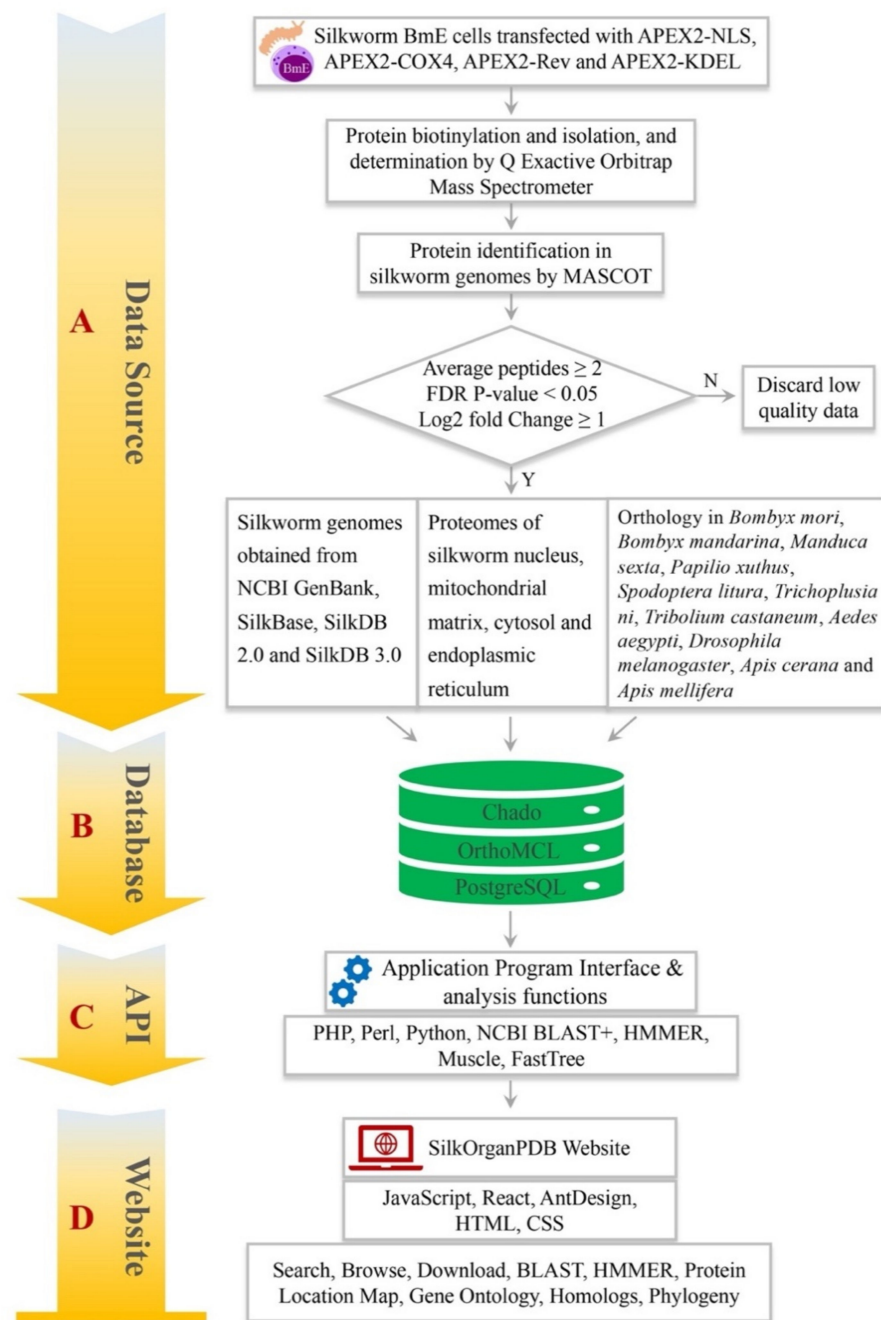


**Figure 7.** GO enrichment for silkworm organelle proteins of the nucleus (A), MM (B), cytosol (C), ER (D), and MitoMax homologs (E).

## 2.7. SilkOrganPDB Architecture and Implementation

The identified organelle proteins and silkworm genome datasets obtained from GenBank, SilkBase, SilkDB 2.0 and SilkDB 3.0 were parsed and imported into the Chado schema, which is managed by the PostgreSQL (Figure 8A). Homologous proteins among *B. mori*, *B. mandarina*, *M. sexta*, *P. xuthus*, *S. litura*, *T. ni*, *T. castaneum*, *A. aegypti*, *D. melanogaster*, *A. cerana* and *A. mellifera* were predicted with the OrthoMCL, and also imported into the database (Figure 8B). Application programming interface (API) and analysis tools were developed using PHP, Perl, and Python to search data in Chado, and execute analysis programs such as NCBI BLAST+, HMMER, and FastTree (Figure 8C). Based on the Chado and API, the SilkOrganPDB website was designed with JavaScript, React, AntDesign, HTML, and CSS. The website provides users an interface to perform browse, text search, homologous sequence search, download, and phylogenetic analysis (Figure 8D).





**Figure 8.** Workflow to build the SilkOrganPDB. (A) Isolation and identification of organelle proteins from silkworm nucleus, MM, cytosol, and ER. (B) Database schema of SilkOrganPDB. (C) Development of application program interface (API) and analysis tools. (D) Design of SilkOrganPDB website.

The SilkOrganPDB website (<https://silkorgan.biodb.org/>, accessed on 30 March 2021) comprises two main panels: the query and functional modules (Figure 9). The query panel provides the user an interface to search organelle proteins by entering keywords and selecting the organelles of an organism, the results of which would be presented right below the input box (Figure 9A). The functional module panel allows the user to browse organism data, download datasets, and analyze data (Figure 9B).

