



## OPEN ACCESS

## EDITED BY

Eustachio Tarasco,  
University of Bari Aldo Moro, Italy

## REVIEWED BY

Shubha Subramanyam,  
Agricultural Research Service (USDA),  
United States  
Shanchun Yan,  
Northeast Forestry University, China

## \*CORRESPONDENCE

Arun Jagannath  
✉ jagannatharun@yahoo.co.in

## †PRESENT ADDRESS

Rubina Chongtham,  
Department of Botany, Deshbandhu College,  
University of Delhi, Delhi, India

RECEIVED 10 April 2024

ACCEPTED 21 October 2024

PUBLISHED 02 December 2024

## CITATION

Sharma M, Oraon PK, Srivastava R,  
Chongtham R, Goel S, Agarwal M and  
Jagannath A (2024) Comparative  
transcriptomics of a generalist aphid, *Myzus  
persicae* and a specialist aphid, *Lipaphis  
erysimi* reveals molecular signatures  
associated with diversity of their feeding  
behaviour and other attributes.  
*Front. Plant Sci.* 15:1415628.  
doi: 10.3389/fpls.2024.1415628

## COPYRIGHT

© 2024 Sharma, Oraon, Srivastava,  
Chongtham, Goel, Agarwal and Jagannath. This  
is an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Comparative transcriptomics of a generalist aphid, *Myzus persicae* and a specialist aphid, *Lipaphis erysimi* reveals molecular signatures associated with diversity of their feeding behaviour and other attributes

Manvi Sharma, Praveen Kumar Oraon, Rakesh Srivastava, Rubina Chongtham<sup>†</sup>, Shailendra Goel, Manu Agarwal and Arun Jagannath\*

Department of Botany, University of Delhi, Delhi, India

**Introduction:** Aphids are phloem sap-sucking insects and are a serious destructive pest of several crop plants. Aphids are categorized as “generalists” or “specialists” depending on their host range. *Myzus persicae* (Sulz.) is a generalist aphid with a broad host range while *Lipaphis erysimi* (Kalt.), a specialist aphid, has a narrow host range. Aphid infestation involves several sequential stages including host recognition and selection, overcoming primary plant defence barriers, feeding on phloem sap and detoxification of host defence responses. Information on the molecular basis of variations between generalist and specialist aphids with reference to the above processes is limited.

**Methods:** In the current study, we generated transcriptome data of *M. persicae* and *L. erysimi* from adult and nymph stages and analysed the differential expression of genes between adults of the generalist and specialist aphid and similarly, between nymphs of the two aphid species. We categorized these differentially expressed genes into nine different categories namely, chemosensation-related, plant cell wall degrading enzymes, detoxification-related, digestive enzymes, peptidases, carbohydrate-, lipid-, amino acid-metabolism and reproduction. We also identified putative effector molecules in both *M. persicae* and *L. erysimi* from the transcriptome data,

**Results and discussion:** Gene expression analysis identified 7688 and 8194 differentially expressed unigenes at adult and nymph stages, respectively of *M. persicae* and *L. erysimi*. *M. persicae* showed significantly higher levels of expression in a greater number of unigenes (5112 in adults and 5880 in nymphs) in contrast to the specialist, *L. erysimi* (2576 in adults and 2314 in nymphs) in both developmental stages. In addition, *M. persicae* displayed a greater number (350 in adults and 331 in nymphs) of upregulated unigenes

involved in important processes such as host recognition, plant cell wall degradation, detoxification, digestion and metabolism, which correlate with its dynamic and polyphagous nature in contrast to the specialist (337 in adults and 251 in nymphs). We also observed a greater number of putative effectors in *M. persicae* (948 in adults and 283 in nymphs) than *L. erysimi* (797 in adults and 245 in nymphs). Based on our analysis, we conclude that the generalist aphid, *M. persicae* has a more diversified and stronger arsenal of genes that influence its polyphagous feeding behaviour and effective response to plant defence mechanisms against insect-herbivory. Our study provides a compendium of such candidate genes that would be most useful in studies on aphid biology, evolution and control.

#### KEYWORDS

aphids, generalist, specialist, transcriptome, differential expression, effectors

## Introduction

Aphids are small, soft-bodied, phloem sap-sucking insects belonging to the Order Hemiptera. They consist of approximately 4000 species which collectively infect ~25% of plant species worldwide (Dixon, 1985). Of these, around 100 aphid species infect a wide range of crop plants including oilseeds, cereals, pulses, vegetables and fruits leading to significant losses in crop productivity and yield (Blackman and Eastop, 2007). They also function as vectors for transmission of plant viruses with ~50% of insect-borne plant viruses being transmitted by aphids (Katis et al., 2007). Aphids can multiply by sexual reproduction but more importantly, they also have the ability to reproduce parthenogenetically which results in rapid colonisation, multiplication and spread rendering them highly destructive to standing crops. Parthenogenetic females give rise to young nymphs which develop into adults through four instar stages. Aphids are classified as generalists or specialists based on their host range. Specialists exhibit a narrow and selective host range while generalists are polyphagous and capable of infesting a wide diversity of plant species allowing them access to a wider range of resources (Capinera, 2008). Around 99% of all aphid species have a high degree of host specificity and the remaining 1% are highly polyphagous (Egas et al., 2005).

Infestation by aphids begins with detection of a suitable host using their chemosensory system, which includes various protein/gene families such as odorant binding proteins, odorant binding receptors, chemosensory proteins and gustatory receptors (Zhang et al., 2020). Following identification of a potentially favourable host, aphids initiate the feeding process by inserting a highly specialised stylet into the host tissue through which, they release a watery saliva. Aphids overcome the first line of host defence i.e., the plant cell walls by disrupting cell wall polymers using plant cell-wall

degrading enzymes (PCWDEs) present in the saliva (Guangxi et al., 2006; Thorpe et al., 2016; Silva-Sanzana et al., 2020). A gelling saliva is also secreted around the stylet to facilitate penetration and to minimise physiological contact with host cells (Tjallingii, 2006). The penetrating stylet also punctures adjacent cells during its progression towards the phloem to analyse the host cell sap for its pH, sugar concentration and amino acid contents to determine suitability of the host plant for feeding. The watery saliva contains salivary effectors that are proteins or small molecules which modify host cell physiology and can induce and/or suppress defence responses in the host plant (Hogenhout et al., 2009; Mondal, 2017, 2020). Several studies have identified effectors in various aphids viz., *Myzus persicae* (Harmel et al., 2008), *Acyrtosiphon pisum* (Carolan et al., 2011), *Schizaphis graminum* (Cooper et al., 2011), *Sitovian avenae* (Rao et al., 2013), *Diuraphis noxia* (Nicolis et al., 2022), *Pseudoregma bambucicola* (Zhang et al., 2023) and some of these have been functionally characterised. The phloem sap contains secondary metabolites and other proteins viz., redox regulators, phytohormones-related proteins, protease inhibitors, lectins, etc. that are involved in plant defence (Kehr, 2006). Aphids combat the effects of such proteins using detoxification enzymes, digestive enzymes, and peroxidases (Ramsey et al., 2010). Aphids also show physiological adaptations to feed continuously on a sugar-rich and nitrogen-deficient phloem sap that contains only a few free amino acids, mainly, asparagine, serine, glutamate, aspartate and glutamine as the nitrogen source (Douglas, 2006). They produce proteolytic enzymes to allow digestion of ingested proteins (Pyati et al., 2011).

Several studies based on transcriptome profiling of aphids have attempted to analyse the molecular basis of aphid infestation, growth and development. Transcriptomic studies of bamboo aphid by Zhang et al. (2023) identified putative effectors important for plant-aphid interactions such as those involved in detoxification, digestion and

antioxidant enzymes. Transcriptome profiles of alimentary canal of *Sitobion avenae* at pre- and post-feeding stages identified five novel candidate genes that led to higher aphid mortality and delayed development when used as targets in plant-mediated RNAi for aphid control in wheat (Zhang et al., 2013). To study the differences in two generalists feeding on the same host plant, transcriptome studies were conducted by Koch et al. (2019) between *S. graminum* and *Sipha flava* in which, one of the generalists, *S. graminum* upregulated more stress-responsive genes as compared to the other when fed on the same variety of switchgrass. In another study by Wang et al. (2014), comparison of the transcriptome of *S. avenae* (a specialist) and mRNA sequences of *A. pisum* predicted positive selection of 186 pairs of orthologous groups involved in xenobiotics and secondary metabolism which might be responsible for their divergence. In an integrative study combining transcriptomics and proteomics approaches on three different aphid species, *M. persicae*, *M. cerasi* and *Rhopalosiphum padi*, core effectors comprising of previously identified candidate effectors and species-specific effectors were identified (Thorpe et al., 2016).

To the best of our knowledge, comparative transcriptome profiling between generalist and specialist aphids feeding on the same host plant has not been reported till date. Additionally, information on differential expression profiles during infestation in feeding adults and nymphs of generalist and specialist aphids is limited. Comparison between the repertoire of effectors and their expression profiles between a generalist and a specialist aphid has also not been reported. In the current study, we generated and analysed transcriptome data at both adult and nymph stages of development for a generalist, *M. persicae* and a specialist, *Lipaphis erysimi* both of which are devastating pests of the important oilseed crop, *Brassica juncea* (Indian mustard). *L. erysimi* (mustard aphid) specifically feeds on *Brassica* sp. while *M. persicae* infests host plants of over 50 families encompassing more than 400 species (Weber, 1985; Blackman and Eastop, 2007). These insects also demonstrate significant differences in their infestation behaviour on the same susceptible host (*B. juncea*) during the crop growing season (October – March). In contrast to infestation by the specialist, infestation by the generalist is more regular in successive growing seasons, more widespread (more number of plants are infested with *M. persicae*) and higher numbers of these aphids are found on infested plants. The specialist (*L. erysimi*) was more susceptible to temperature variations and the duration of infestation was lesser compared to the generalist which was able to tolerate fluctuating weather conditions and survive on the host for a longer duration. To understand the basis of these differences in insect behaviour and to decipher the molecular basis of variations in their host range, feeding behaviour, detoxification mechanisms and evolutionary divergence in important genes, we analysed the differential expression of genes between the two aphid species in both adults and nymphs. We also identified putative effector proteins using a bioinformatic pipeline to study differences between the two types of aphids, their role in countering plant defence compounds and the degree of divergence between them. Our study has identified several species-specific novel effectors and generated resources for both the aphids at two different developmental stages which provide interesting leads on aphid biology, divergence and evolution.

## Materials and methods

### Aphid rearing and collection

Apterous adults of *L. erysimi* and *M. persicae* were collected from *Brassica juncea* var. Varuna plants during the growing season and placed inside clip-cages attached to the leaves. After 10 days, offspring thus produced by the parthenogenetic apterous females were picked with a fine brush and transferred to a new clip cage to generate a single clonal lineage. Nymphal (N1 to N4) stages were pooled together while apterous female adults were harvested separately. Aphids were collected directly from clip cages, immediately flash-frozen in liquid nitrogen, and stored at -80°C for RNA extraction. Three biological replicates of both the stages i.e., adult and nymph were harvested.

### Total RNA extraction and sequencing

Total RNA was extracted from ~50 morphs in each sample using RNeasy<sup>®</sup> Mini Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. Total extracted RNA was quantified using Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and its integrity and quality were assessed using TapeStation 2200 (Agilent, CA, USA). Purified total RNA was used for transcriptome library preparation using NEBNext<sup>®</sup> Ultra<sup>™</sup> II Directional RNA Library Prep Kit for Illumina<sup>®</sup> (New England Biolabs, Ipswich, MA, USA). cDNA library was quantified using Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and quality assessment was done by TapeStation 2200 (Agilent Technologies, Santa Clara, CA, USA). Transcriptome sequencing was performed on Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with 151-base pair (bp) paired-end reads.

### Data assembly and annotation

Raw reads were processed to filter low-quality reads and adaptor sequences using Fastp (v0.23.2.0) (Chen et al., 2018) with a phred score cut-off of 30 and a minimum length of 75 bp. High-quality reads thus retained of all the three replicates of adults and nymphs of both the aphid species were catenated together (forward R1, reverse R2 separately) and were assembled *de novo* using Trinity assembler (v2.11.0) (Grabherr et al., 2011) for adult and nymph samples of *L. erysimi*. For adult and nymph samples of *M. persicae*, the reads were first aligned to publicly available reference genome of *M. persicae* MPER\_G0061.0 (GCF\_001856785.1). Aligned reads were assembled using Trinity assembler in -genome\_guided\_bam mode. Final transcriptome assembly was generated after performing clustering using CD-HITest (v4.8.1) (Li and Godzik, 2006) at 80% similarity. The completeness of the transcriptome assembly was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) (v5.4.7) (Simão et al., 2015) against insecta\_odb10 database. Assembled unigenes were functionally annotated using

the Blast2GO program against public databases NCBI non-redundant (Nr) and AphidBase ('local database' was created by combining Nr and aphidbase data against which blast was performed). Functional classification of all the unigenes including Gene Ontology (GO) was performed using Blast2GO. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using DAVID (Sherman et al., 2007).

