Signatures of adaptation to a monocot host in the plant-parasitic cyst nematode Heterodera sacchari

Somnath S. Pokhare1,2, Peter Thorpe3,4, Pete Hedley3, Jennifer Morris3, Samer S. Habash1, Abdelnaser Elashry1,†, Sebastian Eves-van den Akker5, Florian M. W. Grundler1 and John T. Jones3,6,*

1Department of Molecular Phytomedicine, Institute of Crop Science and Resource Conservation, University of Bonn, Bonn 53115, Germany, 2Crop Protection Division, ICAR-National Rice Research Institute, Cuttack, Odisha 753006, India, 3The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK, 4School of Medicine, University of St Andrews, North Haugh, St Andrews KY16 9TZ, UK, 5Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK, and 6School of Biology, University of St Andrews, North Haugh, St Andrews KY16 9TZ, UK

Received 18 January 2019; accepted 18 June 2020; published online 4 July 2020.

*For correspondence (e-mail john.jones@hutton.ac.uk).
†Present address: Strube Research GmbH & Co., Neue Str. 11, Schlanstedt, 38838, Germany

SUMMARY

Interactions between plant-parasitic nematodes and their hosts are mediated by effectors, i.e. secreted proteins that manipulate the plant to the benefit of the pathogen. To understand the role of effectors in host adaptation in nematodes, we analysed the transcriptome of Heterodera sacchari, a cyst nematode parasite of rice (Oryza sativa) and sugarcane (Saccharum officinarum). A multi-gene phylogenetic analysis showed that H. sacchari and the cereal cyst nematode Heterodera avenae share a common evolutionary origin and that they evolved to parasitise monocot plants from a common dicot-parasitic ancestor. We compared the effector repertoires of H. sacchari with those of the dicot parasites Heterodera glycines and Globodera rostochiensis to understand the consequences of this transition. While, in general, effector repertoires are similar between the species, comparing effectors and non-effectors of H. sacchari and G. rostochiensis shows that effectors have accumulated more mutations than non-effectors. Although most effectors show conserved spatiotemporal expression profiles and likely function, some H. sacchari effectors are adapted to monocots. This is exemplified by the plant-peptide hormone mimics, the CLAVATA3/EMBRYO SURROUNDING REGION-like (CLE) effectors. Peptide hormones encoded by H. sacchari CLE effectors are more similar to those from rice than those from other plants, or those from other plant-parasitic nematodes. We experimentally validated the functional significance of these observations by demonstrating that CLE peptides encoded by H. sacchari induce a short root phenotype in rice, whereas those from a related dicot parasite do not. These data provide a functional example of effector evolution that co-occurred with the transition from a dicot-parasitic to a monocot-parasitic lifestyle.

Keywords: Peptide hormone mimics, effectors, Heterodera sacchari, transcriptomics.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.14933.
invasion of the host (Sobczak and Golinowski, 2011). Syncytium formation is associated with profound changes in host gene expression and modulation of the cell cycle (e.g. Siddique and Grundler, 2018).

The interactions between plants and their pathogens, including cyst nematodes, are mediated by effectors, i.e. secreted proteins that manipulate the host to the benefit of the pathogen. Effectors from cyst nematodes are primarily produced in the dorsal and subventral pharyngeal gland cells and are secreted into the host via the stylet. The availability of genome and/or transcriptome resources from a range of cyst nematodes (e.g. Gao et al., 2003; Cotton et al., 2014; Kumar et al., 2014; Zheng et al., 2015; Eves-van den Akker et al., 2016a,b) has facilitated identification of effectors from these species. Strategies for identifying effectors from genome and transcriptome resources include identifying secreted proteins that are upregulated at parasitic stages of the nematode (e.g. Thorpe et al., 2014; Espada et al., 2016) and direct sequencing of mRNA extracted from aspirated gland cells (Maier et al., 2013), followed in both cases by in situ hybridisation to confirm expression in the pharyngeal gland cells. More recently, it has been shown that promoters associated with genes expressed in the gland cells can be used to identify comprehensive lists of effectors from diverse plant-parasitic nematode species (Eves-van den Akker and Birch, 2016; Espada et al., 2018).

As a result of these studies, effectors have been identified from a wide range of plant-parasitic nematodes with subsequent functional studies showing that they have roles in various stages of the plant–nematode interaction, including metabolism of the plant cell wall to facilitate invasion and migration, suppression of host defences and initiation of syncytium formation (reviewed in Gheysen and Mitchum, 2011; Jones and Mitchum, 2018).

