INTRODUCTION

Almost all major crops surrender yield losses to parasitic nematodes with annual damages exceeding US one billion worldwide (Nicol et al., 2011). The most well-known and complex group of these plant parasites comprises root-knot (Meloidogyne spp) and cyst nematodes (Globodera spp. and Heterodera spp), which manipulate the host to develop a long-term feeding site. The soybean cyst nematode...
(Heterodera glycines) is of particularly great economic importance due to its prominent role in reducing soybean yields worldwide (Davis & Tylka, 2000; Jones et al., 2013; Koennig & Wrather, 2010; Nicot et al., 2011). Overcoming this crop pest requires scrutinizing the H. glycines lifestyle and the molecular exchange at the core of this problem.

The lifecycle of H. glycines begins as an egg that is queued to hatch. The emerged juvenile nematode migrates to the host root zone, where it penetrates the outer layers of roots using a combination of mechanical and enzymatic processes, and eventually induces a single root cell near the vascular cylinder to form a feeding site, known as a syncytium (Pogorelko et al., 2016). The syncytium becomes metabolically active and expands to incorporate hundreds of adjacent cells through cell wall breakdown and protoplast fusion. The syncytium matures to an efficient nutrient sink with enlarged host nuclei and pronounced cytoplasmic streaming (Abad & Williamson, 2010; Lilley et al., 2005).

Successful feeding site development depends upon the parasite’s ability to manipulate a complex interaction with its host via the transfer of nematode gland cell-produced effector proteins into or around host root cells (Hewezi & Baum, 2013; Mitchum et al., 2013; van den Akker et al., 2014). During juvenile nematode migration within the root, plant cell walls are digested by an abundance of secreted enzymes including cellulases, pectate lyases and other hydrolases (De Boer et al., 2002; Gao et al., 2003; Rai et al., 2015). In later parasitic stages, the nematode manipulates plant metabolism (Bekal et al., 2003), development (Matthews et al., 2014; Mitchum et al., 2008; Wang et al., 2010, 2011), and elicits a dramatic and long-term suppression of host defences (reviewed by Hewezi, 2015; Hewezi & Baum, 2013; Mitchum et al., 2013; Noon et al., 2015). While the functional mechanisms of many effector proteins remain elusive, a variety of functions have been attributed to previously characterized effector proteins secreted from the esophageal gland cells of root-knot and cyst nematodes (Chronis et al., 2013; Elling & Jones, 2014; Eves-van den Akker et al., 2014; Gao et al., 2001, 2003; Haegeman et al., 2012; Hamamouch et al., 2012; Hewezi, 2015; Hewezi et al., 2010; Lee et al., 2011; Maier et al., 2013; Noon et al., 2015; Patel et al., 2010; Vanholme et al., 2006; Wang et al., 2001, 2005). For example, a chorate mutase protein, typically absent in animals, is secreted by root-knot and cyst nematodes to manipulate the host cell’s shikimate pathway, a pathway involved in producing aromatic amino acids, plant hormones, cell wall components, and plant defence metabolites (Bekal et al., 2003; Lu et al., 2008; Noon & Baum, 2016; Vanholme et al., 2009). Signalling peptides, like CLAVATA3 plant peptide mimics, can affect plant developmental pathways (Lu et al., 2009; Olsen & Skriver, 2003; Replogle et al., 2011; Wang et al., 2005, 2010, 2011). While these effectors have led to a better understanding of plant-nematode interactions, only a small portion have been functionally characterized.

Understanding the totality of effector proteins in the nematode genome and how they manipulate the host will shed light on this molecular interplay, inspiring the development of novel mechanisms to defend plants from these important pests. To accomplish this goal for the soybean cyst nematode, two annotated genome assemblies were published from two different nematode strains: a partially virulent TN10 line (Masonbrink et al., 2019) and a highly virulent X12 line (Lian et al., 2019). Here, we improve upon the existing published H. glycines genomes by reassembling the TN10 PacBio reads de novo and scaffolding with Chicago and Hi-C reads to obtain the highest quality plant-parasitic nematode genome assembly to date with nine complete pseudomolecule chromosomes and zero unscaffolded contigs. We went to great lengths to ensure the integrity of the assembly, by mapping input genome assembly reads back to the assembly and by using synteny in among H. glycines assemblies. While large rearrangements exist between the TN10 and X12 pseudomolecule assemblies, technological improvements in Hi-C scaffolding software (Lachesis vs. Juicer) revealed that these differences can be attributed to many small and a few large chromosomal misjoins in the X12 assembly. Although the X12 and this latest TN10 assembly have similar assembly metrics and size, 141 versus 158 Mb, choices in gene prediction created a large disparity in gene frequency between annotations. Here, we have attempted to bridge this gap with an extensive gene annotation that uses multiple prediction pipelines and lines of evidence to generate an annotation that is complete and comparable to other parasitic nematode species. We also limited our gene homology input to include only genes of related Tylenchida species to prevent the homology-driven over-simplification of gene structure when using more distant and nonparasitic relatives. Using this vastly improved genomic resource, we explore the nature of previously published effectors and other secreted proteins to address the heart of H. glycines genomics, to understand the adaptive evolution involved in the constant battle between host resistance and parasite virulence.

2 | MATERIALS AND METHODS

2.1 | Dovetail Chicago library preparation and sequencing

A Chicago library was prepared as described previously (Putnam et al., 2016). Briefly, ~500 ng of high molecular weight (HMW) gDNA (mean fragment length ~ 75 kbp) was reconstituted into chromatin in vitro and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5’ overhangs filled in with bacterially synthesized nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size, and sequencing libraries were generated using NEBNext ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X to produce 500 million 151 base paired-end reads.

2.2 | Dovetail Hi-C library preparation and sequencing

A Dovetail Hi-C library was prepared in a similar manner as described previously (Lieberman-Aiden et al., 2009). Chromatin was fixed in the nucleus with formaldehyde, extracted, and DpnII digested. 5’
overhangs were filled with biotinylated nucleotides, and then free blunt ends were ligated. Crosslinks were reversed for DNA purification from proteins. Purified DNA was treated to remove biotin outside of ligated fragments. The DNA was then sheared to ~350 bp, and libraries were generated using NEBNEx Ultra enzymes and Illumina-compatible adapters. Biotinylated fragments were isolated with streptavidin beads before PCR enrichment of libraries and sequenced on an Illumina HiSeq X to produce 531 million 2 × 151 bp paired end reads.