## Identification of differentially expressed genes

Differentially expressed transcripts between adults and nymphs (adult vs nymph) of both the species as well as between the same stages of the two species (adult vs adult; nymph vs nymph) were identified using R-based tool DESEQ2 (v1.36.0.5) (Love et al., 2014) by comparing normalised Fragments Per Kilobase of transcripts per Million mapped reads (FPKM) count generated by RSEM (Li and Dewey, 2011). Transcripts that showed expression levels in at least two of the replicates were selected for differential gene analysis. The resulting p-values were adjusted to control the false discovery rate using Benjamini-Hochberg method. Among the adults or nymphs of two species, transcripts with log2fold change  $\geq 2$  or  $< 2$   $P_{\text{adjusted}}$ -value less than 0.05 were considered as upregulated or downregulated respectively among differentially expressed genes (DEGs).

## In silico identification of effectors

To predict putative effector molecules from transcriptome data of *L. erysimi* and *M. persicae*, we used the pipeline described by Prajapati et al. (2020) and Min (2010). All the assembled transcripts were subjected to clustering using CDHIT-est to remove any redundancy. These unigenes were analysed using TransDecoder (<https://github.com/TransDecoder/TransDecoder.git>) for ORF calling with a cut off value of minimum 100 amino acids and all the possible ORFs were translated to proteins. Predicted proteins thus obtained were used for the identification of candidate effector molecules in both the aphid species. The predicted proteins were filtered for the presence of signal peptide using SignalP (Petersen et al., 2011) (SignalP package in InterProScan). Filtered proteins were further analysed for the presence of transmembrane domain using InterProScan (v5.62.94) (Jones et al., 2014) (TMHMM package in InterProScan). Finally, subcellular localisation was predicted by TargetP-2.0 (Emanuelsson et al., 2000) and WolfPSort (Horton et al., 2007). Protein sequences which showed the presence of signal peptide in their sequence, absence of transmembrane domain and predicted to be extracellular (ext >17, Prajapati et al., 2020) using WolfPSort were considered putative effector proteins.

## Orthology and Ka/Ks analysis

The coding sequence (CDS) of each unigene was determined using TransDecoder and Blastx was conducted (with cutoff E-value

of 1e-5, 50% of query coverage) to compare the CDS of unigenes against predicted genes of pea aphid. Single copy orthologous genes between *L. erysimi* and *M. persicae* were identified using OrthoFinder (v2.5.5) (Emms and Kelly, 2019). The CDS region of predicted single copy orthologs were aligned by MAFFT (v7.505) (Katoh and Standley, 2013) using - auto parameter which automatically ran on L-INS-i mode and the alignments were trimmed using trimAl (v1.4.rev15) (Capella-Gutiérrez et al., 2009) with "gappout" option to remove poorly aligned sequences. The substitution rates between these single copy orthologous pairs were determined separately for non-synonymous sites i.e. Ka and synonymous sites i.e. Ks using approximate method in KaKs-Calculator3 (Zhang, 2022). YN method was used to perform pairwise approximate method. Since the sequencing errors were distributed equally among synonymous and non-synonymous sites, they were not expected to influence the results in this analyses (Tiffin and Hahn, 2002).

## Real-time quantitative PCR

To conduct qRT-PCR analysis, total RNA was extracted from two biological replicates of four aphid samples (*M. persicae* adult, *L. erysimi* adult, *M. persicae* nymph, *L. erysimi* nymph) and its integrity was checked using agarose gel electrophoresis after performing DNase I (Thermo Fischer Scientific, Inc., Waltham, MA, USA) treatment. First-strand cDNA synthesis was performed from 2µg of purified RNA using iScript<sup>TM</sup> cDNA synthesis kit (Bio-Rad, Hercules, CA, US) following manufacturer's protocol. Primer sequences for transcripts to be validated were designed using IDT OligoAnalyzer<sup>TM</sup> Tool (Integrated DNA Technologies, Newark, New Jersey, US) software and are provided in Supplementary File 3. The *40S ribosomal protein S23* gene was used as endogenous gene control (primer sequence provided in Supplementary File 3) to stabilise expression levels among all the replicates of aphid samples (Yang et al., 2015), each with two technical replicates. After normalisation with endogenous control gene, the relative expression levels of target transcripts were analysed by  $\Delta\Delta C_t$  method by pairwise comparison between *M. persicae* adult with *L. erysimi* adult and *M. persicae* nymph with *L. erysimi* nymph. qRT-PCR was performed on CFX Connect Real-Time System (Bio-Rad, Hercules, CA, US). To determine the statistical significance of qRT-PCR data, *t-test* was conducted ( $P_{\text{value}} < 0.05$ ).

## Results and discussion

### Illumina sequencing, transcriptome assembly and functional annotation

We generated transcriptome data of *L. erysimi* and *M. persicae* from two stages of aphid development - adults and nymphs. We obtained an average of 37.16, 28.53, 29.48 and 32.67 million raw reads of *L. erysimi* adult, *M. persicae* adult, *L. erysimi* nymph and *M. persicae* nymph, respectively. The raw reads were processed for



TABLE 1 Summary of transcriptomic data generated in *Myzus persicae* and *Lipaphis erysimi* at adult and nymph stage.

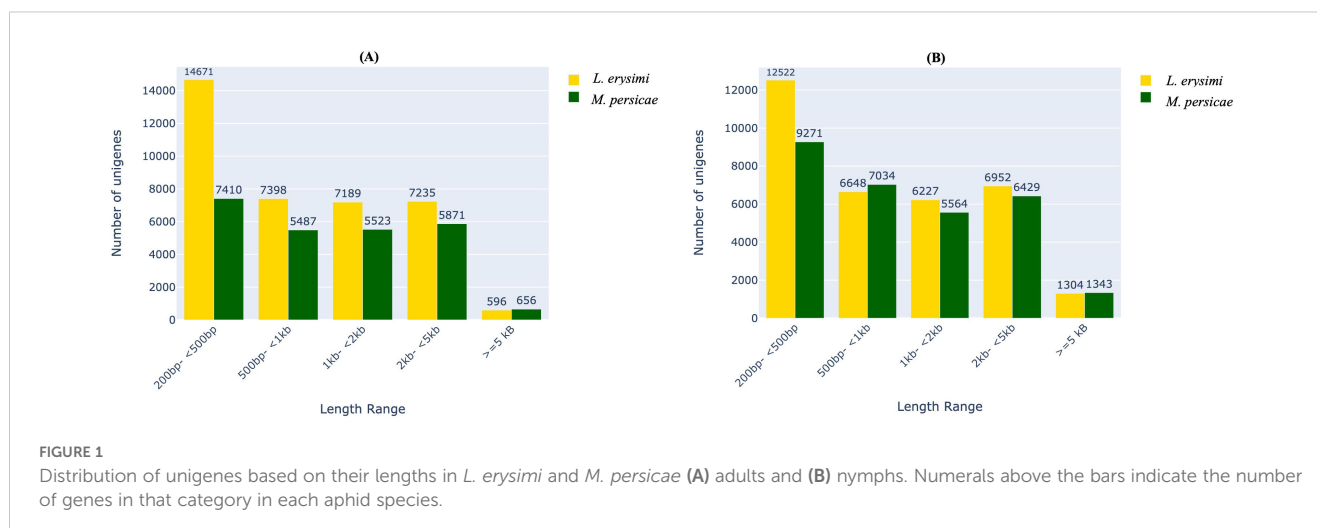
	<i>M. persicae</i> adult	<i>M. persicae</i> nymph	<i>L. erysimi</i> adult	<i>L. erysimi</i> nymph
Average Raw reads (bp)	28533626	32677230	37166324	29486830
Average Filtered read (bp)	27338676	31670028	33655556	28691401
Total unigenes	27420	29642	42963	33656
Total assembled bases	36894745	45431928	47127835	48485979
Average sequence length (bp)	1345.54	1532.69	1096.94	1440.63
N50 value (bp)	2367	2763	2126	2634
GC%	35.52	33.59	34.97	33.98
Complete BUSCOs	90.9%	85.0%	90.5%	86.3%
Complete and single copy BUSCOs	75.3%	68.7%	64.8%	61.0%
Complete and duplicated BUSCOs	15.6%	16.3%	25.7%	25.3%
Fragmented BUSCOs	1.0%	1.5%	4.0%	2.0%
Missing BUSCOs	8.1%	13.5%	5.5%	11.7%

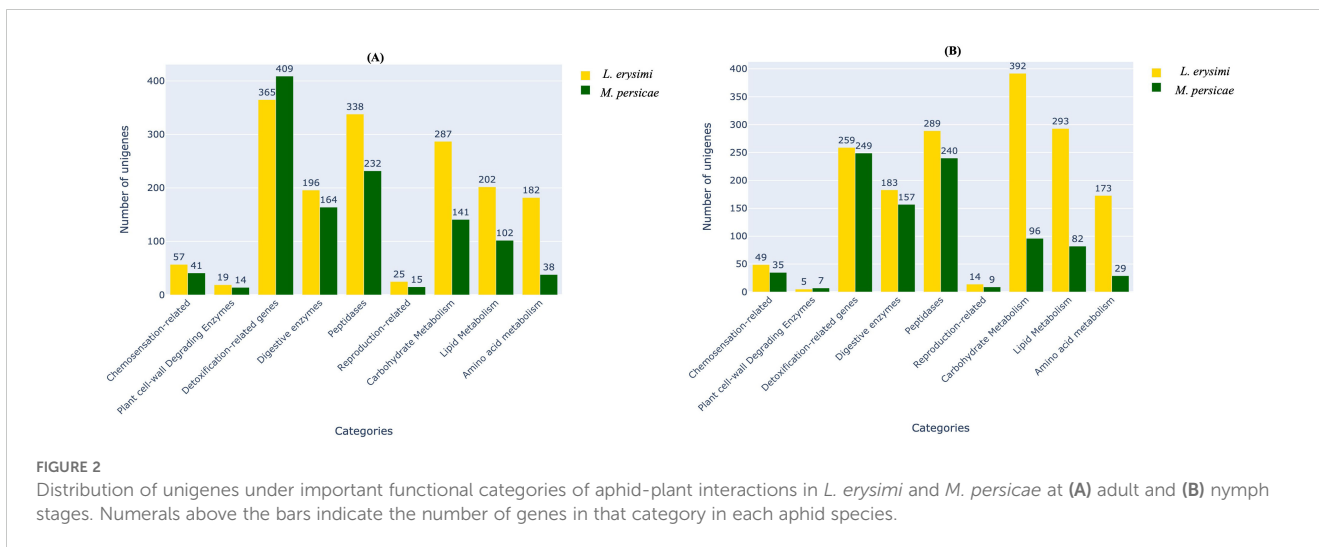
quality screening. The clean, filtered reads thus obtained were used for assembly and further clustering of transcripts that resulted in unigenes with an average length of 1096.94 bp, 1345.54 bp, 1440.63 bp, 1532.69 bp for *L. erysimi* adult, *M. persicae* adult, *L. erysimi* nymph and *M. persicae* nymph, respectively (Table 1). High BUSCO scores for all samples indicated a high degree of completeness and quality of the assembly. BUSCO scores were comparable for *M. persicae* and *L. erysimi* adults (90.9% and 90.5%, respectively) and nymphs (85% and 86.3%, respectively) (Table 1). N50 values were also greater than 2 kb indicating the superior quality of all the assemblies.

Analysis of unigenes in different categories of transcript lengths indicated that *L. erysimi* had a larger number of unigenes in most categories as compared to *M. persicae* in both adults and nymphs (Figure 1). This could be attributed to *de novo* assembly of *L. erysimi* transcriptome in contrast to the *M. persicae* transcriptome for which, reference-based assembly was done. These results are in consonance with an earlier study on aphid genomes (Mathers et al.,

2017) which showed higher number of genes in *Acyrtosiphon pisum*, a specialist aphid, due to extensive duplication of genes from conserved gene families and four-times the number of lineage-specific genes than in *M. persicae*, a generalist. However, for transcripts  $\geq 5$ kb, we observed more unigenes in both adults and nymphs of the generalist, *M. persicae* than *L. erysimi*. Among transcripts in the 500bp -1kb length range, more unigenes were found in nymphs of *M. persicae* as compared to *L. erysimi* (Figure 1). Unigenes were annotated against multiple databases including merged NCBI-Nr and aphidbase, GO, KEGG, PFAM and EC number databases. Approximately 64.7%, 64.2%, 61.8% and 53.1% of total unigenes of *L. erysimi* adult, *M. persicae* adult, *L. erysimi* nymph and *M. persicae* nymph, respectively were annotated by at least one of the above databases (Supplementary Table 1). Many unigenes were annotated with multiple databases.

