Current Research in Insect Science 4 (2023) 100064

Contents lists available at ScienceDirect

Current Research in Insect Science

journal homepage: www.elsevier.com/locate/cris

# Characterization of the swede midge, *Contarinia nasturtii*, first instar larval salivary gland transcriptome

Boyd A. Mori [a,*], Cathy Coutu [b], Martin A. Erlandson [b], Dwayne D. Hegedus [b,*]

[a] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada
[b] Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, Saskatoon, SK S7N 0×2, Canada

ARTICLE INFO

*Keywords:*
Gall midge
RNA-Seq
Extra-oral digestion
Effectors
Plant hormones

ABSTRACT

Proteins in saliva of gall-forming insect larvae govern insect-host plant interactions. *Contarinia nasturtii*, the swede midge, is a pest of brassicaceous vegetables (cabbage, cauliflower, broccoli) and canola. We examined the salivary gland (SG) transcriptome of first instar larvae reared on *Brassica napus* and catalogued genes encoding secreted proteins that may contribute to the initial stages of larval establishment, the synthesis of plant growth hormones, extra-oral digestion and evasion of host defenses. A significant portion of the secreted proteins with unknown functions were unique to *C. nasturtii* and were often members of larger gene families organized in genomic clusters with conservation patterns suggesting that they are undergoing selection.

## 1. Introduction

Insect salivary gland (SG) secretions allow intimate host associations. Larval SG secretions of gall midges cause changes in host tissues resulting in specialized structures (Giron et al., 2016). These range from nutritive tissues to enclosed structures in which insects develop (Stone and Schönrogge, 2003); however, no SG proteins are associated with gall development. Gall formation is influenced by phytohormones (Tooker and Helms, 2014) from the insect (Yamaguchi et al., 2012; Suzuki et al., 2014), host (Takeda et al., 2021) or endosymbionts (Kaiser et al., 2010).

Effectors are molecules that augment insect-host plant interactions. In this context, gall insects release effectors that aid infestation; however, they may also induce or suppress defense responses. In Hessian fly, *Mayetiola destructor*, compatible/incompatible interactions are mediated by host resistance proteins (Prather et al., 2022) recognizing insect salivary proteins (Stuart et al., 2012). 60% of *M. destructor* SG mRNAs encode secreted proteins (Chen et al., 2008) and 5% of the genome encodes effectors (Zhao et al., 2015). Effector genes are under selective pressure (Chen et al., 2010), as would all genes involved in host-pathogen or plant-herbivore interactions based on reciprocal adaptations, and correlated with deployment of resistant varieties (Johnson et al., 2014).

*Contarinia nasturtii* (Figure 1) affects brassicaceous crops in Eurasia and North America (Hallett and Heal, 2001; Chen et al., 2011). Little is known about its feeding, but salivary secretions are involved in extra-oral digestion and plant manipulation (Hallett and Heal, 2001). Larvae infest meristematic tissue leading to unmarketable vegetables (Stratton et al., 2018) or reduced canola yield (Hallett, 2017). Here, we use a SG transcriptome and genome sequence (Mori et al., 2021) to catalogue secreted SG proteins (SSGPs) and effectors involved in host interactions.

## 2. Materials and Methods

### 2.1. Insects

*Contarinia nasturtii* were obtained from a colony established from insects collected in Ontario, Canada (43.6409°, -80.3979°) and maintained on *Brassica napus* var. AC Excel under controlled conditions (21 ± 2°C, ~70% RH, 16:8 h light:dark) for ca. 18 generations. To obtain larvae, two *B. napus* seedlings were grown in 15.25 cm diameter pots

*Abbreviations:* ARMET, Arginine-rich mutated in early stages of tumor; BLAST, basic local alignment and search tool; BUSCO, benchmarking universal single-copy Orthologs; EC, enzyme classification; GO, gene ontology; IAA, indole acetic acid; Iaa1d, indole-3-acetaldehyde; IAOx, indole-3-acetaldoxime; KEGG, kyoto encyclopedia of genes and genomes; ORF, open reading frame; Pfam, protein family; SSGP, secreted salivary gland proteins; TMD, transmembrane domain; TPM, transcripts per million.

* Corresponding authors.
*E-mail addresses:* bmori@ualberta.ca (B.A. Mori), dwayne.hegedus@canada.ca (D.D. Hegedus).

https://doi.org/10.1016/j.cris.2023.100064
Received 27 October 2022; Received in revised form 9 June 2023; Accepted 11 July 2023
Available online 13 July 2023

with soil-less potting mix (Stringham, 1971) for 21 days (4-5 leaf stage) under the same conditions as the colony. A single pot was exposed to 20 female and 10 male adults for 24 hours in a cage (47.5 cm x 47.5 cm X 47.5 cm) (Bugdorm, Megaview Science Co., Ltd., Taiwan). Newly eclosed, first instar larvae that were feeding or in the process of establishing a feeding site were dislodged from the apical meristem after 3 days with Ringer's solution. The SGs from 1000 larvae were dissected in Ringer's solution + 0.1% Triton X100, pooled as a single replicate in 500 µl of TRIzol Reagent (Invitrogen, Life Technologies, Waltham, MA, USA) and frozen at -80°C.

### 2.2. RNA extraction and sequencing

Total RNA was extracted using TRIzol® Reagent and cleaned using an Illustra™ RNAspin Mini Kit with DNase treatment (GE Healthcare Life Sciences, Chicago, IL, USA). Total RNA was quantified using a NanoDrop® One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and quality assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Total RNA (1 µg) was sent to the National Research Council of Canada (Saskatoon, SK, Canada) for library preparation and sequencing. The library was prepared with the Illumina TruSeq stranded mRNA kit (Illumina, San Diego, CA, USA). A single paired-end, 125 bp cDNA library was sequenced on an Illumina HiSeq 2500. Sequences were deposited in the National Center for Biotechnology Information (NCBI) short read archive (Accession SRR10186704).

### 2.3. Transcriptome assembly and annotation

Raw sequence read quality was assessed with FastQC v0.11.5 (Andrews 2010). rRNA contaminants, and low-quality reads and adaptor sequences, were removed with SortMeRNA v.2.1 (Kopylova et al., 2012) and Trimmomatic v.0.36 (Bolger et al., 2014), respectively (Figure S1). Clean reads were used to assemble a *de novo* transcriptome with Trinity v.2.6.6 (default parameters) (Haas et al., 2013). Summary statistics were generated by a script included in the Trinity program (TrinityStats.pl). Assembly quality and completeness were evaluated by mapping paired-reads back to the assembly using Bowtie2 v.2.3.4.1 (default parameters) (Langmead and Salzberg, 2012) and BUSCO (Benchmarking Universal Single-Copy Orthologs) v.2.0 (Diptera odb9 gene set) (Waterhouse et al., 2018), respectively. To reduce redundancy in the assembly, TransDecoder v.5.0.2 (Haas et al., 2013) (default parameters) with incorporated homology searches using HMMER v.3.1b2 against the Pfam-A database (v.31.0) (El-Gebali et al., 2019) and BLASTp (Altschul et al., 1990) against the UniProtKB SwissProt database (v. 28 February 2018) (UniProt Consortium, 2019), was used to extract open reading frames (ORF) at least 100 amino acids in length from the transcripts. The predicted proteins were condensed to non-redundant proteins using CD-HIT v.4.6 (Fu et al., 2012) at 100% amino acid identity.

