View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Ghent University Academic Bibliography

N-glycosylation site analysis reveals sex-related differences in protein N-glycosylation in the rice brown planthopper (*Nilaparvata lugens*)

Freja Scheys[1,2], Els J. M. Van Damme[2], Jarne Pauwels[3,4], An Staes[3,4], Kris Gevaert[3,4,£], Guy Smagghe[1,£]*

[1]Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

[2]Department of Biotechnology, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

[3]VIB-UGent Center for Medical Biotechnology, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

[4]Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

[£]: Equal last author

*Corresponding author: Email: Guy.Smagghe@UGent.be, Tel: +32 9 2646150

Running title: Protein N-glycosylation is sex-related in a rice pest insect

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Abbreviations

ABC – ammonium bicarbonate

ASA – *Allium sativum* agglutinin

BPH – brown planthopper

ConA – Concanavalin A

FASP – Filter-aided sample preparation

Gal – galactose

GalNAc – N-acetylgalactosamine

GO – Gene Ontology

PBS – phosphate buffered saline

PNGase A – Peptide-*N*-glycosidase A

LF – long-winged female

LFQ – label-free quantification

LM – long-winged male

RSA – *Rhizoctonia solani* agglutinin

SF – short-winged female

SM – short-winged male

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Abstract

Glycosylation is a common modification of proteins and critical for a wide range of biological processes. Differences in protein glycosylation between sexes have already been observed in humans, nematodes and trematodes, and have recently also been reported in the rice pest insect *Nilaparvata lugens*. Although protein N-glycosylation in insects is nowadays of high interest due to its potential for exploitation in pest control strategies, the functionality of differential N-glycosylation between sexes is yet unknown. In this study, therefore, the occurrence and role of sex-related protein N-glycosylation in insects were examined. A comprehensive investigation of the N-glycosylation sites from the adult stages of *N. lugens* was conducted, allowing a qualitative and quantitative comparison between sexes at the glycopeptide level. N-glycopeptide enrichment via lectin capturing using the high mannose/paucimannose-binding lectin Concanavalin A, or the *Rhizoctonia solani* agglutinin which interacts with complex N-glycans, resulted in the identification of over 1,300 N-glycosylation sites derived from over 600 glycoproteins. Comparison of these N-glycopeptides revealed striking differences in protein N-glycosylation between sexes. Male- and female-specific N-glycosylation sites were identified, and some of these sex-specific N-glycosylation sites were shown to be derived from proteins with a putative role in insect reproduction. In addition, differential glycan composition between males and females was observed for proteins shared across sexes. Both lectin blotting experiments as well as transcript expression analyses with complete insects and insect tissues confirmed the observed differences in N-glycosylation of proteins between sexes. In conclusion, this study provides further evidence for protein N-glycosylation to be sex-related in insects. Furthermore, original data on N-glycosylation sites of *N. lugens* adults are presented, providing novel insights into planthopper's biology and information for future biological pest control strategies.

Keywords: N-linked glycosylation, N-glycosylation sites, *Nilaparvata lugens*, pest insect

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

1. Introduction

Nilaparvata lugens Stål (Hemiptera: Delphacidae) or the brown planthopper (BPH) is one of the most notorious pest insects in the rice producing countries of Asia (1). Being a sap-sucking insect, *N. lugens* causes direct damage to the rice plant by feeding on the phloem. In addition, BPH can also act as a vector for viruses causing indirect plant damage (2). *N. lugens* exhibits wing polymorphism, and adults appear either as short-winged or long-winged morphs. Whereas the short-winged morph is flightless and specialized in reproduction, the long-winged morph has a fully developed flight apparatus, allowing it to migrate and infest new rice fields (3). Recently, an extensive N-glycome analysis for the different life stages of *N. lugens* revealed that protein N-glycosylation is sex-related in this insect (4). Female N-glycomes show a 10-fold increase in high mannose N-glycans compared to males, while in adult males, mono- and difucosylated paucimannose N-glycans are more abundant. Furthermore, male glycoproteins can carry complex N-glycans with terminal galactose (Gal), which was not detected in females. The functionality of this differential N-glycosylation between sexes is yet unknown. Therefore, it is of interest to identify the protein moieties carrying these differential carbohydrate structures in *N. lugens* males and females. Since male and female adults can occur either as long-winged or short-winged morphs, it is also of interest to investigate whether protein N-glycosylation is wing type-dependent.

Using lectin enrichment, glycoproteins from several insect species have already been described (5-7), but little information is available on the actual insect N-glycosylation sites (8, 9). Furthermore, all these studies focused on adult insects without distinguishing for sexes or wing forms. Here, the N-glycosylation sites from *N. lugens* adults, comprising both sexes and wing types, were studied using the N-glyco-FASP method as developed by Zielinska *et al*. (10). This technique was recently optimized for insect studies by our group (8), and N-glycosylation sites from insects belonging to different insect orders were analyzed and compared. Two lectins, Concanavalin A (ConA) and *Rhizoctonia solani* agglutinin (RSA), were used for the enrichment of peptides carrying specific carbohydrate structures. ConA specifically interacts with high mannose and paucimannose N-glycans, while RSA binds to complex N-glycans carrying terminal Gal and N-acetylgalactosamine (GalNAc) residues. Previous glycomics studies revealed that especially high mannose and paucimannose N-

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

glycans occur in insects, together comprising over 90% of the total N-glycan pool, while complex glycans represent less than 10% of the total amount of detected N-glycans (4, 11-15). Enrichment of N-glycopeptides with ConA or RSA will thus allow enriching for abundant and rather scarce N-glycopeptides, respectively.

Our experiments allowed both a qualitative as well as a quantitative comparison of the N-glycosylation sites between the two sexes and wing types of *N. lugens* adults. For a selection of proteins, the proteomic data were validated at transcript level, using complete adult insects as well as adult tissues. Our comprehensive overview of N-glycosylation sites in *N. lugens* adults can help to address the importance of protein N-glycosylation in adult sexes and wing types, and provides novel insights into planthopper's biology. Furthermore, this information will have to be taken into account when assessing the possibilities of protein N-glycosylation disruption as a novel strategy to control insect pests.

2. Experimental Procedures

2.1. Experimental design and statistical rationale

N-glycosylation sites were analyzed for *N. lugens* adult morphs; short-winged males (SM), short-winged females (SF), long-winged males (LM) and long-winged females (LF). Protein extracts were prepared in three biological replicates for each adult morph. Glycopeptides were enriched from these protein extracts, and for each biological replicate the eluent from three filter units was pooled. All samples were purified on C18 Ziptips (10 µl, Millipore) prior to analysis on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (16) partner repository with the dataset identifier PXD013657.

2.2. Insect colonies and lectins

All insects were collected from a continuous colony, kept on 4- to 8-week-old rice plants (*Oryza sativa* japonica, Nipponbare; USDA-Agricultural Research Service, Genetic Stocks-Oryza Collection, Stuttgart, AR) in a climate chamber under standard conditions of 27 ± 1°C, 80 ± 5% relative humidity and a 16:8 light/dark photoperiod, and stored at -80°C until further analysis. The plant lectin ConA

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

was purified from Jack bean meal (*Canavalia ensiformis*) (17), and the fungal lectin RSA was isolated from sclerotia of the *R. solani* strain AG 1-1B (18).