Aphid infestation is associated with variations in expression of genes involved with important functions such as host plant recognition, degradation of plant cell wall, detoxification, digestion,





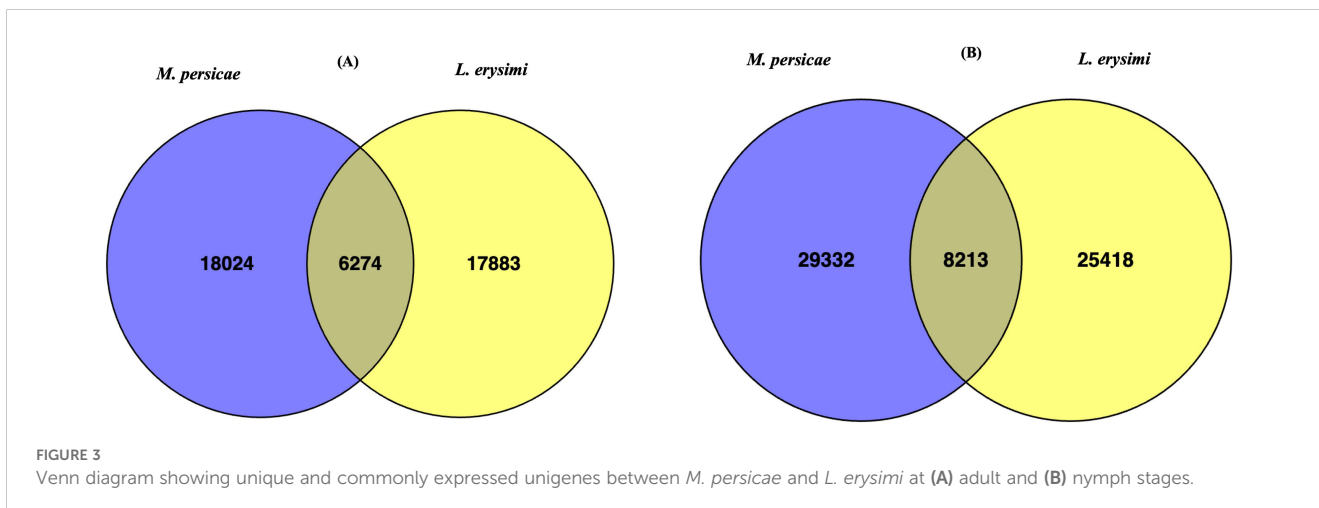
**FIGURE 2** Distribution of unigenes under important functional categories of aphid-plant interactions in *L. erysimi* and *M. persicae* at (A) adult and (B) nymph stages. Numerals above the bars indicate the number of genes in that category in each aphid species.

and metabolism (Simon et al., 2015; Silva-Sanzana et al., 2020). These categories of genes were identified in both the generalist and specialist aphids in adult as well as nymph stages (Figure 2). Earlier studies (Cristofolletti et al., 2003; Pyati et al., 2011; Roy et al., 2016; Hilliou et al., 2021) described limited number of categories such as detoxification enzymes, digestive enzymes, transporters, immunity but did not specifically focus on the variations between a specialist and a generalist. Therefore, our study aims to highlight the key differences among these two categories of aphid species and their infestation and feeding strategies. Hence, we classified all the unigenes obtained in the current into nine categories viz., chemosensation-related, plant cell-wall degrading enzymes, detoxification-related genes, digestive enzymes, peptidases, reproduction-related, carbohydrate metabolism, lipid metabolism and amino acid metabolism, all of which play an important role during aphid infestation on host plants. For most functional categories, the number of unigenes was higher in *L. erysimi* as compared to *M. persicae*. However, a significantly higher number of transcripts under the detoxification-related genes category was obtained in adults of *M. persicae* (409) as compared to *L. erysimi* (365) (Figure 2A). This category of genes may play an important role in extending the host

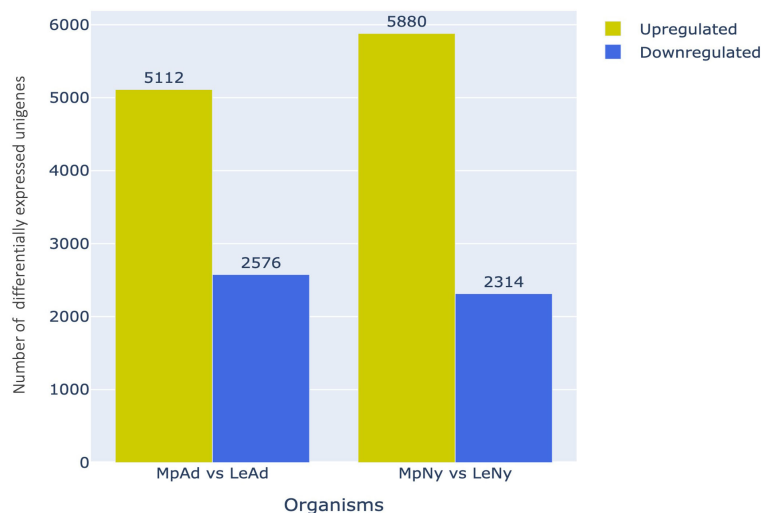
range of the generalist aphid by allowing neutralisation of a wider repertoire of defence-related secondary metabolites produced by a host plant. Higher number of unigenes were also detected in Plant Cell Wall Degrading Enzymes (PCWDEs) category at nymph stage (Figure 2B). Unigenes identified in *L. erysimi* and *M. persicae* at adult and nymph stages were categorised into transcripts that are unique for each aphid and those which are commonly expressed in both species. The generalist, *M. persicae* had greater number of unique transcripts at both developmental stages as compared to the specialist, *L. erysimi* (Figure 3). All the predicted unigenes of *L. erysimi* and *M. persicae* were subjected to gene ontology analysis. Maximum number of unigenes were majorly involved in protein binding, metabolic processes and in different cellular entities (Supplementary Figure 1).

### Differential expression of genes between generalist and specialist aphids

To identify genes potentially contributing to the generalist or specialist nature of an aphid, we performed differential expression analysis to identify transcripts that showed variations in expression



**FIGURE 3** Venn diagram showing unique and commonly expressed unigenes between *M. persicae* and *L. erysimi* at (A) adult and (B) nymph stages.



**FIGURE 4**  
Number of differentially regulated unigenes (upregulated/downregulated) in *M. persicae* relative to *L. erysimi* (Mp vs Le) at adult (Ad) and nymph (Ny) stages.

levels between the two aphid species. This analysis resulted in the identification of 15882 DEGs between *M. persicae* and *L. erysimi* of which, 7688 DEGs were identified in the adult stage and 8194 DEGs were identified in nymphs. Of the 7688 differentially regulated adult unigenes, 5112 and 2576 unigenes were upregulated and downregulated respectively in *M. persicae* relative to *L. erysimi* (Figure 4). Similarly, in nymphs, 5880 unigenes were upregulated and 2314 unigenes were downregulated in *M. persicae* relative to *L. erysimi* (Figure 4). These observations indicate that the generalist aphid has significantly higher levels of expression in a greater number of unigenes relative to the specialist.

Among the differentially expressed transcripts, 92.6% of adult transcripts and 93.9% of nymph transcripts could be annotated using at least one of the databases viz., RefSeq, KEGG pathways, Gene Ontology (GO), Pfam and E.C. number databases (Supplementary Figure 2). Among the differentially expressed unigenes, 244 adult and 194 nymph unigenes could be functionally annotated with all the above databases (Supplementary Figure 2). Among the DEGs identified between the generalist and specialist adults, 572 unigenes remained un-annotated and 131 unigenes were annotated as ‘uncharacterised proteins’. Among the nymph DEGs, 499 unigenes remained unannotated and 94 DEGs were uncharacterised proteins. The species-based distribution of annotated DEGs against Nr and aphidbase databases are depicted in Supplementary Figure 3 indicating maximum correspondence to *M. persicae* (35.4% in adults and 40% in nymphs).

## Gene ontology, pFAM and, KEGG pathway annotation of DEGs

According to Gene Ontology classification, 5004 and 5306 differentially expressed unigenes at adult and nymph stages,

respectively were categorised into biological process, cellular components, and molecular functions. The distribution of top 15 GO classes indicated that the categories of upregulated and downregulated unigenes were similar in both adults and nymphs (Supplementary Figure 4). However, the category of ‘oxidoreductases’, which also includes unigenes involved in detoxification, was unique to the upregulated category of unigenes in *M. persicae* adults and absent in *L. erysimi* adults indicating a more robust response of the generalist aphid in detoxification. Another category of ‘protein synthesis or translation’ in molecular function also included a higher number of up-regulated unigenes at both the stages relative to down-regulated unigenes in *M. persicae* relative to *L. erysimi* (Supplementary Figure 4).

To infer differences, if any, in enzymatic functions between the generalist and specialist aphids, E.C. numbers along with the nomenclature of differentially expressed unigenes between *M. persicae* and *L. erysimi* were studied and a total of 1809 and 1948 unigenes were assigned E.C. numbers in adult and nymph stages, respectively (Supplementary Figure 5). The top 15 enzyme categories included those that play an important role during aphid feeding and combating host defences. Examples include oxidoreductases, hydrolases (Sadek et al., 2013), peptidases and enzymes involved in oxidation-reduction reactions. Our data indicates a significantly higher number of upregulated unigenes in these categories in the generalist aphid as compared to the specialist. Further, higher expression levels of these unigenes in the generalist aphid indicates a more vigorous system of sequestering secondary metabolites produced by host plants against insect herbivory at both developmental stages when compared to the specialist aphid.

Differentially expressed unigenes between the two aphid species across different biochemical pathways assigned using the KEGG database were also studied. In total, 1473 differentially expressed unigenes were assigned 33 KEGG pathways at the adult stage and 31 pathways were assigned to 957 differentially expressed unigenes at

the nymph stage of both the species. The top 15 differentially expressed pathways between the generalist and the specialist aphid at both the stages (Supplementary Figure 6) included metabolic pathways, carbohydrate metabolism, signal transduction, xenobiotics biodegradation and immune system.

## Classification of DEGs into categories of interest

From host recognition to initiation of the feeding process, aphids undergo various physiological changes related to multiple processes *viz.*, olfactory signal transduction, sap-sucking, detoxification, digestion, combating host defences and procreation (Powell et al., 2006; Simon et al., 2015; Wang et al., 2020b). Concurrently, they also alter the host system by eliciting and/or curbing defence responses (Giordanengo et al., 2010; Elzinga and Jander, 2013; Mondal, 2017). We had earlier categorised the unigenes of *M. persicae* and *L. erysimi* into nine functionally important categories associated with the above processes and reported variations in the number of unigenes in each category between the two aphid species (Figure 2). Based on differential expression data from our transcriptomes, we analysed the number and distribution of differentially expressed genes in each of these nine categories (Supplementary File 1). Our data indicates that in almost all categories of interest except carbohydrate metabolism, lipid metabolism and amino acid metabolism, the generalist aphid, *M. persicae* had greater number of unigenes that showed significantly higher expression levels compared to the specialist, *L. erysimi* (Figure 5). These unigenes might be important candidate genes associated with the generalist nature of *M. persicae*.

We also analysed the distribution of genes among the nine important functional categories of interest vis-à-vis their fold-change in expression in both adults and nymphs (Figure 6). Those unigenes which showed differential expression in *M. persicae* at both developmental stages ranged from 2- to >17 log<sub>2</sub>-fold changes while the corresponding range for downregulation was -2 to -8 log<sub>2</sub>-fold changes relative to *L. erysimi*. This indicated that the range of

upregulation in the generalist, *M. persicae* was significantly higher than that observed in the specialist, *L. erysimi*. In *M. persicae* adults, maximum number of unigenes (1998) showed an upregulation in the 12- to 17 log<sub>2</sub>-fold range. We observed a -2- to -4 log<sub>2</sub>-fold range reduction in expression levels for 2248 genes in *M. persicae* (Figure 6). Similar trends were observed in nymphs for both species with 3494 genes of *M. persicae* being upregulated in the 12- to 17 log<sub>2</sub>-fold range and 2154 genes were downregulated in the 2- to 4- log<sub>2</sub>-fold range when compared to *L. erysimi*. A significantly high number of genes (608 in adults and 803 in nymphs) were in the maximum range of upregulation (>17 log<sub>2</sub>-folds) in *M. persicae*. Many differentially regulated unigenes from both adults and nymphs could not be annotated (Supplementary Table 2). All the un-annotated unigenes which are differentially expressed at higher fold-changes could be novel candidate genes which might be functionally associated with the generalist nature of *M. persicae* and/or specialist behaviour of *L. erysimi*. These candidate genes would need to be studied further to understand their functional role(s), if any, in the biology of both species.