BLASTp (e-value $< 10^{-25}$) was used to annotate protein sequences against the NCBI non-redundant (nr) Arthropoda (taxa:6656) subset. Sequences without a hit were searched against the complete NCBI nr database. All sequences were then assigned InterPro entries (with Pfam ID), gene ontology (GO) terms, enzyme classification (EC) codes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with Blast2GO v.5.0.13 (default parameters) (Gotz et al., 2008). BLAST and InterPro GO terms were merged in Blast2GO with the built-in algorithm. Expression analysis was carried out with RNA-Seq by expectation maximum (RSEM v.1.3.0) (Li and Dewey, 2011) using a built-in Trinity script (align_and_estimate_abundance.pl).

### 2.4. Identification of secreted proteins

To identify secreted proteins, sequences were annotated with SignalP v.4.1 (Nielsen, 2017) and TMHMM v.2.0 (Krogh et al., 2001) to predict signal peptides and transmembrane domains (TDs), respectively. Sequences with a signal peptide, 0-1 TDs, encoding a complete or 5'-partial ORF, and a transcripts per million (TPM) > 0.01 were examined. First, sequences with a TD downstream of the signal peptide were removed. Remaining sequences were annotated with Phobius, which combines signal peptide and TD prediction (Käll et al., 2004).

Non-annotated SSGPs were grouped according to Al-jbory et al., (2018). A database was created from non-annotated SSGPs in Geneious Prime v.2020.1 and subjected to a BLASTx search against the local database (e-value $< 10^{-3}$, all other default settings). Sequences with > 90% signal peptide identity and > 30 % overall identity in the mature protein were considered to be in the same group. To confirm groupings, pairwise or multiple sequence alignments were conducted with Clustal Omega v.1.2.2 (default settings, Sievers et al. 2011) in Geneious. Transcripts encoding non-annotated SSGPs were searched against the *C. nasturtii* genome (Mori et al., 2021) using BLASTp (default settings).

## 3. Results

### 3.1. Sequencing and transcriptome assembly

The first instar SG library yielded 37.3 million clean reads resulting in 99,215 transcripts and 67,734 Trinity 'genes' (Contig N50 = 1,214). 98.6% of the reads mapped back to the transcriptome indicating that most were used in the assembly. BUSCO analysis determined that 74.6 % of core genes were complete (42.6 % single copy, 32.0 % duplicated), 9.1 % fragmented, and 16.3 % missing (Figure 2A). Transdecoder predicted that the assembly represented 48,693 proteins, of which 40,665 were non-redundant (less than 100% amino acid identity), more than is typical for a dipteran [*Drosophila melanogaster* = 13,920 (Adams et al., 2000) *Musca domestica* = 23,884 (Scott et al., 2014)].

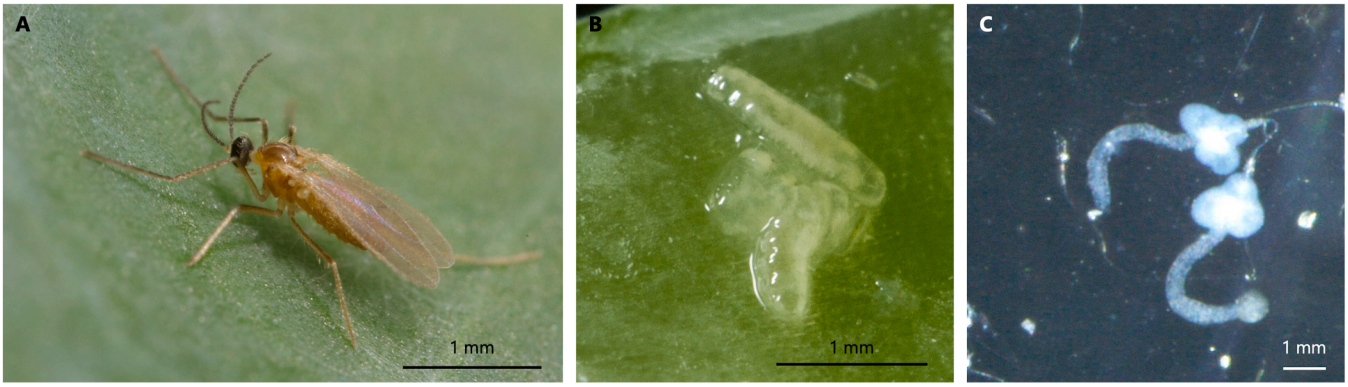


**Fig. 1.** Swede midge life stages and salivary glands. (A) Adult female, (B) First instar larvae, (C) Dissected first instar larval salivary glands.