The screenshot displays the SilkOrganPDB homepage. At the top, the title is 'SilkOrganPDB - Silkworm Organelle Proteome Database' with a 'Version 1.0' badge. Below the title is a navigation menu with 'Browse', 'Analysis', 'Downloads', 'Links', and 'Documents'. The main content area is divided into two panels. Panel (A) is the query panel, featuring a search bar with 'clathrin' entered and a list of protein results. Panel (B) is the functional module panel, showing a central 'OrganPDB' logo surrounded by icons for 'Search & Browse', 'eFP', 'Sequences', 'Phylogeny', 'Gene Ontology', and 'Orthology'.

**Figure 9.** The homepage of the SilkOrganPDB. (A) Query panel. (B) Functional module panel.

### 2.7.1. Search and Browse Modules

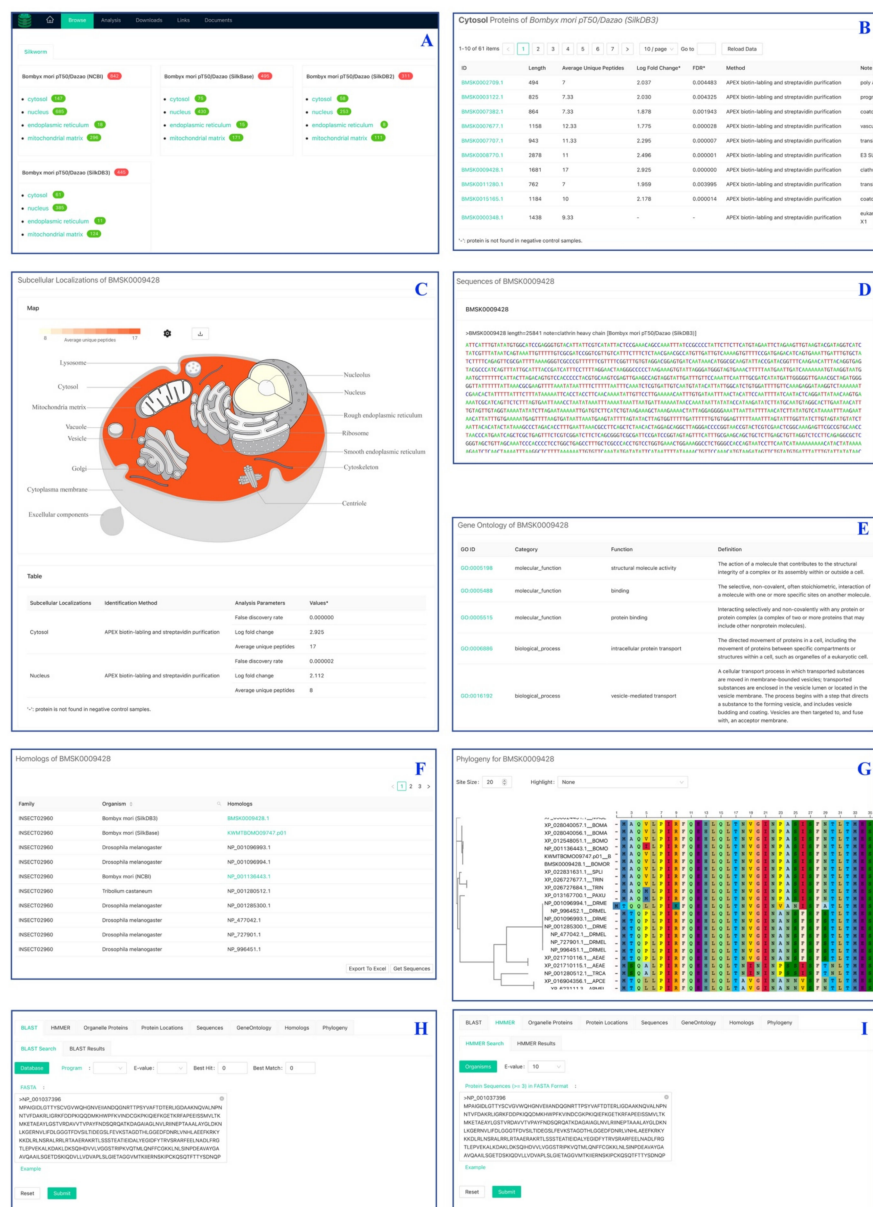
The organelle proteins in the database can be searched by typing protein identifiers or text keywords in the input box (Figure 9A). The query results would be shown in a table, items in which could be exported to an Excel table and checked to be listed below the input box. On the other hand, the proteome of an organelle can be browsed by clicking the name listed on the index page (Figure 10A). All identified proteins of the selected organelle would be listed in a table, which shows basic information about the protein annotation and identification (Figure 10B). Clicking on a protein ID listed in the querying results and proteome table would initiate analysis of the selected protein.

### 2.7.2. eFP Module

The subcellular localization of a protein is shown as an electronic fluorescent pictograph (eFP) (Figure 10C), which responsively colors the organelles based on the average number of unique peptides determined by the mass spectrometry. The eFP map is drawn as a vector image, which can be magnified indefinitely and downloaded for further editing to make high-resolution figures. The color of the eFP map is configurable in absolute and relative color mode. The relative mode can be used to compare the relative quantities among different proteins.

### 2.7.3. Sequence and GO Module

The Sequence module shows basic information for a protein, including sequences of gene, mRNA and protein, and sequence length and annotation (Figure 10D). The GO module provides the ontology terms for the protein queried, including molecular function, biological process, and cellular component (Figure 10E). By clicking on a GO ID, users can visit the detailed term information in the AmiGO database [33].



**Figure 10.** The main functional modules in the SilkOrganPDB. (A) Index organelle proteomes. (B) Browse organelle proteome. (C) eFP. (D) Sequence features. (E) Gene ontology. (F) Protein homologs in insects. (G) Protein phylogeny. (H) BLAST. (I) HMMER.

