LETTER

doi:10.1038/nature11651

# A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens

Shiming Liu[1]*, Pramod K. Kandoth[2]*, Samantha D. Warren[3], Greg Yeckel[2], Robert Heinz[2], John Alden[2], Chunling Yang[4], Aziz Jamai[1], Tarik El-Mellouki[1], Parijat S. Juvale[4], John Hill[4], Thomas J. Baum[4], Silvia Cianzio[5], Steven A. Whitham[4], Dmitry Korkin[3], Melissa G. Mitchum[2] & Khalid Meksem[1]

**Soybean (*Glycine max* (L.) Merr.) is an important crop that provides a sustainable source of protein and oil worldwide. Soybean cyst nematode (*Heterodera glycines* Ichinohe) is a microscopic roundworm that feeds on the roots of soybean and is a major constraint to soybean production. This nematode causes more than US$1 billion in yield losses annually in the United States alone[1], making it the most economically important pathogen on soybean. Although planting of resistant cultivars forms the core management strategy for this pathogen, nothing is known about the nature of resistance. Moreover, the increase in virulent populations of this parasite on most known resistance sources necessitates the development of novel approaches for control. Here we report the map-based cloning of a gene at the *Rhg4* (for resistance to *Heterodera glycines* 4) locus, a major quantitative trait locus contributing to resistance to this pathogen. Mutation analysis, gene silencing and transgenic complementation confirm that the gene confers resistance. The gene encodes a serine hydroxymethyltransferase, an enzyme that is ubiquitous in nature and structurally conserved across kingdoms. The enzyme is responsible for interconversion of serine and glycine and is essential for cellular one-carbon metabolism. Alleles of *Rhg4* conferring resistance or susceptibility differ by two genetic polymorphisms that alter a key regulatory property of the enzyme. Our discovery reveals an unprecedented plant resistance mechanism against a pathogen. The mechanistic knowledge of the resistance gene can be readily exploited to improve nematode resistance of soybean, an increasingly important global crop.**

The first quantitative trait loci (QTL) for resistance to *Heterodera glycines* (*rhg*) were identified in the early 1960s[2,3]. Resistance QTL on chromosomes 18 (*rhg1*) and 8 (*Rhg4*) have been consistently mapped in a variety of soybean germplasm and represent the major sources of resistance in soybean cultivars[4]. In the soybean cultivar (cv.) Forrest, resistance to soybean cyst nematode (SCN) requires both *rhg1* and *Rhg4*, with *Rhg4* exhibiting dominant gene action[5]. Roots of plants carrying *Rhg* genes are penetrated by infective juveniles, but feeding cells ultimately degenerate, causing the nematodes to die before reaching adult stages[6]. Genetic variability in *H. glycines* is prevalent, and nematodes that survive on resistant cultivars carry the undefined *ror* (reproduction on a resistant host) alleles[7], leading to population shifts in the field as a consequence of resistant soybean monoculture[8]. So far, our understanding of resistance to SCN remains limited because the genes underlying resistance QTL have not been cloned[9,10].

We report here the positional cloning of the *Rhg4* gene from soybean cv. Forrest (Fig. 1a). For this purpose, three $F_{2:6}$ recombinant inbred line (RIL) populations segregating for SCN resistance were developed. From a total of 355 recombinant lines identified with chromosomal breakpoints at the *Rhg4* locus, two recombinants (ExF74 and FxW5093) were crucial in defining the interval carrying the *Rhg4* gene. Both lines carry the resistance allele at the *rhg1* locus and are double recombinants for an 8-kilobase (kb) interval carrying the *Rhg4* resistance allele. Two genes, one coding for a serine hydroxymethyltransferase (*SHMT*) and the other a subtilisin-like protease (*SUB1*), were identified in the interval. An alignment of *SUB1* cDNA sequences from Forrest, Essex and Williams 82 cultivars indicates that the two amino acid differences between Forrest and Essex do not correlate with SCN resistance (Supplementary Fig. 1). In addition, no nucleotide differences were identified in the 1,766 base pairs (bp) of sequence 5′ of the predicted translational start site for *SUB1* between Forrest and Essex. A comparison between the genomic DNA sequences of *SHMT* from Forrest and Essex identified five nucleotide differences (three single nucleotide polymorphisms (SNPs) and two insertions/deletions (indels)) between the resistant and susceptible alleles (Fig. 1a). Two of the nucleotide differences found between the Forrest and Essex *SHMT* cDNAs result in an amino acid change in the predicted protein sequences (R130P and Y358N) (Fig. 1a). Furthermore, three nucleotide differences were identified in the 2,339 bp of sequence 5′ of the predicted start site for *SHMT* between Forrest and Essex (Supplementary Fig. 2a). On the basis of these findings, *SHMT* was characterized further for a role in SCN resistance.

Using TILLING[11,12], we identified two Forrest mutants, F6266 and F6756, carrying mutations in the *SHMT* gene on chromosome 8 that lead to missense mutations at amino acid positions 61 (E61K) and 125 (M125I), respectively (Fig. 1b). The M125I mutation is predicted to be deleterious (sorting intolerant from tolerant (SIFT) score = 0.02) to the protein. Both mutants are more susceptible to SCN (Fig. 1c). In the segregating M3 F6756 (M125I) mutant plants, the mutation is directly correlated with the SCN resistance phenotype of individual plants.

To establish a link between *SHMT* alleles and soybean resistance to SCN, we scored 81 soybean lines representing 90% of the sequence diversity in soybean[13] for SCN female index and then determined their SNP-based *SHMT* haplotype (Supplementary Table 1). The *SHMT* gene was fully sequenced from 28 selected soybean lines (Fig. 2). Eight different *SHMT* haplotypes were identified. Soybean lines with haplotypes H1–H3 carry resistant alleles at *SHMT* and *rhg1* and are resistant to SCN. These include soybean lines PI 548402 (Peking), Forrest, PI 90763, PI 437654 and PI 89772, all of which exhibit 'Peking-type' resistance, which requires both *rhg1* and *Rhg4* (ref. 5).

Further evidence that the *SHMT* gene confers SCN resistance comes from knockdown studies using virus-induced gene silencing (VIGS) and targeted RNA interference (RNAi). We cloned 328 bp of *SHMT* into bean pod mottle virus (BPMV) RNA-2 (ref. 14) and generated infected tissue. Silencing of the *SHMT* gene in the SCN-resistant RIL ExF67 by inoculation with BPMV-*SHMT* results in an increase in susceptibility to SCN compared to ExF67 inoculated with BPMV only (Fig. 3a; $P < 0.0001$). At the time of nematode inoculation,

[1]Department of Plant, Soil and Agricultural Systems, Southern Illinois University, Carbondale, Illinois 62901, USA. [2]Division of Plant Sciences, Christopher S. Bond Life Sciences Center and Interdisciplinary Plant Group, University of Missouri, Columbia, Missouri 65211, USA. [3]Department of Computer Science, Christopher S. Bond Life Sciences Center and Informatics Institute, University of Missouri, Columbia, Missouri 65211, USA. [4]Department of Plant Pathology and Microbiology, Iowa State University, Ames, Iowa 50011, USA. [5]Department of Agronomy, Iowa State University, Ames, Iowa 50011, USA.
*These authors contributed equally to this work.