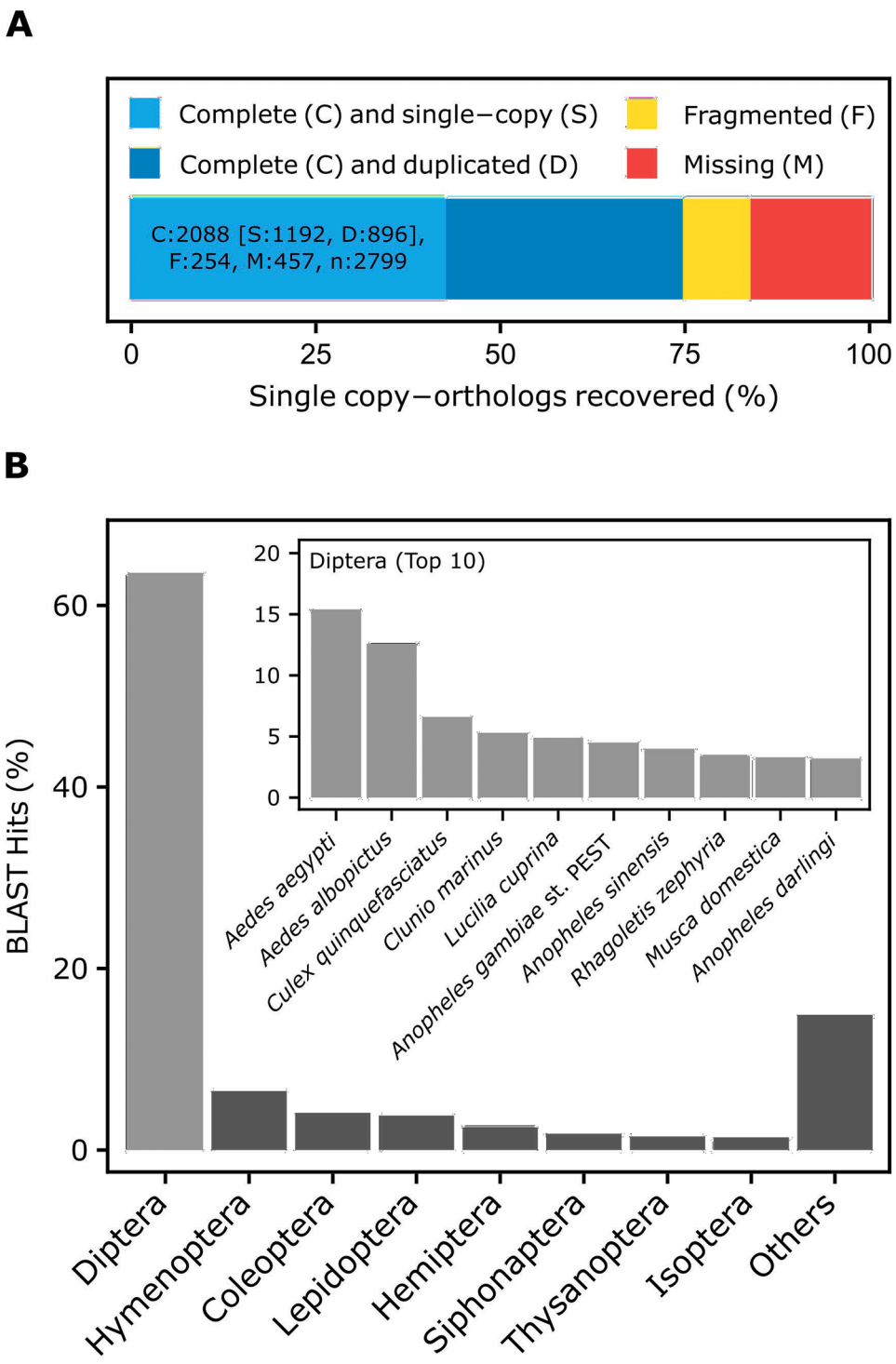


**Fig. 2.** Summary of Swede midge (*C. nasturtii*) first instar larval salivary gland transcriptome results. (A) Percentage of the Diptera odb9 Benchmarking Universal Single-Copy Orthologs (BUSCO) identified in the transcriptome classified as complete (C), single-copy (S), duplicated (D), fragmented (F) and missing (M). (B) Homology of insect orders and top 10 Diptera species (inset) distribution annotated with Blast2GO (NCBI non-redundant (n.r.) Arthropoda (taxa:6656) and full n.r. databases).

### 3.2. Functional annotation

BLASTp analysis annotated 33,198 of the predicted proteins, after removing those from non-arthropods (Table S1), resulting in 28,850 protein isoforms (21,383 with and 7,467 without BLAST hits), originating from 19,540 Trinity genes (Table S2). The predominant species from which annotations were obtained were *Aedes aegypti* (Diptera: Culicidae) (10%) and *A. albopictus* (Diptera: Culicidae) (8%) (Figure 2B, Table S3). GO annotations were available for 68.4% of the proteins with 14,346 assigned to a biological process, 17,210 to a molecular function and 10,992 to a cellular component (Figure 3A, Table S4). KEGG metabolic pathway analysis assigned 4,033 proteins to 145 pathways (Table S5). The majority mapped to purine and thiamine metabolism, followed by biosynthesis of antibiotics (Figure 3B, Table S5). Other notable pathways included carbohydrate metabolism (starch, sucrose and galactose), detoxification and xenobiotic metabolism by cytochrome P450. Pfam assigned 4,190 domains to 19,609 proteins resulting in 25,337 annotations (Table S6). The most abundant were associated with cellular signaling (protein kinase domain, protein tyrosine kinase, ras family), digestive enzymes (trypsin, lipase), transport (ABC transporter, sugar transport, major facilitator superfamily) and detoxification (cytochrome P450, carboxylesterase, flavin-binding monooxygenase-