### 2.7.4. Homologs and Phylogeny Module

The homologs module provides homology information of a *B. mori* protein in *B. mandarina*, *M. sexta*, *P. xuthus*, *S. litura*, *T. ni*, *T. castaneum*, *A. aegypti*, *D. melanogaster*, *A. cerana*, and *A. mellifera* (Figure 10F). The information table can be exported to an Excel file. All homologous sequences can be exported by clicking on the Get Sequences button. The phylogeny module performs phylogenetic analysis and constructs a phylogenetic tree based on the homologous sequences in the homologs module (Figure 10G).

### 2.7.5. BLAST Search and HMMER Search Modules

The BLAST Search built with the NCBI BLAST+ is used to search organelle proteins in the databases for query sequences that users enter in FASTA format (Figure 10H). The search results would be shown in a table, displaying detailed alignment information for all target sequences. The HMMER Search developed with HMMER software is designed for

searching homologous proteins with user-provided seed sequences (Figure 10I). The target proteins identified in the organelles would be shown as a hyperlink, by which users can explore protein features and perform analysis of them.

### 3. Discussion

The silkworm *B. mori* is an economic insect and a great model for studying lepidoptera with plenty of mutant phenotypes and fine genomic and genetic maps, as well as massive functional omics data including transcriptomes, tissue proteomes, and epigenomes. However, little proteomics data were available at the organelle level in silkworms. This was mainly limited by the lacking effective technology for targeting isolation of proteins from subcellular organelles. For example, Wang et al. identified only 87 proteins from the cytosol, mitochondria, and microsomes of silkworm midguts using two-dimensional gel electrophoresis [23]. Instead, with the APEX technology, we obtained many more organelle proteins from only one type of cell line. Therefore, it is necessary to perform large-scale isolation and identification of organelle proteomes from silkworm cells, tissues, and organs using the APEX, to make the silkworm a better lepidopteran model.

Some proteins we identified are present in multiple organelles (Figure 4B). This was also found in *Caenorhabditis elegans* organelle proteins identified by Reinke et al. [31]. Out of 3180 proteins identified in *C. elegans* cytosol and nucleus, 2048 showed multiple subcellular location (MSL), leaving only 1132 that are organelle-specific. On the one hand, the MSL indicates protein translocation among organelles. For example, the tumor suppressor factor p53 keeps a balance between the cytosol and nucleus under normal circumstances, but aggregates in the nucleus during cell apoptosis [34]. On the other hand, the MSL suggests potential protein interactions. This is especially common for hub proteins, which interact with numerous partners [35]. In CompPPI v2.1.1, the compartmentalized protein-protein interaction database, 1413 out of 1989 mitochondrial proteins of *D. melanogaster* display MSL [36]. We found 58 silkworm MM proteins homologous to the fly proteins with MSL (Table S7). The silkworm proteins with MSL provide clues for understanding their function and interplay. Although the transfection validation showed high accuracy (Figure 5), a low false-positive rate might exist, as shown in the study by Chen et al. [29]. This is probably caused by the defects of streptavidin purification. High throughput of transgenic verification is necessary to ultimately confirm the protein localizations, even though not all proteins could express in the transfected cells.

The SilkOrganPDB is the first database that provides silkworm organelle proteomes identified with an organelle-targeted method. The SilkDB 3.0 is a comprehensive data source for silkworm genomics [37]. However, the protein subcellular localizations in this database were predicted using bioinformatic methods and lack experimental verification. In the future, we will keep identifying silkworm organelle proteomes using the APEX, and developing the SilkOrganPDB for integrating more data determined experimentally and predicted with bioinformatics methods. The SilkOrganPDB provides a proteomics resource that is helpful for understanding gene functions and the better modeling of the silkworm.

## 4. Materials and Methods

### 4.1. Construction of APEX-Transgenic BmE Cell

Silkworm *B. mori* embryo cells (BmE) were cultured in Grace media containing 10% fetal bovine serum (Gibco, Australia Origin), 50 µg/mL streptomycin, and 50 unites/mL penicillin at 28 °C. The coding sequence of APEX fused with an organelle-targeting sequence, and a FLAG-tag was synthesized and cloned to a piggyBac vector for transfecting BmE cells. The amino acid sequences of nuclear localization signal (NLS) and nuclear export signal (Rev) were PKKKRKV and LQLPPLRLRLD. The MM targeting sequence was the first 30 amino acids of silkworm COX4 protein (GenBank accession no. NM\_001079652.1), and the ER retention sequence was KDEL. BmE cells expressing the APEX fusion proteins were screened with the geneticin G418.

#### 4.2. Biotinylation of Organelle Proteins

The transgenic cells were incubated with culture media containing biotin-phenol (final concentration 500  $\mu$ M) at 28 °C for 30 min. Protein-labeling was initiated by adding 1 mM H<sub>2</sub>O<sub>2</sub> for 1 min with shaking gently. After a 1 min reaction, the quencher solution (the final concentration of 5 mM Trolox and 10 mM sodium ascorbate in DPBS) was added to halt the labeling. The cells were washed three times with 5 mM Trolox and 10 mM sodium ascorbate in DPBS.

#### 4.3. Western Blotting and Streptavidin Blotting

The transgenic cells were collected and lysed with RIPA lysis buffer for 10 min at 4 °C to detect the expression of APEX fusions using Western blotting. The APEX-mediated biotinylation was verified by streptavidin blotting. The cells were collected by centrifugation at 3000 $\times$  g for 5 min and stored at –80 °C. The cell pellet was lysed with RIPA lysis buffer (Sigma, Shanghai, China), Lot # SLBQ9770V) by additionally adding buffers including 10 mM sodium azide, 10 mM sodium ascorbate, 5 mM Trolox, 1 $\times$  protease cocktail, and 1 mM PMSF at 4 °C for 10 min, and centrifuged at 14,000 $\times$  g for 15 min at 4 °C. The protein in the supernatants was checked with 10% SDS-PAGE gel, transferred to a PVDF membrane at constant voltage 25 v for 23 min and blocked with 5% skimmed milk in PBS-buffer saline at 37 °C for 1 h. The blots were immersed in a 1:1000 dilution of anti-FLAG at 4 °C overnight and incubated in 1:5000 peroxidase labeled goat anti-mouse IgG for 1 h at 37 °C or just immersed in a 1:1000 streptavidin-HRP at 37 °C for 1 h after blocking. Before imaging using an Azure Biosystems C300, the blots were incubated with ECL Western Blotting Substrate (Bio-Rad) for 1 min.

