scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Diploid chromosome-level genome assembly and annotation for *Lycorma delicatula*

Anthony A. Snead[1,2,5], Fang Meng[1,2,5], Nicolas Largotta[3], Kristin M. Winchell[1,2] & Brenna A. Levine[3,4] ✉

The spotted lanternfly (*Lycorma delicatula*) is a planthopper species (Hemiptera: Fulgoridae) native to China but invasive in South Korea, Japan, and the United States where it is a significant threat to agriculture. Genomic resources are critical to both management of this species and understanding the genomic characteristics of successful invaders. We report an annotated, haplotype-phased, chromosome-level genome assembly for the spotted lanternfly using PacBio long-read sequencing, Hi-C technology, and RNA-seq. The 2.2 Gbp genome comprises 13 chromosomes, and whole genome resequencing of eighty-two adults indicated chromosome four as the sex chromosome and a corresponding XO sex-determination system. We identified over 12,000 protein-coding genes and performed functional annotation, facilitating the identification of candidate genes that may hold importance for spotted lanternfly control. The assemblies and annotations were highly complete with over 96% of BUSCO genes complete regardless of the database (i.e., Eukaryota, Arthropoda, Insecta). This reference-quality genome will serve as an important resource for development and optimization of management practices for the spotted lanternfly and invasive species genomics as a whole.

Background & Summary

The spotted lanternfly (*Lycorma delicatula*) is a phloem-feeding invasive planthopper species of the family Fulgoridae (Hemiptera) that is native to Asia[1]. This species displays an ability to rapidly invade new habitats[2]. In 2004, the spotted lanternfly was detected in South Korea[3,4] where it damages grape, apple, and stone fruit crops[3]. It was subsequently detected in Japan in 2009[5] before invading the United States via a bridgehead population in South Korea ca. 2012–2014[6]. Since its introduction to the United States, it has rapidly spread across the northeast[7] and is predicted to establish in California by 2033[8]. Its continued spread may be due to its broad environmental tolerances, lack of natural predators, and capacity for long-distance dispersal via anthropogenic activities[2,9–11]. Spotted lanternfly spread may also be facilitated by the prevalence of its preferred host, the invasive tree-of-heaven (*Ailanthus altissima*)[12], which is present on every continent except Antarctica[13]. Importantly, spotted lanternfly activity is not restricted to the tree-of-heaven, and it has been associated with 172 plant taxa with documented feeding behavior on 103[14]. Hence, the spotted lanternfly is a highly invasive pest species with the potential to cause significant agricultural damage.

The spotted lanternfly likely invaded the United States through a single introduction event originating from South Korea[6], potentially resulting in multiple bottlenecks (i.e., China → South Korea, South Korea → United States) and reduced genetic diversity. Reductions in genetic diversity are expected to limit a species' ability to adapt to new environments[15,16]; yet, the spotted lanternfly apparently thrives in novel landscapes and, in the United States, spread rapidly after introduction[17]. The often low genetic diversity characteristic of invasive species juxtaposed with the success of invasive species within novel environments is known as the 'genetic paradox of invasion' and is a central area of research in invasion biology and genomics[18,19]. While more empirical work is required to explicitly test whether populations of the spotted lanternfly have lower genetic diversity within the introduced range compared to the native range, the spotted lanternfly offers a system with a well-documented invasion history to interrogate the genomics of invasion.

[1]Department of Biology, New York University, New York, NY, USA. [2]Center for Genomics & Systems Biology, New York University, New York, NY, USA. [3]Department of Biology, Kean University, Union, New Jersey, USA. [4]Chiricahua Desert Museum, Rodeo, New Mexico, USA. [5]These authors contributed equally: Anthony A. Snead, Fang Meng. ✉e-mail: brlevine@kean.edu

While recent research has primarily focused on modeling the spread, economic consequences, and control of the spotted lanternfly[1,7,20], genetics has also been employed to understand the demographic history and population genetics of the spotted lanternfly across Asia and the United States[6,11,17]. Historically, this work has been limited to microsatellite and mitochondrial markers[4,6,17,21], with the first contig-level genome assembly for the spotted lanternfly published in 2019[22]. However, without a contiguous and annotated genome, researchers are limited in both methodology and inference. A contiguous genome assembly for the spotted lanternfly will allow researchers to determine the genomic location of genes which is essential for accurate annotation[23,24], identification of co-regulated gene clusters[25], categorization of structural rearrangements[26], identification of repetitive elements[25], and characterization of patterns of synteny across species[27]. Likewise, accurate genome annotations will allow researchers to infer functional consequences of genomic variation.

From an applied perspective, a chromosome-scale annotated assembly can yield critical information for the development of spotted lanternfly control methods via the pinpointing of genes related to mating and reproduction. Pheromones in particular likely play a key role in the reproductive biology of the spotted lanternfly by mediating mate-finding, aggregation, and synchronization of oviposition during the mating season[28–31], and recent research has explored the use of pheromonal lures to disrupt the mating activity of this species[31]. A highly contiguous, well-annotated assembly can allow for identification of genes that may be involved in pheromone synthesis and signaling and which could be the focus of control methodology development.

In this study, a functionally-annotated haplotype-phased chromosome-level genome assembly was constructed for a single female spotted lanternfly collected in Union, New Jersey, USA using Pacbio sequencing, Illumina sequencing, and Hi-C technology. Using 279.3 Gigabases (Gb) of Pacbio long reads, the genome size and heterozygosity were estimated before integrating 95.6 Gb of Omni-C reads to generate a chromosome-level assembly. Approximately 260 Gb of paired-end RNAseq data from 12 individuals were juxtaposed with public protein databases to structurally annotate the assembly prior to functional annotation. After assessing genome quality and completeness, the assembly was found to be both highly contiguous and complete. Further, whole-genome resequencing data from 82 adult spotted lanternflies were used to infer an XO sex determination system[32] and identify the X chromosome. Finally, gene ontology enrichment analysis was conducted to identify candidate genes with potential importance for control. Hence, this work resulted in the first annotated, chromosome-level, haplotype-phased reference genome for *L. delicatula* and the family Fulgoridae, yielding an essential genomic resource to inform management of this species and study the genomic architecture of invasion in general.

Methods

Sample collection. For assembly, a single adult female spotted lanternfly was collected on September 20, 2023, from Kean University's campus (Union, New Jersey, United States) which is located in the US invasive range. The female was held in an empty aquarium and starved for approximately 36 hours. After fasting, the female was flash-frozen with liquid nitrogen and stored in a −80 °C freezer until being shipped on dry ice overnight to Dovetail Genomics (Cantata Bio LLC) for DNA extraction, library preparation, sequencing, Hi-C assisted scaffolding, and assembly.

