

Proteomic Analysis of Histone Crotonylation Suggests Diverse Functions in *Myzus persicae*

Manlin Xu,[∇] Yi Xie,[∇] Ying Li, Lili Shen, Kun Huang, Zhonglong Lin, Bin Li, Changjian Xia, Xia Zhang, Yucheng Chi, Bin Zhang,* and Jinguang Yang*



Cite This: *ACS Omega* 2021, 6, 16391–16401



Read Online

ACCESS |



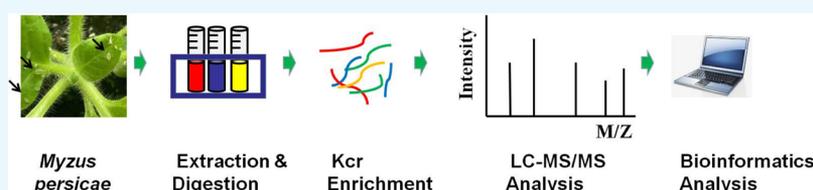
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: *Myzus persicae* is one of the most important economic pests of cultivated crops. In the present study, we used an integrated approach involving high-performance liquid chromatography fractionation, affinity enrichment, and mass spectrometry-based proteomics to carry out a comprehensive proteomic analysis of lysine crotonylation in *M. persicae*. Altogether, 7530 lysine crotonylation sites were identified in 2452 protein groups. Intensive bioinformatic analyses were then carried out to annotate those lysine crotonylated targets identified in terms of Gene Ontology annotation, domain annotation, subcellular localization, Kyoto Encyclopedia of Genes and Genomes pathway annotation, functional cluster analysis, etc. Analysis results showed that lysine-crotonylated proteins were involved in many biological processes, such as the amino acid metabolism, aminoacyl-tRNA biosynthesis, spliceosomes, ribosomes, and so forth. Notably, the interaction network showed that there were 199 crotonylated proteins involved in the amino acid metabolism and numerous crotonylation targets associated with fatty acid biosynthesis and degradation. The results provide a system-wide view of the entire *M. persicae* crotonylome and a rich data set for functional analysis of crotonylated proteins in this economically important pest, which marks an important beginning for the further research.

1. INTRODUCTION

In the nucleosome, two molecules each of the core histones H2A, H2B, H3, and H4 form a histone octamer, around which about 145–147 base pairs of deoxyribonucleic acid (DNA) are wrapped (to form the nucleosome “core”) in about 1^{3/4} supercoiled left-handed turns.^{1–4} Histone post-translational modifications (PTMs) play an important role in the epigenetic regulation of the chromatin structure, in which the functions of eukaryotic genomic DNA, such as transcription, cell differentiation, and organismal development, are regulated by the dynamics of the nucleosome structure.⁵ Many PTMs, such as phosphorylation, lysine acylation, ubiquitylation, succinylation, crotonylation, malonylation, 2-hydroxyisobutyrylation, and β -hydroxybutyrylation, can change protein functions by modulating diverse cellular processes.⁶ In recent years, histone lysine crotonylation (Kcr) has been identified and described as a novel evolutionarily conserved PTM, which is primarily associated with transcriptional regulation and which is mechanistically and functionally distinct from histone lysine acetylation (Kac).^{7–9}

Kcr is widespread from yeast to humans and is enriched at active gene promoters and potentially at enhancers.⁷ There exist many similarities between histone Kac and Kcr; for example, crotonylation was also found on the ϵ -amino group on the side chain of lysine residues and can also be catalyzed

by p300/CBP, a well-known histone acetyltransferase, while crotonylation overlaps with acetylation and other types of acylations mediated by acyltransferases and de-acylases, respectively.^{10,11} Nevertheless, many important structural and functional markers distinguish crotonylation from acetylation, including planar orientation and four-carbon length,^{7,12} preferentially marked “escapee genes” during post-meiotic sex inactivation in mouse testes,⁹ and several Kcr-specific and selective readers.^{13–15} Moreover, histone crotonylation catalyzed by p300, which is probably metabolically regulated,^{16,17} stimulates a more potent transcriptional activator than does p300-catalyzed acetylation.¹² Kcr was first reported in the genomes of human somatic cells and mouse male germ cells, which have unique structures and genomic localizations, and it is a specific marker of active sex chromosome linked genes in post-meiotic male germ cells.⁷ Recently, research has been

Received: March 4, 2021

Accepted: June 10, 2021

Published: June 18, 2021



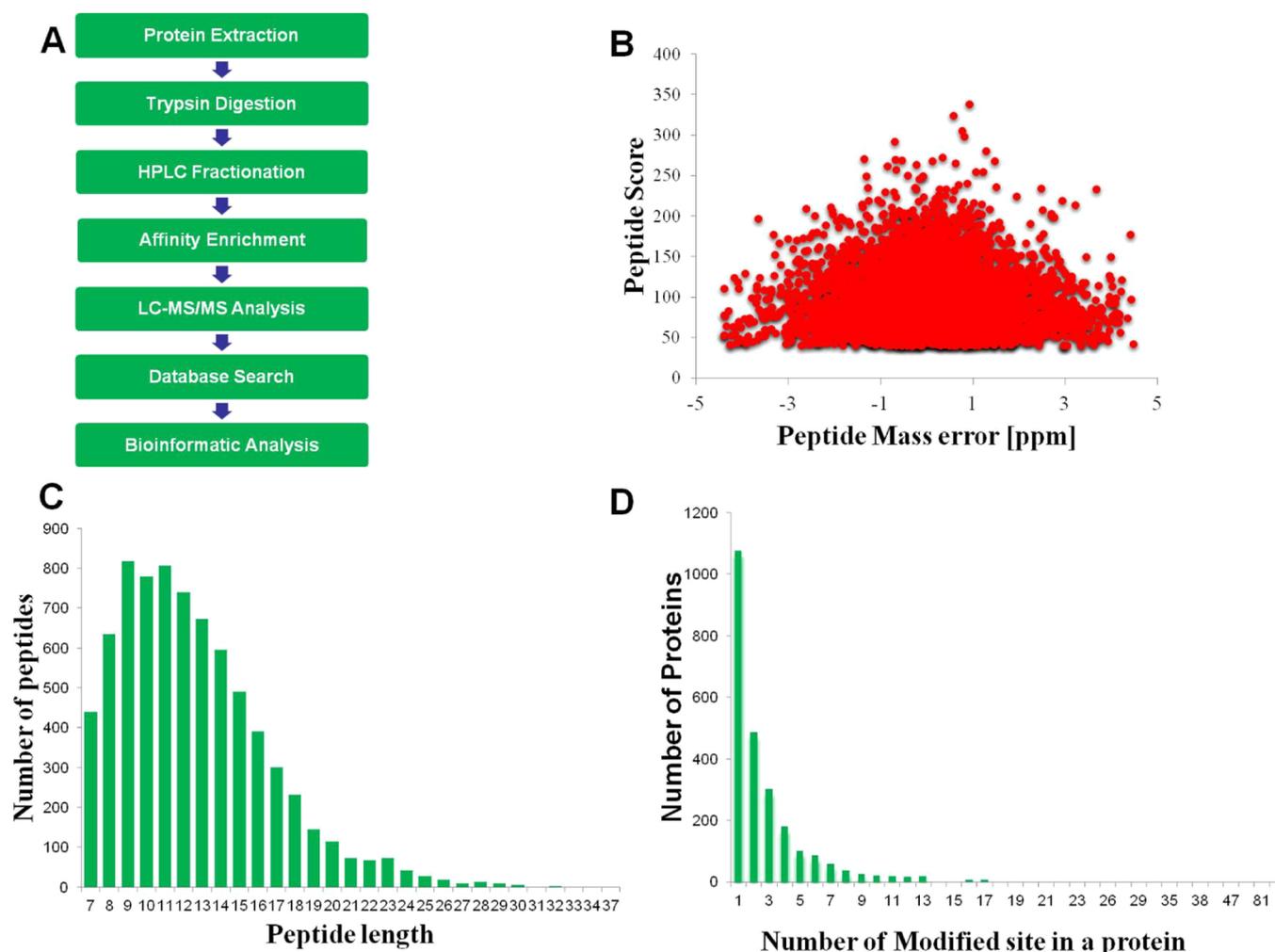


Figure 1. Proteome-wide identification of lysine crotonylation sites in *M. persicae*. (A) Overview of experimental procedures used in this study. (B) Mass error distribution of all identified peptides. (C) Peptide length distribution. (D) Number of modified sites in a protein.

published on histone lysine crotonylation in mammals,^{18,19} fish,²⁰ yeast,²¹ fungi,²² and plants^{23–25} but none on insects.