2.3. Enrichment of N-glycopeptides

Protein extracts were prepared from approximately 100 mg of adult insects (50-100 insects) as previously described in Scheys *et al*. (2018). In short, complete insects were crushed in liquid nitrogen and insect powder was dissolved in extraction buffer (2% SDS, 50 mM DTT in 0.1 M Tris-HCl, pH 7.8). Extracts were processed using the FASP method, as previously described (19), with adjustments (8). Briefly, an extract containing approximately 3 mg protein was transferred to Amicon Ultra-0.5 centrifugal filter devices (±500 µg/filter; 30 K; Millipore, Burlington, MA), washed with urea buffer and alkylated with 50 mM iodoacetamide, and washed with 40 mM ammonium bicarbonate (ABC). Proteins were digested overnight using trypsin in a ratio of 1:100 (w:w) at 37°C and collected in 100 µl of 40 mM ABC.

Subsequently, glycopeptides were enriched from the collected peptide mixture. In short, the peptide mixture (150 µg) was mixed with a single lectin (ConA or RSA, 300 µg) in 2x binding buffer (2 mM $CaCl_2$, 2 mM $MnCl_2$, 1 M NaCl in 20 mM Tris-HCl, pH 7.4), and the mixture was added on a Microcon Centrifugal device (30 K; Millipore). Unbound peptides were eluted by centrifugation followed by 4 washes with binding buffer, and 2 washes with 50 mM citrate-phosphate buffer, pH 5 (0.2 M dibasic sodium phosphate set to pH 5 with 0.1 M citric acid). Peptide-*N*-glycosidase A (PNGase A, 0.2 mU) from almonds (ProGlycAn, Vienna, Austria) in 100 µl of 50 mM citrate-phosphate buffer (pH 5) was added on top of the filter unit. Incubation of the filter unit for 3 h at 37°C allowed deglycosylation of the lectin-bound glycopeptides. The deglycosylated peptides were collected in 50 µl of 40 mM ABC.

2.4. Mass spectrometry

Samples were purified on C18 Ziptips (10 µl, Millipore) prior to analysis on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). For each sample, 10 µl was introduced into an LC-MS/MS system through an Ultimate 3000 RSLC nano LC (Thermo Fisher Scientific). The sample mixture was first loaded on a trapping column (made in-house, 100 µm internal diameter (I.D.) × 20 mm, 5 µm-beads C18 Reprosil-HD; Dr. Maisch, Ammerbuch-Entringen, Germany). After

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

flushing from the trapping column, the sample was loaded on an analytical column (made in-house, 75 µm I.D. × 400 mm, 1.9 µm beads C18 Reprosil-HD; Dr. Maisch). Peptides were loaded with loading solvent (0.1% TFA, 2% ACN) and eluted with a non-linear 150 min gradient of 2-56% solvent B (0.1% formic acid in water/acetonitrile, 20/80 (v/v)) at a flow rate of 250 nl/min, followed by a 10 min wash reaching 99% solvent B and re-equilibration with solvent A (0.1% formic acid in water). The column temperature was kept constant at 50°C (CoControl 3.3.05, Sonation). The mass spectrometer was operated in data-dependent, positive ionization mode, automatically switching between MS and MS/MS acquisition for the 16 most abundant peaks in a given MS spectrum. The source voltage was set to 2.8 kV and the capillary temperature was 250°C. One MS1 scan (m/z 375-1500, AGC target 3E6 ions, maximum ion injection time of 60 ms) acquired at a resolution of 60,000 (at 200 m/z) was followed by up to 16 tandem MS scans (resolution 15,000 at 200 m/z, AGC target 1E5 ions, maximum ion injection time of 80 ms, isolation window of 1.5 m/z, fixed first mass of 145 m/z, spectrum data type: centroid) of the most intense ions fulfilling predefined selection criteria (underfill ratio 2%, intensity threshold 1.3E4, exclusion of unassigned, singly and >7 charged precursors, peptide match preferred, exclude isotopes on, dynamic exclusion time of 12 s). The HCD collision energy was set to 28% Normalized Collision Energy and the polydimethylcyclosiloxane background ion at 445.12002 Da was used for internal calibration (lock mass).

2.5. Data analysis

Raw data from each LC run were searched with MaxQuant software (version 1.5.8.3) (20) against the unannotated *N. lugens* proteome (27,571 entries, BioProject: PRJNA177647) (21) with a precursor mass tolerance set to 20 ppm for the first search and to 4.5 ppm for the main search. Trypsin/P was selected as enzyme setting, allowing for two missed cleavages, and cleavage was allowed when arginine or lysine was followed by proline. S-carbamidomethylation of cysteines was set as a fixed modification, whereas methionine oxidation, acetylation of the protein N-terminus and asparagine deamidation were set as variable modifications. The false discovery rate for peptide, protein, and site identification was set to 1%, and the minimum peptide length to 7. The minimum score threshold for both modified and unmodified peptides was set to 40, and known contaminants were excluded. Match between runs was set to True. The presence of remaining glycopeptides was assessed by tracing

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

glycan specific oxonium ions. For each sample we extracted the trace in MS2 for the masses 163.06 and 204.09 which are specific for the fragmentation of glycopeptides containing hexoses and N-acetylhexosamines, respectively. These raw traces are depicted in supplementary (Figure S1).

Collected data were analyzed using a KNIME workflow (version 3.3.2) (22). First, former N-glycopeptides were identified as peptides containing one or more N-glycosylation site(s), either N-X!-S, N-X!-T, N-X!-V or N-X!-C, where the asparagine residue at position 1 in the N-glycosylation sequon is deamidated, and X! represents any amino acid apart from proline. Only N-glycopeptides that were detected in at least two biological replicates were withheld for further analysis. Glycopeptides were grouped by protein to assess the number of unique identified N-glycoproteins. Identified N-glycoproteins from *N. lugens* adults were annotated by means of the *Drosophila melanogaster* proteome (UniProt; 21,979 entries) using BLAST+ (version 2.2.3.1, E-value threshold = 0.00001) (23).

Gene Ontology (GO) was studied for all annotated N-glycoproteins using the online tool from UniProt (www.uniprot.org). For the ConA and RSA unique N-glycoproteins, gene ontologies were analyzed using the online database DAVID (24), allowing the visualization of overrepresented cellular components and biological processes in relation to the *D. melanogaster* proteome. The molecular weight from glycoproteins was estimated using the online 'compute pI/Mw' tool from ExPASy (https://web.expasy.org/compute_pi/).

After log2 transformation, the label-free quantification (LFQ) intensities of the identified glycopeptides were compared between groups (sex and wing form) using Two-way ANOVA to reveal proteins with a significantly different intensity ($p<0.001$) using the 'peptides' file from MaxQuant in Perseus software version 1.6.2.2. The ratios for the (male/female) of the log2 transformed LFQ intensities of the proteins identified in the shotgun proteome analysis were calculated as well, and proteins detected with higher intensity in males or females were selected as regulated.

2.6. Dissection of adult tissues

Adult insects (SM and SF) were collected and kept on ice. Insects were dissected in phosphate buffered saline (PBS) under a microscope (Leica) using forceps. Collected tissues (head, gut, ovaries, testes) were rinsed three times in PBS to minimize hemolymph contamination, and were subsequently

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

collected in an Eppendorf tube containing 600 µl of lysis buffer with β-mercaptoethanol (for RNA extraction, Qiagen RNeasy mini kit) or in 600 µl PBS (for protein extraction) on ice. Tissues were collected from 50 male and 50 female insects.

2.7. Far-western blot analysis

Protein extracts from complete adult insects (LF, LM, SF, SM; 30 µg) and adult tissues (ovaries, testes; 50 µg) were loaded on two SDS-PAGE gels. One gel was used for Coomassie Brilliant blue staining, the other one for blotting. After blotting, the PVDF membrane was blocked overnight with 5% (w/v) bovine serum albumin. Then, the PVDF membrane was incubated for 3 h with ConA-solution (5 µg/ml). Subsequently, the blot was incubated for 1 h with primary antibody (anti-ConA, 1:4,000), secondary antibody (goat-anti-rabbit, 1:20,000) and peroxidase-antiperoxidase (1:400). Finally, the glycoproteins were detected with 3,3'-diaminobenzidine staining.