#### 4.4. Immunofluorescence Assay (IFA) of APEX Location and Biotinylated Proteins

The cells were plated into a 12-well plate for one day, followed by staining with MitoTracker Deep Red and ER dyes at 28 °C for 20 to 30 min, and then fixed in 4% paraformaldehyde at room temperature for 15 min. After fixation, the cells were washed with PBS three times for 5 min each time, permeabilized with 0.5% TritonX-100 for 20 min, and washed three times with PBS. Then, cells were blocked with 10% goat serum and 5% BSA in PBS for 1–2 h and incubated with anti-FLAG (1:800 dilution) in PBS or blocking buffer for 2 h at room temperature. After being washed five times with PBST (0.05% Tween in PBS), the cells were incubated with Alexa Flour 488 conjugate Goat anti-Mouse IgG (1:1000 dilution) and Streptavidin-AF (Alexa Fluor)–568 (1:1000 dilution) at room temperature for 1 h, and then washed again for five times with PBST. Finally, the samples were stained with DAPI (4'6-diamidino-2-phenylindole, Sigma) at room temperature for 10–20 min. Before imaging with an Olympus FV1200 laser scanning confocal microscope, the Fluoromount™ Aqueous Mounting Medium (Sigma) was added.

#### 4.5. Isolation of Biotinylated Proteins

The biotinylated proteins were isolated using streptavidin-coated magnetic beads. Before enrichment, the streptavidin-coated magnetic beads were washed twice with 1 mL RIPA lysis buffer, and 4 mg total proteins were mixed and incubated with 500  $\mu$ L streptavidin-coated magnetic beads for 1 h at room temperature. Then, the beads were washed with 1 mL RIPA lysis buffer two times, and 1 M KCl, 0.1 M Na<sub>2</sub>CO<sub>3</sub>, 2 M urea for one time, respectively, and again twice with RIPA lysis buffer to elute the non-specific binding proteins. Biotinylated proteins were eluted by adding 60  $\mu$ L SDT buffer (4% SDS, 100 mM Tris-HCl, 1 mM DDT pH 7.6) and boiling for 5–10 min. 5  $\mu$ L supernatant was used for Western blotting and the rest for mass spectrometry.

#### 4.6. Identification of Biotinylated Proteins

Protein digestion was performed according to the FASP procedure described in [38]. Briefly, the protein pellet (about 30  $\mu$ g) was solubilized in 30  $\mu$ L SDT buffer (4% SDS, 100 mM DTT, 150 mM Tris-HCl (pH 8.0)) at 90 °C for 5 min. The detergent, DTT, and



other low-molecular-weight components were removed using 200  $\mu$ L UA buffer (8 M Urea, 150 mM Tris-HCl pH 8.0) by repeated ultrafiltration (Microcon units, 30 kD). Then, 100  $\mu$ L 0.05 M iodoacetamide in UA buffer was added to block reduced cysteine residues and the samples were incubated for 20 min in darkness. The filter was washed with 100  $\mu$ L UA buffer three times and then 100  $\mu$ L 25 mM  $\text{NH}_4\text{HCO}_3$  twice. Finally, the protein suspension was digested with 2  $\mu$ g trypsin (Promega) in 40  $\mu$ L 25 mM  $\text{NH}_4\text{HCO}_3$  overnight at 37 °C, and the resulting peptides were collected as a filtrate.

LC-MS/MS determination was performed using a Q Exactive mass spectrometer (Thermo Scientific). The mass spectrometer was operated in positive ion mode. MS data was acquired using a data-dependent top10 method dynamically choosing the most abundant precursor ions from the survey scan (300–1800 m/z) for HCD fragmentation. The MS spectra data were then searched using MASCOT (Matrix Science, London, UK; version 2.2) against silkworm proteins, which were downloaded from databases of the GenBank (<https://www.ncbi.nlm.nih.gov/genome/?term=ASM15162v1> (accessed on 15 October 2020)), SilkBase [4], SilkDB 2.0 [39] and SilkDB 3.0 [37], respectively. Proteins were identified with the following parameters: peptide mass tolerance = 20 ppm, MS/MS tolerance = 0.1 Da, Enzyme = Trypsin, Missed cleavage = 2, Fixed modification = Carbamidomethyl (C), Variable modification = Oxidation (M).

The protein candidates identified were then statistically screened using qprot\_v1.3.5 [40]. The peptide spectral count of each protein was used to calculate the false discovery rate (FDR) *P*-value and log<sub>2</sub> fold change (LFC) between the non-APEX and APEX-transgenic samples. Then, proteins were filtered with the following criteria: at least 2 unique peptides detected in the APEX-transgenic samples, and no peptide determined in the control samples otherwise FDR *p*-value < 0.05 and LFC  $\geq$  1, which means if a peptide is detected in the non-APEX samples the protein should have an FDR *P*-value less than 0.05 and LFC equal or greater than 1.

