Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean

David E. Cook *et al.*

Science **338**, 1206 (2012);
DOI: 10.1126/science.1228746

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean

David E. Cook,[1]* Tong Geon Lee,[2]* Xiaoli Guo,[1]*† Sara Melito,[1]‡ Kai Wang,[3] Adam M. Bayless,[1] Jianping Wang,[2]§ Teresa J. Hughes,[1]|| David K. Willis,[4] Thomas E. Clemente,[5] Brian W. Diers,[2] Jiming Jiang,[3] Matthew E. Hudson,[2,6]¶ Andrew F. Bent[1]¶#

The *rhg1-b* allele of soybean is widely used for resistance against soybean cyst nematode (SCN), the most economically damaging pathogen of soybeans in the United States. Gene silencing showed that genes in a 31-kilobase segment at *rhg1-b*, encoding an amino acid transporter, an α-SNAP protein, and a WI12 (wound-inducible domain) protein, each contribute to resistance. There is one copy of the 31-kilobase segment per haploid genome in susceptible varieties, but 10 tandem copies are present in an *rhg1-b* haplotype. Overexpression of the individual genes in roots was ineffective, but overexpression of the genes together conferred enhanced SCN resistance. Hence, SCN resistance mediated by the soybean quantitative trait locus *Rhg1* is conferred by copy number variation that increases the expression of a set of dissimilar genes in a repeated multigene segment.

Soybean (*Glycine max*) is the world's most widely used legume crop, providing 68% of world protein meal as well as food oil and a renewable source of fuel, with a farm gate value of more than $35 billion in the United States alone (www.soystats.com). Soybean cyst nematode (SCN; *Heterodera glycines*) is the most economically damaging pathogen of soybean, causing more than $1 billion in annual losses. SCN has infested most major soybean producing areas worldwide, and there are no practical means of eradication (*1*).

SCN molts through multiple juvenile and adult life stages, including obligate endoparasitic stages on plant roots, to complete its life cycle (*1*). Infective J2 juveniles invade roots of both susceptible and resistant soybean hosts, then reprogram host root cells to form feeding sites using highly evolved, secreted nematode effectors (*2*, *3*).

The soybean *Rhg1* (resistance to *H. glycines*) quantitative trait locus on chromosome 18 consistently contributes much more effective SCN resistance than any other known loci (*4*, *5*). *Rhg1* disrupts the formation and/or maintenance of most potential nematode feeding sites (*1*). Roughly 90% of the commercially cultivated soybean varieties marketed as SCN-resistant in the central United States use the *rhg1-b* allele (haplotype), derived from the soybean line PI 88788, as the main SCN resistance locus. The molecular basis of this SCN resistance has remained unclear.

Genetic mapping has placed *rhg1-b* in an interval that corresponds to a 67-kb segment carrying 11 predicted genes in the genome of the SCN-susceptible but fully sequenced Williams 82 soybean variety (*6*, *7*). It was recently suggested that an amino acid polymorphism in the *Glyma18g02590*-encoded α-SNAP protein in this interval contributes to SCN resistance (*8*), although the authors indicated that this polymorphism does not account for *rhg1-b*–mediated resistance. None of the gene products within the *rhg1-b* genetic interval resemble canonical plant immune receptors (*9*).

In the present study, genes from the *rhg1-b* interval (*6*) were silenced to test for impacts on SCN resistance (*10*). Transgenic soybean roots expressing artificial microRNA (amiRNA) or hairpin (RNA interference) constructs were produced using *Agrobacterium rhizogenes* (*11*–*13*). Soybean resistance to SCN was measured 2 weeks after root inoculation by determining the proportion of the total nematode population that had advanced past the J2 stage in each root (Fig. 1A) relative to known resistant and susceptible controls (*14*). Silencing any one of three closely linked genes at the *rhg1-b* locus of the SCN-resistant soybean variety Fayette significantly reduced SCN resistance (Fig. 1B). Depletion of resistance was dependent on target transcript reduction (fig. S1). Silencing other genes in and around the locus did not affect SCN resistance (e.g., Fig. 1B, genes *Glyma18g02570* and *-2620*) (*10*). The three *Rhg1* genes that were found to contribute to SCN resistance encode a predicted amino acid transporter (*Glyma18g02580*), an α-SNAP protein predicted to participate in disassembly of SNARE membrane trafficking complexes (*Glyma18g02590*), and a protein with a WI12 (wound-inducible protein 12) region but no functionally characterized domains (*Glyma18g02610*) (*15*–*17*).