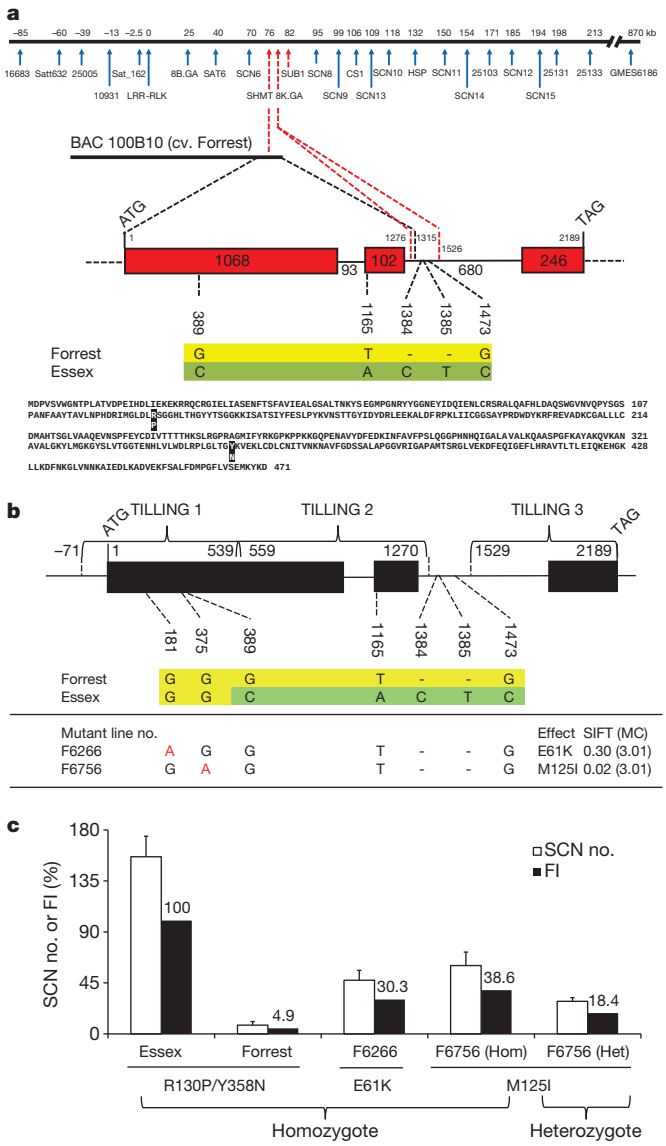


**Figure 1 | *Rhg4* positional cloning and functional validation of *SHMT* by TILLING.** **a**, Top: high-density genetic map of the *Rhg4* locus showing double recombinants ExF74 and FxW5093 using Williams 82 sequence as reference. Red arrows represent DNA markers with Forrest alleles. BAC clone 100B10 encompasses a partial *SHMT* sequence. Middle: *SHMT* gene model and polymorphisms between Forrest and Essex. Bottom: SHMT predicted protein sequence showing amino acid differences (R130P and Y358N) between Forrest and Essex. **b**, Identification of two missense TILLING mutations (E61K and M125I) within the *SHMT*. MC, median sequence conservation. **c**, SCN phenotype of the TILLING mutants. FI, female index. Bars, s.e.m.

*SHMT* transcript levels were determined by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) to have decreased by an average of 74% in the roots of plants inoculated with BPMV-*SHMT* compared with those inoculated with BPMV only (Fig. 3b). For targeted RNAi, a 338-bp double-stranded RNA (dsRNA) corresponding to *SHMT* was expressed under the control of an SCN-inducible zinc finger transcription factor promoter[15] in soybean hairy roots. Nematode reproduction on hairy roots of the SCN-resistant RIL ExF67 transformed with pZF-SHMTi is greater than on ExF67 hairy roots transformed with pZF-GUSi control (Fig. 3c; $P < 0.01$).

Next we fused 2.3 kb of *SHMT* promoter sequence from Forrest and Essex cultivars with the β-glucuronidase (*GUS*) reporter gene. In uninfected roots, *SHMT* expression is observed within the vasculature of roots (Supplementary Fig. 3). Nematode infection experiments of soybean hairy roots of ExF67 and ExF63 transformed with the Forrest pSHMT-GUS construct confirm expression of *SHMT* within syncytia of both lines (Fig. 3d–f). The same pattern of GUS expression is observed in nematode-infected soybean hairy roots of ExF63 transformed with the Essex pSHMT-GUS construct (Supplementary Fig. 2b). These data verify *SHMT* expression in nematode feeding sites of both resistant and susceptible soybean and indicate that differences in *Rhg4* expression are unlikely to explain the resistance phenotype.

To confirm that *SHMT* is the resistance gene at the *Rhg4* locus, we transformed the SCN-susceptible RIL ExF63 with a 5.1-kb genomic fragment that includes the Forrest *SHMT* gene and the 2.3 kb of sequence upstream of the start and the 0.57 kb downstream of the stop codon (gSHMT). SCN resistance was restored in the complemented, transformed hairy roots (Fig. 3g; $P < 0.0001$), confirming that *SHMT* is the *Rhg4* gene.

To gain insight into the mechanistic basis of resistance, a structural model of SHMT was obtained by applying homology modelling. The predicted structure was then used to examine how the variant genotypes may be affecting the structural and functional properties of the enzyme. Forrest P130R and N358Y are co-localized with the tetrahydrofolate (THF)/5,10-methylene tetrahydrofolate (MTHF)/5-formyltetrahydrofolate (FTHF) binding site and are in close proximity to pyridoxal 5′-phosphate (PLP)-serine (PLS), PLP-glycine (PLG), and one of the two glycine binding sites. The position of the E61K mutation overlaps with the PLS and THF/MTHF/FTHF binding sites (Fig. 4a, b, Supplementary Data and Supplementary Video 1). The findings suggest that these residue changes may directly affect the reversible interconversion of serine and THF to glycine and MTHF. On the other hand, the M125I mutation in the Forrest F6756 TILLING mutant is found in an interior β-sheet (Fig. 4a, b), indicating that there may be a different mechanism altering the function of SHMT in this mutant, perhaps through the structural instability of the region affected by the TILLING mutation.

We further tested the ability of the Forrest, Essex, F6266 and F6756 Shmt proteins to complement an *Escherichia coli* glycine auxotroph[16]. Essex Shmt fully restores growth, Forrest Shmt partially restores growth, and the F6266 and F6756 Shmt proteins are unable to restore growth of the mutant (Fig. 4c, d). Although there seems to be a difference in the ability of the Forrest and Essex Shmt proteins to complement the *E. coli* glycine auxotroph, this finding establishes that both the Forrest and Essex alleles encode functional Shmt enzymes. Consistent with this observation, kinetic studies show different reaction kinetics of the Forrest and Essex Shmt proteins. Essex Shmt reaction velocity increases with increasing concentrations of THF, peaks, and then declines to a lower rate and remains steady (Fig. 4e). On the other hand, Forrest Shmt seems to follow typical Michaelis–Menten kinetics (Fig. 4f and Supplementary Fig. 4). The inhibition of Essex Shmt activity at high THF concentrations might have regulatory implications in a cellular environment[17]. The residue changes in Forrest Shmt seem to change this regulatory property of the enzyme.