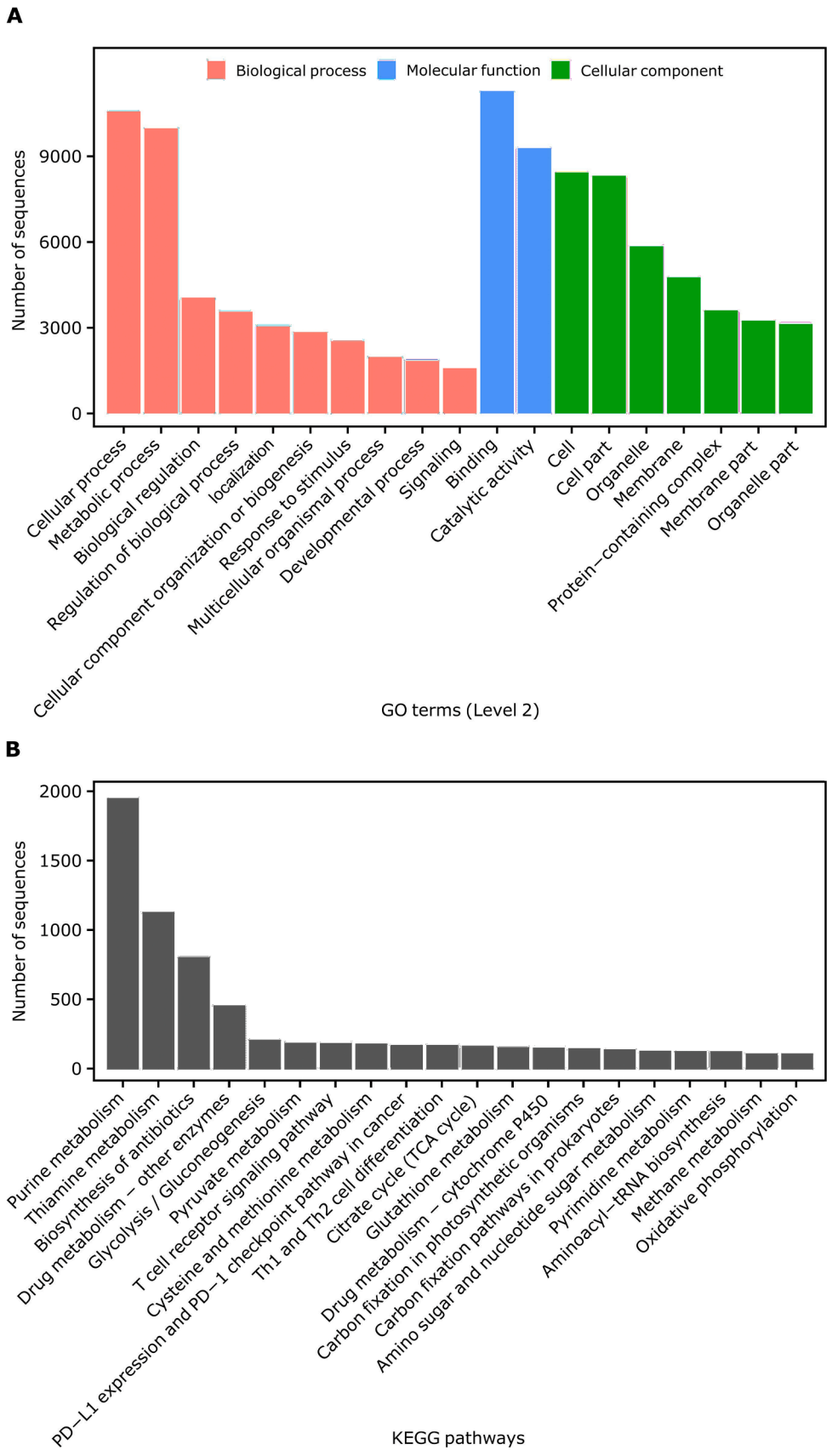


**Fig. 3.** Summary of Swede midge (*C. nasturtii*) first instar larval salivary gland transcriptome results. (A) Gene Ontology (GO) classification (level 2) of transcripts into three main categories (biological process, molecular function, and cellular component). (B) Distribution of sequences in the salivary gland transcriptome that mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Top 20 only).

like).

### 3.3. Secreted salivary gland proteins (SSGPs)

Of the 2,434 proteins with a signal peptide, 2,165 were predicted to be SSGPs (Table S7) and 1,147 could be annotated. Non-annotated SSGPs were classified into 788 groups (Table S8), with 157 containing multiple members that represent alleles from individual insects or paralogues (Figure S2). A search of the *C. nasturtii* genome revealed that 60 were members of expanded, unlinked gene families, while 145 were from gene clusters having 2-8 members (Table S9, S10). Differences between members of clusters were generally deletions centered on specific regions (Figure S3), while single amino acid differences or truncations were less common.

The annotated SSGPs were curated into categories based on events for larval establishment, including penetration, nutrient acquisition, and interaction with defenses (Table S11). Enzymes involved in synthesis of hormones associated with formation of feeding structures, namely indole acetic acid (IAA; auxin) and cytokinins, were identified (Table S11). It is noted that the implied functions of these proteins is based solely on the annotations and work done with similar proteins in other systems. Additional studies are required to verify the assertions below.

## 4. Discussion

### 4.1. Host penetration

The initial barrier to larvae and the first opportunity to access nutrients is the plant cuticle, a layer of C:16 and C:18 fatty acids that form cutin and surface waxes (Yeats and Rose, 2013). Lipase genes were highly expressed in the SG (Table S11) and may disrupt the cuticle facilitating entry. Phospholipase genes were expressed, these enzymes cleave fatty acids from membrane lipids and digest cellular debris.

The second physical barrier is the plant cell wall, which is composed of cellulose, hemicellulose, pectin, glycoproteins and phenolic esters. Cellulose, beta(1,4)-linked D-glucose, is the main carbohydrate polymer. Transcripts encoding cellulose-hydrolyzing endo-glucanases were not abundant, even though carbohydrate active hydrolase genes from 12 glycoside hydrolase families (CAZY, http://www.cazy.org/) were expressed (Table S11). However, enzymes affecting cell wall rigidity were expressed. Hemicellulose is a heteropolymer of glucose and other sugars (xylose, mannose, galactose, rhamnose, and arabinose), and interacts with cellulose and lignin. Genes encoding enzymes that release monosaccharides from hemicellulose (alpha-fucoside, beta-mannosidase, beta-galactosidase and beta-glucuronidase) were expressed. Pectin is a polymer of galacturonic acid that interacts with cellulose and hemicellulose, and pectin-lyase/esterase genes were expressed. This milieu of cell wall hydrolytic enzymes suggests that dissolution of host tissues is not vital for penetration and may be related to re-shaping host morphology to establish the feeding structure.

### 4.2. Nutrient acquisition

*C. nasturtii* larval mouthparts provide few clues about feeding; however, they likely inject saliva via specialized mandibles and then re-ingest fluids (Chen et al., 2011). Whether they penetrate phloem sieve elements is unknown, though evidence for extra-oral digestion was apparent in the SG transcriptome.

Starch, the major storage carbohydrate, is digested by amylases into trisaccharides (maltotriose) and disaccharides (maltose) and then to glucose. Transcripts encoding several alpha-glucosidases (maltases) were in the SG (Table S11). Amylase is produced in the SGs of hemipteran insects (Li et al., 2017) and lepidopterans (Da Lage, 2018). Sucrose is the main phloem carbohydrate and is hydrolyzed into glucose and fructose by sucrases. An alpha-glucosidase with sucrase activity hydrolyzes sucrose in aphids (Price et al., 2007). alpha-glucosidases with sucrase activity are within glycoside hydrolase family 31 and several were expressed at high levels in the SG.

Nucleic acids may be used directly or broken down into carbon, nitrogen and phosphate. Deoxyribonuclease and adenosine deaminase genes were expressed in the SG. Phosphate is a limiting nutrient (Perkins et al., 2004) and obtained from phospholipids, nucleotides, or sugar-phosphates. The SG expressed genes encoding nucleoside triphosphate hydrolases, nucleoside triphosphate dihydrolases, apyrase, pyrophosphatase, acid phosphatases, inositol phosphate phosphates and prostatic acid phosphatases (Table S11). Several were expressed at very high levels attesting to the importance of phosphate acquisition. In Hessian fly, a nucleotide diphosphate kinase effector supplies ATP to the larvae (Wang et al., 2018) linking phosphate to energy requirements.

Insects with acidic guts (coleopterans, hemipterans and hymenopterans) possess aspartic and cysteine proteases, while serine proteases are present in lepidopterans (Terra and Ferreira, 1994). The SG expressed many serine proteases (Table S11) and a few encoding cysteine protease genes, as did Hessian fly (Chen et al., 2013). Serine proteases and high pH are found in saliva and SGs of piercing insects that employ extra-oral digestion (Zhu et al., 2003; Lomate and Bonning, 2018). Like *C. nasturtii*, aphid saliva contains cathepsin B-type cysteine proteases (Rispe et al., 2008).