Concurrent study of the physical structure of the *rhg1-b* locus revealed an unusual genomic configuration. A 31.2-kb genome segment encoding the above-noted genes is present in multiple copies in SCN resistant lines (Figs. 2 and 3). The DNA sequence of fosmid clone inserts carrying genomic DNA from the *rhg1-b* genetic interval identified a unique DNA junction, not present in the published Williams 82 soybean genome, in which the intergenic sequence downstream of (centromeric to) *Glyma18g02610* is immediately adjacent to a 3′ fragment of *Glyma18g02570* (Fig. 2A, fosmids 3, 4, and 5). The genomic repeat contains full copies of *Glyma18g02580*, *-2590*, *-2600*, and *-2610*, as well as the final two exons of *Glyma18g02570*. Whole-genome shotgun sequencing of a line containing *rhg1-b* revealed greater depth of coverage of this interval by a factor of 10 relative to the surrounding chromosomal region or homeologous regions on other chromosomes (Fig. 2B), suggesting the presence of multiple repeats. Further polymerase chain

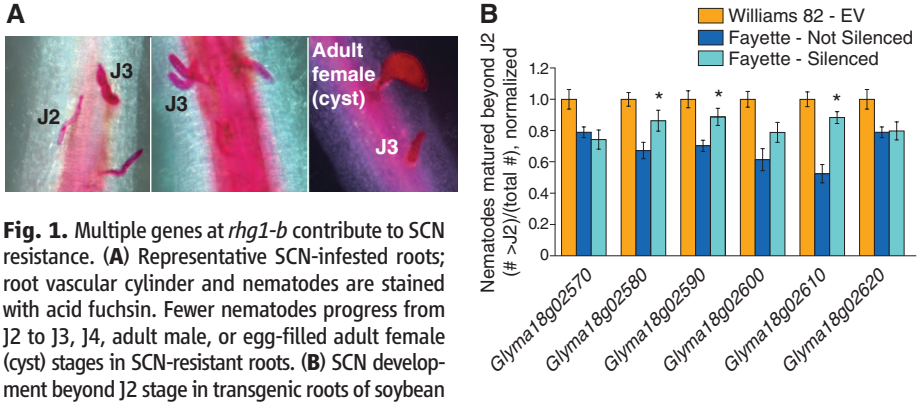

Fig. 1. Multiple genes at *rhg1-b* contribute to SCN resistance. (**A**) Representative SCN-infested roots; root vascular cylinder and nematodes are stained with acid fuchsin. Fewer nematodes progress from J2 to J3, J4, adult male, or egg-filled adult female (cyst) stages in SCN-resistant roots. (**B**) SCN development beyond J2 stage in transgenic roots of soybean variety Fayette with the designated gene silenced, relative to Williams 82 (SCN-susceptible) and nonsilenced Fayette (SCN-resistant) controls. Data are means ± SE. *$P < 0.05$, Fayette (silenced) significantly different from Fayette (not silenced) based on analysis of variance (ANOVA) Tukey test; $P > 0.1$ for *Glyma18g02600*. EV, transformed with empty vector.

[1]Department of Plant Pathology, University of Wisconsin, Madison, WI 53706, USA. [2]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [3]Department of Horticulture, University of Wisconsin, Madison, WI 53706, USA. [4]USDA-ARS and Department of Plant Pathology, University of Wisconsin, Madison, WI 53706, USA. [5]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588, USA. [6]Institute of Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

*These authors contributed equally to this work.
†Present address: Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA.
‡Present address: Dipartimento di Agraria, Università di Sassari, 07100 Sassari, Italy.
§Present address: Department of Agronomy, University of Florida, Gainesville, FL 32610, USA.
||Present address: USDA-ARS and Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA.
¶These authors contributed equally to this work.
#To whom correspondence should be addressed. E-mail: afbent@wisc.edu

reaction (PCR) and sequencing tests confirmed the presence of the *Glyma18g02610-2570* junction in DNA from multiple SCN-resistant soybean accessions, whereas the junction was not detected in four tested SCN-susceptible varieties including Williams 82 (Fig. 2C and fig. S2). The shared identity of the junction sites in disparate sources of SCN resistance suggests a shared origin of the initial resistance-conferring event at *Rhg1*.

