Analysis of a Horizontally Transferred Pathway Involved in Vitamin B₆ Biosynthesis from the Soybean Cyst Nematode *Heterodera glycines*

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

Department of Crop Sciences, University of Illinois, Urbana

*Heterodera glycines* is an obligate plant parasite capable of biochemically and developmentally altering its host’s cells in order to create a specialized feeding cell. Although the exact mechanism of feeding cell morphogenesis remains a mystery, the nematode’s ability to manipulate the plant is thought to be due in part to horizontal gene transfers (HGTs). A bioinformatic screen of the nematode genome has revealed homologues of the genes *SNZ* and *SNO*, which comprise a metabolic pathway for the de novo biosynthesis of pyridoxal 5’-phosphate, the active form of vitamin B₆ (VB₆). Analysis of the 2 genes, *H₂SNZ* and *H₂SNO*, show that they contain nematode-like introns, generate polyadenylated mRNAs, and map to the soybean cyst nematode genetic linkage map, indicating that they are part of the nematode genome. However, gene synteny, protein homology, and phylogenetic evidence suggest prokaryotic origin. This would represent the first case of the HGT of a complete pathway into a nematode or terrestrial animal. VB₆ acts as a cofactor in over 140 different enzymes, and recent studies point toward an important role as a potent quencher of reactive oxygen species. With *H. glycines*’ penchant for acquiring parasitism genes through HGT along with the absence of this pathway in other land-based animals suggests a specific need for VB₆ which may involve the parasite–host interaction.

Introduction

*Heterodera glycines*, known as the soybean cyst nematode (SCN), is a plant parasite that has the unique ability to seize control over plant root cells by injecting them with a complex cocktail of proteins and metabolites (Niblack et al. 2006). What ensues is a dramatic and highly orchestrated transformation of the plant root cell into a structure called the syncytium. As this syncytium grows, it merges with the surrounding root cells and becomes dense and metabolically active (Endo 1998). This highly specialized cell functions as a nutrient sink from which the nematode can feed for the rest of its 30-day life cycle. The success of SCN has made it the primary pest of soybean plants.

Unfortunately, the mechanism by which this destructive nematode creates or sustains the syncytia remains a mystery. The complex interaction between plant and nematode, along with the lack of a defined transcriptome, has yielded an incomplete picture. However, some important inroads into this process have been made: chorismate mutase, cellulases, pectinases, ubiquitin extension proteins, and clavata3 are but a few of the candidate parasitism genes identified so far from plant parasitic nematodes (Smant et al. 1998; Lambert et al. 1999; De Boer et al. 2002; Doyle and Lambert 2002; Olsen and Skriver 2003; Tytgat et al. 2004). Interestingly, many of these parasitism proteins are closely related to bacterial proteins, suggesting that they were horizontally transferred. However, these transfer events have always been difficult to prove. Naturalization of genes and convergent evolution often blurs the line between native and horizontally transferred genes. Still, the idea of horizontal gene transfer (HGT) as an important driving force in the evolution of parasitism in nematodes remains (Bird et al. 2003; Jain et al. 2003). Following this lead, a specifically designed bioinformatic screen recently identified several horizontal gene candidates in the root-knot nematode (Scholl et al. 2003).

Other studies have identified HGT candidates and parasitism genes by focusing upon secreted proteins, genes expressed in the dorsal and subventral esophageal glands (Lambert et al. 1999; Meutter et al. 2001; Wang et al. 2001). Although this method has proven fruitful, it can overlook important mRNAs that are not expressed in the glands. It can also overlook metabolites—a class of molecules that likely has an important effect upon the host plant. Currently, nod factor NodL is the only known metabolite implicated in nematode parasitism of plants (McCartter et al. 2003; Scholl et al. 2003; Weerasinghe et al. 2005).

To facilitate the study of this organism, the *H. glycines* genome was recently shotgun sequenced (Bekal et al. 2008). Looking beyond the well-studied gland genes, we searched for novel genes using a genome-wide bioinformatics approach. We performed a screen with Hidden Markov Models (HMMs) from plants not previously found in nematodes. This screen identified homologues of the genes *SNZ* (snooze) and *SNO* (snz-proximal open reading frame [ORF]) (Padilla et al. 1998). In this paper, we show that these genes are of nematode origin and encode a functional pyridoxal 5’-phosphate (PLP) synthase that produces the vital metabolite vitamin B₆ (VB₆). We show that although the genes appear to be typical nematode genes at the DNA level, evidence suggests that they were acquired through a horizontal gene transfer event. If true, this is the first example of an HGT of an entire metabolic pathway into a nematode or any terrestrial animal.

Materials and Methods

HMM Queries

The SCN genome, biotype TN20 and TN10, was randomly fragmented and sequenced with a highly parallel microbead pyrosequencing system developed by 454 Life Sciences (Margulies et al. 2005). The sequences were converted into their 6 ORFs using a BioPerl-based perl script. ORFs greater than 20 amino acids in length were stored in

Key words: *Heterodera glycines*, horizontal gene transfer, soybean cyst nematode, pyridoxal 5’-phosphate, vitamin B₆, antioxidant.

E-mail: Knlamber@uiuc.edu.

doi:10.1093/molbev/msn141

Advance Access publication June 26, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.

