Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism

Charles H. Opperman[a,b,c], David M. Bird[a,b], Valerie M. Williamson[d], Dan S. Rokhsar[e], Mark Burke[a], Jonathan Cohn[a], John Cromer[a], Steve Diener[a,f], Jim Gajan[a], Steve Graham[a], T. D. Houfek[a,g], Qingli Liu[d,h], Therese Mitros[i], Jennifer Schaff[a,j], Reenah Schaffer[a], Elizabeth Scholl[a], Bryon R. Sosinski[k,l], Varghese P. Thomas[d], and Eric Windham[a]

[a]Center for the Biology of Nematode Parasitism, Department of Plant Pathology, North Carolina State University, Raleigh, NC 27695; [d]Department of Nematology, University of California, Davis, CA 95616; [e]Joint Genome Institute, Department of Energy, Walnut Creek, CA 94598; [i]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; and [k]Genome Research Laboratory, North Carolina State University, Raleigh, NC 27695

Edited by Roger N. Beachy, Donald Danforth Plant Science Center, St. Louis, MO, and approved August 18, 2008 (received for review June 20, 2008)

We have established *Meloidogyne hapla* as a tractable model plant-parasitic nematode amenable to forward and reverse genetics, and we present a complete genome sequence. At 54 Mbp, *M. hapla* represents not only the smallest nematode genome yet completed, but also the smallest metazoan, and defines a platform to elucidate mechanisms of parasitism by what is the largest uncontrolled group of plant pathogens worldwide. The *M. hapla* genome encodes significantly fewer genes than does the free-living nematode *Caenorhabditis elegans* (most notably through a reduction of odorant receptors and other gene families), yet it has acquired horizontally from other kingdoms numerous genes suspected to be involved in adaptations to parasitism. In some cases, amplification and tandem duplication have occurred with genes suspected of being acquired horizontally and involved in parasitism of plants. Although *M. hapla* and *C. elegans* diverged >500 million years ago, many developmental and biochemical pathways, including those for dauer formation and RNAi, are conserved. Although overall genome organization is not conserved, there are areas of microsynteny that may suggest a primary biological function in nematodes for those genes in these areas. This sequence and map represent a wealth of biological information on both the nature of nematode parasitism of plants and its evolution.

compaction | dauer | development | horizontal gene transfer | gene

Nematodes are an abundant and species-rich animal phylum. They share a common body plan on which various adaptations have evolved, enabling Nematoda to occupy essentially all ecological niches, including being parasites of many other organisms (1). Parasitism of plants appears to have arisen independently in three of the major 12 nematode clades (2) and results in annual losses to world agriculture estimated to exceed $US100 billion (3, 4). The majority of damage is caused by sedentary endoparasitic forms in the order Tylenchida, which includes the root-knot nematodes (*Meloidogyne* spp., RKN). RKN have a cosmopolitan distribution and a host range that spans most crops, although individual RKN species exhibit a more restricted host range. Mature female RKN release hundreds of eggs onto the surface of the root that hatch in the soil as second-stage larvae (L2) and typically reinfect the same plant. RKN L2 are similar in function to dauer larvae (5), which were first described as an adaptation to parasitism to overcome adverse environmental conditions and facilitate dispersal (6), but have been best studied in the free-living nematode *Caenorhabditis elegans* (7). These larvae are developmentally arrested, motile, nonfeeding, nonaging, and long-lived. Like *C. elegans* dauers, RKN L2 are detergent-resistant (5), use the glyoxylate pathway (8), and exhibit intestinal morphology with sparse luminal microvilli and numerous lipid storage vesicles that permit long-term survival in the soil. RKN L2 penetrate the root and migrate intercellularly into the vascular cylinder. Migration is accompanied by extensive protein secretion via the stylet, and, because changes in morphology of the pharyngeal glands correlate with the establishment of the parasitic interaction, secretions from these glands are believed to play a central role in host–parasite interactions (9). Various enzymatic functions for the secretions have been proposed, and cloning and sequencing of individual genes encoding gland proteins have permitted their natures to be discerned with confidence.

Genome sequencing of free-living nematodes *C. elegans* (10), *C. briggsae* (11), and *Pristionchus pacificus* (12) has provided reference genomes for comparison with the parasitic nematodes, and recently a draft genome sequence was published for the human parasite *Brugia malayi* (13). Here, we present the genome of the RKN species *Meloidogyne hapla*. Although many RKN species reproduce by mitotic parthenogenesis (1), many isolates of *M. hapla* reproduce by facultative meiotic parthenogenesis where sexual crosses occur, but, in the absence of males, the diploid state is restored by reuniting sister chromosomes of a single meiosis (14). We have exploited this unique genetic system for construction of a linkage map that we have anchored to the sequence and as a means to suppress DNA sequence heterogeneity to facilitate genome assembly from a whole-genome shotgun (WGS) approach. Indeed, the ability to produce highly inbred lines greatly enhanced our assembly fidelity.