To verify subcellular locations, we randomly selected candidates from the identified proteins. Transfection vectors of pSL1180 with a fusion of the candidate protein and EGFP were constructed. Silkworm BmE cells were seeded into a 12-well plate by  $5 \times 10^5$  cells per well. 1  $\mu$ g constructed vectors were transiently transfected into the BmE cells according to the instructions of X-tremeGENE HP DNA transfection reagents (Roche, product number: XTGHP-RO). The medium was replaced with fresh Grace's Insect medium contains 10% serum after 5 hours. Two days later, the transfected cells were stained with DAPI to label the nucleus and then examined by confocal microscopy (Olympus, FV1200).

#### 4.7. Homologous Search in Mitochondrial Protein Databases and Insect Proteomes

The identified mitochondrial proteins in silkworm proteomes from GenBank, SilkBase, SilkDB 2.0, and SilkDB 3.0 were searched for homologs in fly, human, and mouse using BLASTP [41] with an E-value cutoff less than  $1e-5$  against mitochondrial protein databases of fly MitoMax [29], Human MitoCarta3.0 and Mouse MitoCarta3.0 [42], respectively. The MitoMax is a database for fly mitochondrial proteins, containing 5128 candidates identified by APEX labeling and isolation methods. The MitoCarta3.0 is an inventory of 1136 human and 1140 mouse proteins with the strong support of mitochondrial localization. From the BLASTP results, homologs of silkworm proteins in fly, human, and mouse were identified and counted.

Sequences of *D. melanogaster* proteins with MSL in ComPPI v2.1.1 [36] were retrieved from the UniProt [43]. The identified mitochondrial proteins of silkworm were aligned against the fly proteins with MSL using the BLASTP with E-value <  $1e-5$ .

To find orthologs of silkworm proteins in other insects, we downloaded the genome data of *Bombyx mandarina* [44], *Manduca sexta* [45], *Papilio xuthus* [46], *Spodoptera litura* [47], *Trichoplusia ni* [48], *Tribolium castaneum* [49], *Aedes aegypti* [50], *D. melanogaster* [51], *Apis cerana* [52], and *Apis mellifera* [53] from GenBank. All protein sequences were all-to-all aligned using BLASTP [41] with an E-value cutoff less than  $1e-5$ . The BLASTP results were parsed and imported into a MySQL database, tables in which were created by the

OrthoMCL [54] for identifying homologs with thresholds of the percentMatchCutoff = 50 and evaluateExponentCutoff =  $1e-6$ .

#### 4.8. Gene Ontology Enrichment

The identified organelle proteins were mapped to the Entrez Gene database in NCBI to retrieve ENTREZ IDs, with which gene ontology (GO) for each ENTREZ ID was retrieved from a local annotation database of *Bombyx mori* constructed with the AnnotationHub in Bioconductor [55], which is an R framework for bioinformatics and computational biology. The GOs were then enriched using the enrichGO in clusterProfiler software [56] with thresholds of pvalueCutoff = 0.05 and qvalueCutoff = 0.2. The GO terms of fly proteins were enriched by assigning the parameter “OrgDb = org.Dm.eg.db” when running the enrichGO. Subsequently, the enrichment was plotted using the dotplot in the clusterProfiler.

#### 4.9. Development of Silkworm Organelle Proteome Database (SilkOrganPDB)

The genomic datasets of *Bombyx mori* pT50/Dazao downloaded from the GenBank, SilkBase, SilkDB 2.0, and SilkDB 3.0 were parsed into the Generic Feature Format version 3 (GFF3, <https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.md> (accessed on 15 October 2020)) according to the Sequence Ontology [57] using the BioPerl script bp\_genbank2gff3.pl ([https://metacpan.org/pod/distribution/BioPerl/bin/bp\\_genbank2gff3](https://metacpan.org/pod/distribution/BioPerl/bin/bp_genbank2gff3) (accessed on 15 October 2020)). Then, the data in GFF3 was imported into the Chado [58], which is a relational database schema managed by the PostgreSQL and widely used to store biological data from a wide variety of organisms. The organelle proteins identified in the four versions of silkworm genomes were imported into the Chado schema as gene features, including information about the experiment, sampling, determination method, subcellular localization, and quantity.

The SilkOrganPDB server is implemented on a platform mainly composed of Linux CentOS Server 7.9 (<https://www.centos.org> (accessed on 10 October 2020)), PostgreSQL 9.6 (<https://www.postgresql.org> (accessed on 10 October 2020)), MySQL 5.7 (<https://www.mysql.com> (accessed on 10 October 2020)), Apache 2.4 (<https://www.apache.org> (accessed on 10 October 2020)), PHP 7.1, Perl 5.16 (<https://www.perl.org> (accessed on 10 October 2020)), and Python 3.6 (<https://www.python.org/> (accessed on 10 October 2020)). The platform is installed in a computer cluster composed of Dell PowerEdge R920, which provides 1 TB random access memory and 22 TB hard drive capacity. The high performance of the system promises a high speed of analysis on big datasets.

The SilkOrganPDB website was developed using React (<https://reactjs.org> (accessed on 20 October 2020)), which is a JavaScript library supplying a variety of packages and plugins for rendering data and producing friendly interfaces. The BLAST tool was built with NCBI BLAST+ 2.11.0 [41], which supports searching against selectable and multiple databases. The other homologous search tool was built using the HMMER [59], which implements methods using probabilistic models known as profile hidden Markov models. The phylogenetic tool was developed by aligning homologous protein sequences with muscle [60] and then building phylogeny using FastTree [61] with default parameters, the results of which were visualized using software GraPhlAn [62].

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms22095051/s1>, Table S1: The identified nucleus proteins in silkworm BmE cells with engineered ascorbate peroxidase, Table S2: The identified mitochondrial matrix proteins in silkworm BmE cells with engineered ascorbate peroxidase, Table S3: The identified cytosol proteins in silkworm BmE cells with engineered ascorbate peroxidase, Table S4: The identified endoplasmic reticulum proteins in silkworm BmE cells with engineered ascorbate peroxidase, Table S5: The orthology for the organelle proteins of silkworm in multiple insect genomes, Table S6: The orthology for the organelle proteins of silkworm found in all insect genomes selected, Table S7: Silkworm mitochondrial matrix proteins homologous to *D. melanogaster* proteins with multiple subcellular location in ComPPI v2.1.1.