Gene expression analysis using quantitative PCR (qPCR) determined that the three genes found to significantly affect SCN resistance show higher transcript levels in roots of SCN-resistant varieties than in susceptible lines (Fig. 2D and fig. S2). This suggested that elevated expression of one or more of the SCN-affecting genes could be a primary cause of elevated SCN resistance. Full-length transcripts were confirmed for *Glyma18g02580*, *-2590*, and *-2610*, no transcript was detected for *Glyma18g02600*, and no hybrid repeat-junction transcript was detected for *Glyma18g02570* (fig. S2).

Fiber-FISH (fluorescence in situ hybridization) was used to directly view the arrangement and copy number of the 31-kb repeat segment in different haplotypes of the *Rhg1* locus. The hybridization pattern and DNA fiber length estimates (Fig. 3 and table S1) indicate a single copy of the repeat in Williams 82, as in the reference soybean genome sequence (*7*). In Fayette, fiber-FISH revealed 10 copies of the repeat segment per DNA fiber, in the same configuration throughout the multiple nuclei sampled, in a pattern consistent with 10 direct repeats abutting head-to-tail (Fig. 3 and table S1). In samples from Peking, another common source of SCN resistance, three copies per DNA fiber were present in direct repeat orientation (Fig. 3). No additional copies (e.g., at other loci) were evident. *Rhg1* repeat copy number expansion is likely to have occurred by unequal-exchange meiotic recombination events between homologous repeats.

Amino acid polymorphism or overexpression of any one of the three identified *rhg1-b* genes did not account for SCN resistance. From all available *rhg1-b* sequence reads (across multiple repeat copies), no predicted amino acid polymorphisms relative to Williams 82 were identified for *Glyma18g02580*, *Glyma18g02600*, or *Glyma18g02610*. Some copies of *Glyma18g02590* from *rhg1-b* resembled the Williams 82 sequence, whereas others contained a set of polymorphisms, notably at the predicted C-terminal six amino acids of the predicted α-SNAP protein (table S2, confirmed by cDNA sequencing). However, expression of this non–Williams 82–type *Glyma18g02590* downstream of a strong constitutive promoter or native promoter sequence did not increase the SCN resistance reaction of Williams 82 transgenic roots (Fig. 4 and fig. S3), which suggests that *rhg1-b* SCN resistance requires more than this 2590–amino acid polymorphism. Overexpression of *Glyma18g02580* or *Glyma18g02610* also failed to increase SCN resistance (Fig. 4).

Given the above, simultaneous overexpression of genes within the 31-kb repeat segment was tested as a possible source of SCN resistance. In two separate experiments that together tested >25 independent transgenic events for each DNA construct, resistance to SCN was significantly increased in SCN-susceptible Williams 82 by simultaneous overexpression of the set of genes (Fig. 4; see also fig. S4A). A DNA construct overexpressing *Glyma18g02580*, *-2590*, and *-2610* (but not *-2600*) also conferred enhanced SCN resistance (fig. S4B). The collective findings indicate that *Rhg1*-mediated SCN resistance is attributable to elevated expression of *Glyma18g02580*, *-2590*, and *-2610*.