## Variations in host recognition and selection genes

To settle on a host plant, it is important for herbivorous insects to perceive cues related to a suitable host. This is achieved by activating a complex chemosensory pathway of host recognition that includes various kinds of proteins like odorant binding receptor/protein, pheromone-binding protein, olfactory receptors, gustatory receptors and chemosensory receptors (Zhang et al., 2020; Wang et al., 2020b; Huang et al., 2023; Powell et al., 2006). An earlier study by He et al. (2023) revealed the presence of fewer chemosensory genes, odorant binding proteins and gustatory receptors in *Schlechtendalia chinensis*, a specialist aphid in contrast to *Myzus persicae*. A lower number of chemosensory genes in a specialist aphid may be the cause of its narrow host range requiring fewer host recognition genes as compared to the generalist aphid. Likewise in our study, among the differentially expressed unigenes between *M. persicae* and *L. erysimi*, ‘chemosensation-related’ properties were present in 18 and

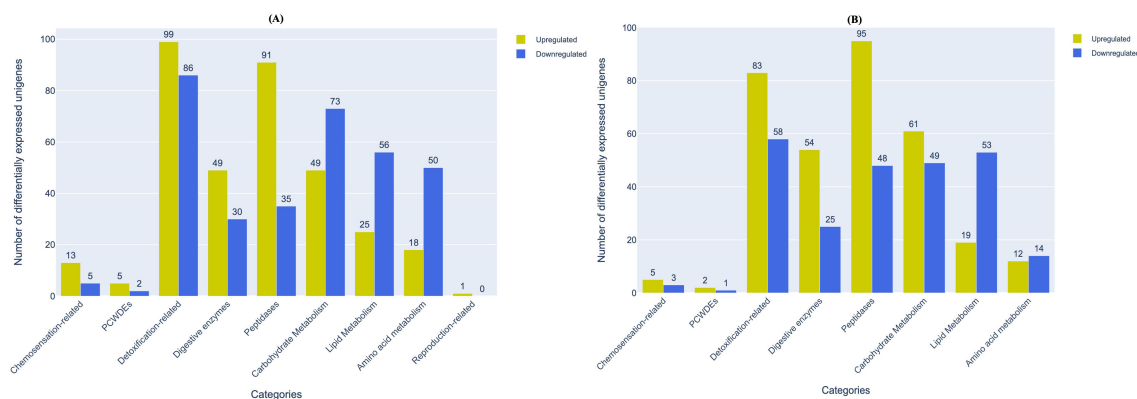


FIGURE 5

Distribution of differentially expressed unigenes between *M. persicae* and *L. erysimi* in important functional categories of interest in (A) adults and (B) nymphs. Numerals above the bars indicate the number of genes in that category in each aphid species.



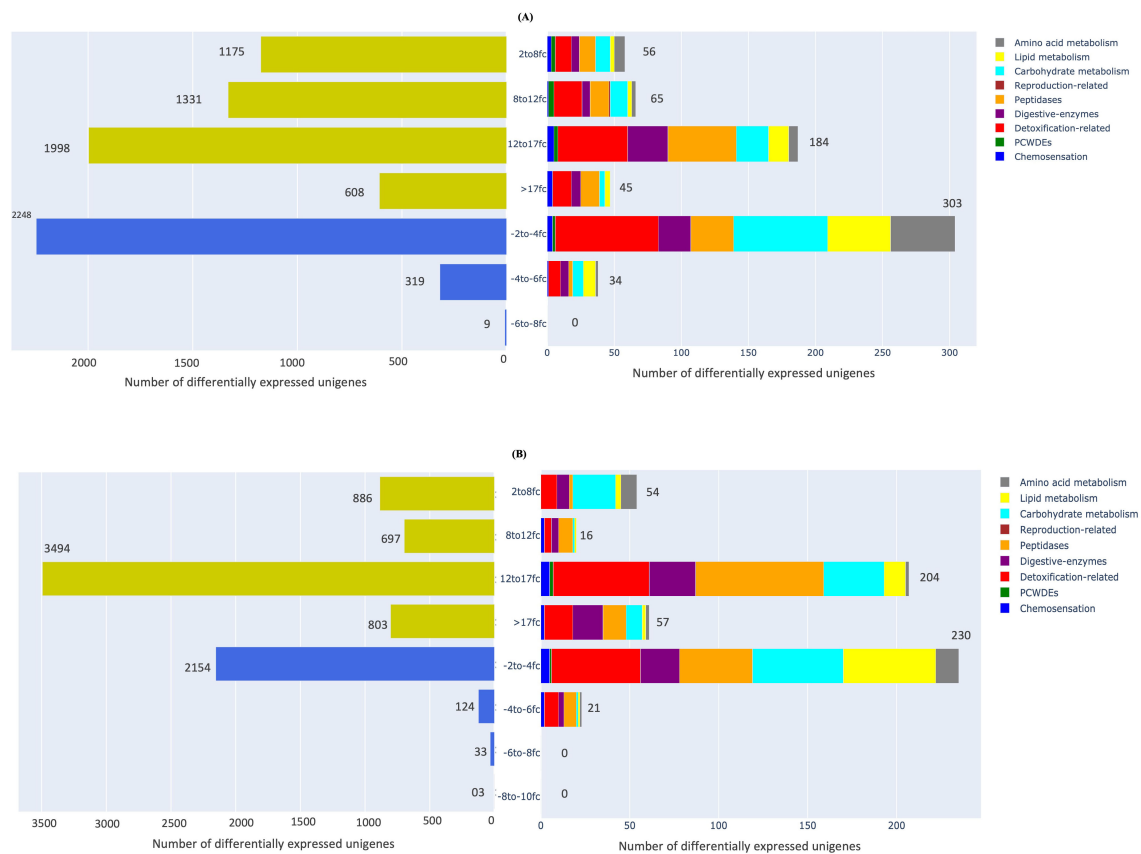


FIGURE 6

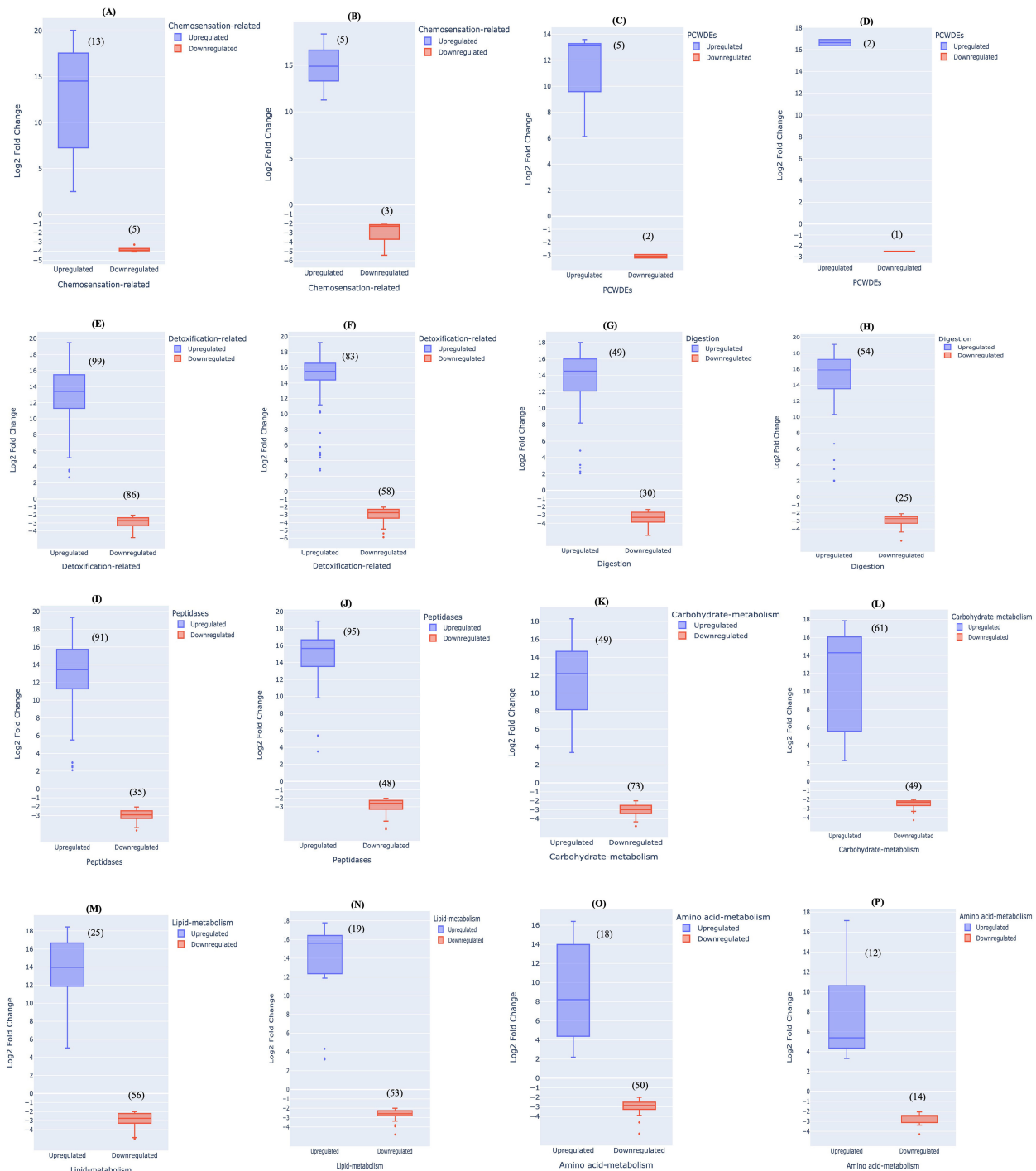
Distribution of differentially expressed unigenes between *M. persicae* and *L. erysimi* adults and nymphs in different ranges of log<sub>2</sub>-fold changes (fc). Yellow-coloured bars of the graph on the left-hand side represent upregulated range in *M. persicae* relative to *L. erysimi* whereas blue-coloured bars represent downregulated range in (A) adults and (B) nymphs. Bar graph on the right-hand side of the figure represents different ranges of fold change among genes specific to the nine functionally important categories in (A) adults (B) nymphs.

8 unigenes at adult and nymph stages, respectively. Among these, 13 unigenes showed higher expression in the range of 2- to 20- log<sub>2</sub>-fold changes relative to *L. erysimi* and 5 genes were downregulated in adults of *M. persicae* (Figure 7A). Similarly, 5 unigenes showed higher expression in *M. persicae* nymphs as compared to *L. erysimi* ranging from 11- to 18- log<sub>2</sub>-fold changes while 3 unigenes were downregulated in range of -2 to -5 log<sub>2</sub>-fold changes (Figure 7B). Generalist aphids are considered to be more flexible in their host preferences and these differences with the specialist could influence their search for appropriate hosts (Tapia et al., 2015). Therefore, upregulation of a larger number of unigenes involved in 'chemosensation' in *M. persicae* might indicate a more robust, active and diversified host-recognition mechanisms in the generalist aphid as compared to the specialist.

## Variations in plant cell wall degrading enzyme genes

Aphid saliva contains PCWDEs which facilitate stylet penetration into host tissues. These enzymes are important for the infestation process and also play a role in eliciting plant defence responses (Silva-Sanzana et al., 2020). Studies conducted on aphids

and related hemipterans have identified cell-wall degrading enzymatic activities in their saliva. Pectinase and cellulase activity have been observed in the saliva of *S. avenae* (Guangxi et al., 2006) while a transcriptome study on *M. persicae* reported the presence of cellulase transcripts (Thorpe et al., 2016). In our study, we identified differentially expressed unigenes for PCWDEs including pectinacetylsterases, glucanases, glucosidases and beta-mannosidase, all of which are responsible for cell wall degradation and aid in stylet penetration (Zhang et al., 2023). In the generalist aphid, 5 and 2 unigenes were upregulated with 10- to 13- log<sub>2</sub> fold-changes and 13 log<sub>2</sub> fold changes at adult and nymph stages, respectively (Figures 7C, D). Downregulation in expression levels was observed in two unigenes with -2 to -3 log<sub>2</sub> fold-changes in *M. persicae* adults and at nymph stage, one unigene was downregulated as compared to *L. erysimi* (Figures 7C, D). In both developmental stages of *M. persicae*, we observed a greater number of upregulated unigenes with higher fold changes as compared to *L. erysimi* that may provide a greater ability to the generalist aphid for altering the cell wall of host plant as compared to the specialist. Furthermore, since a generalist aphid has a broader host range, involvement of more number of cell wall-altering genes may further assist the generalist aphid to commence the penetration and sap-testing process more rapidly as compared to the specialist.