### 2.3 | Genome assembly

To create a more reproducible *H. glycines* genome assembly, we used Falcon with previously deposited Pacbio sequencing (SRX2692203 - SRX2692222), rather than use the previous assembly (Masonbrink et al., 2019). Falcon unzip 0.4.0 (Chin et al., 2016) was then used to reduce the heterozygosity in the assembly. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. This Dovetail assembly was further scaffolded with SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise. Dovetail Genomics scaffolded this assembly with Chicago and Hi-C reads using a modified SNAP read mapper (http://snap.cs.berkeley.edu) and an iterative assembly with HiRise.

### 2.4 | Gene prediction

Genes were predicted using a MIKADO 1.24 pipeline (Venturini et al., 2018) that picked consensus transcripts from seven transcriptome assemblies and gene predictions. First, the genome was masked using REPEATMODELER 1.0.11 (Smit et al., 2014) and REPEATMASKER 4.0.9 (Smit et al., 2015). Previously published Illuma RNA-seq reads (SRX3339090–SRX3339098) were processed with TRIM GALORE 0.4.5 (Krueger, 2012), HISAT2 2.1.0 (Kim et al., 2015), and SAMTOOLS 1.9 (Li et al., 2009) on both a masked and an unmasked genome. Previously published NCBI expressed sequence tags (downloaded 06-17-19), previously published effectors (Table 1) (Gao et al., 2001, 2003; Noon et al., 2015), and IsoSeq (SRX3702373) were aligned using GMAP (version 2018-03-25) (Wu et al., 2016). These data were utilized with BRAKER 2.1.0 (Hoff et al., 2018) using all three data sources, annotating both an unmasked assembly and a masked assembly to compensate for parasitic genes associated with high copy numbers.

Transcriptomes were assembled using the guidance of a masked genome with TRINITY 2.3.2 (Haas et al., 2013; Henschel et al., 2012), CLASS2 2.1.7 (Song et al., 2016), STRINGTIE 1.3.4a (Pertea et al., 2015, 2016), and SPADES 3.13.1 (Bankevich et al., 2012). This first Mikado prediction was utilized in a second round of Mikado, supplemented with masked braker prediction and a MAKER 2.31.10 (Cantarel et al., 2008) gene annotation from a 368-scaffold version of the assembly. All resulting predictions from the second round of Mikado were collapsed into gene loci via using shared intron/exon borders with Cufflinks gffread (CUFFLINKS 2.2.1) (Trapnell et al., 2012).

The Maker annotation mentioned previously was run over four rounds, with Maker’s internal algorithm first, then AUGUSTUS 3.2.1, then SnapHMM 2006-07-28, followed by GENEMARK-ES 4.32. REPEATMODELER 1.0.11 and REPEATMASKER 4.0.9 were used to perform the softmasking used in the annotation. Maker utilized all transcripts and proteins from related species genomes (Coordinators, 2016; Howe et al., 2017) and UNIPROT (Consortium, 2019), including: Bursaphelenchus xylophilus (Kikuchi et al., 2011), Ditylenchus destructor (Zheng et al., 2016), Globodera pallida (Cotton et al., 2014), Globodera rostochiensis (Eves-van den Akker et al., 2016), Globodera ellingtonae (Phillips et al., 2017), Meloidogyne floridensis (Lunt et al., 2014), Meloidogyne hapla (Opperman et al., 2008), Meloidogyne incognita (Abad et al., 2008), Parastrongyloides trichosuri, Rhabditophanes KR3021, Strongyloides papillosus, Strongyloides ratti, Strongyloides stercoralis (Hunt et al., 2016); all *H. glycines* ESTs from NCBI (Coordinators, 2016), and a BRAKER 1.9 (Hoff et al., 2015) annotation on this unmasked assembly using published RNA-seq (SRX3339090–SRX3339098).

To obtain the final gene counts, mRNAs were classified via repetitiveness and expression. mRNAs with at least a 30% overlap in two CDS sequences with a REPEATMASKER repeat were removed from the final gene prediction (8313). Another 643 transcripts were filtered from the annotation that had a functional containing the term, heli-tron, transposase, or transposon. These genes were placed in a separate annotation track, representing transposable elements. Another 11,268 predicted mRNAs with one or fewer unique mapping RNA-reads were placed in a separate annotation track, representing unexpressed genes.

### 2.5 | Functional gene annotations

Gene annotations were compiled from INTERPROSCAN 5.27-66.0 (Finn et al., 2016) and BLAST (Madden, 2013) searches to NCBI NT and nr databases downloaded on 10-23-19 (Coordinators, 2016), as well as swissprot/uniprot databases downloaded on 12-09-2019 (Consortium, 2019). Genes encoding transposable element-associated proteins were identified using BEDTOOLS 2.27.1 (Quinlan, 2014) with exon overlaps to EDTA and Repeatmodeler-predicted transposable elements.
TABLE 1 Published effector locations and putative duplications between TN10 and X12 genome assemblies. These are nucleotide alignment with GMAP of 80 previously published effector genes. Secreted status of predicted effector proteins indicates the presence of a signal peptide and the absence of a transmembrane domain past the signal peptide cleavage site.