For permissions, please e-mail: journals.permissions@oxfordjournals.org
a database of potential nematode proteins. The database was then queried using HMMER2.3.2 against a list of 1,013 plant HMMs not found in Caenorhabditis elegans. The HMMs were downloaded from the Sanger Institute Pfam Web site (http://pfam.sanger.ac.uk/search?tab = searchSequenceBlock). The results were parsed and the sequences screened against a plant and bacterial nucleotide database created from the National Center for Biotechnology Information (NCBI). Those with high nucleotide identity to plants or bacteria were removed from further analysis.

Nematode Culture

Inbred nematode lines of biotype TN10 and TN20 were cultured on soybean cultivar UC82 plants, and mature female cysts were harvested as described previously (Niblack et al. 1993). Eggs were released and collected by gently crushing female cysts over a 0.25-mm sieve. As the eggs hatched, the juvenile nematodes (J2s) were crawled through several layers of Kimwipes suspended over a water-filled petri dish (Ho et al. 1992). The J2s were then filtered and collected onto a 10 micron polycarbonate membrane (GE Osmonics Inc., Minnetonka, MN).

To obtain surface-sterilized nematodes, the collected J2s went through a second round of cleaning. They were allowed to crawl through a sterile sand column equilibrated with 0.1% sodium dodecyl sulfate and suspended over a water reservoir (Lambert 1987; Painter and Lambert 2003). Worms that crawled through the sand column into the reservoir were considered largely free of plant and fungal debris and other microbes associated with the cuticle. When over 200 surface-sterilized worms were placed on Murashige and Skoog (1962) media (Phytotechnology Laboratories, Shawnee Mission, KS), only a few colony-forming units were seen.

Single Nematode Real Time Quantitative-Polymerase Chain Reaction Assay

Primers: (all oligos ordered from Invitrogen, Carlsbad, CA)

SNZsyr-F: 5’-ATCACCAGCGCCACAACT-3’  
SNZsyr-R: 5’-TGCCCTGGTCGCTTACATC-3’  
SNOSyr-F: 5’-AGGCAACGTGCGACCAAT-3’  
SNOSyr-R: 5’-CTGATCGCCAGTCTTCACTATGA-3’  
HgCM1-F: 5’-CCAAAGGCCTGCTTCAATTAC-3’  
HgCM1-R: 5’-CCCTGCCGCGAAACAT-3’  
EF1a-F: 5’-CTTTGGGGTGCTAACAGCCTC-3’  
EF1a-R: 5’-GGCGGTCTAGTGTTGCTTACATC-3’

Surface-sterilized worms were obtained as described above. Due to their resilient cuticle, nematodes were individually digested in 10 μl of buffer (50 mM Tris–HCL pH7.5, 50 mM NaCl) containing 4 mg/ml of fungal protease K (Invitrogen). After 24 h, the protease K was inactivated at 80 °C for 30 min. Quantitative-polymerase chain reaction (Q-PCR) primers were designed on ABI’s Primer Express 2.0, and the efficiency of each primer pair was verified on a dilution series of genomic SCN DNA. The Sybr Green real time Q-PCR assay was performed on an ABI series 7900HT using 10 pmoles of primers and the 10 μl of digested worm. The following conditions were used: 50 °C (2 min); 94 °C (5 min); 40 cycles 94 °C (30 s), 60 °C (1 min), dissociation stage. Five replicates were performed for each gene tested.

Cloning of SCN gDNA

Primers:

SNZ–SNO–F: 5’-ATTGCGCTTCTGCTCCGCTACT-3’  
SNZ–SNO–R: 5’-GCTGCTCCGCGCTTACATC-3’

Sequences that matched the SNZ and SNO HMMs were aligned and assembled using the program Sequencher 4.6. In order to obtain further sequence data, we used the GenomeWalker kit from Clontech (Mountain View, CA). To obtain DNA for this experiment, surface-sterilized nematodes were spun down and pelleted in sterile 1% carboxymethylcellulose. The worms were freeze fratured with a tissue pulverizer and DNA extracted as described in Lambert et al. (1999). The DNA was used to create a genomic library with the enzymes DraI, PvuII, StuI, and EcoRV as instructed by the GenomeWalker kit. A second library was made using the enzymes XmnI, FspI, SwaI, and HincII (New England BioLabs, Ipswich, MA).

Each library was PCR amplified using the kit primers and gene-specific primers according to the manufacturer’s instructions. The amplicons were sequenced at the University of Illinois W.M. Keck Biotechnology Center. Three rounds of genome walking resulted in a 5-kb genomic contig encompassing both HgSNZ and HgSNO. The entire contig was amplified with the Elongase Amplification System (Invitrogen) using the primers SNZ–SNO–F/R and resequenced.

Cloning of HgSNZ and HgSNO cDNA

Primers:

3’ RacePolyT: 5’-ACCTGGACTCGGCGTACCTTTTTTTTTTTTTTTTTT-3’  
3’ RaceComp: 5’-GGCGCGATGAGATGTACATC-3’

Sequences that matched the SNZ and SNO HMMs were aligned and assembled using the program Sequencher 4.6. In order to obtain further sequence data, we used the GenomeWalker kit from Clontech (Mountain View, CA). To obtain DNA for this experiment, surface-sterilized nematodes were spun down and pelleted in sterile 1% carboxymethylcellulose. The worms were freeze fratured with a tissue pulverizer and DNA extracted as described in Lambert et al. (1999). The DNA was used to create a genomic library with the enzymes DraI, PvuII, StuI, and EcoRV as instructed by the GenomeWalker kit. A second library was made using the enzymes XmnI, FspI, SwaI, and HincII (New England BioLabs, Ipswich, MA).