The plant cell wall is the first significant barrier that any invading pathogen, including plant-parasitic nematodes, will need to overcome to infect a plant and plant-parasitic nematodes are well equipped with proteins that allow them to metabolise the plant cell wall. The first effector identified from any plant-parasitic nematode was a beta-1,4-endoglucanase (cellulase) from the potato (Solanum tuberosum) cyst nematode Globodera rostochiensis (Smant et al., 1998), and a range of cell wall-degrading and -modifying proteins have subsequently been identified as cyst nematode effectors, including pectate lyase (Popeijus et al., 2000), GHF43 Arabinase (Cotton et al., 2014), GH53 Arabino-ogalactan endo-1,4-beta-galactosidase (Vanholme et al., 2009), expansins (Qin et al., 2004) and proteins encoding carbohydrate-binding domains (Hewezi et al., 2008). All of these genes, as well as others encoding chorismate mutase (Jones et al., 2003) and proteins potentially involved in vitamin biosynthesis (Craig et al., 2008), have been acquired by horizontal gene transfer from bacteria (reviewed in Kikuchi et al., 2017). Cyst nematode effectors have also been identified that suppress host defence responses, most notably several members of the SPRYSEC family of effectors (Postma et al., 2012; Mei et al., 2015) and a modified ubiquitin extension protein (Chronis et al., 2013). The details of how cyst nematodes induce the formation of their syncytium in the roots of their hosts are less clear, although effectors that are likely to be important in this process have been characterised. Another novel effector (19C07) has been identified from Heterodera glycines and Heterodera schachtii that interacts with the LAX3 auxin influx transporter (Lee et al., 2011). In addition, all cyst nematodes studied to date produce effectors that include variable numbers of C-terminal repeats encoding peptides similar to CLAVATA3/EMBRYO SURROUNDING REGION-like (CLE) peptides. Functional studies have shown that the nematode peptides can complement mutant Arabidopsis lacking these peptides (Wang et al., 2005). The CLE proteins are modified by plant cell machinery in a manner similar to that of the endogenous proteins and the CLE peptides themselves subsequently interact with the CLAVATA2 receptor protein, which is required for nematode parasitism (Replogle et al., 2011; Replogle et al., 2013). In plants, CLE peptides regulate cell differentiation and thus contribute to the control of meristem maintenance in shoots, roots and vascular tissues. The ability to produce endogenous CLE peptides is likely to be a key factor in the ability to induce a feeding structure in plants.

Many plant species are parasitised by cyst nematodes, including monocots and dicots. However, each individual species of cyst nematode tends to have a relatively narrow host range, with some exceptions. Although little is known about the molecular determinants of host range in plant-parasitic nematodes, effectors have been shown to have a central role in this process in other pathogens (reviewed by Stam et al., 2014). A view has emerged that ‘non-host resistance’ of closely related species is most likely due to recognition of effectors by a resistance gene, while failure to infect a more distantly related species is most likely due to the incompatibility of effectors with their cognate targets (Schulze-Lefert and Panstruga, 2011). For example, comparisons of the genomes of Phytophthora infestans and a closely related species Phytophthora mirabilis, which infects Mirabilis jalapa, revealed that 82 of 345 genes which showed signs of positive selection could encode effector sequences (Raffaele et al., 2010). Subsequent work on orthologous effectors that encode protease inhibitors from the two species showed that the protease inhibitor effectors from each of these species interact specifically with protease targets from their respective host plants (Dong et al., 2014). Host–pathogen co-evolution is therefore reflected in adaptations of effectors for function in the host.

Heterodera sacchari is an increasingly economically important pathogen of several monocot species, including rice (Oryza sativa) and sugarcane (Saccharum officinarum).
Crop losses due to *H. sacchari* can exceed 40%, with more severe damage reported under rain-fed upland conditions (Kyndt *et al.*, 2014). Compared to other characterised cyst nematodes, the biology of *H. sacchari* is unusual in that it is restricted to monocots and reproduces mainly by mitotic parthenogenesis (CABI, 2014). Here we sequenced and assembled the transcriptome of *H. sacchari* and used these data to reconstruct the evolutionary history of this and related species. These data suggest that *H. sacchari* and a related monocot parasite *Heterodera avenae* evolved to parasitise monocot plants secondarily, from a dicot-parasitic ancestor. To explore the genetic changes associated with the evolutionary transition from dicot to monocot parasite, we identified homologues of previously identified effectors in *H. sacchari*. We show that in general the effectors have diversified in sequence more than non-effectors when compared to their most similar homologue in a dicot parasite, and that while some effectors show conserved expression profiles and likely function, specific aspects of the effector repertoire of *H. sacchari* appear to be adapted to monocots.

**RESULTS AND DISCUSSION**

**The transcriptome of *H. sacchari***

A total of 17,086,132 paired-end reads were obtained from the cDNA extracted from second stage juveniles (J2) and parasitic stage nematodes 15 days after infection. These sequence reads have been submitted to SRA (accession number PRJEB28025). The sequence reads were pooled and assembled into a single reference transcriptome of 44,230 transcripts after filtering with Transrate and removing contaminants. Assemblies are available at https://zenodo.org/deposit/1324265. CEGMA analysis showed that 88% of the core eukaryotic genes were present as full-length transcripts, with a further 6% represented by partial-length transcripts. BUSCO analysis using the metazoan data set suggested that the assembly contains 76% complete BUSCO sequences and a further 6.0% fragmented sequences. Seventeen percent of the BUSCO genes were not identified. Comparisons with published gene models from cyst nematode genome sequences showed that 34,203 (77.3%) and 30,482 (68.9%) of the *H. sacchari* transcripts matched sequences in *G. rostochiensis* and *Globodera pallida*, respectively (e value, *1* × 10⁻¹⁰). Taken together, these data suggest that the assembly produced represents a large proportion of the *H. sacchari* transcriptome of these life stages.

*H. sacchari* evolved to parasitise monocots from a dicot-parasitic ancestor

We used a subset of 96 core eukaryotic genes conserved in *H. sacchari* and 18 related species to reconstruct a multi-gene phylogeny (Figure 1). This phylogeny robustly positions *H. sacchari* and the related cereal cyst nematode *H. avenae* in a monophyletic subclade of monocot parasites, nested within a clade of related dicot parasites of the genera *Heterodera* and *Globodera*. The most parsimonious explanation for this is that (i) *H. sacchari* and *H. avenae* share a common, monocot-parasitic, ancestor and (ii) this last common monocot-parasitic ancestor evolved to parasitise a monocot host secondarily, from a dicot-parasitic ancestor. This provides the comparative framework to explore the genes conserved, and the genes involved in the transition to parasitise monocots.