These results reveal a novel mechanism for disease resistance: an expression polymorphism

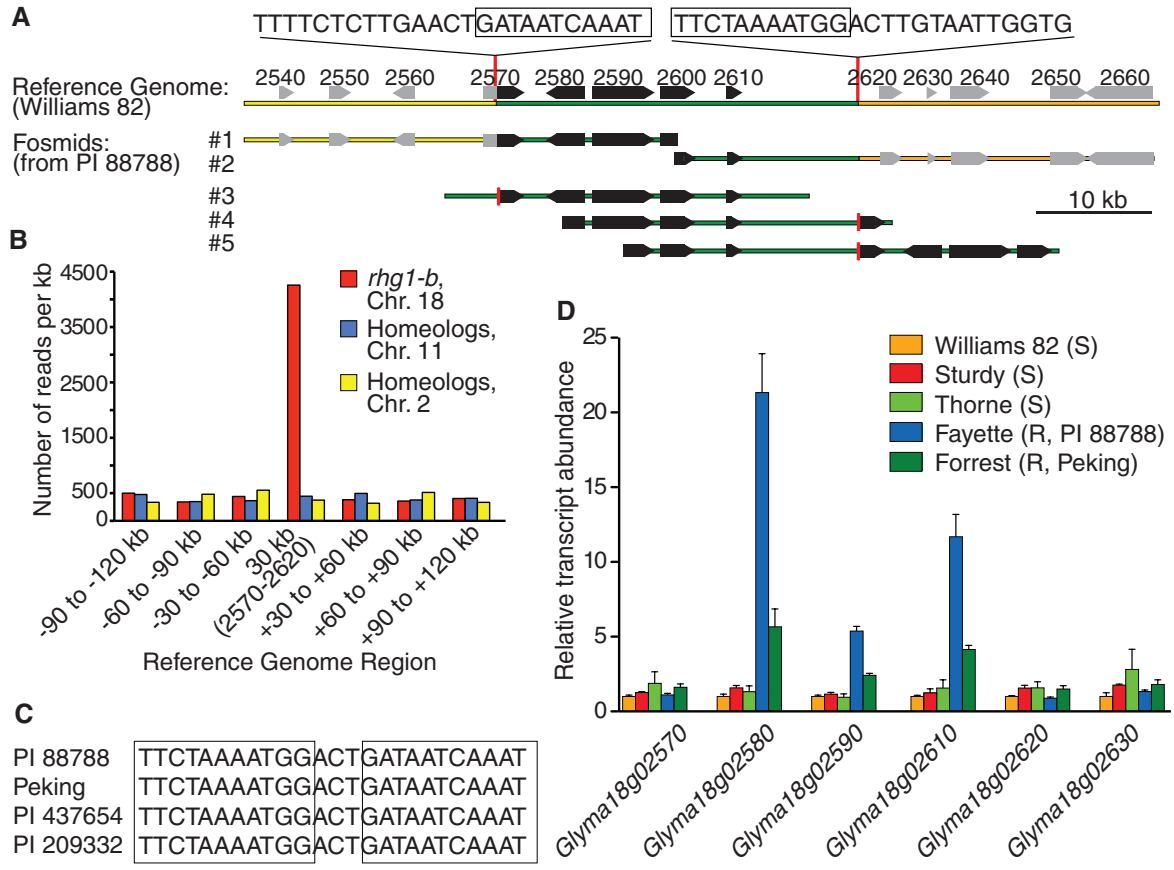

Fig. 2. A 31.2-kb repeat that elevates expression of the encoded genes is present in SCN-resistant haplotypes of the *Rhg1* locus. (**A**) Diagram of *Rhg1* locus of Williams 82 (top) and five fosmid inserts from *rhg1-b* haplotype. Numbers and block icons refer to soybean genes (e.g., *Glyma18g02540*). Fosmids 3, 4, and 5 carry *rhg1-b* genome segments that span repeat junctions. (**B**) Number of whole-genome shotgun sequencing reads corresponding to *Rhg1* [green region shown in (A)] was greater by a factor of 10 than for adjacent genomic regions on chromosome 18 or for homeologous regions. (**C**) Sequence of unique *Rhg1* repeat junction not found in reference genome, from four different sources of SCN resistance. (**D**) Transcript abundance of genes encoded in 31-kb repeat region is greater in roots from SCN-resistant varieties than from SCN-susceptible varieties. Means ± SE shown for qPCR; results for *Glyma18g02600* were at limit of detection.