**FIGURE 7** Range of upregulation and downregulation of differentially expressed unigenes between *M. persicae* and *L. erysimi* at adult stage in different categories- Chemosensation-related (A) adults (B) nymphs; plant cell wall modifying enzymes (PCWMEs) (C) adults (D) nymphs; detoxification-related (E) adults (F) nymphs; digestive-enzymes (G) adults (H) nymphs; peptidases (I) adults (J) nymphs; carbohydrate-metabolism (K) adults (L) nymphs; lipid-metabolism (M) adults (N) nymphs; amino-acid metabolism (O) adults (P) nymphs.

## Variations in detoxification-related genes

Selection of host plants by aphids is influenced by the diversity of plant metabolites. While a specialist aphid chooses the host based on few metabolites (mainly secondary metabolites), a generalist aphid screens a much larger scale of both primary and secondary

metabolites (Kennedy and Booth, 1951; Bernays, 2001; Powell et al., 2006). Secondary metabolites released by host plants are an important component of plant defence against insect herbivory. Moreover, these metabolites also have antixenotic properties that either repel aphids or attract their natural enemies (Wagner et al., 2004; Walling, 2008; Sadek et al., 2013). Insects respond to such

metabolites by releasing detoxifying enzymes and oxidoreductases (Cai et al., 2009). In a study by Ramsey et al. (2010), comparison of transcriptome data generated for a generalist, *M. persicae* (green peach aphid) with data available for a specialist, *A. pisum* (pea aphid) indicated 40% higher number of cytochrome P450 detoxifying enzyme genes in the generalist compared to the specialist. No significant differences were detected in other enzymes between the two species. In our study, we identified 185 and 141 detoxification-related, differentially expressed unigenes in adult and nymph stages, respectively between *M. persicae* and *L. erysimi*. Examples of such genes include catalase, superoxide dismutase, cytochrome P450, glutathione S-transferase, peroxidase, UDP-glucuronosyltransferase, genes involved in cellular oxidant detoxification, oxidoreductase activity and monooxygenase activity (Sadek et al., 2013; Wang et al., 2020a; Durak et al., 2021; Hilliou et al., 2021). Of all the differentially regulated unigenes between *L. erysimi* and *M. persicae* that might be involved in detoxification process, 99 and 83 unigenes showed significantly higher expression levels in adults and nymphs, respectively of *M. persicae*. Further, these genes had substantial log<sub>2</sub>-fold changes of 2- to 19- log<sub>2</sub> fold changes in adults and 4- to 18- log<sub>2</sub> fold changes in nymphs, respectively. On the other hand, 86 and 58 unigenes showed downregulation of -2 to -4 and -2 to -5 log<sub>2</sub> fold changes in adults and nymphs, respectively (Figures 7E, F). These results indicate that the extent of upregulation of detoxification genes in *M. persicae* was significantly higher than the downregulation levels of detoxification genes in *M. persicae*. A greater number of unigenes involved in detoxification and more importantly, a much higher extent of up-regulation in *M. persicae* might be attributed to the fact that *M. persicae* being a generalist, is exposed to varied types of plant defensive metabolites. To counter a larger repertoire of plant metabolites and overcome challenges like plant toxins and digestive inhibitors, generalists may have to adopt a wider and a more robust system for detoxification (Bansal and Michel, 2018). On the other hand, specialists are well-adapted to their host plants and therefore might have a limited but streamlined detoxification system suitable for neutralising specific metabolites of their host plants. This was demonstrated in a study by Govind et al. (2010) on a wild type and a defence-less transgenic plant of *Nicotiana attenuata*, wherein the specialist herbivore (*Manduca sexta*), being adapted to its host metabolite, nicotine, regulated only specific detoxifying genes according to the defence response of the host plant. However, the generalist herbivore (*Heliothis virescens*) continued to generate a robust response even against the defence-less transgenic plant i.e., irrespective of the status of its host plant.

## Variations in digestion-related genes

Differentially expressed unigenes involved as digestive enzymes were included in this category. This category comprises cathepsin, maltase, aminopeptidase, trehalase, alpha-glucosidase and trypsin (Figures 7G, H). In this category too, *M. persicae* had a larger number of significantly upregulated unigenes with 49 and 54 unigenes (including variants of the above enzymes) being upregulated in adults and nymphs, respectively. The upregulation

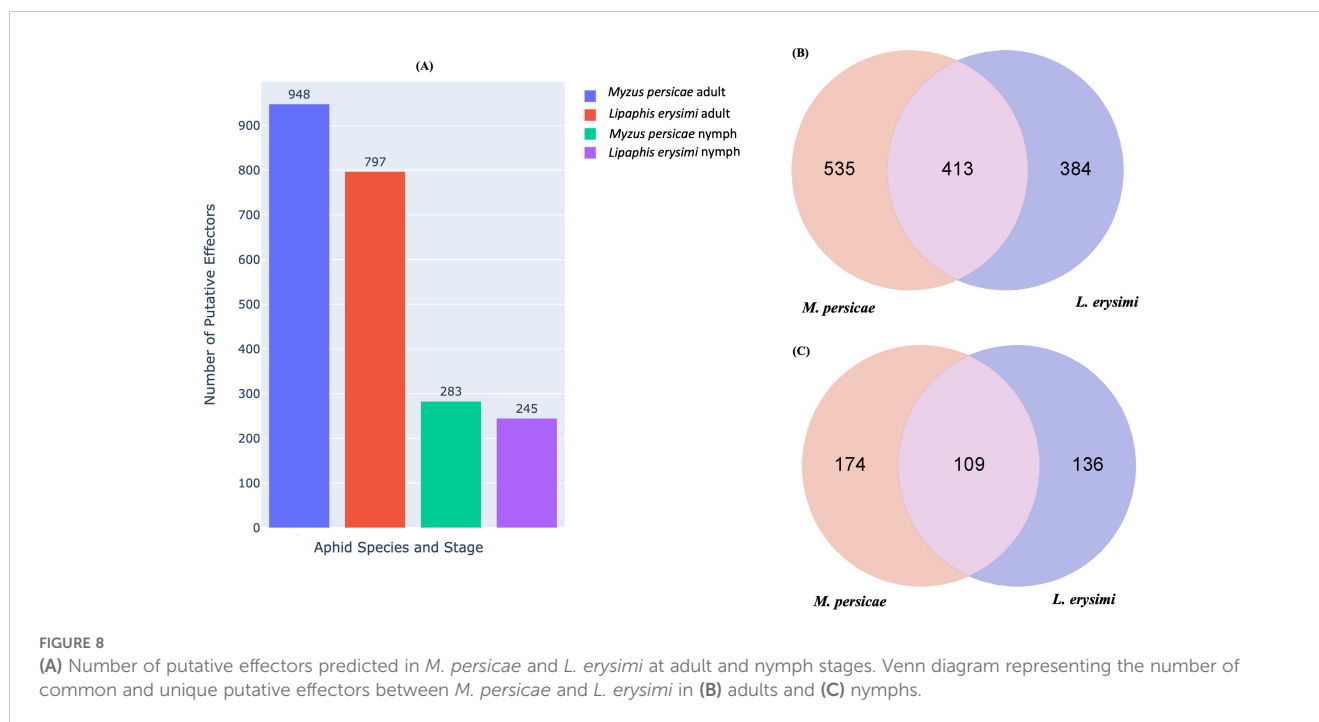
varied from 2- to 18- fold change and 2- to 19- log<sub>2</sub> fold change in adults and nymphs, respectively (Figures 7G, H). The downregulated unigenes included 30 and 25 unigenes with a range of -2 to -5 log<sub>2</sub> fold change and -2 to -4 log<sub>2</sub> fold change in adults and nymphs, respectively (Figures 7G, H). In our study more number of upregulated genes were observed in all categories of digestive enzymes (except Maltase) in both adults and nymphs of *M. persicae* as compared to *L. erysimi* (Supplementary Figure 7). Due to narrower host range, a specialist aphid might require fewer specialised enzymes to process chemical compounds present in their diet. On the other hand, our results also indicate that polyphagy shown by a generalist aphid requires a broader range of adaptations to digest varied components of diet from multiple hosts and correlates with the study of Roy et al. (2016).

## Variations in proteolysis-related genes

In addition to high sugar content, phloem sap also contains proteins that can vary between 0.3 to 60mg/ml depending upon the host species (Pyati et al., 2011). It was earlier believed that ingested plant-sap does not undergo proteolysis in hemipterans. However, recent studies have shown that proteolysis in sap-feeding insects is essential for their proper nutrition (Salvucci et al., 1998; Foissac et al., 2002; Cristofolletti et al., 2003). The category of 'Peptidases' comprised of all unigenes that have enzymatic action on peptide bonds and included a total of 126 and 143 differentially expressed unigenes between *L. erysimi* and *M. persicae* at adult and nymph stage. Of these, 91 and 95 unigenes were upregulated in *M. persicae* at adult and nymph stages, respectively while 35 and 48 unigenes showed lower levels of expression in adults and nymph stages, respectively (Figures 7I, J). Upregulation of these differentially expressed unigenes ranged from 2- to 19- log<sub>2</sub> fold change and 3- to 19- log<sub>2</sub> fold changes in adults and nymphs, respectively (Figures 7I, J). On the other hand, downregulation ranged from -2 to -4 log<sub>2</sub> fold changes and -2 to -6 log<sub>2</sub> fold changes in adults and nymphs, respectively (Figures 7I, J). Phloem sap ingested by aphids also contains many proteins with anti-insect properties as a component of plant defence (Kehr, 2006). Consequently, peptidases allow aphids to neutralise the effects of these defence proteins to facilitate infestation of the host plant. The presence of a larger number and an increased level of upregulation of peptidases in *M. persicae* may be advantageous in broadening its host range and allowing rapid infestation.

## Variations in reproduction-related genes

As soon as favourable conditions are available during the growing season of *Brassica juncea*, both *M. persicae* and *L. erysimi* become prevalent. In order to establish its population on the host plant, the aphids need to expedite the process of feeding and reproduction. In case of a generalist aphid, higher number/expression of reproduction-related genes would allow rapid establishment of these aphids on a broad range of host plants. Interestingly, our study identified more number of reproduction-related genes in *L. erysimi* as compared to *M. persicae* (Figure 2).



However, in consonance with the above hypothesis, one unigene related to ‘reproduction-process’ had significantly higher expression levels in *M. persicae* adults (more than 11 log<sub>2</sub> fold change) as compared to adults of *L. erysimi*. None of the other reproduction-related genes showed differential expression between the two aphid species. It would be interesting to study the functional role of this gene in aphid biology and ascertain its contribution, if any, to the increased occurrence, fecundity, persistence and distribution of the generalist aphid on host plants, including *B. juncea*.

## Variations in metabolism-related genes

Categories such as metabolism of carbohydrates, lipids and amino acids included a total of 122, 81 and 68 differentially regulated unigenes, respectively between *M. persicae* and *L. erysimi* adults (Figures 7K, M, O). The corresponding numbers in nymphs were 110, 72, 26 differentially regulated unigenes (Figures 7L, N, P). In these categories, most differentially expressed unigenes showed significantly lower expression levels in *M. persicae* at both the developmental stages in the range of -2 to -6 log<sub>2</sub> fold changes in adults and nymphs. Though the number of upregulated unigenes in *M. persicae* was lower than that of *L. erysimi* in all categories except carbohydrate metabolism in nymphs (Figure 5), the extent of upregulation was much higher (from 2- to >17 log<sub>2</sub> fold change) in the generalist aphid indicating an increased ability to digest these categories of molecules from phloem sap. However, greater number of upregulated unigenes in *L. erysimi* might be an indication that specialists are capable of processing their diet from specific host plant more efficiently and completely than generalists. Since generalists have the option to feed upon multiple hosts, they might have a more generalised mechanism of

metabolism for all kinds of host plants which may further lead to incomplete digestion of metabolites (Olazcuaga et al., 2023).

Our study was targeted towards studying variations in the molecular responses displayed by a generalist aphid and a specialist aphid growing under the same conditions and feeding on the same host plant i.e., *Brassica juncea* var. Varuna. It allowed analysis of the similarities and differences in expression of genes involved in aphid-plant interactions and more importantly, it facilitated identification of differences in the responses of the generalist and the specialist aphid against defence mechanisms of the host. It generated an exhaustive repertoire of genes in the generalist and specialist aphids and allowed their comparison between the corresponding adult and nymph developmental stages. To further validate and substantiate the identified trends in our study between generalist and specialist aphids, further studies can include a larger number of generalist aphids with a broader host range and corresponding specialists, which would provide more comprehensive information about the molecular signatures that confer advantage to generalist aphids and the evolutionary mechanisms.

