G3, 2026, 16(3), jkag003
https://doi.org/10.1093/g3journal/jkag003
Advance Access Publication Date: 13 January 2026
Genome Report

A draft genome assembly of the agricultural pest *Leucoptera coffeella* and analysis of its dsRNA processing machinery is a key step toward RNAi-based biopesticides in Lepidoptera

Jay K. Goldberg [1,2,*] Leonardo A. Vidal,[3] Erick S. L. Queiroz,[4] Eliza F. M. B. Nascimento,[3,5] Marcos J. A. Viana,[6] Wellington R. Clarindo,[5] Andrea Q. Maranhao,[4] Natália F. Martins [6] Érika V. S. Albuquerque [3,*]

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287, United States
[2]Department of Crop Genetics, John Innes Centre, Norwich, Norfolk NR4 1UD, United Kingdom
[3]Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF 70770-917, Brasil
[4]Departamento de Biologia Molecular, Universidade de Brasília, Brasília, DF 70910-900, Brazil
[5]Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil
[6]Embrapa Agroindústria Tropical, Rua Dra Sara Mesquita 2270, Planalto do Pici, Fortaleza, CE 60511-110, Brazil

*Corresponding authors: Jay K. Goldberg, School of Life Sciences, Arizona State University, Tempe, AZ 85287, United States. Email: jay.goldberg@jic.ac.uk; jay.goldberg@asu.edu; Érika V. S. Albuquerque, Brasil Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF 70770-917, Brazil. Email: erika.albuquerque@embrapa.br

The Coffee Leaf Miner (Lepidoptera: Lyonetiidae: *Leucoptera coffeella*) is a specialist herbivore and major global pest of coffee plants. Current pest control strategies primarily rely on chemical pesticides which in turn negatively impact both human health and ecological stability. Additionally, the emergence of insecticide-resistant populations underscores the urgent need for more specific and efficient pest management strategies. The development of novel techniques for controlling this insect pest requires rigorous interrogation of its physiology and interactions with host plants at a molecular/genetic level. To enable future research in this vein, we sequenced and assembled a draft *L. coffeella* genome using PacBio highly accurate long-reads (HiFi). Our assembly is comprised of 1,615 contigs showing fragmentation, yet the majority of gene content is represented (BUSCO complete = 91.7%). We annotated 17,467 protein-coding genes within our assembly, seven of which are core components of the small interfering RNA machinery. The expression of these genes was further confirmed via qPCR. This analysis—and the underlying genomic data—highlights potential targets for RNAi-based biopesticide development and will serve as the foundation for important future research aimed at protecting global coffee production from one of its most destructive pests.

Keywords: coffee; insect; leaf miner; pest control; interfering RNA; genome assembly

Introduction

Lepidopteran insects, which include butterflies and moths, are estimated to represent more than 70% of agricultural pests. Current control practices rely on the application of chemical pesticides to the detriment of off-target species and, in some cases, human health (Pathak et al. 2022), causing global biodiversity losses (Raven and Wagner 2021) and triggering insecticide resistance in important crop pests (Barathi et al. 2024; Willow and Smagghe 2025).

The Coffee Leaf Miner (CLM; Lepidoptera: Lyonetiidae: *Leucoptera coffeella*; Fig. 1) is a major pest of one of the most traded crops in the world. This monophagous Lepidopteran is currently a cosmopolitan pest present in all coffee producing countries, causing up to 87% losses on both *Coffea arabica* and *Coffea canephora* plantations (Picanço Filho et al. 2024). Leaf damage is caused by larval feeding behavior, which consists of burrowing mines into the mesophyll (Fig. 1), leading to necrosis and consequently reducing photosynthesis, accelerating senescence, and causing defoliation that weaken plant growth and reduces yield (Dantas et al. 2021). CLM reproduction is favored by high temperatures and dry climatic conditions, leading to worsening infestations that are further exacerbated by anthropogenic climate change (Leite et al. 2020).