Shmt is a ubiquitous enzyme in nature with a key role in one-carbon folate metabolism that is conserved across kingdoms. Although the enzyme has multiple catalytic activities, one of its main roles is to catalyse the reversible conversion of serine and THF to glycine and MTHF to supply one-carbon units for *de novo* purine, thymidylate and methionine synthesis, underlying its importance in DNA synthesis and cellular methylation reactions. Consequently, in humans, mutations in SHMT and folate deficiency have been linked to a wide range of disease states[18–21].

Transcriptional and metabolic profiling studies of syncytia, the feeding cells formed by cyst nematodes in plant roots, support a high demand on folate one-carbon metabolism for their development and maintenance[22,23]; perturbations to which could severely compromise their activity. Syncytia induced in plants resistant to SCN degenerate by what has been described as a hypersensitive response, a form of localized programmed cell death (PCD) in plants to ward off invading pathogens. Molecular analysis has identified increased defence-related

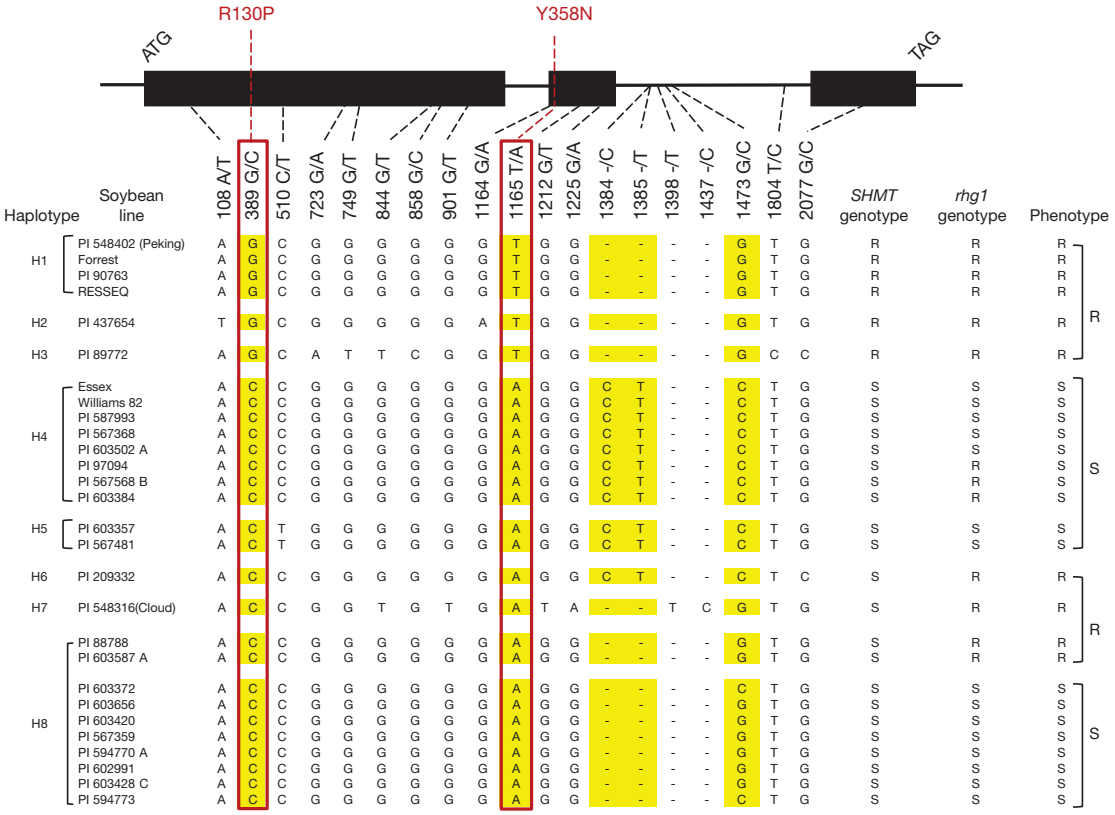


**Figure 2 | Haplotypes identified at *SHMT* in 28 soybean lines.** The *SHMT* coding region for the 28 lines shown here was sequenced. Lines are classified resistant (R) to SCN if FI ≤ 10% and susceptible (S) if FI > 10%. Polymorphic sites were positioned relative to the first nucleotide of the start codon in Forrest. Red boxes indicate the G-to-C and T-to-A transitions resulting in the amino acid substitutions R130P and Y358N, respectively, which are correlated with the SCN phenotype.

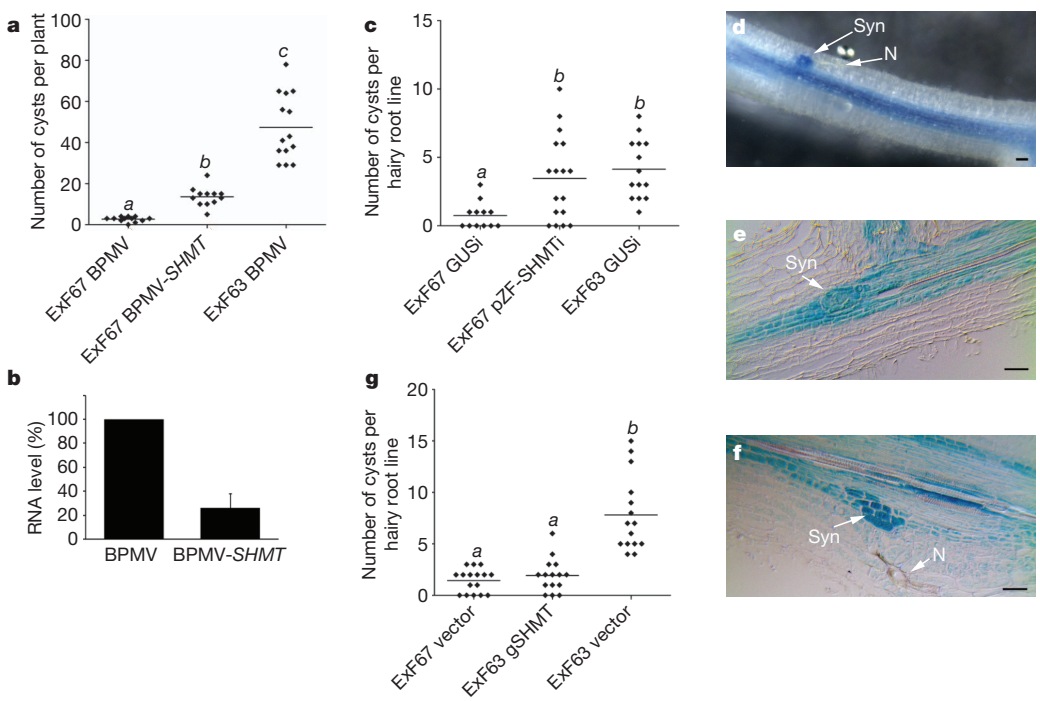


**Figure 3 | Functional validation of *SHMT* by VIGS, RNAi and complementation.** **a**, Nematode reproduction on SCN-resistant RIL ExF67 inoculated with BPMV or BPMV-*SHMT* ($P < 0.0001$). **b**, qPCR of *SHMT* transcript levels in control and *SHMT*-silenced roots (mean and s.e. of five samples). **c**, Nematode reproduction on ExF67 hairy root lines transformed with a *SHMT* RNAi construct ($P < 0.01$). **d**–**f**, Forrest *SHMT* promoter-*GUS* analysis in ExF67 (**d**, **e**) and ExF63 (**f**) showing expression in syncytial feeding cells at 3 days after inoculation with SCN. N, nematode; Syn, syncytium. Scale bars, 50 μm. **g**, SCN reproduction on ExF63 hairy root lines transformed with a 5.1-kb Forrest *SHMT* gene fragment ($P < 0.001$). In graphs, diamonds represent the number of cysts on a single root system or hairy root line and the bars indicate the mean values. At least three independent experiments were performed showing similar results. Data from one experiment are presented. Letters indicate a significant difference at the indicated *P* value.