Each library was PCR amplified using the kit primers and gene-specific primers according to the manufacturer’s instructions. The amplicons were sequenced at the University of Illinois W.M. Keck Biotechnology Center. Three rounds of genome walking resulted in a 5-kb genomic contig encompassing both HgSNZ and HgSNO. The entire contig was amplified with the Elongase Amplification System (Invitrogen) using the primers SNZ–SNO–F/R and resequenced.

Cloning of SCN gDNA

Primers:

SNZ–SNO–F: 5’-ATTGCGCTTCTGCTCCGCTACT-3’  
SNZ–SNO–R: 5’-GCTGCTCCGCGCTTACATC-3’

Sequences that matched the SNZ and SNO HMMs were aligned and assembled using the program Sequencher 4.6. In order to obtain further sequence data, we used the GenomeWalker kit from Clontech (Mountain View, CA). To obtain DNA for this experiment, surface-sterilized nematodes were spun down and pelleted in sterile 1% carboxymethylcellulose. The worms were freeze fratured with a tissue pulverizer and DNA extracted as described in Lambert et al. (1999). The DNA was used to create a genomic library with the enzymes DraI, PvuII, StuI, and EcoRV as instructed by the GenomeWalker kit. A second library was made using the enzymes XmnI, FspI, SwaI, and HincII (New England BioLabs, Ipswich, MA).

Each library was PCR amplified using the kit primers and gene-specific primers according to the manufacturer’s instructions. The amplicons were sequenced at the University of Illinois W.M. Keck Biotechnology Center. Three rounds of genome walking resulted in a 5-kb genomic contig encompassing both HgSNZ and HgSNO. The entire contig was amplified with the Elongase Amplification System (Invitrogen) using the primers SNZ–SNO–F/R and resequenced.
Genetic Mapping—SNP Ligation Assay

Primers: (universal primer tags underlined)

\[
\begin{align*}
HgSNOSnpA & : 5' - CGCCTCGCCGGCTTTTGGAAACG- GTGCCGGGAAATGTA-3' \\
HgSNOSnpG & : 5' - CGCCTCGCCGGCTTTTGGAAACG- GTGCCGGGAAATGTA-3' \\
HgSNOR & : 5' - P04-GAAAACTGAGAATGTAACGC- ATTACCATAGGCCACGAGCC-3' \\
\text{UniversalPrimerF} & : 5' - CGCCTCGCCGGCTTTTGG-3' \\
\text{UniversalPrimerR} & : 5' - GGTCTCTCGCCGCTCTATGG-3'
\end{align*}
\]

An A/G SNP within \( HgSNOr \) was identified by aligning sequences from the nematode biotypes TN16 and TN20. Genotyping probes were created surrounding the SNP using the program NetPrimer (Premier BioSoft, Palo Alto, CA). Using single nucleotide polymorphism (SNP) as a molecular marker, the mapping population was genotyped using a variation of the platform described in Macdonald et al. (2005). For each DNA sample from the mapping population (Attibalenti et al. 2005), a pair of 5 pmol ligation reactions was set up, one using the genotyping oligos \( HgSNOSnpA/HgSNOSnpR \) and the other using \( HgSNOSnpG/HgSNOSnpR \). The following reaction concentrations were used: 0.02 pmol of each genotyping oligo, 5 ng of template gDNA, 1.6 units of Amphiligase thermostable DNA ligase (Epicentre Biotechnologies, Madison, WI), and 1× ligase buffer (Epicentre Biotechnologies). The ligation reactions were performed using the following cycle: initial denaturation for 5 min at 95 \( ^\circ \text{C} \), 4 cycles of 30 s at 95 \( ^\circ \text{C} \) and 25 min at 65 \( ^\circ \text{C} \), and 4 \( ^\circ \text{C} \) storage. Ligation products were analyzed using Q-PCR. The following primers were cloned separately into the vector pCDF Duet-1 and transformed into the kanamycin-resistant strain obtained from the Coli Genetic Stock Center at Yale (Keio Collection: JW1630-1). Selection was performed on kanamycin (30 mg/ml) and spectinomycin (50 mg/ml) LB plates.

This process was repeated with \( HgSNOr \), only the cDNA was amplified with the primers SNOstart-BglII/SNOend-EcoRV. The cDNA was excised from pCR8-TOPO with EcoRV and BglII and ligated into the second expression position of pCDF Duet-1-SNZ. The created strains were then transformed with the ampicillin-resistant vector TargeTron pAR1219 (Sigma, St Louis, MO). This vector contains the T7 RNA polymerase gene and allows isopropyl-beta-D-thiogalactopyranoside (IPTG) induction of the Lac-T7 promoters on pCDF-Duet-1.

For the complementation experiments, the strains were plated onto minimal media M9 plates with kanamycin 30 mg/ml, spectinomycin 50 mg/ml, and ampicillin 50 mg/ml. The cultures were grown at 37 \( ^\circ \text{C} \). Due to leaky expression of \( HgSNZ \) and \( HgSNOr \), IPTG was not required for induction. For the control plate, pyridoxal was added to a concentration of 1 \( \mu \text{M} \). For the growth curves, the strains were grown in liquid M9 media with the required antibiotics and monitored at 600 nm on a spectrophotometer.