### 4.3. Interaction with host defenses

Injury to sieve elements induces callose and sieve element occlusion protein synthesis to restrict flow of phloem (Furch et al., 2007) leading to resistance to piercing/sucking insects (Peng and Walker, 2020). In planthoppers, salivary beta-1,3-glucanase removes callose from sieve pores (Hao et al., 2008). Even if *C. nasturtii* larvae do not pierce sieve tubes, callose would restrict phloem movement to feeding structures and genes encoding callose-hydrolyzing enzymes (glucan endo-1, 3-beta-glucosidase and beta-1,3-glucan-binding protein with hydrolase activity) were expressed (Table S11).

Callose and sieve element occlusion protein production depends on calcium signaling (Chen and Kim, 2009). Calcium causes occlusion protein complexes to enlarge and block sieve tubes (Knoblauch et al., 2012). Disruption of calcium channels prevents this (Peng and Walker, 2020) and calcium-binding proteins in aphid saliva lead to sieve tube re-opening (Will et al., 2007). Regucalcin binds calcium and activates calcium pumps and is found in aphid (Will et al., 2007) and planthopper (Hattori et al., 2015) saliva. Genes encoding regucalcin or other calcium-binding proteins (calnexin, calreticulin, calumenin, sarcalumenin and SPARC) were expressed in the *C. nasturtii* SG (Table S11). ARMET is critical for aphid feeding (Wang et al., 2015) and interferes with calcium release in sieve element cells (van Bel and Will, 2016). An ARMET gene was expressed at high levels in the SG implying that phloem is a nutrient source for *C. nasturtii* larvae. Aphid saliva contains zinc metalloproteases that degrade sieve tube proteins to prevent blockage (Furch et al., 2015) and metalloprotease genes were expressed in the *C. nasturtii* SG (Table S11).

ATP primes signaling cascades for defense compound production, such as nitrous oxide and reactive oxygen species (Giron et al., 2016). Insects secrete ATP-hydrolyzing enzymes in saliva to suppress this (Wu et al., 2012) and apyrase and ATPase genes were expressed in the SG (Table S11). Carbonic anhydrase genes were also expressed; in planthopper saliva this enzyme suppresses defenses by sequestering salicylic acid (Huang et al., 2018). Calcium-binding proteins and odorant-binding proteins suppress defense pathways (Tian et al., 2021; Liu et al., 2021) and many were expressed in the SG.

Flavin-dependent monooxygenases detoxify chemical defenses and several genes were expressed in the SG (Table S11), including senecionine N-oxygenases that modify this alkaloid (Langel and Ober, 2011). Glucosinolate detoxification is important for insects that specialize on Brassicaceae (Jeschke et al., 2016). Myrosinase, sulfatase

and glutathione-S-transferase implicated in glucosinolate detoxification (Mori et al., 2021; Chen et al., 2022) were expressed in the SG, as were those that detoxify other chemical defenses (esterases, peroxidases, superoxide dismutase).

### 4.4. Plant hormones synthesis

Changes associated with feeding structure formation are induced by plant growth hormones (IAA and cytokinins) (Tooker and Helms, 2014). Enzymes in aphid (Duspiva 1954), plant-sucking hemipteran (Miles and Lloyd, 1967) and sawfly (Yamaguchi et al., 2012) saliva convert tryptophan to IAA via two intermediates, indole-3-acetaldoxime (IAOx) and indole-3-acetaldehyde (Iaa1d) (Suzuki et al. 2014). In *Bombyx mori*, indole-3-acetaldehyde oxidase converts Iaa1d to IAA (Takei et al., 2019). A search of the *C. nasturtii* genome with the *B. mori* Iaa1d oxidase identified five genes. XM_031768326 was expressed in the SG (Table S11) and was similar to the *B. mori* (56%) and *T. castaneum* (59.4%) indole-3- acetaldehyde oxidase enzymes (Figure S4). In sawfly, a flavin-dependent monooxygenase (PonFMO1) converts tryptophan to IAOx, while aromatic aldehyde synthase (PonAAS2) converts tryptophan to Iaa1d (Yamaguchi et al., 2012). A search with PonAAS2 identified two genes encoding similar enzymes (XP_031627721 and XP_031628133) annotated as 3,4-dihydroxyphenylacetaldhehyde synthase-like enzymes. XP_031627721 was expressed in the SG (Table S11) and the enzyme (XM_031771861) had conserved residues (Tyr80, Ser147 and Asn192) (Figure S5) present in dihydroxyphenylacetaldehyde synthases that differentiate it from dihydroxyphenylacetaldehyde decarboxylase (XM_031782752) (Liang et al., 2017). A search using PonFMO1 identified 57 genes encoding flavin-dependent monooxygenases. This suggests that *C. nasturtii* larval SG can produce IAA *de novo*.

In insects, trans-zeatin and isopentenyladenine are the most common cytokinins and are derived from tRNA (Andreas et al., 2020). Initially, tRNA dimethylallyltransferase transfers an isoprenyl group to adenine to form cytokine ribose phosphate. Two tRNA dimethylallyltransferases genes (XM_031779266 and XM_031779268) were expressed in the SG. In plants, cytokinin is released via a cytokinin-specific phosphoribohydrolase (Kurakawa et al., 2007). Transcripts encoding cytokinin riboside 5'monophosphate phosphoribohydrolases were found in the SG (Table S1); however, these were from the *B. napus* (*LOG8*) gene or similar to an *Acinetobacter junii* sequence in the silk orb spider (PRD28654) transcriptome. This indicates that insects employ a different pathway for cytokinin synthesis than plants or bacteria.

### 4.5. Effectors

Secreted effectors that manipulate host morphology to form feeding structures or evade host defenses are addressed above. Half of all gall midge (*M. destructor* and *Sitodiplosis mosellana*) (Chen et al., 2010; Al-jbory et al., 2018) and 10% of *C. nasturtii* SG transcripts encode secreted proteins. 47% of *C. nasturtii* SSGPs could not be fully annotated with most not having orthologues (Table S8). Many were members of gene clusters with variation in length of specific regions or carboxyl termini (Figure S3); non-synonymous mutations were less common. Hessian fly SG effectors are under selective pressure as evidenced by the massive expansion of gene families (Zhao et al., 2015). Cereal midge SSGP gene clusters have conserved regulatory and signal peptide regions, but extreme diversity elsewhere (Chen et al., 2010; Al-jbory et al., 2018) and may be related to deployment of resistant varieties (Schmid et al., 2018). This pattern was not noted in *C. nasturtii*.

