



Original article

SCNBase: a genomics portal for the soybean cyst nematode (*Heterodera glycines*)

Rick Masonbrink¹, Tom R. Maier², Arun S. Seetharam¹, Parijat S. Juvele², Levi Baber³, Thomas J. Baum² and Andrew J. Severin^{1,*}

¹Genome Informatics Facility, Iowa State University, Osborne Dr, Ames, IA 50011, USA, ²Department of Plant Pathology and Microbiology, Iowa State University, Pammel Dr, Ames, IA 50011, USA and ³Research IT, Iowa State University, Osborne Dr, Ames, IA 50011, USA

*Corresponding author: Phone: 5152941320; Fax: 5152941337; E-mail: severin@iastate.edu

Citation details: Masonbrink,R., Maier,T.R., Seetharam,A.S. *et al.* SCNBase: a genomics portal for the soybean cyst nematode (*Heterodera glycines*). *Database* (2019) Vol. 2019: article ID baz111; doi:10.1093/database/baz111

Received 29 May 2019; Revised 22 July 2019; Accepted 9 August 2019

Abstract

Soybean is an important worldwide crop, and farmers continue to experience significant yield loss due to the soybean cyst nematode (SCN), *Heterodera glycines*. This soil-borne roundworm parasite is rated the most important pathogen problem in soybean production. The infective nematodes enter into complex interactions with their host plant by inducing the development of specialized plant feeding cells that provide the parasites with nourishment. Addressing the SCN problem will require the development of genomic resources and a global collaboration of scientists to analyze and use these resources. SCNBase.org was designed as a collaborative hub for the SCN genome. All data and analyses are downloadable and can be analyzed with three integrated genomic tools: JBrowse, Feature Search and BLAST. At the time of this writing, a number of genomic and transcriptomic data sets are already available, with 43 JBrowse tracks and 21 category pages describing SCN genomic analyses on gene predictions, transcriptome and read alignments, effector-like genes, expansion and contraction of genomic repeats, orthology and synteny with related nematode species, Single Nucleotide Polymorphism (SNPs) from 15 SCN populations and novel splice sites. Standard functional gene annotations were supplemented with orthologous gene annotations using a comparison to nine related plant-parasitic nematodes, thereby enabling functional annotations for 85% of genes. These annotations led to a greater grasp on the SCN effectorome, which include over 3324 putative effector genes. By designing SCNBase as a hub, future research findings and genomic resources can easily be uploaded and made available for use by others with minimal needs for further curation. By providing these resources to nematode research community, scientists will be empowered to develop novel, more effective SCN management tools.

Database URL: <https://scnbase.org/>

Introduction

Worldwide, soybean serves as a major nutritive component of human and livestock diets, providing high-quality protein and oil. Demand for soybean is likely to continue to rise with increasing populations and recent shifts in dietary trends. It is essential to maximize soybean production to support global food security. Yet, many biotic and abiotic stresses continually impact soybean yield worldwide. The most prominent of these is caused by a soil-borne parasitic worm *Heterodera glycines*, the soybean cyst nematode (SCN). SCN is the most economically damaging soybean pest, causing upwards of \$1.2 billion dollars of yield loss each year in the USA (1). SCN management has traditionally relied upon a combination of crop rotation, the use of SCN-resistant soybean varieties and nematicides. However, the current practice is losing its potency. SCN-resistant soybean varieties are increasingly becoming vulnerable to SCN infections, as genetic variability, numerous progeny and a relatively short life cycle have enabled the evolution of SCN populations to overcome resistant soybean varieties. The narrow genetic basis of soybean's SCN resistance and the overuse of these varieties have exacerbated the problem by stimulating the selection of virulent SCN populations throughout many soybean production areas (2–8). Thus, there is an immediate need to identify and develop novel management approaches to combat this pest.

