Insect Molecular Biology

Insect Molecular Biology (2009) **18**(5), 649–660 doi: 10.1111/j.1365-2583.2009.00903.x

Shotgun proteomics approach to characterizing the embryonic proteome of the silkworm, *Bombyx mori*, at labrum appearance stage

J-Y. Li*, X. Chen†, S. H. Hosseini Moghaddam*‡, M. Chen†, H. Wei* and B-X. Zhong*

**College of Animal Sciences, Zhejiang University, Hangzhou 310029, P. R. China; †College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China; and ‡Agriculture Faculty, UniversityHangzhou310058,Rasht, Iran*

Abstract

The shotgun approach has gained considerable acknowledgement in recent years as a dominant strategy in proteomics. We observed a dramatic increase of specific protein spots in two-dimensional electrophoresis (2-DE) gels of the silkworm (*Bombyx mori*) embryo at labrum appearance, a characteristic stage during embryonic development of silkworm which is involved with temperature increase by silkworm raiser. We employed shotgun liquid chromatography tandem mass spectrometry (LC-MS/MS) technology to analyse the proteome of *B. mori* embryos at this stage. A total of 2168 proteins were identified with an in-house database. Approximately 47% of them had isoelectric point (p*I*) values distributed theoretically in the range p*I* 5–7 and approximately 60% of them had molecular weights of 15–45 kDa. Furthermore, 111 proteins had an p*I* greater than 10 and were difficult to separate by 2-DE. Many important functional proteins related to embryonic development, stress response, DNA transcription/translation, cell growth, proliferation and differentiation, organogenesis and reproduction were identified. Among them proteins related to nervous system development were noticeable. All known heat shock proteins (HSPs) were detected in this developmental stage of *B. mori* embryo. In addition, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed energetic metabolism at this stage. These results were expected to provide more information for proteomic monitoring of the insect embryo and better understanding of the spatiotemporal expression of genes during embryonic developmental processes.

Keywords: *Bombyx mori*, embryo, proteomics, Gene Ontology, pathway.

Introduction

One of the best model systems for lepidopteran biological studies is the domesticated silkworm, *Bombyx mori*, with its known developmental and physiological characters. There are four morphologically distinct developmental stages in the life cycle of *Bombyx mori*: egg, larva, pupa and adult. In the case of the *B. mori* embryo, diapause is a conspicuous character which appears at the late gastrula stage of embryogenesis, when the embryonic cell cycle becomes arrested in the G2 phase (Nakagaki *et al.*, 1991). Development of the diapaused egg is stopped at the early stage of embryonic growth and then the egg becomes dormant. The dormant embryo is woken up to grow by either warm weather or acid treatment (artificial hatching). The diapaused embryos usually go through 12 morphologically specific developmental stages in 11 days from the time of diapause release to hatching, under the conditions of 22–25 °C and 75–80% relative humidity (RH) (Zhong *et al.*, 2005). These stages are: critical development I, critical development II, neural groove appearance, abdominal outgrowth appearance, labrum appearance, shortening, head thorax differentiation, embryonic reverse, tubercle appearance, head pigmentation, body pigmentation, and hatching. The period from the neural groove appearance to tubercle appearance is known as organogenesis. The labrum appearance is one of the most important stages in embryonic development. During this stage the temperature of the incubator, or room containing silkworm eggs for incubation purposes, should be increased from 22 °C to 25 °C for both bivoltine and polyvoltine eggs of *B. mori* (Lü *et al.*, 1990), as it is vital for the hatching percentage, larval health, silk quantity and quality and also diapause character of *B. mori* eggs in the next generation

Correspondence: Prof. Bo-Xiong Zhong, College of Animal Sciences, Zhejiang University, Hangzhou 310029, P. R. China. Tel/Fax: +86-571-8697-1302; e-mail: bxzhong@zju.edu.cn

(reviewed by Chen, 2000). The morphology of embryos drastically changes after this stage.

The level of gene expression can be predicted by gene expression profiling experiments. For instance, the gene expression profiles of *B. mori* embryo in some developmental stages have been characterized (Hong *et al.*, 2006; Oh *et al.*, 2006). Nevertheless, gene expression profile alone is not sufficient to determine gene functions, and mRNA levels may or may not correlate with the protein level (Gygi *et al.*, 1999; Pandey & Mann, 2000) because of the variety and dynamics of gene translational productions. Proteomics is, therefore, indispensable for entomological research (Collins *et al.*, 2006; Zhou *et al.*, 2008a,b; Li *et al.*, 2009). Identifying the proteome of the insect embryo is an essential step towards monitoring the embryonic developmental processes. Two-dimensional gel electrophoresis (2-DE), combined with mass spectrometry (MS), has been used in insect embryonic research (Zhong, 1999; Zhong *et al.*, 2005; Chen *et al.*, 2008). We have already observed that the specific protein spots in 2-DE gels were greatly increased at embryonic labrum appearance stage (Supplementary Fig. 1). Chen *et al.* (2008) found several larva-related functional proteins were highly expressed at different developmental stages showing logical expression of these genes. Gong *et al.* (2004) identified more than 50 proteins were differentially expressed in *Drosophila* ventralized vs. lateralized embryos using 2-D differential in gel electrophoresis (DIGE).