The peach potato or green peach aphid (*Myzus persicae*) is one of the most economically important pests of cultivated crops.²⁶ *M. persicae* is a significant herbivorous pest with a worldwide distribution (showing a preference for cold and wet environments), including China. It is a polyphagous species with a host range of more than 400 species in 40 different plant families, including many economically important crop plants. Like other aphids, *M. persicae* causes damage to its host by piercing the plant tissue to access the phloem for direct feeding, by the transmission of plant viruses (it is capable of transmitting over 100 different plant viruses), and by the production of honeydew.²⁶ The exceptional ability of *M. persicae* to adapt to new host plants, to disperse, and to evolve resistance to insecticides has further enhanced the status of this species as a serious pest.²⁷ In the present study, we carry out a comprehensive proteomic analysis of lysine crotonylation in *M. persicae*. Altogether, 7530 lysine crotonylation sites were identified in 2452 protein groups and revealed that crotonylation widely exists in *M. persicae*. We analyzed the lysine crotonylation sites and proteins, characteristics of the crotonylation site motif, the functional classification and enrichment, proteins involved in the central metabolism and the protein–protein interaction (PPI). The results provide a

system-wide view of the *M. persicae* whole crotonylome and offer an interesting perspective as well as a potential application as a novel control strategy by which this pest is managed.

2. RESULTS

2.1. Proteomic Analysis of Lysine Crotonylation Sites and Proteins in *M. persicae*. In order to map lysine crotonylation sites in *M. persicae*, an integrated approach involving high-performance liquid chromatography (HPLC) fractionation, affinity enrichment, and mass spectrometry-based proteomics was used to identify the whole crotonylome of *M. persicae* (Figure 1A). A total of 7530 lysine crotonylation sites were identified distributed in 2452 protein groups (Tables S1 and S2). A total of 42,249 proteins have been estimated as being encoded by the *M. persicae* genome [uniprot_133076_Aphidinae (<https://www.uniprot.org>)], which indicated that the crotonylated proteins identified accounted for 5.8% (2452/42,249) of the proteome. The mass errors of all the identified peptides were checked to validate the MS data, and the results are shown in Figure 1B. The distribution of the mass error was near 0, and most of them were less than 0.02 Da, which implied that the mass accuracy of the MS data fit the requirement. The length of most peptides was distributed between 8 and 20, a range which agreed with the properties of

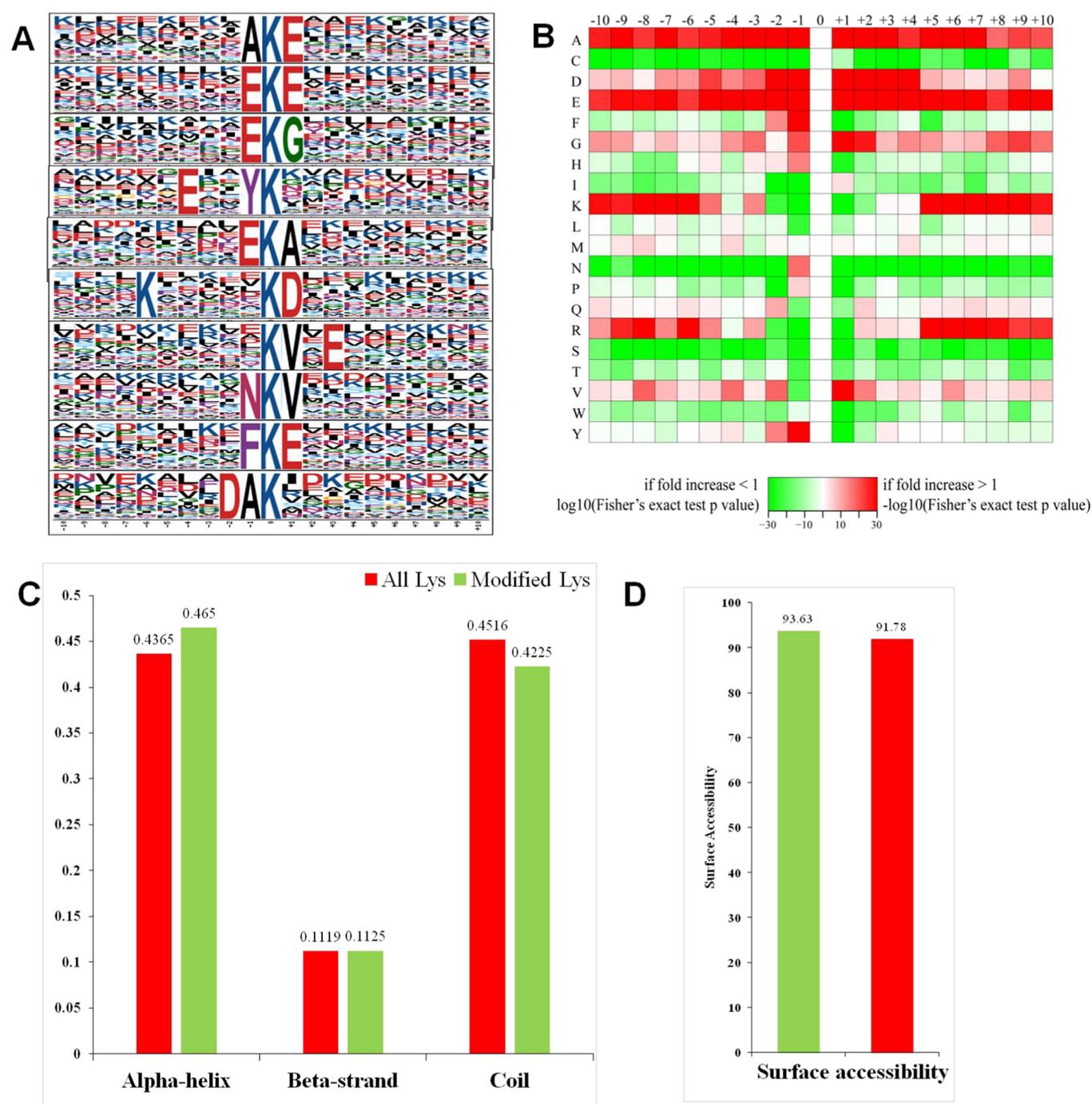


Figure 2. Motif analysis of lysine crotonylated (Kcr) peptides, secondary structures, and surface accessibility of crotonylation sites. (A) Motif analysis shows crotonylated peptide motifs and conservation of Kcr sites. The intensity map shows enrichment of amino acids in particular positions around the crotonylated lysine residues (10 amino acids upstream and downstream of the targeted lysine site). (B) Heat map of the amino acid compositions around the lysine crotonylation sites. Red indicates enrichment, and green indicates depletion. (C) Probabilities of lysine crotonylation in different protein secondary structures (alpha helix, beta-strand, and coil structures). (D) Predicted surface accessibility of crotonylation sites.

tryptic peptides (Figure 1C, Table S3), meeting the standard. Of all the crotonylated proteins, 43.8% (1075/2452) contained 1 modified site, 19.8% (485/2452) and 12.2% (299/2452) of all the crotonylated proteins contained 2 or 3 modified sites, respectively, whereas 4.6% (114/2452) of the proteins had more than 10 modification sites (Figure 1D, Table S3).

