High throughput sequencing and annotation of *Hellula undalis* (Fabr.) (Lepidoptera: Pyralidae)

Malyaj R Prajapati[1], Pankaj Kumar[1*] and Jitender Singh[2*]

[1]College of Biotechnology, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut, Uttar Pradesh, India-250110
[2] Department of Microbiology, Chaudhary Charan Singh University, Meerut
Presenting author's email: malyajrprajapati@gmail.com

Introduction

The cabbage webworm, *Hellula undalis*, ravages brassicas and cruciferous plants worldwide, originating in Europe and causing devastating outbreaks in warm regions like India, Malaysia, and Vietnam. Its larvae bore into growing shoots and later mine leaves, encasing feeding areas in silk webs filled with waste, often leading to plant death or the production of non-marketable heads [4]. RNA sequencing (RNA-seq) has transformed transcriptome analysis, but only 17 nucleotide records exist for *Hellula undalis* in the NCBI database. This study utilized Illumina NovaSeq6000 technology to sequence and assemble the transcriptome, identifying genes related to metabolism, development, and reproduction. This comprehensive analysis offers insights into the molecular mechanisms governing the insect's lifecycle and provides a foundation for understanding gene expression and regulatory dynamics [5, 6]. The transcriptome analysis of *Hellula undalis* is pivotal, unlocking vital insights into its developmental pathways and reproductive processes. Beyond elucidating these mechanisms, this research serves as a crucial resource for identifying target genes, essential for devising sustainable strategies to control this destructive pest, potentially revolutionizing agricultural pest management.

Methodology

- ❖ Collection of *Hellula undalis* larvae.
- ❖ RNA extraction by RNA Purification Kit.
- ❖ Preparation cDNA library and Illumina sequencing for Next Generation Sequencing (NGS).
- ❖ *De novo* assembly of transcriptome (Prajapati et al., 2021)
- ❖ BLASTX alignments between unigenes and protein databases, including Nr, Swiss-Prot.
- ❖ The functional annotation by gene ontology terms will be performed using the BLAST2GO.
- ❖ The COG annotation will be performed using the BLASTX algorithm against the Cluster of Orthologous Groups database.
- ❖ The KEGG pathways annotation will be performed by sequence comparisons against the Kyoto Encyclopedia of Genes and Genomes database using BLASTX (Kanehisa et al., 2004).
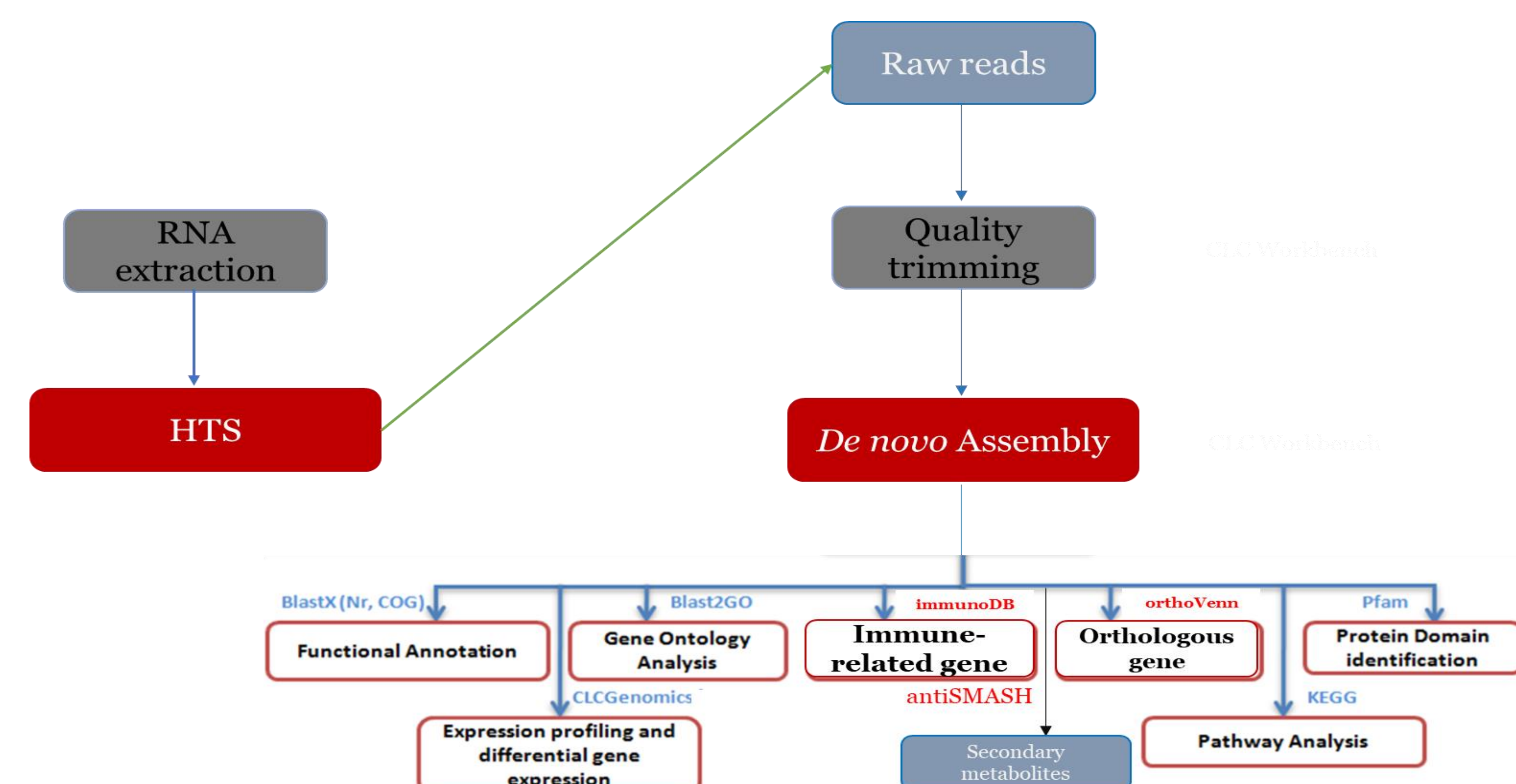
Results