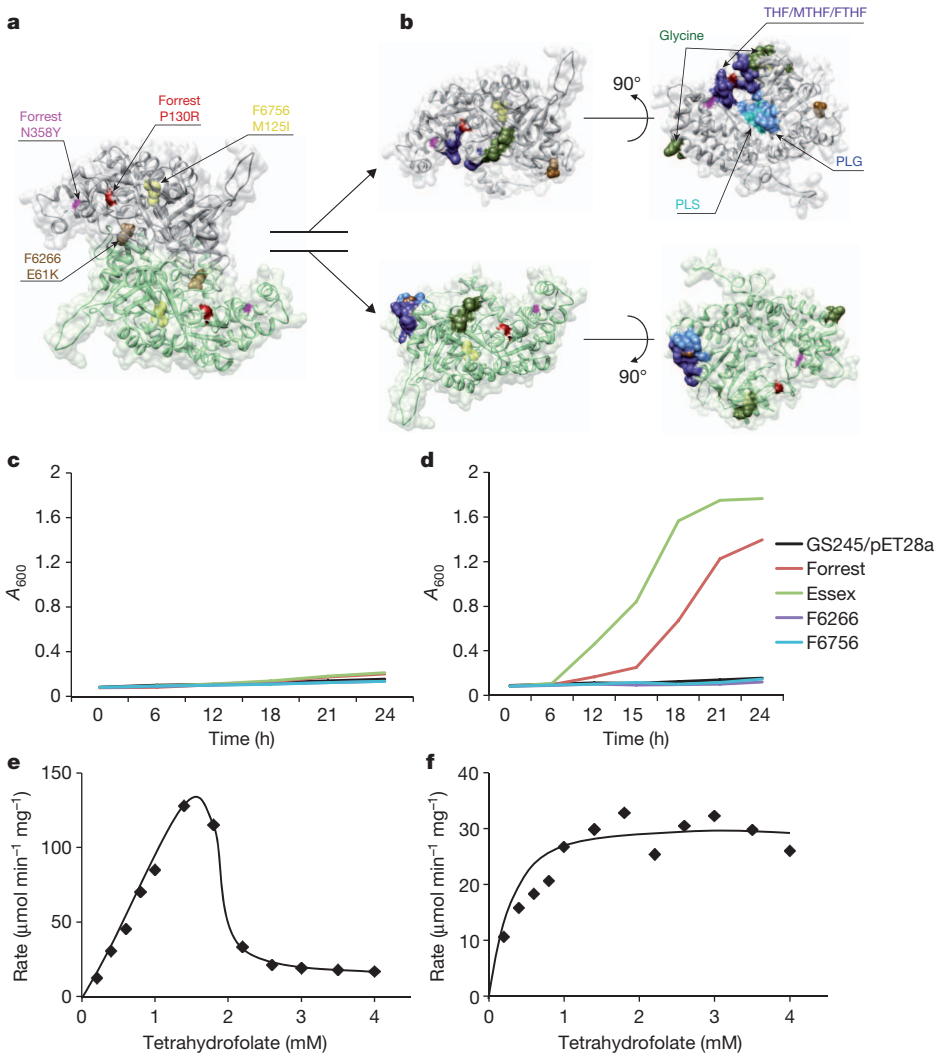


**Figure 4 | Modelled structure and biochemical analysis of SHMT.** **a**, Homology model of the Essex SHMT homodimer. Forrest polymorphisms, P130R and N358Y, are located on the dimer surface; the Forrest F6266 (E61K) mutation is buried in the dimerization interface; the Forrest F6756 (M125I) mutation is located in the core of each monomeric subunit. **b**, Split-open view of the SHMT dimer with the ligand binding sites located at the interface. **c**, **d**, Complementation of an *E. coli* glycine auxotroph (GS245(DE3)pLysS; *shmt*−) by the Essex, Forrest, F6266 and F6756 Shmt proteins. Absorbance at 600 nm ($A_{600}$) of cultures was measured at different time points as shown, uninduced (**c**) or induced (**d**) with 0.25 mM IPTG. GS245(DE3)pLysS carrying the pET28a vector was used as a control. **e**, **f**, Enzyme kinetics of Essex (**e**) and Forrest (**f**) Shmt proteins. The graphs were plotted as rate of product formation per min per mg of enzyme at varying concentrations of tetrahydrofolate and 2 mM serine. Graphs shown are representative of three experiments.

gene expression associated with oxidative stress, apoptotic cell death, and the plant hypersensitive response in syncytia formed in resistant plants[15].

Our findings are consistent with the dominant nature of *Rhg4*-mediated resistance and support a model wherein the Forrest Shmt has acquired a new or altered function. Computational analysis predicts that the polymorphisms in Forrest Shmt reside near ligand-binding sites, and our biochemical analyses suggest that these residue changes probably impair a key regulatory property of the enzyme. Any changes in Shmt function can have wide-ranging affects on one-carbon folate metabolism. Alterations to folate homeostasis leading to folate deficiency, which can induce apoptosis in mammalian cells[24,25], may trigger hypersensitive-response-like programmed cell death of the developing feeding cells and subsequent death of the nematode. Alternatively, the Forrest Shmt may lead to the production of a nematicidal compound or serve as the target of a nematode-secreted small molecule or effector protein to trigger a resistance signalling pathway. Although the nutritional requirements of plant-parasitic nematodes are not well defined, it is assumed that nematodes, like other animals, acquire folate from their diet. Thus, the nematode's nutritional requirements may also be influencing folate metabolism in developing feeding cells. Perturbations to the plant's folate pathway may lead to a nutritional deficiency that starves the nematode. Upon nematode death, the loss of stimuli required for maintenance of the feeding cells results in their degeneration. The work described here provides a foundation for future molecular and biochemical studies aimed at achieving a mechanistic understanding of how Shmt functions in resistance to SCN.

## METHODS SUMMARY

Three $F_{2:6}$ recombinant inbred line (RIL) populations segregating for resistance to SCN Hg type 0 were developed from crosses of SCN-resistant cv. Forrest (F) with the SCN-susceptible soybean cultivars Williams 82 (W) or Essex (E) and used in mapping studies. Because Forrest resistance to SCN requires both *rhg1* and *Rhg4* (ref. 5), genotyping was conducted using DNA markers flanking both loci to detect informative recombinants at the *Rhg4* locus (Supplementary Table 1). The two $F_{2:6}$ RILs, ExF67 ($rhg1_F rhg1_F Rhg4_F Rhg4_F$) and ExF63 ($rhg1_F rhg1_F Rhg4_E Rhg4_E$), used in functional analyses are resistant and susceptible, respectively, to SCN[10]. An ethylmethane sulphonate (EMS)-mutagenized M2 population of Forrest containing 1,920 M2 families was used to screen for mutations as previously described[12]. VIGS and hairy root assays were carried out as previously described[10,14]. A 5,103-bp Forrest *SHMT* genomic DNA fragment (GenBank accession number JQ714083) was cloned and sequenced for complementation analysis. Homology

modelling was done using MODELLER-9[26]. *Escherichia coli* complementation analysis[16] and SHMT assays[27] were performed as described.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 4 June; accepted 27 September 2012.**
**Published online 15 October 2012.**