2.2. Motif Analysis of Lysine Crotonylation Sites and Structural Analysis of all the Crotonylated Proteins. To understand the nature of the crotonylated lysines in *M.*

persicae, the sequence motifs in all the crotonylated peptides identified were analyzed using the Motif-X program. Substantial bias in the amino acid distribution was observed from the -10 to $+10$ positions surrounding the crotonylated lysine in the 6940 peptides (Figure 2A,B, Tables S4 and S5). Among the 52 conserved motifs, motifs AK^{cr}E, EK^{cr}E, EK^{cr}G, ExxYK^{cr}, and EK^{cr}A (K^{cr} indicates the crotonylated lysine, with x indicated as a random amino acid residue) were strikingly conserved (with a motif score > 25). As shown by the heat

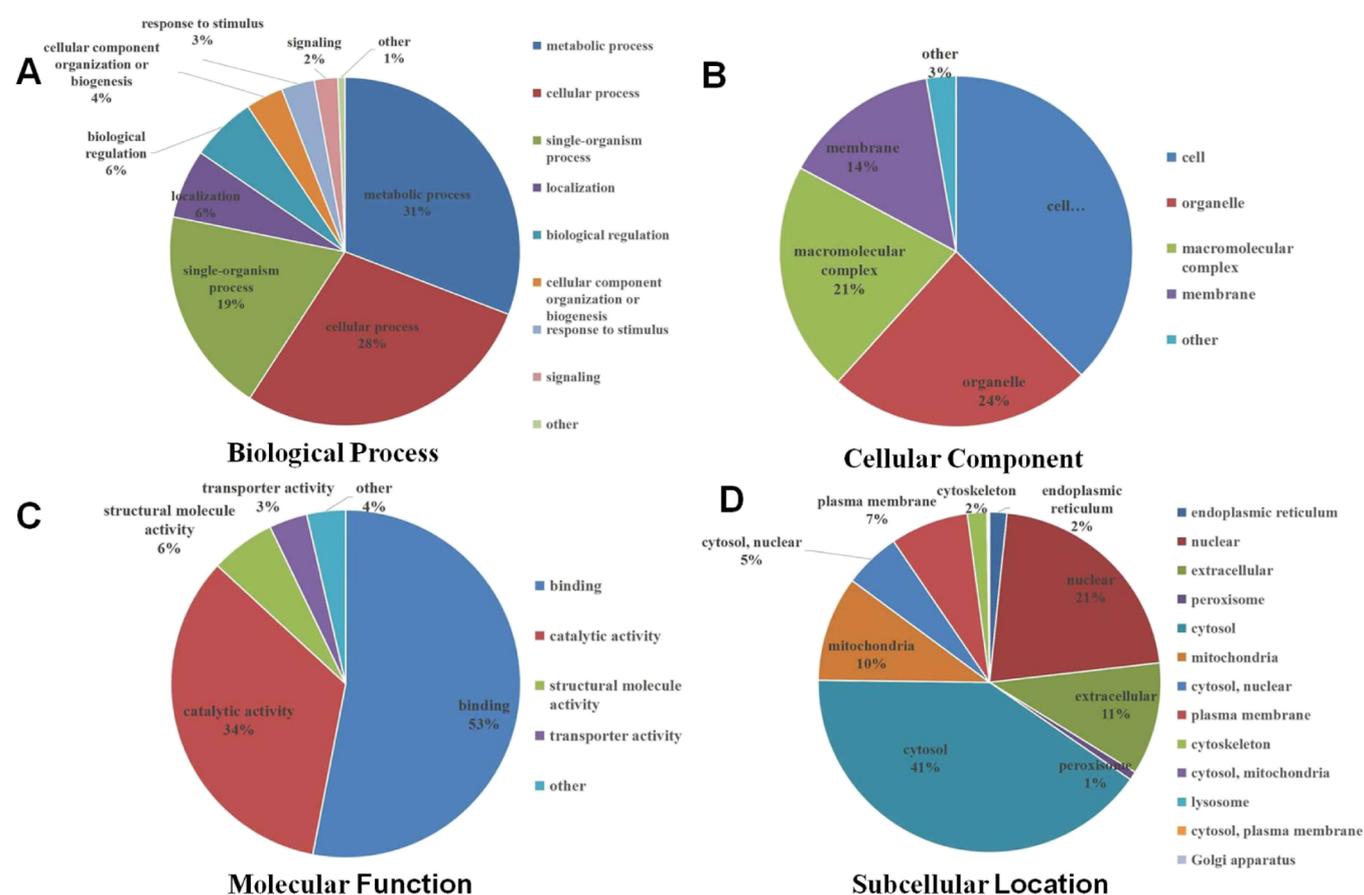


Figure 3. Pie charts showing the distribution of all the identified crotonylated proteins. (A) Crotonic acylation proteins categorized according to biological process. (B) Crotonic acylation proteins categorized according to the cellular component. (C) Crotonic acylation proteins categorized according to the molecular function. (D) Subcellular localization.

map of motif analysis of all the identified sites, the greatest enrichment in the -10 to $+10$ positions was found for a residue with hydrophobic side chain groups, namely, alanine (A) (Figure 2B). In addition, two negatively charged residues, aspartic acid (D) and glutamic acid (E), were observed to have the highest frequency of occurrence. However, a neutral amino acid, serine (S), and two positively charged residues, lysine (K) and arginine (R), were all observed to have the lowest enrichment in the position -1 to $+1$ (Figure 2B). Thus, proteins with A, D, or E and without S, K, or R in the corresponding positions would be the preferred substrates for lysine crotonylation in the cell. It also suggested that amino acid residues with negative charges and hydrophobic side chains might be functionally important for crotonylation to occur. The secondary structures of all the crotonylated proteins were analyzed to understand the relationship between Kcr and the protein structure. In *M. persicae*, 46.5% of the crotonylated sites were located at the beta-strand region, 42.3% were located at the coil region, and 11.3% were located at the alpha-helix region (Figure 2C). In addition, the result showed that the surface accessibility between the properties of the crotonylated sites and all lysine residues (Figure 2D) suggested that Kcr may not affect the surface properties of the modified proteins.

2.3. Functional Annotation and Cellular Localization of Crotonylated Proteins in *M. persicae*. To better understand the lysine crotonylome in *M. persicae*, Gene Ontology (GO) functional classification of all the crotonylated

proteins was investigated based on the categories biological process, cellular component, and molecular function (Figure 3A–C). In the biological process classification, the largest group of crotonylated proteins consisted of enzymes that are related to metabolic processes (30.9%) (Figure 3A, Table S6), while another large crotonylated protein group, classified as the biological process, was assigned to the cellular process (28.3%). In the cellular component classification, 38.0, 24, 21, and 14% of the crotonylated proteins were associated with the cell, organelle, macromolecular complex, and membrane, respectively (Figure 3B, Table S6). With respect to the molecular function classification, most crotonylated proteins were shown to be associated with binding (53.0%) and catalytic activity (33.9%), whereas others were mainly related to structural molecule activity (6.0%) and transporter activity (3.5%) (Figure 3C, Table S6). The subcellular localization of the crotonylated proteins was also analyzed, and the results showed that most of the crotonylated proteins identified were located in the cytosol (40.5%) or nuclear (21.5%) (Figure 3D). Several proteins were predicted to be located in the extracellular space (10.7%), mitochondrion (10.0%), plasma membrane (7.4%), cytoskeleton (1.8%), endoplasmic reticulum (1.6%), and peroxisome (0.8%) (Figure 3D, Table S7).