A striking finding of this project is that *M. hapla* encodes ≈5,500 fewer protein-coding genes than does *C. elegans*. We have noted several gene families that are significantly smaller in number than seen in *C. elegans*, and we have further substantiated the hypothesis that horizontal gene transfer played a role in evolution of parasitism. Collectively, the acquisition of the *M. hapla* sequence represents a major step in the understanding of

Author contributions: C.H.O., D.M.B., and V.M.W. designed research; V.M.W., D.S.R., Q.L., and V.P.T. performed research; D.S.R. and M.B. contributed new reagents/analytic tools; C.H.O., D.M.B., V.M.W., D.S.R., M.B., J. Cohn, J. Cromer, S.D., J.G., S.G., T.D.H., Q.L., T.M., J.S., R.S., E.S., B.R.S., V.P.T., and E.W. analyzed data; and C.H.O. and D.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank database (project accession no. ABLG00000000). The version described in this paper is the first version, ABLG01000000.

[b]C.H.O. and D.M.B. contributed equally to this work.

[c]To whom correspondence should be addressed. E-mail: warthog@ncsu.edu.

[f]Present address: BASF Plant Sciences, Research Triangle Park, NC 27709.

[g]Present address: Center for Stem Cell Biology, Vanderbilt University, Nashville, TN 37232.

[h]Present address: Department of Biology, University of North Carolina, Chapel Hill, NC 27599.

[j]Present address: Genomic Sciences Laboratory, North Carolina State University, Raleigh, NC 27695.

[l]Present address: Department of Horticultural Science, North Carolina State University, Raleigh, NC 27695.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805946105/DCSupplemental.

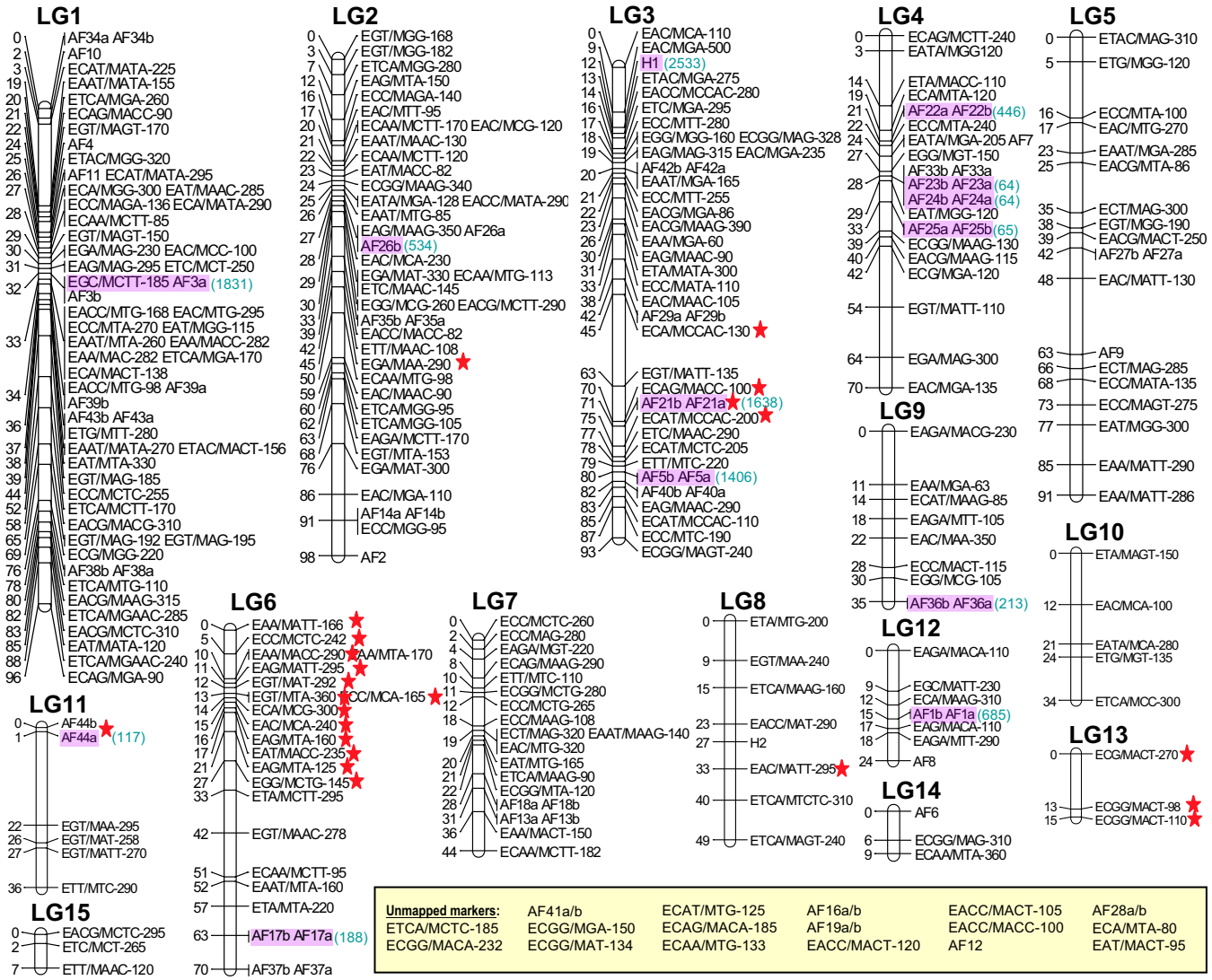

Fig. 1. Genetic linkage map of *M. hapla*. Linkage groups (LG) with three or more markers at LOD8 are shown. At this LOD, 17 of the 293 markers could not be assigned to linkage groups (yellow box). H1 and H2 are PCR markers used to monitor crosses. Markers with the same name and suffix a or b segregate as codominant markers. Stars indicate markers that deviate from 1:1 segregation at $P > 0.01$. Numbers to the left of the linkage groups indicate genetic distances in cM. Markers highlighted in pink are codominant and have been merged with the genomic sequence of VW9. Numbers in blue indicate contig number from the assembly.

