Evidence for Horizontally Transferred Genes Involved in the Biosynthesis of Vitamin B_1, B_5, and B_7 in *Heterodera glycines*

James P. Craig, Sadia Bekal, Terry Niblack, Leslie Domier, Kris N. Lambert

Abstract: *Heterodera glycines* is a nematode that is highly adapted to manipulate and parasitize plant hosts. The molecular players involved in these interactions have only recently begun to be identified. Here, the sequencing of the second stage juvenile transcriptome, followed by a bioinformatic screen for novel genes, identified seven new genes involved in biosynthesis and salvage of vitamins B_1, B_5, and B_7. With no confirmed reports in the literature, each of these biosynthesis pathways is believed to have been lost in multicellular animals. However, eukaryotic-like introns in the genomic sequences of the genes confirmed eukaryotic origin and nematode-specific splice leaders found on five of the cDNAs confirmed their nematode origin. Two of the genes were found to be flanked by known nematode sequences and quantitative polymerase chain reactions on individual nematodes showed similar and consistent amplification between the vitamin B biosynthesis genes and other known *H. glycines* genes. This further confirmed their presence in the nematode genome. Similarity to bacterial sequences at the amino acid level suggested a prokaryotic ancestry and phylogenetic analysis of the genes supported a likely horizontal gene transfer event, suggesting *H. glycines* re-appropriated the genes from the prokaryotic kingdom. This finding complements the previous discovery of a vitamin B_6 biosynthesis pathway within the nematode. However, unlike the complete vitamin B_6 pathway, many of these vitamin B pathways appear to be missing the initial enzymes required for full *de novo* biosynthesis, suggesting that initial substrates in the pathways are obtained exogenously. These partial vitamin B biosynthesis enzymes have recently been identified in other single-celled eukaryotic parasites and on rhizobia symbiosis plasmids, indicating that they may play an important role in host-parasite interactions and survival within the plant environment.

Key words: *Heterodera glycines*, vitamin B_1, vitamin B_5, vitamin B_7, thiamin, pantothenate, biotin, biosynthetic pathways, horizontal gene transfer.

Heterodera glycines, commonly known as the soybean cyst nematode (SCN), is an obligate plant parasite and an important pest of soybean plants. The nematodes survive by creating a feeder cell, called a syncytium, within its plant host. This process is presumably started by injecting a mixture of metabolites and proteins into a plant cell, causing them to become highly metabolically active and to merge with nearby cells. The syncytia then form close associations with the phloem, facilitating the inward draw of nutrients and metabolites (Endo, 1998; Jones, 1981). In the end, the plant cells are essentially transformed into a metabolic sink from which the nematode feeds for the rest of its life.

Studies involving this plant parasitic nematode-host interaction have largely focused on identifying and characterizing genes encoding secreted parasitism-associated proteins (Bellafiore, et al., 2008; Curtis, 2007; Gao, et al., 2001; Meutter, et al., 2001; Tucker, et al., 2005; Vanholme, et al., 2004; Wang, et al., 2001). This has exposed a rich array of secreted proteins that could potentially manipulate plant physiology, metabolism, and defenses. A few of these include enzymes that can alter the cell wall (cellulase, pectinase, expansins), protease machinery (ubiquitin extensins), reactive oxygen state (peroxidases, superoxide dismutase), cell differentiation (*CLAVATA3*), the shikimate pathway (chorismate mutase), as well as nuclear targeted proteins that may directly reprogram the plant cell (Doyle and Lambert, 2002; Elling, et al.,

2007; Lambert, et al., 1999; Molinari and Miacola, 1997; Qin, et al., 2004; Smant, et al., 1998; Tytgat, et al., 2004; Wang, et al., 2005). However, this has left non-secreted proteins, a class of proteins that can also have a large impact on host-parasite interactions, largely unexplored. One such group of non-secreted proteins may be enzymes involved in the nutritional status of the nematode.

The need for nutrition may have been one of the driving forces behind the evolution of the feeding cell. Thus, a better understanding of the specific nutrient requirements of the worm could provide important clues as to how the nematode manipulates the plant to obtain them. Yet, little is known about the nutritional requirements of *H. glycines*. Studies of the nematode *Caenorhabditis elegans* have shown that like all animals, there is a strict requirement for B vitamins and other essential amino acids, which are absorbed in the intestine and salvaged (Liu and Lu, 2008; Nicholas, et al., 1962; Szewczyk, et al., 2003; Vanfleteren, 1973). It was presumed that SCN required the same essential nutrients common to animals. However, we recently reported the presence of the horizontally transferred gene pair *HgSNZ* and *HgSNO*, a complete functional pathway capable of synthesizing vitamin B_6 *de novo* (Craig, et al., 2008). Because most animals have long since lost the ability to synthesize vitamin B_6, this discovery led to interesting questions about the nematode's nutritional requirements.

Here, we report that not only does *H. glycines* possess genes to synthesize its own vitamin B_6, it has additional genes that express proteins and enzymes involved in the *de novo* biosynthesis of vitamin B_1, vitamin B_5, and vitamin B_7; hereafter referred as the vitamin B biosynthesis genes and proteins. These genes were revealed in a bioinformatic screen of the second-stage

juvenile (J2) transcriptome and like the vitamin B_6 pathway, they are most closely related to bacterial and fungal genes and may have arisen in the *H. glycines* genome through horizontal gene transfer (HGT).