## Identification of effectors and their variations between generalist and specialist aphids

Aphids, while probing and feeding on host plants, release effector molecules many of which, are species-specific (Elzinga and Jander, 2013). Effectors play a major role in aphid-plant interactions including host recognition (Mondal, 2017, 2020). Many effectors are recognised by host plants which then initiate defence responses against the insect and this effectors-based interaction determines the consequences of aphid and host plant interactions (Bos et al., 2010;

TABLE 2 Number of total and unique effectors in *M. persicae* and *L. erysimi* under different categories of interest at adult and nymph stage.

Categories of interest	Total number of putative effectors in <i>M. persicae</i>	Total number of putative effectors in <i>L. erysimi</i>	Number of unique effectors in <i>M. persicae</i>	Number of unique effectors in <i>L. erysimi</i>
<b>Adults</b>				
Chemosensation-related	9	13	1	5
PCWDEs*	3	5	1	3
Detoxification-related	21	11	18	8
Digestive enzymes	64	37	45	18
<b>Nymphs</b>				
Chemosensation-related	6	6	5	5
PCWDEs*	1	0	1	0
Detoxification-related	2	4	2	4
Digestive enzymes	13	8	6	1

\*PCWDEs, Plant Cell Wall Degrading Enzymes.

Elzinga and Jander, 2013; Jaouannet et al., 2014; Mondal, 2017). *In silico* identification of putative effectors identified 948 effectors in adults and 283 effectors in nymphs of *M. persicae* while in *L. erysimi*, 797 and 245 putative effectors showed expression at adult and nymph stages, respectively (Figure 8A). Among the identified effectors, 535 and 174 effectors were unique to *M. persicae* adults and nymphs, respectively while the corresponding figures for *L. erysimi* adults and nymphs were 384 and 137, respectively (Figures 8B, C). The larger repertoire of effectors, including a higher number of unique effectors in the generalist aphid compared to the specialist might facilitate feeding, modulate hosts' defence responses more effectively and more importantly, confer on the generalist, an ability to infest multiple hosts.

Many putative effectors identified in the current study showed homology with effectors of other aphid species available in public domain. Protein sequences of 1094 published effectors of *A. pisum*, *M. persicae* and *D. noxia* (Carolan et al., 2011; Bos et al., 2010; Nicolis et al., 2022) were obtained using transdecoder. Following manual curation, a total of 1044 protein sequences were retrieved and used to perform BlastP analysis with identified putative effectors of adults and nymphs of both species. Among adults, out of 1403 identified putative effectors, 854 (60.86%) showed homology with the published effectors. Of the 420 effectors detected in nymphs, 281 (66.9%) were found to be homologous to effectors of other aphid species. Most effectors identified in our study [652 (46.47%) in adults and 220 (52.38%) in nymphs] were homologous to *D. noxia*. The remaining effectors showed homology with *A. pisum* and *M. persicae*. (Supplementary File 2).

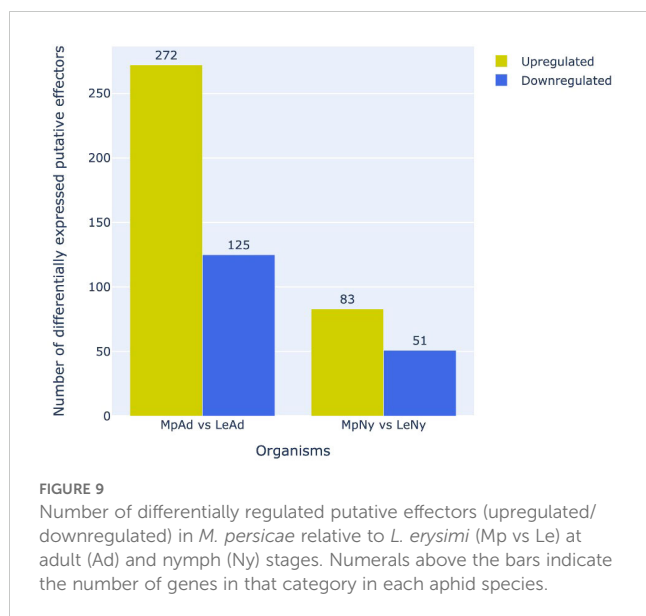
Our study also identified several effectors from both species which did not show any homology with above-mentioned published data of effectors. In *M. persicae*, out of 535 and 174 uniquely expressed putative effectors in adults and nymphs, 190 and 62 putative effectors, respectively did not show any homology with effectors of other three species. Similarly, in *L. erysimi*, out of 384 and 137 uniquely expressed putative effectors at adult and nymph stages, 176 and 54 effectors showed no homology with other

effectors. These effectors can be studied further for their functional roles which might allow identification of molecular signatures that govern variations between the generalist and specialist aphids.

Annotation of putative effectors identified in our study against multiple databases resulted in annotation of 95.2% and 100% of adult and nymph effectors, respectively. Significant progress has been achieved towards understanding the potential functions of many effectors. Some of them have been functionally characterised and have been used as candidates for introducing aphid resistance in agricultural crops. For example, *c002*, an effector protein that was first described in *A. pisum* (Mutti et al., 2006, 2008) was shown to be essential for successful and prolonged feeding on the host plant. Overexpression of *M. persicae c002* was found to promote aphid fecundity and proliferation on the host plant (Bos et al., 2010; Pitino et al., 2011). Putative effectors identified from our transcriptome data include antioxidant and detoxifying enzymes such as superoxide dismutase, peroxidase, glucose dehydrogenase and some effectors aiding the feeding process and manipulating host defence and immunity viz., odorant-binding proteins, calcium-ion binding, insect pheromone-binding and chemosensory protein. The identified effectors also included digestive enzymes such as maltase, cathepsin-B and L, aminopeptidase, trypsin, serine protease and a few plant cell wall modulating enzymes like beta-mannosidase and glucosidase II beta subunit. Sheath proteins such as mucin-2 and mucin-5AC were also identified in adults and nymphs, respectively. We also detected ARMET, a Ca<sup>2+</sup>-binding salivary protein which is known to interfere in Ca<sup>2+</sup> trafficking from ER- membrane to the sieve element. Suppression of ARMET had a negative impact on the feeding and life span of aphids (Wang et al., 2015; Hafke et al., 2009).

To analyse variations, if any, in the functional repertoire of effectors between generalist and specialist aphids, we studied their annotations from both aphids at adult and nymph stages. Adults of *M. persicae* showed a significantly greater number of effectors under the categories of 'detoxification-related' and 'digestive enzymes' (Table 2). The generalist aphid thus appears to demonstrate a





stronger response in neutralising host defence responses and digesting components acquired during feeding through its effectors, both of which could play an important role in widening its host range. Adults of *L. erysimi* showed more number of putative effectors related to host recognition (chemosensation-related) and onset of feeding process (PCWDEs) (Table 2). In case of nymphs, higher number of ‘detoxification-related’ effectors were observed in *L. erysimi* while nymphs of *M. persicae* had more number of effectors in ‘PCWDEs’ and ‘digestive enzymes’ categories (Table 2). Thus, we may conclude that effectors in the generalist adults are more active in detoxification and digestion while *L. erysimi*, being a specialist, has more effectors that aid in host localisation and also in interrupting the first line of host defence i.e. plant cell wall.

Analysis of differential expression profiles of putative effectors between *L. erysimi* and *M. persicae* indicated that a significantly higher number of effectors were upregulated in *M. persicae* in both adult and nymph stages as compared to *L. erysimi* (Figure 9). Interestingly, as many as 10 detoxification-related genes that were common to both the aphids were found to be significantly upregulated in the generalist aphid. A list of all the differentially expressed effectors along with their functional annotation is provided in Supplementary File 2. These results indicate that, in addition to the higher number of effectors, generalist aphids also have significant up-regulation of such genes (including those in important functional categories) as compared to the specialist. These molecular signatures could play an important role in determining increased efficacy of the generalist aphid in infesting a large number and wider range of host plants.

## Validation of differentially expressed unigenes by qRT-PCR

For validation of expression levels of transcripts by qRT-PCR, 14 differentially expressed transcripts with fold changes of >2 to <2

at  $p_{\text{adjusted}}$  value <0.05 were selected which also included differentially regulated putative effectors (which did not show homology with other organisms). Of these 14 transcripts, 12 transcripts showed statistically significant results at  $p_{\text{value}} < 0.05$  and the remaining two transcripts (TRINITY\_DN354\_c0\_g2\_i1, TRINITY\_GG\_4338\_c8\_g1\_i2) showed non-significant results (Figure 10). Significantly upregulated transcripts included one transcript (TRINITY\_2422\_c0\_g1\_i1) involved in reproductive process and showed high levels of expression in *M. persicae* adults as compared to adults of *L. erysimi*. This gene could be an interesting candidate for further studies on its potential contribution to higher reproductive rates of generalist aphids and could also be a potential target for aphid resistance in crop improvement strategies. Among other upregulated transcripts were two putative effector transcripts (TRINITY\_GG\_5521\_c2\_g1\_i1, TRINITY\_7609\_c1\_g1\_i1). Details of remaining transcripts are provided in Supplementary File 3. On the other hand, downregulated transcripts (in *M. persicae* adults and nymphs) included a putative effector transcript (TRINITY\_DN121\_c3\_g1\_i2), cuticular protein 14 precursor (TRINITY\_DN1411\_c0\_g2\_i1), cathepsin-B (TRINITY\_DN5469\_c0\_g1\_i1), protein yellow (TRINITY\_DN2330\_c0\_g1\_i1) transcripts and others, details of which are provided in Supplementary File 3. qRT-PCR results for all the selected differentially expressed transcripts between *M. persicae* and *L. erysimi* at adult and nymph stages were in consonance with the expression patterns concluded from transcriptome data.

## Ka/Ks analysis

In order to study the genes that have contributed significantly towards the evolutionary divergence between the two species, we performed Ka/Ks analysis between orthologous groups of the two species. Ka/Ks ratio is broadly used to provide information about the intensity of evolutionary force and its mode of selection, acting on a coding gene between organisms.  $Ka/Ks > 1$  represents a positive selection,  $Ka/Ks = 1$  indicates neutral selection and  $Ka/Ks < 1$  reflects negative selection. To understand whether a gene is undergoing purifying or diversifying selection, rate of substitutions is measured within orthologous pairs of two species. We calculated Ka/Ks ratio between *L. erysimi* and *M. persicae*. From the transcriptome data of *L. erysimi* and *M. persicae*, 2904 one-to-one orthologs were detected of which, 7 pairs of groups had  $Ka/Ks > 1$  indicating that these genes of both aphid species were under positive selection i.e., they are undergoing diversification (Figure 11). These 7 pairs of include unigenes involved in lipid metabolism, transcription, protein sumoylation which further can studied to understand their functional relevance in evolutionary divergence. Out of these 7 pairs of genes, one unannotated unigene which was identified as a putative effector showed expression only in *L. erysimi* adults. Three unigenes were differentially expressed between the two aphid species. In the generalist aphid, *M. persicae*, two of these unigenes were upregulated with 14- and 16-  $\log_2$  fold changes at nymphal stage whereas one unigene was downregulated with -2  $\log_2$  fold changes. Swanson et al. (2004), in their study re-adjusted