❖ **High-throughput Sequencing and Assembly:**
  - ❖ Utilized Illumina NovaSeq6000 to sequence Hullula undalis cDNA, generating 48 million raw reads (2 x 150 bp).
  - ❖ After stringent filtering, approximately 97% of the raw data remained, ensuring reliability in sequence assembly.
  - ❖ Generated 30,451 contigs divided into scaffolded and non-scaffolded regions, deposited in the NCBI Sequence Read Archive (SRR24041435).

❖ **Functional Annotation:**
  - ❖ BLASTX annotation against NCBI's nr database revealed 21,806 contigs with significant similarity.
  - ❖ Predominant blast hits were observed with species like *Spodotera litura*, *Spodotera littoralis*, *Spodoptera frugiperda* and *Spodoptera exigua*.
  - ❖ Gene Ontology (GO) mapping showed prevalent contigs in biological processes (30.58%), followed by molecular functions (35.15%) and cellular components (24.93%).

❖ **Cluster of Orthologous Groups (COG) Analysis:**
  - ❖ Classified 9,646 contigs (31.67%) into 21 COG functional classes, with major proportions in "Signal transduction," "Post-translational modification," and "Transcription."

❖ **KEGG Pathway Analysis:**
  - ❖ Mapped contigs against the KEGG database, assigning 123 pathways to *Hullula undalis* transcripts, with metabolism being the most prevalent category.

❖ **Comparative Analysis of Orthologous Gene Families:**
  - ❖ Compared *Hullula undalis* with related species, identifying 5,948 common orthologous gene families and 391 unique orthologous gene families in *Hullula undalis*.

❖ **Molecular Marker Identification:**
  - ❖ Detected 1,913 potential SSRs within *Hullula undalis* contigs, highlighting 31 distinct sequence motifs with mononucleotide repeats (A and T) being most abundant.

These findings lay a comprehensive foundation for understanding *Hullula undalis* molecular landscape, offering insights into gene functions, pathways, and evolutionary relationships. Additionally, the identification of molecular markers serves as a valuable genomic resource for genetic enhancement and pest management strategies.

Conclusion

- ❖ **Comprehensive Transcriptomic Insight:** This study provides a detailed understanding of *Hellula undalis* transcriptome, unraveling genetic complexities and functional elements.
- ❖ **Annotation Enriching Functional Understanding:** Annotation through NR, GO, COG, and KEGG databases elucidates key biological processes, molecular functions, and pathways.
- ❖ **Insights into Evolution and Divergence:** Comparative analysis with related species unveils shared and unique genetic components, shedding light on evolutionary relationships.
- ❖ **Foundation for Future Research:** The transcriptomic elucidation forms a robust basis for further investigations into *Hellula undalis* biology, genetics, and functionality.
- ❖ **Pest Management Implications:** The identified molecular markers and genomic insights pave the way for innovative strategies in pest control and agricultural management.

Acknowledgment

References

- Kunjwal, N.; Srivastava, R. M. Insect Pests of Vegetables. Pests Manag. 2018, 163–221.
- Prajapati, M. R.; Singh, J.; Kumar, P.; Dixit, R. De Novo Transcriptome Analysis and Identification of Defensive Genes in Garlic (*Allium sativum* L.) Using High-Throughput Sequencing. J. Genet. Eng. Biotechnol. 2023, 21 (1), 56. DOI: 10.1186/s43141-023-00499-5
- Duan, X.; Wang, K.; Su, S.; Tian, R.; Li, Y.; Chen, M. De Novo Transcriptome Analysis and Microsatellite Marker Devel-opment for Population Genetic Study of a Serious Insect Pest, Rhopalosiphum Padi (L.)(Hemiptera: Aphididae). PLOS ONE 2017, 12 (2), e0172513. DOI: 10.1371/journal.pone.0172513