2.4. Functional Enrichment Analysis. In order to study which types of proteins are preferred targets for lysine crotonylation, GO (the biological process, molecular function, and cellular component), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and protein domain enrich-

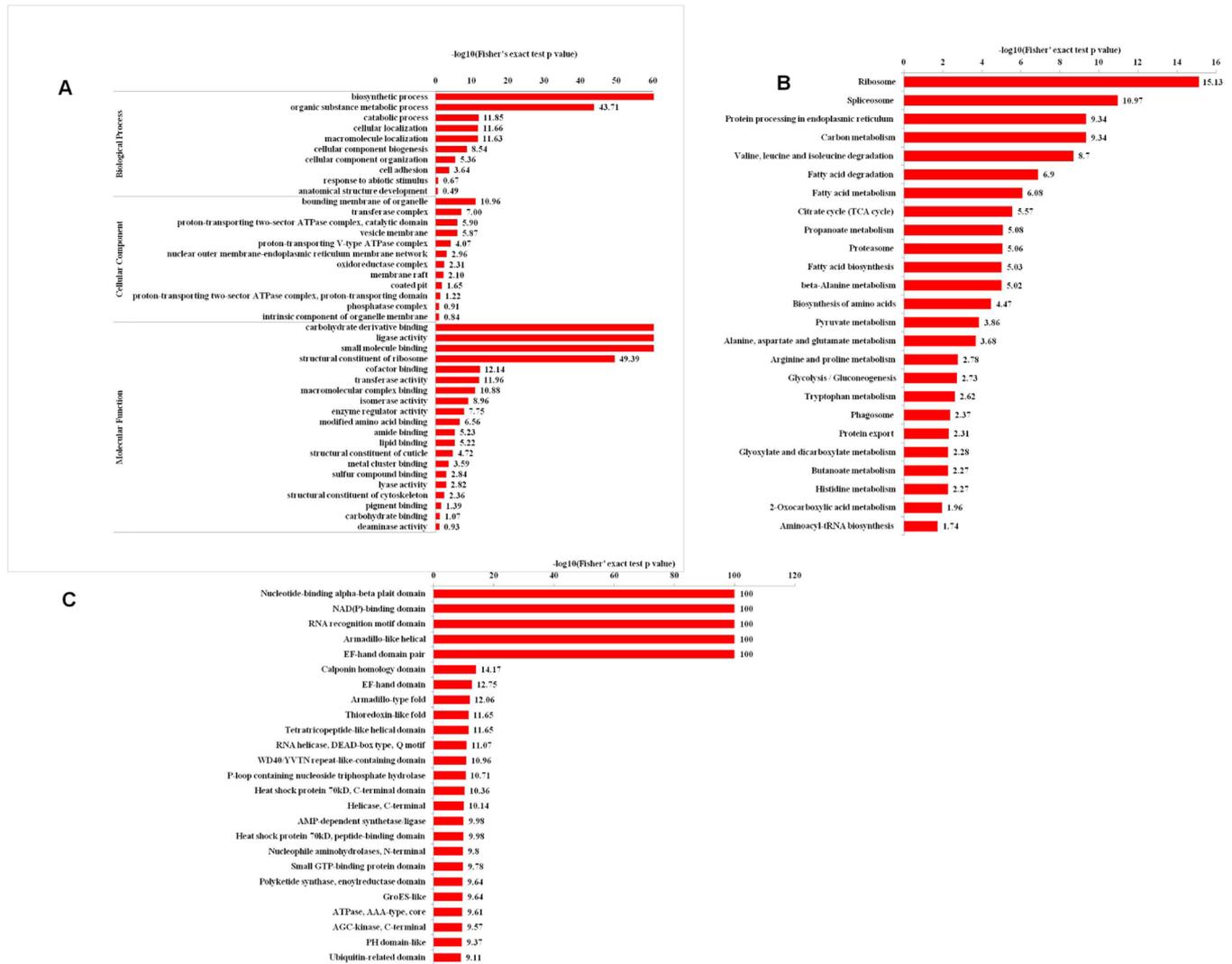
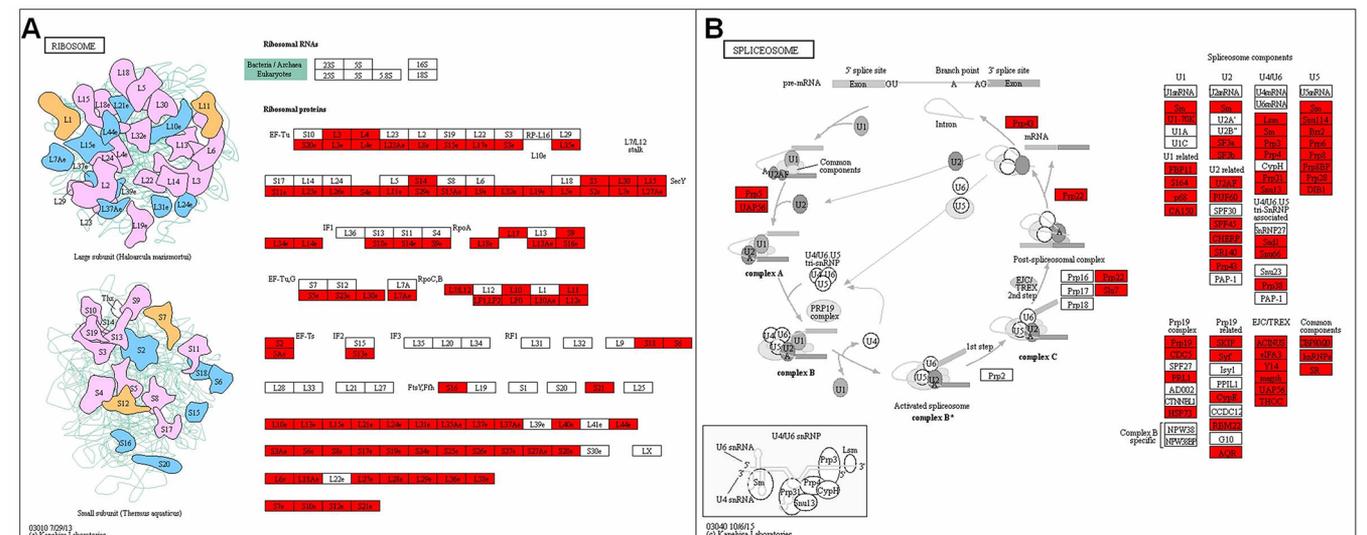


Figure 4. Enrichment analysis of the crotonylated proteins in *M. persicae*. (A) GO-based enrichment analysis of the crotonic acylation proteins. (B) KEGG pathway enrichment analysis. (C) Protein domain enrichment analysis.