RNA interference (RNAi) and post-transcriptional gene silencing continue to offer promising solutions for developing alternative, species-specific biopesticides (Ortolá and Daròs 2024). RNAi-based technologies, which are centered around the application of double-stranded RNA (dsRNA) constructs to silence genes in target species (Niu et al. 2024), include low-risk biopesticides that reduce pest loads on crops while avoiding unwanted environmental consequences (Willow and Smagghe 2025). RNAi constructs can be directly engineered into host plants themselves or delivered by non-transgenic spray-induced methods that may avoid the regulatory complications surrounding genetic modification (Chen et al. 2025). In order to design effective RNAi-based pesticides, we must

Received on 28 July 2025; accepted on 30 November 2025

Fig. 1. Symptoms and developmental stages of *Leucoptera coffeella* in coffee plants. The pictures show: (a) Coffee plant (*C. arabica*) heavily infested by the CLM, showing chlorosis and defoliation; (b) Leaf damage caused by larval feeding, with characteristic necrotic mines along the leaf blade; (c) Larva L2 isolated from leaf tissue after removal from mine; (d) Pupa within its silk protective structure adhered to the abaxial leaf surface; (e) Adult moth of *L. coffeella*, showing characteristic moth morphology and microlepidoptera size. Scale bars = 1 mm.

enhance our fundamental understanding of these mechanisms in insect species of interest; a process that begins with the generation of reference genome resources (Schoville et al. 2018; Sparks et al. 2020).

To enable molecular biological studies of CLM and the development of targeted RNAi-based pest control solutions, we sequenced and assembled a draft genome for *L. coffeella*. Due to the small physical size of this insect, we used a pooled-sample of multiple individuals that we sequenced with PacBio highly accurate long-reads (HiFi). To improve knowledge about the RNAi machinery genes in Lepidopteran pests, we identified in our assembly multiple genes that are involved with exogenous dsRNA processing pathway and validated their active expression across development stages. Additionally, BUSCO categories allow the search of species-specific biopesticide targets. Our analysis, and the underlying genomic and transcriptomic datasets, will serve as a firm foundation for future research—fundamental and applied—on CLM and its interactions with host plants.

Materials and methods

Sample collection, preparation, and sequencing

Genomic DNA (gDNA) was obtained from a pool of individuals at the pupal stage (Fig. 1), collected from *C. arabica* leaves originating from Embrapa Cerrados, Planaltina, D.F., Brazil. A modified protocol of the EZNA insect DNA kit (Omega BioTek, Cat. No.: D0926-02) was used for the extractions, as described (Nascimento et al. 2022). DNA quantity and quality were analyzed using NanoDrop, Qubit, agarose gel electrophoresis, and Femto Pulse. High molecular weight gDNA was stored at 4 °C until sequencing began. Genomic sequencing was carried out by Precision Genomics (Gangseo-gu, Seoul, South Korea). The samples were prepared using a PacBio HiFi Express Prep kit and sequenced using the Sequel II platform.

Total RNA was extracted from CLM at different developmental stages: three larval instars (L2, L3, and L4), pupae, male and female adults. Extracted RNA was cleaned with a ReliaPrep RNA Tissue Miniprep System kit (Promega, Cat. No.: PRZ6111), applying

the modified protocol described (Nascimento et al. 2022). After extraction, the concentration and quality of the RNA were evaluated using NanoDrop, Qubit, and Bioanalyzer, and the sample was stored immediately at −80 °C until sequencing using Illumina technology Novaseq 6000.

Genome size estimate via flow cytometry

Adult insects (male and female) were collected from two populations in Brazil (Viçosa-MG and Barreiras-BA) and dissected in commercial saline solution. Nuclear extraction and isolation were processed in a BD Accuri™ C6 Flow Cytometer (Accuri, Belgium) and the flow cytometry histograms were analyzed considering G0/G1 fluorescent peaks with coefficient of variation below 5% for nuclear 2C value measurement. Mean nuclear 2C values were converted to Mbp, considering that 1C pg is equivalent to 978 Mbp (Praça-Fontes et al. 2011).