nematode niche evolution. A draft genome sequence has recently been acquired for the aneuploid, mitotically parthenogenetic species *Meloidogyne incognita* (15), and the simultaneous finishing of these genomes will permit future comparative genomic approaches to study nematode parasitism and evolution. Plant-parasitic nematodes are among the most damaging and difficult-to-control pests of world agriculture. Meeting current and future worldwide demands for food, fiber, and bioenergy will necessitate minimizing these losses and will require development of new control paradigms. These genome sequences provide a new first step toward this goal.

Results and Discussion

Genetic Linkage Map of *M. hapla*. Analysis of 293 AFLP DNA markers exhibiting polymorphism between the VW8 and VW9 parents permitted assembly of a genetic map of *M. hapla* with 15 linkage groups (Fig. 1). Based on cytological examination, the parental strains have a complement of 16 chromosomes (16) and thus, most of the chromosomes may be represented in this map. The sum of the genetic distances of markers in current linkage groups is 771 cM, and from this we estimate a total genetic distance of ≈1,000 cM and an average physical to genetic distance of ≈50 kb/cM. This linkage map is the densest map for any plant-parasitic nematode. As previously noted (14), segregation is 1:1 for most markers, reflecting the novel parthenogenetic mechanism, although we found several areas where segregation substantially deviated from a 1:1 ratio, possibly indicating that traits affecting survival or parasitism map to these regions. We have merged this iteration of the map with the genome sequence by using codominant markers from the VW9 × VW8 cross (Fig. 1).

Draft Genome Sequence and *M. hapla* Gene Discovery. Multiple-sized insert libraries were constructed from DNA from the inbred line VW9, and whole-genome shotgun sequencing reactions were analyzed by 1.04 million reads yielding 587 million base pairs of sequence (supporting information (SI) Table S1). By using Arachne (17), these were assembled into 1,523 scaffolds, resulting in 10.4× coverage of the ≈54-Mb genome (18) and spanning 99% of the VW9 genome (Table 1).

Table 1. Comparison of *M. hapla* genome statistics with *C. elegans* and *B. malayi* (26)

	*C. elegans**	*M. hapla*	*B. malayi*
Genome size, Mb	100	54	90–95
Scaffolds	N/A	1,523	8,810
Scaffold N50, bp	N/A	83,645	93,771
Assembled bp	100,267,623	53,578,246	70,837,048
Sequence coverage, %	100	99.2	76.5
Gene models	21,193	14,420	11,515
Gene density	235	270	162
Median exon, bp	147	145	140
Exon/gene	6	6	7
Median intron, bp	68	55	219
G + C, %	35.4	27.4	30.5
Chromosomes	6	16	6
Predicted peptides	23,662	16,676	11,500

**C. elegans* wormbase assembly release WS185, November 2007.

Despite having an unusually low G+C content (Table 1), 83% of the VW9 genome represents unique sequence, likely contributing to the robust genome assembly. We examined the repetitive sequence and found mainly simple repeats. Approximately 1% of the repetitive sequence encodes characterized repeats, including DNA transposons (Fig. S1, Table S2). We found 323 copies of the *M. hapla* equivalent of the Tc1 transposon (19), a number almost identical to that found in *C. elegans*. Similarly, in both the number of genes and predicted structure, the SL-1 transsplice leader (20) in *M. hapla* is equivalent to *C. elegans*, but the SL-1 loci appear dispersed throughout the *M. hapla* genome. Additionally, small groups of satellites were found, as were clusters of (5S, 16S-5.8S-28S) rRNA sequences.

We used Glimmer and FgenesH (both independently trained on ≈3,000 hand-curated *M. hapla* gene models) for *ab initio* gene predictions from genomic sequence and identified 14,420 protein-coding genes in *M. hapla*. Alignment of 6,711 *M. hapla* EST unigenes derived from 26,707 ESTs showed correct alignment in orientation and order, suggesting that the sequence assembly is robust and misassembly of regions encoding genes is limited at most. Gene density is the highest yet reported for a nematode, although median gene length appears very similar to other completely sequenced larger genomes (Table 1). The average intron and exon sizes (55 and 145 bp, respectively) are very similar to those in *C. elegans* (Table 1). The protein dataset (*HapPep3*) deduced from the 14,420 predicted *M. hapla* genes was used as queries in an HMM search against the Pfam22 database, and the top 10 matches with an E value of less than E-05 were recovered. A similar analysis was done by using the wormpep185 build (www.wormbase.org) for comparison between *M. hapla* and *C. elegans*. There were 4,943 matches in *M. hapla* and 13,207 matches in *C. elegans*. Eight of the top 20 pfam domains found in *M. hapla* are also in the top 20 for *C. elegans* (Fig. 2*A*).