## 5. Conclusion

The *C. nasturtii* SSGPs provided insight into its biology and interaction with its host plant. Similarities/differences to other gall forming cecidomyids with respect to the amount of transcription devoted to SSGPs and conservation of gene clusters was noted. Similar to cereal gall midges, secreted insect avirulence and host R proteins may mediate compatible and compatible interactions and could be an avenue for resistance in Brassicas afflicted by *C. nasturtii*.

## Data availability

The *C. nasturtii* first instar larval SG sequencing data underlying this article are available in the NCBI Short Read Archive at https://www.ncbi.nlm.nih.gov/ and can be accessed with SRX6907226. All other data/files are included in the Supplement.

Supplementary Fig. S1. Swede midge (*C. nasturtii*) first instar larval salivary gland transcriptome assembly, annotation, and secreted salivary gland protein prediction pipeline.

Supplementary Fig. S2. *C. nasturtii* salivary gland transcripts with non-annotated SSGPs with more than member.

Supplementary Fig. S3. Alignment of proteins encoded by the *C. nasturtii* gene clusters.

Supplementary Fig. S4. Alignment of indole-3-acetaldehyde oxidases from *C. nasturtii* (putative XP_031636493.1), *T. castaneum* (XP_015834263.1) and *B. mori* (XP_004925960.2).

Supplementary Fig. S5. Alignment of aromatic putative aldehyde synthases from *C. nasturtii* (putative XP_031627721.1, XP_015834263.1 and XP_031638612.1), *Pontania* species (BCT26320.1) and *D. melanogaster* (NP_724162.1 and NP_724164.1).

Supplementary Table S1. Unfiltered *C. nasturtii* salivary gland transcripts.

Supplementary Table S2. *C. nasturtii* salivary gland transcripts.

Supplementary Table S3. Top BLAST hits for *C. nasturtii* salivary gland transcripts.

Supplementary Table S4. Gene ontologies for *C. nasturtii* salivary gland transcripts.

Supplementary Table S5. KEGG Reports for *C. nasturtii* salivary gland transcripts.

Supplementary Table S6. Pfam domains associated with proteins encoded by *C. nasturtii* salivary gland transcripts.

Supplementary Table S7. *C. nasturtii* secreted salivary gland proteins.

Supplementary Table S8. Non-annotated *C. nasturtii* secreted salivary gland proteins.

Supplementary Table S9. BLAST reports for transcripts encoding non-annotated *C. nasturtii* secreted salivary gland proteins against the C. nasturtii genome sequence.

Supplementary Table S10. BLAST reports showing clusters of genes encoding related non-annotated *C. nasturtii* secreted salivary gland proteins.

Supplementary Table S11. *C. nasturtii* secreted salivary gland proteins involved in host-plant interactions.

## CRediT authorship contribution statement

**Boyd A. Mori:** Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **Cathy Coutu:** Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **Martin A. Erlandson:** Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing. **Dwayne D. Hegedus:** Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing.

## Declaration of Competing Interest

The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

## Data availability

Data will be made available on request.

*Funding*

This work was funded by the Agriculture & Agri-Food Canada Genomics Research and Development Initiative. Financial support was provided to BAM by a NSERC Industrial Research Chair (545088) and partner organizations Alberta Wheat Commission, Alberta Barley Commission, Alberta Canola Producers Commission, and Alberta Pulse Growers Commission, and a NSERC Discovery grant (RGPIN-2021-02479) during the preparation of the manuscript.

*Acknowledgments*

We thank Stephanie Harris for collecting SGs, Shane Hladun and Jennifer Holowachuk for insect rearing, Doug Baldwin for RNA extraction, and Jonathon Williams, Shelley Barkley, and Shane Hladun for the photographs.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cris.2023.100064.

## References

Adams, M.D., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195. https://doi.org/10.1126/science.287.5461.21.

Al-jbory, Z., Anderson, K.M., Harris, M.O., Mittapalli, O., Whitworth, R.F., Chen, M.S., 2018. Transcriptomic analyses of secreted proteins from the salivary glands of wheat midge larvae. J. Insect Sci. 18, e17. doi.org/10.1093/jisesa/iey009.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Andreas, P., Kisiala, A., Emery, R.J.N., De Clerck-Floate, R., Tooker, J.F., Price, P.W., Miller III, D.G., Chen, M.S., Connor, E.F., 2020. Cytokinins are abundant and widespread among insect species. Plants 9, e208. https://doi.org/10.3390/plants9020208.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Version 0.11.8 ).

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Chen, X.Y., Kim, J.Y., 2009. Callose synthesis in higher plants. Plant Sig. Behav. 4, 489–492. https://doi.org/10.4161/psb.4.6.8359.

Chen, M.S., Zhao, H.Z., Zhu, Y.C., Scheffler, B., Liu, X., Liu, X., Hulbert, S., Stuart, J.J., 2008. Analysis of transcripts and proteins expressed in the salivary glands of Hessian fly (*Mayetiola destructor*) larvae. J. Insect Physiol. 54, 1–16.

Chen, M.S., Liu, X., Yang, Z., Zhao, H., Shukle, R.H., Stuart, J.J., Hulbert, S., 2010. Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge. BMC Evol. Biol. 10, e296. https://doi.org/10.1186/1471-2148-10-296.

Chen, M., Shelton, A.M., Hallett, R.H., Hoepting, C.A., Kikkert, J.R., Wang, P., 2011. Swede midge (Diptera: Cecidomyiidae), ten years of invasion of crucifer crops in North America. J. Econ. Entomol. 104, 709–716.

Chen, H., Zhu, Y.C., Whitworth, R.J., Reese, J.C., Chen, M.S., 2013. Serine and cysteine protease-like genes in the genome of a gall midge and their interactions with host plant genotypes. Insect Biochem. Mol. Biol. 43, 701–711. https://doi.org/10.1016/j.ibmb.2013.05.006.

Chen, W., Saqib, H.S.A., Xu, X., Dong, Y., Zheng, L., Lai, Y., Jing, X., Lu, Z., Sun, L., You, M., He, W., 2022. Glucosinolate sulfatases-sulfatase-modifying factors system enables a crucifer-specialized moth to pre-detoxify defensive glucosinolate of the host plant. J. Agric. Food Chem. 70, 11179–11191. https://doi.org/10.1021/acs.jafc.2c03929.

Da Lage, J.L., 2018. The amylases of insects. Int. J. Insect Sci. 10, e1179543318804783 https://doi.org/10.1177/1179543318804783(2018).