1. Koenning, S. R. & Wrather, J. A. Suppression of soybean yield potential in the continental United States from plant diseases estimated from 2006 to 2009. *Plant Health Prog.* http://dx.doi.org/10.1094/PHP-2010-1122-01-RS (2010).
2. Caldwell, B. E., Brim, C. A. & Ross, J. P. Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines. Agron. J.* **52,** 635–636 (1960).
3. Matson, A. L. & Williams, L. F. Evidence of a fourth gene for resistance to the soybean cyst nematode. *Crop Sci.* **5,** 477 (1965).
4. Concibido, V. C., Diers, B. W. & Arelli, P. R. A decade of QTL mapping for cyst nematode resistance in soybean. *Crop Sci.* **44,** 1121–1131 (2004).
5. Meksem, K. *et al.* 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. *Theor. Appl. Genet.* **103,** 710–717 (2001).
6. Endo, B. Y. Histological responses of resistant and susceptible soybean varieties, and backcross progeny to entry development of *Heterodera glycines. Phytopathology* **55,** 375–381 (1965).
7. Dong, K. & Opperman, C. H. Genetic analysis of parasitism in soybean cyst nematode *Heterodera glycines. Genetics* **146,** 1311–1318 (1997).
8. Niblack, T. L., Colgrove, A. L., Colgrove, K. & Bond, J. P. Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Online. *Plant Health Prog.* http://dx.doi.org/10.1094/PHP-2008-0118-01-RS (2008).
9. Melito, S. *et al.* A nematode demographics assay in transgenic roots reveals no significant impacts of the *Rhg1* locus LRR-Kinase on soybean cyst nematode resistance. *BMC Plant Biol.* **10,** 104 (2010).
10. Liu, X. *et al.* Soybean cyst nematode resistance in soybean is independent of the *Rhg4* locus LRR-RLK gene. *Func. Integr. Gen.* **11,** 539–549 (2011).
11. Cooper, J. L. *et al.* TILLING to detect induced mutations in soybean. *BMC Plant Biol.* **8,** 9 (2008).
12. Meksem, K. *et al.* TILLING: A reverse genetics and a functional genomics tool in soybean. In *The handbook of Plant Functional Genomics: Concepts and Protocols* (eds Kahl, G. & Meksem, K.) 251–265 (Wiley, 2008).
13. Hyten, D. *et al.* Impacts of genetic bottlenecks on soybean genome diversity. *Proc. Natl Acad. Sci. USA* **103,** 16666–16671 (2007).
14. Zhang, C. Q., Bradshaw, J. D., Whitham, S. A. & Hill, J. H. The development of an efficient multipurpose *Bean pod mottle virus* viral vector set for foreign gene expression and RNA silencing. *Plant Physiol.* **153,** 52–65 (2010).
15. Kandoth, P. K. *et al.* The soybean *Rhg1* locus for resistance to the soybean cyst nematode *Heterodera glycines* regulates expression of a large number of stress- and defense-related genes in degenerating feeding cells. *Plant Physiol.* **155,** 1960–1975 (2011).
16. Alfadhli, S. & Rathod, P. K. Gene organization of a *Plasmodium falciparum* serine hydroxymethyltransferase and its functional expression in *Escherichia coli. Mol. Biochem. Parasitol.* **110,** 283–291 (2000).
17. Reed, M. C., Lieb, A. & Nijhout, H. F. The biological significance of substrate inhibition: A mechanism with diverse functions. *Bioessays* **32,** 422–429 (2010).
18. Skibola, C. F. *et al.* Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. *Blood* **99,** 3786–3791 (2002).
19. Lim, U. *et al.* Polymorphisms in cytoplasmic serine hydroxymethyltransferase and methylenetetrahydrofolate reductase affect the risk of cardiovascular disease in men. *J. Nutr.* **135,** 1989–1994 (2005).
20. Heil, S. G. *et al.* Is mutated serine hydroxymethyltransferase (SHMT) involved in the etiology of neural tube defects? *Mol. Genet. Metab.* **73,** 164–172 (2001).
21. Kim, Y. I. Role of folate in colon cancer development and progression. *J. Nutr.* **133,** 3731S–3739S (2003).
22. Ithal, N. *et al.* Developmental transcript profiling of cyst nematode feeding cells in soybean roots. *Mol. Plant Microbe Interact.* **20,** 510–525 (2007).
23. Hofmann, J. *et al.* Metabolic profiling reveal local and systemic responses of host plants to nematode parasitism. *Plant J.* **62,** 1058–1071 (2010).
24. Novakovic, P., Stempak, J. M., Sohn, K.-J. & Kim, J.-I. Effects of folate deficiency on gene expression in the apoptosis and cancer pathways in colon cancer cells. *Carcinogenesis* **27,** 916–924 (2006).
25. Bagnyukova, T. V. *et al.* Induction of oxidative stress and DNA damage in rat brain by a folate/methyl-deficient diet. *Brain Res.* **1237,** 44–51 (2008).
26. Sali, A. & Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
27. Zhang, Y., Sun, K. & Roje, S. An HPLC-based fluorometric assay for serine hydroxymetyltransferase. *Anal. Biochem.* **375,** 367–369 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank X. Yang for technical assistance with soybean hairy root propagation and M. Kroll for proofreading the manuscript. We thank R. Hussey, W. Gassmann, P. Gresshoff, A. Bendahmane, D. Xu and B. McClure for critical reading of the manuscript. We thank K. Sharma and S. Puthur for use of the HPLC facility and technical help. This work was supported by the Illinois-Missouri Biotechnology Alliance (project 2005-3 to M.G.M. and K.M.), United Soybean Board (project 0253 to K.M., S.C. and M.G.M.; project 3253 to K.M. and S.C.; project 2268 to K.M. and M.G.M; project 1251 to J.H. and S.A.W.), USDA-NIFA (grant 2006-35300-17195 to K.M.), the National Science Foundation Plant Genome Research Program (grant 0820642 to S.A.W., J.H., M.G.M. and T.J.B.), the National Science Foundation CAREER Program (DBI-0845196 to D.K.), Missouri Soybean Merchandising Council (project 258 to M.G.M.), Illinois Soybean Association, North Central Soybean Research Program, Iowa Soybean Association, and a Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship (to S.D.W.).

**Author Contributions** S.L. and P.K.K. contributed equally as first authors. K.M. and M.G.M. contributed equally as senior authors. S.L. carried out mapping and haplotyping studies. S.L., A.J. and T.E.-M. developed the TILLING population used in this study. S.C. developed two of the RIL populations used in this study. S.L. identified the TILLING mutants and conducted the zygosity analysis. G.Y. and R.H. carried out SCN phenotyping of all RILs, TILLING mutants, and soybean lines. G.Y. collected leaf tissues for RIL and TILLING studies. P.K.K. collected leaf tissues haplotyping analyses. P.K.K. and S.L. confirmed the mutations, cloned the genes, and conducted sequence analyses. P.K.K., C.Y., R.H. and P.S.J. developed and carried out VIGS assays. P.K.K. conducted RNAi experiments. P.K.K. carried out the promoter analysis. P.K.K. and J.A. carried out the complementation analysis. S.D.W. and D.K. performed the computational analysis. P.K.K. performed the biochemical studies. J.H., S.A.W. and T.J.B. provided materials and advice for VIGS analysis. K.M., M.G.M. and D.K. designed the research, and together with S.L., P.K.K. and S.D.W. wrote the manuscript. All authors reviewed and commented on the manuscript.