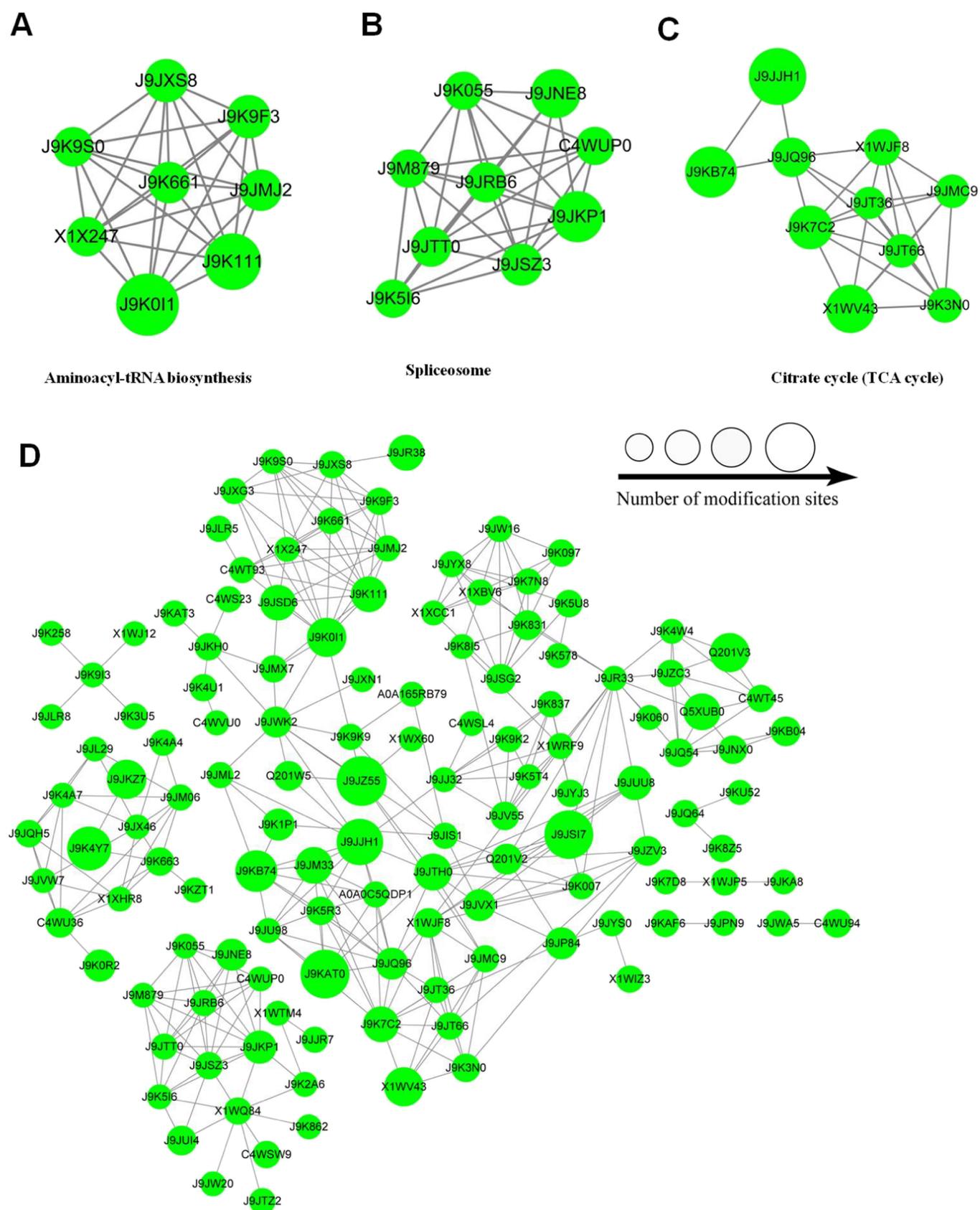


Figure 6. All crotonylated proteins were identified which were related to the amino acid metabolic process. The interaction network from STRING was visualized in Cytoscape (Version 3.3.0). A graph of the theoretical clustering algorithm, molecular complex detection (MCODE), was utilized to analyze densely connected regions. The size of the circles represents the number of Kcr modifications in each figure. (A) Proteins related to the amino acid metabolism. (B) Aminoacyl-tRNA biosynthesis. (C) Spliceosome. (D) Citrate cycle (TCA cycle).

ment analysis of all crotonylated proteins identified were conducted. The results indicated that most of the crotonylated proteins were involved in the biosynthetic process and organic substance metabolism process in the biological process category. Consistent with these findings, the enrichment analysis based on the molecular function showed that the majority of the modified proteins were associated with carbohydrate derivative binding and ligase activity. Additionally, proteins enriched in the bounding membrane of the organelle, transferase complex, proton-transporting two-sector ATPase complex, and catalytic domain are more likely to be modified by lysine crotonylation (Figure 4A, Table S8). KEGG pathway-based enrichment analysis showed that a large number of crotonylated proteins were involved in ribosomes, spliceosomes, protein processing in the endoplasmic reticulum, the carbon metabolism, valine, leucine and isoleucine degradation, fatty acid degradation, the fatty acid metabolism, the citrate cycle (TCA cycle), the propanoate metabolism, proteasome, fatty acid biosynthesis, the beta-alanine metabolism, the biosynthesis of amino acids, etc. (Figure 4B, Table S9). Furthermore, protein domain enrichment analysis showed that lysine crotonylation occurred on many proteins involved in the energy metabolism and stress defense response domains, such as the NAD(P)-binding domain, AMP-dependent synthetase/ligase, the thioredoxin-like fold, the Heat shock protein 70 kDa, and the C-terminal domain (Figure 4C, Table S10). These results suggested that crotonylated proteins played an important role in biosynthesis, metabolic pathways, and adaption to various stresses.

2.5. Lysine Crotonylated Proteins are Involved in the Central Metabolism. The data from the KEGG pathway-based enrichment analysis showed that Kcr was associated with a large number of proteins involved in ribosomes and spliceosomes. A large number of subunits of ribosomes were crotonylated (Figure 5A): 15 of the 81 large subunits and 39 of the small subunits were crotonylated, accounting for 63.0 and 75% of the ribosomes subunits, respectively. The spliceosome is a large molecular complex responsible for catalyzing RNA splicing, which is very important for the transmission of information from DNA to proteins. The results showed that crotonylation was widespread in various components of the spliceosome. In the subcomplex, small nuclear ribonucleoproteins (snRNP) like U1, U2, U4, U5, U6, the Prp19 complex, Prp19-related components, EJC/TREX, and common components all have several crotonylated subunits (Figure 5B) These results suggested that crotonylation in proteins was involved in translation and protein synthesis.

2.6. Protein Interaction Network Analysis. To investigate how the interaction between crotonylated sites in *M. persicae*, we established the interaction networks using the Cytoscape software. As shown in Figure 6C, there were 199 crotonylated proteins related to the amino acid metabolism process, which were mapped in the PPI network database, which provided a global view of how crotonylated proteins were involved in a wide range of pathways in *M. persicae*. In the aminoacyl-tRNA biosynthesis network, 26 crotonylated proteins were identified as nodes in the protein interaction database. In the spliceosome network, 145 proteins were mapped to the protein interaction database, whereas in the citrate cycle network, 26 crotonylated proteins were identified as nodes in the protein interaction database (Figure 6, Table S11).

3. DISCUSSION

Lysine crotonylation is recognized to be a novel, evolutionarily conserved PTM with diverse biological functions. This type of modification has been reported in a limited number of species, including examples of mammals, fish, yeast, fungi, and plants. In this paper, we described a proteomic study of lysine crotonylation in the peach potato aphid, one of the most important pests of cultivated crops, and this represents the first report of crotonylation in the insect. A total of 2452 crotonylated protein groups, with 7530 lysine crotonylation sites were identified through a combination of HPLC fractionation and liquid chromatography–tandem mass spectrometry (LC–MS/MS), showing that 5.8% of the proteins in *M. persicae* were crotonylated. We added common acetylation as variable modifications, and the results showed that only 109 acetylation sites were found (Table S12); the confidence of crotonylation events was improved. The crotonylated proteins were distributed in multiple cellular compartments and belonged to a variety of functional groups, suggesting that lysine crotonylation may play a wide range of important roles in regulating diverse cellular processes in *M. persicae*. Additionally, a large number of crotonylated proteins were related to the metabolic process in the biological process and binding in the molecular function, which implies that lysine crotonylation probably participates in the primary and secondary metabolism activities of the aphid. Many crotonylated proteins were related to the amino acid metabolism, aminoacyl-tRNA biosynthesis, and the spliceosome network, indicating that crotonylation is important in regulation of protein turnover. In addition, crotonylated proteins were involved in glycolysis and TCA, were related to the glucose catabolism, acetyl-CoA biogenesis, NAD(P)H production, fatty acid biosynthesis, fatty acid degradation, and all aspects of the metabolism, which showed that crotonylation also regulates the carbon metabolism.

Motif analysis of lysine crotonylation sites was also carried out in our study. It was found that amino acid residues with a negative charge and a hydrophobic side chain could be functionally important for crotonylation to occur. However, analysis of lysine acetylation motifs showed a high frequency of this modification on the amino acid residues with a positive charge and a hydrophobic side chain.²⁸ These results suggested that the motifs associated with different PTMs were diverse, which may be responsible for their individual functions.

Protein domain enrichment analysis showed that lysine crotonylation occurred on many proteins involved in stress defense response domains, with a large number of thioredoxin-like fold and heat shock protein 70 kDa, the C-terminal domain involved in crotonylation (Figure 4C). Whitefly (*Bremia tabaci*) is an important global insect pest which transmits an extremely important plant virus. Heat shock protein 70 has profound effects on whitefly development and its ability to transmit *tomato yellow leaf curl virus* (TYLCV).²⁹ By buffering mutations to favor cyclical parthenogenesis in the peach potato aphid, heat shock protein 90 can reduce the pest damage incurred by disrupting cyclical parthenogenesis.³⁰ Because cyclical parthenogenesis can improve the aphid survival rate through the winter by reproducing sexually and producing winter tolerant eggs, heat shock protein 90 could be important for *M. persicae* adapting to various abiotic stress factors, and the results suggested that crotonylated proteins could also be involved in these functions.