2.8. Gene expression analysis

Approximately 15 mg of adult insects and insect tissues (head, gut, ovaries, testes) were collected for RNA extraction. RNA was isolated using Qiagen RNeasy mini kit according to the manufacturer's protocol. The TURBO DNA-free Kit (Ambion, Life technologies, Carlsbad, CA) was used to remove trace quantities of DNA. cDNA was synthesized with SuperScript IV First-Strand Synthesis System (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA) using 1 µg of total RNA. The cDNA was diluted 1:10 for subsequent real-time qPCR, which was performed using GoTaq qPCR Master Mix (Promega, Madison, WI) and Bio-Rad CFX 96 Connect robot. The program for thermal cycling included 10 s denaturation at 95°C, 30 s annealing/extension at 60°C with melt curve analysis at the end of the run. NlRps15 and NlActin were used as reference genes. Detailed information on primer sequences and amplification efficiency can be found in Supplementary Table S1. The whole experiment was performed in four independent biological replicates for complete adult insects and in three independent biological replicates for adult tissues. Relative mRNA expression of the target genes was quantified by the BioRad CFX Manager software. Statistical analysis was conducted in qbase+ (Biogazelle, Zwijnaarde, Belgium). Primers for qPCR were designed using Primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/).

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

3. Results

3.1. Far-western blot analysis reveals differences in protein N-glycosylation between adult sexes and reproductive tissues

To evaluate whether differences in protein glycosylation are present between wing forms and/or adult sexes, as was previously observed at the N-glycome level (4), far-western blot analysis was conducted using anti-ConA as the primary antibody to visualize glycoproteins carrying high mannose and/or paucimannose N-glycans. First, complete protein extracts for the adult morphs were studied (Figure S2a). Coomassie Brilliant blue staining revealed very similar protein patterns for all adult morphs. After anti-ConA staining, visual differences between males and females can be observed, while no such differences were noted between wing forms. Next, a comparative analysis was made between extracts from ovaries and testes to examine whether the observed differences at the level of the complete insect arise from differences in the reproductive tissues (Figure S2b). Indeed, the male and female reproductive tissues show a completely different pattern of high mannosylated and/or paucimannosylated glycoproteins, with more glycoproteins observed in ovaries.

In summary, these observations indicate that there are clear differences at the glycoprotein level between adult sexes, but not between wing forms, and that ovaries contain more glycoproteins compared to testes.

3.2. Enrichment of N-glycosylation sites from *N. lugens* adults

To further compare protein glycosylation between adult sexes and wing forms of *N. lugens*, a comprehensive N-glycosylation site analysis was conducted for the different adult morphs. Complete protein extracts were enriched for glycopeptides using the lectins ConA and RSA to select peptides carrying high mannose/paucimannose N-glycans and complex N-glycans, respectively. After deglycosylation these peptides were analyzed using mass spectrometry to identify the N-glycosylation sites. ConA enrichment resulted in the detection of over 1,200 N-glycosylation sites, derived from about 600 different glycoproteins (Table 1). Enrichment with RSA led to the identification of over 300 N-glycosylation sites, from close to 200 glycoproteins. In total, over 1,300 N-glycosylation sites were identified in *N. lugens* adults (Figure 1). Almost 40% or 528 N-glycosylation sites were detected in all adult morphs and thus were present in both sexes and both wing morphs. On the total number of N-

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

glycosylation sites, 216 (or 16%) were identified both after enrichment with ConA and RSA (Supplementary Table S2). These results indicate the micro-heterogeneity of N-glycan structures in *N. lugens* proteins, where certain N-glycosylation sites can carry either high mannose/paucimannose N-glycans or complex N-glycans. A complete list of all identified N-glycosylation sites is provided in Supplementary Table S3.

GO terms were analysed using the online tool from UniProt to assess the cellular localization of the identified glycoproteins and the biological processes in which they are involved. Glycoproteins from *N. lugens* adults are mostly present in membranes and the extracellular space, as is expected for N-glycosylated proteins, and play a role in diverse biological processes such as metabolic, developmental and reproductive processes. Glycopeptides enriched with different lectins, hence decorated with different glycan structures, were most likely derived from glycoproteins with differential functions. Based on GO analysis using DAVID to obtain an overview of the overrepresentation of cellular localization and biological processes, glycoproteins carrying high mannose and/or paucimannose N-glycans were found to be mostly membrane proteins that are involved in proteolysis (GO:0006508, p=2.2E-4) and cell adhesion (GO:0007155, p=7.4E-7), while glycoproteins decorated with complex glycans are more involved in transmembrane transport (GO:0055085, p=6.0E-2) and are enriched in membranes and lipid particles.

3.3. Sex-specific glycopeptides reveal a putative role for glycosylation in reproduction

Quantitative comparisons of glycopeptides between sexes and wing forms were made based on log2 transformed LFQ intensities to evaluate the putative involvement of protein N-glycosylation in sex- and wing type-specific processes. Glycopeptides detected with differential intensity ($p<0.001$) between the two sexes were abundant for both lectins, with 68 glycopeptides enriched with ConA, and 33 with RSA (Figure 2). Glycopeptides carrying complex N-glycans were mostly male-specific (29 male, 4 female), while there was a more equal sex distribution for high mannosylated and/or paucimannosylated glycopeptides (38 male, 30 female). These sex-specific glycopeptides were, amongst others, derived from proteins important for male (seminal fluid protein) and female reproduction (putative vitellogenin receptor, vitellogenin-like protein, Nasrat). Glycopeptides detected with differential intensity between the two wing forms were scarce (1 detected for ConA and 4 for

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

RSA). The complete list of glycopeptides with differential intensity between either sexes or wing forms is provided in supplementary Table S4.

A comparison can also be made between the information obtained from the mass spectrometry data and the far-western blot analyses. Given that the far-western blots visualized the high mannosylated and/or paucimannosylated glycoproteins, only the sex-related glycopeptides detected with ConA can be evaluated. The 68 sex-related glycopeptides were derived from 50 unique glycoproteins; among which 13 female-specific and 37 male-specific glycoproteins. The molecular weights of these 50 glycoproteins were calculated based on their amino acid sequences, allowing to determine the molecular weights of the non-glycosylated proteins. Taking into account that the molecular weight of the high mannose N-glycan $GlcNAc_2Man_9$ is approximately 2 kDa, the molecular weight of each glycoprotein was estimated (taking into account the number of identified N-glycosylation sites for each protein) (Supplementary Table S5). For the males, a faint band around 38 kDa can be observed in complete insects and is also clearly present in the testes, but not observed in females or ovaries. In testes, two other distinct bands are visible at 25 kDa and 65 kDa. All these polypeptides react with ConA and could possibly correspond to the identified male-specific glycoproteins James bond (37 kDa), Dumpy (26 kDa) and Amino acid transporter (67 kDa), respectively. For females, the high number of glycoproteins visualized in the ovaries is prominent, with most glycoproteins appearing at molecular weights ranging from 130 to 180 kDa, whereas no proteins smaller than 37 kDa were detected. These observations correspond well to the female-specific glycoproteins identified after mass spectrometry analysis, which are almost all glycoproteins with a molecular weight of over 100 kDa. At present, only assumptions can be made with respect to interpretation of the bands detected on the western blots and the molecular masses of the glycoproteins. Further post-translational processing of the proteins could occur, which will affect their electrophoretic behavior on SDS-PAGE. Therefore, future experiments should focus on the identification and characterization of the ConA reactive glycoproteins.