| Effector annotation | TN10 location | X12 location | TN10 duplication | X12 duplication | Secreted? | Secreted? |
|---------------------|---------------|--------------|-----------------|-----------------|-----------|-----------|
| GLAND17             | Chromosome_1:12774011-12775451 chr4:8520525-8521897 | Yes | Yes |
| flhggfha21E12       | Chromosome_1:13503194-13503710 chr4:8260552-8261068 | No | No |
| GLAND11             | Chromosome_1:14411530-14414106 chr4:7744175-7746752 | No | Yes |
| flhggfha10C02       | Chromosome_1:14531951-14532560 chr4:5880824-5881776 | No | No |
| GLAND5              | Chromosome_1:15579816-15580580 chr4:6040213-6040972 | No | No |
| GLAND10             | Chromosome_1:16136574-16137525 chr4:5613194-5614144 | Yes | No |
| GLAND10             | Chromosome_1:16310843-16311797 chr4:5580824-5581776 | Yes | No |
| GLAND18             | Chromosome_1:18105545-18144598 chr4:4539362-4543392 | Yes | Yes |
| GLAND8              | Chromosome_1:18145074-18146760 chr4:4395204-4403690 | Yes | No |
| flhggfha4G12        | Chromosome_1:18238568-18239414 chr4:4438743-4439553 | Yes | No |
| flhggfha2B10        | Chromosome_1:18238603-18256409 chr4:4438743-4439553 | Yes | No |
| flhggfha20E03       | Chromosome_1:186637920-18693063 #N/A | Yes | Yes |
| flhggfha20E03       | Chromosome_1:18846674-18701829 chr4:4355808-4356943 | Yes | No |
| flhggfha6F06        | Chromosome_1:19654918-19691572 chr2:16042063-16043220 | Yes | No |
| flhggfha4D06        | Chromosome_1:21886622-21887572 chr2:16043772-16088548 | Yes | No |
| GLAND6              | Chromosome_1:22083063-22083708 chr2:16090328-16119698 | Yes | Yes |
| GLAND6              | Chromosome_1:22086326-22093978 chr2:16086361-1608994 | Yes | No |
| flhggfha4D06        | Chromosome_1:22206326-22106675 chr2:16044512-1604718 | Yes | No |
| flhggfha2E04        | Chromosome_1:22252047-22253314 chr2:16087112-1608995 | Yes | No |
| GLAND13             | Chromosome_1:22848387-22853278 chr4:805461-8010347 | Yes | No |
| flhggfha22C12       | Chromosome_1:22948580-22949324 chr2:203467-204198 | Yes | No |
| flhggfha18H08       | Chromosome_1:22975439-22977224 chr2:1617299-1619054 | Yes | No |
| GLAND2              | Chromosome_1:23010405-23011244 chr4:388212-389051 | Yes | Yes |
| flhggfha3B05        | Chromosome_1:23190203-23191077 chr2:213693-214571 | Yes | No |
| flhggfha3D11        | Chromosome_1:23196555-23198657 chr2:205749-207718 | Yes | Yes |
| flhggfha16B09       | Chromosome_1:231989842-23199786 #N/A | Yes | Yes |
| flhggfha16B09       | Chromosome_1:23200183-23201041 chr4:203379-204228 | Yes | No |
| flhggfha13A06       | Chromosome_2:11158343-11164210 #N/A | Yes | No |
| flhggfha6E07        | Chromosome_2:11538648-11539386 chr2:17153143-17154440 | No | No |
| flhggfha2E04        | Chromosome_2:15955547-15956802 chr4:505759-507043 | No | No |
| GLAND4              | Chromosome_2:18006498-18007341 chr2:10650950-10651793 | No | No |
| flhggfha23G12       | Chromosome_2:20777375-20798370 chr2:8047407-8067005 | No | No |
| flhggfha33E05       | Chromosome_2:20815132-20869509 chr2:8038615-8045899 | No | No |
| flhggfha34B08       | Chromosome_2:20863705-20865825 chr2:8064031-8123400 | No | No |
| esthggfha15A10      | Chromosome_2:3001552-3001788 chr9:4666570-4692682 | Yes | No |
| esthggfha15A10      | Chromosome_2:30099104-3020158 #N/A | Yes | No |
| GLAND7              | Chromosome_2:3052306-3057600 chr9:4636652-4639994 | No | No |
| flhggfha17G01       | Chromosome_2:637539-1206335 chr9:7372804-7374001 | Yes | No |
| flhggfha4D07        | Chromosome_2:8104908-8106196 #N/A | Missing | Yes |
| flhggfha32E03       | Chromosome_2:8910595-9214871 chr9:54514-55502 | Yes | No |
| flhggfha32E03       | Chromosome_2:8987579-9033715 chr26 pilon:115362-350944 | Yes | Yes |
| flhggfha13C08       | Chromosome_2:9967702-9970940 chr287 pilon:159998-164262 | No | Yes |
| GLAND9              | Chromosome_3:13612076-13612361 chr3:7616125-7616372 | No | No |

(Continues)
| Effector   | TN10 location            | X12 location    | TN10 duplication | X12 duplication | Secreted? | Secreted? |
|-----------|--------------------------|----------------|------------------|-----------------|-----------|-----------|
| flhggfha30C02 | Chromosome_3:14140055-14167132 | chr3:7305317-7305966 | Yes              | No              |
| flhggfha5D08   | Chromosome_3:14838109-14838979  | chr3:6572178-6572975  | No                | No              |
| flhggfha31A08   | Chromosome_3:20866442-20867481  | chr3:1130997-1132339  | No               | Yes             |
| flhggfha4F01   | Chromosome_3:38679834-8682148  | chr1:5889248-5891557  | No               | Yes             |
| flhggfha2A05   | Chromosome_3:38979144-8980194  | chr3:11880332-11881382 | Yes              | Yes             |
| flhggfha4F01   | Chromosome_3:9456452-9587578   | chr1:11969912-11972219 | Yes              | No              |
| flhggfha3H07   | Chromosome_3:39995145-9995746  | chr3:10927840-10928561 | No               | No              |
| GLAND3      | Chromosome_4:10048233-10049714 | chr1:11494423-11495904 | Yes              | Yes             |
| flhggfha17G06 | Chromosome_4:10115834-10116732 | chr1:11423138-11422214 | No               | No              |
| flhggfha17G06 | Chromosome_4:10142225          | #N/A                   | No               | No              |
| flhggfha26D05 | Chromosome_4:11770056-11773116 | chr1:13070201-13074976 | No               | No              |
| flhggfha30D08 | Chromosome_5:13907699-13908483 | chr1:455013-455561    | No               | No              |
| flhggfha30D08 | Chromosome_5:13928823-13929607 | #N/A                   | Yes              | No              |
| flhggfha17G01 | Chromosome_5:14135965-14136046 | #N/A                   | Yes              | No              |
| flhggfha33A09 | Chromosome_5:1716302-1717253   | chr2:1828161-1829151   | Yes              | No              |
| flhggfha17C07 | Chromosome_5:3269361-3271060   | chr2:1415540-1435621   | No               | No              |
| flhggfha19B10 | Chromosome_5:7409103-7409982   | chr6:6148639-6149518   | No               | Yes             |
| GLAND15     | Chromosome_5:7579174-7579472   | chr2:2082930-2088049   | Yes              | No              |
| GLAND14     | Chromosome_5:7579149-7579445   | chr2:6293813-6294664   | No               | No              |
| GLAND14     | Chromosome_5:7666023-7666874   | #N/A                   | Yes              | No              |
| GLAND14     | Chromosome_5:7783829-7798775   | #N/A                   | No               | No              |
| esthggfha8C06 | Chromosome_6:10658809-10659701 | chr5:10952046-10952940 | No               | No              |
| flhggfha4D09 | Chromosome_6:1375619-1377266   | #N/A                   | Yes              | No              |
| flhggfha5D06 | Chromosome_6:1665201-2317339   | chr5:422170-427526     | Yes              | No              |
| GLAND1      | Chromosome_6:2572115-2575451   | chr5:1059003-1061827   | No               | No              |
| GLAND1      | Chromosome_6:2749021-2751861   | chr5:926950-929793     | No               | No              |
| flhggfha2A12 | Chromosome_6:4522904-4579539   | chr5:3276655-3277393   | Yes              | No              |
| flhggfha11A06 | Chromosome_6:4578751-4579620   | chr5:3275497-3276365   | Yes              | No              |
| flhggfha11A06 | Chromosome_6:4579841-4580710   | chr4:12561870-12562738 | Yes              | No              |
| flhggfha2D01 | Chromosome_6:5139498-5140297   | chr5:3065476-3066261   | Yes              | No              |
| flhggfha2D01 | Chromosome_6:5150305-5256318   | chr5:3065639-3067426   | Yes              | Yes             |
| flhggfha11A06 | Chromosome_6:5582402-5583265   | #N/A                   | Yes              | Yes             |
| flhggfha11A06 | Chromosome_6:5583486-5584350   | chr5:3066537-3277437   | No               | Yes             |