Although insect embryonic proteins have been reported by many researchers, there are few reports on large-scale characterization of the embryonic proteome. Shotgun proteomics is an eligible method for large-scale screening of peptides and proteins of a complex biological sample in order to generate rapidly a global profile of the protein components in cells, tissues and organs (MacCoss *et al.*, 2002). The liquid chromatography tandem mass spectrometry (LC-MS/MS) is a highly sensitive and high-throughput method based on the shotgun proteomics approach. In a typical LC-MS/MS experiment, proteins in a sample are first digested into peptides, separated by an LC system and then subjected to MS analysis. The peptides and proteins are identified based on the matching of experimentally generated tandem mass spectra to the theoretical best match from a protein database. Recently this technique has been employed for some model insects, such as *Drosophila* (Baggerman *et al.*, 2005) and *Anopheles* (Kalume *et al.*, 2005). In this information approach, a database search is a bottleneck for many shotgun proteomics experiments, especially when the organism database has been not fully developed which is the case for *B. mori*. Fortunately, the large amount of genomic resources available and protein sequences of other organisms provrote-42rChe f

ofr this

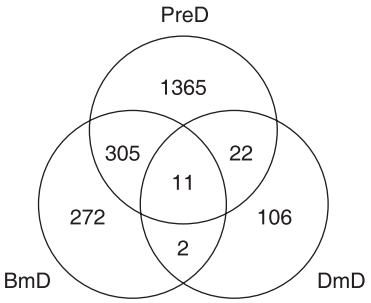


Figure 1. Venn diagram for the numbers of identified protein entries from three sources. The overlap indicates the number of entries that belong to the two or three sources. PreD, protein database of *Bombyx mori* predicted from the genome sequences; BmD, protein database of *Bombyx* from NCBInr; DmD, protein database of *Drospophila melanogaster* from NCBInr.

which protein or proteins were real identifications. To decrease redundancy of the results, these proteins were selected manually according to the explanation in experimental procedures. Finally, 2168 proteins were identified with 1917 overlaps between SEQUEST and X!Tandem (supplementary Table S1). By this method, we identified about 6% (135/2033, 116/2052) more proteins than by only using SEQUEST or X!Tandem.

Comparison of the results from three sources

The protein isoforms, or the homologous proteins from the same or different sources, were always assigned into one entry. Among the 2083 identified protein entries (Fig. 1), 1365 entries (65.5%) belonged to predicted database (PreD). The sequences of these proteins were subjected to BLASTP against the newly *B. mori* protein database predicted from the combining of two genome sequences (Xia *et al.*, 2008). The results showed 23 proteins without matching, 264 proteins with identical sequences, and 66 matched proteins without common peptides used in the identification (supplementary Table S1). In addition, there were 272 unique protein entries from BmD. The common entries of PreD and *Bombyx* database (BmD) (305 entries) were far more than that of PreD and *D. melanogaster* database (DmD) (22 entries) or BmD and DmD (2 entries). This proved the usefulness of *Bombyx* databases for our study, although the number of protein sequences in DmD (20 735) exceeded 13 times that of BmD (1510). In addition, 106 unique protein entries for DmD and 35 entries from shared parts were also found. Thus DmD, combined with the other databases, provided more detected peptides and was more useful.

Theoretical two-dimensional distribution of the identified proteins

The theoretical isoelectric point (p*I*) and molecular weight (MW) of the identified proteins were calculated using the Compute pI/Mw tool (http://cn.expasy.org/tools/pi_tool.html) according to predicted amino acid sequences. Although the

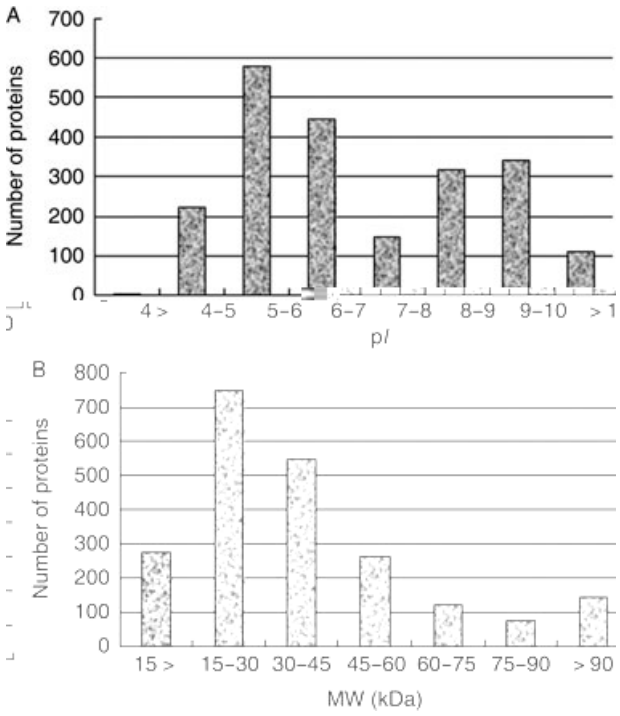


Figure 2. Theoretical two-dimensional distribution of the identified proteins. (A) distribution of p*I*. (B) distribution of MW. The theoretical isoelectric point (p*I*) and molecular weight (MW) of the proteins were calculated using the Compute pI/Mw tool (http://cn.expasy.org/tools/pi_tool.html) according to predicted amino acid sequences.

theoretical p*I* and MW of proteins should be different from the actual amount as a result of the modifications of proteins as well as the quality of databases, especially the PreD, it can give an overview of the distribution of proteome components. It showed that 47% of the total proteins were distributed in a range of p*I* 5–7 (Fig. 2a). About 60% of the proteins had molecular weights in the range 15–45 kDa (Fig. 2b). Moreover, there were 111 proteins with p*I* > 10 which were difficult to separate by 2-DE. From the SDS-PAGE pattern (supplementary Fig. S3) we observed an enrichment protein band at around 30 kDa (25–31 kDa) which might contain a large number of 30 K protein family members. Indeed, we detected 285 proteins in this range.

Embryonic diapause release and development

Embryonic diapause of *B. mori* is primarily determined by the accumulation of sorbitol in eggs (Horie *et al.*, 2000), which is indirectly induced by diapause hormone that secreted from the suboesophageal ganglion of the female pupa and then transferred to her hemolymph during ovarian development. The break of diapause is often by complex mechanisms involved in regulatory cascades of related factors, such as environmental, endocrinous, circadian, transcriptional, and proteomic changes. Termination of *B. mori* embryonic diapause usually requires 2–3 months of

low temperature (5 °C). In our study, the diapause-determined eggs were treated with hydrochloric acid at 46 °C to terminate the diapause. It has been shown that diapause termination requires a decrease in sorbitol concentration and activation of ecdysteroid which are regulated by extracellular signal-regulated kinase (ERK) through enhancing expression of sorbitol dehydrogenase (SDH) and ecdysteroid-phosphate phosphatase (EPPase) genes,