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

3.4. Comparison of glycopeptides between adult males and females reveals altered glycosylation of proteins shared across sexes

Complete protein extracts were analyzed using mass spectrometry for the different adult morphs of *N. lugens*, and this shotgun proteome analysis resulted in the identification of over 3,300 proteins (Table 1). The log2 transformed LFQ intensities from the proteins identified and quantified in both male and female adult samples were compared between the sexes. Further comparison with the identified sex-specific glycopeptides allowed to determine whether differences between sexes on the glycopeptide level resulted from differential glycosylation between the sexes or from differences in protein levels. The 99 sex-specific glycopeptides were derived from 69 unique glycoproteins, of which 49 (71%) were also present in the shotgun proteomics data. For 22 of these glycoproteins, no statistically significant differences ($p>0.05$) could be observed between the protein intensities from male and female samples, indicating that the observed differences on the glycopeptide level result from differential glycosylation between sexes (Table 2). Upregulation of the protein in either males or females was observed for 13 glycoproteins (Table 2), whereas complete absence in either of the two sexes was observed for 14 other glycoproteins. After correction of the regulation of the protein, differential levels of glycopeptides between sexes observed for these 27 proteins could be assigned to the (up)regulation of the protein. All glycoproteins involved in the reproductive process belong to this latter group.

3.5. Transcript expression analyses confirm observed differences between sexes on N-glycopeptide level

Striking differences in protein N-glycosylation were observed between sexes, while there were no obvious discrepancies between wing forms. Both sex-specific glycoproteins as well as proteins that are differentially glycosylated in the two sexes were identified. To confirm this observation, transcript expression was studied for adult morphs and adult tissues (head, gut, ovaries and testes) for 5 selected glycoproteins. These glycoproteins were identified by glycopeptides only detected in either male or female adults or with a statistically different glycopeptide intensity ($p<0.001$) between sexes, after enrichment with ConA, RSA or both (Table 3). Transcript analysis from complete adult insects showed no differences between wing forms, since both long- and short-winged adults cluster together

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

and no statistical differences could be observed (One-way ANOVA, n=4) (Figure S3). Two transcripts were found to be female-specific; Nasrat (p=9.36E-10) and Flyers-cup (p=1.97E-7), while James bond was found to be male-specific (p=3.20E-8) (Figure 3). The other two transcripts showed no significant differences between males and females at the level of the whole insect.

Since no differences on transcript level were observed between wing types, expression in adult tissues was only evaluated for short-winged insects. For all studied transcripts, differences between tissues were observed (One-way ANOVA, n=3) (Figure 4). For the female-specific transcripts, expression of Flyers-cup was high in all female tissues (p=2.60E-4), while Nasrat was found to be ovary-specific (p=4.97E-6). For two other transcripts, the expression was the highest in the testes; James bond (p=4.97E-6) and Aminopeptidase (p=2.19E-4). CG5276 was expressed at the highest levels in the male gut (p=1.56E-3) and in testes.

4. Discussion

Sex-related N-glycosylation has already been reported in several organisms, including humans (25), parasitic nematodes (26) and trematodes (27). Recently, differences in protein N-glycosylation between sexes were also observed for the first time in an insect species, in particular the pest insect *N. lugens* (4). All these studies only focused on the N-glycomes from males and females, while here we opted to identify the proteins carrying these differential carbohydrate structures to unravel the relevance of these observed differences between sexes at the glycoprotein level. Experiments were conducted in *N. lugens*, a hemipteran model insect, where male and female adults can appear as two wing morphs; the flightless short-winged morph and the migratory long-winged morph. N-glycosylation sites were identified from the different adult morphs, including both sexes and wing types, allowing for the evaluation of the putative involvement and significance of protein N-glycosylation in sex- and wing type-specific processes.

Far-western blot experiments were first conducted to test the hypothesis that protein N-glycosylation can be sex- and/or wing type-dependent in *N. lugens* adults. Complete protein extracts from the adult insects as well as male and female reproductive tissues were investigated. Obvious differences were observed in the presence of glycoproteins between males and females, independent of their wing form. Furthermore, clear differences between glycoproteins present in ovaries and testes

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

were visualized, where ovaries seem to contain a high amount of high mannosylated and/or paucimannosylated glycoproteins. This western blot result is a first indication that there is no differential glycosylation between wing types, but protein N-glycosylation is sex-related in *N. lugens* adults and could play a role in (female) reproduction.

Glycopeptides were enriched by the N-glyco-FASP method using two lectins interacting with different N-glycan structures, allowing for the further identification of the N-glycosylation sites. ConA specifically interacts with high mannose and paucimannose N-glycans, while RSA binds to complex N-glycans with terminal Gal or GalNAc residues. In total, over 1,300 N-glycosylation sites were identified, most of them via enrichment with ConA. This result is not surprising since high mannose and paucimannose N-glycans together make up >90% of the insect N-glycan profile (4, 15, 28, 29). Overlap between the datasets obtained after enrichment with ConA and RSA indicates that certain glycoproteins may occur in different glycoforms, suggesting micro-heterogeneity of N-glycan structures in *N. lugens* proteins.

The N-glycosylation sites identified were derived from glycoproteins playing a role in diverse biological processes such as metabolic, developmental and reproductive processes, demonstrating the importance of protein N-glycosylation in *N. lugens* adults. Interestingly, N-glycosylation sites that were uniquely identified after enrichment with ConA or RSA were derived from glycoproteins with differential functions. Glycoproteins decorated with high mannose and/or paucimannose N-glycans are mostly membrane proteins that are involved in adhesion and proteolysis. While glycoproteins carrying complex N-glycans, as enriched with RSA, play a role in transmembrane transport, and are present in lipid particles and membranes. A similar observation was made in *Spodoptera* midgut cells; high and oligomannose N-glycans appeared to be more abundant in the basal region of the cells, which is involved in cell-cell and cell-matrix interactions (30). On the other hand, glycans with terminal Gal/GalNAc residues were more abundant in the apical region of the columnar midgut cells, the primary site for many physiological, biological and biochemical interactions (e.g. release of digestive enzymes, uptake of nutrients). Furthermore, the addition of carbohydrate structures to proteins can influence their properties (31), and thus the biological activity of a particular protein can differ between the glycosylated and non-glycosylated form, as well as between different glycoforms of the

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

same glycoprotein. For example in *Drosophila*, diminished N-glycosylation of the protein chaoptin resulted in an abolished activity of the protein (32), while sialylation of the voltage gated K^+ channel SBb demonstrated that sialylation may increase gating efficiency (33).

A quantitative comparison of N-glycopeptides between the two sexes and wing types of *N. lugens* adults revealed the presence of almost 100 sex-specific glycopeptides, while wing type-specific glycopeptides were scarce. The sex-specific glycopeptides carried high mannose/paucimannose N-glycans, complex N-glycans or both. Glycopeptides carrying complex N-glycans were mostly male-enriched, which is in agreement with the presence of more Gal-terminal N-glycans in males (4). Comparative analyses between glycopeptides and the presence of glycoproteins in complete protein extracts allowed to determine whether differences between sexes at the glycopeptide level resulted from differences in glycoprotein levels or from differential glycosylation between the sexes. This comparison revealed that differences in protein glycosylation between sexes are caused for approximately 55% by differences observed at protein level, while 45% can be attributed to differential glycosylation. For example, glycoproteins involved in insect reproduction belong to the first category of proteins, and are either unique for one sex or much higher expressed in one of both sexes. For males, only one glycoprotein (seminal fluid protein), both carrying high mannose/paucimannose N-glycans and complex N-glycans, was identified that plays a role in male reproduction. Insect seminal fluid proteins are transferred to females during mating and induce several physiological and behavioral post-mating changes in females (34). In contrast to males, several high abundant glycoproteins were identified in females, carrying high mannose/paucimannose N-glycans that are involved in female reproduction; putative vitellogenin receptor (35), vitellogenin-like protein (36), and Nasrat (37-39). The vitellogenin-like protein also carries complex N-glycans, next to the high mannose/paucimannose N-glycans. These findings are in line with the detection of a high amount of glycoproteins in ovaries by western blot analyses, and support the hypothesis that female adults obtain a female-unique N-glycan fingerprint with increased levels of high mannose N-glycans due to the presence of female-unique glycoproteins that play a role in female reproduction (4). Glycoproteins that are equally produced in both sexes, but show differential glycosylation between males and females are of high interest. Of these, 22 glycoproteins were identified of which two are yet

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

uncharacterized. The other glycoproteins are putatively involved in metabolism, stress response and developmental processes. Further investigation of the function of these glycoproteins in male and female adults could indicate whether this differential protein glycosylation also results in a different biological function for these glycoproteins between sexes.