Genome assembly and annotation

Circular consensus sequencing (CCS) output (i.e. HiFi reads; $N_{reads} = 1{,}712{,}831$; N50 = 14,954) were found to have a highly inflated number of low-coverage contigs (Supplementary Fig. 1) which could be due to genetic heterogeneity or the presence of contamination. To identify contaminant reads, we first produced a metagenomic assembly of our data using metaMDBG (Benoit et al. 2024). We then used the blobtoolkit pipeline (Challis et al. 2020), which calls on minimap2 (Li 2018) and blastn (Camacho et al. 2009), to identify contaminant contigs and filter our PacBio HiFi reads to remove contigs identified as anything other than Lepidoptera ($N = 63{,}334$). Kmer analysis of both raw and filtered reads was conducted using jellyfish (Marçais and Kingsford 2011) and GenomeScope2.0 (Ranallo-Benavidez et al. 2020). Filtered reads ($N = 1{,}649{,}497$) were then assembled using hifiasm 2.4.0 (Cheng et al. 2021) with the self-scaffolding function (–dual-scaf) turned on. We assessed the completeness of this assembly using BUSCO v5.4.7 within the lepidoptera_odb10 dataset and determined that our genome was heavily duplicated due to the presence of alternative haplotigs resulting from our pooled sample of multiple insects. To remove duplicated regions, we ran the purge_dups algorithm v1.2.6 (Guan et al. 2020) five times with the -e option removed to include both duplications at the ends and interiors of contigs. Genome statistics were obtained from Bandage v0.8.1 (Wick et al. 2015) after each round of purge_dups. Our purged assembly was then polished with Inspector v1.3.1 (Chen et al. 2021), an error-correction module developed for long-read genome assemblies. Structural annotation was performed using Helixer v0.3.4 using the pre-trained invertebrate model (Stiehler et al. 2021; Holst et al. 2025). The completeness of the annotation was determined using BUSCO v.5.4.7 (Seppey et al. 2019) in protein mode and by aligning RNA-seq reads of multiple life stages (larvae, pupae, and adults; 1× sample per stage) to the genome using STAR v2.7 (Dobin et al. 2013) on the default settings. Functional annotation of the gene models was done using the eggnog-mapper (v2) web portal (eggnog-mapper.embl.de; Cantalapiedra et al. 2021). Repetitive element content of our genome was assessed using RepeatModeler and RepeatMasker (Tarailo-Graovac and Chen 2009).

Structural modeling and qPCR

Gene models that eggnog-mapper identified as RNAi pathway components in *L. coffeella* data were submitted to structural predictions using InterproScan and AlphaFold3 to confirm both gene annotations and structural-functional conservation (Blum et al. 2021; Abramson et al. 2024). To obtain the multiple sequence alignments (MSAs), the selected RNAi genome represented were used as queries in BLASTP searches against the reference protein sequences database at the National Center for Biotechnology Information (NCBI) and organisms *Manduca sexta*, *Plutella xylostella*, and *Drosophila melanogaster*. Homologous sequences from related insect species alignments obtained by COBALT. For the structural alignment, we submitted candidate sequences and their homologues to AlphaFold2.0 server to provide a basis for comparison and validation, experimentally solved structures of homologous RNAi proteins from other organisms (e.g. from *Drosophila melanogaster* or other *Lepidoptera*) were downloaded from the Protein Data Bank (PDB). Both the predicted models and the experimental reference structures were imported into Chimera. This structural superposition allowed for the visualization and analysis of conserved structural cores and functional domains critical for the RNAi machinery, assessing the quality of the predictions and revealing evolutionarily maintained architectural features.

Primers for quantitative real-time PCR (RT-qPCR) reactions (Supplementary Table 1) were synthesized by Sigma® (*Rpl10*, *Rpl18*, *Sid1*, and *C3pO*) or Exxtend® (*Ago1*, *Ago2*, *Dcr1*, *Dcr2*, and *R2d2*) and used in amplification reactions performed in a QuantStudio 3® system (Applied Biosystems, Waltham, MA, USA). Amplification data were analyzed using LinRegPCR, removing potential biases from the sample data (Ramakers et al. 2003). Fold Change (FC) of the relative gene expression was quantified using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). The ΔC_T values of pupa samples were compared with larva/male/female samples from three biological experiments. Statistical significance was calculated with two-way ANOVA, followed by Tukey's post-hoc test (P-value < 0.01).

Results and discussion

Genome assembly and annotation

Blobtools analysis of our preliminary meta-genomic assembly identified only a handful of contigs ($N = 21$) belonging to phyla other than Arthopoda (Supplementary Fig. 2). Of those identified as Arthropoda, 600 were found to correspond to classes other than Lepidoptera (Supplementary Figs 2 and 3). Removing all non-Lepidoptera reads did not meaningfully reduce the low-coverage kmers in our dataset (Supplementary Fig. 4), suggesting that they were the result of pooling multiple genetically heterogeneous individuals.

Assembling only the Lepidoptera-identified reads with hifiasm produced an assembly that was free of contamination (Supplementary Fig. 5) yet highly duplicated (>90% of odb10_lepidoptera BUSCOs; Fig. 2). Five iterations of purge_dups were able to reduce the duplicated BUSCO rate to 10.5% (Fig. 2 and Supplementary Fig. 6), consistent with previous studies that found it suitable for reducing spurious duplication in pooled-sample assemblies (Goldberg et al. 2024); however, this final duplication rate is still much higher than expected indicating that some haplotigs remain in our final assembly. This is further suggested by the presence of some low coverage contigs (<10×; Supplementary Fig. 5).