**Author Contributions:** Conceptualization, T.L. and Z.Z.; Experimental Analysis, C.X., J.X., J.L., B.Y., and X.M.; Data Analysis, T.L., C.X., and C.L.; Writing—Original Draft Preparation, T.L., and C.X.; Writing—Review & Editing, T.L., G.P., and Z.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant numbers 31772678, 31770159 and 31472151, and Natural Science Foundation of Chongqing, China, grant number cstc2019yszx-jcyjX0010.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Laboratory Animal Ethics Committee of Southwest University (protocol code IACUA-20190301, March 1, 2019).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org> (accessed on 6 May 2021)) via the iProX partner repository [63] with the dataset identifier PXD025816. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We would like to thank Tongbao Liu for helpful comments on the manuscript. We would like to thank all the authors who published the manuscripts included in this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Xia, Q.; Zhou, Z.; Lu, C.; Cheng, D.; Dai, F.; Li, B.; Zhao, P.; Zha, X.; Cheng, T.; Chai, C.; et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **2004**, *306*, 1937–1940. [[CrossRef](#)] [[PubMed](#)]
- International Silkworm Genome, C. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* **2008**, *38*, 1036–1045. [[CrossRef](#)]
- Xia, Q.; Guo, Y.; Zhang, Z.; Li, D.; Xuan, Z.; Li, Z.; Dai, F.; Li, Y.; Cheng, D.; Li, R.; et al. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* **2009**, *326*, 433–436. [[CrossRef](#)] [[PubMed](#)]
- Kawamoto, M.; Jouraku, A.; Toyoda, A.; Yokoi, K.; Minakuchi, Y.; Katsuma, S.; Fujiyama, A.; Kiuchi, T.; Yamamoto, K.; Shimada, T. High-quality genome assembly of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* **2019**, *107*, 53–62. [[CrossRef](#)] [[PubMed](#)]
- Li, J.; Ye, L.; Lan, T.; Yu, M.; Liang, J.; Zhong, B. Comparative proteomic and phosphoproteomic analysis of the silkworm (*Bombyx mori*) posterior silk gland under high temperature treatment. *Mol. Biol. Rep.* **2012**, *39*, 8447–8456. [[CrossRef](#)]
- Xia, Q.; Cheng, D.; Duan, J.; Wang, G.; Cheng, T.; Zha, X.; Liu, C.; Zhao, P.; Dai, F.; Zhang, Z.; et al. Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. *Genome Biol.* **2007**, *8*, R162. [[CrossRef](#)] [[PubMed](#)]
- Zhang, P.; Aso, Y.; Yamamoto, K.; Banno, Y.; Wang, Y.; Tsuchida, K.; Kawaguchi, Y.; Fujii, H. Proteome analysis of silk gland proteins from the silkworm, *Bombyx mori*. *Proteomics* **2006**, *6*, 2586–2599. [[CrossRef](#)]
- Li, J.; Hosseini Moghaddam, S.H.; Chen, X.; Chen, M.; Zhong, B. Shotgun strategy-based proteome profiling analysis on the head of silkworm *Bombyx mori*. *Amino Acids* **2010**, *39*, 751–761. [[CrossRef](#)]
- Li, J.Y.; Ye, L.P.; Che, J.Q.; Song, J.; You, Z.Y.; Wang, S.H.; Zhong, B.X. Proteome identification of the silkworm middle silk gland. *Data Brief.* **2016**, *6*, 903–907. [[CrossRef](#)]
- Dong, Z.; Zhao, P.; Zhang, Y.; Song, Q.; Zhang, X.; Guo, P.; Wang, D.; Xia, Q. Analysis of proteome dynamics inside the silk gland lumen of *Bombyx mori*. *Sci. Rep.* **2016**, *6*, 21158. [[CrossRef](#)]
- Bovilla, V.R.; Padwal, M.K.; Siripurapu, P.; Basu, B.; Mamillapalli, A. Developmental proteome dynamics of silk glands in the 5th instar larval stage of *Bombyx mori* L. (CSR2xCSR4). *Biochim. Biophys. Acta* **2016**, *1864*, 860–868. [[CrossRef](#)] [[PubMed](#)]
- Jin, Y.; Chen, Y.; Jiang, Y.; Xu, M. Proteome analysis of the silkworm (*Bombyx mori* L.) colleterial gland during different development stages. *Arch. Insect Biochem. Physiol.* **2006**, *61*, 42–50. [[CrossRef](#)]
- Hou, Y.; Zou, Y.; Wang, F.; Gong, J.; Zhong, X.; Xia, Q.; Zhao, P. Comparative analysis of proteome maps of silkworm hemolymph during different developmental stages. *Proteome Sci.* **2010**, *8*, 45. [[CrossRef](#)]
- Li, J.Y.; Chen, X.; Hosseini Moghaddam, S.H.; Chen, M.; Wei, H.; Zhong, B.X. Shotgun proteomics approach to characterizing the embryonic proteome of the silkworm, *Bombyx mori*, at labrum appearance stage. *Insect Mol. Biol.* **2009**, *18*, 649–660. [[CrossRef](#)] [[PubMed](#)]
- Wang, X.; Li, Y.; Liu, Q.; Xia, Q.; Zhao, P. Proteome profile of spinneret from the silkworm, *Bombyx mori*. *Proteomics* **2017**, *17*. [[CrossRef](#)]
- Ye, X.; Tang, X.; Wang, X.; Che, J.; Wu, M.; Liang, J.; Ye, L.; Qian, Q.; Li, J.; You, Z.; et al. Improving Silkworm Genome Annotation Using a Proteogenomics Approach. *J. Proteome Res.* **2019**, *18*, 3009–3019. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y.; Xia, D.; Zhao, Q.; Zhang, G.; Zhang, Y.; Qiu, Z.; Shen, D.; Lu, C. Label-free proteomic analysis of silkworm midgut infected by *Bombyx mori* nuclear polyhedrosis virus. *J. Proteomics* **2019**, *200*, 40–50. [[CrossRef](#)]