SCN has evolved complex and fine-tuned molecular interactions with the soybean plant that culminate into the long-term suppression of effective defense responses and the development of a feeding structure in the soybean root, the syncytium. The SCN lifecycle begins with second-stage juveniles (J2) hatching from eggs in the soil and migrating toward host roots. J2s destructively enter and migrate through the root to find a preferred feeding location, usually in the vicinity of the vascular tissue. Once found, the nematode becomes sedentary and elicits host cell reprogramming to create cells that sequester nutrients from the vascular system, a process likely initiated by delivering secreted proteins (effectors) into the plant tissue using its hollow mouth spear, the stylet (1–8). Many long-standing questions in the biology of this parasite remain. It is particularly important to identify and functionally characterize all SCN effectors and to understand their variability among nematode populations with differing levels of virulence. Such analyses likely will reveal the genetic components responsible for SCN's ability to overcome defense mechanisms in previously resistant soybean cultivars. In pursuit of these questions, a number of diverse secreted effector proteins have been identified that enable SCN to (6, 9–16) mimic host regulatory peptides (17, 18), to affect hormone transport and signaling (19, 20), to modify polyamine

metabolism (21), to suppress the host immune system (6, 22–24), remodel host chromatin (25), to modify cell walls (26) and to utilize the host proteolytic pathway (27). Thus, understanding the genetic basis for these SCN phenotypes will better inform soybean researchers and breeders on how to develop soybean cultivars with lasting SCN resistance.

The assembly of an SCN genome sequence has enabled the application of a number of new omics technologies that may prove insightful to SCN research. While WormBase (28) and ParaSite (29) websites are great resources for all currently sequenced nematode genomes, gene annotations and other general omics data, species-specific databases are needed to mitigate the deluge of data detailing the intricacies of the host–parasite relationships. To fully exploit such a wealth of data, we created SCNBase, a globally accessible web portal to centralize SCN omics. We incorporated all currently available public data into SCNBase, including the supporting data from the SCN genome and transcriptome (30, 31). Many pre-existing analyses are available, including gene predictions, promoters, raw data alignments, genomic SNPs from 15 populations, synteny and ortholog families for related species, mitochondrial genome annotations, functional protein domains and effector predictions and alignments.

We improved existing genomic resources by including new analyses to improve current genome annotations. The functional annotation of genes in draft genomes can be extremely low compared to model species (32). However, we addressed this issue in the SCN genome, as exceptional detail has been added to gene annotations that uniquely incorporates orthologous annotations from nine related plant-parasitic nematodes. Repeat expansions and contractions were further surveyed using RepeatExplorer, adding depth to repeat annotations that may have been removed during the purging of heterozygous scaffolds during assembly. SCNBase will continue to evolve by facilitating community involvement and curation through an interface that is easy to use, view and search. Our goal with SCNBase is to bridge the gap between SCN and soybean researchers by increasing collaboration to better understand the extant diversity and adaptive mechanisms of SCN.

Materials and Methods

Data sets

Genomes, proteins and GFF files were downloaded from WormBase (29) for *Globodera rostochiensis* (33), *Globodera pallida* (34), *Meloidogyne hapla* (35), *Meloidogyne incognita* (36), *Meloidogyne floridensis* (37), *Meloidogyne javanica* (37), *Ditylenchus destructor* (38) and *Bursaphelenchus xylophilus* (39). The *Globodera ellingtonae* (40)

genome was downloaded from NCBI, and gene annotations were created in the *H. glycines* genome (31).

Ortholog-based annotation

Proteins were renamed using Maker scripts (41). Orthologous proteins were identified using Orthofinder (42) under default conditions using primary isoforms only. Once ortholog families were established, all annotations from each protein were extracted using custom Bash scripts. The custom ortholog-based annotations were then counted using matches to the column nine annotations of each GFF file and then sorted by predominance. Scripts are available at <https://github.com/remkv6/SCNBase> and SCNBase.org.