Our final assembly remained highly fragmented by the standards of modern long-read sequencing (number of contigs = 1615; largest contig = 6.3Mb; total size = 370.6Mb; Table 1, Supplementary Table 2) and slightly smaller than our kmer estimate (399.7 Mb; Supplementary Fig. 4); however, it is important to note that our estimate is likely inflated by the presence of multiple haplotypes in our dataset and the true genome size is much smaller than our total assembly size. Indeed, the size of our genome assembly is

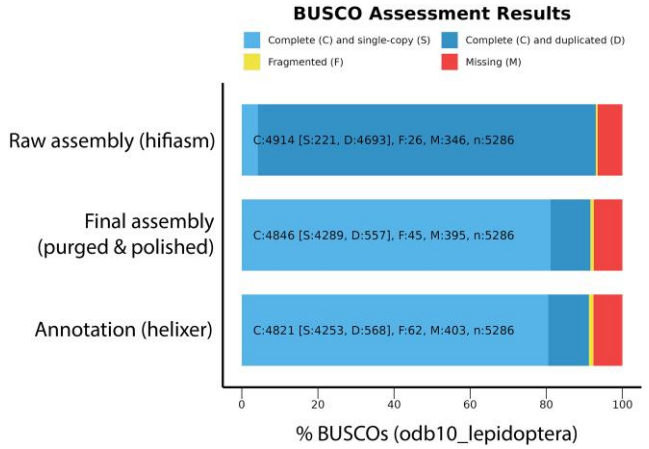

Fig. 2. Results of BUSCO analysis for the raw assembly produced by hifiasm (top bar), the final assembly after 5× rounds of purge_dups and polishing with inspector (middle bar), and the structural annotation produced by helixer (bottom row).

Table 1. Statistics describing our final (purged & polished) assembly.

Feature	Value
Assembly size	370,558,664
Number of contigs	1615
Largest contig	6,298,745
Contig N50	544,430
Repeat content (%)	46.44
GC content (%)	36.07
Assembly BUSCO complete (odb10_lepidoptera)	91.6
Annotation BUSCO complete (odb10_lepidoptera)	91.2
Number of genes (helixer)	17,467
Functionally annotated genes (eggnog-mapper)	13,725

larger than our flow cytometry measurement (294.6225 Mbp; mean nuclear value = 2C = 0.603pg) consistent with the presence of haplotypic duplication. Both our flow cytometry measurement and the size of our assembly are well within the range expected based on other lepidopteran genomes (e.g. *Manduca sexta* = 470 Mb, Gershman et al. 2021; *Plutella xylostella* = 323 Mb, Boyes et al. 2023; *Pieris rapae* = 246 Mb, Shen et al. 2016).

Inspector, which assesses genome quality and polishes errors by mapping raw reads to the assembly, found that our assembly was error-prone ($N_{\text{small-scale errors}}$ = 2.03M; 5,482 per Mb; Supplementary Table 1), which is typical of pooled-sample assemblies (Goldberg et al. 2024), yet the polishing algorithm was able to significantly reduce this ($N_{\text{small-scale errors}}$ = 190k; 512 per Mb). Despite these errors, BUSCO analysis found that most gene content was still represented (BUSCO complete = 91.6%; Table 1). This was further supported by high mapping rates for our six RNA-seq samples (mean mapping rate = 86%; Supplementary Table 3). RepeatMasker found that 46.44% of our assembly was composed of repetitive elements (Supplementary Table 4), predominantly retroelements (20.91% of total length) and unclassified elements (18.14% of total length); however, our aggressive purging of duplicate haplotigs is likely to have collapsed repetitive regions of the genome, as evidenced by the presence of some very high coverage contigs (>100×; Supplementary Fig. 5), thus the real repeat content of the *L. coffeella* genome may be higher.

Helixer was able to annotate 17,467 genes in our assembly, with 91.2% of protein BUSCOs complete; 13,725 of these genes were able to be functionally annotated. This speaks to the ability of helixer to identify lepidopteran genes, and is likely due to the overrepresentation of holometabola (especially lepidopteran and dipteran) genomes within its training corpus (details regarding Helixer training data sets can be found at Supplemental Figure S20 in Holst et al. 2025).