For annotation, twelve adult spotted lanternflies were collected in Queens, New York and divided into three treatment groups, with four individuals per group. To annotate genes that may only be expressed under different environmental conditions, each group was subjected to one of three temperature conditions: room temperature (24–25 °C), cold (10 °C), or warm (45 °C) for two hours. Following the treatments, insects were immediately dissected to separate the head and legs. The head and legs were subsequently stored in Zymo research DNA/RNA shield buffer until RNA extraction.

To identify potential sex chromosomes, eighty-two adult spotted lanternflies were opportunistically collected from October 6 to November 7, 2022 from locations in Pennsylvania, New Jersey, New York, and Connecticut, with the majority of samples collected from New York City. Adults were identified as female (n = 71) or male (n = 11) by inspecting the abdomens for the presence or absence of red valvifers, and whole genome resequencing was performed on the samples. Samples were stored in a −80 °C freezer until DNA extraction.

DNA extraction and sequencing. High Molecular Weight (HMW) DNA was extracted from the head, thorax and appendages of the starved adult female spotted lanternfly using the Qiagen Midi Kit and quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The abdomen was avoided to reduce the potential for sequencing endosymbionts.

For long-read sequencing, a PacBio SMRTbell library (~20 kb) was constructed with the SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using the manufacturer's recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio, Menlo Park, CA, USA) before being sequenced on PacBio Sequel II 8 M SMRT cells to generate 17,184,679 PacBio CCS reads totaling 279.3 Gb (Gigabases) of data.

Omni-C libraries were generated for Hi-C assisted scaffolding. Briefly, chromatin was fixed with formaldehyde prior to digestion with DNAse I. Chromatin ends were repaired and ligated to a biotinylated bridge adaptor before proximity ligation. After reverse crosslinking, the DNA was purified and treated to remove unligated biotin. NEBNext Ultra enzymes and Illumina-compatible adaptors were used to construct sequencing libraries before biotin-containing fragments were extracted with streptavidin beads and enriched via PCR. The Omni-C libraries were sequenced on an Illumina HiSeq X to ~30x coverage resulting in 95.6 Gb of Hi-C data. Library preparation and genome assembly were conducted by Dovetail Genomics (Cantata Bio LLC).

RNA sequencing was conducted to facilitate annotation. Total RNA was extracted from RNA shield-preserved insects using Qiagen's RNeasy Fibrous Tissue Mini Kit following the manufacturer's protocol. RNA purity and concentration were assessed using a Nanodrop (Thermo Fisher Scientific, Waltham,

Species	K-mer	Genome Size		Heterozygous Ratio	
		Min	Max	Min	Max
L. delicatula	31	2.292	2.294	0.42%	0.43%

Table 1. Estimated genome size and heterozygosity of the *L. delicatula* genome (K-mer = 31).

MA), Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), and an Agilent 4200 TapeStation System (Agilent, Santa Clara, CA). All samples had an RIN number above 8 to limit prepping samples with polyA tail degradation. Libraries were constructed with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs E7760S) with the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs E7490S) for mRNA enrichment and NEBNext Multiplex Oligos for Illumina (New England Biolabs 96 Unique Dual Index Primer Pairs Set 1). Libraries were pooled and sequenced on an SP cell on an Illumina NovaSeq. 6000; however, a head sample in the cold-temperature treatment and a leg sample in the room-temperature treatment had low quality and were excluded from sequencing. The reads were basecalled using Picard IlluminaBasecallsToFastq version 2.23.8[33] with APPLY_EAMSS_FILTER set to false. The reads were demultiplexed using Pheniqs version 2.1.0[34] with one mismatch in the sample index sequence allowed. The basecalling and demultiplexing were executed using a custom nextflow pipeline, GENEFLOW[35]. Sequencing resulted in 868,805,636 paired-end 150 bp reads for annotation. Adaptors and low-quality base pairs were removed before filtering out any reads that were less than 50 bp in length with trimgalore![36].

Genomic DNA was extracted from male and female adult spotted lanternflies to identify potential sex chromosomes. DNA was extracted from the thorax and head tissues using the Qiagen DNeasy Blood & Tissue Kit, following the manufacturer's protocol. The DNA concentration was quantified using the Qubit 1X dsDNA High Sensitivity Assay. Genomic libraries were constructed using the NEBNext® Ultra™ II FS DNA Library Prep Kit, with each sample uniquely labeled using NEBNext® Multiplex Oligos (96 Unique Dual Index Primer Pairs Set 1 & 2) for Illumina®. Quality and concentration of the libraries were assessed with the 4200 TapeStation System (Agilent Technologies), producing libraries with an average insert size of 475 bp. Sequencing was performed using 150 bp paired-end reads on the Illumina NovaSeq. 6000 platform, generating a minimum of 40 million reads per sample.

Predicted genome size and heterozygosity. Genome size and heterozygosity were predicted using Meryl v1.4.1[37] and GenomeScope v2[38]. Long-read sequencing data was used with Meryl v1.4.1[37] and a k value of 31 to generate the k-mer frequency distribution table. The k-mer distribution table was used as input for Genomescope v2[38] to predict genome size and heterozygosity which suggested the genome is between 2.292 and 2.294 Gb in size with a heterozygosity between 0.42% and 0.427% (Table 1).

***De-novo* genome assembly.** A *de novo* assembly of *L. delicatula* was constructed using the PacBio HiFi long reads with hifiasm v15.4[39,40]. Contaminated sequences were identified and removed using BLAST[41] against the NCBI nt database and Blobtools2 v1.1.1[42]. Haplotigs and contig overlaps were removed with purge_dups v1.2.5[43]. The resulting primary genome assembly was 2.45 Gb long in 6,650 contigs with an N50 of 51,941,045 bp, auN 70,493,069.18 bp, L50 of 13, L90 of 101 and a GC content of 31.26% (Table 2). Primary genome assembly quality was assessed by using the Eukaryota, Arthropoda, and Insecta Benchmarking Universal Single-Copy Orthologs (BUSCO) datasets[44,45] with compleasm[46]. Results indicated that the primary assembly was highly complete with 255 (100%), 999 (98.62%) and 1,347 (98.53%) BUSCO groups identified as complete out of the 255 groups for Eukaryota, 1,013 groups for Arthropoda, and 1,367 groups for Insecta (Table 3; Fig. 1).