Gene Families in *M. hapla*. Examination of the *M. hapla* gene repertoire reveals that the single largest gene family in *C. elegans*, the G protein-coupled receptor family (GPCR: 1,011 genes) (21, 22), is drastically reduced in *M. hapla* (147 genes) (Fig. S2). Perhaps this reduced gene count represents gene loss observed during niche specialization to become an internal parasite of plants (a homeostatic environment compared with soil), with the life stages outside the plant being restricted to egg and L2 dauer (stages both with restricted neuronal access to the environment). Alternatively, this disparity may reflect gene expansion in *C. elegans* for its unique niche. A similar pattern exists for other gene families in *M. hapla* when compared with *C. elegans*. For example, the number of genes in the nuclear steroid hormone receptor family (284 genes, *C. elegans*) is ≈25% in *M. hapla* (76 genes), and *M. hapla* encodes 81 collagen genes, compared with 165 in *C. elegans*. Because we assume that nuclear steroid hormone receptors and collagens likely play key roles in basic nematode biology, these reductions are surprising and reflect either the result of strong pressures reducing the *M. hapla* genome, or expansion in *C. elegans*. Not surprisingly, almost half of the genes to which we can ascribe function in *M. hapla* (based on database similarity) show the highest similarities to *C. elegans* genes (Fig. 2*B*), with the second largest group showing similarity to animal-parasitic nematodes. Other categories are *M. hapla* genes with significant (and unique) matches to other animal and plant genes. Where possible, genes were fitted to the GO hierarchy (Fig. 2*C*).

***M. hapla* Has Suites of Genes Acquired from Bacteria via Horizontal Gene Transfer (HGT).** Because the literature contains compelling arguments to support the hypothesis that during its evolutionary history RKN has acquired genes via HGT (23, 24), we searched for *M. hapla* genes encoding proteins with significant matches only to bacterial proteins but not to other animal proteins. Since their first discovery in plant-parasitic nematodes (25, 26), many of these genes have been subjected to intensive individual scrutiny. Not surprisingly, we found many of the HGT candidates that had been identified in tylenchids, including, but not limited to, hydroxymuconicsemialdehyde hydrolase, endoglucanases, chorismate mutase, exo-polygalacturonase, glutamine synthetase, isochorismatase, L-threonine aldolase, NodL, pectate lyase, and other pectinases (9). We also discovered candidates that appear to be present in bacteria and plants, but not in animals or other eukaryotes, including a cyanate lyase and a fructofuranosidase.

In addition to identifying all of the known HGT candidates previously ascribed to RKN, we examined their genome-wide distribution, revealing, for example, that *M. hapla* encodes cellulases at six loci (Fig. S3), but that only four of those have EST support, suggesting either that two of the copies are pseudogenes or that they are expressed in a manner not sampled by ESTs. Perhaps the most fascinating example of an RKN gene apparently acquired by HGT is that of pectate lyase. Originally discovered as secreted proteins (9, 27), our sequence revealed that *M. hapla* encodes a family of 22 pectate lyases. Common in plant-pathogenic fungi, this enzyme is responsible for depolymerization of pectin, primarily in the middle lamella of the plant cell wall, and likely plays a role in nematode migration and possibly a regulatory role in feeding site formation and maintenance. Protein distance and neighbor-joining analysis of a multiple sequence alignment of these deduced proteins suggest two principal clades (Fig. 3*A*), possibly reflecting two discrete HGT events. Both clades have subsequently undergone expansion, presumably reflecting diversification of gene function during RKN evolution. In three instances, genes encoding pectate lyases most similar to each other are adjacent in the genome (Fig. 3), consistent with recent tandem duplication. Thus, despite the apparent compaction of the *M. hapla* genome compared with that of *C. elegans*, genes central to parasitic interaction have undergone expansion.

Cataloging the *M. hapla* Secretome. It is widely accepted that secretions produced by plant-parasitic nematodes play a role in pathogenesis on plants (27). Our initial examination of known secretion genes from the public databases revealed a total of 70 putative orthologs to previously identified plant-parasitic nematode-secreted proteins (Table S3). To identify new putative *M. hapla*-secreted proteins, we searched all 14,420 predicted proteins from *HapPep3* for signal sequences by using SignalP (28, 29). A total of 1,534 proteins were predicted by both available SignalP search