To verify the observations from the proteomic analysis, a selection of glycoproteins identified by male- and female-specific glycopeptides, carrying high mannose/paucimannose N-glycans and/or complex N-glycans, was also analyzed for transcript expression.

Three male-specific glycoproteins were selected: James bond, CG5276, and Aminopeptidase. For the first two male-specific glycoproteins it was previously shown that they could play a role in male reproduction. The fatty acid elongase James bond was already proven to be essential for sex pheromone synthesis and male fertility in *Drosophila* (40). CG5276 is a calcium-activated nucleosidase with apyrase activity from *D. melanogaster*, found on the cell surface and as a secreted protein (41). The expression of this enzyme in *Drosophila* salivary glands and testes suggests an important role in these organs, and a putative role in male reproduction. Aminopeptidase was previously shown to have metalloendopeptidase activity (42), but there is no evidence for a role in reproduction.

Two female-specific glycoproteins were further investigated: Nasrat and Flyers-cup. The role of Nasrat in female reproduction has already been extensively studied in *D. melanogaster* (35-37), and the protein was proven to be essential for the vitelline membrane formation. In *D. melanogaster*, Flyers-cup transcripts were detected in spermatids (43), and its involvement in memory formation has also previously been implicated (44). In contrast to *Drosophila*, Flyers-cup is much more abundant in *Nilaparvata* female adults compared to males, based on information from the shotgun proteome analysis. Expression of the selected glycoproteins was studied at transcript level for both complete adult insects and adult tissues. Expression analysis from complete adult insects confirmed the sex-specific expression as seen on glycopeptide level for Nasrat, Flyers-cup and James bond, while no differences in transcript levels were observed for aminopeptidase and CG5276. For Aminopeptidase, James bond and Nasrat, gene expression was statistically highest in the reproductive tissues, supporting a potential role for these proteins in *Nilaparvata* reproduction. CG5276 showed highest

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

expressed in male gut and testes. Flyers-cup revealed a higher expression in all female tissues compared to the male tissues. The combination of transcript expression data from complete insects and adult tissues confirms that the observed sex-differences in *N. lugens* protein glycosylation are the result from both the presence of sex-specific glycoproteins as well as from differential glycosylation of proteins occurring in both male and female adults. Given the contradictions for the tissue specificity of Flyers-cup between *D. melanogaster* and *N. lugens*, it would be of interest to focus further experiments on the functionality of this protein in *N. lugens*.

In summary, our study presents original data on the N-glycosylation sites from *N. lugens* adults over sexes and wing types. Although almost no differences in N-glycopeptides were observed between wing types, several sex-specific glycopeptides have been identified. Next to the N-glycome, differences between males and females have now also been observed on N-glycosylation site level, strengthening the belief that protein N-glycosylation is sex-related in this pest insect. This sex-related protein glycosylation in *N. lugens* adults is believed to result from both the presence of sex-specific glycoproteins as well as from altered glycosylation of proteins shared across sexes. Furthermore, the discovery that several proteins involved in (female) reproduction are glycosylated implies that protein N-glycosylation can play a role during insect reproduction. Our research provides novel insights into planthopper's biology, and offers information for future research on planthopper's development, reproduction and potential control methods.

Acknowledgments

This research was funded by the Research Foundation-Flanders (Belgium) and the Special Research Fund of Ghent University (Belgium).

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013657. All spectra and search results were uploaded to MS-Viewer with following accessions; ke88onqwpl (shotgun samples), hgenowvedt (ConA-enriched samples), and gfrwdgysub (RSA-enriched samples).

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

References

1. Bao, Y. Y., Qu, L. Y., Zhao, D., Chen, L. B., Jin, H. Y., Xu, L. M., Cheng, J. A., Zhang, C. X. (2013) The genome- and transcriptome-wide analysis of innate immunity in the brown planthopper, *Nilaparvata lugens*. BMC Genomics. 14:10.1186/1471-2164-14-160

2. Li, J., Chen, Q., Lin, Y., Jiang, T., Wu, G., Hua, H. (2011) RNA interference in *Nilaparvata lugens* (Homoptera: Delphacidae) based on dsRNA ingestion. Pest Manag. Sci. 67, 852-859

3. Zera, A. J. (2004) The endocrine regulation of wing polymorphism in insects: State of the art, recent surprises, and future directions. Integr. Comp. Biol. 43, 607-616

4. Scheys, F., De Schutter, K., Shen, Y., Yu, N., Smargiasso, N., De Pauw, E., Van Damme, E. J. M., Smagghe, G. (2019) The N-glycome of the hemipteran pest insect *Nilaparvata lugens* reveals unexpected sex differences. Insect Biochem. Mol. Biol. 107, 39-45

5. Koles, K., Lim, J. M., Aoki, K., Porterfield, M., Tiemeyer, M., Wells, L., Panin, V. (2007) Identification of N-glycosylated proteins from the central nervous system of *Drosophila melanogaster*. Glycobiology. 17, 1388-1403

6. Vandenborre, G., Van Damme, E. J. M., Ghesquière, B., Menschaert, G., Hamshou, M., Rao, R. N., Gevaert, K., Smagghe G. (2010) Glycosylation signatures in *Drosophila*: fishing with lectins. J. Prot. Res. 9, 3235-3242

7. Vandenborre, G., Smagghe, G., Ghesquière, B., Menschaert, G., Rao, R. N., Gevaert, K., Van Damme, E. J. M. (2011) Diversity in protein glycosylation among insect species. PLoS ONE. 10.1371/journal.pone.0016682

8. Scheys, F., Van Damme, E. J. M., De Schutter, K., Staes, A., Gevaert, K., Smagghe, G. (2018) Evolutionarily conserved and species-specific glycoproteins in the N-glycoproteomes of diverse insect species. Insect Biochem. Mol. Biol. 100, 22-29

9. Zielinska, D. F., Gnad, F., Schropp, K., Wiśniewski, J. R., Mann, M. (2012) Mapping N-glycosylation sites across seven evolutionary distant species reveals a divergent substrate proteome despite a common core machinery. Mol. Cell. 46, 542-548

10. Zielinska, D. F., Gnad, F., Wiśniewski, J. R., Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell. 141, 897-907

11. Katoh, T., Tiemeyer, M. (2013) The N's and O's of *Drosophila* glycoprotein glycobiology. Glycoconj. J. 30, 57-66

12. Kurz, S., Aoki, K., Jin, C., Karlsson, N. G., Tiemeyer, M., Wilson, I. B., Paschinger, K. (2015) Targeted release and fractionation reveal glucoronylated and sulphated N- and O-glycans in larvae of dipteran insects. J. Proteomics. 126, 172-188