(Continues)
2.6 | Differential gene expression

The strandedness of the RNA-seq was evaluated with rseqc V4.0 (Wang et al., 2012, 2016), followed by alignment to the genome with hisat (2.2.0) (Kim et al., 2015), and converted to bam with samtools (1.1.0) (Li et al., 2009). Read counts were calculated with FeatureCounts from subread package (1.6.0) (Liao et al., 2013), followed by deseq2 (1.20.0) (Love et al., 2014) with \( p \)-value cutoffs at 0.05 to determine differential expression between the samples.

2.7 | BUSCO analysis

Universal single copy orthologous genes were assessed using BUSCO 5.12 (Seppey et al., 2019; Simão et al., 2015; Waterhouse et al., 2017) on both the predicted proteins and the genome against the nematoda ODB10 data set. In genome mode, --augustus and --long options were used to survey genomes with the nematodea_odb10 data set. Revised missing counts were determined by taking total missing counts from the consensus of both genome and protein BUSCO scans, divided by the total number of BUSCOs available (3131–666 = 2465).

2.8 | Effector gene prediction

Effector proteins were mapped to the predicted proteome using diamond 0.9.23 (Buchfink et al., 2015). Effector genes were mapped to the genome using gmap (2018-03-25) (Wu et al., 2016). Secreted proteins were identified with signalp 5.0 (Armenteros et al., 2019). Secreted proteins were truncated at the signal peptide cleavage site and then subjected to transmembrane domain prediction with tmhmm 2.0 (Krogh et al., 2001).

2.9 | Repeat prediction

Multiple repeat predictions were pursued to finely detail genome structure. To comprehensively predict the structure of transposable
elements in the genome with Extensive de novo TE Annotator, EdTA
1.7 (Ou et al., 2019). Tandem Repeat Finder 4.0.9 (Benson, 1999) was run
on the genome to identify tandem repeats. A repeat prediction sensitive
to copy number variation was also pursued with RepeatModeler
1.0.11 (Smit et al., 2014) and RepeatMasker 4.0.9 (Smit et al., 2015).

2.10 | Synteny

Genome alignments were performed using Minimap2.2 (Li, 2016)
for display with Dotplotly (https://github.com/tpoor/dotPlotly). By deriving gene orthology from primary mapping sites of the pre-
dicted transcripts from our genome with gmap 2018-03-25 (Wu et al.,
2016), we inferred gene-based synteny with iAdHoRe 3.0.01 (Proost
et al., 2011) for display in Circos 0.69-6 (Krzywinski et al., 2009). Total
genomic proportions of synteny were inferred by subtracting the be-
inning from the end of each syntenic block generated by iAdHoRe,
and added together to get a final genomic proportion of synteny.

3 | RESULTS AND DISCUSSION

3.1 | Genome quality metrics

The H. glycines TN10 genome assembly comprises 2109 contigs, all
of which were incorporated into the expected nine pseudomolecule
scaffolds using the Juicer pipeline, in agreement with cytological
observations (Cotten, 1965; Lian et al., 2019) (Figure 1). Our goal
was to generate a genome using easy-to- replicate de novo methods
to facilitate comparisons with future H. glycines assemblies, as the
previous draft assembly utilized synteny among contigs to generate
scaffold-size contigs (Masonbrink et al., 2019). With this approach
we used Falcon to assemble and Falcon Unzip (Chin et al., 2016) to
remove putative haplotigs from the assembly. Dovetail genomics
scaffolded the genome with Chicago and HiC reads, reducing the
assembly to 509 scaffolds (Table S1). This assembly was scaffolded
further using Sspace (Boetzer & Pirovano, 2014) with PacBio sub-
reads, gaps were filled with gmcloser (Kosugi et al., 2015), and Pilon
(Walker et al., 2014) was used to polish prior to implementing the
Juicer pipeline. Using the Juicer pipeline (Dudchenko et al., 2017;
Durand et al., 2016; Lieberman-Aiden et al., 2009) we were able to
generate the expected nine pseudomolecules, although uncertainties
in small contig orientation necessitated an approach to fix these
potential misassemblies. To address this and potentially fill the gaps
created in the Juicer assembly, we polished the genome with Pilon
(Walker et al., 2014) using short reads, then 38 times with Pacbio
CCS reads, and another short read polish (Table S1).

The genome size of the new assembly was 157,982,452 bp,
within the expected range for this clade of species (Table 2); how-
ever, the genome size was 34 Mb larger than the previous TN10
assembly (Table 2). This increased genome size may be due to the
incorporation of repetitive haplotigs, an assumption supported by
the inflation of repeats (61.4 Mb) compared to the previous draft
assembly of TN10 (42.1 Mb) (Masonbrink et al., 2019). Total repeat
content (38.9%) was still within the published range (34%–47.7%),
and yet maintains a lower repeat content than the X12 assembly
(67.3 Mb) (Table S2) (Lian et al., 2019). Interestingly, the HiC heat-
map shows intense signals near what could represent the chromatin
compaction of H. glycines centromeres (Figure 1a); however, we did
not find a single repeat, but six different repeats enriched in these
regions, including mutator like elements, hAT-like elements, a gypsy
retroelement, and a putative MITE (Figure S1). Another interesting
phenomenon was the observation that the high density HiC signals
were associated with regions of high gene density (Figure 1B). While
we are unsure of the implications of this result, we infer that this
effect may be due to the holocentric nature of H. glycines chromo-
somes, or from the large population of individuals used to generate the
HiC sequencing.