**Figure 1.** Phylogenetic analysis of 96 CEGMA genes conserved in 19 species of nematode shows that *H. sacchari* and *H. avenae* share a common monocot-parasitic ancestor that most likely evolved from a parasite of dicots (indicated by the arrow). The scale bar indicates substitutions per base. Node labels indicate support values for 1000 bootstraps.
diverged, during adaptation to monocot parasitism by nematodes.

**Genes encoding cell wall-modifying enzymes acquired via horizontal gene transfer in the *H. sacchari* transcriptome**

The plant cell wall is the first significant barrier to an invading pathogen, and while largely similar between dicots and monocots, there are notable differences in composition. Plant-parasitic nematodes in general are well equipped with proteins that allow them to modify and degrade specific components of the plant cell wall. Many of these genes were acquired via horizontal gene transfer from bacteria (Danchin et al., 2010).

The transcriptome of *H. sacchari* contains representatives of most previously described cases of horizontal gene transfer in related plant-parasitic nematodes, including a wide range of cell wall-degrading enzymes (Table 1) and several other sequences putatively acquired by horizontal gene transfer (e.g. the GH32 inverases and chorate mutase (Jones et al., 2003; Danchin et al., 2016)). Analysis of the expression profiles of one of the *H. sacchari* GH5 cellulases and the chorate mutase showed that, as in other cyst nematodes, expression was restricted to the subventral pharyngeal gland cells in J2s, and while the cellulase was upregulated at J2, the chorate mutase was expressed throughout the life cycle (Figure 2). These sequences may therefore play a similar role in the biology of *H. sacchari* and other cyst nematodes.

Notably, however, both the *H. sacchari* and *H. avenae* transcriptomes lack sequences similar to GH53 Arabinogalactan endo-1,4-beta-galactosidase, in spite of the fact that all other cyst nematodes analysed to date which parasitise dicots have such proteins. While the absence of evidence in transcriptome data sets is not necessarily evidence of absence in the genome, it nevertheless reflects the host range of these species. Cell walls of commenliroid monocots (which include the main hosts of *H. sacchari* and *H. avenae*) have a different composition compared to those of dicots and contain relatively low amounts of pectic polysaccharides, including the substrate of the GH53 enzymes (Vogel, 2008). Without genomic resources we cannot conclude whether the conspicuous absence in the transcriptomes of the monocot parasites is because these genes have been lost entirely or they are not expressed under these conditions.

**An overview of *H. sacchari* effector-like sequences**

Effectors modulate plant processes to promote disease, and are often finely tuned to their host. To determine whether the effector repertoire of *H. sacchari* reflects its secondary adaptation to a monocot host, we first identified and characterised effectors in the *H. sacchari* transcriptome by building on a detailed genome-wide analysis of cyst nematode effectors performed for *G. rostochiensis* (Eves-van den Akker et al., 2016a). Using *G. rostochiensis* effectors as a starting point for comparative purposes, 185 of the 295 identified a similar sequence in the *H. sacchari* transcriptome (at e value < 1 x 10^-10). We compared the similarity of *H. sacchari* effector-like sequences and non-effector-like sequences to their most similar homologues in the *G. rostochiensis* genome. This showed that, on average, putative effectors of *H. sacchari* are more different from their closest homologue in *G. rostochiensis* than non-effectors are to their corresponding closest homologue (Figure 3). Taken together, these data suggest that while most effectors are apparently conserved, they have nevertheless accumulated more mutations than non-effectors since these species diverged. Further analysis of the differences between the variation in effectors and non-effectors was undertaken. Twenty-one percent of the mutations were classed as non-synonymous for the putative effectors, versus 22% for other transcripts; however, non-synonymous single nucleotide polymorphisms in the putative effectors were more often predicted to have a significant effect on the coding sequence compared to other transcripts: 0.93% versus 0.82% were classed as ‘high impact’ for putative effectors and other transcripts, respectively. All of the following high impact categories were more common in the putative effectors compared to other transcripts: gain of premature start codon, in-frame deletions, frameshift variants, start loss

*Table 1* *H. sacchari* sequences similar to putative cell wall-degrading and -modifying proteins

| Substrate Family | Cellulose | Xylan | Arabinan | Pectin | Arabinogalactan | Expansins | CBM2 domains |
|------------------|-----------|------|----------|--------|----------------|-----------|--------------|
|                  | GH5       | GH30 | GH43     | GH28   | PL3            | GH53      |              |
| *H. sacchari*    | 8         | 0    | 0        | 0      | 1              | 0         | 2            |
| *H. avenae*      | 16        | 0    | 0        | 0      | 2              | 0         | 2            |
| *H. schachti*    | 9         | 0    | 0        | 1      | 9              | 3         | 4            |
| *G. rostochiensis* | 11      | 0    | 0        | 1      | 3              | 1         | 7            |
| *G. pallida*     | 16        | 0    | 0        | 1      | 7              | 2         | 9            |
| *M. incognita*   | 21        | 0    | 6        | 2      | 30             | 0         | 20           |
| *N. aberrans*    | 2         | 0    | 3        | 1      | 8              | 3         | 2            |
| *Bursaphelenchus xylophilus* | 0 | 11 | 0 | 0 | 15 | 0 | 8 |

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. The Plant Journal, (2020), 103, 1263–1274
and stop gain. Moreover, the rate of variants (i.e. length of transcript/number of variants) was significantly higher for the putative effectors versus non-effectors (mean = 1754 versus 1243 for putative effectors versus other transcripts, respectively; Mann–Whitney U test P value: 1.98 \times 10^{-41})

Consistent with this observation, analysis of effectors from a range of pathogens, including plant-parasitic nematodes, has shown that they are under strong diversifying selection pressure. For example, whole genome resequencing of five pathotypes of *G. rostochiensis* showed that effectors contain significantly more variants and more non-synonymous variants per gene than do randomly selected non-effector genes (Eves-van den Akker et al., 2016a). On a more detailed scale, the SPRYSEC effector RBP1 from *G. pallida*, which is recognised by the Gpa2 resistance gene of potato, has been subjected to positive selection at several different residues, including the residue that determines recognition or evasion by Gpa2 (Sacco et al., 2009).