Identification and analysis of RNAi machinery proteins

Eggnog-mapper revealed several putative gene sequences associated with small RNA processing and transport pathways in the CLM genome (Supplementary Fig. 8): Double-stranded RNA-specific endoribonuclease (*Dcr*), Argonaute RISC Component 1 (*Arg*), Systemic RNAi-deficient-1, RNAi regulator-component 3 promoter of RISC (*C3po*), and dsRNA-binding protein R2D2 (*R2d2*). Additionally, we performed homology-based molecular modeling of selected RNAi core protein candidates to validate genomic predictions by elucidating conserved structural and functional features. Our results showed that the predicted proteins LcDCR1, LcDCR2, LcAGO1, LcAGO2, SID1, C3PO, and R2D2 (Fig. 3 and Supplementary Fig. 8) retain not only sequence homology but also their expected conformational function (Davis-Vogel et al. 2018; Arraes et al. 2021). Moreover, in order to assess the fit of the predicted genes to their models, we compared the *L. coffeella* sequences and structures to other lepidopteran insects, being *Plutella xylostella* the closest well studied relative within the same superfamily (Yponomeutoidea), and the more distantly related *Manduca sexta* (Bombycoidea) *and Spodoptera frugiperda* (Glossata), besides *Drosophila melanogaster* as an outgroup. We observed two distinct patterns on structural superposition: (i) highly conserved at the critical catalytic and functional domains, pinpointed at SID1 (Supplementary Fig. 9a and b) and DCR (Supplementary Fig. 9c and d); and (ii) variables regions at the flexible structural segments that potentially reflect recent evolutionary pressures exampled by LOQ (Supplementary Fig. 9e and f) and C3PO (Supplementary Fig. 9k and l). Structural analysis revealed that SID1, DCR, and AGO (Supplementary Fig. 9i and j) as the most conserved proteins across species. In contrast, HEN1 (Supplementary Fig. 9m and n) and R2D2 (Supplementary Fig. 9g and h) showed moderate conservation, while C3PO and LOQ were the least similar.

Although RNAi is a conserved mechanism in eukaryotes, variations in gene duplication, deletion, and expression across insect species affect its efficiency (Cooper et al. 2019). RNAi efficacy is influenced more by the expression levels of key enzymes—such as Dicer, Argonaute, and RISC components—than by the sheer number of core genes. The reported high variability in RNAi responses among insects suggests that analyzing the expression of RNAi-related genes could enhance our understanding of interspecies differences in RNAi efficiency and help identify key factors influencing silencing success (Koo et al. 2024).

The expression patterns of larvae, male and female compared to the pupal stage are depicted in Supplementary Fig. 7. Our qPCR results show that even candidate genes coding for closely related proteins (i.e. DCR2 and AGO2) presented different melting temperature of the amplicon and expression profiles, suggesting that the mined sequences were distinct enough to be assigned to *L. coffeella* as *LcDcr1*, *LcDcr2*, *LcAgo1*, and *LcAgo2*.

There are several fascinating findings within our analysis of dsRNA processing in *L. coffeella*. Firstly, we detected R2D2—a dsRNA-binding protein which forms a complex with Dicer to