The phased assemblies were further scaffolded into pseudo-chromosome molecules using the Omni-C sequencing data with HiRise[47]. First, Omni-C sequences were aligned to the draft assemblies with BWA[48] before HiRise used the alignments to produce a likelihood model for the genomic distance between read pairs, identify and break misjoins, score prospective joins, and make joins above a threshold. For haplotype one, HiRise made 188 joins and three breaks using 179,563,002 read pairs, while HiRise made 138 joins and two breaks for haplotype 2 using 181,850,961 read pairs. The scaffolded haplotype one assembly was 2.63 Gb long in 7,995 scaffolds with a scaffold N50 of 165,984,238 bp, auN 179,668,902.19, bp, L50 of 6, L90 of 1,166 and a GC content of 31.48% (Table 2). The scaffolded haplotype two assembly was 2.5 Gb long in 3,046 scaffolds with a scaffold N50 of 180,825,106 bp, auN 191,095,999.14 bp, L50 of 5, L90 of 24 and a GC content of 31.19% (Table 2). Completeness of the haplotype-phased assemblies were assessed against the Eukaryota, Arthropoda, and Insecta BUSCO datasets[44,45] with compleasm[46]. Results indicated that both haplotype 1 and 2 were of high quality with 254 (99.61%), 1,008 (99.5%), and 1,357 (99.27%) complete BUSCO groups identified for haplotype 1 and 255 (100%), 1,010 (99.71%), 1,360 (99.49%) for haplotype 2 out of the 255 groups for Eukaryota, 1013 groups for Arthropoda, and the 1367 groups for Insecta (Table 3; Fig. 1).

The resulting Hi-C contact maps suggested 13 chromosomes (Fig. 2A,C). Both assemblies were cleaned to only retain the 13 largest scaffolds (i.e., pseudo-chromosomal molecules; Table 4) with pseudo-chromosomes ranging from 55,466,916 bp to 404,011,691 bp (Table 4). To ensure cleaning the assemblies to the 13 pseudo-chromosomal molecules did not impact genome quality, the assemblies were again assessed for Eukaryota, Arthropoda, and Insecta BUSCO datasets[44,45] with compleasm[46]. The cleaned assemblies were highly similar to the uncleaned assemblies in genome completeness with >99% of BUSCO genes identified as complete regardless of the database, and the number of complete and duplicate genes decreased, and the number of complete and single copy genes increased indicating the cleaned assemblies are better quality than unfiltered

Statistic Type	Final Contig Assembly	Hi-C Scaffolded		Cleaned Assemblies (13 Chr)	
		Haplotype 1	Haplotype 2	Haplotype 1	Haplotype 2
N50	51,941,045	165,984,238	180,825,106	182,426,808	180,825,106
N90	628,811	63,206	1,690,275	123,181,216	121,173,264
auN	70,493,069.18	179,668,902.19	191,095,999.14	214,304,014.63	215,404,078.82
L50	13	6	5	5	5
L90	101	1166	24	10	10
Shortest Length	16,219	6,849	13,932	55,466,916	56,512,573
Longest Length	220,252,607	404,011,691	403,805,378	404,011,691	403,805,378
Total length	2,455,973,448	2,625,905,697	2,500,931,066	2,201,051,051	2,217,731,716
Number of Fragments	4,613	7,995	3,046	13	13
GC Content (%)	31.26%	31.48%	31.19%	30.74%	30.75%

Table 2. Statistics for the initial hifiasm assembly after decontamination and purging duplicates (Final ContigAssembly), both haplotypes after scaffolding with Hi-C (Hi-C Scaffolded), and the cleaned assembly trimmed to the13 pseudo-chromosomal molecules (Cleaned Assemblies) for *Lycorma delicatula*. All metrics are relative to the scaffold except for the Final Contig Assembly which is based on contigs.

Type	Assembly	Dataset	Results
Assembly	Unphased	Eukaryota	C:100%[S:87.06%, D:12.94%], F:0%,M:0%,n:255
Assembly	Unphased	Arthropoda	C:98.6%[S:84.21%, D:14.41%], F:0.2%,M:1.18%,n:1013
Assembly	Unphased	Insecta	C:98.53%[S:83.61%, D:14.92%], F:0.37%,M:1.1%,n:1367
Assembly	Hi-C Scaffolded Haplotype 1	Eukaryota	C:99.61%[S:78.43%, D:21.18%], F:0.39%,M:0%,n:255
Assembly	Hi-C Scaffolded Haplotype 1	Arthropoda	C:99.5%[S:79.66%, D:19.84%], F:0.39%,M:0.1%,n:1013
Assembly	Hi-C Scaffolded Haplotype 1	Insecta	C:99.27%[S:78.57%, D:20.7%], F:0.22%,M:0.51%,n:1367
Assembly	Hi-C Scaffolded Haplotype 2	Eukaryota	C:100%[S:88.42%, D:11.76%], F:0%,M:0%,n:255
Assembly	Hi-C Scaffolded Haplotype 2	Arthropoda	C:99.71%[S:88.65%, D:11.06%], F:0.2%,M:0.11%,n:1013
Assembly	Hi-C Scaffolded Haplotype 2	Insecta	C:99.49%[S:87.42%, D:12.07%], F:0.22%,M:0.29%,n:1367
Assembly	Cleaned Haplotype 1	Eukaryota	C:99.22%[S:97.65%, D:1.57%], F:0%,M:0.78%,n:255
Assembly	Cleaned Haplotype 1	Arthropoda	C:99.31%[S:98.32%, D:0.99%], F:0.2%,M:0.49%,n:1013
Assembly	Cleaned Haplotype 1	Insecta	C:98.91%[S:97.81%, D:1.1%], F:0.29%,M:0.8%,n:1367
Assembly	Cleaned Haplotype 2	Eukaryota	C:99.61%[S:98.04%, D:1.57%], F:0%,M:0.39%,n:255
Assembly	Cleaned Haplotype 2	Arthropoda	C:98.01%[S:97.83%, D:1.18%], F:0.1%,M:0.89%,n:1013
Assembly	Cleaned Haplotype 2	Insecta	C:98.83%[S:97.59%, D:1.24%], F:0.22%,M:0.95%,n:1367
Annotation	Cleaned Haplotype 1	Eukaryota	C:99.21%[S:86.27%, D:12.94%], F:0%,M:0.78%,n:255
Annotation	Cleaned Haplotype 1	Arthropoda	C:98.22%[S:86.28%, D:11.94%], F:0.39%,M:1.38%,n:1013
Annotation	Cleaned Haplotype 1	Insecta	C:97.08%[S:85.52%, D:11.56%], F:0.51%,M:2.41%,n:1367
Annotation	Cleaned Haplotype 2	Eukaryota	C:98.43%[S:87.06%, D:11.37%], F:1.18%,M:0.39%,n:255
Assembly	Cleaned Haplotype 2	Arthropoda	C:97.92%[S:85.98%, D:11.94%], F:0.3%,M:1.78%,n:1013
Annotation	Cleaned Haplotype 2	Insecta	C:96.63%[S:85.44%, D:11.19%], F:0.51%,M:2.85%,n:1367

Table 3. The *Lycorma delicatula* genome completeness results compared to the Eukaryota, Arthropoda, and Insecta BUSCO datasets with compleasm for the initial assembly (unphased), the Hi-C Scaffolded Haplotype 1 assembly, the Hi-C Scaffolded Haplotype 2 assembly, the Cleaned Haplotype 1 assembly, the Cleaned Haplotype 2 Assembly, as well as the annotations for the cleaned assemblies. The cleaned assemblies refer to the assemblies after only chromosome-level scaffolds were retained. The results are reported in common BUSCO notation where C refers to complete, S refers to single-copy, D refers to duplicated, F refers to fragmented, M refers to missing, and n indicates the total number within the database.

assemblies (Table 3; Fig. 1). The alignment of long-read data with Minimap2 v2.2[49] indicated 99% alignment rate for the hifi data.