RepeatExplorer

H. glycines strain TN20 150 bp PE DNA-seq reads were downloaded from NCBI (SRR1800548) trimmed with Trimmomatic (43), converted to fasta with the FASTX-Toolkit (44), sorted and interlaced using Seqtk (45). After trimming adapters and low-quality sequence, 200 000 read pairs were retained that maintained the maximum length (150 bp). Classified repeats were also obtained using RepeatModeler (46) on the *H. glycines* genome (31). These reads were submitted to Repeatexplorer.org webserver (47) with paired = true, rename sequences = true, min overlap for clustering = 55, cluster size threshold for detailed analysis = 0.002, RepeatMasker database = Metazoa, custom repeat database = true, sequence filtering = false and minimal overlap for assembly = 50. The fasta output from RepeatExplorer was aligned to the genome using Gmap (default parameters) (48).

Database development

SCNBase uses GMOD (49), Tripal (50) and Drupal's luggage extension to manage the genomics data and empower efficient web development by multiple users. Tripal is based on a community-derived database schema named Chado (51, 52) and employs the use of controlled vocabularies such as Sequence Ontology (53), Gene Ontology (54) and others to ensure standardization of data storage.

Results and Discussion

Homepage

Because a functional, user-friendly and attractive homepage is likely to garner more interest from the end-users, we created a simple interface (Figure 1). The menu at the top of the homepage includes six choices: 'Content By Category', 'People', 'Research', 'Learn', 'Downloads' and 'Tools'. The

'Content By Category' menu provides quick links to categories of data on the website, including 'Comparative Biology', 'Effector Biology', 'Omics', 'Research', 'Statistics' and 'Tools'. The 'People' choice describes the researchers currently collaborating in maintenance and database curation. The 'Research' link provides introductory information on current research in the SCN community. 'Learn' provides a link to frequently asked questions, as well as the possibility to ask a question. The 'Downloads' option directs the user to a list of downloadable genomics content, while 'Tools' links the user to all available genomics tools.

The information embedded within the 'Content By Category' menu details different genomic analyses and provides supporting information for alignment tracks. 'Comparative Biology' leads the user to synteny and orthology data and experiments, while 'Effector Biology' provides data that are relevant to effector/parasite research. The 'Omics' category contains all general genomics analyses, including repeats, genome contamination, mitochondrial sequences, genomic structures, gene subsets and novel splicing. 'Research' leads the user to active and ongoing SCN research projects. 'Statistics' encompasses numerical data on the genome assembly and annotation, and 'Tools' leads the user to available genomics tools.

The homepage is designed to draw the eye by displaying different SCN life stages, 'parasitized soybean roots', 'egg-filled cyst', 'eggs' and 'infecting J2', but also supplies the user with immediate access to a number of tools. The top right of the page hosts a search bar integrated with all site content, including research projects, people, tools, frequently asked questions, file metadata and functional gene annotations. The most frequently used tools are present at the lower end of the homepage: JBrowse, BLAST and Feature Search. The JBrowse genome browser comes with a comprehensive selection of informative and downloadable tracks (Table 1). BLAST is available to search the genome, CoDing Sequence (CDS) and predicted proteome using all four BLAST algorithms: BLASTn, BLASTx, BLASTp and tBLASTn. The Feature Search tool provides an avenue to investigate the gene features and functions of SCN gene annotations, including sequences, functional cross-references, orthologous gene annotations and gene ontology. Altogether, with the simplicity of the layout, the far-reaching functionality of the search tool, and the in-depth list of tutorials, we hope to inspire users to use SCNBase for their research.

Genomic resources available on SCNbase

The recently published *H. glycines* TN10 genome (31) provides a foundation for 43 JBrowse tracks and 21 category pages detailing genomic analyses/resources of gene pre-

IOWA STATE UNIVERSITY

SCNBBase

Search

Content By Category ▾ People Research Learn Downloads Tools

BLAST

Feature Database

JBrowse

IOWA STATE UNIVERSITY

Genome Informatics Facility
207 Science
2237 Osborn Dr Ames IA
50011-1027
gifhelp@iastate.edu

Twitter
Site Index
Site map

Copyright © 2019
Iowa State University of Science
and Technology. All rights
reserved.