In fatty acid biosynthesis and degradation in *M. persicae*, Kcr serves as a diverse regulatory factor. In eukaryotes, fatty acids are important cellular components with most of them being polyunsaturated fatty acids (PUFAs). They play crucial roles in membrane biology and signaling processes in different organisms and are involved in the regulation of a range of cellular physiological functions, including cold tolerance, ion channel regulation, endocytosis, and immunity to pathogens.³¹ PUFAs are an indispensable component of phospholipids in cell membranes, and the type and content ratio determine the fluidity of cell membranes, which, in turn, has a great influence on the function of immune cells.³² Among the PUFAs, eicosanoids are crucial mediators of insect cellular immunity, as illustrated when inhibition of eicosanoid biosynthesis in tobacco hornworms was inhibited, which prevented clearance of injected bacteria from the circulating hemolymph.³³ Second, eicosanoids influence specific cellular actions, such as mediating micro-aggregation and nodulation reactions to bacterial challenges.³⁴ Third, eicosanoids influence humoral immune reactions. In *Drosophila*, this is achieved by functional coupling between the phospholipase A2-generated fatty acid cascade and lipopolysaccharide-dependent activation of the immune deficiency (imd) pathway in its immune response.³⁵ Last but not the least, the invaders target eicosanoid biosynthesis to suppress host immunity.³⁶ Many lysine-crotonylated proteins were shown to be involved in fatty acid biosynthesis and degradation in *M. persicae* (Figure S1), which suggested that Kcr plays an important role in the regulation of the *M. persicae* immune response. *M. persicae* causes heavy global crop losses per year, and Kcr targeting of eicosanoid signaling may have potential as a pest management technology.

The spliceosome is a large ribonucleoprotein complex composed of five small nuclear RNAs (snRNAs), namely, U1, U2, U4, U5, and U6, associated with other accessory proteins to catalyze splicing of transcribed pre-mRNA.³⁷ There have been reports showing that spliceosome dysfunction is associated with many human genetic diseases, such as diseases affecting the intron recognition step, diseases affecting the formation of catalytically active spliceosomes, and diseases involving mutations in the splice sites or spliceosome mutations affecting minor spliceosome activity.³⁸ Dysregulation of the spliceosome can be found at every stage of the splicing process through downregulation of every subunit of spliceosome in early stage Alzheimer's disease.³⁹ Proteomics research in aging human skeletal muscles found changes in the spliceosome,⁴⁰ inhibiting the spliceosome function in breast cancer cells, which may be potential entry points for therapy against aggressive MYC-driven cancer.⁴¹ Our results showed that almost all the *M. persicae* spliceosome units were crotonylated. From this finding, spliceosomes could be used as a new target of insecticides by regulating PTMs in insects to interfere with insect gene transcription and metamorphosis to achieve the prevention of *M. persicae* outbreak.

In conclusion, our research gave global maps of lysine crotonylation of *M. persicae*. Function analysis showed that a majority of the crotonylated proteins not only were involved in a large amount of basic life activities such as biosynthetic and metabolic processes but perhaps also play a crucial part for adapting to abiotic stress, regulating immune response and genetic activities. Our research provided a foundation and protein candidates; these involved processes can be used as potential targets for *M. persicae* control, and we will choose

some candidates and focus on them to search for new tools in order to better control the economically important pest *M. persicae*.

4. EXPERIMENTAL SECTION

4.1. Insect Rearing. The peach potato or green peach aphids (*M. persicae*) used in the study were collected from a tobacco field in Jimo, Shandong Province, China. The aphids used in the study were raised on *Nicotiana tabacum* cv. K326, with the temperature set constantly at 25 °C under a photoperiod of 16 h light/8 h dark. The wingless adult aphids were selected and used immediately for protein extraction.

4.2. Protein Extraction and Trypsin Digestion. The sample was grinded in liquid nitrogen, and the cell powder was sonicated three times on ice using a high-intensity ultrasonic processor (Scientz, Ningbo Scientz Biotechnology, Ningbo, China) in the lysis buffer [8 M urea, 2 mM ethylenediaminetetraacetic acid (EDTA), 3 μM trichostatin A, 50 mM nicotinamide1, 10 mM dithiothreitol (DTT), and 1% protease inhibitor cocktail]. The insoluble debris was removed by centrifugation at 20,000g at 4 °C for 10 min. The protein in the supernatant was precipitated with cold 15% trichloroacetic acid (TCA) for 2 h at −20 °C. After centrifugation at 4 °C for 10 min, the supernatant was discarded. The precipitate pellet was washed with cold acetone three times. The protein was redissolved in the buffer (8 M urea, 100 mM NH₄CO₃, pH 8.0) and the protein concentration was determined with the 2-D Quant Kit (GE Healthcare, Beijing, China) according to the manufacturer's instructions. For digestion, the protein solution was reduced with 10 mM DTT for 1 h at 37 °C and alkylated with 20 mM IAA for 45 min at room temperature in darkness. For trypsin digestion, the protein sample was diluted by adding 100 mM NH₄CO₃ to a final urea concentration of less than 2 M. Finally, trypsin was added at a 1:50 trypsin-to-protein mass ratio for the first digestion overnight and at a 1:100 trypsin-to-protein mass ratio for the second 4 h digestion.

4.3. HPLC Fractionation. The sample was then fractionated by high-pH reverse-phase HPLC using an Agilent 300Extend (Agilent, Shanghai, China) C18 column (5 μm particle size, 4.6 mm ID, 250 mm length). Briefly, peptides were first separated into 80 fractions with a linear gradient of 2–60% acetonitrile (ACN) in 10 mM ammonium bicarbonate pH 10 over 80 min. Then, the peptides were combined into six fractions and dried by vacuum centrifugation.

4.4. Affinity Enrichment. To enrich the tryptic peptide sample with respect to crotonylated peptides, tryptic peptides dissolved in NETN buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl, 0.5% NP-40, pH 8.0) were incubated with pre-washed antibody beads (PTM Biolabs, Hangzhou, China) at 4 °C overnight with gentle shaking. The beads were washed four times with NETN buffer and twice with ddH₂O. The bound peptides were eluted from the beads with 0.1% TFA. The eluted fractions were combined and dried under vacuum. The resulting peptides were cleaned with C18 ZipTips (Millipore, Sanit Louis, USA) according to the manufacturer's instructions, followed by LC-MS/MS analysis.

4.5. LC-MS/MS Analysis. The peptides were dissolved in 0.1% formic acid (FA) (Fluka, Shanghai, China) and then directly loaded onto a reversed-phase pre-column (Acclaim PepMap 100, Thermo Scientific, Waltham, USA). Peptide separation was performed using a reversed-phase analytical column (Acclaim PepMap RSLC, Thermo Scientific, Waltham, USA). Elution was carried out using a linear gradient from 6 to

22% solvent B [0.1% FA in 98% ACN (Fisher Chemical, Waltham, USA)] for 24 min and then from 22 to 40% solvent B for 8 min, climbing to 80% in 5 min, then being maintained at 80% for the last 3 min, all at a constant flow rate of 300 nL/min on an EASY-nLC 1000 UPLC system. The eluted peptides were analyzed using a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, USA). The peptides were subjected to a nanospray (NSI) source, followed by tandem mass spectrometry (MS/MS) in a Q Exactive Plus MS (Thermo Scientific, Waltham, USA) coupled online to the UPLC system. Intact peptides were detected in the Orbitrap at a resolution of 70,000. Peptides were selected for MS/MS using NCE setting at 30; ion fragments were detected in the Orbitrap at a resolution of 17,500. A data-dependent procedure, which alternated between one MS scan, followed by 20 MS/MS scans, was applied for the top 20 precursor ions above a threshold ion count of 5×10^3 in the MS survey scan with a 15.0 s dynamic exclusion. The electrospray voltage applied was 2.0 kV. Automatic gain control was used to prevent overfilling of the Orbitrap; 5×10^4 ions were accumulated for generation of the MS/MS spectra. For MS scans, the m/z scan range was 350–1800. The fixed first mass was set at 100 m/z .