13. North, S. J., Koles, K., Hembd, C., Morris, H. R., Dell, A., Panin, V. M., Haslam, S. M. (2006) Glycomic studies of *Drosophila melanogaster* embryos. Glycoconj. J. 23, 345-354

14. Stanton, R., Hykollari, A., Eckmair, B., Malzl, D., Dragosits, M., Palmberger, D., Wang, P.,

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Wilson, I. B., Paschinger, K. (2017) The underestimated N-glycomes of lepidopteran species. Biochim. Biophys. Acta. 1861, 699-714

15. Walski, T., Van Damme, E. J. M., Smargiasso, N., Christiaens, O., De Pauw, E., Smagghe, G. (2016) Protein N-glycosylation and N-glycan trimming are required for postembryonic development of the pest beetle *Tribolium castaneum*. Sci. Rep. 10.1038/srep35151

16. Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., Hermjakob, H. (2016) 2016 update of the PRIDE database and related tools. Nucleic Acids Res. 44, D447-D456

17. Van Damme, E. J. M., Peumans, W. J., Pusztai, A., Bardocz, S. (1998) Handbook of plant lectins: Properties and biomedical applications. Wiley, Chichester

18. Hamshou, M., Van Damme, E. J. M., Smagghe, G. (2014) Entomotoxic effects of fungal lectin from *Rhizoctonia solani* towards *Spodoptera littoralis*. Fungal Biology. 114, 34-40

19. Wiśniewski, J. R., Zougman, A., Nagaraj, N., Mann, M. (2009) Universal sample preparation method for proteome analysis. Nat. Methods. 6, 359-362

20. Cox, J., Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367-1372

21. Xue, J., et al. (2014) Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome biology. 15, 10.1186/s13059-014-0521-0

22. Warr, W. A. (2012) Scientific workflow systems: Pipeline Pilot and KNIME. J. Comput. Aided. Mol. Des. 26, 801-804

23. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L. (2009) BLAST+: Architecture and applications. BMC Bioinformatics. 10.1186/1471-2105-10-421

24. Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., Lempicki, R. A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, R60

25. Qin, Y., Zhong, Y., Zhu, M., Dang, L., Yu, H., Chen, Z., Chen, W., Wang, X., Zhang, Y., Li, Z. (2013) Age- and sex-associated differences in the glycopatterns of human salivary glycoproteins and their roles against Influenza A virus. J. Proteome Res. 12, 2742-2754

26. Jiménez-Castells, C., Vanbeselaere, J., Kohlhuber, S., Ruttkowski, B., Joachim, A., Paschinger, K. (2017) Gender and developmental specific N-glycomes of the porcine parasite *Oesophagostomum dentatum*. BBA Gen. Subj. 1861, 418-430

27. Wuhrer, M., Koeleman, C. A. M., Fitzpatrick, J. M., Hoffmann, K. F., Deelder, A. M., Hokke, C. H. (2006) Gender-specific expression of complex-type N-glycans in schistosomes. Glycobiology 16, 991-1006

28. Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., Tiemeyer, M. (2007) Dynamic

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo. J. Biol. Chem. 282, 9127-9142

29. Fabini, G., Freilinger, A., Altmann, F., Wilson, I. B. (2001) Identification of core alpha 1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from *Drosophila melanogaster*. Potential basis of the neural anti-horseradish peroxidase epitope. J. Biol. Chem. 276, 28058-28067

30. Walski, T., De Schutter, K., Cappelle, K., Van Damme, E. J. M., Smagghe, G. (2017) Distribution of glycan motifs at the surface of midgut cells in the cotton leafworm (*Spodoptera littoralis*) demonstrated by lectin binding. Front. Physiol. 8, 10.3389/fphys.2017.01020

31. Staudacher, E. (2015) Mucin-type O-glycosylation in invertebrates. Molecules. 20, 10622-10640

32. Hiraj-Fujita, Y., Yamamoto-Hino, M., Kanie, O., Goto, S. (2008). N-glycosylation of the *Drosophila* neural protein Chaoptin is essential for its stability, cell surface transport and adhesive activity. FEBS Lett., 582, 2572-2576.

33. Johnson, D., Bennett, E.S. (2008) Gating of the shaker potassium channel is modulated differentially by N-glycosylation and sialic acids. Eur. J. Appl. Physiol. 456, 393-405.

34. Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D., Wolfner, M. F. (2011) Insect seminal fluid proteins: identification and functions. Annu. Rev. Entomol. 56, 21-40

35. Lu, K., Shu, Y., Zhou, J., Zhang, X., Zhang, X., Chen, M., Yao, Q., Zhou, Q., Zhang, W. (2015) Molecular characterization and RNA interference analysis of vitellogenin receptor from *Nilaparvata lugens* (Stål). J. Insect. Physiol. 73, 20-29

36. Tufail, M., Naeemullah, M., Elmogy, M., Sharma, P. N., Takeda, M., Nakamura, C. (2010) Molecular cloning, transcriptional regulation, and differential expression profiling of vitellogenin in two wing-morphs of the brown planthopper, *Nilaparvata lugens* Stål (Hemiptera: Delphacidae). Insect. Mol. Biol. 19, 787-798

37. Cernilogar, F. M., Fabbri, F., Andrenacci, D., Taddei, C., Gargiulo, G. (2001) *Drosophila* vitelline membrane cross-linking requires the *fs(1)Nasrat*, *fs(1)polehole* and chorion genes activities. Dev. Genes. Evo. 211, 573-580

38. Degelmann, A., Hardy, P. A., Mahowald, A. P. (1990) Genetic analysis of two female-sterile loci affecting eggshell integrity and embryonic pattern formation in *Drosophila melanogaster*. Genetics. 126, 427-434

39. Jiménez, G., Gonzáles-Reyes, A., Casanova, J. (2002) Cell surface proteins Nasrat and Polehole stabilize the Torso-like extracellular determinant in *Drosophila* oogenesis. Genes Dev. 16, 913-918

40. Ng, W. C., Chin, J. S. R., Tan, K. J., Yew, J. Y. (2015) The fatty acid elongase Bond is essential for *Drosophila* sex pheromone synthesis and male fertility. Nat. Comm. 6, 10.1038/ncomms9263

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

41. Fenckova, M., Hobizalova, R., Fric, Z. F., Dolezal, T. (2011) Functional characterization of ecto-5'-nucleotidases and apyrases in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. 41, 956-967

42. Gaudet, P., Livstone, M. S., Lewis, S. E., Thomas, P. D. (2011) Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief. Bioinform. 12, 449-462

43. Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., White-Cooper, H. (2008) Post-meiotic transcription in *Drosophila* testes. Development. 135, 1897-1902

44. Walkinshaw, E., Gai, Y., Farkas, C., Richter, D., Nicholas, E., Keleman, K., Davis, R. L. (2015) Identification of genes that promote or inhibit olfactory memory formation in *Drosophila*. Genetics. 199, 1173-1182

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Tables

Table 1. Enrichment of N-glycopeptides from *N. lugens* adult morphs.