Non-discrimination Policy
Privacy Policy
Digital Access & Accessibility

Figure 1. SCNBBase.org homepage. The homepage provides a quick access to the most frequently used functions, bioinformatics tools and search functions. Content pages are found in the 'Content by Category' links.

diction, genomic structure, transcripts, genomic and transcriptional read alignments, effector predictions, repeats, orthologs, synteny, SNPs and novel splicing. Some of the most powerful public expression data sets are present with 230 million RNA-seq reads enabling the scrutinization of gene expression changes across pre-parasitic and parasitic nematodes on resistant and susceptible soybean cultivars (30). A long-read transcript (IsoSeq) data set improves gene models and enables the confident assessment of splicing isoforms across multiple life stages in virulent and avirulent populations (31).

Expanding genomic data sets

The diversity within the phylum Nematoda, and speed with which nematodes evolve, results in large numbers of novel

and unannotated genes. To fully exploit these resources and identify genes involved in parasitism, we created a robust gene annotation that leverages numerous databases and related species. Interproscan identified ontological terms and domains associated with 22 718 genes/proteins. Functional information was also obtained for 20 305 genes using orthology-based annotations to 9 sister species. Orthologous families were assigned with Orthofinder, and gene families were assigned a count of every unique gene function (Figure 2). This analysis adds depth to the annotations by capitalizing on the unique functional annotations of nine related *Tylenchia* species (*B. xylophilus*, *D. destructor*, *M. hapla*, *M. incognita*, *M. arenaria*, *M. javanica*, *M. floridensis*, *G. pallida* and *G. rostochiensis*), altogether providing functional annotations for 85% of genes (25 295/29769). Both transposable element genes (1853) and genes with

Table 1. The data in SCNBase.org allocated by data type, name, JBrowse alignment and category page content. A total of 43 JBrowse alignments and 21 category pages are available

Data Type	Data Name	Alignment	Page
Gene	High Confidence Horizontal Gene Transfers	+	
Gene	Alignments of genes producing secreted proteins	+	+
Gene	Genes with promoters containing DOG-box motifs	+	+
Gene	Alignment of 80 published known effector genes	+	+
Gene	Mitochondrial Genes	+	+
Gene	Annotated Gene Models Heterodera Glycines (updated v2)	+	
Gene	Annotated Gene Models Heterodera Glycines (v2) IPRSCAN	+	
Genes	Soybean Cyst Nematode (v2) gene models	+	
Genome Structure	Tandem Duplications	+	+
Genomic-seq	NCBI Nucleotide Database Entries	+	
Genomic-seq	Falcon Assembled Pacbio RS II subreads (pReads)	+	
Genomic-seq	Circular Consensus Pacbio RS II subreads (CCS Reads)	+	
Genomic-seq	Camtech Raw Pacbio RS II subreads	+	
Iso-seq	W82_PA3_Eggs Isoseq reads	+	
Iso-seq	W82_PA3_J2/J3 Isoseq reads	+	
Iso-seq	W82_PA3_J4/Adult Isoseq reads	+	
Iso-seq	W82_TN19_Eggs Isoseq reads	+	
Iso-seq	W82_TN19_J4/Adult Isoseq reads	+	
Iso-seq	W82_TN19_J2/J3 Isoseq reads	+	
Motif	80 Effector MEME->FIMO motifs	+	+
Orthologs	Globodera pallida Opscan orthologs with H. glycines		+
Orthologs	Globodera rostochiensis Opscan orthologs with H. glycines		+
Orthologs	Globodera ellingtonae Opscan orthologs with H. glycines		+
Orthologs	Meloidogyne hapla Opscan orthologs with H. glycines		+
Orthologs	Meloidogyne incognita Opscan orthologs with H. glycines		+
Orthologs	Bursephelenchus xylophilus Opscan orthologs with H. glycines		+
Repeat	RepeatModeler Repeats	+	+
Repeat	LTR Retroelements	+	
Repeat	DNA Transposons	+	
Repeat	Helitrons	+	
Repeat	Repeat Explorer of TN20	+	+
RNA-seq	Parasitic J2 (PA3) Forrest (Resistant Soybean) Biorep 2	+	
RNA-seq	Parasitic J2 (PA3) EXF63 (Susceptible soybean cultivar) Biorep 1	+	
RNA-seq	Pre-parasitic J2 (PA3) Biorep 1	+	
RNA-seq	Parasitic J2 (PA3) Forrest (Resistant Soybean) Biorep 1	+	
RNA-seq	Parasitic J2 (PA3) EXF63 (Susceptible soybean cultivar) Biorep 2	+	
RNA-seq	Pre-parasitic J2 (PA3) Biorep 2	+	
RNA-seq	SCN Soybean Root Exposure 3h-Control T01	+	
RNA-seq	SCN Soybean Root Exposure 3h-Control T02	+	
RNA-seq	SCN Soybean Root Exposure 3h-Control T11	+	
RNA-seq	SCN Soybean Root Exposure 3h-Attraction T03	+	
RNA-seq	SCN Soybean Root Exposure 3h-Attraction T04	+	
RNA-seq	SCN Soybean Root Exposure 3h-Attraction T09	+	
RNA-seq	SCN Soybean Root Exposure 24h-Control of T05	+	
RNA-seq	SCN Soybean Root Exposure 24h-Control of T06	+	
RNA-seq	SCN Soybean Root Exposure 24h-Control of T10	+	
RNA-seq	SCN Soybean Root Exposure 24h-Attraction T07	+	
RNA-seq	SCN Soybean Root Exposure 24h-Attraction T08	+	
RNA-seq	SCN Soybean Root Exposure 24h-Attraction T12	+	
SNP	SNPs across 15 populations of H.glycines	+	