Protein Purification

\( HgSNZ \) and \( HgSNOr \) protein were His-tagged and purified on TALON Resin columns (Clontech). The 2 genes were cloned separately into the vector pCDF Duet-1 and transformed into the \( E. coli \) strain Nova Blue DE3 (Novagen). The cells were grown as instructed by Novagen in LB with 50 \( \mu \text{g/ml} \) spectinomycin and 12.5 \( \mu \text{g/ml} \) tetracycline. Once an OD of 0.6 at 600 nm was obtained, the cells were induced with IPTG (1 mM) for 3 h. In all, 50-ml aliquots were pelleted and incubated for 45 min in 4 ml of lysis buffer (50 mM Tris pH 8.0, 10% sucrose, 140 mM NaCl) and 5 mg/ml lysozyme. The cells were freeze-thawed 3 times in liquid nitrogen and centrifuged for 30 min. The supernatant was applied to the TALON resin and the protein extracted according to the protocol for nondenaturing conditions. The purity and size of the protein were visually verified on a Tris–HCL 4–20% Ready Gels (Biorad, Hercules, CA). The concentration of the purified proteins was determined with BioRad Protein Assay Dye using a serial dilution of bovine serum albumin as a standard curve.

Enzyme Assays

\( HgSNOr \) Enzyme Assay

The glutaminase assay was carried out via a coupled reaction with glutamate dehydrogenase (GDH) (Sigma). In this setup, the SNO hydrolyzes glutamine to glutamate, which in turn is converted to alpha-ketoglutarate by GDH along with the reduction of 3-acylpyridine adenine dinucleotide (APAD) (Sigma), an analog of nicotinamide adenine dinucleotide (NAD\(^+\)). The reduction of APAD to APADH is detected by a spectrophotometer at 363 nm and can be used as a measure of glutaminase activity by SNO (Raschle et al. 2005; Gengenbacher et al. 2006).
The reaction was set up in a 300 µl reaction as follows: 5 uM of HgSNO, 5 uM HgSNZ, 10 mM glutamine, 7 units GDH, 10 mM Tris–HCl (pH 8.0), and 0.5 mM APAD were combined. The reaction was incubated together for 10 min before the substrate, glutamine, was added.

**HgSNZ Enzyme Assay**

This assay was performed as described in Raschle et al. (2005) with slight modifications. Within a 300 µl reaction, 10 uM HgSNZ, 10 uM HgSNO, 20 uM glutamine (Invitrogen), 2 mM ribose 5'-phosphate (Sigma), and 4 mM glyceraldehydes 3'-phosphate (Sigma) were combined. As before, HgSNZ and HgSNO were incubated together in buffer for 10 min before adding the substrates. During the reaction presumably, a Schiff’s base is formed between the Tris buffer and the product. This can be monitored by measuring its absorbance at 414 nm in a spectrophotometer and is used as a measure of PLP synthase activity.

**Phylogenetic Analysis**

Blasts to the NCBI protein and nucleotide databases were performed to obtain homologues to HgSNZ and HgSNO (Altschul et al. 1990). Nucleotide sequences were initially assembled in Sequencher and imported into the phylogenetic program MEGA-4 (Tamura et al. 2007). MEGA-4 translated the sequences into protein and aligned them according to their amino acids. Bayesian analysis was performed upon the protein alignments using the program MrBayes (Ronquist and Huelsenbeck 2003). The analysis was performed with 4 parallel chains for 1 million generations or until convergence below 0.05. During the analysis, trees were sampled every 100th generation. From these, a consensus tree was created with a burn-in of 25%; branches whose posterior probabilities dropped below 50% were collapsed into a polytomy. Ten representative sequences were chosen from each kingdom based upon the following criteria: complete protein known, a SNO or SNO homologue was PCR amplified, cloned, and sequenced (GenBank accession number EU380677). The cDNAs for HgSNZ and HgSNO were PCR amplified, cloned, and sequenced as well (GenBank accession numbers EU747297 and EU747298, respectively). In accordance with the current nomenclature for genes identified from parasitic nematodes, the SNZ/SNO homologues from *H. glycines* were named *HgSNZ* and *HgSNO* (Bird and Riddle 1994).

**Nematode Origin of the Gene Pair**

Due to the highly conserved nature of SNZ and SNO proteins across all the kingdoms, it was important to verify their nematode origin. By aligning the cDNA to genomic DNA, 5 introns were revealed in *HgSNZ* and 1 intron in *HgSNO* (fig. 1). In addition, a few of the *HgSNZ* cDNA clones were alternatively spliced with an extended intron or an extra intron near the 3' end. The splice junctions themselves were most similar to eukaryotes, specifically, the splice consensus of *C. elegans* (5'-AGIGURAGUUU and 3' -UUUUCAGIG) (Blumenthal and Stewart 1997). The 3' Race also showed that both mRNAs had a poly-A tail with a putative poly-A signal (AAUUAAG) located 12 bases upstream. Although unconventional, the polyadenylation signal has been shown to be functional in eukaryotes (Unterman et al. 1981). The presence of introns and a poly-A tail removes the possibility of contamination from internal or external prokaryotes.