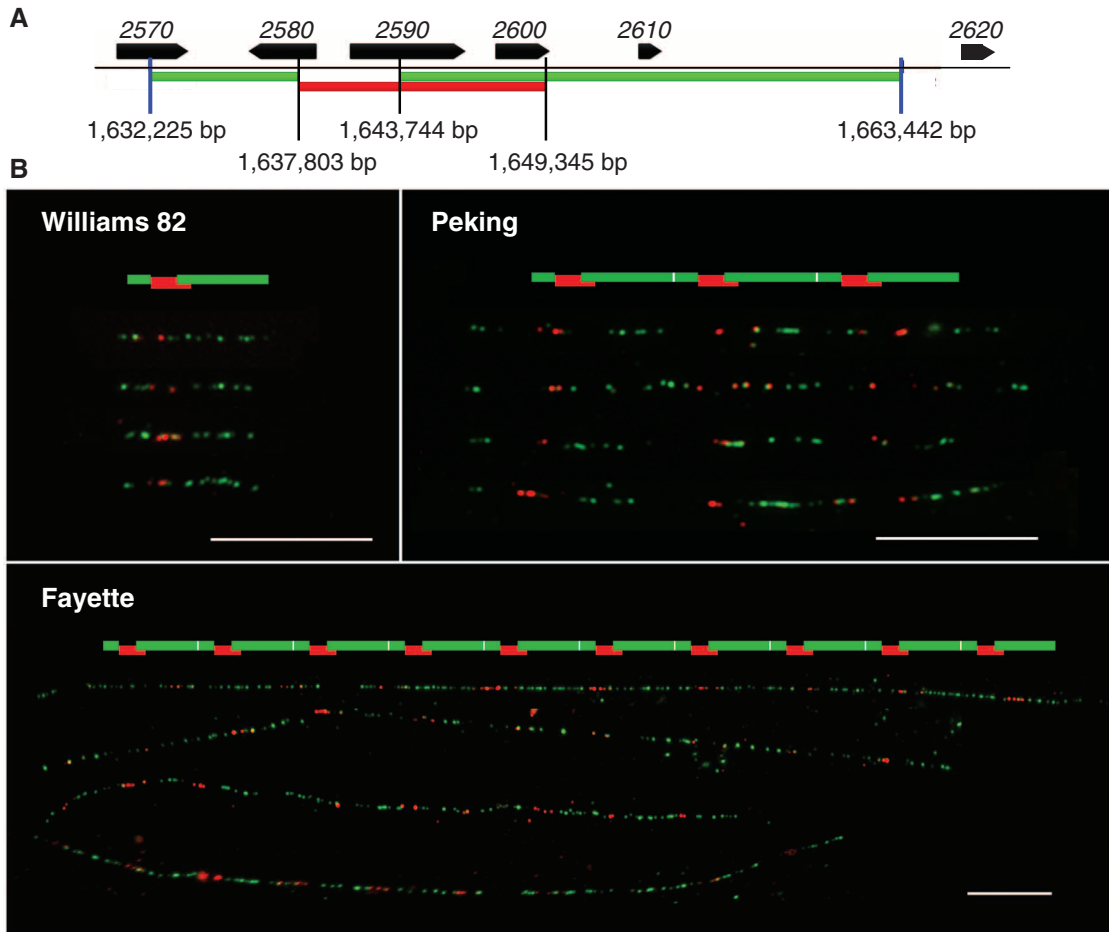

Fig. 3. Fiber-FISH detection of *Rhg1* copy number variation in widely used soybean lines. (**A**) Two adjacent probes were isolated from a single PI88788 (*rhg1-b*) genomic DNA fosmid clone whose insert spans a repeat junction, generating the 25.2-kb (green label) and 9.7-kb (red label) probes that correspond to the Williams 82 chromosome 18 sequence as shown. (**B**) Probe diagram and composite of four Fiber-FISH images (four DNA fibers) per genotype, revealing 10 or 3 direct repeat copies of the 31-kb *Rhg1* segment in SCN-resistant Fayette and Peking and one copy per *Rhg1* haplotype in SCN-susceptible Williams 82. White bars = 10 μm, which correspond to approximately 32 kb using a 3.21 kb/μm conversion rate (*32*).