Table 1. Annotated proteins identified at labrum appearance stage of *Bombyx mori* embryos

Protein	Theor. pI/MW(kDa)	Biological Process	Molecular Function
Embryonic development#			
14–3-3zeta	4.90/28.17	activation of tryptophan 5-monooxygenase activity	diacylglycerol-activated phospholipid-dependent protein kinase C inhibitor activity
almondex CG12127-PA*	7.88/31.36	ectoderm development	
alpha Spectrin CG1977-PA*	5.08/27.83	body morphogenesis	
brother of odd with entrails limited CG10021-PD, isoform D*	6.73/79.83	hindgut morphogenesis	RNA polymerase II transcription factor activity
COP9 complex homolog subunit 4 CG8725-PA*	5.93/46.47	compound eye photoreceptor cell differentiation	NEDD8-activating enzyme activity
DnaJ domain-containing protein	9.94/12.75	genitalia development	
fumarylacetoacetate hydrolase	5.93/47.37	embryonic development ending in birth or egg hatching	
innexin 2	6.60/41.81	foregut morphogenesis	gap junction channel activity
low molecular lipoprotein 30 K precursor	6.11/29.73	positive regulation of growth rate	
nemo CG7892-PG, isoform G*	7.29/48.20	compound eye development	protein serine/threonine kinase activity
nonclathrin coat protein zeta 1-COP	4.92/20.61	lens development in camera-type eye	
ovarian serine protease	5.55/20.36	dorsal/ventral axis specification	peptidase activity
Pop2 CG5684-PC, isoform C*	4.96/33.48	muscle development	general RNA polymerase II transcription factor activity
prohibitin protein WPH	6.45/30.08	instar larval or pupal development	
promoting protein	8.37/17.12	mesoderm development	
RAB6A, member RAS oncogene family	5.53/23.63	compound eye morphogenesis	GTPase activity
rotamase Pin1	5.90/19.29	epidermal growth factor receptor signaling pathway	peptidyl-prolyl cis-trans isomerase activity
sex-lethal	9.16/37.49	alternative nuclear mRNA splicing, via spliceosome	growth factor activity
squid CG16901-PC, isoform C*	6.47/33.05	dorsal/ventral axis specification, ovarian follicular epithelium	mRNA 3′-UTR binding
thymosin isoform 1	4.95/19.02	brain development	actin binding
xanthine dehydrogenase	7.78/14.99	compound eye pigmentation	xanthine dehydrogenase activity
Stress responses#			
90-kDa heat shock protein	4.99/82.42	positive regulation of nitric oxide biosynthetic process	nitric-oxide synthase regulator activity
apolipophorin III	9.04/20.73	defense response to Gram-positive bacterium	
c-Jun NH2-terminal kinase	6.49/45.17	antibacterial humoral response	JUN kinase activity
Cu/Zn SOD	5.78/15.84	aging	antioxidant activity
endoplasmic reticulum protein	8.84/24.81	immune response	receptor binding
glutathione S-transferase sigma	5.85/23.34	response to oxidative stress	glutathione peroxidase activity
heat shock cognate protein	5.33/71.18	axon guidance	ATPase activity
heat shock protein 25.4	5.15/25.39	protein stabilization	protein binding
heat shock protein 70	5.47/73.59	determination of adult life span	unfolded protein binding
heat shock protein hsp19.9	6.53/19.89	embryonic development	
heat shock protein hsp20.1	5.46/20.14	embryonic development	
heat shock protein hsp20.4	6.54/20.43	embryonic development	
heat shock protein hsp20.8	5.98/20.80	embryonic development	
heat shock protein hsp21.4	5.79/21.40		protein binding
heat shock protein hsp23.7	5.21/23.57	embryonic development	
Hsc70/*Hsp*90-organizing protein HOP	6.14/62.14	response to stress	
*Hsp*40 protein	9.25/38.93	response to heat	
p38 map kinase	5.93/41.55	defense response	MAP kinase activity
serine protease zymogen (proBAEEase)	5.21/40.74	defense response	peptidase activity
Cell growth, proliferation and differentiation#			
actin-depolymerizing factor 1	6.17/17.01	actin filament depolymerization	actin binding
receptor for activated protein kinase C RACK 1 isoform 1	8.07/36.04	cuticle development	protein kinase C binding
ribonuclease L inhibitor homolog	8.78/69.38	cell growth	ATP binding
selenophosphate synthetase 1	6.27/44.13	cell proliferation	selenide, water dikinase activity
stathmin	7.74/33.48	germ cell migration	microtubule binding
translationally controlled tumor protein	4.66/19.86	positive regulation of cell size	guanyl-nucleotide exchange factor activity

Table 1. *Continued*

Protein	Theor. pI/MW(kDa)	Biological Process	Molecular Function
Spermatogenesis/oogenesis/zygogenesis#			
AGO3 protein	9.12/10.46	spermatogenesis	protein binding
ALY	10.63/27.27		mRNA binding
atypical protein kinase C	5.55/67.32	apical protein localization	atypical protein kinase C activity
infertile crescent CG9078-PA*	9.18/37.21	spermatogenesis	sphingolipid delta-4 desaturase activity
nascent polypeptide associated complex protein alpha subunit	4.65/22.67	oogenesis	protein binding
piwi protein	9.13/10.4	spermatogenesis	protein binding
β-N-acetylglucosaminidase	5.25/61.55		beta-N-acetylhexosaminidase activity
Apoptosis/anti-apoptosis#			
apoptosis-linked protein 2	4.97/20.57	apoptosis	calcium ion binding
double-stranded RNA-binding zinc finger protein JAZ	5.5/48.54	apoptosis	double-stranded RNA binding
Extracellular regulated MAP kinase	5.87/41.99	anti-apoptosis	JUN kinase activity
ras oncoprotein	6.33/21.83	anti-apoptosis	GDP binding
DNA repair#			
14–3-3 epsilon protein	4.66/29.67	DNA damage checkpoint	diacylglycerol-activated phospholipid-dependent protein kinase C inhibitor activity
double-time protein	9.53/39.92	DNA repair	protein binding
RFC40	6.96/37.37	DNA repair	ATP-dependent DNA helicase activity
Sex-determination#			
SNF	9.75/24.02	female germ-line sex determination	mRNA binding
Others			
AP-2sigma CG6056-PA*	5.82/16.96	neurotransmitter secretion	
estrogen sulfotransferase	6.48/37.98		retinol dehydratase activity
mago nashi	6.06/17.24	cell-cell signaling	
PKG-II	5.79/83.98	feeding behavior	cGMP-dependent protein kinase activity
Rtnl1 CG33113-PA, isoform A*	9.18/24.72	olfactory behavior	
vitellogenin precursor	6.85/20.31	determination of adult life span	

#Classifications of the proteins according to the published documents and GO annotation (only two GO categories are shown). In fact, many of the proteins are a multifunction.
*The proteins from DmD.