Sample	Identified proteins	Enriched glycosites		Enriched glycopeptides		Enriched glycoproteins	
		ConA	RSA	ConA	RSA	ConA	RSA
LF	2,251	870	124	863	119	501	96
LM	2,242	1,012	225	997	214	532	161
SF	2,421	679	232	677	217	411	158
SM	2,807	905	249	893	237	507	179
Total	**3,322**	**1,269**	**315**	**1,216**	**297**	**606**	**205**
		1,368		**1,352**		**658**	

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Table 2. Sex-specifc glycopeptides result both from increased glycosylation of their corresponding proteins as well as from increased protein levels in one the sexes

Sex	Accession	Protein description	Male/Female
Female	NLU001824.1	Putative vitellogenin receptor	ND male
Female	NLU012863.1	Uncharacterized protein	ND male
Female	NLU006222.1	Female sterile (1) Nasrat	ND male
Female	NLU004755.1	Flyers-cup	0.75
Female	NLU019204.1	Vitellogenin-like	0.82
Female	NLU011477.1	Vitellogenin-like	0.86
Female	NLU011555.1	Uncharacterized protein	0.94
Female	NLU024350.1	Uncharacterized protein	0.96
Female	NLU010113.1	Alpha-mannosidase	0.96
Female	NLU020778.1	Transmembrane GTPase Marf	1.00
Female	NLU002989.3	Receptor expression-enhancing protein	1.02
Male	NLU020415.2	Maltase A8	0.99
Male	NLU001320.1	LD23292p	1.00
Male	NLU002176.1	Aminopeptidase	1.00
Male	NLU023838.1	Maltase A3	1.00
Male	NLU027145.1	CG5276	1.00
Male	NLU014188.3	Amino acid transporter	1.01
Male	NLU003276.1	Cardinal	1.01
Male	NLU008295.1	RNA-binding protein cabeza-like	1.01
Male	NLU011751.3	RH72323p	1.01
Male	NLU007664.2	UDP-glucuronosyltransferase	1.01
Male	NLU001964.1	Basigin	1.01
Male	NLU024229.1	Neurotactin	1.02
Male	NLU016734.1	Kazachoc	1.03
Male	NLU020709.1	Dipeptidase	1.03
Male	NLU011804.1	RE14947p	1.04
Male	NLU005014.1	Neural lazarillo	1.05
Male	NLU011657.1	Spaetzle-processing enzyme	1.07
Male	NLU017497.1	Carboxypeptidase B-like	1.11
Male	NLU001528.1	Uncharacterized protein	1.12
Male	NLU001918.1	Uncharacterized protein	1.12
Male	NLU027835.1	Dumpy	1.12
Male	NLU024819.2	RE52890p	1.17
Male	NLU012529.1	Uncharacterized protein	1.18
Male	NLU021589.1	Uncharacterized protein	1.23
Male	NLU002851.1	Seminal fluid protein	1.23
Male	NLU012421.1	Carboxylic ester hydrolase	1.23
Male	NLU014493.1	Carboxylic ester hydrolase	1.32
Male	NLU003617.2	ATPase ASNA1 homolog	ND female
Male	NLU003798.1	Kekkon-2	ND female
Male	NLU003811.1	Transmembrane 9 superfamily member	ND female
Male	NLU006117.1	Esterase P	ND female
Male	NLU012109.1	Uncharacterized protein	ND female

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Sex	Accession	Protein description	Male/Female
Male	NLU012692.1	Nuclear factor NF-kappa-B p110 subunit	ND female
Male	NLU014036.1	Uncharacterized protein	ND female
Male	NLU018639.1	Vesicular acetylcholine transporter	ND female
Male	NLU020349.1	LD28763p	ND female
Male	NLU020863.1	Thioredoxin reductase 1	ND female
Male	NLU028220.1	Amino acid transporter	ND female

The column 'Male/Female' contains the ratios of the log2 transformed LFQ intensities from the proteins identified in the shotgun proteome samples from male and female adults. The 95% confidence interval is [0.93; 1.07], where <0.93 means upregulated in females compared to males (yellow), and >1.07 means upregulated in males compared to females (blue). ND male – not detected in male samples, ND female- not detected in female samples.

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Table 3. Selection of sex-specific glycoproteins for transcript expression analysis.

N. lugens **accession**	***D. melanogaster*** **accession**	**Protein name**	**Sex specificity**	**Lectin**	**Sex-specific N-glycosites**
NLU002176.1	Q8INH5	Aminopeptidase	Male	RSA	1
NLU004755.1	Q9V3X1	Flyers-cup	Female	ConA, RSA	7
NLU006222.1	O76904	Nasrat	Female	ConA	2
NLU020486.2	Q9VCY7	James bond	Male	ConA	2
NLU027145.1	Q9VGN8	CG5276	Male	ConA	1

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

Figures and figure legends

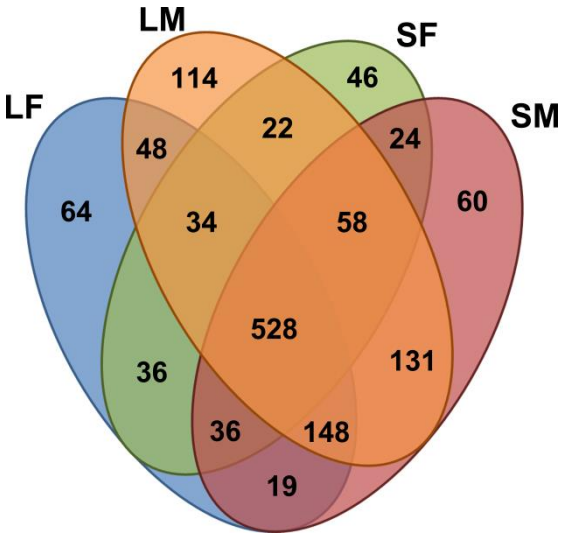

Figure 1. Distribution of N-glycosylation sites in the different adult morphs of *N. lugens*.

N-glycosylation sites were identified from *N. lugens* adult morphs using mass spectrometry. The Venn diagram shows the distribution of the N-glycosylation sites over the different adult morphs; long-winged females (LF, 913), long-winged males (LM, 1,083), short-winged females (SF, 784), and short-winged males (SM, 1,004).