To test if these genes were derived from other surface contaminants, Q-PCR was performed upon individual, surface-sterilized nematodes. Both HgSNZ and HgSNO were detected in all individuals tested. The level of amplification was similar to positive Q-PCR controls for the known SCN genes chorismate mutase and elongation factor 1a. Water controls did not amplify (table 1). This suggests that *HgSNZ* and *HgSNO* are not from surface contaminants because amplification from such a source would vary from nematode to nematode.

To verify *HgSNZ* and *HgSNO* as part of the SCN genome, an SNP within *HgSNO* was mapped to an existing SCN amplified fragment length polymorphism (AFLP)-based genetic map. The map was constructed from a population of F2 progeny using the inbred parental lines TN16 and TN20 (Atibalentja et al. 2005). Analysis of the
Genotyped population with the program JoinMap placed the HgSNO-SNP within linkage group 2 (fig. 2) (Stam 1993). These data verify HgSNO as part of the SCN genome.

Phylogenetic Analysis

On the amino acid level, SNZ and SNO represent 2 of the most highly conserved protein families and have been identified in all the kingdoms. Homologues of SNO have been published under the names YaaE, PdxT, PDX2, and SNZB. SNZ homologues are also known as YaaD, PdxS, PDX1, SOR1, PYROA, and HEVER. To understand the potential origin of HgSNO and HgSNZ, homologues of SNZ and SNO were downloaded from NCBI's protein and nucleic acid databases for each of the 6 kingdoms: protists, viridiplantae, fungi, bacteria, archaea, and a few aquatic animals. Due to the large number of homologues identified, Bayesian analysis was first performed upon each kingdom separately (Supplementary Material online). From each tree, the most closely related sequence to H. glycines along with 9 other homologues from representative clades were pooled together. These representatives of horizontal gene transfer in SCN...

FIG. 1.—Genomic DNA encompassing HgSNZ and HgSNO (GenBank accession number EU380677). An overview of the genomic fragment is shown on top with exons represented as block arrows. The 2 genes are separated by 442 bp. The flanking gDNA and introns are lower case, whereas the exons are capitalized. Minor alternatively spliced regions within HgSNO are capitalized and italicized. Predicted splice recognition sites are in bold. Start and stop codons are shown as bold and underlined. The stop codons are followed by an underlined putative poly-A tail signal.
SNZ and SNO were then used to create a more broadly based tree encompassing all the kingdoms (fig. 3).

After a million generations, the SNZ Bayesian analysis converged below 0.0061. Within the resulting phylogenetic tree, plants, aquatic animals, and fungi form clearly separated clades. The bacteria and archaea together form another clade. HgSNZ is found grouped with the bacteria. The SNO analysis converged below 0.023, with the tree topology providing a similar picture to that of SNZ. Fungi, plants, aquatic animals, and prokaryotes clearly fall within their own strongly supported clades. Here, HgSNO is also grouped with the prokaryotes. In both trees, the protists form a separate clade, however, some plant-like protists separate with the plants and some fungal-like protists separate with the fungi. Trees generated using maximum likelihood resulted in largely congruent tree topologies (SNZ loglk = −18265; SNO loglk = −22425). Most importantly, they support the Bayesian result with HgSNZ and HgSNO grouping with prokaryotes (Supplementary Material online).

Analysis of HgSNZ and HgSNO Proteins: A PLP Synthase

To determine if HgSNO and HgSNZ could have their predicted function, the proteins were analyzed for conservation of functionally critical amino acid residues. Representatatives from plant, fungi, marine animals, and bacteria were compared and aligned together using ClustalW (fig. 4) (Thompson et al. 1994). These alignments show the high level of conservation between the kingdoms. Specifically, the catalytic triad Cys–His–Glu indicative of class I glutaminases was preserved within HgSNO (Zalkin and Smith 1998). HgSNZ also retains amino acids shown to be essential to activity in Bacillus subtilis (Lys81, Lys149, and Asp24) (Strohmeier et al. 2006; Raschle et al. 2007). This analysis predicts that HgSNZ and HgSNO are functional and should synthesize VB6.

Further analysis of the proteins by PSORT II predicts a cytoplasmic localization for both enzymes (Horton and Nakai 1997). No signal peptide or any glycosylation sites were predicted; this correlates with previous studies which have localized the proteins to the cytoplasm, associated with the cytoplasmic plasma membrane, endomembranes, and nucleus (Chen and Xiong 2005; Matsuyama et al. 2006; Denslow et al. 2007).

Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Ct Avg.</th>
<th>Ct SD</th>
<th>Water Control Ct</th>
</tr>
</thead>
<tbody>
<tr>
<td>HgSNZ</td>
<td>23.83</td>
<td>0.36</td>
<td>&gt;45</td>
</tr>
<tr>
<td>HgSNO</td>
<td>25.45</td>
<td>1.1</td>
<td>&gt;34</td>
</tr>
<tr>
<td>HgCM1</td>
<td>24.86</td>
<td>0.48</td>
<td>&gt;38</td>
</tr>
<tr>
<td>HgEF1α</td>
<td>23.77</td>
<td>0.86</td>
<td>40.42</td>
</tr>
</tbody>
</table>

Note.—The data are reported as the name of the gene amplified, followed by the average (avg) cycle threshold (Ct) of 5 replicates per gene, and the standard deviation (SD) from that average.