The SPRY domain family is greatly expanded in cyst nematodes, with 295 sequences predicted in *G. pallida*, of which approximately 30 have a signal peptide (and thus encode SPRYSEC effectors) (Cotton et al., 2014), and 71 sequences in *G. rostochiensis*, of which 17 may be SPRYSEC effectors (Eves-van den Akker et al., 2016a). While we found a similar proportion of SPRYs to SPRYSECs in *H. sacchari* (80 SPRY domain-encoding transcripts and 6 SPRYSECs), with similar spatial expression in the dorsal pharyngeal gland cell (Figure 4), the temporal expression is unusual. At least two *H. sacchari* SPRYSECs are upregulated in parasitic stage nematodes (Figure 4), and this contrasts with the *G. pallida* SPRYSECs, which tend to be upregulated in J2s (Mei et al., 2015). Several SPRYSECs have been shown to suppress host defence signalling (Postma et al., 2012; Mei et al., 2015; Mei et al., 2018). What
role the \textit{H. sacchari} SPRYSECs play in infection is yet to be determined.

Other notable peculiarities in the effector repertoire of \textit{H. sacchari} are the HYP effectors. The HYP effectors, first identified in \textit{G. pallida} and \textit{G. rostochiensis}, are secreted from the gland cells surrounding the main anterior sense organs and show unprecedented variability between individuals (Eves-van den Akker et al., 2014a,b). The HYP effectors are strongly upregulated in parasitic stages of \textit{G. pallida} and can be subdivided into three subfamilies based on the presence and type of subfamily-specific tandem repeats. The transcriptome of \textit{H. sacchari} contains two full-length HYP-like transcripts (primarily represented in the 15 days post-infection library) with unusual characteristics. A phylogenetic analysis of the \textit{H. sacchari} HYP-like sequences with all other known full-length HYP effectors (\(n = 65\)) shows that the \textit{H. sacchari} HYP-like sequences group in a separate subfamily (named subfamily HYP-0, Figure 5). Comparison of the \textit{H. sacchari}-predicted sequences with those from \textit{G. pallida} shows that these sequences have both highly conserved regions that flank the variable domain, but they are the first to encode a major structural variant outside the hypervariable domain, and the region corresponding to the ‘hypervariable domain’ contains a novel sequence that has a very limited repeat structure (sequence features are summarised in Figure 5).

Specialisation of \textit{H. sacchari} CLE-like effectors to parasitism of a monocot host

One class of \textit{H. sacchari} effectors shows signs of adaptation to a monocot host. The CLE effectors mimic plant-peptide hormones and have been characterised in a number of plant-parasitic nematodes. The transcriptome of \textit{H. sacchari} contains several partial transcripts that encode proteins with similarity to CLE-like effectors from other plant-parasitic nematodes. Using a novel approach based on identifying reads that map to partial transcripts and carrying out a local assembly, we were able to computationally assemble six unique transcripts to recapitulate the full-length open reading frame, resulting in five unique polypeptide sequences from methionine to the stop codon. As for other cyst nematodes, the \textit{H. sacchari} sequences each encoded a signal peptide at the N-terminus followed by an N-terminal domain (Figure S2), which in other cyst nematodes enables translocation into the apoplast after the protein is secreted into the plant cell (Wang et al., 2010a). All six of these \textit{H. sacchari} transcripts encode a single canonical CLE domain at their C terminus. Another transcript (DN37996_c0_g2_i1) encodes a tandemly repeated motif with no clear homology to canonical CLE domains (despite the similarity of the rest of the protein sequence to CLE effectors of other cyst nematodes). The six \textit{H. sacchari} CLE effector-like sequences can be divided into two groups based at least in part on their signal peptide and CLE domains. The CLE domains within each group are identical in protein and nucleic acid sequence (Figure 6(a)).
Given that CLE peptides vary between plant species in general, and monocots and dicots in particular, we hypothesised that the CLE domains of CLE-like effectors in *H. sacchari* may have specialised prior to/concurrent with the transition from dicot to monocot parasite. To test this hypothesis, we analysed a database of 391 CLE peptides from 20 plant species collated from Zhang et al. (2014) and Oelkers et al. (2008). We created an all-by-all matrix of similarity between plant CLE peptides and *H. sacchari* CLE peptides based on a normalized BLOSUM62 score. We then used this matrix to generate a CLE similarity network (Figure 6(b)), highlighting the host (*Oryza sativa*) and the nematode (*H. sacchari*) CLEs. Three of the *H. sacchari* CLEs form part of a well-connected cluster containing 14 other CLEs, eight of which are from rice. On closer inspection, we found that three of these *H. sacchari* CLEs (Hsac_DN37996_c0_g4_i1, Hsac_DN35920_c0_g3_i1 and Hsac_DN35920_c0_g2_i1) were sequence-identical mimics of eight rice CLEs (OS_GEN_Os01g55080_1, OSEST_NP890021_1, OSEST_TC269510_1, OS_TA_AK108976_1, OS_TA_CA758496_1, OS_GEN_Os01g48230_1, OS_GEN_Os01g48260_1 and OSEST_TC271220_1) in the terminal 13 amino acids of its CLE domain. The six non-rice plant CLEs within this cluster are from *Arabidopsis* and *Populus*, but none have direct connections with *H. sacchari* CLEs. Three other *H. sacchari* CLEs (Hsac_DN37996_c0_g1_i1, Hsac_DN49341_c0_g1_i1 and Hsac_DN35920_c0_g1_i1) form a second small cluster with one other rice CLE (OS_GEN_Os05g48730_1). *H. sacchari* CLEs are the only nematode CLEs with a connection to a rice CLE and the only connections between *H. sacchari* CLEs and plant CLEs are with rice. Interestingly, there are several other sequence-divergent CLE clusters in rice with no corresponding *H. sacchari* CLE, suggesting selective mimicry of a subset of this family. It is known that CLE family members in dicots have diverse roles (e.g. Mitchum et al., 2008) and mimicry of a subset of rice CLEs by *H. sacchari* may reflect the need to target the function of a subset of the full rice CLE complement.