Spermatogenesis, oogenesis and zygogenesis

Quantitative analyses demonstrated that Piwi subfamily genes *SIWI* and *BmAGO3* were abundantly expressed in the larval testis, pupal ovary, and adult eggs, suggesting that they might be involved in spermatogenesis and oogenesis of *B. mori* (Kawaoka *et al.*, 2008). In *Drosophila*, the Aly protein regulates both male meiotic cell cycle progression and the terminal differentiation of spermiogenesis by activating the transcription of genes required for both processes (White-Cooper *et al.*, 2000). While *Drosophila* atypical protein kinase C (aPKC) is known to play conserved roles in the generation of cell polarity in the germ line, as well as in epithelial and neural precursor cells within the embryo, it is also essential for two key aspects of oocyte determination: the posterior translocation of oocyte specification factors and the posterior establishment of the microtubule organizing center within the oocyte (Wodarz *et al.*, 2000; Cox *et al.*, 2001). In addition, β-N-acetylglucosaminidase is a major glycosidase involved in several physiological processes, such as fertilization, metamorphosis, glycoconjugate degradation, and glycoprotein biosynthesis in insects. The β-N-acetylglucosaminidase of the plasma membrane covering the acrosome functions is a receptor for the glycoconjugates on the egg surface that plays a crucial role in sperm-egg recognition (Perotti *et al.*, 2001). Although these reproduction-related proteins were detected at labrum appearance stage of *B. mori* embryo in this work, they were also observed in other periods of insect life cycle (Kawaoka *et al.*, 2008), suggesting that they may be involved in multiple processes at different stages other than spermatogenesis, oogenesis and zygogenesis.

Effect of incubation temperature on embryonic proteome

The 2-DE protein expression profiles of embryo at different developmental stages were analysed by Image Master software (V2002.01, Amersham Biosciences) and a dramatic increase in the number of specific protein spots was detected at the labrum appearance stage (Fig. 3 and supplementary Fig. S1). Zhong *et al.* (2005) compared and identified 2-D patterns of the 30 K protein family of embryos in normal and temperature-sensitive mutant *B. mori* strains. They found five 30 K proteins, mainly existing in normal strain, implying their possible relation to embryonic development. Among these, we detected proteins in the range MW 25–31 kDa, some of them have connection to the regulation of embryonic development, such as the low

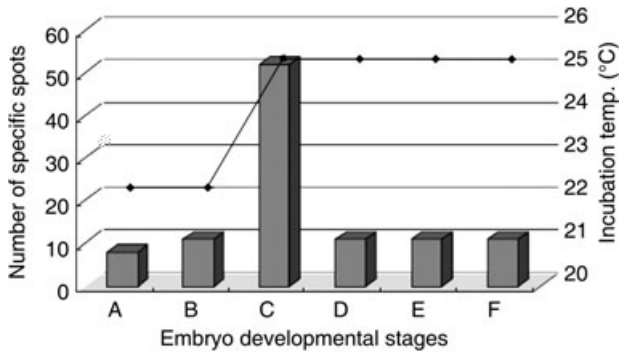


Figure 3. The incubation temperature *vs.* number of specific protein spots in two-dimensional electrophoresis (2-DE) gels of the *Bombyx mori* embryo at different developmental stages. A, neural groove appearance stage; B, abdominal outgrowth appearance stage; C, labrum appearance stage; D, shortening stage; E, head thorax differentiation stage; F, embryonic reverse stage. The histogram shows the number of specific protein spots. The diagram reveals the trends of incubation temperature.

molecular lipoprotein 30 K precursor, 14–3-3 proteins, insulin-related peptide binding protein and juvenile hormone esterase binding protein. Furthermore, the nine proteins detected by Zhong *et al.* (2005) in low expression in temperature-sensitive strain at different developmental stages were in the range of p*I* 5–8 and MW 25–28 kDa. We also found 67 proteins, including 19 annotated from BmD distributed in this range. Interestingly, among the 19 known proteins, three of them were involved in the response to stresses, including C-type lectin, glutathione peroxidase and heat shock protein 25.4. In addition, there were two proteins, casein kinase 2 beta subunit and insulin-related peptide binding protein, which may be involved in the insulin signalling pathway (Coverley *et al.*, 2000). All these proteins may play important roles in normal embryonic development even in an adverse environment. Therefore, the LC-MS/MS is more useful to an overall understanding of the molecular mechanism of embryonic development.