To assess quality and completeness, the input sequences were
aligned to the assembly. High rates of alignment: 97.3%, 97.2%, and
73.5%, were observed for Pacbio subreads, Pacbio CCS reads, and
260 bp PE Illumina reads, respectively. To evaluate the genic com-
plement of the annotation, we ran Benchmarking Universal Single
Copy Orthologues (busco) 5.12 (Waterhouse et al., 2017) on both the
gene, its predicted proteins, and for all related species’ assem-
bies. Interestingly, of the 3031 possible Nematoda BUSCO genes,
666 genes from the Nematoda_odb10 data set were missing from all
surveyed Tylenchida genomes. Even among the three H. glycines ge-
nome assemblies, we found a consistent 813 missing BUSCO genes.
This is not necessarily surprising due to the underrepresentation of
plant parasitic nematodes in nematoda_odb10, though this finding
will improve the interpretation of genome quality for species in this
clade. While BUSCO protein complete scores were at 60.2% for the
new TN10 assembly, a subtraction of 666 genes from the total 3131
BUSCO genes reveal a clearer picture of BUSCO genes that are bio-
logically missing from the species assessed, and/or are hard to assem-
bles in this clade. Although complete scores across all species were
relatively low due to this fact, a proxied missing score could be com-
puted by dividing the missing total by the reduced number of BUSCO
genes possible (2466). This revealed that only 12.1% BUSCO genes
were missing in the new TN10 genome, though this same statistic was
applied to all surveyed species’ genomes, revealing that the G. rosto-
chensis genome assembly had the lowest rate of missing BUSCOs at
8.7%. Strangely, even though the TN10 genome improved in contigu-
ity, BUSCO scores declined, which may be indicative of the number of
gaps introduced to get all contigs incorporated into the pseudomole-
cules (Table 2). Overall, the high proportion of input read mapping, rel-
atively high BUSCO scores, and complete incorporation of all contigs,
suggests this latest H. glycines assembly is of high quality.

3.2 | Improvements over existing soybean cyst
tematode assemblies

This pseudomolecule assembly is a massive step forward in the
genomics of plant-parasitic nematodes, increasing the ability of
interspecies comparisons. To assess the contiguity and accuracy of our new assembly we used direct chromosome alignments using minimap2 and dotplotly, as well as gene-based synteny with i-ADHoRe (Proost et al., 2011) (Figure 2). To avoid contending with differences in gene prediction, we proxied synteny by mapping TN10 genes to the TN10 draft and the X12 genome. While 712 and 1760 genes did not align to the TN10 draft genome (Masonbrink et al., 2019) and the X12 genome (Lian et al., 2019), we found 65.6 Mb of synteny to 462 of TN10 draft genomes’ 738 contigs (Masonbrink et al., 2019) (Figure 2a,b). This number increases with the X12 genome, with the 80.6 Mb of synteny to the X12 genomes’ nine pseudomolecules and 19 unscaffolded contigs (Lian et al., 2019). We further investigated synteny by performing whole genome alignments with minimap2 and dotplotly, thereby verifying the scaffolding successes of the new TN10 draft and the X12 genome. While 712 and 1760 genes did not align to the TN10 draft genome (Masonbrink et al., 2019) and the X12 genome (Lian et al., 2019), we found 65.6 Mb of synteny to 462 of TN10 draft genomes’ 738 contigs (Masonbrink et al., 2019) (Figure 2a,b). This number increases with the X12 genome, with the 80.6 Mb of synteny to the X12 genomes’ nine pseudomolecules and 19 unscaffolded contigs (Lian et al., 2019). We further investigated synteny by performing whole genome alignments with minimap2 and dotplotly, thereby verifying the scaffolding successes of the new TN10 assembly (Figure 2a) and clarifying the differences in chromosomal structure between the TN10 and X12 pseudomolecule assemblies (Figure 2c,d). Considering that more than 61 Mb of repeats are in the TN10 genome (38.6%), synteny to 41.5% and 51% the genome in the TN10 draft and X12 assemblies, respectively, is high.

Assignment of contigs to chromosomes was improved in this assembly compared to existing X12 assembly. These differences resulted in the identification of a number of large chromosomal misjoins in the X12 assembly: including multiple interchromosomal translocations and the misassignment of chromosome 9 (Figure 2c,d; Table S3). Surprisingly, after adjusting for these large chromosomal misjoins in X12, there were very few chromosomal rearrangements between the two lines of a highly adaptable species (Figure 2c,d).

3.3 | Gene annotation

The gene annotation resulted in 22,465 gene models, encoding 23,933 transcripts with an average gene length of 4569 bp, values that are comparable to related species (Table 3). While the frequency of genes is substantially larger than the previously published X12 annotation (11,882), the propensity for parasites to duplicate genes involved in host-parasite interactions requires a novel approach to gene prediction. To prevent the obliteration of parasitism genes thought to be maintained at high copies in the nematode, we developed an annotation approach to predict all transcribed elements in the genome, including repetitive elements. A genome without repeat masking was used to allow highly similar, high-copy number genes to be identified. However, because repetitive elements frequently reside in noncoding regions of genes, multiple genome-guided transcriptomes and gene predictions enabled the dissection of high-confidence gene models (Table S4). This improvement in gene prediction is indicated by our total gene count (22,265), a figure similar to the previous TN10 draft when considering that 9000 repetitive elements were among the 29,769 genes examined in the previous assembly (Masonbrink et al., 2019). Our analyses included known parasitism genes (see Methods) and repeats missing from X12 (11,882) and produced a more highly
|                          | **Heterodera glycines** | **Heterodera glycines** | **Globodera rostochiensis** | **Globodera pallida** | **Globodera ellingtonae** | **Meloidogyne hapla** | **Meloidogyne incognita** |
|--------------------------|-------------------------|-------------------------|-----------------------------|----------------------|--------------------------|----------------------|--------------------------|
| **Number of scaffolds**  | 9                       | 738                     | 267                         | 4281                 | 6873                     | 2246                 | 3452                     | 12,091                  |
| **Number of contigs**    | 2109                    | 738                     | 888                         | 17,236               | 18,270                   | 13,843               | 3452                     | 13,148                  |
| **Percent scaffolded**   | 100.0%                  | N/A                     | 92.6%                       | 94.2%                | 90.1%                    | 94.8%                | N/A                      | 25.0%                   |
| **Genome size (base pairs)** | 157,982,452             | 123,847,574             | 141,354,287                 | 95,876,286           | 123,625,196              | 105,964,814          | 53,017,507               | 183,531,997             |
| **N50 scaffold length (base pairs)** | 17,908,190             | 304,127                 | 16,265,615                  | 88,688               | 120,481                  | 327,189              | 37,608                   | 62,516                  |
| **L50 scaffolds**        | 4                       | 109                     | 4                           | 278                  | 296                      | 85                   | 372                      | 1209                    |
| **N's (%)**              | 1.06%                   | 0.00%                   | 0.73%                       | 4.61%                | 16.17%                   | 0.85%                | 0.00%                    | 1.81%                   |
| **Gaps**                 | 4328                    | 0                       | 1244                        | 47,076               | 35,940                   | 26,075               | N/A                      | 10,490                  |
| **BUSCO Proteins**       | C:78.0% [S:69.0%, D:9.0%, F:7.8%, M:14.2%, n:255] | C:89.0% [S:58.8%, D:5.9%, F:13.7%, M:21.6%, n:255] | C:63.2% [S:62.4%, D:0.6%, F:18.4%, M:18.4%, n:255] | C:48.2% [S:43.9%, D:4.3%, F:22.4%, M:29.4%, n:255] | C:59.6% [S:59.6%, D:0%, F:18.4%, M:22.0%, n:255] | C:65.1% [S:64.7%, D:0.4%, F:16.9%, M:18.0%, n:255] | C:70.2% [S:18.0%, D:52.2%, F:12.2%, M:17.6%, n:255] |
| **BUSCO Genome**         | C:55.6% [S:52.5%, D:3.1%, F:2.3%, M:42.1%, n:3131] | C:55.1% [S:49.0%, D:3.2%, F:1.8%, M:46.0%, n:3131] | C:60.0% [S:58.9%, D:1.1%, F:1.9%, M:38.1%, n:3131] | C:46.7% [S:43.5%, D:3.2%, F:2.5%, M:50.8%, n:3131] | C:54.4% [S:52.8%, D:1.6%, F:2.4%, M:43.2%, n:3131] | C:45.4% [S:44.5%, D:0.9%, F:4.7%, M:49.9%, n:3131] | C:63.6% [S:24.0%, D:39.6%, F:1.0%, M:35.4%, n:3131] |
| **BUSCO Proteins**       | C:60.2% [S:51.5%, D:8.7%, F:1.5%, M:38.3%, n:3131] | C:66.5% [S:49.4%, D:4.5%, F:2.2%, M:43.9%, n:3131] | C:66.7% [S:64.3%, D:2.4%, F:2.1%, M:31.2%, n:3131] | C:45.6% [S:41.9%, D:3.7%, F:2.8%, M:51.6%, n:3131] | C:65.4% [S:56.5%, D:8.9%, F:2.0%, M:32.6%, n:3131] | C:59.6% [S:58.5%, D:1.1%, F:2.2%, M:38.2%, n:3131] | C:68.5% [S:13.9%, D:54.6%, F:1.7%, M:29.8%, n:3131] |
| **BUSCO Max Missing**    | 299/2465 (12.1%)        | 274/2465 (11.1%)        | 476/2465 (19.3%)            | 215/2465 (8.7%)      | 711/2465 (28.8%)         | 300/2465 (11.4%)     | 384/2465 (15.6%)         | 223/2465 (9%)           |
| **References**           | N/A                     | Masonbrink et al. (2019) | Lian et al. (2019)          | Eves-van den Akker et al. (2016) | Cotton et al. (2014) | Phillips et al. (2017) | Opperman et al. (2008) | Abad et al. (2008)       |
contiguous genome with fewer fragmented genes than the previous TN10 assembly (29,769). Our average gene and transcript lengths are the largest among the compared species, while exon count per transcript has also increased relative to earlier annotations of *H. glycines* and other related species (Table 3). Another line of evidence to support these gene predictions lies with the high proportion of genes that have functional annotations with 85.2% of predicted proteins or transcripts having homology to sequences in Interpro, Swissprot, NCBI NR, or NCBI NT databases (Tables S5 and S6).