MATERIALS & METHODS

Nematode Culture and RNA Extraction for Sequencing: *H. glycines* nematodes (isolates TN10 and TN16) were raised on *Glycine max* (Essex cultivar) and cysts were harvested from roots as described by Niblack et al., (1993). The cysts were crushed over a 0.25 mm sieve to release the eggs and hatched on tissue paper suspended over a water reservoir. Hatched J2 were collected on a 10 μm polycarbonate membrane (GE Osmonics Inc., Minnetonka, MN), pelleted in 1% carboxymethylcellulose, and stored at -80°C until needed. RNA was extracted following the procedure described by Bekal et al., (2003). Briefly, the frozen J2 were freeze-fractured with a tissue pulverizer and digested for 12 hours in a buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl) containing 2 mg/ml protease K (Invitrogen, Carlsbad, CA). After acid-phenol extraction, chloroform extraction, and ethanol precipitation, the quality of the RNA was checked on a formaldehyde gel and a Bioanalyzer (Agilent, Santa Clara, CA).

cDNA Library Construction: Approximately 30 μg of the total SCN RNA was used to select mRNA with the Oligotex mRNA mini kit (Qiagen, Germantown, CA). cDNAs were synthesized from mRNA with the Creator SMART cDNA synthesis kit (Clontech, Mountain View, CA) with two modifications: 1) the oligo-dT for priming the reverse transcription was modified (CDSIII-First: 5'-TAGAGACCGAGGCGGCCGACATGTTTTGTTTTTTTTTCTTTTTTTTTTTTVN -3') to break the poly-A run of the mRNA in the final cDNA; this modification has been found to overcome the high failure rate observed upon pyrosequencing cDNAs with long homopolymer runs; 2) after amplification, the double-stranded cDNA was size selected on a 2% agarose gel to eliminate fragments smaller than 400 bp.

Normalization of the cDNA Library: The cDNA library was normalized with the Trimmer Direct Kit (Evrogen, Moscow, Russia). In brief, 300 ng of cDNA was incubated at 95°C for 5 min followed by incubation at 68°C for 4 hours in the hybridization buffer included in the kit (50 mM HEPES, pH 7.5 and 0.5 M NaCl). After incubation, the reaction was treated with 0.25 units of duplex specific nuclease (DSN). The normalized cDNA was amplified from 1 μl of DSN-treated cDNA by PCR with the CDSII-First and SMARTIV primers with the following conditions: 98°C (30 sec); 10 cycles of 98°C (10 sec), 68°C (30 sec), 72°C (30 sec); followed by extension at 72°C (5 min).

454 Library Construction, Emulsion PCR, and Sequencing: 3 μg of amplified normalized cDNA was sheared into fragments 300-800 bp in size, blunt-ended, ligated to adaptors, and converted to a single-stranded template DNA library with the GS FLX General Library Prep Kit (Roche, Basel, Switzerland). Libraries were quantified with Picogreen Quantitation reagents (Invitrogen, CA) and average fragment sizes were determined by analyzing 1 μl on the Bioanalyzer (Agilent, CA) with a DNA 7500 chip. The library was diluted to 1×10^6 molecules/μl and processed for emulsion PCR, titrated, and sequenced on the GS FLX machine following the manufacturer's protocols (Roche, Switzerland).

Bioinformatic Screen: The cDNA reads were assembled in CLC Bio Genomic Workbench (CLC Bio, Cambridge, MA). Using a Perl script, the contigs and unassembled reads were translated into amino acid sequences according to their six open reading frames (ORFs), keeping only those at least 50 amino acids in length. The *H. glycines* amino acid sequences were compared to Hidden Markov models (HMM) from the Sanger database (http://pfam.sanger.ac.uk) that were found in archaea, bacteria, plants, and fungi, but not in metazoans with the program HMMER 2.3.2 (Finn, et al., 2006). The basic local alignment search tool (BLAST) suite of programs was used to find homologues in the National Center for Biotechnology Information (NCBI) databases (Altschul, et al., 1990).

Sequencing the Genomic DNA of the Vitamin B Synthetic Genes: Sequences that matched a HMM were assembled in Sequencer 4.9, and the contigs bioinformatically expanded using BLAST and a SCN nucleotide database obtained from NCBI. To determine the complete ORF of each gene, contigs missing a start or stop codon were extended and sequenced using the Genome Walker kit (ClonTech, CA) as described by Craig et al., 2008. To genome walk, the AP1 primer and a gene-specific genwalk primer (supplemental Table 1) were used. Primer pairs (supplemental Table 1) were then designed around the complete ORFs and polymerase chain reaction (PCR) amplified from SCN genomic DNA. The PCR products were T/A-TOPO cloned into a pCR2.1 or pCR4 TOPO vector and transformed into *Escherichia coli* TOP10 chemically competent cells (Invitrogen, CA). Sanger sequencing of the inserts was performed at the University of Illinois W.M. Keck Biotechnology Center.

The *thi4* and *bioB* genomic sequences were amplified and cloned in pieces due to their large size. In addition, the thiM-thiE-F/R primer pair (supplemental Table 1) was used to clone and sequence the intergenic region between the *thiE* and *thiM* genes, while the thiE-thiD-F/R primer pair sequenced the intergenic region between *thiD* and *thiE*.

Cloning the Vitamin B cDNA: RNA was extracted from SCN as described above, and each gene was transcribed into cDNA using AMV reverse transcriptase (Promega, WI) and a gene-specific primer (supplemental Table 2). Each gene was PCR amplified with a primer pair in supplemental Table 2, cloned, and sequenced as described above.

Phylogenetic Analysis: The NCBI protein database was searched for homologues using the predicted amino acid sequences of each SCN vitamin B biosynthesis gene. The top 10 protein hits representing a unique genus were collected from plants, fungi, single-cell eukaryotes, bacteria, archaea, and animals, and aligned with CLUSTALW (Thompson, et al., 1994). Maximum likelihood phylogenetic trees were generated from the amino acid sequence alignments using the program PHYML 3.0 (Guindon and Gascuel, 2003). The rate of amino acid substitution was estimated using a Whelan and Goldman (WAG) model. The JTT, Dayhoff, and Blossom62 models were also evaluated, though in general, WAG produced the highest log likelihoods. The gamma distribution parameter and proportion of invariable sites were estimated, and a random BioNJ tree was used as the starting point for PHYML analysis. The reliability of the resulting trees was estimated with the approximate likelihood ratio test (aLRT) with SH-like support.