Repeat sequences and genome annotation. Prior to structural annotation, the repetitive genomic elements of both assemblies were identified and masked with RepeatModeler v2.0.5[50] and RepeatMasker v4.1.5[51] in Dfam TE-tools v1.87 with the LTR pipeline, resulting in 52.47% and 51.93% of bases identified as repeats and masked for haplotype 1 and haplotype 2, respectively (Table 5). While many repeats remained uncategorized (Haplotype 1: 18.03%; Haplotype 2: 17.56%), the repeats successfully classified were primarily retroelements (Haplotype 1: 15.84%; Haplotype 2: 15.34%), specifically LINEs (Haplotype 1: 14.25%; Haplotype 2: 13.55%), and transposons (Haplotype 1: 15.21%; Haplotype 2: 15.29%; Table 5).

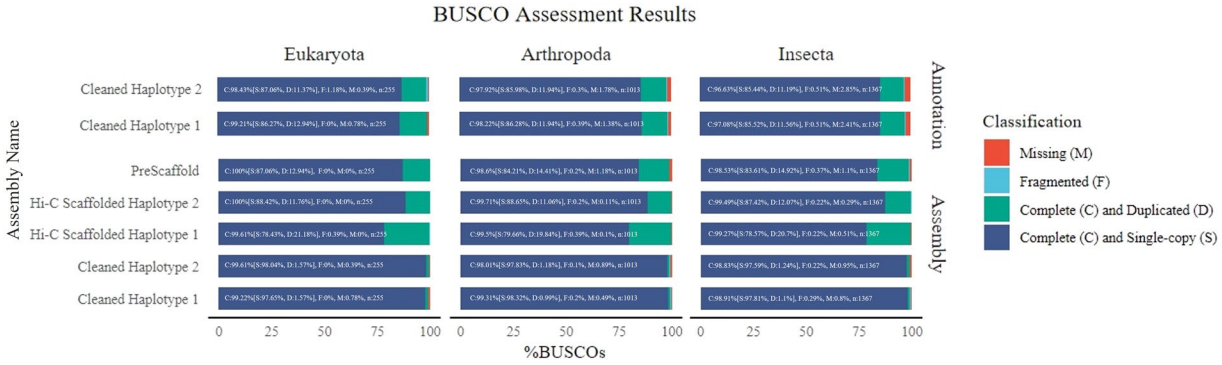

Fig. 1 Horizontal stacked bar graph visualizing the *Lycorma delicatula* genome completeness based on BUSCO genes for the Eukaryota, Arthropoda, and Insecta databases filled with the standard BUSCO notation where C refers to complete, S refers to single-copy, D refers to duplicated, F refers to fragmented, M refers to missing, and n indicates the total number of genes within the database.

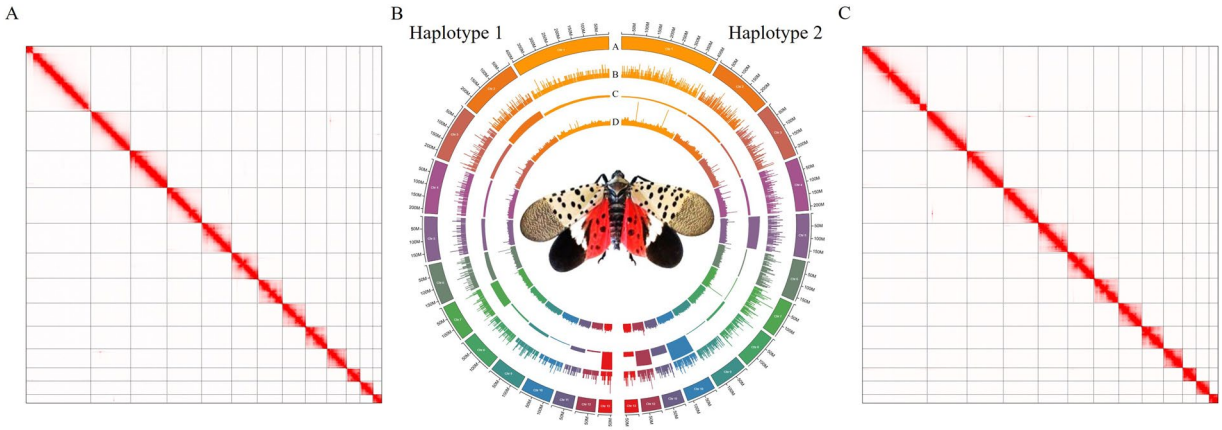

Fig. 2 A multi-paneled figure showing the Hi-C contact map for haplotype 1 (**A**) and haplotype 2 (**B**) for *Lycorma delicatula*. The Hi-C maps indicate 13 chromosomes. A circos plot is shown in the center with haplotype 1 on the left and haplotype 2 on the right. An adult spotted lanternfly (*Lycorma delicatula*) is displayed in the center of the circos plot. In the circos plot, the chromosomes are labeled in the outermost ring (A) along with the gene density (B), gc content (C), and repeat density (D) per 50 kb windows across the genome.

No	Scaffold Name	Length (bp)	
		Haplotype 1	Haplotype 2
1	chr_1	404,011,691	403,805,378
2	chr_2	238,731,119	247,834,938
3	chr_3	227,282,743	228,742,886
4	chr_4	217,703,105	217,132,706
5	chr_5	182,426,808	180,825,106
6	chr_6	165,984,238	165,490,700
7	chr_7	149,132,726	149,373,276
8	chr_8	142,605,837	147,200,923
9	chr_9	133,192,330	135,265,946
10	chr_10	123,181,216	121,173,264
11	chr_11	83,140,728	84,343,789
12	chr_12	78,191,594	80,030,231
13	chr_13	55,466,916	56,512,573

Table 4. Lengths of pseudo-chromosomal molecules for the finalized *Lycorma delicatula* assemblies.