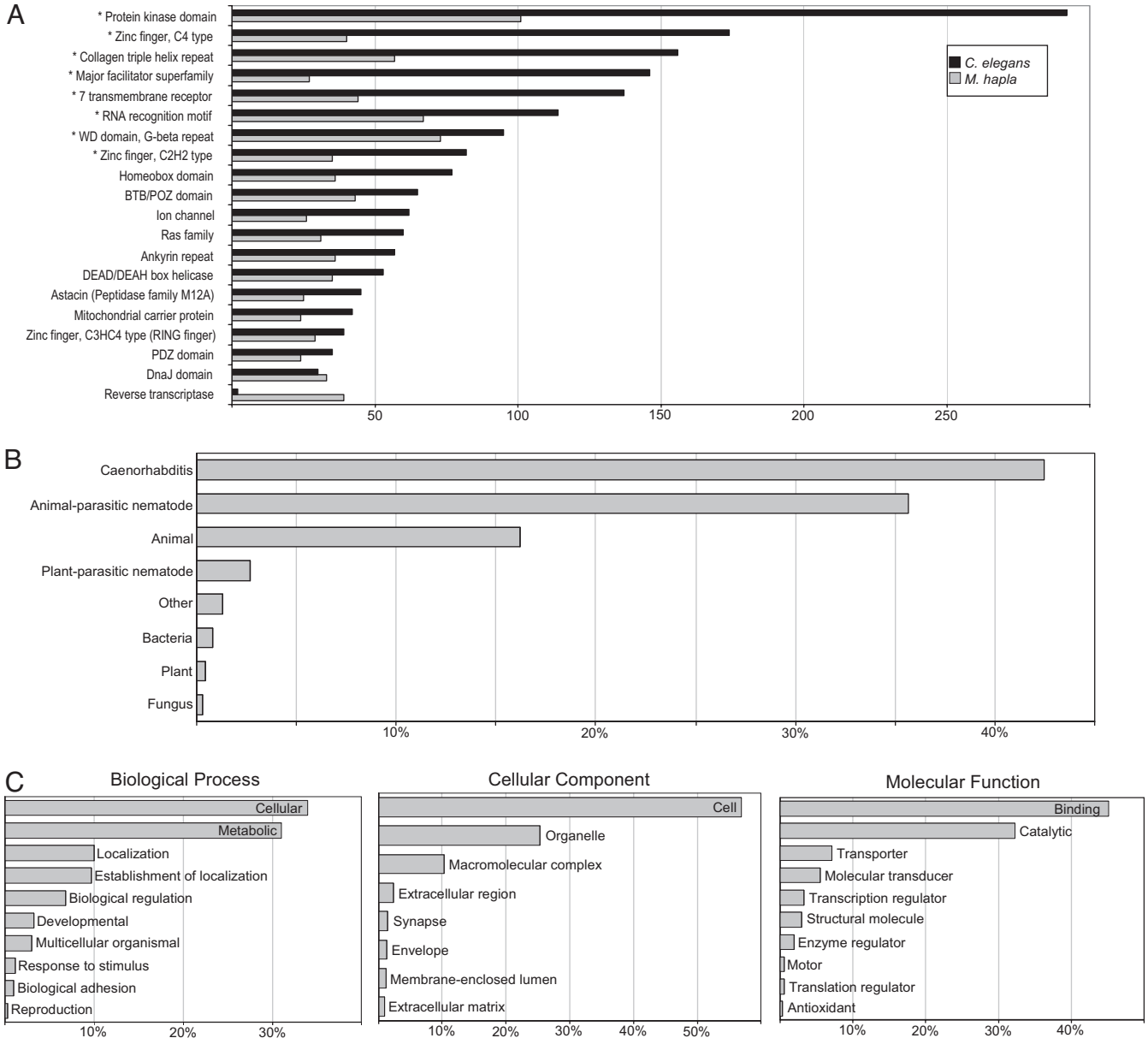

Fig. 2. Protein analysis for *M. hapla* genome. (*A*) The 20 most common protein domains found in *M. hapla* based on an HMM search of Pfam22 compared with the number of occurrences of each domain found in *C. elegans*. Domains that are within the 20 most common for *C. elegans* are indicated with a *. Number of domains detected are indicated. (*B*) Percent similarity of genomic sequence to known proteins categorized based on best BlastX match to NCBI's nonredundant protein database. (*C*) Summary of the three major GO categories based on a comparison of protein sequences to the Uniprot swissprot+trembl database.

algorithms to have signal sequences and were searched for putative transmembrane-spanning regions by using TM-HMM (30), which identified a total of 832 proteins with a secretion signal, but lacking membrane-spanning helices. These proteins were compared with the total set of *C. elegans* proteins in WormPep release 185 (http://www.wormbase.org), revealing 360 to have a significant match to a *C. elegans* protein. The remaining 472 proteins were compared with the National Center for Biotechnology Information (NCBI) nonredundant database of peptide sequences, revealing a total of 38 sequences with significant matches. Many of these matches are to plant-parasitic nematode proteins suggested to play a role in parasitism, as well as to proteins from *B. malayi* (13). The remainder are candidates for proteins that may play roles in defining the host–parasite interface; their discovery helps validate the power of whole-genome analysis. Consistent with their characters, many of these secreted proteins were not captured in our GO mapping (Fig. 2*C*).