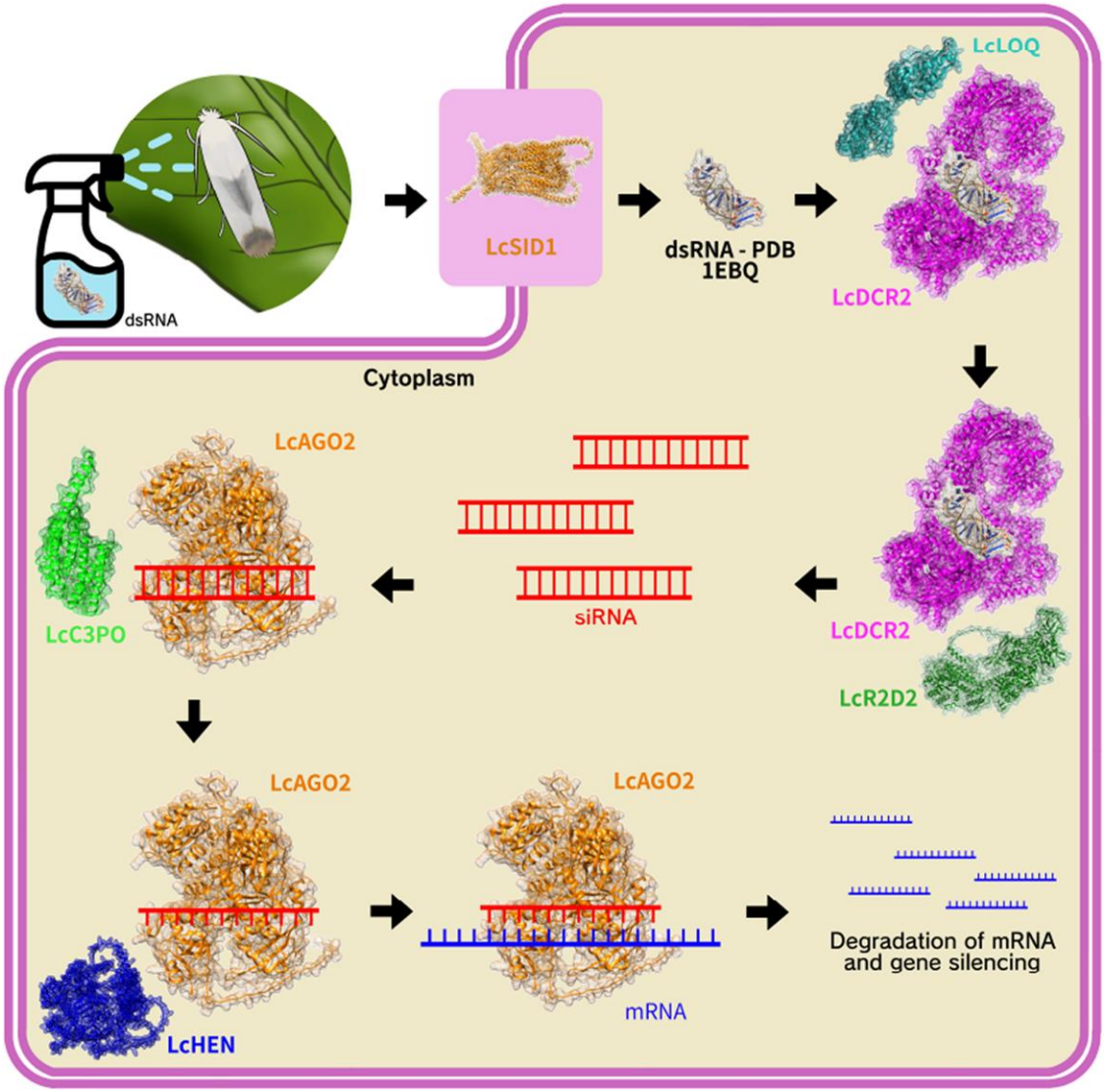

Fig. 3. Graphical representation of the main genes in the exogenous dsRNA processing pathway involved in siRNA formation for gene silencing. AlphaFold3-predicted structures of key RNAi proteins found in *Leucoptera coffeella*: LcSID1 (ocher), LcDCR2 (magenta), LcLOQ (turquoise), LcR2D2 (dark green) LcAGO2 (orange), LcC3PO (light green), LcHEN (dark blue) all modeled from the *L. coffeella* genome. General representation of substrate molecules: dsRNA—PDB 1EBQ (multicolored), siRNA (red), mRNA (blue).

process dsRNA into siRNA thereby initiating RNAi through the loading of siRNA duplexes onto ARG proteins (Hameed et al. 2024). *R2d2* has not been detected in several lepidopteran transcriptomes, suggesting a possible loss or reduced function of this gene in the order. In *Bombyx mori*, the R2D2 homolog is expressed at very low levels (Swevers et al. 2011). Interestingly, our study revealed *LcR2d2* has a higher expression in males compared to larva, pupa and female. Although R2D2's role in the RNAi pathway is well established, its involvement in male-specific traits or processes remains largely uncharacterized. There is currently no information about the implication of R2D2 in male-specific development or reproduction. Second, *LcSid1*—a transmembrane protein involved in double-stranded RNA uptake (Saakre et al. 2024), exhibited an overexpression peak in adult males, with mean fold change values significantly higher than those observed in immature stages and females. Conversely, expression analyses in *Ostrinia nubilalis* showed that *R2d2* transcripts were present across all developmental stages, with expression levels significantly lower in males than in females (Cooper et al. 2021). Also, in contrast to *L. coffeella*, *Sid-1* expression in *Nilaparvata lugens* was lower in males compared to females (Zha et al. 2011). Altogether, these results suggest that RNAi pathways may be involved in sex-specific regulatory dynamics that warrant further investigation, especially when determining possible targets for novel biopesticides.

Concluding remarks

In light of rapid environmental changes and biodiversity loss, especially within the highly diverse tropics, the need to protect agriculture with novel pest management solutions has never been greater. RNAi-based biopesticides are a promising solution but require a substantial amount of research into the biology of target species to develop. By generating a draft genome assembly with PacBio long-reads, we have taken a crucial first step toward the goal of protecting coffee production from its most damaging pest, the CLM. Furthermore, we characterized the structures and expression of key RNAi pathway genes within our genome

and confirmed the presence of RNAi machinery within *L. coffeella*. This establishes targeted biopesticides as a solution worthy of continued research and development. Notably, the serendipitous discovery of sex-biased expression in these genes points to previously undiscovered roles for RNAi in lepidopteran biology. As such, our study—and the underlying data—will serve as a valuable resource for future research, both fundamental and applied.