Type	Haplotype 1			Haplotype 2		
	No.	Length Occupied	Percent	No.	Length Occupied	Percent
Retroelements	896,572	348,556,561	15.84	873,261	340,225,439	15.34
SINEs	89,222	9,202,900	0.42	67,525	8,046,017	0.36
LINEs	685,876	313,680,656	14.25	651,773	300,564,174	13.55
L2/CR1/Rex	94,195	17,920,889	0.81	109,977	23,204,373	1.05
R1/LOA/Jockey	189,607	112,700,280	5.12	160,060	114,614,274	5.17
R2/R4/NeSL	9,555	3,129,876	0.14	4,888	2,667,872	0.12
RTE/Bov-B	228,607	128,614,373	5.84	179,319	90,067,052	4.06
L1/CIN4	9,531	1,019,222	0.05	944	80,475	0
LTR elements	121,474	25,673,005	1.17	153,963	31,615,248	1.43
BEL/Pao	24,345	9,406,615	0.43	10,795	7,952,108	0.36
Ty1/Copia	17,946	2,925,204	0.13	13,094	2,032,996	0.09
Gypsy/DIRS1	31,141	8,519,873	0.39	54,370	10,100,123	0.46
DNA transposons	969,583	334,854,306	15.21	955,596	339,152,694	15.29
hobo-Activator	50,164	19,444,235	0.88	64,505	21,345,620	0.96
Tc1-IS630-Pogo	731,858	251,287,521	11.42	735,882	258,373,452	11.65
MULE-MuDR	945	382,137	0.02	623	189,610	0.01
PiggyBac	11,649	4,098,249	0.19	5,802	3,086,036	0.14
Tourist/Harbinger	2,026	627,988	0.03	1154	539,720	0.02
Other	171	139,536	0.01	472	184,675	0.01
Rolling-circles	111,189	24,779,754	1.13	117,995	30,471,950	1.37
Unclassified	2,156,858	396,817,622	18.03	2,199,425	389,382,311	17.56
Total interspersed repeats		1,080,228,489	49.08		1,068,760,444	48.19
Small RNA	72,084	9,019,510	0.41	67,828	10,451,332	0.47
Simple repeats	869,466	40,508,678	1.84	878,286	40,588,635	1.83
Low complexity	164,746	8,263,177	0.38	170,250	8,554,648	0.39

Table 5. The repeat content for each *Lycorma delicatula* haplotype assembly. Penelope, CRE/SLACS, Retroviral, and En-SPM type repeats were not identified and are not included in the table. Further, the term "Other" refers to Mirage, P-element, and Transib repeat classes.

The masked genomes were structurally annotated using the BRAKER v3.0.8 annotation pipeline[52] with the aforementioned RNA-seq and publicly available protein data. Within the pipeline, RNA-seq samples were aligned to the genomes with HiSat2 v2.2.1[53] and reference-guided transcriptome assemblies constructed with StringTie v2.2.1[54]. The assembled transcripts and protein databases were used to train GeneMark-ETP[55] followed by AUGUSTUS v3.5[56] training and prediction using the same RNA-seq and protein databases to produce both homology-based and *ab initio* predictions. The protein databases included the Orthodb[57] Arthropod dataset (4,307,558 proteins) and the UniProt[58] entries for Arthropods (6,349,339 proteins) totaling 10,656,897 proteins. TSEBRA v1.2[59] was used to combine the AUGUSTUS and GeneMark-ETP results. After filtering to retain the longest transcript per gene, 12,410 genes were predicted for haplotype 1, while haplotype 2 was predicted to have 12,794 genes. Annotation completeness and quality were assessed for each haplotype using the Eukaryota, Arthropoda, and Insecta Benchmarking Universal Single-Copy Orthologs (BUSCO) datasets[44,45] with compleasm[46]. Both haplotype annotations were highly complete with 253 (99.21%), 995 (98.22%), and 1,327 (97.08%) BUSCO groups identified as complete from the predicted proteins from haplotype 1 along with 251 (98.43%), 992 (97.92%), and 1,321 (96.63%) complete BUSCO genes from the haplotype 2 structural annotation out of the 255 groups for Eukaryota, 1,013 groups for Arthropoda, and the 1,367 groups for Insecta (Table 3; Fig. 1).

The longest transcripts were functionally annotated for each haploytpe using a combination of eggnog-mapper v2.1.12[60], interproscan v5.68–100[61], and funannotate v1.8.17[62] to add Pfam domains, CAZYmes[63], secreted proteins, proteases (MEROPS[64]), BUSCO groups[45], InterPro terms[61], Gene Ontology terms[65], Eggnog annotations, and Orthologous Groups[66] (OG) along with gene names and product descriptions from UniProt[58]. For haplotype 1, 9,311, 10,516, 11,103, 9,262, 225, 582, and 1,651 genes were annotated with GO terms, InterPro classifications, Eggnog annotations, Pfam domains, Cazyme terms, Merop results, and Busco groups, respectively accounting for 75%, 84.7%, 89.5%, 74.6%, 1.8%, 4.7%, and 13.3% of the total number of genes in the *L. delicatula* haplotype 1 genome. For haplotype 2, 9,320, 10,682, 11,481, 9,374, 224, 584, and 11,660 genes were annotated with GO terms, InterPro classifications, Eggnog annotations, Pfam domains, Cazyme terms, Merop results, and Busco groups, respectively, accounting for 72.9%, 83.5%, 89.7%, 73.3%, 1.8%, 4.6%, and 13% of the total number of genes in the *L. delicatula* haplotype 2 genome.

Sex chromosome identification. The raw sequence data were processed to remove adapter sequences and low-quality bases using trimgalore![36]. The cleaned reads were then mapped to the two haplotype assemblies using BWA[48]. Coverage was calculated for each chromosome to identify potential sex chromosomes using samtools[67].