Exploring Developmental Pathways in *M. hapla*. Many important developmental pathways in *C. elegans* are partially conserved in the *M. hapla* genome. Sex determination clearly is a key developmental event in all nematodes, yet *M. hapla* carries a very limited number of recognizable orthologs of *C. elegans* sex determination genes, most notably *tra-1* and *tra-2*, although several dosage compensation genes are also conserved. In one case, the *C. elegans* operon containing the dosage compensation gene *dpy-30* is also conserved in *M. hapla* (Fig. S4), suggesting that, although lifestyle is different between these nematodes, certain downstream events may be conserved. Genes upstream in the pathway, such as *xol-1*, *sdc-1*, *sdc-3*, and *her-1*, were not detected, suggesting that the signals that trigger these pathways

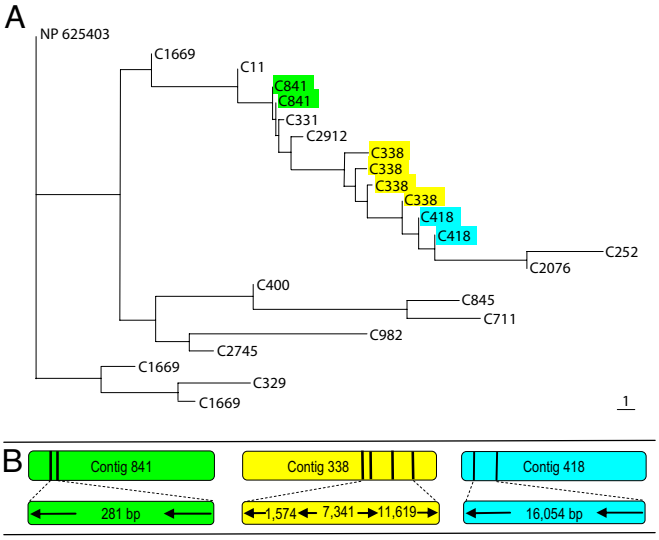

Fig. 3. The pectate lysae gene family in *M. hapla*. (*A*) Neighbor joining analysis of pectate lysases encoded by 22 genes throughout the *M. hapla* genome; these enzymes are not found in *C. elegans*. Secreted pectate lysase from *Streptomyces coelicolor* was used as an outgroup. Colors map to genomic locations shown in *B*.

are substantially diverged. In contrast, many of the genes in other *C. elegans* pathways have clear orthologs in *M. hapla*, including genes for programmed cell death, perhaps reflecting their primary roles in generalized nematode developmental biology as opposed to response to the environment.

Examination of the *M. hapla* genome indicates substantial conservation of the RNAi pathway. For example, genes involved in function of small RNAs, including *drsh-1*, *pash-1*, *dcr-1*, *drh-2*, *drh-3*, *agl-1*, *agl-2*, *rrf-3*, *eri-1*, and *pir-1*, can unequivocally be discerned in *M. hapla*, with the sole exception of *rde-4*, which is not well conserved across phylogenetic distances. The use of RNAi for examination of gene function has become an important research tool, and given appropriate selection, it is persistent over several RKN generations (31). The potential for delivery of small RNA molecules via transgenic plants for nematode control also has recently been demonstrated (32), underscoring the importance of better understanding this process in RKN.

A key biological event in the parasite life cycle is the formation of the infective stage, which for nematodes generally corresponds to the dauer larvae stage; the ability to form dauers is broadly conserved across the Nematoda. In *C. elegans*, dauer entry and exit is controlled by the environmental cues of "food signal," temperature, and nematode population density, measured based on a secreted pheromone; the precise natures of the cues used by *M. hapla* are unknown but presumably differ from those for *C. elegans*. Genetic analysis in *C. elegans* identified 32 genes as dauer-affecting (*daf*) (7). Interestingly, many of these genes are also related to life span; in some cases, mutations in *daf* genes result in exponentially longer life span in *C. elegans*. Multiple genetic pathways in *C. elegans* control both entry into and exit from the dauer stage. These pathways combine to form a system conceptually analogous to quorum sensing in bacteria and likely play a role in general environmental sensing. The molecular nature of 20 of the 32 genetically characterized *daf* genes has been discerned. *Caenorhabditis briggsae* encodes 19 of these 20, but lacks the beta-insulin molecule involved in signal transduction encoded by *daf-28*. *M. hapla* encodes strong orthologs of 14 *C. elegans daf* genes and weak orthologs of three more (Table S4). Like *C. briggsae*, *M. hapla* lacks an ortholog of *daf-28*. The molecular identities of those genes not found in *M. hapla* are associated with specific developmental cues related to lifestyle, and this demonstrates that, although the basic mechanical aspects of development are conserved, response to environment in parasite versus free-living nematode is substantially diverged.