Gene Ontology functional categories

To further understand the functions of the identified proteins, they were functionally categorized based on universal GO annotation terms (Ashburner *et al.*, 2000) using the online GO tool WEGO [Web Gene Ontology Annotation Plot, (http://wego.genomics.org.cn/)]. Among the identified proteins, 2053 ones with annotation terms were linked to the GO cellular component, molecular function and biological process categories (Fig. 4). From the cellular component categories we could learn the subcellular location of identified proteins that 150 were annotated as membrane proteins, 321 located on intracellular organelles and 174 came from membrane-bounded organelles (supplementary Table S3). Among the 278 macromolecular complexes, there were 126 ribonucleoprotein complexes and 7 protein-DNA complexes. Molecular function ontology showed that the binding and catalytic activity proteins were 43.7% (897) and 34.2% (702), respectively. A large number of proteins showed nucleotide binding (332), nucleic acid binding (287), protein binding (154) and ion binding (146). Therefore they probably were a major transporter of material for protein biosynthesis. In addition, the majority of proteins possessed hydrolase activity (253), oxidoreductase activity (155) or transferase activity (143), which exhibited vigorous metabolic action in the cells. With regard to the biological process, 648 (31.6%) proteins were involved in cellular metabolic process. Moreover, there were 108 proteins involved in development, of which 101 proteins were related to anatomical structure morphogenesis, organization and

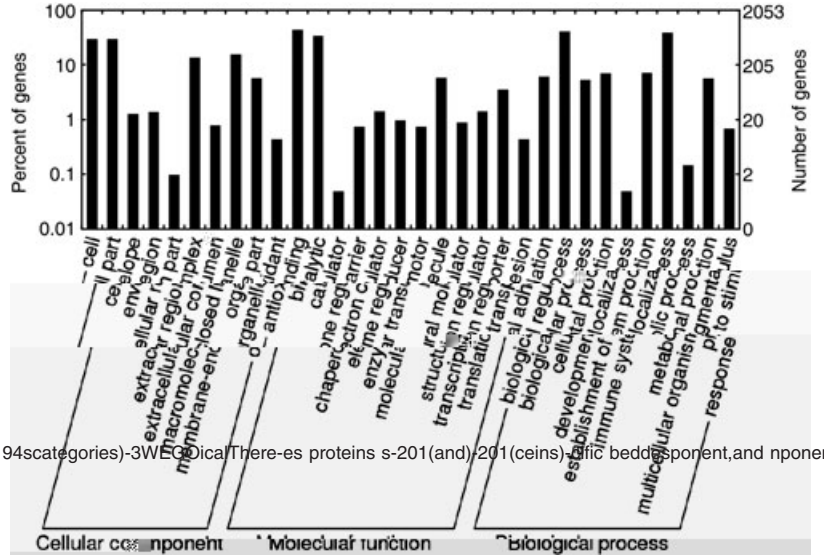


Figure 4. Gene ontology (GO) categories of the identified proteins. The identified proteins were classified into cellular component, molecular function and biological process categories b(into)-333.6(c.4(the)-194scategories)-3WEGOical There-es proteins s-201(and)-201(ceins)-ific bedds ponent,and nponent, proteinst-19-3

development. This is consistent with extensive morphological changes after the labrum appearance stage.

KEGG pathway analysis

KEGG is a knowledge base for systematic analysis of gene functions, linking genomic information with higher order functional information which is stored in the PATHWAY database (Kanehisa & Goto, 2000). Among the 2168 identified proteins, 534 proteins had the matched Enzyme Commission (EC) numbers with the *E*-value $\leq e^{-15}$. These enzymes were involved in 128 KEGG pathways with at least three EC numbers in each, including 20 signalling pathways such as insulin signalling, MAPK signalling and Calcium signalling. Insulin signalling pathway proteins not o

complex for protein assembly and it mainly depends on bioinformatic methods. With the development of genomics and bioinformatics, the shotgun LC-MS/MS will be a promising strategy in proteomics research.

Experimental procedures

Embryos collection

The diapause-determined eggs of polyvoltine *B. mori* strain P50 were exposed to hydrochloric acid (specific gravity 1.10, 20.01% HCl) 20 h after egg laying at 46 °C for 5 min in order to prevent entering of diapause. The diapause-relieved eggs were incubated at 22 °C, with 75% RH for 4 days and then at 25 °C with 75–80% RH in an incubator. Embryos were staged by careful observation of morphological criteria (Lu *et al.*, 1981) under a stereoscopic microscope (Olympus SZX16, Japan). Embryos were examined at different stages from neural groove appearance to embryonic reverse. To this end, eggs were immersed in the boiling potassium hydroxide solution (20% KOH, m/v) for about 2 s until the colour turned to mahogany. They were then washed thrice with pre-cold ddH_2O and were exposed to a gentle stream of air to remove the egg shells. The collected embryos in the Eppendorf tubes (Axygen Scientific, CA, USA) were washed with 95% ethanol. After centrifugation the samples were stored at –20 °C for further use.

Sample preparation and SDS-PAGE and 2-DE separation

The total protein of embryos was extracted on ice in 10 μl lysis buffer (containing 2.5% SDS, 10% glycerin, 5% β-mercaptoethanol and 62.5 mM Tris-HCl pH 6.8) per mg of sample weight with a motor-driven plastic grinder. The homogenate extraction was kept for 10 min at room temperature and then subjected to four times continued sonication treatment in an ice-bath, each time 30 s with a 30 s interval. After centrifugation twice at 20 000 *g*, 25 °C for 10 min, the supernatant was aliquoted and stored at –20 °C. The sample was boiled for 2 min and centrifuged at 20 000 *g* for 10 min before subjected to SDS-PAGE. The sample was separated by SDS-PAGE using 5% stacking gel and 12.5% resolving gel with constant current. Electrophorized gel was stained with Coomassie Brilliant Blue R250 (CBB-R250, Sigma St. Louis, MO, USA). For 2-DE, the sample preparation and separation were carried out according to the methods described by Zhou *et al.* (2008b). The samples were firstly separated by 24 cm, pH 3–10 IEF strips and then were transferred to 15% SDS-PAGE gels for 2nd dimensional separation.

In-gel digestion

Each gel lane was manually cut into 6 bands according to the deepness of Coomasie staining (supplementary Fig. S3), and each band was diced into small pieces (~2 mm^2). Then they were subjected to in-gel tryptic digestion process as described by Shevchenko *et al.* (2006). Briefly, the gel pieces were washed thrice using MilliQ water (Millipore, Bedford, MA, USA) and de-stained twice with 25 mM NH_4HCO_3 in 50% acetonitrile (ACN, Amersham, UK) at 37 °C until the colour disappeared completely. Reduction and alkylation of proteins were performed by incubating the gel pieces with 50 mM Tris[2-carboxyethyl]phosphine (TCEP, Sigma, St Louis, MO, USA) in 25 mM NH_4HCO_3 at 56 °C for 1 h followed by 100 mM iodoacetamide (IAA, Amersham) in 25 mM NH_4HCO_3 for 0.5 h at room temperature in the dark. Gel pieces were washed twice with 25 mM ammonium bicarbonate in 50% acetonitrile solution, dehydrated twice with 100% acetonitrile, and dried in a vacuum centrifuge. The proteins were digested with 20 ng/μl modified proteomics grade trypsin (Sigma) overnight at 37 °C. The resulting tryptic peptides mixtures were extracted twice from the gel pieces with 5% trifluoroacetic acid (TFA, Fluka, Milwaukee, WI, USA) in 50% ACN solution. The pooled extracts were evaporated in the vacuum centrifuge, and re-suspended with 0.1% methanoic acid (Sigma) prior to LC-MS/MS analysis.