Continued

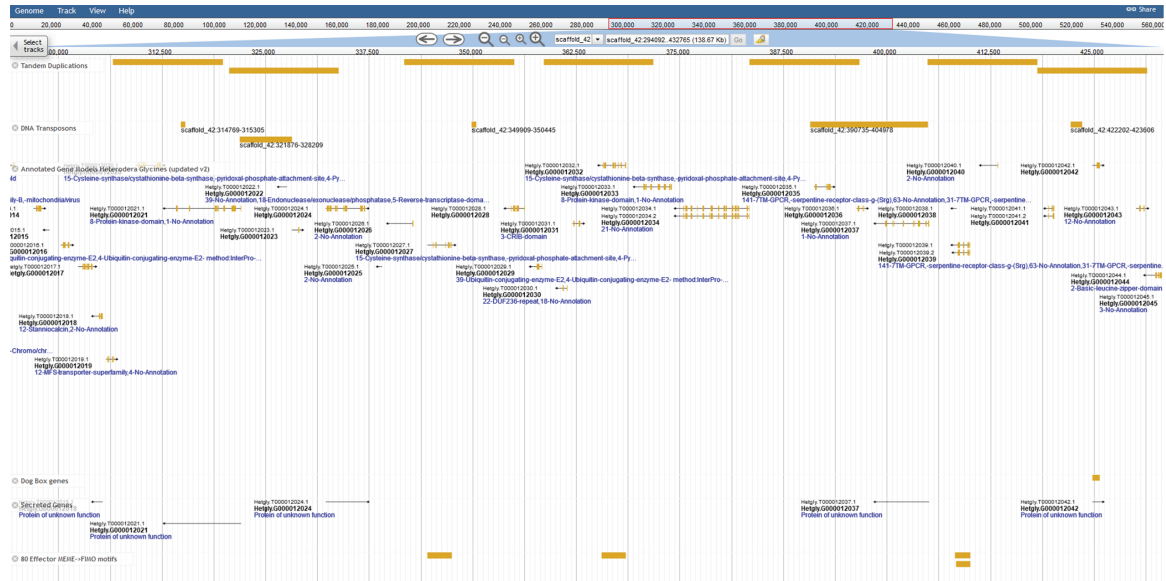


Figure 3. Duplication of effector genes through transposition/tandem duplication. Three types of effectors appear duplicated by tandem duplications, a DOG-box gene, multiple secreted genes and genes containing motifs derived from 80 previously published SCN effector sequences. JBrowse annotations are in the following order: tandem duplications, DNA transposons, ortholog-annotated genes, DOG-box genes, secreted genes and effector MEME motifs. Ortholog names represent the number of orthologous annotations among nine related plant-parasitic nematodes in the Tylenchia subclass.

dent. Thus, by combining the numerous data sets on SCN-Base, scientists can create the crucial contrast needed to identify interesting gene targets.

Conclusion and future perspective

SCN is considered the greatest threat to soybean yield worldwide but was lacking many genomic resources until now. SCNBase fills this void and will open avenues of research and increase the efficiency of many experimental approaches. To incorporate ever-increasing quantities of data, SCNBase is designed for continual improvement by encouraging extensive community-based curation through the use of standardized forms. Furthermore, an improved selection of bioinformatic tools will continue to be implemented, thereby increasing online accessibility and analyses.

Acknowledgements

We thank Anju Verma and Melissa Mitchum for testing and providing feedback for SCNBase.org.

Funding

North Central Soybean Research Program (A.J.S. and T.J.B.); this work used the Extreme Science and Engineering Discovery Environment (57), which is supported by the National Science Foundation (ACI-1548562); specifically, it used the Bridges System (58), which is supported by National Science Foundation (ACI-1445606); Hatch Act and State of Iowa Funds.

Conflict of interest. None declared.

References

1. Davis,E.L., Hussey,R.S. and Baum,T.J. (2004) Getting to the roots of parasitism by nematodes. *Trends Parasitol.*, 20, 134–141.
2. Niblack,T., Lambert,K. and Tylka,G. (2006) A model plant pathogen from the kingdom animalia: Heterodera glycines, the soybean cyst nematode. *Annu. Rev. Phytopathol.*, 44, 283–303.
3. Mitchum,M.G., Wang,X. and Davis,E.L. (2008) Diverse and conserved roles of CLE peptides. *Curr. Opin. Plant Biol.*, 11, 75–81.
4. Mitchum,M.G., Wang,X., Wang,J. et al. (2012) Role of nematode peptides and other small molecules in plant parasitism. *Annu. Rev. Phytopathol.*, 50, 175–195.
5. Davis,E.L., Hussey,R.S., Mitchum,M.G. et al. (2008) Parasitism proteins in nematode–plant interactions. *Curr. Opin. Plant Biol.*, 11, 360–366.
6. Mitchum,M.G., Hussey,R.S., Baum,T.J. et al. (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol.*, 199, 879–894.
7. Kikuchi,T., Eves-van den Akker,S. and Jones,J.T. (2017) Genome evolution of plant-parasitic nematodes. *Annu. Rev. Phytopathol.*, 55, 333–354.
8. Jones,M. (1981) Host cell responses to endoparasitic nematode attack: structure and function of giant cells and syncytia. *Ann. Appl. Biol.*, 97, 353–372.
9. Govere,A. and Bird,D. (2011) The role of plant hormones in nematode feeding cell formation. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Springer, Dordrecht Holland, pp. 325–347.
10. Quentin,M., Abad,P. and Favry,B. (2013) Plant parasitic nematode effectors target host defense and nuclear functions to establish feeding cells. *Front. Plant Sci.*, 4, 53.