**Author Information** All sequences have been deposited in GenBank/EMBL/DDBJ under the following accession numbers: Forrest *SHMT* full-length genomic DNA, JQ714083; Essex *SHMT* full-length genomic DNA, JQ714084; Forrest *SHMT* cDNA sequence, JQ714080; Essex *SHMT* cDNA sequence, JQ714079; Forrest *SHMT* TILLING mutant F6266 sequence, JQ714081; Forrest *SHMT* TILLING mutant F6756 sequence, JQ714082; Forrest *SUB1* cDNA sequence, JQ762395; Essex *SUB1* cDNA sequence, JQ762396; Forrest *SUB1* promoter sequence, JQ762397; Essex *SUB1* promoter sequence, JQ904711. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M. (meksemk@siu.edu) or M.G.M. (goellnerm@missouri.edu).

## METHODS

**Nematode and plant material.** The SCN (*Heterodera glycines* Ichinohe) inbred population PA3 (Hg type 0) used in this study was mass-selected on soybean cv. Williams 82 according to standard procedures[28] at the University of Missouri. The soybean cultivar Forrest[29] is resistant to SCN PA3. The soybean cultivars Essex[30] and Williams 82 (ref. 31) are susceptible to SCN PA3. Forrest was used to develop an ethyl methane sulphonate (EMS)-mutagenized M2 population of 1,920 lines for TILLING[12]. The two $F_{2:6}$ RILs, ExF67 ($rhg1_F rhg1_F Rhg4_F Rhg4_F$) and ExF63 ($rhg1_F rhg1_F Rhg4_E Rhg4_E$), used in this study are resistant and susceptible, respectively, to PA3[10]. These two RILs differ at the majority of markers assigned to the *Rhg4* region and seem to be nearly opposite recombination events. The collection of soybean lines used in this study was obtained from the USDA Soybean Germplasm Collection, University of Illinois.

**Map-based positional cloning of the *Rhg4* gene.** Three genetic populations segregating for resistance to SCN PA3 (Hg type 0) were used for mapping. These included an $F_{2:6}$ RIL population from a cross between Forrest and Essex[5] (98 individuals) and two large $F_{2:6}$ RIL populations generated from crosses between Forrest and either Essex (1,755 lines) or Williams 82 (2,060 lines), to enrich the chromosomal interval carrying the *Rhg4* gene with recombinants. SCN phenotyping was conducted according to ref. 32.

Because Forrest resistance to SCN requires both *rhg1* and *Rhg4* genes[5], genotyping was conducted using DNA markers flanking both loci to detect informative recombinants at the *Rhg4* locus (Supplementary Table 2). The SSR markers Satt632, Sat_162 (http://soybase.org) and GMES6186 (ref. 33) were used to identify chromosomal breakpoints at the *Rhg4* locus. PCR amplifications were performed using DNA from individuals from each of the three genetic populations. Cycling parameters were as follows: 35 cycles of 94 °C 30 s, 50 °C 30 s and 72 °C 30 s with 7 min of extension at 72 °C. The PCR products were separated on 3–4% metaphor agarose gels. The identified recombinants were subject to a second screening by using the SSR markers Sat_210 and Satt309 (http://soybase.org) and SIUC-SAT143 to identify the *rhg1* genotype of each recombinant.

To enrich the chromosomal regions carrying the *Rhg4* locus with DNA markers, the GenBank published sequences AX196297 and AX197417 spanning this region were used to design PCR primers every 5 to 10 kb of the 300 kb carrying the *Rhg4* locus (Supplementary Table 2). DNA from Forrest, Essex and Williams 82 were tested with each primer by using a modified EcoTILLING protocol to find and map polymorphic sequences at the *Rhg4* locus[10,12]. The identified SNP and indel DNA markers were integrated into the informative recombinants to identify chromosomal breakpoints and the interval that carried the *Rhg4* gene.

The closest DNA markers harbouring the *Rhg4* locus were used to screen three Forrest BAC libraries[34]. The BAC clone 100B10 was identified, integrated with the developed genetic map, and partially sequenced[5,10].

**Isolation of the *SHMT* genomic and cDNA sequences.** A 5,103-bp Forrest *SHMT* genomic DNA fragment (GenBank accession number JQ714083) was cloned and sequenced. The fragment spans 2,339-bp of sequence 5′ of the ATG start site, 2,189 bp of sequence from start to stop including 3 exon and 2 introns, and 0.575 kb of sequence 3′ of the stop codon. Because the BAC clone 100B10 contains only a partial *SHMT* gene sequence that includes the 2,339 bp of sequence 5′ of the ATG start site and 1,315 bp downstream of the ATG start site (Fig. 1a), we used an internal SacI site at position 108 from the ATG start for a PCR-based cloning approach of the full-length genomic sequence. First, a 2,447-bp fragment, including the 2,339 bp of sequence 5′ of the ATG start site and 108 bp of exon 1, was amplified by PCR using both a forward primer designed with an AscI site and a reverse primer spanning the internal SacI site. The fragment was digested and cloned into the CGT35S vector[35] by using AscI and SacI. An SbfI site was also introduced into the forward primer internal to AscI for subsequent subcloning for complementation analysis (Supplementary Table 2; see below). The remainder of the *SHMT* gDNA fragment, including the unique internal SacI site, was amplified from Forrest genomic DNA by PCR with a forward primer spanning the internal SacI site and a reverse primer designed with a KpnI site. The fragment was digested and cloned into the SacI and KpnI sites downstream of the 5′ fragment in the above CGT35S clone. The reverse primer was designed with an SbfI site internal to KpnI for the purpose of subsequent subcloning for complementation analysis (Supplementary Table 2). The fragments were ligated together using the internal SacI restriction site to generate the 5,103-bp *SHMT* genomic DNA fragment and were sequence-verified for accuracy. Primers designed to the Forrest genomic DNA sequence were used to clone the Essex *SHMT* genomic DNA sequence. PCR primers based on the Forrest and Essex genomic DNA sequences were used to amplify the corresponding cDNA sequences. Genomic DNA was isolated from young leaves using the DNeasy plant mini kit (Qiagen). Total RNA was isolated from roots using the RNeasy plant mini kit (Qiagen), and cDNA was synthesized using a cDNA synthesis kit (Invitrogen).

**Mutation screening of *SHMT*.** Primers specific for *SHMT* (Supplementary Table 2) were used to screen a population of 1,920 EMS-mutagenized M2 lines from the SCN-resistant cv. Forrest[11,12]. The gene was divided into three intervals (Fig. 1b), and TILLING was performed as previously described[12]. The *SHMT* gene of each mutant was sequenced to characterize the identified allele and its subsequent amino acid changes within the predicted protein sequences. SIFT predictions were performed on identified mutations. SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids[36]. SIFT predictions with median sequence conservation (MC) $<3.25$ are considered confident. Changes with a SIFT score $<0.05$ are predicted to be damaging to the protein. Both missense mutations identified had MC $<3.25$ (Fig. 1b), thus the SIFT predictions can be considered confident.