Duspiva, F., 1954. Weitere Untersuchungen i.iber Stoffwechsel-physiologische Beziehungen zwischen Rhynchoten und ihren. Wirtspflanzen. Mitt. Biol. Bundesanstalt 80, 155–162. Berlin Dahlem.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucl. Acids Res. 47, D427–D432. https://doi.org/10.1093/nar/gky995.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Furch, A.C.U., Hafke, J.B., Schulz, A., van Bel, A.J.E., 2007. Ca2+-mediated remote control of reversible sieve tube occlusion in *Vicia faba*. J. Exp. Bot. 58, 2827–2838. https://doi.org/10.1093/jxb/erm143.

Furch, A.C.U., van Bel, A.J.E., Will, T., 2015. Aphid salivary proteases are capable of degrading sieve-tube proteins. J. Exp. Bot. 66, 533–539. https://doi.org/10.1093/jxb/eru487.

Giron, D., Huguet, E., Stone, G.N., Body, M., 2016. Insect-induced effects on plants and possible effectors used by galling and leaf-mining insects to manipulate their host-plant. J. Insect Physiol. 84, 70–89.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucl. Acids Res. 36, 3420–3435. https://doi.org/10.1093/nar/gkn176.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hallett, R.H., Reddy, G.V., 2017. The challenge of swede midge management in canola. Integrated Management of Insect Pests on Canola and Other Brassica Oilseed Crops. Centre for Agriculture and Biosciences International, Wallingford, United Kingdom, pp. 44–67.

Hallett, R.H., Heal, J.D., 2001. First nearctic record of the swede midge (Diptera: Cecidomyiidae), a pest of cruciferous crops from Europe. Can. Ent. 133, 713–715.

Hao, P., Liu, C., Wang, Y., Chen, R., Tang, M., Du, B., Zhu, L., He, G., 2008. Herbivore-induced callose deposition on the sieve plates of rice: an important mechanism for host resistance. Plant Physiol. 146, 1810–1820. https://doi.org/10.1104/pp.107.111484.

Hattori, M., Komatsu, S., Noda, H., Matsumoto, Y., 2015. Proteome analysis of watery saliva secreted by green rice leafhopper, *Nephotettix cincticeps*. PLOS One 10, e0123671. https://doi.org/10.1371/journal.pone.0123671.

Huang, H.J., Lu, J.B., Li, Q., Bao, Y.Y., Zhang, C.X., 2018. Combined transcriptomic/proteomic analysis of salivary gland and secreted saliva in three planthopper species. J. Proteomics 172, 25–35. https://doi.org/10.1016/j.jprot.2017.11.003.

Jeschke, V., Gershenzon, J., Vassao, D.G., 2016. Insect detoxification of glucosinolates and their hydrolysis products. Adv. Bot. Res. 80, 199–245.

Johnson, A.J., Shukle, R.H., Chen, M.S., Srivastava, S., Subramanyam, S., Schemerhorn, B.J., Weintraub, P.G., Abdel Moniem, H.E., Flanders, K.L., Buntin, G.D., Williams, C.E., 2014. Differential expression of candidate salivary effector proteins in field collections of Hessian fly, *Mayetiola destructor*. Insect Mol. Biol. 24, 191–202. https://doi.org/10.1111/imb.12148.

Kaiser, W., Huguet, E., Casas, J., Commin, C., Giron, D., 2010. Plant green island phenotype induced by leaf-miners is mediated by bacterial symbionts. Proc. Royal Soc. B Biol. Sci. B 277, 2311–2319.

Käll, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

Knoblauch, M., Stubenrauch, M., Van Bel, A.J.E., Peters, W.S., 2012. Forisome performance in artificial sieve tubes. Plant Cell Environ. 35, 1419–1427. https://doi.org/10.1111/j.1365-3040.2012.02499.x.

Kopylova, E., Noe, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. https://doi.org/10.1093/bioinformatics/bts611.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., Sakakibara, H., Kyozuka, J., 2007. Direct control of shoot meristem activity by a cytokinin-activating enzyme. Nature 445, 652–655. https://doi.org/10.1038/nature05504.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Langel, D., Ober, D., 2011. Evolutionary recruitment of a flavin-dependent monooxygenase for stabilization of sequestered pyrrolizidine alkaloids in arctiids. Phytochemistry 72, 1576–1584. https://doi.org/10.1016/j.phytochem.2010.12.014.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, e323. https://doi.org/10.1093/bioinformatics/btl158.

Li, W., Zhao, X., Yuan, W., Wu, K., 2017. Activities of digestive enzymes in the omnivorous pest Apolygus lucorum (Hemiptera: Miridae). J. Econ. Entomol. 110, 101–110.

Liang, J., Han, Q., Ding, H., Li, J., 2017. Biochemical identification of residues that discriminate between 3,4-dihydroxyphenylalanine decarboxylase and 3,4-dihydroxyphenylacetaldehyde synthase-mediated reactions. Insect Biochem. Mol. Biol. 91, 34–43. https://doi.org/10.1016/j.ibmb.2017.10.001.

Liu, H., Wang, C., Qiu, C.L., Shi, J.H., Sun, Z., Hu, X.J., Liu, L., Wang, M.Q., 2021. A salivary odorant-binding protein mediates *Nilaparvata lugens* feeding and host plant phytohormone suppression. Int. J. Mol. Sci. 22, e4988. https://doi.org/10.3390/ijms22094988.

Lomate, P.R., Bonning, B.C., 2018. Proteases and nucleases involved in the biphasic digestion process of the brown marmorated stink bug, *Halymorpha halys* (Hemiptera; Pentatomidae). Arch. Insect Biochem. Physiol. 98, e21459.

Miles, P.W., Lloyd, J., 1967. Synthesis of a plant hormone by the salivary apparatus of plant-sucking Hemiptera. Nature 213, 801–802.

Mori, B.A., Coutu, C., Chen, Y.H., Campbell, E.O., Dupuis, J.R., Erlandson, M.A., Hegedus, D.D., 2021. *De novo* whole-genome assembly of the Swede midge (*Contarinia nasturtii*), a specialist of Brassicaceae, using linked-read sequencing. Genome Biol. Evol. 13, evab036. https://doi.org/10.1093/gbe/evab036.

Nielsen, H., 2017. Predicting secretory proteins with SignalP. Methods Mol. Biol. 1611, 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6.

Peng, H.C., Walker, G.P., 2020. Sieve element occlusion provides resistance against *Aphis gossypii* in TGR-1551 melons. Insect Sci. 27, 33–48. https://doi.org/10.1111/1744-7917.12610.

Perkins, M.C., Woods, H.A., Harrison, J.F., Elser, J.J., 2004. Dietary phosphorus affects the growth of larval *Manduca sexta*. Arch. Insect Biochem. Physiol. 55, 153–168. https://doi.org/10.1002/arch.10133.