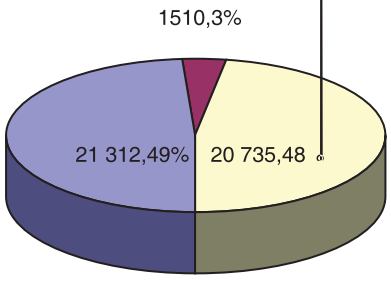
Shotgun LC-MS/MS analysis

All digested peptide mixtures were separated by reverse phase (RP) HPLC followed by tandem MS analysis. RP-HPLC was performed on a surveyor LC system (Thermo Finnigan, San Jose, CA). Samples were loaded into a trap column (Zorbax 300SB-C18 peptide traps, 300 μm × 65 mm, Agilent Technologies, Wilmington, DE) at a 3 μl/min flow rate before the split. After flow-splitting down to about 1.5 μl/min, peptides were transferred to the analytical column (RP-C18, 150 μm × 150 mm, Column Technology Inc., Fremont, CA) for separation with a 195-min linear gradient from 96% buffer A (0.1% methanoic acid in water) to 50% buffer B (84% ACN, 0.1% methanoic acid in water) at a flow rate of 250 nl/min.

Spectra were acquired by LTQ linear ion trap mass spectrometer (Thermo Electron Corporation) in data-dependent mode using Xcalibur software. The mass spectrometer was operated in positive ion mode employing collision-induced dissociation (CID) with a source temperature of 160 °C. The spray voltage was 3.0 kV and normalized collision energy was set at 35.0% for MS/MS. The MS analysis was performed with one full MS scan (m/z 400–1800) followed by ten MS/MS scans on the ten most intense ions from the MS spectrum with the dynamic exclusion settings: repeat count 2, repeat duration 30 s, exclusion duration 90 s.

In-house database search by using TurboSEQUEST and X!Tandem

An in-house database was constructed (Fig. 6) with the protein sequences downloaded from NCBInr protein database (http://www.ncbi.nlm.nih.gov/) including the domesticated silkworm (*B. mori*), the wild silkworm (*B. mandarina*) and the fruit fly (*D. melanogaster*), and the data published by Xia *et al.* (2004) based on



prediction from the *B. mori* genome sequences. The PreD is the database of proteins predicted from the genome sequences of the *B. mori*. The BmD is the protein database of *Bombyx* including *B. mori* and *B. mandarina*. The DmD is the protein database of *D. melanogaster*. The protein sequences in BmD and DmD were all from NCBInr.

The six raw datasets from LC-MS/MS were searched against the in-house database on a local server using the TurboSEQUEST (Bioworks version 3.2, Thermo Electron) and an open-source software X!Tandem, separately. The monoisotopic precursors and fragments were selected for the SEQUEST search. Besides, the mass tolerances of precursor ion and fragmentation ion were set to 1.5 Da and 1.0 Da, respectively. The trypsin-cleavage was at both ends of the protein and two missing cleavage sites were allowed. Only *b* and *y* fragment ions were taken into account. Static (carbamidomethyl) modification on cysteine, and variable modifications (oxidation) on methionine were set for all searches. The results for each dataset were stored as .out format files. The parameters for X!Tandem search were the same to that for SEQUEST.

The target-decoy database was constructed from the original database using an easy-to-use DecoyDDB software tool (Reidegeld *et al.*, 2008). The target-decoy database is a combination of the original (target) and reversed (decoy) databases. The raw data were searched against the target-decoy database under the same parameters as above.

Validation and combination of the database search results

The identified peptides from the database (original and target-decoy databases) search were subjected to TPP (v4.0 JETSTREAM rev 2) for further validation according to the manual. The peptide and protein probability thresholds for running PeptideProphet and ProteinProphet were both set at 0.9. The FDR (FDR = number of validated decoy entries/number of the total entries $\times$ 100) of the final results was evaluated using the data from target-decoy database searching. The final results protXML files from SEQUEST and X!Tandem were subjected to CPAS (LabKey Server 8.2) to compare and export in an .xls file. The criterion for acceptance of the identified proteins was the a probability of more than 0.9 in both or either algorithm. Because of similarity between proteins of the different databases, if multiple proteins from different sources were assigned into a single protein entry by ProteinProphet, only the protein or proteins of one source were accepted according to the higher priority of BmD, DmD and PreD respectively. If these homologous proteins were from the same source, for example, the isoforms generated by alternative splicing, all of them were reported in the final protein list.

Bioinformatics analysis

The new version of protein database predicted by the Beijing Genome Institute (BGI) from the recently combined genome sequences was downloaded from http://silkworm.swu.edu.cn/silkdb/, and used for reciprocal BLASTP searches to identify the BGI-predicted protein corresponding to each Bmb prediction from PreD. The signatures of all identified proteins were queried against InterPro member databases by InterProScan searching (http://www.ebi.ac.uk/InterProScan/). GO classification of the matched proteins was conducted with WEGO (http://wego.genomics.org.cn/) according to the method described by Ye *et al.* (2006). The EC numbers of the identified proteins were acquired (if available) with *E*-value $\leq e^{-15}$ by using KEGG GENES BLASTP (http://blast.genome.jp/) and then were subjected to search against KEGG reference pathway database (http://www.genome.jp/kegg/tool/search_pathway.html). The pathways with at least three EC numbers were accepted and classified according to the definition of KEGG (http://www.genome.ad.jp/kegg/pathway.html).