Operon Structure and Conserved Synteny Between *M. hapla* and *C. elegans*. One intriguing feature of the *C. elegans* genome is that numerous genes are organized into operons (20). We queried *HapPep3* for orthologs of the 3,539 *C. elegans* proteins encoded by genes found in operons and identified 2,536 matches in the *M. hapla* assembly. By examining the positions of the matches for the individual members of each operon, we identified 140 operons from *C. elegans* that in *M. hapla* are at least partially conserved, which we define as having at least two genes from an operon within close proximity (Fig. S4, Tables S5 and S6). The largest operon fully conserved consists of three genes, although a larger cassette of four genes from the five-gene operon CEOP3272 also has been conserved. This analysis does not preclude the possibility that other *M. hapla* genes may be organized into operons, but suggests that the genes coregulated by virtue of such organization in *C. elegans* are not thusly regulated in *M. hapla*. Because of the small genome size, *M. hapla* intergenic regions also tend to be short, making *ab initio* prediction of operon organization difficult. However, comparison of genes in operons has shed considerable light on conservation of microsynteny between *M. hapla* and *C. elegans*. In many cases, lack of operon conservation with *C. elegans* results simply from the orthologous genes not being present in *M. hapla* (Fig. S4). Significantly, the corresponding genes in *C. elegans* generally do not have an RNAi phenotype (www.wormbase.org), suggesting either redundant or dispensable function. We also observed that although areas of microsynteny do occur in the *M. hapla* genome, for the most part, synteny is either broken or nonexistent. Taken collectively, these data may point to *M. hapla* as having a minimal and nonexpanded genome, possibly because of its obligate biotrophic lifestyle.

Conclusions

The complete genome sequence of *M. hapla* has immediate and important implications for research on both plant nematodes and for broader biology studies. The ability to perform comparative genomics by using *C. elegans*, *C. briggsae*, *P. pacificus*, *B. malayi*, and *M. incognita* as well as other plant- and animal-parasitic nematode genomes will lead to deeper understanding of the evolution of parasitic ability and comparative nematode development. Because parasitic forms exhibit developmental aspects distinct from free-living forms, these studies may reveal critical junctures in the life cycle of the obligate parasites that may be unique and specific targets for anti-nematode therapies. In particular, events such as arrested development and response to host and environmental cues may be examined in detail previously not achievable. Although the precise machinery may differ, these developmental stages represent key points in the parasite's life cycle, and we expect that some mechanisms will be conserved between species, whereas others may be vastly different. Beyond that, a better understanding of biological transitions in RKN such as the degeneration of the somatic musculature during the transition from a migratory L2 dauer to the sedentary feeding stages, and the subsequent reacquisition of functional muscles in the migratory adult males, may have implications for broad disciplines, including human health.

Many pathogenic bacteria are evolving to lose genes from pathways for which they can rely on their host to provide and therefore have smaller genomes than their free-living counterparts. The smaller size of the *M. hapla* genome *per se* compared with free-living forms could point to a similar evolutionary phenomenon, and current evidence is pointing to *M. hapla*

possessing fewer genes than its free-living counterparts. Because the parasite has a reduced gene repertoire, those lost or disabled pathways may also provide clues to the complex interaction between host and parasite. Areas such as the identification of pathogenicity islands, virulence operons, and horizontally transferred genes are also amenable to detailed study. On a broader scale, the complete genome from *M. hapla* provides a platform for studies taking advantage of whole-genome transcriptome profiling to ask pathogenicity, evolutionary, and developmental questions. Finally, the sequence provides a springboard for both survey sequencing of other species and genera in the Tylenchida and the Nematoda.

Materials and Methods

Constructing a Genetic Linkage Map of *M. hapla*. We previously established a genetic system for *M. hapla* by exploiting the facultative meiotic parthenogenetic mode of reproduction (14) and described the generation of 183 F2 lines from a cross of two highly inbred lines, VW8 and VW9 (15). Polymorphic AFLP markers were identified as described in ref. 14 and used to genotype these F2 progeny. Data were assembled into a linkage map by using JoinMap 3.0 (32).

Sequence Acquisition and Assembly. Six WGS libraries were constructed from genomic DNA isolated from *M. hapla* strain VW9 eggs, and DNA sequencing was performed at the Department of Energy Joint Genome Institute and at the Center for Biology of Nematode Parasitism on ABI 3730 and MegaBase sequencers. Trimmed sequences were assembled by using the Arachne package (Broad Institute) version 2.0.1. A full description of the libraries and sequence processing methods is provided in *SI Methods*. Refinements to these assemblies will be posted to this site as they become available. Contigs and scaffolds will be available for download, along with readme files listing the updates to the assembly.

Gene Discovery. To train gene-finding programs we first hand-curated 124 EST clusters unequivocally defining full-length *M. hapla* transcripts. We then used PASA (Program to Assemble Spliced Alignments), which assembles full-length genes by aligning EST data with genomic sequence. We parsed the PASA results for gene structure, yielding 2,974 gene models that were used to independently train Glimmer and FgenesH for *ab initio* gene discovery. Together with full-length and partial EST data, the output of Glimmer and FgenesH were manually merged to generate a robust prediction of the *M. hapla* gene complement. A full description of the methods used for discovery and annotation of particular classes and families of genes, genome features, and functional elements is provided in *SI Methods*.

Genome Browsers. Current automated and manual annotation of the *M. hapla* genome is available in Gbrowse format at www.hapla.org and at www.root-knot.org. The *M. hapla* genome also is available through the nematode site (www.wormbase.org). Features and formats for these sites differ. In addition, gene predictions and the latest version of the *M. hapla* protein database, *HapPep*, will be available for download on these sites. All genomic sequence and EST sequence is available at NCBI, as well as the previously listed sites.