The functional significance of the similarity between the *H. sacchari* CLE peptides and those from rice was experimentally investigated by taking advantage of the short root phenotype observed from overexpression or exogenous application of CLEs (e.g. Fiers et al., 2004 and Chen et al., 2015). We synthesised a synthetic version of the *H. sacchari* CLE peptide that had the highest connectivity with rice CLEs in the network (sequence identical to the rice CLEs in the 13 terminal amino acids). In order to confirm that this is indeed a true *H. sacchari* sequence we cloned the gene encoding this peptide from *H. sacchari* gDNA (Figure S3). We analysed the effect of the peptide when applied exogenously to rice seedlings on root growth when compared to a randomised version of this CLE, or a CLE from the dicot-parasitic *H. glycines* (Wang et al., 2005). This analysis showed that the peptide from *H. sacchari* induced a short root phenotype in rice whereas peptides from *H. glycines* and a shuffled peptide used as a control had no effect (n = 24 per condition, P < 0.001, Tukey’s HSD, Figure 7).

**Conclusions**

We used whole transcriptome sequencing to show that *H. sacchari* and the related *H. avenae* evolved to parasitise a monocot host from a last common dicot-parasitic ancestor. We mined these data to identify and characterise the cell wall-degrading enzyme and effector complement of *H. sacchari*. Finally, we showed that while *H. sacchari* has a similar effector arsenal to the related cyst nematodes that parasitise dicots, the CLE effectors

---

**Figure 5.** Overview of the HYP gene family in cyst nematodes. Left, a schematic representation of the *H. sacchari* HYP-like sequences shows the position of a novel structural variant outside the hypervariable domain. The faded region of the hypervariable domain of HYP-1 and -3 represents the region of variable sequence. Right, a midpoint re-rooted phylogenetic construction of HYP sequences from *G. pallida* and *H. sacchari* position those of *H. sacchari* as a distinct subfamily (named here subfamily HYP-0).
The only nematode CLEs with connections to rice CLEs.

The resulting J2s were collected every 5 days and either frozen in liquid nitrogen or used immediately for infection of new plants. The resulting J2s were collected every 5 days and either frozen in liquid nitrogen or used immediately for infection of new plants.

**EXPERIMENTAL PROCEDURES**

**Biological material**

H. sacchari was cultured on rice cv. Nipponbare as described in Pokhare et al. (2019). Briefly, plants were grown in a potting mixture of sand, field soil and organic matter (70:29:1) and were infected with second stage juveniles. After 12 weeks, watering of the plants was stopped and the plants were allowed to dry for 2 weeks. Cysts were collected by Cobb’s decanting and sieving method using standard protocols (Cobb, 1918). The cysts were surface-sterilised and placed in 3 µM ZnCl2 to initiate hatching. The resulting J2s were collected every 5 days and either frozen in liquid nitrogen or used immediately for infection of new plants. Parasitic stage female nematodes were collected by hand 15 days after infection and frozen in liquid nitrogen until use. The original population of H. sacchari was provided by Dr D. Coyne (International Institute for Tropical Agriculture). The population was originally derived from rice in Nigeria and was maintained at The International Institute for Tropical Agriculture on susceptible Oryza sativa. Although the original population was derived from a field population of cysts, this nematode reproduces by mitotic parthenogenesis, meaning that genetic diversity within the population is likely to be low, facilitating much of the bioinformatic analysis.

**RNA extraction from second stage juveniles and parasitic stage female nematodes using a Nucleospin RNA XS kit (Macherey Nagel, Dueren DE, Germany) following the manufacturer’s instructions. The quantity and integrity of RNA were assessed using a Bioanalyzer. Library preparation for RNAseq was performed using the TruSeq RNA Library Prep Kit v2 (Illumina, Cambridge, UK) as recommended by the manufacturer. Illumina: Protocol # 1502645, revision F). Separate libraries were generated from the juvenile (1 µl) and female (500 ng) total RNA samples, using single-end TruSeq Index Adapters AR002 (CGATGTG) and AR004 (TGACCA), respectively. Each library was quality checked using a Bioanalyzer 2100 (Agilent, Stockport, UK) and quantified using a Qubit fluorometer (Thermo Fisher). Equal molarities of each library were combined and run at 12 pm on a MiSeq (Illumina) generating paired-end 2 × 250 bp reads, as recommended. A fastq file was generated for each sample using MiSeq Control Software (version 2.6) for downstream quality control and analysis.