Real-time PCR of Individual Nematodes: Newly hatched J2 were allowed to crawl through a sand column equilibrated with 0.1% sodium dodecyl sulfate to physically remove residual plant debris, fungi, and bacteria associated with the cuticle (Craig, et al., 2008; Painter and Lambert, 2003). Individual nematodes were then digested overnight at 60°C in 5 μl buffer containing 50 mM Tris-HCl (pH7.5), 50 mM NaCl, and 4 mg/ml protease K (Invitrogen, CA). The protease K was heat killed by incubation at 80°C for 20 min and the reaction was transferred into a PCR plate containing Sybr Green Master Mix (Applied Biosystems, Foster City, CA) and 5 picomoles of a primer pair listed in supplemental Table 3. The efficiency of amplification of all primer pairs were shown to be equal by conducting real-time PCR on a genomic DNA dilution series as described in Livak et al., (2001). Real-time PCR was conducted using an ABI 7900HT Sequence Detection System (Applied Biosystems) with the following reaction conditions: 50°C (2 min); 94°C (5 min); 40 cycles of 94°C (30 sec), 60°C (1 min); dissociation stage. Six replicates and two water controls were performed for each of the 10 genes: *bioB, elongation factor 1α (EF1α), HgSNZ, HgSNO, panC, tenA, thi4, thiD, thiE,* and *thiM.* This experiment was conducted twice with similar results.

Results

Identification of the Vitamin B Biosynthesis Genes: To identify unique and unusual genes in *H. glycines,* mRNA from J2 was converted to cDNA and pyrosequenced. The two sequencing runs from SCN isolate TN16 cDNA yielded 253,158 and 293,950 sequences. Runs from the SCN isolate TN10 yielded 220,106 and 276,603 sequences, with an average sequence length of 200 bases, and 189,472,067 million bases in total. This represented approximately 10-fold coverage of the J2 tran-

scriptome, based upon an estimated 20,000 genes, each producing an mRNA with an average length of 1 kb.

The sequences were assembled into 65,535 contigs, from which 201,684 amino acid sequences of at least 50 residues were selected for analysis. To identify genes potentially introduced into the SCN genome by HGT, these amino acid sequences were compared to a database of plant, bacteria, archaea, and fungal HMMs not normally found in animals. The initial search identified five hits to enzymes in the vitamin B_1, B_5, and B_7 biosynthetic pathways, which prompted us to search for additional members of the pathways using bacterial gene sequences as queries in BLAST searches of the generated SCN transcriptome database. This yielded two more genes involved in the salvage of vitamin B_1 that are rarely found in animals. A search of the NCBI database with the seven newly discovered SCN genes returned hits to plant parasitic nematodes, but no other animals, showing that the genes were most similar to bacterial sequences (Table 1).

Origin of the Biosynthetic Genes: Because the genes were most similar to bacterial sequences, it was important to determine whether they originated from the SCN genome or from either external or endosymbiotic contaminants. Initially, all SCN genomic DNA from NCBI that matched the vitamin B genes, were aligned together. The number of contigs formed suggested that most of the genes were single copy; *panC* with two contigs suggested a multi-copy gene. To investigate the context in which the genes resided, the genomic sequence flanking each contig was determined by chromosome walking and bioinformatically contiging outward. Genomic sequences flanking the vitamin B biosynthetic genes contained sequences that were closely related to other nematode genes. The contig homologous to *tenA* was flanked by two nematode genes; a serpin protease inhibitor and a Twik gene belonging to a family of potassium channels. The contig homologous to *bioB* was adjacent to an animal-like gene containing the ELMO/CED-12 domain. The presence of linked animal-like or nematode-like genes strongly suggested that these two vitamin B biosynthetic genes are contained within the SCN genome. Sequencing the flanking genomic DNA also revealed that *thiD, thiE,* and *thiM* are located adjacent to each other.

The cDNA of the vitamin B genes were sequenced next and aligned to their genomic counterparts, which showed that each vitamin B gene contained multiple introns and had poly-A signals (Table 2). The introns followed the canonic eukaryotic GT and AG splice donor and acceptor sites, respectively, which supported their eukaryotic origin. The nematode origin of the genes was investigated by performing real-time PCR on each vitamin B biosynthesis gene from individual, sterilized nematodes. Since each worm has the same number of cells and thus the same amount of genomic DNA, genes that are part of the SCN genome should

TABLE 1.   Vitamin B genes in SCN, their top BLAST hit, and respective biosynthesis pathways.

NCBI Acc #	Gene	Top BLAST Hit	Domain	E-Value	Biosynthesis Pathway
GQ890690	*bioB*	*Rickettsia endosymbiont of Ixodes scapularis*	Bacteria	5e-111	VB_7 - Biotin
GQ890689	*panC*	*Clavibacter michiganensis*	Bacteria	7e-58	VB_5 - Pantothenate
GQ890688	*thiD*	*Methylobacterium extorquens*	Bacteria	6e-49	VB_1 - Thiamin
GQ890687	*thiE*	*Tolumonas auensis*	Bacteria	1e-45	VB_1 - Thiamin
GQ890691	*thi4*	*Schizosaccharomyces japonicus*	Fungi	5e-95	VB_1 - Thiamin
GQ890693	*thiM*	*Pectobacterium carotovorum*	Bacteria	1e-53	VB_1 -Thiamin salvage
GQ890692	*tenA*	*Comamonas testosteroni*	Bacteria	1e-37	VB_1 - Thiamin salvage
EU747298	a*SNO*	*Caldicellulosiruptor saccharolyticus*	Bacteria	4e-36	VB_6 - Pyridoxal Phosphate
EU747297	a*SNZ*	*Ostreococcus lucimarinus*	Bacteria	8e-108	VB_6 - Pyridoxal Phosphate

([a]) Genes discovered and described previously in SCN.

have consistent cycle threshold (Ct) values from one individual to the next. The results in Table 3 show that each vitamin B gene was amplified consistently in each of six individual nematodes examined. The small standard deviations from the average Ct (0.04 to 0.36 cycles) indicated that there was little variability in the copy numbers of the genes among nematodes and among genes. In addition, the average Ct values and standard deviations of the vitamin B biosynthetic genes mirrored closely those of the three positive controls (single copy genes *HgSNZ* and *HgSNO,* and the multi-copy gene *EF1α*). The *panC* and *EF1α* genes had a slightly lower average Ct, which reflects their multi-copy status in the *H. glycines* genome.