ACKNOWLEDGMENTS. This work was supported by a grant from the Microbial Genome Sequencing Program of the Cooperative State, Research, Education, and Extension Service, U.S. Department of Agriculture. ESTs were generated at Washington University's Genome Sequencing Center with the support of the National Science Foundation Plant Genome panel.

1. Blaxter ML, Bird D (1997) Parasitic nematodes. *C. elegans II*, eds Riddle DL, Blumenthal T, Meyer BJ, Priess JR (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 851–878.
2. Holterman M, *et al.* (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Mol Biol Evol* 23:1792–1800.
3. Sasser JN, Freckman DW (1986) A world perspective on nematology: The role of the society. *Vistas on Nematology*, eds Veech JA, Dickson DW (Society of Nematologists, Hyattsville, MD), pp 7–14.
4. Koenning SR, *et al.* (1999) Survey of crop losses in response to phytoparasitic nematodes in the United States for 1994. *J Nematol* 31:587–618.
5. Bird DMcK, Opperman CH (1998) Caenorhabditis elegans: A genetic guide to parasitic nematode biology. *J Nematol* 30:299–308.
6. Fuchs AG (1915) Die Naturgeschichte der nematoden und einiger anderer parasiten. I. Des *Ips typographus* L. 2. Des *Hylobius abietis* L. *Zool Jahrb Abt Systematik Ökol Geogr Tiere Jena* 38:109–122.
7. Riddle DL, Alberts PA (1997) Genetic and environmental regulation of dauer larva development. *C. elegans II*, eds Riddle DL, Blumenthal T, Meyer BJ, Priess JR (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 739–768.
8. McCarter JP, *et al.* (2003) Analysis and functional classification of transcripts from the root-knot nematode *Meloidogyne incognita*. *Genome Biol* 4:R26.
9. Davis EL, Mitchum MG (2005) Sophisticated parasites of legumes. *Plant Physiol* 137:1182–1188.
10. *C elegans* Sequencing Consortium (1998) Genome sequence of the nematode *Caenorhabditis elegans*. A platform for investigating biology. *Science* 282:2012–2018.
11. Stein LD, *et al.* (2003) The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol* 1:E45.
12. Dieterich C, *et al.* (2008) The *Pristionchus pacificus* genome provides a unique perspective on nematode life-style and parasitism. *Nat Genet*, in press.
13. Ghedin E, *et al.* (2007) Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317:1756–1760.
14. Liu QL, Thomas VP, Williamson VM (2007) Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. *Genetics* 176:1483–1490.
15. Abad P, *et al.* (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat Biotech* 26:909–915.
16. Liu QL, Williamson VM (2006) Host-specific pathogenicity and genome differences between inbred strains of *Meloidoygne hapla*. *J Nematol* 38:158–164.
17. Batzoglou S, *et al.* (2002) ARACHNE: A whole-genome shotgun assembler. *Genome Res* 12:177–189.
18. Coghlan A (2005) in *C. elegans WormBook*, ed The *C. elegans* Research Community, 10.1895/wormbook.1.15.1. Available at: http://www.wormbook.org. Accessed June 10, 2008.
19. Bessereau J-L (2006) *C. elegans WormBook*, ed The *C. elegans* Research Community, 10.1895/wormbook.1.70.1. Available at: http://www.wormbook.org. Accessed June 10, 2008.
20. Blumenthal T (2005) *C. elegans WormBook*, ed The *C. elegans* Research Community, 10.1895/wormbook.1.5.1. Available at: http://www.wormbook.org. Accessed June 10, 2008.
21. Robertson HM, Thomas JH (2006) *C. elegans WormBook*, ed The *C. elegans* Research Community, 10.1895/wormbook.1.66.1. Available at: http://www.wormbook.org. Accessed June 10, 2008.
22. Bargmann CI (2006) *C. elegans WormBook*, ed The *C. elegans* Research Community, 10.1895/wormbook.1.123.1. Available at: http://www.wormbook.org. Accessed June 10, 2008.
23. Scholl EH, Thorne JL, McCarter JP, Bird DMcK (2003) Horizontally transferred genes in plant-parasitic nematodes: A high-throughput genomic approach. *Genome Biol* 4:R39.
24. Blaxter ML (2007) Symbiont genes in host genomes: Fragments with a future? *Cell Host Microbe* 2:211–213.
25. Yan Y, *et al.* (1998) Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. *Gene* 220:61–70.
26. Smant G, *et al.* (1998) Endogenous cellulases in animals: Isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci USA* 95:4906–4911.
27. Vanholme B, *et al.* (2004) Secretions of plant-parasitic nematodes: A molecular update. *Gene* 332:13–27.
28. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
29. Nielsen H, Krough A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* 6:122–130.
30. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567–580.
31. Gleason CA, Liu QL, Williamson VM (2008) Silencing a candidate nematode effector gene corresponding to tomato resistance gene *Mi-1* leads to acquisition of virulence. *Mol Plant Microbe Interact* 21:576–585.
32. Huang G, Allen R, Davis EL, Baum TJ, Hussey RS (2006) Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proc Natl Acad Sci USA* 103:14302–14306.
33. Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join map. *Plant J* 3:739–744.