11. Gao,B., Allen,R., Maier,T. *et al.* (2001) Identification of putative parasitism genes expressed in the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Mol. Plant Microbe Interact.*, **14**, 1247–1254.
12. Gao,B., Allen,R., Maier,T. *et al.* (2003) The parasitome of the phytonematode *Heterodera glycines*. *Mol. Plant Microbe Interact.*, **16**, 720–726.
13. Wang,X., Allen,R., Ding,X. *et al.* (2001) Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Mol. Plant Microbe Interact.*, **14**, 536–544.
14. Noon,J.B., Hewezi,T., Maier,T.R. *et al.* (2015) Eighteen new candidate effectors of the phytonematode *Heterodera glycines* produced specifically in the secretory esophageal gland cells during parasitism. *Phytopathology*, **105**, 1362–1372.
15. Noon,J.B., Qi,M., Sill,D.N. (2016) A plasmodium-like virulence effector of the soybean cyst nematode suppresses plant innate immunity. *New Phytol.*, **212**, 444–460.
16. Haegeman,A., Mantelin,S., Jones,J.T. *et al.* (2012) Functional roles of effectors of plant-parasitic nematodes. *Gene*, **492**, 19–31.
17. Wang,J., Lee,C., Replogle,A. *et al.* (2010) Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE effector proteins. *New Phytol.*, **187**, 1003–1017.
18. Wang,J., Replogle,A., Hussey,R. *et al.* (2011) Identification of potential host plant mimics of CLAVATA3/ESR (CLE)-like peptides from the plant-parasitic nematode *Heterodera schachtii*. *Mol. Plant Pathol.*, **12**, 177–186.
19. Lee,C., Chronis,D., Kenning,C. *et al.* (2011) The novel cyst nematode effector protein 19C07 interacts with the Arabidopsis auxin influx transporter LAX3 to control feeding site development. *Plant Physiol.*, **155**, 866–880.
20. Hewezi,T., Juvalle,P.S., Piya,S. *et al.* (2015) The cyst nematode effector protein 10A07 targets and recruits host posttranslational machinery to mediate its nuclear trafficking and to promote parasitism in Arabidopsis. *Plant Cell*, **27**, 891–907.
21. Hewezi,T., Howe,P.J., Maier,T.R. *et al.* (2010) Arabidopsis spermidine synthase is targeted by an effector protein of the cyst nematode *Heterodera schachtii*. *Plant Physiol.*, **152**, 968–984.
22. Hewezi,T. and Baum,T.J. (2013) Manipulation of plant cells by cyst and root-knot nematode effectors. *Mol. Plant Microbe Interact.*, **26**, 9–16.
23. Govere,A. and Smant,G. (2014) The activation and suppression of plant innate immunity by parasitic nematodes. *Annu. Rev. Phytopathol.*, **52**, 243–265.
24. Hewezi,T. (2015) Cellular signaling pathways and posttranslational modifications mediated by nematode effector proteins. *Plant Physiol.*, **169**, 1018–1026.