Finally, in our SCN cDNA database, three genes (*panC, thiD*, and *thiM*) were found to have the nematode splice leader SL-1, attached to their 5' ends (Table 2). Two previously deposited ESTs in NCBI also had splice leaders attached to *thiE* (CD748263) and *tenA* (CB299495)-like sequences. Because the SL-1 sequence is specific only to nematodes, its presence further confirmed the nematode origin of these genes (Davis, 1996).

Protein Analysis: The coding regions of the SCN vitamin B genes were co-linear with their bacterial or fungal homologues, without missing or additional domains. To determine if the predicted amino acids contained conserved active-site residues, each protein was aligned to a homologue whose 3D crystal structure and enzymatic activity had been determined (supplemental data). These alignments showed significant homology between the enzymes, including conservation of amino acids predicted to bind substrates or having catalytic activity in the enzyme active site (Table 4). The few

amino acids with predicted function that diverged in the SCN homologues were not found to be highly conserved among homologues in other species. The recovery of mRNAs from each of the vitamin B genes that contain complete ORFs and the conservation of catalytic amino acid residues suggests that they could be functional.

Phylogenetic Analysis: Because *de novo* vitamin B biosynthesis is normally absent from multicellular animals, and BLAST results point towards bacteria as the nearest homologue, a phylogenetic analysis was performed upon each gene to determine its evolutionary origin. BLAST was used to collect the top 10 protein hits from plants, fungi, bacteria, archaea, and single-celled eukaryotes. A more extensive search through the NCBI databases returned a few hits to multicellular animal sequences. However, they could not be confirmed. The sequences had very high nucleotide sequence identity to bacteria with no identifiable introns, or were from organisms that have known bacterial endosymbionts. Therefore, we concluded that these multicellular animal sequences were likely bacterial contaminants and removed them from subsequent data analyses. The only animal sequences that seemed genuine, due to the presence of introns, were from the closely related plant parasitic nematode genera *Meloidogyne* and *Pratylenchus.*

TABLE 2.   Characteristics of B vitamin biosynthetic genes

Gene	Expressed	# Exons	Poly-A Signal	Splice Leader	Copy #
bioB	Yes	8	Yes	No	1
panC	Yes	6	Yes	Yes	>1
thiD	Yes	4	Yes	Yes	1
thiE	Yes	4	Yes	Yes	1
thi4	Yes	7	Yes	No	1
thiM	Yes	5	Yes	Yes	1
tenA	Yes	3	Yes	Yes	1

TABLE 3.   Real-time PCR on individual sterilized nematodes

Gene Name	Average Ct	Ct Standard Deviation	Relative Copy Number	H_2O Control Ct
Elongation factor 1α	22.79	0.08	2.0	>41
HgSNO	23.69	0.13	1.1	No amp.
HgSNZ	23.48	0.20	1.2	No amp.
bioB	23.61	0.34	1.1	No amp.
panC	23.02	0.36	1.7	No amp.
thi4	23.73	0.11	1.0	No amp.
thiD	23.55	0.11	1.2	No amp.
thiM	23.50	0.08	1.2	39.29
tenA	23.44	0.13	1.3	No amp.
thiE	23.48	0.14	1.2	No amp.
Permease	23.80	0.04	1.0	No amp.

Data is reported as the gene amplified, followed by the average cycle threshold (Ct), the standard deviation from the average Ct, and the relative copy number of each gene based upon their average Ct.

TABLE 4.   Conserved amino acids within vitamin B biosynthesis enzymes

Enzyme	Predicted Function	3D Structure	% Identity	Conserved Amino Acids
BIOB	Biotin synthase	*E. coli* (1R30)	59%	C53, C57, C60, C128, C188, C97, R260, T292
PANC	Pantoate β-alanine ligase	*M. tuberculosis* (3COV)	39%	H44, H44, N69, Q72, K160
THIE	Thiamine-phosphate diphosphorylase	*B. subtilis* (1G69)	31%	Q57, R59, K159, G168, S209, I208, T156, T158, K61
THID	HMP-P kinase/HMP-PP kinase	*S. typhimurium* (1JXI)	45%	E44, C213, G212, K176, E142
THI4	Thiazole synthase/mtDNA damage tolerance	*S. cerevisiae* (2GJC)	49%	F244, V170, E97, G105, H237, R301, G303, D207, G104
THIM	Hydroxyethylthiazole kinase	*B. subtilis* (1ESQ)	37%	R121, T168, D172, M45
TENA	Thiaminase II	*B. subtilis* (1YAF)	28%	Y112, F208, C135, D44, Y163, E205

Once the homologous protein sequences were aligned using CLUSTALW, trees were estimated using maximum likelihood and the branches analyzed with aLRT. Within each tree, *H. glycines* and the other nematode homologues were found together within the prokaryotic clade (Fig. 2). THI4, however, was clearly more closely associated with eukaryotes, specifically fungi. In each tree, plants, fungi, and archaea also largely separated into their own clades; as expected, the single-cell eukaryotes were less monophylogenetic. The archaea were not represented in the PANC tree as they have a separate pathway for pantothenate biosynthesis and do not have PANC homologues (Ronconi, et al., 2008).