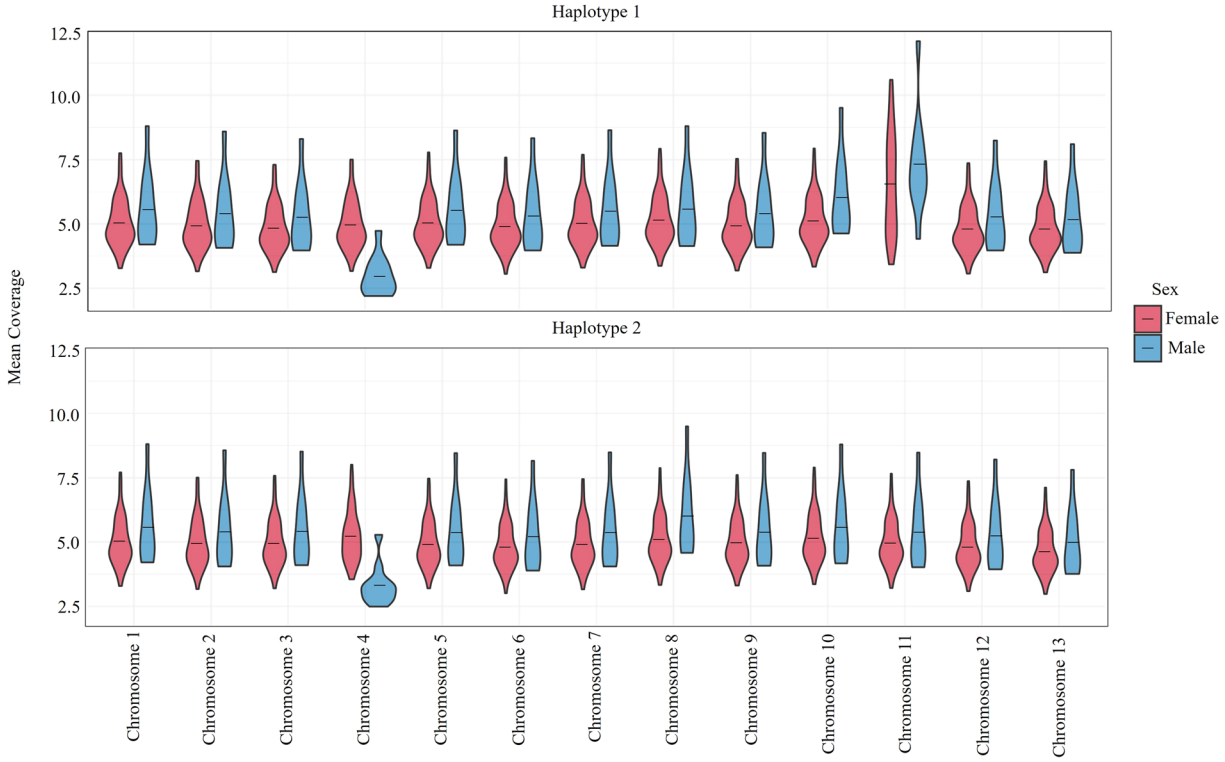

Fig. 3 A violin plot visualizing the mean coverage of each pseudo-chromosome across both *Lycorma delicatula* assemblies colored by sex. The difference between the male and female coverage on pseudo-chromosome four suggests that chromosome four is the sex chromosome (X) and that the sex determination system is XO.

To identify the sex chromosome, read coverage was compared across all pseudo-chromosomes between male and female samples. Pseudo-chromosome 4 consistently exhibited approximately half the coverage in males compared to females (Fig. 3). This reduced coverage suggests that pseudo-chromosome 4 is a sex chromosome, likely representing the hemizygous state in males (XO). This pattern aligns with expectations for sex chromosomes in species with XO sex determination systems in which males possess only one copy of the sex chromosome[32].

Gene ontology enrichment analysis. Gene ontology (GO) enrichment analysis was conducted to identify functional categories significantly overrepresented among genes across pseudo-chromosomes in the spotted lanternfly genome. GO terms were extracted from gene attributes in the GFF3 file, and annotations associated with GO terms were parsed and linked to corresponding gene IDs. Using the *goseq*[68] R package, bias-corrected GO enrichment analysis was conducted using probability weight functions (PWFs) for each chromosome. Significant GO terms ($p < 0.05$) were filtered and visualized as bar plots, with distinct ontological categories (Biological Process, Molecular Function, and Cellular Component) color-coded. The top five GO terms for each pseudo-chromosome were then synthesized into a table to facilitate comparisons among pseudo-chromosomes (Table 6).

The GO Plots for each pseudo-chromosome were inspected to identify terms with potential importance to spotted lanternfly control, with emphasis on development of pheromone-based tools[31]. GO terms related to olfactory receptor activity, odorant binding, and sensory perception of smell, as well as multicellular organism reproduction, were highly represented for pseudo-chromosome 12, suggesting that genes on this pseudo-chromosome may play a role in sex pheromone reception[69–71]. Nine genes within the gustatory receptor (Gr) and G-protein coupled odorant receptor (GPROR) families were identified (i.e., Gr1, Gr64e, GPROR31_2, GPROR31_3, GPROR37, GPROR40, GPROR49_2, GPROR49_3, and GPROR49_4), and 13 genes within the Odorant Receptor (Or) gene family were identified (Or19a_2, Or2a_1, Or2a_2, Or2a_3, Or2a_4, Or2a_5, Or2a_6, Or33a, Or43a, Or63a, Or85d_3, Or85d_4, Or85f_2). Other top GO terms for this pseudo-chromosome include those associated with fatty acid elongation and biosynthetic processes, two activities involved in insect pheromone biosynthesis[72].

Data Records

The raw sequencing data and the annotated haplotype-phased genome assembly of *Lycorma delicatula* have been deposited at the National Center for Biotechnology Information (NCBI). The Hi-C and HiFi data can be found under SRR31402152[73] and SRR31402153[73]. The RNA-seq data can be found under SRR31411873-SRR31411894[73], while the DNA-seq data can be found under SRR29972204-SRR29972212[74]. SRR29972214-SRR29972223[74], SRR29972225-SRR299734[74], SRR29972236-SRR29972255[74], SRR29972257-SRR29972260[74], SRR29972262-SRR299765[74], SRR299767-SRR29972277[74], SRR29972279-SRR29972288[74], SRR29972290-SRR29972293[74]. The chromosome-level genome assemblies have been deposited in the NCBI

Top 5 GO Terms					
Chromosome	**GO Term 1**	**GO Term 2**	**GO Term 3**	**GO Term 4**	**GO Term 5**
Chr 1	nucleosome	structural constituent of chromatin	negative regulation of DNA recombination	nucleosomal DNA binding	nucleosome assembly
Chr 2	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	monooxygenase activity	endothelial cell migration	sulfuric ester hydrolase activity	iron ion binding
Chr 3	cysteine-type peptidase activity	cysteine-type endopeptidase activity	neuron projection membrane	phenylpyruvate tautomerase activity	proteolysis involved in protein catabolic process
Chr 4	flavin adenine dinucleotide binding	oxidoreductase activity, acting on CH-OH group of donors	cell surface receptor signaling pathway	copper ion binding	glutathione-disulfide reductase (NADP) activity
Chr 5	serine-type endopeptidase activity	polyamine oxidase activity	proteolysis	tRNA export from nucleus	cysteine-type endopeptidase inhibitor activity
Chr 6	ATPase-coupled transmembrane transporter activity	glycine dehydrogenase (decarboxylating) activity	lipid catabolic process	ABC-type transporter activity	lipase activity
Chr 7	L-ascorbic acid biosynthetic process	gluconolactonase activity	extracellular matrix	four-way junction DNA binding	1-acylglycerol-3-phosphate O-acyltransferase activity
Chr 8	amino acid transmembrane activity	L-amino acid transmembrane transporter activity	2-(3-amino-3-carboxypropyl)histidine synthase activity	endoplasmic reticulum membrane organization	tissue regeneration
Chr 9	inositol phosphate phosphatase activity	peptidyl-dipeptidase activity	sulfate transmembrane transporter activity	metallopeptidase activity	one-carbon metabolic process
Chr 10	structural constituent of chitin-based larval cuticle	chitin-based extracellular matrix	odorant biding	RNA uridyltransferase activity	RNA 3′ uridylation
Chr 11	acetyktransferase activity, tranferring groups other than amino-acyl groups	positive regulation of phosphatidylinositol 3-kinase/protein kinase B signal transduction	protein serine/threonine kinase activator activity	TORC2 signaling	alpha-galactosidase activity
Chr 12	olfactory receptor activity	detection of chemical stimulus involved in sensory perception of smell	sensory perception of smell	multicellular organism reproduction	odorant binding
Chr 13	amino acid transport	amino acid transmembrane transporter activity	positive regulation of mRNA splicing, via spliceosome	L-ornithine transmembrane transport	L-lysine transmembrane transporter activity