25. Paramasivan,V., Hewezi,T., Pontvianne,F. *et al.* (2018) An effector from the cyst nematode *Heterodera schachtii* derepresses host rRNA genes by altering histone acetylation. *Plant Cell*, **30**, 2795–2812.
26. Hewezi,T., Howe,P., Maier,T.R. *et al.* (2008) Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with Arabidopsis pectin methylesterase: cooperative cell wall modification during parasitism. *Plant Cell*, **20**, 3080–3093.
27. Tytgat,T., Vanholme,B., De Meutter,J. *et al.* (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. *Mol. Plant Microbe Interact.*, **17**, 846–852.
28. Howe,K.L., Bolt,B.J., Cain,S. *et al.* (2015) WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res.*, **44**, D774–D780.
29. Howe,K.L., Bolt,B.J., Shafie,M. *et al.* (2017) WormBase Parasite—a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.*, **215**, 2–10.
30. Gardner,M., Dhroso,A., Johnson,N. *et al.* (2018) Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Sci. Rep.*, **8**, 2505.
31. Masonbrink,R.E., Maier,T.R., Muppirla,U. *et al.* (2019) The genome of the soybean cyst nematode (*Heterodera glycines*) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC genomics* **20.1**, 119.
32. Waterhouse,R.M., Seppey,M., Simão,F.A. *et al.* (2017) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.*, **35**, 543–548.
33. Eves-van den Akker,S., Laetsch,D.R., Thorpe,P. *et al.* (2016) The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biol.*, **17**, 124.
34. Cotton,J.A., Lilley,C.J., Jones,L.M. *et al.* (2014) The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.*, **15**, R43.
35. Opperman,C.H., Bird,D.M., Williamson,V.M. *et al.* (2008) Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 14802–14807.
36. Abad,P., Gouzy,J., Aury,J.-M. *et al.* (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.*, **26**, 909.
37. Szitenberg,A., Salazar-Jaramillo,L., Blok,V.C. *et al.* (2017) Comparative genomics of apomictic root-knot nematodes: hybridization, ploidy, and dynamic genome change. *Genome Biol. Evol.*, **9**, 2844–2861.
38. Zheng,J., Peng,D., Chen,L. *et al.* (2016) The *Ditylenchus destructor* genome provides new insights into the evolution of plant parasitic nematodes. *Proc. Biol. Sci.*, **283**, 20160942.
39. Kikuchi,T., Cotton,J.A., Dalzell,J.J. *et al.* (2011) Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.*, **7**, e1002219.
40. Phillips,W.S., Howe,D.K., Brown,A.M. *et al.* (2017) The draft genome of *Globodera ellingtonae*. *J. Nematol.*, **49**, 127.
41. Cantarel,B.L., Korf,I., Robb,S.M. *et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.*, **18**, 188–196.
42. Emms,D.M. and Kelly,S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, **16**, 157.
43. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