### 3.4 Effector gene prediction

With a high-quality pseudomolecule genome, we can now better understand the molecules that are secreted from the parasite into the host. Though signal peptides provide a reliable method to identify proteins entering the secretory pathway, a number of proteins with nonclassical secretion will be overlooked (Bendtsen et al., 2004; Gahoi & Gautam, 2017; Nielsen et al., 2019). However, to illustrate what can be predicted, we identified transcripts that produce transmembrane-lacking proteins with signal peptides. We identified 1514/23,933 transcripts that produce a secreted protein, originating from 1421 of the 22,465 genes (Table 4). A second step to identifying these molecules lies with attributing the previously published effectors to genes in the genome (Gao et al., 2003; Noon et al., 2015). Using DNA sequence similarity to previously published effectors, 100 potential effector genes were identified in the TN10 genome with a minimum query alignment and sequence identity of 50% (Table 1). We utilized the same parameters to discover effectors in the X12 genome, identifying 98 effector genes and the absence of 45D07 and 4D09 effectors (Table 1). Shifts in the frequency of effector duplication were evident among the two lines, with 12 effector duplications in TN10 and 14 duplications in X12 (Table 1). To expand this putative effector data set, we identified effectors in the predicted proteome of TN10. However, only 43/100 and 25/98 of these putative effectors encode secreted proteins (Table 1), indicating that genes may be variable among SCN lines in their propensity to produce secreted proteins, a variability that may contribute to SCN virulence.

### 3.5 Expression of effector genes

In the hope of further resolving the genes important to the host-parasite exchange, we utilized existing *H. glycines* RNA-seq (SRP122521). All possible comparisons were made between pre-parasitic (i.e., before root penetration) second-stage juvenile nematodes (PP), second-stage parasitic (i.e., after root penetration) nematodes on a susceptible host (C for compatible), and second-stage parasitic nematodes on a resistant host (IC for incompatible).
MASONBRINK et al.

Gene, transcript, and exon statistics for the TN10 pseudomolecule assembly and related species. Total represents the total base pairs of every unit (gene, transcript, exon) in the genome. While count refers to the number of units in the genome.

| Species          | Genes         | Transcripts | Exons        |
|------------------|---------------|-------------|--------------|
| H. glycines     | Total: 71,832,760 | Total: 28,809,923 | Total: 16,809,581 |
| Draft           | Count: 32,271 | Count: 34,904 | Count: 28,809,023 |
| Mean: 4,334     | Mean: 4,549   | Mean: 4,344   | Mean: 192,837  |
| Max: 1,940      | Max: 1,099    | Max: 3,094    | Max: 6154      |
| H. rostochiensis| Total: 15,011,770 | Total: 1,846,821 | Total: 2,966,348 |
| Count: 16,419   | Count: 2,724  | Count: 1,946  | Count: 4,569   |
| Mean: 1,845     | Mean: 1,496   | Mean: 1,946   | Mean: 4,569    |
| Max: 978        | Max: 2,229    | Max: 2,229    | Max: 2,229     |
| G. pallida      | Total: 24,938,405 | Total: 2,874,348 | Total: 2,966,348 |
| Count: 20,704   | Count: 3,487  | Count: 2,966  | Count: 4,569   |
| Mean: 1,940     | Mean: 1,496   | Mean: 1,946   | Mean: 4,569    |
| Max: 55,968     | Max: 2,229    | Max: 2,229    | Max: 2,229     |
| G. elongatae    | Total: 17,623,861 | Total: 2,434,348 | Total: 2,966,348 |
| Count: 14,314   | Count: 5,434  | Count: 2,966  | Count: 4,569   |
| Mean: 1,940     | Mean: 1,496   | Mean: 1,946   | Mean: 4,569    |
| Max: 978        | Max: 7,777    | Max: 2,229    | Max: 2,229     |
| M. hapla        | Total: 25,962,686 | Total: 2,434,348 | Total: 2,966,348 |
| Count: 14,314   | Count: 5,434  | Count: 2,966  | Count: 4,569   |
| Mean: 1,940     | Mean: 1,496   | Mean: 1,946   | Mean: 4,569    |
| Max: 978        | Max: 7,777    | Max: 2,229    | Max: 2,229     |
| M. incognita    | Total: 24,938,405 | Total: 2,874,348 | Total: 2,966,348 |
| Count: 20,704   | Count: 3,487  | Count: 2,966  | Count: 4,569   |
| Mean: 1,940     | Mean: 1,496   | Mean: 1,946   | Mean: 4,569    |
| Max: 55,968     | Max: 2,229    | Max: 2,229    | Max: 2,229     |