Data availability

The genome assembly and raw reads (PacBio HiFi & Illumina RNA-seq) underlying this article are available at NCBI with the BioProject ID PRJNA832598. Genome annotation files and code associated with this project are available at https://github.com/caterpillar-coevolution/Lcoffeella_draft.

Supplemental material available at G3 online.

Acknowledgments

We thank Adriano Veiga from Embrapa for the accession to the field infested plants; Genomix Data, for contracting sequencing facilities. We also thank the University of Arizona HPC team for maintaining computational resources and Professor Judith Bronstein for providing access to J.K.G.

Funding

This work was supported by grants from the Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (20/2018-159) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (440352/2022-3).

Conflicts of interest

None declared.

Author contributions

Conceptualization—J.K.G., N.F.M., E.V.S.A.; Data curation—J.K.G., N.F.M., M.J.A.V., E.V.S.A.; Formal analysis—J.K.G., L.A.V., N.F.M., M.J.A.V.; Funding acquisition—J.K.G., E.V.S.A; Investigation—E.F.M.B.N., L.A.V.; Software—J.K.G., N.F.M., M.J.A.V.; Project administration—E.V.S.A.; Validation—J.K.G., N.F.M. L.A.V.; Visualization—J.K.G., N.F.M., L.A.V., E.S.L.Q.; Resources—E.V.S.A.; Supervision—N.F.M., W.R.C., E.V.S.A.; Writing—original draft—J.K.G., N.F.M., W.R.C., E.V.S.A.; Writing—review and editing—J.K.G., N.F.M., W.R.C., A.Q.M., E.V.S.A.

Literature cited

Abramson J *et al.* 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 630:493–500. https://doi.org/10.1038/s41586-024-07487-w.

Arraes FBM *et al.* 2021. Dissecting protein domain variability in the core RNA interference machinery of five insect orders. RNA Biol. 18:1653–1681. https://doi.org/10.1080/15476286.2020.1861816.

Barathi S, Sabapathi N, Kandasamy S, Lee J. 2024. Present status of insecticide impacts and eco-friendly approaches for remediation-a review. Environ Res. 240:117432. https://doi.org/10.1016/j.envres.2023.117432.

Benoit G *et al.* 2024. High-quality metagenome assembly from long accurate reads with metaMDBG. Nat Biotechnol. 42:1378–1383. https://doi.org/10.1038/s41587-023-01983-6.

Blum M *et al.* 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49:D344–D354. https://doi.org/10.1093/nar/gkaa977.

Camacho C *et al.* 2009. BLAST+: architecture and applications. BMC Bioinform. 10:1–9. https://doi.org/10.1186/1471-2105-10-421.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38:5825–5829. https://doi.org/10.1093/molbev/msab293.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. G3 Genes|Genomes|Genetics. 10:1361–1374. https://doi.org/10.1534/g3.119.400908.

Chen P *et al.* 2025. RNA interference-based dsRNA application confers prolonged protection against rice blast and viral diseases, offering a scalable solution for enhanced crop disease management. J Integr Plant Biol. 67:1633–1648. https://doi.org/10.1111/jipb.13896.

Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. 2021. Accurate long-read de novo assembly evaluation with Inspector. Genome Biol. 22:1–21. https://doi.org/10.1186/s13059-021-02527-4.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cooper AM, Silver K, Zhang J, Park Y, Zhu KY. 2019. Molecular mechanisms influencing efficiency of RNA interference in insects. Pest Manag Sci. 75:18–28. https://doi.org/10.1002/ps.5126.

Cooper AMW *et al.* 2021. Characterization, expression patterns, and transcriptional responses of three core RNA interference pathway genes from Ostrinia nubilalis. J Insect Physiol. 129:104181. https://doi.org/10.1016/j.jinsphys.2020.104181.

Dantas J *et al.* 2021. A comprehensive review of the coffee leaf miner Leucoptera coffeella (Lepidoptera: Lyonetiidae)—a major pest for the coffee crop in Brazil and others Neotropical countries. Insects. 12:1130. https://doi.org/10.3390/insects12121130.

Davis-Vogel C *et al.* 2018. Identification and comparison of key RNA interference machinery from western corn rootworm, fall armyworm, and southern green stink bug. PLoS One. 13:e0203160. https://doi.org/10.1371/journal.pone.0203160.

Dobin A *et al.* 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Goldberg JK, Allan CW, Copetti D, Matzkin LM, Bronstein J. 2024. A pooled-sample draft genome assembly provides insights into host plant-specific transcriptional responses of a Solanaceae-specializing pest, Tupiocoris notatus (Hemiptera: Miridae). Ecol Evol. 14:e10979. https://doi.org/10.1002/ece3.10979.