**Transcriptome sequencing**

DNA was extracted from second stage juvenile and parasitic stage female nematodes using a Nucleospin RNA XS kit (Macherey Nagel, Dueren DE, Germany) following the manufacturer’s instructions. The quantity and integrity of RNA were assessed using a Bioanalyzer. Library preparation for RNAseq was performed using the TruSeq RNA Library Prep Kit v2 (Illumina, Cambridge, UK) as recommended by the manufacturer. Illumina: Protocol # 1502645, revision F). Separate libraries were generated from the juvenile (1 µl) and female (500 ng) total RNA samples, using single-end TruSeq Index Adapters AR002 (CGATGTG) and AR004 (TGACCA), respectively. Each library was quality checked using a Bioanalyzer 2100 (Agilent, Stockport, UK) and quantified using a Qubit fluorometer (Thermo Fisher). Equal molarities of each library were combined and run at 12 pm on a MiSeq (Illumina) generating paired-end 2 × 250 bp reads, as recommended. A fastq file was generated for each sample using MiSeq Control Software (version 2.6) for downstream quality control and analysis.

**Transcriptome assembly, quality control and annotation**

Scripts used to analyse the data are available at https://github.com/peterthorpe5/Hsac_transcriptome. Assemblies are available at https://zenodo.org/deposit/1324265. Raw reads are available under primary accession number PRJEB28025 and secondary accession number ERP110186.

The 17 725 370 read pairs (5 636 376 from juveniles and 11 329 919 from females, with 759 075 undetermined) were first quality controlled using FastQC (Andrews, 2010) and then quality-trimmed using Trimmomatic version 0.32 (Bolger et al., 2014). The resulting 17 086 132 read pairs were assembled using Trinity version 2.1.1 (kmer length 25) (Haas et al., 2013). The resulting assembly was subjected to quality control filtering using TruTransrate version 1.0.1 (Smith-Unna et al., 2016). Low quality/scored transcripts were removed (based on read mapping to the assembly). Coding sequences were predicted using TransDecoder (Using DIAMOND BLASTP versus Swiss prot and HMHR search versus Pfam A domain guides). The predicted coding sequences were DIAMOND BLASTP version 0.7.9 (Buchfink et al., 2014) searched against the GenBank NR database (May, 2017). The results were interrogated for their alien index (AI) score using a lateral gene transfer prediction tool (https://github.com/peterthorpe5/public_scripits/tree/master/Lateral_gene_transfer_prediction_tool), which predicts putative horizontal gene transfer events. Any sequence with an AI greater than 25 and that had a BLAST hit identity greater than 70% to a non-metazoan was flagged as putative contamination. Putative contaminant sequences were removed, and the corresponding transcripts were removed from the assembly. BUSCO version 1.1b (Simão et al., 2015) and CEGMA version 2.4 (Parra et al., 2007) were used to quantify the completeness of the assembly. The resulting coding sequences were annotated using Trinotate (Grabherr et al., 2011), HHMER (Finn et al., 2011), Pfam (Yang et al., 2011), HMMER (Finn et al., 2011), and Pfam (Yang et al., 2011).
Phylogenetics

Ninety-six CEGMA genes conserved in the genome and or transcriptome resources of 18 plant-parasitic nematode species were used for phylogenetic reconstruction. The species used were H. avenae (Kumar et al., 2014), H. glycines (Masonbrink et al., 2019), H. schachtii (S. Eves-van den Akker, pers. comm.), G. pallida (Cotton et al., 2014), G. rostochiensis (Eves-van den Akker et al., 2016a), Globodera ellingtonae (Phillips et al., 2017), Rotylenchulus reniformis (Eves-van den Akker et al., 2018b), Radopholus similis (Jacob et al., 2008), Meloidogyne arenaria (Blanc-Mathieu et al., 2017), Meloidogyne javanica (Blanc-Mathieu et al., 2017), Meloidogyne incognita (Blanc-Mathieu et al., 2017), Meloidogyne hapla (Opperman et al., 2008), Pratylenchus coffeae (Burke et al., 2015), Nacobbus aberrans (Eves-van den Akker et al., 2014), Bursaphelenchus xylophilus (Kikuchi et al., 2019), Caenorhabditis elegans (C. elegans sequencing consortium, 1998), Caenorhabditis briggsae (Hillier et al., 2007) and Lepidioleptolampodogon thompsoni (Chen et al., 2017). The protein sequences of CEGMA genes were aligned and refined individually using MUSCLE (Edgar, 2004). Individual alignments were concatenated and submitted to the IQtree online server with associated partition file. Model selection was carried out on each partition and a concatenated multi-gene phylogeny was generated using the ultra-fast mode and 1000 bootstraps (Trifinopoulos et al., 2016).

CLE effector identification and network analysis

Several partial transcripts encoding proteins with similarity to CLE effectors from various plant-parasitic nematodes were identified in the transcriptome assembly. Partial transcripts were computationally extended using an iterative approach of mapping and overlap assembly using the wrapper script provided with MITObim (Hahn et al., 2013). Only the deduced amino acid sequences of full-length CLEs were used for further analyses. The gene encoding the CLE peptide used for functional analyses was cloned and sequenced in an e value cutoff of 10⁻¹⁰. Similarly, candidate H. sacchari effectors were identified by comparisons with effectors from other cyst nematode species (Gao et al., 2003; Thorpe et al., 2014; Eves-van den Akker et al., 2016a) using an e value cutoff of 10⁻¹⁰.

Cloning and characterisation of candidate effector sequences

The complete open reading frames of selected genes were amplified from cDNA of post-parasitic second stage juveniles, or complete genes from gDNA extracted from cysts. PCR products were purified using the Qiagen PCR Purification kit (Qiagen, Manchester, UK) and cloned into the pGEMT Easy or pCR8/TOPO/GW vectors following the manufacturer’s guidelines.