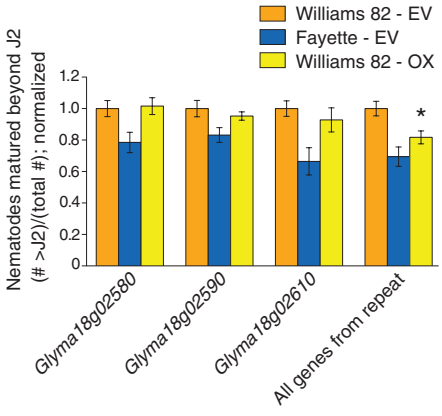

Fig. 4. Elevated SCN resistance conferred by simultaneous overexpression of multiple genes rather than overexpression of individual genes from the 31-kb *rhg1-b* repeat. SCN development beyond J2 stage is reported for transgenic soybean roots (variety Williams 82) overexpressing the designated single genes, or overexpressing all genes encoded within the 31-kb repeat (*Glyma18g02580*, *-2590*, *-2600*, and *-2610*), relative to Williams 82 (SCN-susceptible) and Fayette (SCN-resistant) controls. Data are means ± SE for roots transformed with empty vector (EV) or gene overexpression constructs (OX). *$P < 0.05$, Williams 82 with gene construct significantly different from Williams 82 EV based on ANOVA Tukey test.

for multiple disparate but tightly linked genes, derived through copy number variation at the *Rhg1* locus. This suggests future approaches to enhance *Rhg1*-mediated quantitative resistance against the globally important SCN disease of soybean—for example, through isolation of soybean lines that carry more copies of the 31-kb *Rhg1* repeat. Transgenic overexpression of the native or altered genes may improve SCN resistance and/or be applicable in other species for resistance to other endoparasitic nematodes.

The biochemical mechanisms of *Rhg1*-mediated resistance remain unknown. Other sequenced plant genomes do not carry close homologs of the predicted Glyma18g02610 protein, although a wound-inducible protein in ice plant with 55% identity has been studied (*17*). Modeling of the Glyma18g02610 predicted tertiary structure suggested (*10*) that Glyma18g02610 may participate in the production of phenazine-like compounds that are toxic to nematodes. The Glyma18g02590 α-SNAP protein is likely involved in vesicle trafficking and may influence exocytosis of products that alter feeding site development or nematode physiology (*18*). Because it is one of at least five α-SNAP homologs encoded in the soybean reference genome, Glyma18g02590 may have undergone subfunctionalization or neofunctionalization (*19*). The Glyma18g02580 protein and its most closely related plant transporter proteins are not functionally well characterized, but Glyma18g02580 contains a predicted tryptophan/tyrosine permease family domain. Tryptophan shares structural similarity with and is a precursor of the auxin hormone indole-3-acetic acid, which suggests the intriguing possibility that Glyma18g02580 may affect functionally important auxin levels or distribution (*2*, *3*). Together, these genes create an unfavorable environment at nematode feeding sites.

Growing evidence from metazoa and plants suggests that genome structural variation is a frequent and powerful driver of phenotypic diversity (*20*, *21*). Copy number variation of chromosomal subsegments (beyond simple duplication) can affect gene expression levels (*22*), and single-gene copy number variation contributes to a number of adaptive traits in humans, plants, and insects (*23*–*26*). Recent analyses of genome architecture in sorghum, rice, and soybean have reported high levels of copy number variation and a tendency for these genomic regions to overlap with postulated biotic and abiotic stress-related genes (*27*–*29*).

Our work provides a concrete example of copy number variation in which the repeat encodes multiple gene products that contribute to a valuable disease resistance trait. Single-copy clusters of genes that are functionally related but nonhomologous are highly unusual in multicellular

eukaryotes, but such clusters have been reported in association with plant secondary metabolism (*30*, *31*). Given the repetitive and plastic nature of plant genomes and the relatively underexplored association between copy number variation and phenotypes, it seems likely that a number of other complex traits are controlled by this type of structural variation.