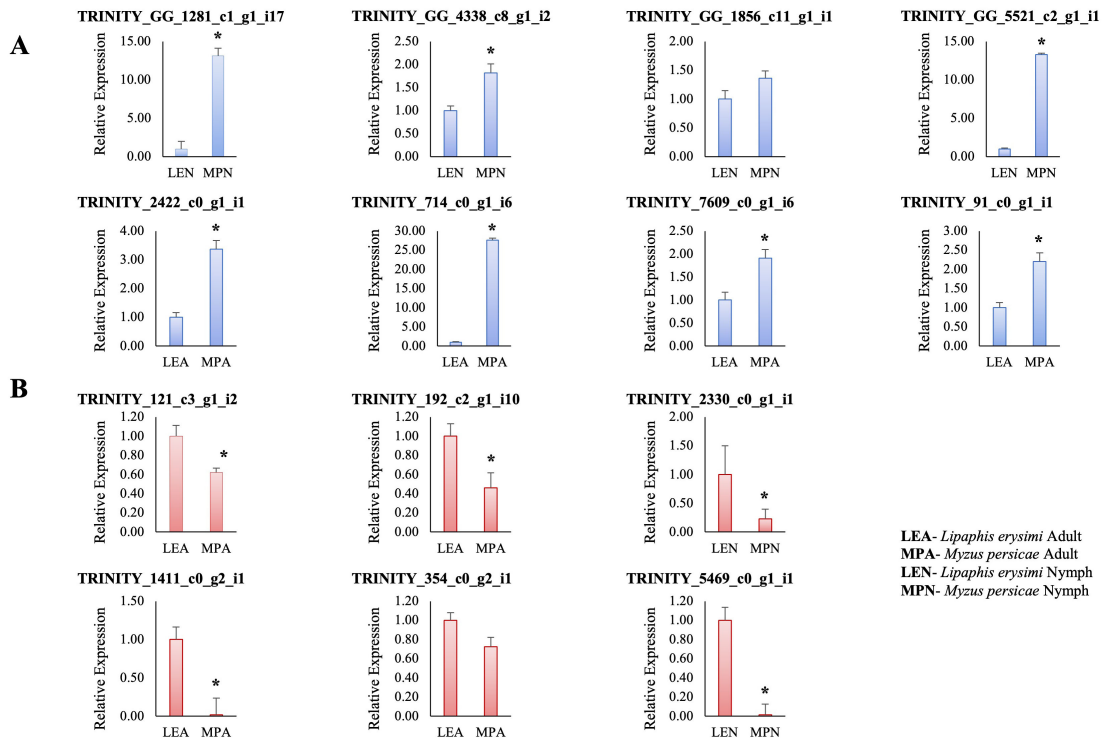


FIGURE 10

Graphs representing the relative expression of 14 differentially expressed unigenes between *Myzus persicae* adult (MPA) and *Lipaphis erysimi* adult (LEA); *Myzus persicae* nymph (MPN) and *Lipaphis erysimi* nymph (LEN). (A) represents the transcripts which are upregulated in *Myzus persicae* (downregulated in *Lipaphis erysimi*) with respect to *Lipaphis erysimi* at both adult and nymph stages in transcriptome data. Among the upregulated transcripts are TRINITY\_2422\_c0\_g1\_i1- involved in Reproduction process, TRINITY\_GG\_1281\_c1\_g1\_i17- plasma membrane calcium-transporting ATPase 2 isoform X1, TRINITY\_GG\_1856\_c11\_g1\_i1-dihydrolipoyl dehydrogenase, TRINITY\_GG\_4338\_c8\_g1\_i2- capon-like protein isoform X2, TRINITY\_DN714\_c0\_g1\_i6- uncharacterised protein. Transcripts-TRINITY\_GG\_5521\_c2\_g1\_i1- protein disulfide-isomerase A3 and TRINITY\_7609\_c1\_g1\_i1-unannotated, are upregulated putative effectors which do not show any homology with other species' effector sequence. (B) represents transcripts which are downregulated in *Myzus persicae* (upregulated in *Lipaphis erysimi*) with respect to *Lipaphis erysimi* at both adult and nymph stages in transcriptome data. Among the downregulated transcripts are, TRINITY\_DN354\_c0\_g2\_i1- repetitive proline-rich cell wall protein 2-like, TRINITY\_DN1411\_c0\_g2\_i1- cuticular protein 41 precursor, TRINITY\_DN5469\_c0\_g1\_i1- cathepsin B-like cysteine proteinase 4 isoform X2, TRINITY\_DN192\_c2\_g1\_i10- DNA-directed RNA polymerase II subunit, TRINITY\_DN2330\_c0\_g1\_i1- protein yellow. TRINITY\_DN121\_c3\_g1\_i2- unannotated, is a downregulated putative effector which does not show any homology with other species' effector sequence. Asterisks present above the bars in graphs show statistically significant differences between the two samples ( $p_{value} < 0.05$ ).

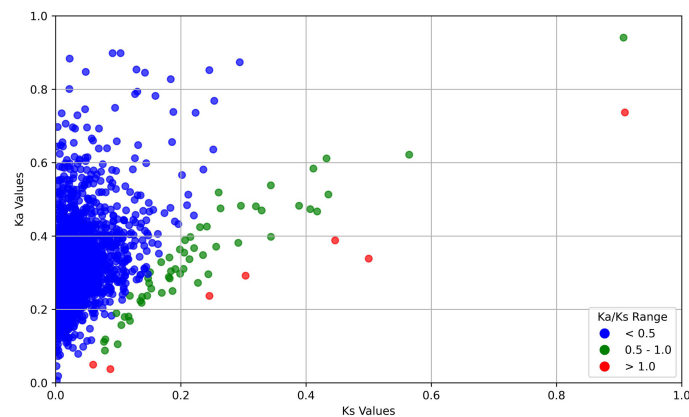


FIGURE 11

Distribution of Ka and Ks between *M. persicae* and *L. erysimi*. Sequences with Ka/Ks > 1 i.e. positive selection are displayed in red colour; sequences with Ka/Ks between 0.5-1.0 are displayed in green and sequences with Ka/Ks less than 0.5 are displayed in blue colour.

the cut-off value of 1 to 0.5 by showing evidence for evolution of genes with  $1 > K_a/K_s > 0.5$ . Additionally, many other studies used this new range for signifying positive selection. In our data, we found 67 pairs of one-to-one orthologs of *M. persicae* and *L. erysimi* having  $1 > K_a/K_s > 0.5$  and thus, these pairs of genes can also be considered as candidates undergoing positive selection. Among these 67 pairs of unigenes, we found 8 unigenes identified as putative effectors of *L. erysimi* and *M. persicae* which further might be responsible for this diversified feeding behaviour of a specialist aphid versus a generalist aphid.

## Concluding remarks

Variations in expression patterns of genes between the generalist and specialist aphids at two developmental stages revealed significant differences in genes that are involved in various stages of the infestation process. Although the specialist, *L. erysimi* had a greater number of unigenes than the generalist, differential expression studies confirmed significantly higher numbers and levels of upregulation of unigenes in various functionally important categories in the generalist aphid. Similarly, putative effectors that were identified in our study were expressed in greater numbers in the generalist than the specialist. We identified significant variations in genes involved in various stages of the infestation process *viz.*, host recognition and selection, plant cell wall degradation, detoxification of plant defence responses, digestion and metabolism between the two aphid species. Our data identifies important candidate genes that govern the robust and polyphagous nature of the generalist aphid. To the best of our knowledge, such a study on analysing variations in gene expression patterns between a generalist and specialist aphid has not been conducted till date. It would serve as an important resource for aphid biologists and facilitate further studies on aphid biology and evolution. Crucifers have a significant economic importance worldwide as a source of vegetables or oilseeds and are severely affected by infestation of both aphid species. Our data builds on the current understanding of aphid biology of these two species and also provides important candidate genes that could be used in crop protection strategies.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/SRR28520211>; <https://www.ncbi.nlm.nih.gov/SRR28520210>; <https://www.ncbi.nlm.nih.gov/SRR28520207>; <https://www.ncbi.nlm.nih.gov/SRR28520206>; <https://www.ncbi.nlm.nih.gov/SRR28520205>; <https://www.ncbi.nlm.nih.gov/SRR28520204>; <https://www.ncbi.nlm.nih.gov/SRR28520203>; <https://www.ncbi.nlm.nih.gov/SRR28520202>; <https://www.ncbi.nlm.nih.gov/SRR28520201>; <https://www.ncbi.nlm.nih.gov/SRR28520200>; <https://www.ncbi.nlm.nih.gov/SRR28520209>; <https://www.ncbi.nlm.nih.gov/SRR28520208>.

## Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

## Author contributions

MS: Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. PO: Formal analysis, Visualization, Writing – review & editing. RS: Writing – review & editing, Methodology, Formal analysis. RC: Supervision, Writing – review & editing. SG: Supervision, Writing – review & editing. MA: Supervision, Writing – review & editing. AJ: Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Institute of Eminence (IOE) - Faculty Research Program (FRP) grant of University of Delhi to AJ (IOE/2021/12/FRP). The study was partially supported by Centre of Excellence grant (BT/01/COE/08/06-II) from Department of Biotechnology, Government of India and University of Delhi - R&D grants to AJ.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1415628/full#supplementary-material>

## SUPPLEMENTARY FIGURE 1

Top 10 Gene Ontology (GO) terms of each category- Biological Process, Molecular Function and Cellular process among unigenes of (A) *M. persicae* adult (B) *L. erysimi* adult (C) *M. persicae* nymph (D) *L. erysimi* nymph. Numerals above the bars indicate the number of genes in that category in each aphid species.

## SUPPLEMENTARY FIGURE 2

Number of differentially expressed unigenes in (A) adults and (B) nymphs annotated with different databases. Numerals above the bars indicate the number of genes in that category in each aphid species.

## SUPPLEMENTARY FIGURE 3

Percent distribution of species showing homology with differentially regulated unigenes in (A) adults and (B) nymphs.

## SUPPLEMENTARY FIGURE 4

Top 15 Gene Ontology (GO) terms of each category - Biological Process, Molecular Function and Cellular Process among differentially expressed unigenes between *M. persicae* and *L. erysimi* (A) upregulated and (B) downregulated unigenes in adults and (C) upregulated and

(D) downregulated unigenes in nymphs. Red arrow in (A) indicates the category of upregulated oxidoreductases. Numerals above the bars indicate the number of genes in that category in each aphid species.

## SUPPLEMENTARY FIGURE 5

Top 15 differentially expressed enzyme categories between *M. persicae* and *L. erysimi* in (A) adults and (B) nymphs. Red arrows indicate important enzyme categories associated with detoxification and feeding. Numerals beside the bars indicate the number of genes in that category in each aphid species.

## SUPPLEMENTARY FIGURE 6

Top 15 KEGG pathways associated with differentially expressed unigenes in (A) adults and (B) nymphs of *M. persicae* and *L. erysimi*. Numerals beside the bars indicate the number of genes in that category in each aphid species.

## SUPPLEMENTARY FIGURE 7

Number of differentially regulated unigenes in different categories of digestive enzymes in *M. persicae* relative to *L. erysimi* at (A) adult (B) nymph stages. Numerals above the bars indicate the number of genes in that category in each aphid species.

## References

- Bansal, R., and Michel, A. (2018). Expansion of cytochrome P450 and cathepsin genes in the generalist herbivore brown marmorated stink bug. *BMC Genomics* 19, 1–14. doi: 10.1186/s12864-017-4281-6
- Bernays, E. A. (2001). Neural limitations in phytophagous insects: implications for diet breadth and evolution of host affiliation. *Annu. Rev. entomology* 46, 703–727. doi: 10.1146/annurev.ento.46.1.703
- Blackman, R. L., and Eastop, V. F. (2007). "Taxonomic issues," in *Aphids as Crop Pests*. Eds. H. F. Van Emden and R. Harrington (CABI, Wallingford, United Kingdom), 1–29. doi: 10.1079/9780851998190.0001
- Bos, J. I., Prince, D., Pitino, M., Maffei, M. E., Win, J., and Hogenhout, S. A. (2010). A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS Genet.* 6, e1001216. doi: 10.1371/journal.pgen.1001216
- Cai, Q. N., Han, Y., Cao, Y. Z., Hu, Y., Zhao, X., and Bi, J. L. (2009). Detoxification of gramine by the cereal aphid *Sitobion avenae*. *J. Chem. Ecol.* 35, 320–325. doi: 10.1007/s10886-009-9603-y
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Capinera, J. L. (2008). *Encyclopedia of Entomology*. 2nd edn (Dordrecht: Springer Science & Business Media), ISBN: .
- Carolan, J. C., Caragea, D., Reardon, K. T., Mutti, N. S., Dittmer, N., Pappan, K., et al. (2011). Predicted effector molecules in the salivary secretome of the pea aphid (*Acyrtosiphon pisum*): a dual transcriptomic/proteomic approach. *J. Proteome Res.* 10, 1505–1518. doi: 10.1021/pr100881q
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, pp.i884–i890. doi: 10.1093/bioinformatics/bty560
- Cooper, W. R., Dillwith, J. W., and Puterka, G. J. (2011). Comparisons of salivary proteins from five aphid (Hemiptera: Aphididae) species. *Environ. Entomology* 40, 151–156. doi: 10.1603/EN10153
- Cristofolletti, P. T., Ribeiro, A. F., Deraison, C., Rahbé, Y., and Terra, W. R. (2003). Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrtosiphon pisum*. *J. Insect Physiol.* 49, 11–24. doi: 10.1016/S0022-1910(02)00222-6
- Dixon, A. F. G. (1985). *Aphid ecology* (London: Chapman & Hall).
- Douglas, A. (2006). Phloem-sap feeding by animals: problems and solutions. *J. Exp. Bot.* 57, 747–754. doi: 10.1093/jxb/erj067
- Durak, R., Dampc, J., Kula-Maximenko, M., Mołoń, M., and Durak, T. (2021). Changes in antioxidative, oxidoreductive and detoxification enzymes during development of aphids and temperature increase. *Antioxidants* 10, 1181. doi: 10.3390/antiox10081181
- Egas, M., Sabelis, M. W., and Dieckmann, U. (2005). Evolution of specialization and ecological character displacement of herbivores along a gradient of plant quality. *Evolution* 59, 507–520. doi: 10.1111/j.0014-3820.2005.tb01011.x
- Elzinga, D. A., and Jander, G. (2013). The role of protein effectors in plant–aphid interactions. *Curr. Opin. Plant Biol.* 16, 451–456. doi: 10.1016/j.pbi.2013.06.018
- Emanuelsson, O., Nielsen, H., Brunak, S., and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 1–14. doi: 10.1186/s13059-019-1832-y
- Foissac, X., Edwards, M. G., Du, J. P., Gatehouse, A. M. R., and Gatehouse, J. A. (2002). Putative protein digestion in a sap-sucking homopteran plant pest (rice brown plant hopper; *Nilaparvata lugens*: Delphacidae)—identification of trypsin-like and cathepsin B-like proteases. *Insect Biochem. Mol. Biol.* 32, 967–978. doi: 10.1016/S0965-1748(02)00033-4
- Giordanengo, P., Brunissen, L., Rusterucci, C., Vincent, C., van Bel, A., Dinant, S., et al. (2010). Compatible plant-aphid interactions: how aphids manipulate plant responses. *Comptes Rendus. Biologies* 333, 516–523. doi: 10.1016/j.crv.2010.03.007
- Govind, G., Mittapalli, O., Griebel, T., Allmann, S., Böcker, S., and Baldwin, I. T. (2010). Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants. *PLoS One* 5, e8735. doi: 10.1371/journal.pone.0008735
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly by RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Guangxi, G., Yong, L., Jingjuan, Y., and Xiangzhen, M. (2006). Identification, activity and function determination of several salivary enzymes secreted by *Macrosiphum avenae*. *Acta Entomologica Sin.* 49, 768–774.
- Hafke, J. B., Furch, A. C., Fricker, M. D., and van Bel, A. J. (2009). Forisome dispersion in *Vicia faba* is triggered by Ca<sup>2+</sup> hotspots created by concerted action of diverse Ca<sup>2+</sup> channels in sieve element. *lant Signaling Behav.* 4, 968–972. doi: 10.4161/psb.4.10.9671
- Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G., Guilloneau, F., et al. (2008). Identification of aphid salivary proteins: a proteomic investigation of *Myzus persicae*. *Insect Mol. Biol.* 17, 165–174. doi: 10.1111/j.1365-2583.2008.00790.x
- He, H., Crabbe, M. J. C., and Ren, Z. (2023). Genome-wide identification and characterization of the chemosensory relative protein genes in *Rhus* gall aphid *Schlechtendalia chinensis*. *BMC Genomics* 24, 222. doi: 10.1186/s12864-023-09322-4
- Hilliou, F., Chertemps, T., Maibèche, M., and Le Goff, G. (2021). Resistance in the genus *Spodoptera*: Key insect detoxification genes. *Insects* 12, 544. doi: 10.3390/insects12060544
- Hogenhout, S. A., van der Hoorn, R. A., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant-Microbe Interact.* 22, 115–122. doi: 10.1094/MPMI-22-2-0115
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- Huang, Y., Hu, W., and Hou, Y. M. (2023). Host plant recognition by two odorant-binding proteins in *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae). *Pest Manage. Sci.* 79, 4521–4534. doi: 10.1002/ps.7654
- Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. J., MacLeod, R., Escudero-Martinez, C., et al. (2014). Plant immunity in plant–aphid interactions. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00663
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031