Analysis of expression profiles of candidate effectors

The spatial expression patterns of candidate effectors from H. sacchari were investigated using in situ hybridisation of digoxigenin-labelled probes to juvenile nematodes as previously described...
ical replicates of each condition. The expression of candidate effectors across the H. sacchari life cycle (eggs, pre-parasitic juveniles and 15 and 25 days post-infective females) was analysed using quantitative reverse transcriptase PCR with gene-specific primers. Around 3000 eggs and pre-parasitic juveniles were collected for each replicate. To collect eggs, cysts were crushed under a binocular microscope and eggs were transferred into an Eppendorf tube in sterile water. Juveniles were harvested every 5 days as described above and flash frozen in liquid nitrogen. For the other life stages (15 and 25 days post-inoculation), approximately 150 parasitic stage nematodes were collected from rice grown in pluronic gel as described above. Nematodes were stored at −80 °C and used subsequently for RNA extraction. Total RNA was extracted as above and were processed using an Agilent 2100 Bioanalyzer system to check the quality and quantity of extracted RNA. The RNA samples with RNA integrity number of more than 8 were used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and oligo-dT primer. The cDNA was tested for expression changes during the nematode life stages using the Stepone Plus Real-Time PCR System (Applied Biosystems) with cycling parameters of 95 °C for 15 sec and 60 °C for 30 sec (40 cycles) for amplification. Each sample well contained 10 μl of Fast SYBR Green qPCR Master Mix (Invitrogen), 9 μl of the gene-specific primer mixture with a final concentration of 0.2 μM for each primer and 1 μl of cDNA. The data were analysed using the Stepone Plus Real-Time PCR software to create Ct values and relative expression was calculated following Pfaffl (2001). Elongation Factor 1 alpha was used as internal control for all experiments. Three biological replicates from each stage and three technical replicates for each biological replicate were used for qRT-PCR studies. Details of the primers used in this study are provided in Table S1.

Functional analysis of H. sacchari CLE sequences

To analyse the in vivo function of CLE peptides, we developed a protocol for exogenous application to rice seedlings, similar to that described for Arabidopsis thaliana (Wang et al., 2010). The terminal 13 amino acid CLE domain of nematode CLE-like sequences was synthesised with hydroxy prolines in positions 5 and 8 for H. sacchari (H-Lys-Arg-Leu-Ser-Hyp-Gly-Gly-Hyp-Asp-Pro-Gln-His-His-OH) and H. glycines (H-Lys-Arg-Leu-Ser-Hyp-Gly-Hyp-Asp-Pro-Gln-His-His-OH) and H. glycines (H-Lys-Arg-Leu-Ser-Hyp-Gly-Hyp-Asp-Pro-Gln-His-His-OH). As a control, a shuffled version of the H. glycines CLE domain (H-Hyp-Lys-Pro-Gly-Leu-OH) was used. Seeds of rice (cv. Nipponbare) were synthesised with hydroxy prolines in positions 5 and 8 for the 13 amino acid CLE domain of nematode CLE-like sequences was also synthesised (H-Asp-His-Ser-Hyp-Gly-Gln-His-Hyp-Arg-Hyp-Lys-Pro-Gly-Leu-OH). Seeds of rice (cv. Nipponbare) were surface-sterilised and allowed to germinate on sterile, wet filter paper. After 7 days, plants of similar size were transferred to plates containing ½ MS supplemented with 10 μM of the relevant peptide or an equivalent volume of sterile distilled water and left for 10 days at 25°C in 16 h light/8 h dark. After this time, plants were removed from the plates and the roots were washed to remove any adhering medium. The length of the longest root was measured for 24 biological replicates of each condition.

ACKNOWLEDGEMENTS

Somnath Pokhare was supported by a fellowship provided by the Indian Council of Agricultural Research (ICAR), Government of India. This work was also supported by Royal Society International Exchanges (award number IE150670) and by the Rural and Environmental Science and Analytical Services Division of the Scottish Government. SEvDA is supported by the Biotechnology and Biological Sciences Research Council (grants BB/R011311/1 and BB/S006397/1). This work benefited from interactions funded through COST Action FA1208. The authors thank Dr Danny Coyne (IITA) for providing the starter culture of the H. sacchari population used for this work. Bioinformatics and Computational Biology analyses were supported by the University of St Andrews Bioinformatics Unit, which is funded by a Wellcome Trust ISSF award (grant 105621/Z/14/Z).

AUTHOR CONTRIBUTIONS

SP, PT and SH carried out the majority of the experimental work and contributed to the writing and revision of the manuscript. PH and JM undertook and assisted in analysis of the sequencing data. JJ, FMW, SEvdA and AE designed the study and contributed to data analysis and interpretation of the results. JJ, PT and SEvdA led the writing of the manuscript, with contributions from FMA and AE.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

All sequence data, including assemblies, raw sequence reads and scripts used for processing, are publicly available as described in the Experimental procedures section. All other materials described in this MS are available upon request.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Negative control in situ hybridisation reactions for chorismate mutase (left panel) and DN36333 (right panel). No staining is seen in these nematodes.

Figure S2. Alignment of CLE effector-like sequences assembled from the H. sacchari transcriptome. Highlighted are the position of the predicted signal peptide (green), the unknown repeat of the terminal CLE domain (yellow) and the canonical CLE domain of the other six sequences (red). The arrow indicates polymorphism in the CLE domain.