Prather, S., Schneider, T., Gaham Godoy, J., Odubiyi, S., Bosque-Perez, N.A., Rashed, A., Rynearson, S., Pumphrey, M.O., 2022. Reliable DNA markers for a previously unidentified, yet broadly deployed Hessian fly resistance gene on chromosome 6B in Pacific northwest spring wheat varieties. Front. Plant Sci. 13, e779096 https://doi.org/10.3389/fpls.2022.779096.

Price, D.R., Karley, A.J., Ashford, D.A., Isaacs, H.V., Pownall, M.E., Wilkinson, H.S., Gatehouse, J.A., Douglas, A.E., 2007. Molecular characterisation of a candidate gut sucrase in the pea aphid, *Acyrthosiphon pisum*. Insect Biochem. Mol. Biol. 37, 307–317. https://doi.org/10.1016/j.ibmb.

Rispe, C., Kutsukake, M., Doublet, V., Hudaverdian, S., Legeai, F., Simon, J.C., Tagu, D., Fukatsu, T., 2008. Large gene family expansion and variable selective pressures for cathepsin B in aphids. Mol. Biol. Evol. 25, 5–17. https://doi.org/10.1093/molbev/msm222.

Schmid, R.B., Knutson, A., Giles, K.L., McCornack, B.P., 2018. Hessian fly (Diptera: Cecidomyiidae) biology and management in wheat. J. Integr. Pest Man. 9, e14. https://doi.org/10.1093/jipm/pmy008.

Scott, J.G., Warren, W.C., Beukeboom, L.W., et al., 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. Genome Biol. 15, 466. https://doi.org/10.1186/s13059-014-0466-3.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.

Stone, G.N., Schönrogge, K., 2003. The adaptive significance of insect gall morphology. Trends Ecol. Evol. 18, 512–522.

Stratton, C.A., Hodgdon, E.A., Zuckerman, S.G., Shelton, A.M., Chen, Y.H., 2018. A single swede midge (Diptera: Cecidomyiidae) larva can render cauliflower unmarketable. J. Insect Sci. 18, e24. https://doi.org/10.1093/jisesa/iey062.

Stringham, G.R., 1971. Genetics of four hypocotyl mutants in *Brassica campestris* L. J. Heredity 62, 248–250.

Stuart, J.J., Chen, M.S., Shulke, R., Harris, M.O., 2012. Gall midges (Hessian flies) as plant pathogens. Annu. Rev. Phytopathol. 50, 339–357.

Suzuki, H., Yokokura, J., Ito, T., Arai, R., Yokoyama, C., Toshima, H., Nagata, S., Asami, T., Suzuki, Y., 2014. Biosynthetic pathway of the phytohormone auxin in insects and screening of its inhibitors. Insect Biochem. Mol. Biol. 53, 66–72. https://doi.org/10.1016/j.ibmb.2014.07.008.

Takeda, S., Hirano, T., Ohshima, I., Sato, M.H., 2021. Recent progress regarding the molecular aspects of insect gall formation. Int. J. Mol. Sci. 22, e9424. https://doi.org/10.3390/ijms22179424.

Takei, M., Kogure, S., Yokoyama, C., Kouzuma, Y., Suzuki, Y., 2019. Identification of an aldehyde oxidase involved in indole-3-acetic acid synthesis in *Bombyx mori* silk gland. Biosci. Biotechnol. Biochem. 83, 129–136. https://doi.org/10.1080/09168451.2018.1525275.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: Properties, compartmentalization and function. Comp. Biochem. Physiol B 109, 1–62.

Tian, T., Ji, R., Fu, J., Li, J., Wang, L., Zhang, H., Yang, S., Ye, W., Fang, J., Zhu-Salzman, K., 2021. A salivary calcium-binding protein from *Laodelphax striatellus* acts as an effector that suppresses defense in rice. Pest Manag. Sci. 77, 2272–2281. https://doi.org/10.1002/ps.6252.

Tooker, J.F., Helms, A.M., 2014. Phytohormone dynamics associated with gall insects, and their potential role in the evolution of the gall-inducing habit. J. Chem. Ecol. 40, 742–753. https://doi.org/10.1007/s10886-014-0457-6.

UniProt Consortium, T.U., 2019. UniProt: a worldwide hub of protein knowledge. Nucl. Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049.

van Bel, A.J.E., Will, T., 2016. Functional evaluation of proteins in watery and gel saliva of aphids. Front. Plant Sci. 7, e1840. https://doi.org/10.3389/fpls.2016.01840.

Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., Sheng, C., Peng, G., Chen, S., Tomich, J.M., Reese, J., Edwards, O., Kang, L., Reeck, G., Cui, F., 2015. Armet is an effector protein mediating aphid–plant interactions. FASEB J. 29, 2032–2045. https://doi.org/10.1096/fj.14-266023.

Wang, Z., Ge, J.Q., Chen, H., Cheng, X., Yang, Y., Li, J., Whitworth, R.J., Chen, M.S., 2018. An insect nucleoside diphosphate kinase (NDK) functions as an effector protein in wheat - Hessian fly interactions. Insect Biochem. Mol. Biol. 100, 30–38. https://doi.org/10.1016/j.ibmb.2018.06.003.

Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35, 543–548. https://doi.org/10.1093/molbev/msx319.

Will, T., Tjallingii, W.F., Thönnessen, A., van Bel, A.J.E., 2007. Molecular sabotage of plant defense by aphid saliva. Proc. Natl. Acad. Sci. U. S. A. 104, 10536–10541. https://doi.org/10.1073/pnas.0703535104.

Wu, S., Peiffer, M., Luthe, D.S., Felton, G.W., 2012. ATP hydrolyzing salivary enzymes of caterpillars suppress plant defenses. PLOS One 7, e41947. https://doi.org/10.1371/journal.pone.0041947.

Yamaguchi, H., Tanaka, H., Hasegawa, M., Tokuda, M., Asami, T., Suzuki, Y., 2012. Phytohormones and willow gall induction by a gall-inducing sawfly. New Phytol. 196, 586–595.

Yeats, T.H., Rose, J.K., 2013. The formation and function of plant cuticles. Plant Physiol. 163, 5–20. https://doi.org/10.1104/pp.113.222737.

Zhao, C., Escalante, L.N., Chen, H., Benatti, T.R., Qu, J., et al., 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. Curr. Biol. 25, 613–620. https://doi.org/10.1016/j.cub.2014.12.057.

Zhu, Y.C., Zeng, F., Oppert, B., 2003. Molecular cloning of trypsin-like cDNAs and comparison of proteinase activities in the salivary glands and gut of the tarnished plant bug Lygus lineolaris (Heteroptera: Miridae). Insect Biochem. Mol. Biol. 33, 889–899. https://doi.org/10.1016/S0965-1748(03)00094-8.