Functional Analysis: Complementation of a PLP Mutant

HgSNZ and HgSNO cDNAs were cloned together into the expression vector pCDF Duet-1 and transformed into the E. coli strain JW1630-1 (Keio collection). This strain is deficient in the enzyme pyridoxine 5'-phosphate oxidase (pdxH). Without pdxH, the cell cannot convert pyridoxine phosphate or pyridoxamine phosphate into the active form of the vitamin, PLP. Thus, the mutant is a VB6 auxotroph that cannot grow on minimal media unless an exogenous source of pyridoxal is supplied (Lam and Winkler 1992). Transformation of the vector expressing HgSNZ and HgSNO successfully complemented the mutant, restoring growth on M9 minimal media (fig. 5). However, transformation with the vector alone did not result in complementation. This result shows that HgSNO and HgSNZ comprise a functional PLP synthase in vivo.

Discussion

Nematode Origin of SNZ and SNO

In the current literature, it is widely assumed that animals lack a de novo synthesis pathway for VB6 (Tanaka
et al. 2005). The only verified exception is the sponge *Suberites domonculus* (Krasko et al. 1999; Seack et al. 2001). However, it is interesting to note that our Blast search through the NCBI expressed sequence tag (EST) database revealed a few hits from aquatic animals such as the sea squirt, planarian, sea anemone, and several mollusks (fig. 3A and B). Also, although the presence of the pyridoxine pathway in nematodes has not been recognized in the published literature, our database searches have shown that previous nematode EST projects have generated incomplete SNO-like cDNA’s. These SNO-like ESTs have been sequenced from the closely related phytoparasitic nematodes *Globodera rostochiensis* (BM343700) and *Meloidogyne hapla* (BQ837630). Sequence and phylogenetic analysis indicates that the *M. hapla* EST may be a fungal contaminant; however, the *G. rostochiensis* protein is closely related to

![Fig. 3.— Unrooted Bayesian phylogenetic trees of SNZ and SNO proteins. Bayesian posterior probability values are located along each branch. Nematode species are highlighted in gray. (A and B) Broad-based phylogenetic trees created from the protein alignments of plants, animals, fungi, protists, bacteria, and archaea. (C and D) Subtrees from the Bayesian phylogenetic analysis of bacterial SNZ and SNO proteins.](https://academic.oup.com/mbe/article-abstract/25/10/2085/1027609)
Fig. 4.—ClustalW alignment of SNZ and SNO proteins from animals, plants, fungi, bacteria, archaea, and protists. Generally conserved sequences are highlighted. Amino acids suspected to be involved in the active site are indicated by an asterisk.

(A) SNO

<table>
<thead>
<tr>
<th>Animal</th>
<th>Plant</th>
<th>Fungi</th>
<th>Bacteria</th>
<th>Archaea</th>
<th>Protist</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. nicotianae</td>
<td>A. thaliana</td>
<td>P. falciparum</td>
<td>B. subtilis</td>
<td>C. intestinalis</td>
<td>S. acidocaldarius</td>
</tr>
<tr>
<td>MTGHSSHLSTLVPGLQLASFSIEKLLRQAAPAL</td>
<td>TADYGTVHMTF111PEPLEG</td>
<td>-EDCA1LPGCSTAEL</td>
<td>-LARBCCGL79</td>
<td>-MLV3GGLQAVREHINLHACG</td>
<td>-AAG</td>
</tr>
<tr>
<td>1056W154</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NQKVQYTHN</td>
<td>-FYTLQIEASATK</td>
<td>-SMTK</td>
<td>-LAMYHN60</td>
<td>-V</td>
<td>-VYR</td>
</tr>
<tr>
<td>-AQP</td>
<td>-VYDTODSG</td>
<td>-SLDG11LPGCSTMR</td>
<td>-TMTK</td>
<td>-LAMYHN60</td>
<td>-V</td>
</tr>
<tr>
<td>-AY</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-G</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-E</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-K</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
</tbody>
</table>

(B) SNZ

<table>
<thead>
<tr>
<th>Animal</th>
<th>Plant</th>
<th>Fungi</th>
<th>Bacteria</th>
<th>Archaea</th>
<th>Protist</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. nicotianae</td>
<td>A. thaliana</td>
<td>P. falciparum</td>
<td>B. subtilis</td>
<td>C. intestinalis</td>
<td>S. acidocaldarius</td>
</tr>
<tr>
<td>HASGTVSPFRSQRGQNAAMAVDPARTPITAASATKNNMTRNIGQSNQSFSSQAPV</td>
<td>4</td>
<td>-LAQLKINGVSAGVYNAXAEGBAAGA</td>
<td>-AAG</td>
<td>-LVYMNBPEGLN</td>
<td>-EYDGIILPSCGSTMR</td>
</tr>
<tr>
<td>1056W154</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NQKVQYTHN</td>
<td>-FYTLQIEASATK</td>
<td>-SMTK</td>
<td>-LAMYHN60</td>
<td>-V</td>
<td>-VYR</td>
</tr>
<tr>
<td>-AQP</td>
<td>-VYDTODSG</td>
<td>-SLDG11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
</tr>
<tr>
<td>-AY</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-G</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-E</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-K</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
<tr>
<td>-E</td>
<td>-LAMH11LPGCSTMR</td>
<td>-M</td>
<td>-SAEAYSG</td>
<td>-T</td>
<td>-VYR</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/mbe/article-abstract/25/10/2085/1027609 by guest on 01 January 2019
that of *H. glycines* (fig. 3B). On the nucleotide level, *G. rostochiensis* and *H. glycines* are 70% identical, much higher than the nearest nonnematode hit with 44% identity. Although SNZ homologues in other nematode species have yet to be found, this may be due to SNZ’s high homology to plants and bacteria, resulting in its accidental removal from databases as a contaminant.