Putative HGT Island and panC: A BLAST search of *H. glycines panC* in the NCBI database returned hits to a 91-kb contig (ABLG01000223) from the related plant parasitic nematode *Meloidogyne hapla*. Located on this contig are sequences similar to the bacterial genes β-*1-4 endoglucanase, NodL,* and *panC,* marking the region as a putative HGT island (Fig. 1B). Also included in the region are two sequences that match the *Meloidogyne incognita* putative gland cell secretory protein 32 and gland cell specific *SXP/RAL-2,* indicating a possible pathogenic island.

DISCUSSION

Nematode Origin of Vitamin B Genes: In this paper, we discovered seven genes that appear to be involved in *de novo* biosynthesis and salvage of biotin (*bioB*), pantothenate (*panC*), and thiamin (*tenA, thi4, thiD, thiE,* and *thiM*). Each of these genes has high homology to bacteria or fungi and is not normally found in animals. Although precedence for unusual vitamin biosynthetic pathways has been reported earlier with the discovery of a vitamin B_6 biosynthesis pathway in SCN, it was important to establish nematode origin.

In addition to our own sequencing projects, EST and genome sequencing projects from SCN and other plant parasitic nematodes have independently sequenced these genes. The nematode-like sequences surrounding *tenA* and *bioB* further strengthens the argument for nematode origin and eukaryotic introns remove the possibility of prokaryotic contamination or endosymbionts. Real-time PCR on individual nematodes also showed consistent amplification at levels comparable to other nematode genes and the nematode specific splice leader SL-1 was found attached to the 5' ends of *panC, tenA, thiD, thiE,* and *thiM* cDNA sequences. Together, these results all point towards localization of the genes within the SCN genome.

Evidence for HGT: Once nematode origin had been established, the significant similarity between bacterial and SCN vitamin B biosynthesis enzymes became a strong case for HGT. HGT has been shown to be rather common in *H. glycines* and other plant parasitic nematodes, and may have contributed significantly to the ability of nematodes to parasitize plants (Bird, et al., 2003; Smant, et al., 1998). In nematodes of the genus
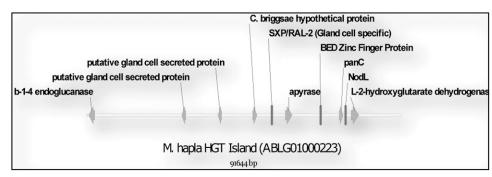

FIG. 1.   A. Graphical representation of the *H. glycines thiM, thiE,* and *thiD* genes. B. Graphical representation of the *M. hapla* 91-kb contig with *panC,* other predicted HGT genes, and gland specific genes.

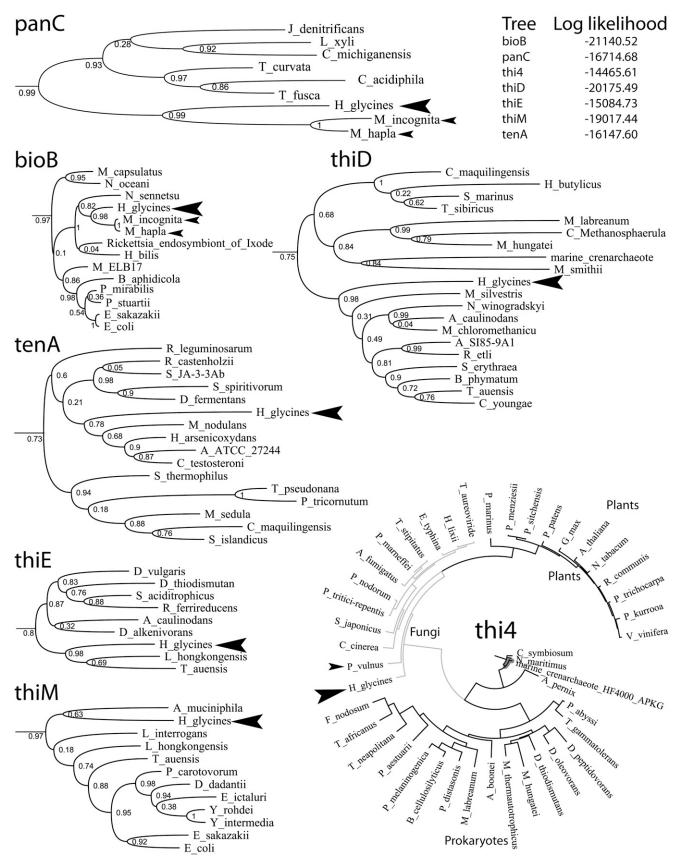| Tree	Log likelihood
bioB	-21140.52
panC	-16714.68
thi4	-14465.61
thiD	-20175.49
thiE	-15084.73
thiM	-19017.44
tenA	-16147.60

FIG. 2.   Prokaryotic subtrees of the vitamin B biosynthesis enzymes. aLRT branch supports are labeled at each node. Complete color coded trees with full species names can be found in the supplemental data; for clarity, a near complete THI4 tree is shown. *H. glycines* is marked with a large arrow, while other nematode species are marked with a small arrow. The log likelihood for each tree is also shown.

Meloidogyne, the bacterium *Mesorhizobium loti* was identified as a potential candidate for the origin of some bacterial-like genes through the transfer of a circular chromosome carrying a symbiosis island (Scholl, et al., 2003). Interestingly, vitamin B biosynthesis genes are also located within this island (Sullivan and Ronson, 1998).

The hypothesis of a HGT event from bacteria is supported by phylogenetic analysis of the nematode vitamin B biosynthesis proteins. Within each phylogenetic tree, the nematode proteins, excluding THI4, were situated in strongly supported bacterial clades (Fig. 2). THI4 is the exception among SCN's bacterial-like vitamin B enzymes. This enzyme has homologues in fungi, plants, archaea, and a few bacteria. By itself, THI4 synthesizes hydroxyethyl thiazole, a process that would normally require six enzymes in most bacteria (Begley, et al., 1999; Park, et al., 2003). Within the THI4 phylogenetic tree, the nematode homologue is only distantly related to archaea and bacteria, instead showing the strongest relation to fungi. It remains unclear if the gene is a remnant of the original animal biosynthesis pathway or the result of HGT from another eukaryote. While HGT between eukaryotes are becoming more widely recognized, this would represent the first such case in SCN (Keeling and Palmer, 2008).