18. Wu, P.; Shang, Q.; Huang, H.; Zhang, S.; Zhong, J.; Hou, Q.; Guo, X. Quantitative proteomics analysis provides insight into the biological role of Hsp90 in BmNPV infection in *Bombyx mori*. *J. Proteomics* **2019**, *203*, 103379. [[CrossRef](#)]
19. Lu, D.; Xu, P.; Hou, C.; Gao, K.; Guo, X. Label-free LC-MS/MS proteomic analysis of the hemolymph of silkworm larvae infected with *Beauveria bassiana*. *J. Invertebr. Pathol.* **2019**, *166*, 107227. [[CrossRef](#)]
20. Chen, S.; Dong, Z.; Ren, X.; Zhao, D.; Zhang, Y.; Tang, M.; Han, J.; Ye, L.; Zhao, P. Proteomic Identification of Immune-Related Silkworm Proteins Involved in the Response to Bacterial Infection. *J. Insect Sci.* **2019**, *19*. [[CrossRef](#)]
21. Mao, F.; Lei, J.; Enoch, O.; Wei, M.; Zhao, C.; Quan, Y.; Yu, W. Quantitative proteomics of *Bombyx mori* after BmNPV challenge. *J. Proteomics* **2018**, *181*, 142–151. [[CrossRef](#)] [[PubMed](#)]
22. Yu, H.; Wang, X.; Xu, J.; Ma, Y.; Zhang, S.; Yu, D.; Fei, D.; Muhammad, A. iTRAQ-based quantitative proteomics analysis of molecular mechanisms associated with *Bombyx mori* (Lepidoptera) larval midgut response to BmNPV in susceptible and near-isogenic strains. *J. Proteomics* **2017**, *165*, 35–50. [[CrossRef](#)]
23. Wang, X.Y.; Yu, H.Z.; Xu, J.P.; Zhang, S.Z.; Yu, D.; Liu, M.H.; Wang, L.L. Comparative Subcellular Proteomics Analysis of Susceptible and Near-isogenic Resistant *Bombyx mori* (Lepidoptera) Larval Midgut Response to BmNPV infection. *Sci. Rep.* **2017**, *7*, 45690. [[CrossRef](#)] [[PubMed](#)]
24. Hung, V.; Udeshi, N.D.; Lam, S.S.; Loh, K.H.; Cox, K.J.; Pedram, K.; Carr, S.A.; Ting, A.Y. Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **2016**, *11*, 456–475. [[CrossRef](#)] [[PubMed](#)]
25. Rhee, H.W.; Zou, P.; Udeshi, N.D.; Martell, J.D.; Mootha, V.K.; Carr, S.A.; Ting, A.Y. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, *339*, 1328–1331. [[CrossRef](#)]
26. Li, J.; Han, S.; Li, H.; Udeshi, N.D.; Svinkina, T.; Mani, D.R.; Xu, C.; Guajardo, R.; Xie, Q.; Li, T.; et al. Cell-Surface Proteomic Profiling in the Fly Brain Uncovers Wiring Regulators. *Cell* **2020**, *180*, 373.e15–386.e15. [[CrossRef](#)]
27. Droujinine, I.A.; Meyer, A.S.; Wang, D.; Udeshi, N.D.; Hu, Y.; Rocco, D.; McMahon, J.A.; Yang, R.; Guo, J.; Mu, L.; et al. Proteomics of protein trafficking by in vivo tissue-specific labeling. *Nat. Commun.* **2021**, *12*, 2382. [[CrossRef](#)] [[PubMed](#)]
28. Hwang, J.; Espenshade, P.J. Proximity-dependent biotin labelling in yeast using the engineered ascorbate peroxidase APEX2. *Biochem. J.* **2016**, *473*, 2463–2469. [[CrossRef](#)] [[PubMed](#)]
29. Chen, C.L.; Hu, Y.; Udeshi, N.D.; Lau, T.Y.; Wirtz-Peitz, F.; He, L.; Ting, A.Y.; Carr, S.A.; Perrimon, N. Proteomic mapping in live *Drosophila* tissues using an engineered ascorbate peroxidase. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12093–12098. [[CrossRef](#)]
30. Bersuker, K.; Olzmann, J.A. Identification of Lipid Droplet Proteomes by Proximity Labeling Proteomics Using APEX2. *Methods Mol. Biol.* **2019**, *2008*, 57–72. [[CrossRef](#)]
31. Reinke, A.W.; Mak, R.; Troemel, E.R.; Bennett, E.J. In vivo mapping of tissue- and subcellular-specific proteomes in *Caenorhabditis elegans*. *Sci. Adv.* **2017**, *3*, e1602426. [[CrossRef](#)]
32. Hung, V.; Lam, S.S.; Udeshi, N.D.; Svinkina, T.; Guzman, G.; Mootha, V.K.; Carr, S.A.; Ting, A.Y. Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* **2017**, *6*, e24463. [[CrossRef](#)]
33. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S.; Ami, G.O.H.; Web Presence Working, G. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289. [[CrossRef](#)]
34. O’Brate, A.; Giannakakou, P. The importance of p53 location: Nuclear or cytoplasmic zip code? *Drug Resist. Updat.* **2003**, *6*, 313–322. [[CrossRef](#)] [[PubMed](#)]
35. Ota, M.; Gonja, H.; Koike, R.; Fukuchi, S. Multiple-Localization and Hub Proteins. *PLoS ONE* **2016**, *11*, e0156455. [[CrossRef](#)] [[PubMed](#)]
36. Veres, D.V.; Gyurko, D.M.; Thaler, B.; Szalay, K.Z.; Fazekas, D.; Korcsmaros, T.; Csermely, P. ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **2015**, *43*, D485–d493. [[CrossRef](#)] [[PubMed](#)]
37. Lu, F.; Wei, Z.; Luo, Y.; Guo, H.; Zhang, G.; Xia, Q.; Wang, Y. SilkDB 3.0: Visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Res.* **2020**, *48*, D749–D755. [[CrossRef](#)]
38. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [[CrossRef](#)]
39. Duan, J.; Li, R.; Cheng, D.; Fan, W.; Zha, X.; Cheng, T.; Wu, Y.; Wang, J.; Mita, K.; Xiang, Z.; et al. SilkDB v2.0: A platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res.* **2010**, *38*, D453–D456. [[CrossRef](#)]
40. Choi, H.; Kim, S.; Fermin, D.; Tsou, C.C.; Nesvizhskii, A.I. QPROT: Statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. *J. Proteomics* **2015**, *129*, 121–126. [[CrossRef](#)] [[PubMed](#)]
41. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
42. Calvo, S.E.; Clauser, K.R.; Mootha, V.K. MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **2016**, *44*, D1251–D1257. [[CrossRef](#)] [[PubMed](#)]
43. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [[CrossRef](#)]
44. Xiang, H.; Liu, X.; Li, M.; Zhu, Y.; Wang, L.; Cui, Y.; Liu, L.; Fang, G.; Qian, H.; Xu, A.; et al. The evolutionary road from wild moth to domestic silkworm. *Nat. Ecol. Evol.* **2018**, *2*, 1268–1279. [[CrossRef](#)]
45. Gershman, A.; Romer, T.G.; Fan, Y.; Razaghi, R.; Smith, W.A.; Timp, W. De novo genome assembly of the tobacco hornworm moth (*Manduca sexta*). *G3* **2021**, *11*, 1–9. [[CrossRef](#)]