4.6. Data Analysis. The resulting MS/MS data were processed using the MaxQuant platform (<http://www.maxquant.org/>) with the integrated Andromeda search engine (v.1.4.2). Tandem mass spectra were searched against the Uniprot_133076_Aphidid database (<https://www.uniprot.org>) concatenated with the reverse decoy database. Trypsin/P was specified as the cleavage enzyme, allowing up to four missing cleavages, five modifications per peptide, and five charges. The mass error was set to 10 ppm for precursor ions and 0.02 Da for fragment ions. Carbamidomethylation on Cys was specified as a fixed modification, and oxidation on Met, crotonylation on Lys, and crotonylation on the protein N-terminus were specified as variable modifications. The false discovery rate (FDR) thresholds for the protein, peptide, and modification site were each specified at 1%. The minimum peptide length was set at 7. All the other parameters in MaxQuant were set to default values. The site localization probability was set at >0.75.

For the MS data validation, we checked the mass error distribution of all the identified peptides and the length distribution of the peptides.

4.7. Bioinformatics Analysis. Soft *motif-x* was used to analyze the model of amino acid sequences consisting of amino acids in specific positions of modify-21-mers (10 amino acids upstream and 10 downstream of the site) in all protein sequences. All the database protein sequences were used as background database parameters, with all other parameters being set at default. Proteins were classified by GO annotation into three categories: biological process, cellular compartment, and molecular function. For each category, a two-tailed Fisher's exact test was employed to test the significance of the enrichment of the identified protein against all database proteins. Correction for multiple hypothesis testing was carried out using standard FDR control methods. The GO with a corrected p -value < 0.05 was considered to be significant. The KEGG database was used to identify enriched pathways by a two-tailed Fisher's exact test to test significance of the enrichment of the identified protein against all database proteins. Correction for multiple hypothesis testing was carried out using standard FDR control methods. The pathway with a

corrected p -value < 0.05 was considered to be significant. These pathways were classified into hierarchical categories according to the KEGG website. For each protein category, the InterPro database (<http://www.ebi.ac.uk/interpro/>) (a resource that provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and important sites) was researched and a two-tailed Fisher's exact test was employed to test the significant of the enrichment of the identified protein against all database proteins. Correction for multiple hypothesis testing was carried out using standard FDR control methods, and domains with a corrected p -value < 0.05 were considered to be significant. For all crotonylated proteins identified (which related to the amino acid metabolism process) name identifiers were searched against the STRING database version 10.0 (<http://string-db.org/>) for PPIs. Only interactions between the proteins belonging to the searched data set were selected, thereby excluding external candidates. STRING defines a metric called the "confidence score" to define interaction confidence; we selected all interactions that had a confidence score ≥ 0.9 (high confidence). The interaction network from STRING was visualized in Cytoscape (Version 3.3.0). A graph of the theoretical clustering algorithm, molecular complex detection (MCODE), was utilized to analyze densely connected regions.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.1c01194>.

Basic analysis of MS identified information, annotation combination of identified proteins, peptide length and number of modified sites, detailed information of motif annotation of identified proteins, detailed information of the motif model of identified proteins, detailed information of GO terms of level 2 distribution of all proteins, detailed information of subcellular localization of the identified protein, GO enrichment analysis of crotonylated proteins, detailed information of the KEGG pathway, detailed information of the protein domain, detailed information of amino acid metabolic process-related protein, MS acetyl identified information, and fatty acid biosynthesis and degradation pathway enriched from KEGG pathway analysis (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Bin Zhang – Qingdao Agricultural University, Qingdao, Shandong 266109, China; Email: yangjinguang@caas.cn

Jinguang Yang – Tobacco Research Institute of CAAS, Qingdao, Shandong 266101, China; orcid.org/0000-0002-9584-9634; Email: binzhang@qau.edu.cn

Authors

Manlin Xu – Tobacco Research Institute of CAAS, Qingdao, Shandong 266101, China; Shandong Peanut Research Institute, Qingdao, Shandong 266100, China; orcid.org/0000-0002-4266-5110

Yi Xie – Tobacco Research Institute of CAAS, Qingdao, Shandong 266101, China

Ying Li – Tobacco Research Institute of CAAS, Qingdao, Shandong 266101, China

Lili Shen – Tobacco Research Institute of CAAS, Qingdao, Shandong 266101, China

Kun Huang – Tobacco Company of Yunnan Province, Honghe Company, Mile, Yunnan 652300, China

Zhonglong Lin – China Tobacco Corporation Yunnan Company, Kunming, Yunnan 650000, China

Bin Li – China Tobacco Corporation Sichuan Company, Chengdu, Sichuan 610000, China

Changjian Xia – Haikou Cigar Research Institute, Hainan Provincial Branch of China National Tobacco Corporation (CNTC), Haikou, Hainan 570100, China

Xia Zhang – Shandong Peanut Research Institute, Qingdao, Shandong 266100, China

Yucheng Chi – Shandong Peanut Research Institute, Qingdao, Shandong 266100, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.1c01194>

Author Contributions

[†]M.X. and Y.X. have contributed equally to this work. Author contributions: writing—original draft preparation, M.L.X., Y.X., B.L., C.J.X., and X.Z.; writing—review and editing, J.G.Y., B.Z., Y.L., L.L.S., K.H., and Z.L.L.; supervision, F.L.W. and Y.C.C.; project administration, J.G.Y. and B.Z.; all authors have read and agreed to the published version of the manuscript.

Notes

The authors declare no competing financial interest.

Institutional review board statement: this study did not involve humans or animals.

The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD 023519.

ACKNOWLEDGMENTS

We are grateful to the funding by the National Natural Science Foundation of China (31801711), Major Tobacco Green Prevention and Control Project 110201901041(ls-04), the Sichuan Provincial Tobacco Company Science and Technology Project (SCYC201804), the Hainan Provincial Tobacco Company Science and Technology Project (201846000024056), and the Agricultural Science and Technology Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2021A42).

REFERENCES

- (1) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **2002**, *319*, 1097–1113.
- (2) Harp, J. M.; Hanson, B. L.; Timm, D. E.; Bunick, G. J. Asymmetries in the nucleosome core particle at 2.5 resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 1513–1534.
- (3) Makde, R. D.; England, J. R.; Yennawar, H. P.; Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **2010**, *467*, 562.
- (4) Luger, K.; Mäder, A. W. Crystal structure of the nucleosome core particle at 2.8 Å resolution. (cover story). *Nature* **1997**, *389*, 251.
- (5) Venkatesh, S.; Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 178–189.
- (6) Ruthenburg, A. J.; Li, H.; Patel, D. J.; David Allis, C. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 983–994.

- (7) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **2011**, *146*, 1016–1028.

- (8) Montellier, E.; Rousseaux, S.; Zhao, Y.; Khochbin, S. Histone crotonylation specifically marks the haploid male germ cell gene expression program: post-meiotic male-specific gene expression. *Bioessays* **2012**, *34*, 187–193.

- (9) Sabari, B. R.; Tang, Z.; Huang, H.; Yong-Gonzalez, V.; Molina, H.; Kong, H. E.; Dai, L.; Shimada, M.; Cross, J. R.; Zhao, Y.; Roeder, R. G.; Allis, C. D. Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation. *Mol. Cell* **2018**, *69*, 533.

- (10) Xu, W.; Wan, J.; Zhan, J.; Li, X.; He, H.; Shi, Z.; Zhang, H. Global profiling of crotonylation on non-histone proteins. *Cell Res.* **2017**, *27*, 946–949.

- (11) Sabari, B. R.; Zhang, D.; Allis, C. D.; Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 90–101.

- (12) Sabari, B. R.; Tang, Z.; Huang, H.; Yong-Gonzalez, V.; Molina, H.; Kong, H. E.; Dai, L.; Shimada, M.; Cross, J. R.; Zhao, Y.; Roeder, R. G.; Allis, C. D. Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol. Cell* **2015**, *58*, 203–215.

- (13) Li, Y.; Sabari, B. R.; Panchenko, T.; Wen, H.; Zhao, D.; Guan, H.; Wan, L.; Huang, H.; Tang, Z.; Zhao, Y.; Roeder, R. G.; Shi, X.; Allis, C. D.; Li, H. Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol. Cell* **2016**, *62*, 181–193.

- (14) Zhao, D.; Guan, H.; Zhao, S.; Mi, W.; Wen, H.; Li, Y.; Zhao, Y.; Allis, C. D.; Shi, X.; Li, H. YEATS2 is a selective histone crotonylation reader. *Cell Res.* **2016**, *26*, 629–632.