Guan D *et al.* 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 36:2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

Hameed MS, Ren Y, Tuda M, Basit A, Urooj N. 2024. Role of Argonaute proteins in RNAi pathway in plutella xylostella: a review. Gene. 903:148195. https://doi.org/10.1016/j.gene.2024.148195.

Holst F *et al.* 2025. Helixer: ab initio prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. *Nat Methods*. 1–8. https://doi.org/10.1038/s41592-025-02939-1.

Koo J, Zhu G-H, Palli SR. 2024. CRISPR-Cas9 mediated dsRNase knockout improves RNAi efficiency in the fall armyworm. Pestic

Biochem Physiol. 200:105839. https://doi.org/10.1016/j.pestbp.2024.105839.
Leite SA *et al.* 2020. Profile of coffee crops and management of the neotropical coffee leaf miner, Leucoptera coffeella. Sustainability. 12:8011. https://doi.org/10.3390/su12198011.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 25:402–408. https://doi.org/10.1006/meth.2001.1262.
Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27:764–770. https://doi.org/10.1093/bioinformatics/btr011.
Nascimento, EFMB *et al.* 2025. Optimization of genomic DNA and total RNA extraction protocols from microlepidopterans for high-fidelity long-read sequencing. BMC Methods. 2:19. https://doi.org/10.1186/s44330-025-00038-3.
Niu J, Chen R, Wang J-J. 2024. RNA interference in insects: the link between antiviral defense and pest control. Insect Sci. 31:2–12. https://doi.org/10.1111/1744-7917.13208.
Ortolá B, Daròs J-A. 2024. RNA interference in insects: from a natural mechanism of gene expression regulation to a biotechnological crop protection promise. Biology (Basel). 13:137. https://doi.org/10.3390/biology13030137.
Pathak VM *et al.* 2022. Current status of pesticide effects on environment, human health and it's eco-friendly management as bioremediation: a comprehensive review. Front Microbiol. 13:962619. https://doi.org/10.3389/fmicb.2022.962619.
Picanço Filho MC *et al.* 2024. Economic injury levels and economic thresholds for Leucoptera coffeella as a function of insecticide application technology in organic and conventional coffee (Coffea arabica), farms. Plants. 13:585. https://doi.org/10.3390/plants13050585.
Praça-Fontes MM, Carvalho CR, Clarindo WR. 2011. C-value reassessment of plant standards: an image cytometry approach. Plant Cell Rep. 30:2303–2312. https://doi.org/10.1007/s00299-011-1135-6.
Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett. 339:62–66. https://doi.org/10.1016/S0304-3940(02)01423-4.
Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11:1432. https://doi.org/10.1038/s41467-020-14998-3.
Raven PH, Wagner DL. 2021. Agricultural intensification and climate change are rapidly decreasing insect biodiversity. Proc Natl Acad Sci U S A. 118:e2002548117. https://doi.org/10.1073/pnas.2002548117.
Saakre M *et al.* 2024. Host-delivered RNA interference for durable pest resistance in plants: advanced methods, challenges, and applications. Mol Biotechnol. 66:1786–1805. https://doi.org/10.1007/s12033-023-00833-9.
Schoville SD *et al.* 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). Sci Rep. 8:1931. https://doi.org/10.1038/s41598-018-20154-1.
Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editors. Gene prediction: methods and protocols, methods in molecular biology. Springer. p. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
Sparks ME *et al.* 2020. Brown marmorated stink bug, Halyomorpha halys (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. BMC Genomics. 21:227. https://doi.org/10.1186/s12864-020-6510-7.
Stiehler F *et al.* 2021. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics. 36:5291–5298. https://doi.org/10.1093/bioinformatics/btaa1044.
Swevers L, Liu J, Huvenne H, Smagghe G. 2011. Search for limiting factors in the RNAi pathway in silkmoth tissues and the Bm5 cell line: the RNA-binding proteins R2D2 and translin. PLoS One. 6:e20250. https://doi.org/10.1371/journal.pone.0020250.
Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 25:4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.
Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
Willow J, Smagghe G. 2025. RNAi applications toward environmentally sustainable food security. Curr Opin Environ Sci Health. 45:100612. https://doi.org/10.1016/j.coesh.2025.100612.
Zha W *et al.* 2011. Knockdown of midgut genes by dsRNA-transgenic plant-mediated RNA interference in the hemipteran insect Nilaparvata lugens. PLoS One. 6:e20504. https://doi.org/10.1371/journal.pone.0020504.

Editor: D. Bergstralh