*Heterodera glycines* being one of the rare animals to have the pathway, as well as being the only land-based animal, led us to closely examine the possibility of contamination from outside sources such as prokaryotes, fungi, or plants. Prokaryotic contamination could be quickly ruled out due to the presence of eukaryotic-like introns, a poly-A tail, and a poly-A signal. Although bacteria and archaea can also have introns, they are self-splicing and do not have any similarity to those within *HgSNZ* and *HgSNO* (Dai and Zimmerly 2002). To further remove the possibility of a contaminant origin, Q-PCR was conducted upon individual surface-sterilized nematodes. The Q-PCR results verified the presence of *HgSNO* and *HgSNZ* within each worm and showed amplification levels similar to known nematode genes. The mapping of the genes within linkage group 2 of the *H. glycines* genetic map also clearly supports SCN origin. The unlikelihood of another organism cosegregating along with the mapping population strongly suggests that *HgSNZ* and *HgSNO* are indeed nematode genes. The draft assembly of the SCN genome released by Monsanto and Divergence also confirms the presence of both genes within the nematode (*gi*-170550926-gb-ABLA01019960.1, *gi*-170556672-gb-ABLA01014214.1, *gi*-170543421-gb-ABLA01027465.1, and *gi*-170546648-gb-ABLA01024238.1). Finally, a cDNA sequencing project within our laboratory has revealed the presence of a splice leader at the 5’ end of *HgSNO*. This splice leader is specific to nematodes and has been found in *Caenorhabditis*, *Heterodera*, *Brugia*, *Globodera*, *Ascaris*, and *Meloidogyne* species and proves that these 2 genes are indeed from *H. glycines* (Huang and Hirsh 1989; Evans et al. 1997; Lambert et al. 1999; McCarter et al. 2003).

**Evidence for Horizontal Transfer**

In some respects, *HgSNZ* and *HgSNO* appear to be of SCN origin and have all the hallmarks of typical eukaryotic genes; however, genetic and phylogenetic evidence points toward a horizontal gene transfer event from prokaryotes. If true, this would mark the first known transfer of an entire functional metabolic pathway into a nematode. Although HGT among eukaryotes has often been regarded as a rare event, widespread HGT was recently demonstrated between the endosymbiont *Wolbachia* and its insect/nematode hosts. In one case, the entire *Wolbachia* genome is thought to have been integrated into the *Drosophila*...
an anassae genome (Hotopp et al. 2007). With the Wolbachia genus infecting over 20% of insects and nematodes, HGT in eukaryotes may be much more widespread than previously thought.

It is believed that animals once had the ability to synthesize VB₆. This pathway has since been lost because they can absorb the vitamin from their diet and use a salvage pathway to convert between the different vitamin forms. SCN too has the salvage pathway represented by homologues to PdxK (CB824492, CK351985), PdxH (CD749566), and PdxY (CK351985, BF013933) (Lam and Winkler 1992; Yang et al. 1996; Yang, Tsui, et al. 1998). Therefore, the nematode should be able to obtain VB₆ from its plant host. The loss of de novo synthesis in animals, including the close relative C. elegans, suggests that SCN also lost the ability to synthesize VB₆ and later reacquired it through HGT once it became advantageous to have the pathway again. There are many potential avenues in which the nematode can employ PLP to its advantage.

This hypothesis is supported by phylogenetic trees of HgSNZ and HgSNO, which place them closer to bacteria rather than eukaryotes. Significantly, the protein sequences do not appear with the other animals which form a single clade. The phylogenetic trees created using solely bacterial proteins point toward a relative of the bacterium Candidatus Protochlamydia amoebophila as a possible genetic source of HgSNZ and HgSNO (fig. 3C and D). Protochlamydia amoebophila is an endosymbiont of amoeba, making it tempting to draw parallels to the Wolbachia endosymbiont (Collingro et al. 2005).

Further support of a HGT event can be found in the gene synteny of HgSNZ and HgSNO. Within prokaryotes, SNZ and SNO are closely linked, within a few hundred bases of each other, and are aligned head to tail. In eukaryotes, SNZ and SNO are either closely linked with the genes aligned in opposite directions or they are unlinked (Tanaka et al. 2005) (fig. 7). Heterodera glycines, with HgSNZ and HgSNO aligned head to tail, has the gene synteny of a bacterium. Given the close proximity of SNZ and SNO, it is simple to imagine a single HGT event preserving this bacterial synteny.

Although it is possible that the genes were preserved over the millennia instead of acquired through horizontal transfer, it seems unlikely that they would have retained their bacterial homology or synteny. We suggest a more likely scenario where SCN originally lost the ability to synthesize VB₆ along with the rest of the animals and later obtained it from bacteria once it became advantageous to have again. This unusual need for HgSNZ and HgSNO, that has caused the worm to retain or reacquire them, calls for a thorough understanding of the genes and their products.