Table 6. Top five gene ontology terms for each chromosome of the *Lycorma delicatula* Haplotype 1 assembly.

assembly under JBLIYR000000000 (GCA_047948215.1)[75] and JBLIYS000000000 (GCA_047948205.1)[76]. The assemblies as well as the annotations for the repeated sequences, gene structure and functional prediction have also been placed in the Figshare repository[77] for immediate use, along with auxiliary GO term graphs.

Technical Validation

To assess the accuracy and completeness of the haplotype-phased genome assembly, compleasm[46] was used to assess the number of single copy orthologs for the Eurkayota, Arthropoda, and Insecta BUSCO databases[44,45] for each cleaned chromosome-level assembly and the associated annotation. The assemblies and their annotations are highly complete. Regardless of the assembly or database, over 99% of BUSCO genes were identified as complete (Table 3, Fig. 1). For the annotation, between 96.5% and 99.2% of BUSCO genes were identified as complete depending on the haplotype and database assessed (Table 3, Fig. 1). The HIFI long reads were subsequently mapped to each haplotype individually to reveal 99.34% and 99.32% for the haplotype one and haplotype two assemblies, respectively.

Usage Notes

The genome and relevant annotation tracks can further be visualized and explored via a public trackhub[78] on the UCSC Genome Browser database[79]. To explore the assembly, annotation, and repeat landscape, visit the UCSC Genome Browser website, select the My Data tab (https://genome.ucsc.edu/cgi-bin/hgHubConnect), and paste the track hub link (https://lycorma-delicatula-tracks.bio.nyu.edu/Spotted_Lanternfly/hub.txt) into the url box under the Connected Hubs tab.

Code availability

The initial haplotype-phased scaffolded genome was assembled by Dovetail Genomics (Cantata Bio LLC) with standard software outlined in the methods with default settings. Scripts for the remaining work including annotation, gene ontology enrichment, and other analyses are located in the Github repository (https://github.com/anthonysnead/SLF-Genome-Assembly).

Received: 11 December 2024; Accepted: 17 March 2025;
Published online: 05 April 2025

References

1. Urban, J. M. Perspective: shedding light on spotted lanternfly impacts in the USA. *Pest Manag. Sci.* **76**, 10–17 (2020).
2. Wakie, T. T., Neven, L. G., Yee, W. L. & Lu, Z. The Establishment Risk of Lycorma delicatula (Hemiptera: Fulgoridae) in the United States and Globally. *J. Econ. Entomol.* **113**, 306–314 (2020).
3. Han, J. M. *et al*. Lycorma delicatula (Hemiptera: Auchenorrhyncha: Fulgoridae: Aphaeninae) finally, but suddenly arrived in Korea. *Entomol. Res.* **38**, 281–286 (2008).
4. Kim, H. *et al*. Molecular comparison of *Lycorma delicatula* (Hemiptera: Fulgoridae) isolates in Korea, China, and Japan. *J. Asia-Pac. Entomol.* **16**, 503–506 (2013).
5. Tomisawa, A., Ohmiya, S., Fukutomi, H., Hayashi, K. & Ishikawa, T. Biological notes on *Lycorma delicatula* (White) (Hemiptera, Fulgoridae) in Ishikawa Prefecture, Japan. *Jpn. J. Entomol.* **16**, 3–14 (2013).
6. Du, Z. *et al*. Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evol. Appl.* **14**, 915–930 (2021).
7. Urban, J. M. & Leach, H. Biology and Management of the Spotted Lanternfly, Lycorma delicatula (Hemiptera: Fulgoridae), in the United States. *Annu. Rev. Entomol.* **68**, 151–167 (2023).
8. Jones, C. *et al*. Spotted lanternfly predicted to establish in California by 2033 without preventative management. *Commun. Biol.* **5**, 1–9 (2022).
9. Jung, J.-M., Jung, S., Byeon, D. & Lee, W.-H. Model-based prediction of potential distribution of the invasive insect pest, spotted lanternfly *Lycorma delicatula* (Hemiptera: Fulgoridae), by using CLIMEX. *J. Asia-Pac. Biodivers.* **10**, 532–538 (2017).
10. Ladin, Z. S., Eggen, D. A., Trammell, T. L. E. & D'Amico, V. Human-mediated dispersal drives the spread of the spotted lanternfly (Lycorma delicatula). *Sci. Rep.* **13**, 1098 (2023).
11. Nakashita, A., Wang, Y., Lu, S., Shimada, K. & Tsuchida, T. Ecology and genetic structure of the invasive spotted lanternfly Lycorma delicatula in Japan where its distribution is slowly expanding. *Sci. Rep.* **12**, 1543 (2022).
12. Uyi, O., Keller, J. A., Swackhamer, E. & Hoover, K. Performance and host association of spotted lanternfly (Lycorma delicatula) among common woody ornamentals. *Sci. Rep.* **11**, 15774 (2021).
13. Kowarik, I. & Säumel, I. Biological flora of Central Europe: *Ailanthus altissima* (Mill.) Swingle. *Perspect. Plant Ecol. Evol. Syst.* **8**, 207–237 (2007).
14. Barringer, L. & Ciafré, C. M. Worldwide Feeding Host Plants of Spotted Lanternfly, With Significant Additions From North America. *Environ. Entomol.* **49**, 999–1011 (2020).
15. Markert, J. A. *et al*. Population genetic diversity and fitness in multiple environments. *BMC Evol. Biol.* **10**, 205 (2010).
16. Sax, D. F. *et al*. Ecological and evolutionary insights from species invasions. *Trends Ecol. Evol.* **22**, 465–471 (2007).
17. Kim, H. *et al*. Tracing the Origin of Korean Invasive Populations of the Spotted Lanternfly, Lycorma delicatula (Hemiptera: Fulgoridae). *Insects* **12**, 539 (2021).
18. Estoup, A. *et al*. Is There a Genetic Paradox of Biological Invasion? *Annu. Rev. Ecol. Evol. Syst.* **47**, 51–72 (2016).
19. Sax, D. F. & Brown, J. H. The paradox of invasion. *Glob. Ecol. Biogeogr.* **9**, 363–371 (2000).
20. Zhang, Y. *et al*. The biology and management of the invasive pest spotted lanternfly, Lycorma delicatula White (Hemiptera: Fulgoridae). *J. Plant Dis. Prot.* **130**, 1155–1174 (2023).
21. Zhang, L., Zhao, W., Wang, F. & Qin, D. Genetic Diversity and Population Structure of Natural Lycorma delicatula (White) (Hemiptera: Fulgoridea) Populations in China as Revealed by Microsatellite and Mitochondrial Markers. *Insects* **10**, 312 (2019).
22. Kingan, S. B. *et al*. A high-quality genome assembly from a single, field-collected spotted lanternfly (Lycorma delicatula) using the PacBio Sequel II system. *GigaScience* **8**, giz122 (2019).
23. Denton, J. F. *et al*. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLOS Comput. Biol.* **10**, e1003998 (2014).
24. Klassen, J. L. & Currie, C. R. Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics* **13**, 14 (2012).
25. Thomma, B. P. H. J. *et al*. Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet. Biol. FG B* **90**, 24–30 (2016).
26. Jonge, R. *et al*. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* **23**, 1271–1282 (2013).
27. Liu, D., Hunt, M. & Tsai, I. J. Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics* **19**, 26 (2018).
28. Cooperband, M. F. & Murman, K. Responses of adult spotted lanternflies to artificial aggregations composed of all males or females. *Front. Insect Sci*. **2**, (2022).
29. Faal, H., Cooperband, M. F., Canlas, I. & Carrillo, D. Evidence of Pheromone Use in a Fulgorid, Spotted Lanternfly. *Forests* **13**, 1639 (2022).
30. Liu, H. Oviposition selection in spotted lanternfly: impact of habitat and substrate on egg mass size and hatchability. *Front. Insect Sci*. **2**, (2022).
31. Cooperband, M. F. & Murman, K. M. Spotted Lanternflies Respond to Natural Pheromone Lures for Mate-Finding and Oviposition. *Insects* **15**, 447 (2024).
32. Tian, R., Zhang, Y. & Yuan, F. Karyotypes of nineteen species of Fulgoroidea from China (Insecta: Homoptera. *Kun Chong Xue Bao. Acta Entomol. Sin.* **47**, 803–808 (2004).
33. Picard Toolkit Broad Institute, Github Repository. https://broadinstitute.github.io/picard/; Broad Institute (2019).
34. Galanti, L., Shasha, D. & Gunsalus, K. C. Pheniqs 2.0: accurate, high-performance Bayesian decoding and confidence estimation for combinatorial barcode indexing. *BMC Bioinformatics* **22**, 359, https://doi.org/10.1186/s12859-021-04267-5 (2021).
35. "GENEFLOW." New York University Center for Genomics and System Biology Genomics Core, GitHub Repository. https://github.com/gencorefacility/GENEFLOW; New York University (2023).
36. Babraham Bioinformatics - Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
37. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
38. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).

39. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
40. Cheng, H. *et al*. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).
41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
42. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research*, **6**, 1287 https://doi.org/10.12688/f1000research.12232.1 (2017).
43. Guan, D. *et al*. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
44. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
45. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
46. Huang, N. & Li, H. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, btad595 (2023).
47. Putnam, N. H. *et al*. Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
48. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
49. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
50. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
51. Chen, N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinforma.* **5**, 4.10.1–4.10.14 (2004).
52. Gabriel, L. *et al*. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
53. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
54. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
55. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
56. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–439 (2006).
57. Zdobnov, E. M. *et al*. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).
58. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
59. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M. & Stanke, M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22**, 566 (2021).
60. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
61. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
62. Palmer, J. M. & Stajich, J. Funannotate v1.8.1: Eukaryotic genome annotation. *Zenodo* https://doi.org/10.5281/zenodo.4054262 (2020).
63. Cantarel, B. L. *et al*. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
64. Rawlings, N. D. *et al*. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632 (2018).
65. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
66. J, H.-C. *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, (2019).
67. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
68. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
69. Sakurai, T. *et al*. Identification and functional characterization of a sex pheromone receptor in the silkmoth Bombyx mori. *Proc. Natl. Acad. Sci.* **101**, 16653–16658 (2004).
70. Touhara, K. & Vosshall, L. B. Sensing odorants and pheromones with chemosensory receptors. *Annu. Rev. Physiol.* **71**, 307–332 (2009).
71. Kang, K. *et al*. Identification of putative fecundity-related gustatory receptor genes in the brown planthopper Nilaparvata lugens. *BMC Genomics* **19**, 970 (2018).
72. Tillman, J. A., Seybold, S. J., Jurenka, R. A. & Blomquist, G. J. Insect pheromones–an overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* **29**, 481–514 (1999).
73. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP546249 (2025).
74. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP522012 (2025).
75. Snead, A. A., Meng, F., Largotta, N., Winchell, K. M. & Levine, B. A. Diploid chromosome-scale genome assembly and annotation of the spotted lanternfly, *Lycorma delicatula*. *GenBank*. https://identifiers.org/ncbi/insdc:JBLIYR000000000 (2025).
76. Snead, A. A., Meng, F., Largotta, N., Winchell, K. M. & Levine, B. A. Diploid chromosome scale genome assembly and annotation of the spotted lanternfly, *Lycorma delicatula*. *GenBank*. https://identifiers.org/ncbi/insdc:JBLIYS000000000 (2025).
77. Snead, A. A., Meng, F., Largotta, N., Winchell, K. M. & Levine, B. A. Haplotype Phased Chromosome-level *Lycorma delicatula* Genome and Annotation. *Figshare*. https://doi.org/10.6084/m9.figshare.28378361 (2025).
78. Snead, A. A. Meng, F. Largotta, N. Winchell, K. M., Levine, B. A. Track Hub [Spotted Lanternfly Assembly Hub] Retrieved from https://lycorma-delicatula-tracks.bio.nyu.edu/Spotted_Lanternfly/hub.txt (2025)
79. Perez, G. *et al*. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res.* **6**, 53 (2025).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number 2312129 to B.A.L. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. A.A.S. is supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2305939. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. We acknowledge the Zegar Family Foundation for their generous support that enabled the RNA sequencing and the whole genome sequencing for sex-chromosome identification. We thank the NYU Center for Genomics and System Biology Genomics Core for their assistance and resources.

Author contributions

Conceptualization: B.A.L., K.M.W. Methodology: A.A.S., F.M., K.M.W., B.A.L. Sample collection: N.L., F.M., K.M.W. Visualization: A.A.S., F.M. Funding acquisition: B.A.L., K.M.W. Writing - original draft: A.A.S., F.M., N.L. Writing - review and editing: A.A.S., F.M., N.L. K.M.W., B.A.L.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.A.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.