- Katis, N. I., Tsitsipis, J. A., Stevens, M., and Powell, G. (2007). *Transmission of plant viruses In H. van Emden and R* (Wallingford, UK: Aphids as crop pests CAB International).
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kehr, J. (2006). Phloem sap proteins: their identities and potential roles in the interaction between plants and phloem-feeding insects. *J. Exp. Bot.* 57, 767–774. doi: 10.1093/jxb/erj087
- Kennedy, J. S., and Booth, C. O. (1951). Host alternation in *Aphis fabae* Scop. I. Feeding preferences and fecundity in relation to the age and kind of leaves. *Ann. Appl. Biol.* 38, 25–64. doi: 10.1111/j.1744-7348.1951.tb07788.x
- Koch, K. G., Scully, E. D., Palmer, N. A., Geib, S. M., Sarath, G., Heng-Moss, T., et al. (2019). Divergent switchgrass cultivars modify cereal aphid transcriptomes. *J. Economic Entomology* 112, 1887–1901. doi: 10.1093/jee/toz053
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 1–16. doi: 10.1186/1471-2105-12-323
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8
- Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., et al. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonize diverse plant species. *Genome Biol.* 18, 1–20. doi: 10.1186/s13059-016-1145-3
- Min, X. J. (2010). Evaluation of computational methods for secreted protein prediction in different eukaryotes. *J. Proteomics Bioinf.* 3, 143–147. doi: 10.4172/jpb.1000133
- Mondal, H. A. (2017). Shaping the understanding of saliva-derived effectors towards aphid colony proliferation in host plant. *J. Plant Biol.* 60, 103–115. doi: 10.1007/s12374-016-0465-x
- Mondal, H. A. (2020). Aphid saliva: A powerful recipe for modulating host resistance towards aphid clonal propagation. *Arthropod-plant Interact.* 14, 547–558. doi: 10.1007/s11829-020-09769-2
- Mutti, N. S., Louis, J., Pappan, L. K., Pappan, K., Begum, K., Chen, M. S., et al. (2008). A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci.* 105, 9965–9969. doi: 10.1073/pnas.0708958105
- Mutti, N. S., Park, Y., Reese, J. C., and Reeck, G. R. (2006). RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J. Insect Sci.* 6, 38. doi: 10.1673/031.006.3801
- Nicolis, V. F., Burger, N. F. V., and Botha, A. M. (2022). Whole-body transcriptome mining for candidate effectors from *Diuraphis noxia*. *BMC Genomics* 23, 493. doi: 10.1186/s12864-022-08712-4
- Olazcuaga, L., Baltenweck, R., Leménager, N., Maia-Grondard, A., Claudel, P., Huguency, P., et al. (2023). Metabolic consequences of various fruit-based diets in a generalist insect species. *Elife* 12, e84370. doi: 10.7554/eLife.84370
- Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Pitino, M., Coleman, A. D., Maffei, M. E., Ridout, C. J., and Hogenhout, S. A. (2011). Silencing of aphid genes by dsRNA feeding from plants. *PLoS One* 6, e25709. doi: 10.1371/journal.pone.0025709
- Powell, G., Tosh, C. R., and Hardie, J. (2006). Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu. Rev. Entomol.* 51, 309–330. doi: 10.1146/annurev.ento.51.110104.151107
- Prajapati, V. K., Varma, M., and Vadassery, J. (2020). In silico identification of effector proteins from generalist herbivore *Spodoptera litura*. *BMC Genomics* 21, 1–16. doi: 10.1186/s12864-020-07196-4
- Pyati, P., Bandani, A. R., Fitches, E., and Gatehouse, J. A. (2011). Protein digestion in cereal aphids (*Sitobion avenae*) as a target for plant defence by endogenous proteinase inhibitors. *J. Insect Physiol.* 57, 881–891. doi: 10.1016/j.jinsphys.2011.03.024
- Ramsey, J. S., Rider, D. S., Walsh, T. K., De Vos, M., Gordon, K. H. J., Ponnala, L., et al. (2010). Comparative analysis of detoxification enzymes in *Acyrtosiphon pisum* and *Myzus persicae*. *Insect Mol. Biol.* 19, 155–164. doi: 10.1111/j.1365-2583.2009.00973.x
- Rao, S. A., Carolan, J. C., and Wilkinson, T. L. (2013). Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. *PLoS One* 8, e57413. doi: 10.1371/journal.pone.0057413
- Roy, A., Walker, III, W. B., Vogel, H., Chattington, S., Larsson, M. C., Anderson, P., et al. (2016). Diet dependent metabolic responses in three generalist insect herbivores *Spodoptera* spp. *Insect Biochem. Mol. Biol.* 71, 91–105. doi: 10.1016/j.ibmb.2016.02.006
- Sadek, R. Z., Elbanna, S. M., and Semida, F. M. (2013). Aphid-host plant interaction. *Open J. Anim. Sci.* 3, 16–27. doi: 10.4236/ojas.2013.32A003
- Salvucci, M. E., Rosell, R. C., and Brown, J. K. (1998). Uptake and metabolism of leaf proteins by the silverleaf whitefly. *Arch. Insect Biochem. Physiol.* 39, 155–165. doi: 10.1002/(SICI)1520-6327(1998)39:4%3C155::AID-ARCH3%3E3.0.CO;2-%23
- Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., et al. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 35, W169. doi: 10.1093/nar/gkm415
- Silva-Sanzana, C., Estevez, J. M., and Blanco-Herrera, F. (2020). Influence of cell wall polymers and their modifying enzymes during plant-aphid interactions. *J. Exp. Bot.* 71, 3854–3864. doi: 10.1093/jxb/erz550
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Simon, J. C., d'Alencón, E., Guy, E., Jacquín-Joly, E., Jaquiere, J., Nouhaud, P., et al. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings Funct. Genomics* 14, 413–423. doi: 10.1093/bfgp/ely015
- Swanson, W. J., Wong, A., Wolfner, M. F., and Aquadro, C. F. (2004). Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168, 1457–1465. doi: 10.1534/genetics.104.030478
- Tapia, D. H., Silva, A. X., Ballesteros, G. I., Figueroa, C. C., Niemeyer, H. M., and Ramírez, C. C. (2015). Differences in learning and memory of host plant features between specialist and generalist phytophagous insects. *Anim. Behav.* 106, 1–10. doi: 10.1016/j.anbehav.2015.04.027
- Thorpe, P., Cock, P. J., and Bos, J. (2016). Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics* 17, 1–18. doi: 10.1186/s12864-016-2496-6
- Tiffin, P., and Hahn, M. W. (2002). Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. *pekinensis*. *J. Mol. Evol.* 54, 746–753. doi: 10.1007/s00239-001-0074-1
- Tjallingii, W. F. (2006). Salivary secretions by aphids interacting with proteins of phloem wound responses. *J. Exp. Bot.* 57, 739–745. doi: 10.1093/jxb/erj088
- Wagner, G. J., Wang, E., and Shepherd, R. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* 93, 3. doi: 10.1093/aob/mch011
- Walling, L. L. (2008). Avoiding effective defenses: strategies employed by phloem-feeding insects. *Plant Physiol.* 146, 859–866. doi: 10.1104/pp.107.113142
- Wang, W., Dai, H., Zhang, Y., Chandrasekar, R., Luo, L., Hiromasa, Y., et al. (2015). Armet is an effector protein mediating aphid-plant interactions. *FASEB J.* 29, 2032–2055. doi: 10.1096/fj.14-266023
- Wang, Z., Gao, C., Liu, J., Zhou, W., and Zeng, X. (2020b). Host plant odours and their recognition by the odourant-binding proteins of *Diaphorina citri* Kuwayama (Hemiptera: Psyllidae). *Pest Manage. Sci.* 76, 2453–2464. doi: 10.1002/ps.5786
- Wang, Y., Huang, X., Chang, B. H., and Zhang, Z. (2020a). Growth performance and enzymatic response of the grasshopper, *Calliptamus abbreviatus* (Orthoptera: Acrididae), to six plant-derived compounds. *J. Insect Sci.* 20, 14. doi: 10.1093/jisesa/ieaa049
- Wang, D., Liu, Q., Jones, H. D., Bruce, T., and Xia, L. (2014). Comparative transcriptomic analyses revealed divergences of two agriculturally important aphid species. *BMC Genomics* 15, 1–14. doi: 10.1186/1471-2164-15-1023
- Weber, G. (1985). Genetic variability in host plant adaptation of the green peach aphid, *Myzus persicae*. *Entomol. Exp. Appl.* 38, 49–56. doi: 10.1111/j.1570-7458.1985.tb03497.x
- Yang, C., Pan, H., Liu, Y., and Zhou, X. (2015). Temperature and development impacts on housekeeping gene expression in cowpea aphid, *Aphis craccivora* (Hemiptera: Aphididae). *PLoS One* 10, e0130593. doi: 10.1371/journal.pone.0130593
- Zhang, Z. (2022). KaKs\_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinf.* 20, 536–540. doi: 10.1016/j.gpb.2021.12.002
- Zhang, H., Chen, J. L., Lin, J. H., Lin, J. T., and Wu, Z. Z. (2020). Odorant-binding proteins and chemosensory proteins potentially involved in host plant recognition in the Asian citrus psyllid, *Diaphorina citri*. *Pest Manage. Sci.* 76, 2609–2618. doi: 10.1002/ps.5799
- Zhang, M., Fang, T., Pu, G., Sun, X., Zhou, X., and Cai, Q. (2013). Xenobiotic metabolism of plant secondary compounds in the English grain aphid, *Sitobion avenae* (F.) (Hemiptera: Aphididae). *Pesticide Biochem. Physiol.* 107, 44–49. doi: 10.1016/j.pestbp.2013.05.002
- Zhang, H., Lin, R., Liu, Q., Lu, J., Qiao, G., and Huang, X. (2023). Transcriptomic and proteomic analyses provide insights into host adaptation of a bamboo-feeding aphid. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1098751