**Phenotype and zygosity analyses of *SHMT* TILLING mutants.** Mutant seeds were planted and scored for their SCN female index as described[32]. DNA from each plant was subjected to TILLING analysis. Wild-type Forrest reference DNA was withheld from the reaction tube before mismatch analyses to detect the zygosity level of the identified mutant.

**Haplotyping of soybean lines.** A total of 81 soybean lines (plant introductions, landraces, and elite cultivars) representing 90% of the genetic variability in soybean[13] were scored for their SCN female index. Lines were classified resistant (R) to SCN if the FI was $\leq$10% and susceptible (S) if the FI was $>$10%. Soybean lines were genotyped at the *Rhg4* locus by using the DNA markers Sat_162 and SUB1 and at the *rhg1* locus by using the DNA markers 560, 570 and Satt309. The coding region of *SHMT* for 28 lines was sequenced (Fig. 2). Common SNPs and indels were identified and used to determine the different *SHMT* haplotypes.

**Virus-induced gene silencing.** Bean pod mottle virus (BPMV, genus *Comovirus*) is an effective virus-induced gene silencing (VIGS) vector for soybean. BPMV has a bipartite positive-strand RNA genome consisting of RNA-1 and RNA-2. We used a DNA-based system that places the cDNAs of BPMV genomic RNA-1 and RNA-2 under the control of the cauliflower mosaic virus (CaMV 35S) promoter. Two BPMV VIGS vectors, pBPMV IA-R1M and pBPMV-IA-D35, were used in this study[14]. pBPMV-IA-D35 is a derivative of pBPMV-IA-R2 containing BamHI and KpnI restriction sites between the cistrons encoding movement protein and the large coat protein subunit. Briefly, a 328-bp fragment (spanning base pairs 210–537) of the *SHMT* cDNA sequence (GenBank accession number JQ714080) was amplified from soybean (cv. Forrest) root cDNA by RT–PCR. PCR products were digested with BamHI and KpnI and ligated into pBPMV-IA-D35 digested with the same enzymes to generate pBPMV-IA-SHMT (BPMV-*SHMT*). Gold particles coated with pBPMV-IA-R1M and pBPMV-IA-SHMT were co-bombarded into soybean leaf tissue as described[14]. At 3–4 weeks after inoculation, BPMV-infected leaves were collected, lyophilized, and stored at −20 °C for future experiments. Infected soybean leaf tissues were ground in a mortar and pestle with 0.05 M potassium phosphate buffer (pH 7.0) and used as virus inoculum for VIGS assays.

The SCN-resistant RIL ExF67 was inoculated with pBPMV-IA-SHMT. Control plants were infected with BPMV only. Each treatment consisted of at least 12 plants. Unifoliate leaves of 9-day-old plants were rub-inoculated with virus using carborundum as described[14]. Plants were grown in a growth chamber set to the following conditions: 20–21 °C, 16 h light/8 h dark, and 100 $\mu mol\, m^{-2}\, s^{-1}$ light intensity. Twenty-one days after virus inoculation, plants were inoculated with 1,500 SCN eggs and maintained at 20 °C for 35 days. Cysts were isolated from the root systems of individual plants by decanting and sieving and counted under a stereomicroscope. The results were plotted and analysed for statistical significance by an unpaired *t*-test using GraphPad PRISM software. To estimate *SHMT* gene silencing in roots, root tissues were collected at 21 days after virus inoculation (the time of nematode inoculation) from two representative plants inoculated with either pBPMV-IA-SHMT or BPMV only and frozen at −80 °C for RNA isolation and qPCR analysis.

**Hairy root RNAi experiments.** A 338-bp fragment (spanning base pairs 205–542) of the *SHMT* cDNA sequence was amplified from soybean (cv. Forrest) root cDNA by RT–PCR, cloned into the pDONR/Zeo gateway cloning vector (Invitrogen), and subsequently moved to a gateway RNAi binary vector under the control of the nematode-inducible Glyma15g04570.1 promoter[15] (pZF-RNAi vector) to generate pZF-SHMTi. The pZF-RNAi vector was constructed by introducing gateway cloning sites flanking the FADR intron downstream of pZF promoter in the pAKK vector[35], which has a GFP selectable marker *in planta*. Transgenic ExF67 hairy roots transformed with pZF-SHMTi were produced from soybean cotyledons[15]. ExF63 and ExF67 hairy roots transformed with pZF-GUSi (the pZF-RNAi vector containing a portion of the *GUS* gene) were used as susceptible and resistant controls, respectively. GFP-positive hairy roots were root-tip propagated three times on media containing antibiotic to clear *Agrobacterium* before nematode inoculation as described previously[15]. Briefly, hairy roots (3–4 cm in length) were grown in square Petri plates and infected with approximately 400 sterile infective second-stage nematode juveniles (J2s) 1 cm above the root tip. The plates were

incubated in the dark at room temperature for 30 days. After 30 days, cysts were counted under a stereomicroscope. The experiment was conducted independently three times with a minimum of 12 independent hairy root lines per treatment. The results were plotted and analysed for statistical significance by an unpaired *t*-test using GraphPad PRISM software.

**Promoter-*GUS* analysis.** A 2,339-bp fragment corresponding to sequence 5′ of the ATG start site of the Forrest *SHMT* gene (GenBank accession number JQ714083) and the same region from the Essex *SHMT* gene (GenBank accession number JQ714084) were amplified by PCR from the BAC clone 100B10 and Essex gDNA, respectively, and cloned into the pYXT1 vector[37] to generate transcriptional fusions with the β-glucuronidase (*GUS*) gene. Soybean hairy roots transformed with these constructs were generated and infected with SCN. At 2 and 4 days after inoculation, root pieces excised from the infection zone were stained for GUS activity[38]. Multiple roots from at least five independent lines were stained for each construct. Root pieces were fixed with 4% v/v paraformaldehyde in phosphate-buffered saline overnight at room temperature, embedded in paraffin, and sectioned longitudinally to a thickness of 10 μm. The sections were observed using differential interference contrast microscopy on a Vanox (Olympus) microscope and photographed with a CMOS colour digital camera.

**Genomic complementation experiments.** The 5,103-bp Forrest *SHMT* genomic DNA fragment was subcloned into the Sbf1 restriction site of the pAKK binary vector, which has a GFP selection for transgenic events. Transgenic hairy roots were produced and infected with SCN as described for RNAi experiments. The SCN-susceptible RIL ExF63 was used for the complementation experiment. Control hairy roots were produced by transforming ExF63 and ExF67 hairy roots with the pAKK binary vector carrying only the *SHMT* promoter sequence. The experiment was conducted independently five times with a minimum of fifteen independent hairy root lines per treatment. The results were plotted and analysed for statistical significance by an unpaired *t*-test using GraphPad PRISM software.

**RNA isolation and qPCR analysis.** Total RNA was isolated from root tissues using the RNeasy plant mini kit (Qiagen) according to the manufacturer's instructions. Real-time qRT–PCR was conducted as described[15]. Samples were normalized relative to soybean ubiquitin and calibrated to the expression in the BPMV control sample.