Acknowledgements

This work was supported by the National Basic Research Program of China (Grant No. 2005CB121003), National Hi-Tech Research and Development Program of China (Grant No. 2006AA10A118), and Program for New Century Excellent Talents in University. We thank the staff of the Research Center for Proteome Analysis, the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences and the Chinese Academy of Sciences for their technological supports on MS analysis. We are thankful to the 985-Institute of Agrobiology and Environmental Sciences of Zhejiang University. We are grateful for the reviewer's careful review and helpful suggestions. We also thank Miss Tianyun Lan and Mr. Wei Fan in our laboratory for their kind help in *B. mori* embryonic incubation and morphological observation.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.

Baggerman, G., Boonen, K., Verleyen, P., De Loof, A. and Schoofs, L. (2005) Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J Mass Spectrom* **40**: 250–260.

Beere, H.M., Wolf, B.B., Cain, K., Mosser, D.D., Mahboubi, A., Kuwana, T. *et al.* (2000) Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. *Nat Cell Biol* **2**: 469–475.

Chen, Y.Y. (2000) Technical standards for commercial egg-production of bivoltine silkworm varieties used in China. *Sericologia* **40**: 185–201.

Chen, J.E., Yang, H.J., Li, J.Y., Shen, Y.W. and Zhong, B.X. (2008) Larva-relevant genes expression in embryonic development stages of domesticated silkworm (*Bombyx mori* L.). *J Zhejiang Univ Agric Life Sci* **34**: 38–44.

Collins, A.M., Caperna, T.J., Williams, V., Garrett, W.M. and Evans, J.D. (2006) Proteomic analyses of male contributions to honey bee sperm storage and mating. *Insect Mol Biol* **15**: 541–549.

Coverley, J.A., Martin, J.L. and Baxter, R.C. (2000) The effect of phosphorylation by casein kinase 2 on the activity of insulin-like growth factor-binding protein-3. *Endocrinology* **141**: 564–570.

Cox, D.N., Seyfried, S.A., Jan, L.Y. and Jan, Y.N. (2001) Bazooka and atypical protein kinase C are required to regulate oocyte differentiation in the *Drosophila* ovary. *Proc Natl Acad Sci U S A* **98**: 14475–14480.

Elias, J.E. and Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**: 207–214.

Fernandez, R., Tabarini, D., Azpiazu, N., Frasch, M. and Schlessinger, J. (1995) The *Drosophila* insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. *EMBO J* **14**: 3373–3384.

Freitak, D., Wheat, C.W., Heckel, D.G. and Vogel, H. (2007) Immune system responses and fitness costs associated with consumption of bacteria in larvae of *Trichoplusia ni*. *BMC Biol* **5**: 56.

Fujiwara, Y., Tanaka, Y., Iwata, K., Rubio, R.O., Yaginuma, T., Yamashita, O. *et al.* (2006a) ERK/MAPK regulates ecdysteroid and sorbitol metabolism for embryonic diapause termination in the silkworm, *Bombyx mori*. *J Insect Physiol* **52**: 569–575.

Fujiwara, Y., Shindome, C., Takeda, M. and Shiomi, K. (2006b) The roles of ERK and P38 MAPK signaling cascades on embryonic diapause initiation and termination of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* **36**: 47–53.

Gong, L., Puri, M., Unlü, M., Young, M., Robertson, K., Viswanathan, S. *et al.* (2004) *Drosophila* ventral furrow morphogenesis: a proteomic analysis. *Development* **131**: 643–656.

Gygi, S., Rochon, Y., Franza, B.R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720–1730.

Hong, S.M., Nho, S.K., Kim, N.S., Lee, J.S. and Kang, S.W. (2006) Gene expression profiling in the silkworm, *Bombyx mori*, during early embryonic development. *Zoolog Sci* **23**: 517–528.

Horie, Y., Kanda, T. and Mochida, Y. (2000) Sorbitol as an arrester of embryonic development in diapausing eggs of the silkworm, *Bombyx mori*. *J Insect Physiol* **46**: 1009–1016.

Hosseini Moghaddam, S.H., Du, X., Li, J., Cao, J.R., Zhong, B.X. and Chen, Y.Y. (2008) Proteome analysis on differentially expressed proteins of the fat body of two silkworm breeds, *Bombyx mori*, exposed to heat shock exposure. *Biotechnol Bioprocess Eng* **13**: 624–631.

Kalume, D.E., Okulate, M., Zhong, J., Reddy, R., Suresh, S., Deshpande, N. *et al.* (2005) A proteomic analysis of salivary glands of female *Anopheles gambiae* mosquito. *Proteomics* **5**: 3765–3777.

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27–30.

Kawaoka, S., Minami, K., Katsuma, S., Mita, K. and Shimada, T. (2008) Developmentally synchronized expression of two *Bombyx mori* Piwi subfamily genes, *SIWI* and *BmAGO3* in germ-line cells. *Biochem Biophys Res Commun* **367**: 755–760.

Keller, A., Eng, J., Zhang, N., Li, X.J. and Aebersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **1**: 1–8.

Kyriakis, J.M. and Avruch, J. (1996) Sounding the alarm: protein kinase cascades activated by stress and inflammation. *J Biol Chem* **271**: 24313–24316.

Li, J.Y., Chen, X., Fan, W., Hosseini Moghaddam, S.H., Chen, M., Zhou, Z.H. *et al.* (2009) Proteomic and bioinformatic analysis on endocrine organs of domesticated silkworm, *Bombyx mori* L. for a comprehensive understanding of their roles and relations. *J Proteome Res* **8**: 2620–2632.

Lohmann, C.M.F. and Riddiford, L.M. (1992) The heat-shock response and heat sensitivity of *Bombyx mori*. *Sericologia* **32**: 533–537.

Lü, H.S. *et al.* (1990) *The sericultural science in China* (in Chinese). Shanghai Scientific and Technical Publisher, Shanghai.

Lu, X.Y. *et al.* (1981) *Silkworm Genetics & Breeding* (in Chinese). Science Press, Beijing.