VB₆ Biosynthesis and Function

VB₆ refers collectively to the vitamers pyridoxal, pyridoxine, pyridoxamine, and their phosphorylated derivatives. The active form of the vitamin is PLP. This essential vitamin acts as a cofactor in over 140 distinct enzymatic activities that participate in a wide variety of pathways from amino acid metabolism to hormone signaling (Ralevic and Burnstock 1998; Percudani and Peracchi 2003).

Two de novo VB₆ biosynthesis pathways and 1 salvage pathway are currently known. A third pathway known as the salvage pathway is present in all organisms, including animals, and allows newly absorbed or created VB₆ to interchange between the 6 vitamer forms as needed. The first de novo pathway discovered is referred to as the deoxyxylulose 5-phosphate (DXP)-dependent pathway; present only in the y subdivision of proteobacteria, it has been extensively studied in E. coli (Zhao et al. 1995; Hill et al. 1996; Yang, Zhao, et al. 1998; Laber et al. 1999). The mechanism of synthesis in plants, fungi, and other organisms was unclear until the second pathway, referred to as DXP-independent, was discovered in the parasitic fungi Cercospora nicotianae (Ehrenshaft et al. 1999).

The discovery initiated a wave of studies attempting to characterize the chemistry of the new DXP-independent pathway. Within less than a decade, homologues of SNZ and SNO have been identified across all 6 kingdoms, revealing an ancient lineage that is among the most highly conserved protein families. The 2 enzymes were soon demonstrated to form a PLP synthase complex, representing a complete pathway for the biosynthesis of PLP (fig. 8). It accomplishes this through a large array of reactions, including pentose and triose isomerizations, imine formation, amine addition, and ring formation (Kondo et al. 2004; Burns et al. 2005; Raschle et al. 2005, 2007; Strohmeier et al. 2006; Zein et al. 2006). The complexity and number of reactions that PLP synthase can perform may account for its high conservation.

Pyridoxal 5’phosphate has many functions beyond its traditional role as a cofactor. In the same way that it binds and acts as a cofactor, PLP can also act as an inhibitor for many enzymes (Hayashi et al. 1990). PLP has also been known to alter gene expression through direct interactions with transcription factors, hormone receptors, and nuclear receptor-interacting proteins (Nishigori et al. 1978; Tully et al. 1994; Oka 2001; Huq et al. 2007). Other studies have implicated the vitamin in binding and regulating calcium channels in rats (Dakshinamurti et al. 1998). In the plant Hevea brasiliensis, the genes are expressed upon application of salicylic acid and ethylene suggesting a role in defenses (Broekaert et al. 1990). Also of note is the recent
discovery of this pathway in the malaria parasite *Plasmodium falciparum* where possible roles in pathogenesis are currently being explored (Wrenger et al. 2005). PLP is also of importance in the medical field where research is currently testing the properties of VB6 to induce apoptosis in cancer cells (Komatsu et al. 2003; Shimada et al. 2006). Perhaps, some of the more intriguing roles of VB6 are its potential as a strong antioxidant (Bilski et al. 2000; Denslow et al. 2005). This role was first discovered in *C. nicotianae*, in which SOR1 (SNZ) was found to be essential for resistance to cercosporin, a generator of singlet oxygen. Other organisms appear to use VB6 in a similar manner. The malaria parasite *P. falciparum* and *B. subtilis* both upregulate the PLP synthase pathway when confronted with oxidative stress (Antelmann et al. 1997; Wrenger et al. 2005).

Arabidopsis SNO and SNZ mutants, in addition to being severely altered in development, were sensitive to oxidative stress and have increased membrane peroxidation (Tambasco-Studart et al. 2005; Wagner et al. 2006; Denslow et al. 2007).

Potential Roles of HgSNO and HgSNZ

Given the vast number of roles that PLP plays, there are many possible functions that the vitamin might play within the nematode as well as the host plant. Indeed, it is difficult to find a pathway that is not in some way affected by VB6. Of the many different avenues in which it could be employed, the most obvious is to maintain a personal supply of a vital cofactor. The antioxidant properties of VB6 could be used in a similar fashion to vitamins C and E and protect the nematode from reactive oxygen species (ROS), specifically oxidants released by plant defenses (Mehdy et al. 1996; Waetzig et al. 1999). Beyond a solely protective role, the antioxidant may be used to actively suppress host defenses by neutralizing ROS which have been implicated in the signal pathway for plant defenses (de Pinto et al. 2002; Denslow et al. 2005). It has also been observed that yeast cells increase SNZ and SNO transcripts and *C. elegans* ramp up their own ROS defenses when entering the stationary and dauer stage, respectively (Anderson 1982; Larsen 1993; Padilla et al. 1998); thus, PLP could
also be important for the long-term survival of SCN eggs in soil (Inagaki and Tsutsumi 1971).

Conclusion

In summary, 2 genes, HgSNZ and HgSNO, have been cloned from the phytoparasitic nematode H. glycines that form a functional PLP synthase pathway. Further analysis of the splice variants will determine if they have the same functionality. Expression patterns within SCN will also aid in elucidating the role of the metabolite within the worm. However, with SCN’s affinity for acquiring parasitism through HGT, it seems likely that SCN is using PLP beyond its traditional role as a cofactor.

Supplementary Material

Supplementary materials are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank Dr Gustavo Caetano-Anolles for his expert advice on phylogenetic analysis.

Literature Cited