Another factor supporting a prokaryotic origin for *panC* is its presence on the bacterial-like HGT island from *M. hapla*. Although large sequence contigs containing *panC* are not available for SCN yet, we were able to verify that the vitamin B biosynthesis genes *thiM*, *thiE*, and *thiD* are adjacent to each other, separated by approximately 200 bases. In bacteria, these genes are often found in the same order and direction as members of an operon (Rodionov, et al., 2002), and like the *Meloidogyne* putative HGT island, we speculate that the three genes were originally transferred into SCN during a single transfer event.

Roles of the Vitamers and their Biosynthesis Enzymes: Once the vitamin B genes were obtained in a HGT event(s), there must have been an evolutionary advantage to maintain them. Instead of degenerating over time, each gene would eventually acquire introns, with some developing a splice site for a SL, and be actively expressed. In order to better understand what this advantage may be, we analyzed the predicted role of the enzymes within their respective pathways as well as the role of the synthesized vitamins.

The vitamins B_1 (thiamin), B_5 (pantothenate), and B_7 (biotin) are essential to all life. As members of the water-soluble vitamin B family, they act as cofactors for a wide range of enzymes which in turn participate in a myriad of metabolic pathways. Thiamin-PP, the main biologically active form of thiamin, plays an important role in carbohydrate and amino acid metabolism as well as the pentose phosphate pathway (Lonsdale, 2006). Biotin functions as a carboxyl carrier for biotin dependent carboxylases, which are critical for fatty acid metabolism and amino acid catabolism. Biotin has also been shown to play a role in cell signaling, epigenetic regulation of genes, and chromatin structure (Zempleni, et al., 2009). Pantothenic acid is necessary for coenzyme A and acyl carrier protein (ACP) biosynthesis. These compounds are vital in the tricarboxylic acid cycle and fatty acid metabolism, and are also used for cholesterol and acetylcholine biosynthesis (Leonardi, et al., 2005).

In general, multicellular animals, including *C. elegans*, have lost the capability to synthesize these essential vitamins. It is assumed the gene loss occurred because animals can simply absorb the vitamins through their diet. Some parasites have taken this one step further, losing many additional pathways that can be supplemented by their host (Mira, et al., 2001). Thus, it seems unusual that SCN, an animal and a parasite, would express genes of the vitamin B biosynthesis pathways. This is all the more perplexing because SCN should have no problem absorbing the vitamins from host plants. Using BLAST, we found that not only does SCN have the normal vitamin B transporters and salvage pathways found in animals, it has a few extra genes previously not identified in multicellular animals. The bacteria-like genes *thiM*, *tenA*, and permease (transporter of cytosine, purine, uracil, thiamin, and allantoin family) are respectively involved in HET salvage, HMP salvage, and thiamin transport (Jenkins, et al., 2008; Karunakaran, et al., 2006; Ren, et al., 2007; Zhang, et al., 1997).

Looking at the biosynthesis pathways, it becomes apparent that SCN only carries part of the biosynthesis pathways for B_5 and B_7 (supplemental data). It also becomes immediately apparent that the gene present is the last key enzyme in the pathway. BIOB (biotin synthase) performs the last reaction to create biotin and PANC (pantoate β-alanine ligase) joins β-alanine and pantoate together to create pantothenate (Leonardi, et al., 2005; Zempleni, et al., 2009). On the other hand, the thiamine biosynthesis pathway is more complete, with all of the enzymes for *de novo* biosynthesis except for the initial enzyme in the pathway: THIC (plants and bacteria) or THI5 (fungi) (supplemental data). Yet, their absence could be functionally replaced by the nematode TENA enzyme. The exact role of TENA is still unclear; in *Bacillus subtilis* it has been shown to have both thiaminase II activity and to salvage base-degraded thiamin (Jenkins, et al., 2008; Toms, et al., 2005). This contrasts with its ability to functionally complement an *E. coli thiC* mutant (Morett, et al., 2003). This functional diversity leaves open the possibility that the SCN *tenA* homologue could complete the thiamin biosynthesis pathway by replacing *thiC*.

Whether the pathways are complete or partial, the presence of these biosynthesis genes is at odds with what is normally observed in multicellular animals. As SCN seems fully equipped for vitamin transport and salvage, it raises the question as to what role these incomplete *de novo* biosynthesis pathways are playing.

Role of Incomplete Pathways: The presence of vitamin B biosynthesis pathways in SCN indicates that the nematode may not be getting enough of the vitamins from the plant. However, having the last key enzyme of each biosynthesis pathway would allow the nematode to acquire precursors from hosts and convert them into the vitamin, thus providing greater access to host nutrient resources while avoiding the high cost of complete *de novo* biosynthesis.

A variety of bacteria and fungi seem to employ this same strategy of recruiting former *de novo* biosynthesis genes to expand their salvage pathways. The parasites *Plasmodium falciparum, Toxoplasma gondii, Cryptosporidium parvum,* and *Perkinsus marinus* also contain similarly expanded B_1, B_5, and B_7 salvage pathways (Muller and Kappes, 2007).

The advantage of having these biosynthesis pathways for survival within the plant root can also be clearly seen in certain rhizobia-plant symbiotic relationships. With a limited supply of essential vitamins provided by the plant, rhizobia bacteria auxotrophic for thiamin or biotin are quickly outcompeted in nodulation by prototrophs or bacteria with expanded salvage pathways (Karunakaran, et al., 2006; Streit, et al., 1996). The importance of these biosynthetic genes is further underscored by their inclusion on symbiosis plasmids. These plasmids are frequently swapped between species of bacteria and confer the ability to form nodules for nitrogen fixation in plants (Rodionov, et al., 2002).