**Computational methods.** A two-pronged computational approach was performed to annotate structurally and functionally the identified Shmt protein and to estimate the effects of the mutations on Shmt function. First, a homology model of Shmt was obtained using Essex sequence as a target. Second, the functional sites were mapped onto the surface of Shmt using the structural information of ligand binding by Shmt homologues. The homology analysis of Shmt determined 43 structurally resolved Shmt homologues from a diverse set of bacterial and mammalian species; no structurally resolved plant Shmt proteins were found. Among the group of four homologues with the highest sequence similarities, the mouse SHMT (Protein Data Bank accession 1EJI) with the largest coverage of the Shmt sequence (sequence identity 57%, template coverage 100%) was selected as a template for homology modelling of Shmt. Homology modelling was done using MODELLER-9[26] and the top-ranked model was selected from the set of candidate models using the MODELLER scoring function. To determine the ligand binding sites for pyridoxal 5′-phosphate (PLP)-serine (PLS), PLP-glycine (PLG), and THF/MTHF/FTHF, the obtained model of Shmt was structurally aligned with each of the orthologous Shmt proteins known to interact with the small ligands, and the ligand binding site from each homologue was mapped on to the surface of the Shmt model through the structural alignment (note that the contributing parts of the ligand binding sites are not identical between the two monomers). The residues constituting the glycine binding sites, GBS1 and GBS2, in the Shmt homologues were identified in the literature and then mapped onto the structure of Shmt in a similar way by using the structural alignment of SHMT with its homologue.

***E. coli* complementation experiment.** An *E. coli* GS245 glyA strain[39] was obtained from the *E. coli* Genetic Stock Center. The GS245 strain was lysogenized with λDE3 phage (Novagen) according to manufacturer's protocols and transformed with pLysS plasmid (Novagen) to generate GS245(DE3)pLysS for complementation analysis[16]. The cDNA sequences corresponding to Forrest, Essex, F6266 and F6756 Shmt proteins were subcloned at NdeI and HindIII sites of pET28a vector (Novagen) for expression of the recombinant proteins with an N-terminal 6×-histidine tag. GS245(DE3)pLysS strains carrying the plasmids were grown in LB medium overnight. The $A_{600}$ value of each culture was measured and an equal absorbance of cells was collected and washed twice in M9 minimal medium (1× M9 salts (Sigma), 10% w/v glucose, 50 mg ml$^{-1}$ phenylalanine, and 10 mg ml$^{-1}$ of thiamine) and finally re-suspended in two flasks of M9 minimal medium with required antibiotics for selection. One of the flasks was supplemented with IPTG at 0.25 mM final concentration. The cultures were grown in a shaker incubator (New Brunswick Scientific) at 37 °C and 180 r.p.m. setting. $A_{600}$ values of cultures were measured at 0 h and 6 h and thereafter every 3 h up to 24 h. The $A_{600}$ measurements were plotted versus time of incubation to obtain growth curves.

**Shmt purification and kinetic analysis.** Shmt proteins were purified from GS245 cultures grown at 30 °C for 4 h after induction with 0.5 mM IPTG at $A_{600}$ of 0.5. His-tagged proteins were purified using His Pur Cobalt Resin (Thermo Scientific) according to manufacturer's instructions. Purified proteins were dialysed against three changes of 50 mM potassium phosphate buffer, pH 7.4, containing 10% v/v glycerol, $10^{-3}$ M tris(hydroxypropyl)phosphine, and $10^{-5}$ M pyridoxal L-phosphate. Protein concentration was estimated by the Bradford method. Shmt assays were performed as described previously[27]. Briefly, 50 μl reactions comprising 2 mM serine, 0.2–4 mM (6*R*,*S*)-$H_4$PteGlu (Tetrahydrofolate, THF) (Schircks Lab), 4 mM tris(hydroxypropyl)phosphine, 0.25 mM PLP, and 6 μg of enzyme in 50 mM potassium phosphate buffer (pH 7.4) were incubated at 30 °C for 20 min. The reaction was stopped by adding 25 μl of 0.1 M dithiothreitol (DTT) and 50 μl of 0.1 M $NaBH_4$. Each reaction was further incubated for 15 min at 37 °C to drive the reduction of 5,10-methylene tetrahydrofolate (MTHF) to 5-methyltetrahydrofolate to completion. The samples were then boiled for 3 min and centrifuged at 20,000 *g* to remove precipitated protein. The supernatant from the centrifugation step was supplemented with 25 μl 0.6 M DTT to prevent product decomposition. A no serine control reaction was performed simultaneously. A 1:10 dilution of this sample was used for HPLC separation and quantification. HPLC injection volume was typically 5 μl. 5-methyltetrahydrofolate in the final reaction mixture was detected and quantified by HPLC coupled with a fluorescent detector set at 289 nm excitation and 359 nm emission wavelengths (Shimadzu Corp). A Restek Pinnacle II C18 column (5 μm, 150 × 4.6mm, Restek US) was used for HPLC separation. The product separation was done isocratically with a mobile phase consisting of 6% v/v acetonitrile in 30 mM phosphate buffer pH 3.0 (adjusted with phosphoric acid) at 0.7 ml min$^{-1}$. The level of 5-methyltetrahydrofolate detected was quantified by comparison with standards. For this, a standard calibration curve was created with known amounts of 5,10-methylene tetrahydrofolate reduced to 5-methyltetrahydrofolate using 0.1 M $NaBH_4$ and 0.1 M DTT and from peak areas of 5-methyltetrahydrofolate eluted. The specific activity of the enzyme at varying concentrations of THF was plotted against THF concentrations to obtain the kinetic curve. Kinetic analysis was repeated three times for each of the enzymes using independent enzyme preparations.


28. Niblack, T. L., Heinz, R. D., Smith, G. S. & Donald, P. A. Distribution, density, and diversity of *Heterodera glycines* in Missouri. *J. Nematol.* **25,** 880–886 (1993).
29. Hartwig, E. F. & Epps, J. M. Registration of 'Forrest' soybeans. *Crop Sci.* **13,** 287 (1973).
30. Smith, T. J. & Camper, H. M. Registration of 'Essex' soybean. *Crop Sci.* **13,** 495 (1973).
31. Bernard, R. L. & Cremeens, C. R. Registration of 'Williams 82' soybean. *Crop Sci.* **28,** 1027–1028 (1988).
32. Brown, S. *et al.* A high-throughput automated technique for counting females of *Heterodera glycines* using a fluorescence-based imaging system. *J. Nematol.* **42,** 201–206 (2010).
33. Hwang, T. Y. *et al.* High density integrated linkage map based on SSR markers in soybean. *DNA Res.* **16,** 213–225 (2009).
34. Meksem, K. *et al.* Two large-inset soybean genomic libraries constructed in a binary vector: Applications in chromosome walking and genome wide physical mapping. *Theor. Appl. Genet.* **101,** 747–755 (2000).
35. Wang, J. *et al.* Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE effector proteins. *New Phytol.* **187,** 1003–1017 (2010).
36. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4,** 1073–1081 (2009).
37. Xiao, Y.-L. *et al.* Analysis of the cDNAs of hypothetical genes on *Arabidopsis* chromosome 2 reveals numerous transcript variants. *Plant Physiol.* **139,** 1323–1337 (2005).
38. Jefferson, R. A. Assaying chimeric genes in plants: the *GUS* gene fusion system. *Plant Mol. Biol. Rep.* **5,** 387–405 (1987).
39. Stauffer, G. V., Plamann, M. D. & Stauffer, L. T. Construction and expression of hybrid plasmids containing the *Escherichia coli glyA* gene. *Gene* **14,** 63–72 (1981).