MacCoss, M.J., McDonald, W.H., Saraf, A., Sadygov, R., Clark, J.M., Tasto, J.J. *et al.* (2002) Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **99**: 7900–7905.

Nakae, J., Kido, Y. and Accili, D. (2001) Distinct and overlapping functions of insulin and IGF-I receptors. *Endocr Rev* **22**: 818–835.

Nakagaki, M., Takei, R., Nagashima, E. and Yaginuma, T. (1991) Cell cycles in embryos of the silkworm, *Bombyx mori*: G2-arrest at diapause stage. *Roux's Arch Dev Biol* **200**: 223–229.

Nesvizhskii, A.I., Keller, A., Kolker, E. and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**: 4646–4658.

Oh, J.H., Jeon, Y.J., Jeong, S.Y., Hong, S.M., Lee, J.S., Nho, S.K. *et al.* (2006) Gene expression profiling between embryonic and larval stages of the silkworm, *Bombyx mori*. *Biochem Biophys Res Commun* **343**: 864–872.

Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405**: 837–846.

Peng, J. and Gygi, S.P. (2001) Proteomics: the move to mixtures. *J Mass Spectrom* **36**: 1083–1091.

Perotti, M.E., Cattaneo, F., Pasini, M.E., Vernì, F. and Hackstein, J.H. (2001) Male sterile mutant *cas*ANOVA gives clues to mechanisms of sperm-egg interactions in *Drosophila melanogaster*. *Mol Reprod Dev* **60**: 248–259.

Picard, D. (2002) Heat-shock protein 90, a chaperone for folding and regulation. *Cell Mol Life Sci* **59**: 1640–1648.

Pimentel, B., de la Rosa, E.J. and de Pablo, F. (1996) Insulin acts as an embryonic growth factor for *Drosophila* neural cells. *Biochem Biophys Res Commun* **226**: 855–861.

Rauch, A., Bellew, M., Eng, J., Fitzgibbon, M., Holzman, T., Hussey, P. *et al.* (2006) Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J Proteome Res* **5**: 112–121.

Reidegeld, K.A., Eisenacher, M., Kohl, M., Chamrad, D., Körting, G., Blüggel, M. *et al.* (2008) An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. *Proteomics* **8**: 1129–1137.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856–2860.

Skoneczna, A., Miciałkiewicz, A. and Skoneczny, M. (2007) *Saccharomyces cerevisiae Hsp*31p, a stress response protein conferring protection against reactive oxygen species. *Free Radic Biol Med* **42**: 1409–1420.

Sobel, A. (1991) Stathmin: a relay phosphoprotein for multiple signal transduction? *Trends Biochem Sci* **16**: 301–305.

Thomas, H.E., Stunnenberg, H.G. and Stewart, A.F. (1993) Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. *Nature* **362**: 471–475.

White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M.T. (2000) Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* **127**: 5463–5473.

Wodarz, A., Ramrath, A., Grimm, A. and Knust, E. (2000) *Drosophila* atypical protein kinase C associates with Bazooka and

controls polarity of epithelia and neuroblasts. *J Cell Biol* **150**: 1361–1374.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B. *et al*. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.

Xia, Q., Wang, J., Zhou, Z., Li, R., Fan, W., Cheng, D. *et al.* (2008) The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* **38**: 1036–1045.

Yan, J. and Hales, B.F. (2008) p38 and c-Jun N-terminal kinase mitogen-activated protein kinase signaling pathways play distinct roles in the response of organogenesis-stage embryos to a teratogen. *J Pharmacol Exp Ther* **326**: 764–772.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z. *et al.* (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* **34**: W293–297.

Yu, X.Q., Gan, H. and Kanost, M.R. (1999) Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem Mol Biol* **29**: 585–597.

Zheng, J.Q. and Poo, M.M. (2007) Calcium signaling in neuronal motility. *Annu Rev Cell Dev Biol* **23**: 375–404.

Zhong, B.X. (1999) Characterization and stage-specific change of proteins during the embryonic development of silkworm *Bombyx mori* (in Chinese). *Acta Genetica Sinica* **26**: 627–633.

Zhong, B.X., Li, J.K., Lin, J.R., Liang, J.S., Su, S.K., Xu, H.S. *et al.* (2005) Possible effect of 30 K proteins in embryonic development of silkworm *Bombyx mori*. *Acta Biochim Biophys Sin* **37**: 355–361.

Zhou, Z.H., Yang, H.J. and Zhong, B.X. (2008a) From genome to proteome: great progress in the domesticated silkworm (*Bombyx mori* L.). *Acta Biochim Biophys Sin* **40**: 601–611.

Zhou, Z.H., Yang, H.J., Chen, M., Lou, C.F., Zhang, Y.Z., Chen, K.P. *et al.* (2008b) Comparative proteomic analysis between the domesticated silkworm (*Bombyx mori*) reared on fresh mulberry leaves and on artificial diet. *J Proteome Res* **7**: 5103–5111.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. 2-DE patterns of the embryos at different developmental stages. A, neural groove appearance stage; B, abdominal outgrowth appearance stage; C, labrum appearance stage; D, shortening stage; E, head thorax differentiation stage; F, embryonic reverse stage.

Figure S2. Two-dimensional density plots of the identified peptides by SEQUEST from the six RAW datasets. The six raw datasets from LC-MS/MS were searched against the in-house database using TurboSEQUEST followed by TPP validation. The minimum probability of the peptides is 0.9. The two-dimensional density plots of the identified peptides could be viewed by a TPP component.

Figure S3. SDS-PAGE pattern of the *Bombyx. mori* embryo at labrum appearance stage. The sample was separated by 12.5% resolving gel. Each gel lane was cut into six bands.

Table S1. Detailed information of the identified peptides and proteins using SEQUEST and X!Tandem followed by TPP validation

Table S2. Identified proteins of *Bombyx mori* embryo at labrum appearance stage from PreD

Table S3. The Gene Ontology (GO) categories of the identified *Bombyx mori* embryonic proteome using WEGO. The GO terms of the main categories are indicated by bold font and the subcategories by non-bold

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