46. Li, X.; Fan, D.; Zhang, W.; Liu, G.; Zhang, L.; Zhao, L.; Fang, X.; Chen, L.; Dong, Y.; Chen, Y.; et al. Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat. Commun.* **2015**, *6*, 8212. [[CrossRef](#)] [[PubMed](#)]
47. Cheng, T.; Wu, J.; Wu, Y.; Chilukuri, R.V.; Huang, L.; Yamamoto, K.; Feng, L.; Li, W.; Chen, Z.; Guo, H.; et al. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat. Ecol. Evol.* **2017**, *1*, 1747–1756. [[CrossRef](#)]
48. Chen, W.; Yang, X.; Tetreau, G.; Song, X.; Coutu, C.; Hegedus, D.; Blissard, G.; Fei, Z.; Wang, P. A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Mol. Ecol. Resour.* **2019**, *19*, 485–496. [[CrossRef](#)] [[PubMed](#)]
49. Tribolium Genome Sequencing, C.; Richards, S.; Gibbs, R.A.; Weinstock, G.M.; Brown, S.J.; Denell, R.; Beeman, R.W.; Gibbs, R.; Beeman, R.W.; Brown, S.J.; et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **2008**, *452*, 949–955. [[CrossRef](#)] [[PubMed](#)]
50. Nene, V.; Wortman, J.R.; Lawson, D.; Haas, B.; Kodira, C.; Tu, Z.J.; Loftus, B.; Xi, Z.; Megy, K.; Grabherr, M.; et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **2007**, *316*, 1718–1723. [[CrossRef](#)]
51. Hoskins, R.A.; Carlson, J.W.; Wan, K.H.; Park, S.; Mendez, I.; Galle, S.E.; Booth, B.W.; Pfeiffer, B.D.; George, R.A.; Svirskas, R.; et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* **2015**, *25*, 445–458. [[CrossRef](#)]
52. Park, D.; Jung, J.W.; Choi, B.S.; Jayakodi, M.; Lee, J.; Lim, J.; Yu, Y.; Choi, Y.S.; Lee, M.L.; Park, Y.; et al. Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics* **2015**, *16*, 1. [[CrossRef](#)]
53. Wallberg, A.; Bunikis, I.; Pettersson, O.V.; Mosbech, M.B.; Childers, A.K.; Evans, J.D.; Mikheyev, A.S.; Robertson, H.M.; Robinson, G.E.; Webster, M.T. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics* **2019**, *20*, 275. [[CrossRef](#)] [[PubMed](#)]
54. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)]
55. Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5*, R80. [[CrossRef](#)] [[PubMed](#)]
56. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
57. Eilbeck, K.; Lewis, S.E.; Mungall, C.J.; Yandell, M.; Stein, L.; Durbin, R.; Ashburner, M. The Sequence Ontology: A tool for the unification of genome annotations. *Genome Biol.* **2005**, *6*, R44. [[CrossRef](#)]
58. Mungall, C.J.; Emmert, D.B.; FlyBase, C. A Chado case study: An ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* **2007**, *23*, i337–i346. [[CrossRef](#)] [[PubMed](#)]
59. Eddy, S.R. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **2009**, *23*, 205–211.
60. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
61. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
62. Asnicar, F.; Weingart, G.; Tickle, T.L.; Huttenhower, C.; Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **2015**, *3*, e1029. [[CrossRef](#)] [[PubMed](#)]
63. Ma, J.; Chen, T.; Wu, S.; Yang, C.; Bai, M.; Shu, K.; Li, K.; Zhang, G.; Jin, Z.; He, F.; et al. iProX: An integrated proteome resource. *Nucleic Acids Res.* **2019**, *47*, D1211–D1217. [[CrossRef](#)] [[PubMed](#)]