Previously published Affymetrix microarrays have analyzed the gene expression of SCN at various time points during infection of the plant (Elling, et al., 2009; Ithal, et al., 2007; Klink, et al., 2009). Five of the genes are included on the microarray chip: the *bioB, panC, tenA,* and *thiE.* These genes generally increased in expression from egg to adulthood, whereas *thi4* generally decreased (Elling, et al., 2009). However, a more stable expression pattern throughout the life cycle is implied in Ithal et al. (2007) and Klink et al. (2009), in which the genes are not listed among those whose expression changed over 1.5 or 2 fold respectively. A more in-depth analysis of the expression patterns of all the vitamin B biosynthesis genes should help elucidate how these pathways are regulated.

The limited supply of vitamins available in the plant, suggested by the presence of these pathways, is unexpected given the nature of the nematode's feeding cell. This could be a consequence of the cell reprogramming that occurs during syncytia formation. Another hypothesis is that the plant is restricting access or biosynthesis of certain essential vitamins to defend against pathogen invasion. Nutrient limitation is a well documented defense that deprives pathogen invaders of resources essential for replication or survival (Expert, 1999; Exton, 1997; Weinberg, 1984, 1996). Biotin chelating using avidin has been shown to occur in the chicken egg and is believed to have antimicrobial properties (Elo, et al., 1980; Green, et al., 1975). This type of broad defense is more difficult to circumvent than gene-for-gene interactions. However, if the vitamin precursors are still available, SCN could circumvent the defense by reacquiring the *de novo* vitamin B biosynthesis genes.

In summary, we have discovered partial pathways for vitamins B_1, B_5, and B_7 biosynthesis in *Heterodera glycines.* This expands SCN's vitamin B repertoire beyond vitamin B_6 and alters our view of the nutritional requirements of this pathogen. Complementation studies will help determine if the proteins are functional and whether the SCN TENA enzyme can complete the thiamin biosynthesis pathway. Further study of these genes will provide clues about the nutrient relationship between SCN and its host and potentially lead to the development of novel methods to protect plants from these parasitic worms.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. 1999. Thiamin biosynthesis in prokaryotes. Archives of Microbiology 171:293–300.

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

Bellafiore, S., Shen, Z., Rosso, M. N., Abad, P., Shih, P., and Briggs, S. P. 2008. Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. PLoS Pathogens 4:e1000192.

Bird, D. M., Opperman, C. H., and Davies, K. G. 2003. Interactions between bacteria and plant-parasitic nematodes: Now and then. International Journal for Parasitology 33:1269–1276.

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 B_6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. Molecular Biology and Evolution 25:2085–2098.

Curtis, R. H. 2007. Plant parasitic nematode proteins and the host parasite interaction. Briefings in Functional Genomics and Proteomics 6:50–58.

Davis, R. E. 1996. Spliced leader RNA trans-splicing in metazoa. Parasitology Today 12:33–40.

Doyle, E. A., and Lambert, K. N. 2002. Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. Molecular Plant-Microbe Interactions 15:549–556.

Elling, A. A., Davis, E. L., Hussey, R. S., and Baum, T. J. 2007. Active uptake of cyst nematode parasitism proteins into the plant cell nucleus. International Journal for Parasitology 37:1269–1279.

Elling, A. A., Mitreva, M., Gai, X., Martin, J., Recknor, J., Davis, E. L., Hussey, R. S., Nettleton, D., McCarter, J. P., and Baum, T. J. 2009. Sequence mining and transcript profiling to explore cyst nematode parasitism. BMC Genomics 10:58.

Elo, H. A., Raisanen, S., and Tuohimaa, P. J. 1980. Induction of an antimicrobial biotin-binding egg white protein (avidin) in chick tissues in septic *Escherichia coli* infection. Experientia 36:312–313.

Endo, B. Y. 1998. *Atlas on ultrastructure of infective juveniles of the soybean cyst nematode, Heterodera glycines*: U.S Department of Agriculture.

Expert, D. 1999. Withholding and exchanging iron: Interactions between *Erwinia* spp. and their plant hosts. Annual Review of Phytopathology 37:307–334.

Exton, M. S. 1997. Infection-induced anorexia: Active host defence strategy. Appetite 29:369–383.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. 2006. Pfam: Clans, web tools and services. Nucleic Acids Research 34:D247–251.

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

Green, N. M., Anfinsen, C. B., Edsall, T. J., and Frederic, M. R. 1975. Avidin. Advances in Protein Chemistry: Academic Press, pp. 85–133.

Guindon, S., and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696–704.

Ithal, N., Recknor, J., Nettleton, D., Hearne, L., Maier, T., Baum, T. J., and Mitchum, M. G. 2007. Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. Molecular Plant-Microbe Interactions 20:293–305.

Jenkins, A. L., Zhang, Y., Ealick, S. E., and Begley, T. P. 2008. Mutagenesis studies on TenA: A thiamin salvage enzyme from *Bacillus subtilis*. Bioorganic Chemistry 36:29–32.

Jones, M. G. K. 1981. Host cell responses to endoparasitic nematode attack: Structure and function of giant cells and syncytia. Annals of Applied Biology 97:353–372.

Karunakaran, R., Ebert, K., Harvey, S., Leonard, M. E., Ramachandran, V., and Poole, P. S. 2006. Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum bv. viciae strain* 3841. Journal of Bacteriology 188:6661–6668.

Keeling, P. J., and Palmer, J. D. 2008. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9:605–618.

Klink, V. P., and Matthews, B. F. 2009. Emerging Approaches to Broaden Resistance of Soybean to Soybean Cyst Nematode as Supported by Gene Expression Studies. Plant Physiology 151:1017–1022.

Lambert, K. N., Allen, K. D., and Sussex, I. M. 1999. Cloning and characterization of an esophageal gland-specific chorismate mutase from the phytoparasitic nematode, *Meloidogyne javanica*. Molecular Plant-Microbe Interactions 4:328–336.