Figure S3. Cloned genomic CLE from H. sacchari. (a) Alignment of the genomic clone and predicted transcript assembled from the transcriptome for the H. sacchari-like gene Hsac_DN37996F. Primers used to clone the gene are highlighted with grey bars. The CLE domain (spanning an intron) is highlighted in purple. (b) Mapping the RNAseq reads back to the genomic clone supports the deduced intron:exon structure and reveals two potential splice variants (a read through intron in exon 4). The shorter splice variant results in a frame shift and consequently no CLE-like domain. The longer splice variant encodes a canonical CLE domain.

Table S1. Primers used in this study.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403-410.
Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 8 September 2017.
Ashburner, M., Ball, C.A., Blake, J.A. et al. (2000) Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25-29.
Eves-van den Akker, S., Laetsch, D.R., Thorpe, P. et al. (2016a) The genome of the yellow potato cyst nematode, Globodera rostochiensis, reveals insights into the bases of parasitism and virulence. *Genome Biol.* 17, 124.

Eves-van den Akker, S., Lilley, C.J., Yusup, H.B., Jones, J.T. and Urwin, P.E. (2016b) Functional C-terminally encoded plant peptide (CEP) hormone domains evolved de novo in the plant parasite *Rotylenchulus reniformis*. *Mol. Plant Pathol.* 17(8), 1265–1275.

Eves-van den Akker, S., Lilley, C.J., Jones, J.T. and Urwin, P.E. (2016c) Identification and characterisation of a hyper variable apoplastic effector gene of *G. pallida*. *PLoS Pathog.* 10(9), e1004391.

Eves-van den Akker, S., Lilley, C.J., Danchin, E.G.J., Rancurel, C., Cock, P.J.A., Urwin, P.E. and Jones, J.T. (2014b) The transcriptome of *Nacobbus aberrans* reveals insights into the evolution of sedentary endoparasitism in plant-parasitic nematodes. *Genome Biol. Evol.* 6(9), 2181–2194.

Fiers, M.A., Hause, G., Boutilier, K.A., Casamitjana-Martinez, E., Weijers, D., Offringa, R., van der Geest, L., Lookeren Campagne, M. and Liu, C.-M. (2004) Mis-expression of the *CLV3/ESR*-like gene *CLE19* in *Arabidopsis* leads to a consumption of root meristem. *Gen.* 237, 37–49.

Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMmer web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37.

Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J. and Hussey, R.S. (2003) The parasitism of the plant nematode *Heterodera glycines*. *Mol. Plant Microbe Interact.* 16, 726–729.

Gheysem, G. and Mittchum, M.G. (2011) How nematodes manipulate plant development pathways for infection. *Curr. Opin. Plant Biol.* 14, 1–7.

Grabher, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.

Haas, B.J., Papanicolaou, A., Yassour, M. et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494.

Hahn, C., Bachmann, L. and Chevreux, B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41(13), e129.

Hewezi, T., Howe, T., Maier, T., Hussey, R.S., Mitchum, M.G., Davis, E.L. and Baum, T.J. (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.

Haus, J.J., Papanicolaou, A., Yassour, M. et al. (2013) A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. *Plant J.* 74, 185–196.

Heterodera schachtii plant parasitism by a cyst nematode. *Genome Biol.* 15, R43.

Heterodera schachtii specific transcriptomes of *planta* parasitism. *The Plant Journal* published by Society for Experimental Biology and John Wiley & Sons Ltd, 2020 The Authors. © 2020 The Authors. *The Plant Journal* published by Society for Experimental Biology and John Wiley & Sons Ltd.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674–3676.

Craig, J.P., Bekal, S., Hudson, M., Domier, L., Niblack, T. and Lambert, K.N. (2008) Analysis of a horizontally transferred pathway involved in vitamin B6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. *Mol. Biol. Evol.* 25, 2085–2096.

Danchin, E.G.J., Guzeeva, E., Mantelin, S., Berepik, A. and Jones, J.T. (2016) Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Heterodera schachtii* to feed on host-derived sucrose. *Proc. Natl Acad. Sci. USA*, 107(41), 17651–17656.

Danchin, E.G.J., Guzeeva, E., Mantelin, S., Berepik, A. and Jones, J.T. (2016) Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Mol. Biol. Evol.* 33, 1571–1579.

Danchin, E.G.J., Guzeeva, E., Mantelin, S., Berepik, A. and Jones, J.T. (2017) The transcriptomes of *Xiphiphinae index* and *Longidorus elongatus* suggest independent acquisition of some plant parasitism genes by horizontal gene transfer in early-branched nematodes. *Genes*, 8(10), 287.

Dong, A., Stam, R., Cano, L.M. et al. (2014) Effector specialization in a lineage of the Irish Potato Famine Pathogen. *Science*, 343, 552–555.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.

Espada, M., Silva, A.C., Eves-van den Akker, S., Cock, P.J.A., Mota, M. and Jones, J.T. (2016) Identification and characterization of parasitism genes from the potato with nematode *Bursaphelenchus xylophilus* xiphidii reveals a multi-layered detoxification strategy. *Mol. Plant Pathol.* 17, 286–295.

Espada, M., Eves-van den Akker, S., Maier, T., Vijaypalani, P., Baum, T.J., Mota, M. and Jones, J.T. (2018) STATAWAARS: a promoter motif associated with the tissue-specific transcriptome of the major effector-producing glands of the plant-parasitic nematode *Bursaphelenchus xylophilus*. *BMC Genom.*, 19, 553.

Eves-van den Akker, S. and Birch, P.R.J. (2016) Opening the effector protein toolbox for plant-parasitic cyst nematode interactions. *Mol. Plant*, 9, 1481–1493.