**Table 3** Gene, transcript, and exon statistics for the TN10 pseudomolecule assembly and related species. Total represents the total base pairs of every unit (gene, transcript, exon) in the genome, while count refers to the number of units in the genome.

Expression during the host-parasite exchange.

**3.6 Conclusion**

In summary, we present the most complete *H. glycines* assembly, with a consensus gene prediction pipeline sensitive to the prediction of high-copy parasitism-related genes. We confirm this with a high percentage of synteny to previous assemblies, high read mapping rates, and the complete integration of all contigs into nine pseudomolecules. Using currently available data, we compiled a comprehensive resource that extensively annotates *H. glycines* genes, a critical resource for the development of advanced technology to combat this pest. This resource will be integrated into SCNBase.org, which further extends the transparency and availability of *H. glycines* genomic data. To demonstrate the utility of this new resource, we assessed the conservation of previously published effectors and leveraged published RNAseq and gene features to further explore effector expression during the host-parasite exchange.
AUTHOR CONTRIBUTIONS
Conceptualization – R.E.M., A.J.S., T.R.M., T.B.; Data curation – R.E.M., M.H.; formal analysis – R.E.M., A.J.S., T.R.M.; funding acquisition – A.J.S., T.R.M., T.B.; investigations – R.E.M., A.J.S., T.R.M., T.B.; methodology – R.E.M., T.R.M., A.J.S., T.B.; resources – A.J.S., T.R.M., M.H., T.B.; Software – R.E.M., A.J.S.; Validation – R.E.M.; Visualization – R.E.M.; Writing – R.E.M.; Review and Editing – R.E.M., T.R.M., M.H., A.J.S., T.B.

DATA AVAILABILITY STATEMENT
All scripts and notes used to prepare this genome are available at https://github.com/ISUgenomics/Dovetail2SCNGenome. The genome, annotation, and Hi-C reads were uploaded to Genbank and SRA under the Bioproject PRJNA603076 and SRR8381095. All genome track data and annotations will also be hosted on SCNBase.org. Once released, genomes and annotations will be uploaded to ParaSite.WormBase.org.

ORCID
Rick E. Masonbrink https://orcid.org/0000-0002-6848-2144

REFERENCES
Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G., J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nature Biotechnology, 26(8), 909. https://doi.org/10.1038/nbt.1482

Abad, P., & Williamson, V. M. (2010). Plant nematode interaction: a sophisticated dialogue. In J.-C. Kader & M. Delseny (Eds.), Advances in Botanical Research (Vol., 53, pp. 147–192). Elsevier.

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology, 37(4), 420–423.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prijibelski, A. D., Pyshkin, A. V., Sirokin, A. V., Yadah, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bekal, S., Niblack, T. L., & Lambert, K. N. (2003). A chorismate mutase from the soybean cyst nematode Heterodera glycines shows polymorphisms that correlate with virulence. Molecular Plant-Microbe Interactions, 16(5), 439–446.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. Protein Engineering Design and Selection, 17(4), 349–356. https://doi.org/10.1099/protein/gzn037

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Research, 27(2), 573–580. https://doi.org/10.1093/nar/27.2.573
Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics, 15(1), 211. https://doi.org/10.1186/1471-2105-15-211

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59–60. https://doi.org/10.1038/nmeth.3176

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research, 18(1), 188–196. https://doi.org/10.1101/gr.6743907

Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Elling, A. A., & Jones, J. T. (2014). Functional characterization of nematode effector proteins cleaved in planta to promote plant parasitism. The Plant Journal, 74(2), 185–196.

Consortium, U. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506–D515.

Coordinators, N. R. (2016). Database resources of the national center for biotechnology information. Nucleic Acids Research, 44(Database issue), D7.

Cotten, J. (1965). Cytological investigations in the genus heterodera. Nematologica, 11(3), 337–342. https://doi.org/10.1163/187529265X00249

Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., Tjaden, T. J., & Wang, X. (2013). A ubiquitin carboxyl extension protein secreted from a plant-plantarastic nematode Globodera rostochiensis is cleaved in planta to promote plant parasitism. The Plant Journal, 74(2), 185–196.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., & Aiden, A. P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, 356(6333), 92–95.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Lander, E. S., & Aiden, A. P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, 356(6333), 92–95.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Lander, E. S., & Aiden, A. P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, 356(6333), 92–95.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., & Aiden, A. P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science, 356(6333), 92–95.

Elling, A. A., & Jones, J. T. (2014). Functional characterization of nematode effectors in plants. In P. Birch, J. Jones & J. Bos (Eds.), Plant-pathogen interactions. Methods in Molecular Biology (Vol. 1127, pp. 113–124). Totowa, NJ: Humana Press.

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G. J., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szütsenberk, A., Grenier, E., Montarry, J., Míme, B., Duceppe, M.-O., Boyes, I., Marvin, J. M. C., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., & Jones, J. T. (2016). The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence. Genome Biology, 17(1), 124. https://doi.org/10.1186/s13059-016-0985-1

Eves-van den Akker, S., Lilley, C. J., Jones, J. T., & Urwin, P. E. (2014). Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. PLoS Path, 10(9), e1004391. https://doi.org/10.1371/journal.ppat.1004391

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., & Fraser, M. (2016). InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Research, 45(D1), D190–D199

Gao, S., & Gautam, B. (2017). Genome-wide analysis of Excretory/Secretory proteins in root-knot nematode, Meloidogyne incognita provides potential targets for parasite control. Computational Biology and Chemistry, 67, 225–233. https://doi.org/10.1016/j.compbiolchem.2017.01.014

Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J., & Hussey, R. S. (2001). Identification of putative parasitism genes expressed in the esophageal gland cells of the soybean cyst nematode Heterodera glycines. Molecular Plant-Microbe Interactions, 14(10), 1247–1254.

Gao, B., Allen, R., Maier, T., Davis, E. L., Baum, T. J., & Hussey, R. S. (2003). The parasitism of the phytonematode Heterodera glycines. Molecular Plant-Microbe Interactions, 16(8), 720–726.