44. Gordon,A. and Hannon,G. (2011) *FASTX-Toolkit*, Computer program distributed by the author.
45. Li,H. (2016) seqtk: a fast and lightweight tool for processing sequences (Broad Inst., Cambridge, MA).
46. Smit, A., Hubley, R., Green, P. (2014) RepeatModeler Open-1.0. 2008–2010. (December, date last accessed).
47. Novák,P., Neumann,P., Pech,J. *et al.* (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, **29**, 792–793.
48. Wu,T.D., Reeder,J., Lawrence,M. *et al.* (2016) GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Statistical Genomics*, Humana Press, New York, NY, 283–334.
49. Papanicolaou,A. and Heckel,D.G. (2010) The GMOD Drupal bioinformatic server framework. *Bioinformatics*, **26**, 3119–3124.
50. Ficklin,S.P., Sanderson,L.-A., Cheng,C.-H. *et al.* (2011) Tripal: a construction toolkit for online genome databases. *Database (Oxford)*, **2011**, bar044, doi: <https://doi.org/10.1093/database/bar044>.
51. Mungall,C.J., Emmert,D.B. and Consortium,F. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.
52. Zhou,P., Emmert,D. and Zhang,P. (2005) Using Chado to store genome annotation data. *Curr. Protoc. Bioinformatics*, **12**, 9.6.1–9.6.28.
53. Eilbeck,K., Lewis,S.E., Mungall,C.J. *et al.* (2005) The sequence ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.
54. Ashburner,M., Ball,C.A., Blake,J.A. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25.
55. Bird,D.M. and Koltai,H. (2000) Plant parasitic nematodes: habitats, hormones, and horizontally-acquired genes. *J. Plant Growth Regul.*, **19**, 183–194.
56. Noon,J.B. and Baum,T.J. (2016) Horizontal gene transfer of acetyltransferases, invertases and chorismate mutases from different bacteria to diverse recipients. *BMC Evol. Biol.*, **16**, 74.
57. Towns,J., Cockerill,T., Dahan,M. *et al.* (2014) XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.*, **16**, 62–74.
58. Nystrom,N.A., Levine,M.J., Roskies,R.Z. *et al.* (2015) Bridges: a uniquely flexible HPC resource for new communities and data analytics. In: *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*. ACM New York, NY, USA, p. 30.