Leonardi, R., Zhang, Y.-M., Rock, C. O., and Jackowski, S. 2005. Coenzyme A: Back in action. Progress in Lipid Research 44:125–153.

Liu, C., and Lu, N. 2008. Biotin requirement and its biosynthesis blockage in the free-living nematode, *Caenorhabditis elegans*. Federation of American Societies for Experimental Biology 22:1102–1104.

Livak, K. J., and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402–408.

Lonsdale, D. 2006. A review of the biochemistry, metabolism and clinical benefits of thiamin(e) and its derivatives. Evidence-based Complementary and Alternative Medicine 3:49–59.

Meutter, J.d, Vanholme, B., Bauw, G., Tytgat, T., Gheysen, R., and Gheysen, G. 2001. Preparation and sequencing of secreted proteins from the pharyngeal glands of the plant parasitic nematode *Heterodera schachtii*. Molecular Plant Pathology 2:297–301.

Mira, A., Ochman, H., and Moran, N. A. 2001. Deletional bias and the evolution of bacterial genomes. Trends in Genetics 17:589–596.

Molinari, S., and Miacola, C. 1997. Antioxidant enzymes in phytoparasitic nematodes. Journal of Nematology 29:153–159.

Morett, E., Korbel, J. O., Rajan, E., Saab-Rincon, G., Olvera, L., Olvera, M., Schmidt, S., Snel, B., and Bork, P. 2003. Systematic discovery of analogous enzymes in thiamin biosynthesis. Nature Biotechnology 21:790–795.

Muller, S., and Kappes, B. 2007. Vitamin and cofactor biosynthesis pathways in *Plasmodium* and other apicomplexan parasites. Trends Parasitology 23:112–121.

Niblack, T. L., Heinz, R. D., Smith, G. S., and Donald, P. A. 1993. Distribution, density, and diversity of *Heterodera glycines* in Missouri. Journal of Nematology 25:880–886.

Nicholas, W. L., Hansen, E., and Dougherty, E. C. 1962. The B-vitamins required by *Caenorhabditis briggsae* (*Rhabditidae*). Nematologica 8:129–135.

Painter, J. E., and Lambert, K. N. 2003. *Meloidogyne javonica* chorismate mutase transcript expression porfile using real-time quantitative RT-PCR. Journal of Nematology 35:82–87.

Park, J.-H., Dorrestein, P. C., Zhai, H., Kinsland, C., McLafferty, F. W., and Begley, T. P. 2003. Biosynthesis of the thiazole moiety of thiamin pyrophosphate (vitamin B_1). Biochemistry 42:12430–12438.

Qin, L., Kudla, U., Roze, E. H., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J. T., Schots, A., Smant, G., Bakker, J., and Helder, J. 2004. Plant degradation: A nematode expansin acting on plants. Nature 427:30.

Ren, Q., Chen, K., and Paulsen, I. T. 2007. TransportDB: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Research 35:D274–279.

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. 2002. Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. Journal of Biological Chemistry 277:48949–48959.

Ronconi, S., Jonczyk, R., and Genschel, U. 2008. A novel isoform of pantothenate synthetase in the Archaea. Federation of the Societies of Biochemistry and Molecular Biology 275:2754–2764.

Scholl, E. H., Thorne, J. L., McCarter, J. P., and Bird, D. M. 2003. Horizontally transferred genes in plant-parasitic nematodes: A high-throughput genomic approach. Genome Biology 4:R39.

Smant, G., Stokkermans, J. P., Yan, Y., de Boer, J. M., Baum, T. J., Wang, X., Hussey, R. S., Gommers, F. J., Henrissat, B., Davis, E. L., Helder, J., Schots, A., and Bakker, J. 1998. Endogenous cellulases in animals: Isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proceedings of the National Academy of Sciences USA 95:4906–4911.

Streit, W. R., Joseph, C. M., and Phillips, D. A. 1996. Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. Molecular Plant-Microbe Interactions 9:330–338.

Sullivan, J. T., and Ronson, C. W. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proceedings of the National Academy of Sciences USA 95:5145–5149.

Szewczyk, N. J., Kozak, E., and Conley, C. A. 2003. Chemically defined medium and *Caenorhabditis elegans*. BMC Biotechnology 3:19.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673–4680.

Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. 2005. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. Biochemistry 44:2319–2329.

Tucker, M. L., Xue, P., Raina, A., Ehrenfried, M. L., Asif, M., and Thai, V. K. 2005. Characterization of several *Heterodera glycines* mRNA that encode small proteins with putative signal peptides. Journal of Nematology 37:422–428.

Tytgat, T., Vanholme, B., De Meutter, J., Claeys, M., Couvreur, M., Vanhoutte, I., Gheysen, G., Van Criekinge, W., Borgonie, G., Coomans, A., and Gheysen, G. 2004. A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. Molecular Plant-Microbe Interactions 17:846–852.

Vanfleteren, J. R. 1973. Amino acid requirements of the free-living nematode *Caenorhabditis briggsae*. Nematologica 19:93–99.

Vanholme, B., De Meutter, J., Tytgat, T., Van Montagu, M., Coomans, A., and Gheysen, G. 2004. Secretions of plant-parasitic nematodes: A molecular update. Gene 332:13–27.

Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., de Boer, J. M., Baum, T. J., Hussey, R. S., and Davis, E. L. 2001. Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. Molecular Plant-Microbe Interactions 14:536–544.

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

Weinberg, E. D. 1984. Iron withholding: A defense against infection and neoplasia. Physiological Reviews 64:65–102.

Weinberg, E. D. 1996. Iron withholding: A defense against viral infections. Biometals 9:393–399.

Zhang, Y., Taylor, S. V., Chiu, H. J., and Begley, T. P. 1997. Characterization of the *Bacillus subtilis* thiC operon involved in thiamine biosynthesis. Journal of Bacteriology 179:3030–3035.