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

Haegeman, A., Mantelin, S., Jones, J. T., & Gheyse, G. (2012). Functional roles of effectors of plant-parasitic nematodes. Gene, 492(1), 19–31. https://doi.org/10.1016/j.gene.2011.10.040

Hamamouch, N., Li, C., Hewezi, T., Baum, T. J., Mitchum, M. G., Hussey, R. S., Vodkin, L. O., & Davis, E. L. (2012). The interaction of the novel 30C02 cyst nematode effector protein with a plant β-1,3-endoglucanase may suppress host defence to promote parasitism. Journal of Experimental Botany, 63(10), 3683–3695. https://doi.org/10.1093/jxb/ers058

Henschel, R., Lieber, M., Wu, L.-S., Nista, P. M., Haas, B. J., & LeDuc, R. D. (2012). Trinity RNA-Seq assembler performance optimization. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond.

Hewezi, T. (2015). Cellular signaling pathways and posttranslational modifications mediated by nematode effector proteins. Plant Physiology, 169(2), 1018–1026. https://doi.org/10.1104/pp.115.00923

Hewezi, T., & Baum, T. J. (2013). Manipulation of plant cells by cyst and root-knot nematode effectors. Molecular Plant-Microbe Interactions, 26(1), 9–16. https://doi.org/10.1094/MPMI-05-12-0106-FI

Hewezi, T., Howe, P. J., Maier, T. R., Hussey, R. S., Mitchum, M. G., Davis, E. L., & Baum, T. J. (2010). Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode Heterodera schachtii. Plant Physiology, 152(2), 968–984.

Hoff, K. J., Lange, S., Lomaszde, A., Borodovsky, M., & Stanke, M. (2015). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics, 31(5), btv661.

Hoff, K. J., Lomaszde, A., Stanke, M., & Borodovsky, M. (2018). BRAKER2: incorporating protein homology information into gene prediction with GeneMark-EP and AUGUSTUS. Plant and Animal Genomes XXVI.

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., & Berriman, M. (2017). WormBase ParaSite—A comprehensive resource for helminth genomics. Molecular and Biochemical Parasitology, 215, 2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005

Hunt, V. L., Tsai, I. J., Coghlan, A., Reid, A. J., Holroyd, N., Foth, B. J., Tracey, A., Cotton, J. A., Stanley, E. J., Beasley, H., Bennett, H.
new candidate effectors of the phytomastomate Heterodera glycines produced specifically in the secretory esophageal gland cells during parasitism. *Phytopathology*, 105(10), 1362–1372.

Olsen, A. N., & Skriver, K. (2003). Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. *Trends in Plant Science*, 8(2), 55–57. https://doi.org/10.1016/S1360-1385(03)00003-7

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Ou, S., Su, W., Liao, Y. I., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Dere, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, 20(1), 1–18. https://doi.org/10.1186/s13059-019-1905-y

Patel, N., Hamamouch, N., Li, C., Hewezi, T., Husse, R. S., Baum, T. J., Mitchum, M. G., & Davis, E. L. (2010). A nematode effector protein similar to annexins in host plants. *Journal of Experimental Botany*, 61(1), 235–248. https://doi.org/10.1093/jxb/erp293

Pertea, M., Kim, D., Pertea, G. M., Lee, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT. *StringTie and Ballgown. Nature Protocols*, 11(9), 1650. https://doi.org/10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290. https://doi.org/10.1038/nbt.3122

Phillips, W. S., Howe, D. K., Brown, A. M., Eves-Van Den Akker, S., Dettwyler, L., Peetz, A. B., Denver, D. R., & Zasada, I. A. (2017). The Draft Genome of Globodera elliottiae. *Journal of Nematology*, 49(2), 127. https://doi.org/10.2137/jofnem-2017-054

Pogorelko, G., Juvala, P. S., Rutter, W. B., Hewezi, T., Husse, R., Davis, E. L., Mitchum, M. G., & Baum, T. J. (2016). A cyst nematode effector binds to diverse plant proteins, increases nematode susceptibility and affects root morphology. *Molecular Plant Pathology*, 17(6), 832–844. https://doi.org/10.1111/mpp.12330

Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., & Vandepoele, K. (2011). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *GigaScience*, 44(9), giy093. https://doi.org/10.1111/j.1365-313X.2010.04433.x

Wang, J., Lee, C., Replogle, A., Joshi, S., Korkin, D., Husse, R., Baum, T. J., Davis, E. L., Wang, X., & Mitchum, M. G. (2010). Dual roles for the variable domain in protein trafficking and host-specific recognition of Heterodera glycines CLE effector proteins. *New Phytologist*, 187(4), 1003–1017.

Wang, J., Replogle, A., Husse, R., Baum, T., Wang, X., Davis, E. L., & Mitchum, M. G. (2011). Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode Heterodera schachtii. *Molecular Plant Pathology*, 12(2), 177–186. https://doi.org/10.1111/j.1364-3730.2010.00660.x

Wang, L., Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J. E., Dasari, S., Vedell, P. T., Barman, P., Wang, L., Weinsiboum, R., Jen, J., Huang, H., Kohl, M., & Kocher, J.-P. A. (2016). Measure transcript integrity using RNA-seq data. *BMC Bioinformatics*, 17(1), 58. https://doi.org/10.1186/s12859-016-0922-z

Wang, L., Wang, S., & Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, 28(16), 2184–2185. https://doi.org/10.1093/bioinformatics/bts356
Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., de Boer, J. M., Baum, T. J., Hussey, R. S., & Davis, E. L. (2001). Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions*, 14(4), 536–544.

Wang, X., Mitchum, M. G., Gao, B., Li, C., Diab, H., Baum, T. J., Hussey, R. S., & Davis, E. L. (2005). A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. *Molecular Plant Pathology*, 6(2), 187–191.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3), 543–548. https://doi.org/10.1093/molbev/msx319

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. In Mathé, E., & S. Davis (2016). *Statistical Genomics: Methods and Protocols* (pp. 283–334). Springer New York.

Zheng, J., Peng, D., Chen, L., Liu, H., Chen, F., Xu, M., Ju, S., Ruan, L., & Sun, M. (2016). The Ditylenchus destructor genome provides new insights into the evolution of plant parasitic nematodes. *Proceedings of the Royal Society B: Biological Sciences*, 283(1835), 20160942.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

_How to cite this article:_ Masonbrink, R. E., Maier, T. R., Hudson, M., Severin, A., & Baum, T. (2021). A chromosomal assembly of the soybean cyst nematode genome. *Molecular Ecology Resources*, 00, 1–16. https://doi.org/10.1111/1755-0998.13432