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

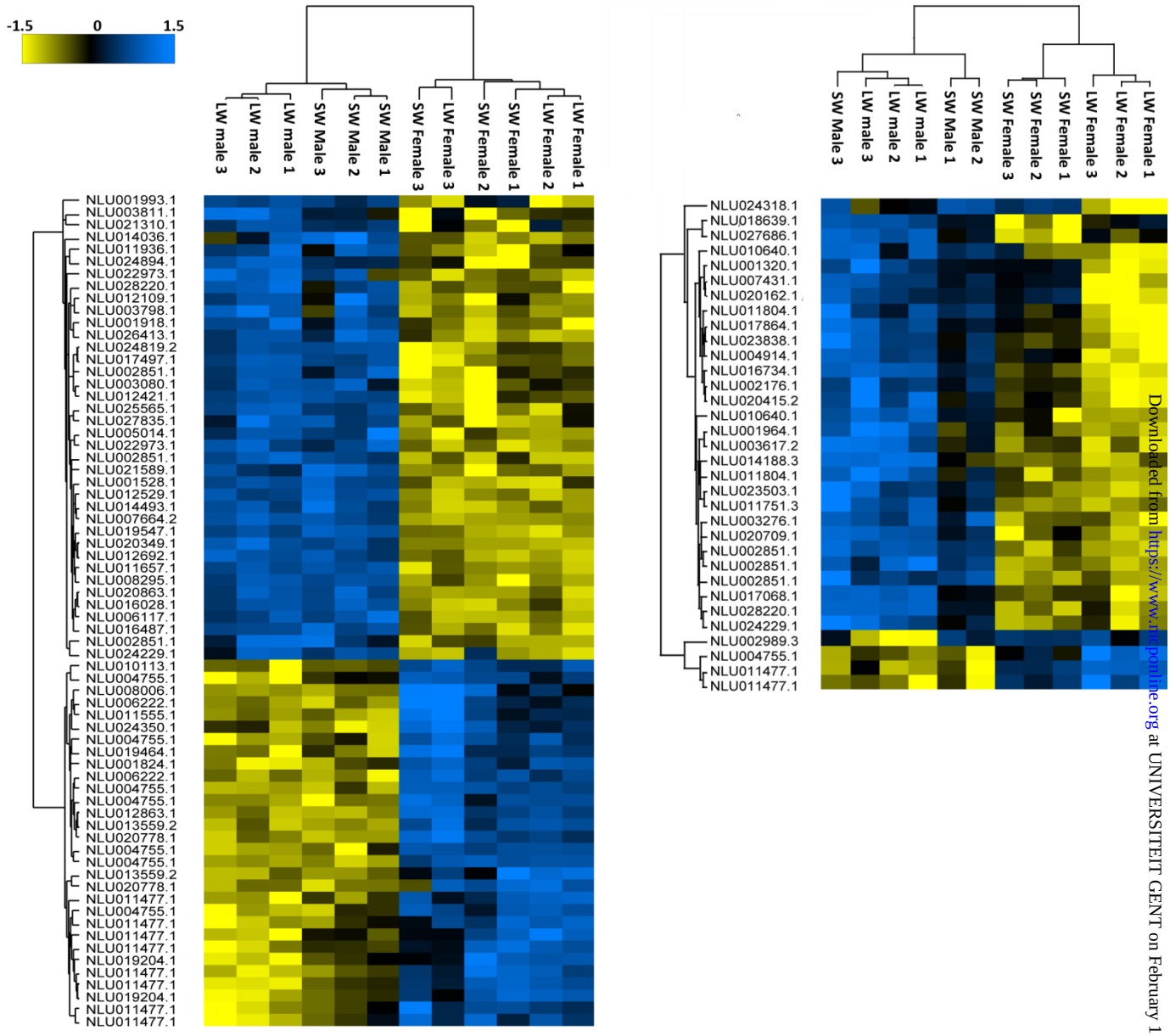

Figure 2. Heatmap of sex-specific glycopeptides.

Quantitative comparison based on LFQ intensities between male and female *N. lugens* adults using Two-way ANOVA in Perseus. For each adult morph, three biological replicates were analysed. Glycopeptides with a statistical different intensity ($p<0.001$) between sexes are visualized for glycopeptides enriched with ConA (left) and RSA (right) and clustered by sample. The protein accession is represented of the protein from which the glycopeptide is derived. Protein sequences are available on ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100139/. Accession numbers that occur more than once, indicate multiple glycosylation sites on the given protein. The scale represents the Z-score of the log2 LFQ intensities.

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

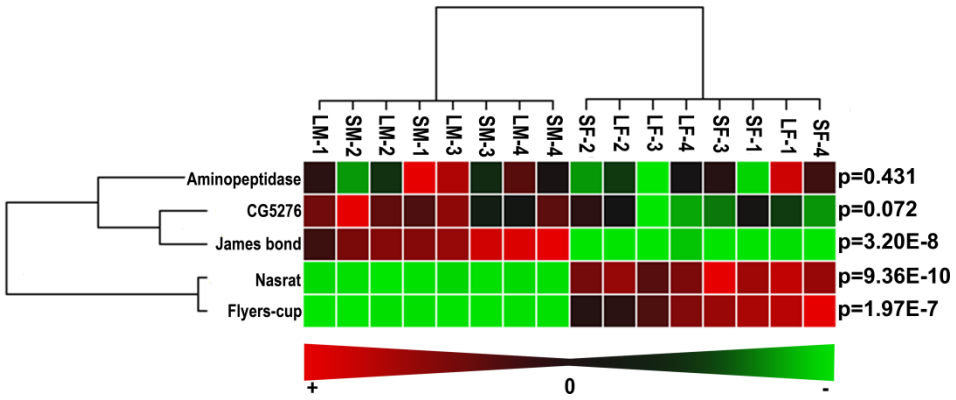

Figure 3. Clustergram of transcript expression of sex-specific glycoproteins from *N. lugens* adult morphs.

Transcript levels for 5 selected glycoproteins were analysed in *N. lugens* adult insects (long-winged females - LF, long-winged males - LM, short-winged females - SF, short-winged males - SM). NlRps15 and NlActin were used as reference genes. Statistical differences were evaluated using One-way ANOVA test (n=4). The data are shown in a hierarchy based on the degree of similarity of expression for different samples and targets; upregulation (red), downregulation (green) or no regulation (black).

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020

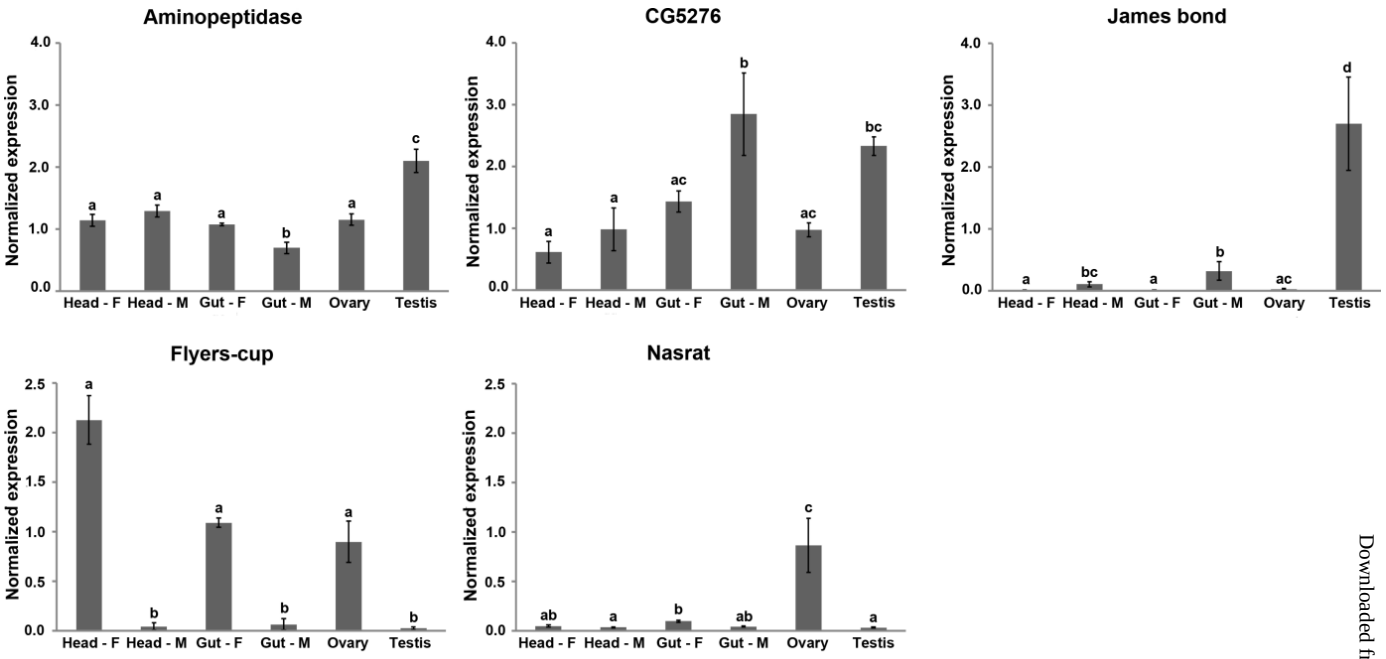

Figure 4. Transcript expression of sex-specific glycoproteins in tissues from *N. lugens* short-winged adults.

Transcript levels for 5 selected glycoproteins were analysed in *N. lugens* short-winged adult tissues (head, gut, ovaries, testes). NlRps15 and NlActin were used as reference genes. The bar chart shows the average ± standard deviation (n = 3). Statistical differences were evaluated using One-way ANOVA with post-hoc Tukey test, and are marked with a different letter (a, b, c, d).

Downloaded from https://www.mcponline.org at UNIVERSITEIT GENT on February 11, 2020