- (15) Andrews, F. H.; Shinsky, S. A.; Shanley, E. K.; Bridgers, J. B.; Gest, A.; Tsun, I. K.; Krajewski, K.; Shi, X.; Strahl, B. D.; Kutateladze, T. G. The Taf14 YEATS domain is a reader of histone crotonylation. *Nat. Chem. Biol.* **2016**, *12*, 396–398.

- (16) Lin, H.; Su, X.; He, B. Protein lysine acylation and cysteine succination by intermediates of energy metabolism. *ACS Chem. Biol.* **2012**, *7*, 947–960.

- (17) Bao, X.; Wang, Y.; Li, X.; Li, X. M.; Liu, Z.; Yang, T.; Wong, C. F.; Zhang, J.; Hao, Q.; Li, X. D. Identification of 'erasers' for lysine crotonylated histone marks using a chemical proteomics approach. *Elife* **2014**, *3*, No. e02999.

- (18) Chen, W.; Tang, D.; Xu, Y.; Zou, Y.; Sui, W.; Dai, Y.; Diao, H. Comprehensive analysis of lysine crotonylation in proteome of maintenance hemodialysis patients. *Medicine* **2018**, *97*, No. e12035.

- (19) Liu, J. F.; Wu, S. F.; Liu, S.; Sun, X.; Wang, X. M.; Xu, P.; Chen, H. Z.; Yang, J. T. Global Lysine Crotonylation Profiling of Mouse Liver. *Proteomics* **2020**, *20*, No. e2000049.

- (20) Kwon, O. K.; Kim, S. J.; Lee, S. First profiling of lysine crotonylation of myofibrillar proteins and ribosomal proteins in zebrafish embryos. *Sci. Rep.* **2018**, *8*, 3652.

- (21) Yang, Q.; Li, Y.; Apaliya, M. T.; Zheng, X.; Serwah, B. N. A.; Zhang, X.; Zhang, H. The Response of *Rhodotorula mucilaginosa* to Patulin Based on Lysine Crotonylation. *Front. Microbiol.* **2018**, *9*, 2025.

- (22) Zhang, N.; Yang, Z.; Liang, W.; Liu, M. Global Proteomic Analysis of Lysine Crotonylation in the Plant Pathogen *Botrytis cinerea*. *Front. Microbiol.* **2020**, *11*, 564350.

- (23) Sun, H.; Liu, X.; Li, F.; Li, W.; Zhang, J.; Xiao, Z.; Shen, L.; Li, Y.; Wang, F.; Yang, J. First comprehensive proteome analysis of lysine crotonylation in seedling leaves of *Nicotiana tabacum*. *Sci. Rep.* **2017**, *7*, 3013.

- (24) Liu, K.; Yuan, C.; Li, H.; Chen, K.; Lu, L.; Shen, C.; Zheng, X. A qualitative proteome-wide lysine crotonylation profiling of papaya (*Carica papaya* L.). *Sci. Rep.* **2018**, *8*, 8230.

- (25) Liu, S.; Xue, C.; Fang, Y.; Chen, G.; Peng, X.; Zhou, Y.; Chen, C.; Liu, G.; Gu, M.; Wang, K.; Zhang, W.; Wu, Y.; Gong, Z. Global

Involvement of Lysine Crotonylation in Protein Modification and Transcription Regulation in Rice. *Mol. Cell. Proteomics* **2018**, *17*, 1922–1936.

(26) Kettles, G. J.; Drurey, C.; Schoonbeek, H. J.; Maule, A. J.; Hogenhout, S. A. Resistance of *Arabidopsis thaliana* to the green peach aphid, *Myzus persicae*, involves camalexin and is regulated by microRNAs. *New Phytol.* **2013**, *198*, 1178–1190.

(27) Gupta, V. Aphids on the world's crops. An identification and information guide. *Orient. Insects* **2001**, *35*, 104.

(28) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. *Sci. Rep.* **2016**, *6*, 21069.

(29) Mandrioli, M.; Zanetti, E.; Nardelli, A.; Manicardi, G. C. Potential role of the heat shock protein 90 (hsp90) in buffering mutations to favour cyclical parthenogenesis in the peach potato aphid *Myzus persicae* (Aphididae, Hemiptera). *Bull. Entomol. Res.* **2019**, *109*, 426–434.

(30) Kanakala, S.; Kontsedalov, S.; Lebedev, G.; Ghanim, M. Plant-Mediated Silencing of the Whitefly *Bemisia tabaci* Cyclophilin B and Heat Shock Protein 70 Impairs Insect Development and Virus Transmission. *Front. Physiol.* **2019**, *10*, 557.

(31) Wallis, J. G.; Watts, J. L.; Browse, J. Polyunsaturated fatty acid synthesis: what will they think of next? *Trends Biochem. Sci.* **2002**, *27*, 467.

(32) Schumann, J. It is all about fluidity: Fatty acids and macrophage phagocytosis. *Eur. J. Pharmacol.* **2016**, *785*, 18–23.

(33) Stanley-Samuelson, D. W.; Jensen, E.; Nickerson, K. W.; Tiebel, K.; Ogg, C. L.; Howard, R. W. Insect immune response to bacterial infection is mediated by eicosanoids. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1064–1068.

(34) Miller, J. S.; Nguyen, T.; Stanley-Samuelson, D. W. Eicosanoids mediate insect nodulation responses to bacterial infections. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12418–12422.

(35) Yajima, M.; Takada, M.; Takahashi, N.; Kikuchi, H.; Natori, S.; Oshima, Y.; Kurata, S. A newly established in vitro culture using transgenic *Drosophila* reveals functional coupling between the phospholipase A2-generated fatty acid cascade and lipopolysaccharide-dependent activation of the immune deficiency (*imd*) pathway in insect immunity. *Biochem. J.* **2003**, *371*, 205–210.

(36) Fang, Q.; Wang, L.; Zhu, J.; Li, Y.; Song, Q.; Stanley, D. W.; Akhtar, Z.-r.; Ye, G. Expression of immune-response genes in lepidopteran host is suppressed by venom from an endoparasitoid, *Pteromalus puparum*. *BMC Genom.* **2010**, *11*, 484.

(37) Hoskins, A. A.; Rodgers, M. L.; Friedman, L. J.; Gelles, J.; Moore, M. J. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *Elife* **2016**, *5*, No. e14166.

(38) Verma, B.; Akinyi, M. V.; Norppa, A. J.; Frilander, M. J. Minor spliceosome and disease. *Semin. Cell Dev. Biol.* **2018**, *79*, 103–112.

(39) Nuzzo, D.; Inguglia, L.; Walters, J.; Picone, P.; Di Carlo, M. A Shotgun Proteomics Approach Reveals a New Toxic Role for Alzheimer's Disease A β Peptide: Spliceosome Impairment. *J. Proteome Res.* **2017**, *16*, 1526–1541.

(40) Ubaida-Mohien, C.; Lyashkov, A.; Gonzalez-Freire, M.; Tharakan, R.; Shardell, M.; Moaddel, R.; Semba, R. D.; Chia, C. W.; Gorospe, M.; Sen, R.; Ferrucci, L. Discovery proteomics in aging human skeletal muscle finds change in spliceosome, immunity, proteostasis and mitochondria. *Elife* **2019**, *8*, No. e49874.

(41) Hsu, T. Y.-T.; Simon, L. M.; Neill, N. J.; Marcotte, R.; Sayad, A.; Bland, C. S.; Echeverria, G. V.; Sun, T.; Kurley, S. J.; Tyagi, S.; Karlin, K. L.; Dominguez-Vidaña, R.; Hartman, J. D.; Renwick, A.; Scorsone, K.; Bernardi, R. J.; Skinner, S. O.; Jain, A.; Orellana, M.; Lagisetti, C.; Golding, I.; Jung, S. Y.; Neilson, J. R.; Zhang, X. H.-F.; Cooper, T. A.; Webb, T. R.; Neel, B. G.; Shaw, C. A.; Westbrook, T. F. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature* **2015**, *525*, 384–388.