References and Notes

1. T. L. Niblack, K. N. Lambert, G. L. Tylka, *Annu. Rev. Phytopathol.* **44**, 283 (2006).
2. E. L. Davis, R. S. Hussey, M. G. Mitchum, T. J. Baum, *Curr. Opin. Plant Biol.* **11**, 360 (2008).
3. G. Gheysen, M. G. Mitchum, *Curr. Opin. Plant Biol.* **14**, 415 (2011).
4. M. Kim, D. L. Hyten, T. L. Niblack, B. W. Diers, *Crop Sci.* **51**, 934 (2011).
5. V. C. Concibido, B. W. Diers, P. R. Arelli, *Crop Sci.* **44**, 1121 (2004).
6. M.-S. Kim, D. L. Hyten, A. F. Bent, B. W. Diers, *Plant Genome* **3**, 81 (2010).
7. J. Schmutz *et al.*, *Nature* **465**, 120 (2010).
8. P. D. Matsye *et al.*, *Plant Mol. Biol.* **80**, 131 (2012).
9. P. N. Dodds, J. P. Rathjen, *Nat. Rev. Genet.* **11**, 539 (2010).
10. See supplementary materials on *Science* Online.
11. N. A. Smith *et al.*, *Nature* **407**, 319 (2000).
12. S. Ossowski, R. Schwab, D. Weigel, *Plant J.* **53**, 674 (2008).
13. R. Narayanan, R. Atz, R. Denny, N. Young, D. Somers, *Crop Sci.* **39**, 1680 (1999).
14. S. Melito *et al.*, *BMC Plant Biol.* **10**, 104 (2010).
15. S. Okumoto, G. Pilot, *Mol. Plant* **4**, 453 (2011).
16. R. Jahn, R. H. Scheller, *Nat. Rev. Mol. Cell Biol.* **7**, 631 (2006).
17. S. Yen, P. Chen, H. C. E. Yen, *Plant Physiol.* **119**, 1147 (1999).
18. N. Frei dit Frey, S. Robatzek, *Curr. Opin. Plant Biol.* **12**, 437 (2009).
19. B. Conrad, S. E. Antonarakis, *Annu. Rev. Genomics Hum. Genet.* **8**, 17 (2007).
20. J. Sebat *et al.*, *Science* **305**, 525 (2004).
21. N. M. Springer *et al.*, *PLoS Genet.* **5**, e1000734 (2009).
22. B. E. Stranger *et al.*, *Science* **315**, 848 (2007).
23. G. H. Perry *et al.*, *Nat. Genet.* **39**, 1256 (2007).
24. L. U. Wingen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7115 (2012).
25. S. Pearce *et al.*, *Plant Physiol.* **157**, 1820 (2011).
26. J. M. Schmidt *et al.*, *PLoS Genet.* **6**, e1000998 (2010).
27. L. Y. Zheng *et al.*, *Genome Biol.* 10.1186/gb-2011-12-11-r114 (2011).
28. P. Yu *et al.*, *BMC Genomics* **12**, 8 (2011).
29. L. K. McHale *et al.*, *Plant Physiol.* **159**, 1295 (2012).
30. M. Frey *et al.*, *Science* **277**, 696 (1997).
31. D. DellaPenna, S. E. O'Connor, *Science* **336**, 1648 (2012).
32. Z. Cheng, C. R. Buell, R. A. Wing, J. Jiang, *Chromosome Res.* **10**, 379 (2002).

Acknowledgments: Supported by the United Soybean Board (A.F.B.), USDA CSREES (M.E.H.), Illinois Soybean Association (M.E.H. and B.W.D.), NSF grant DBI-0922703 (J.J.), Wisconsin Experiment Station Hatch Award (A.F.B.), and the Pioneer Fellowship in Plant Pathology awarded to D.E.C. by the American Phytopathological Society through a gift from Pioneer Hi-Bred. We thank A. E. MacGuidwin for suggestions regarding SCN experiments and J. M. Palmer for assistance with RNA blots. DNA sequence data are available at NCBI GenBank under accession codes JX907804 to JX907808 and at the NCBI Short Read Archive, study no. SRA059285. A.F.B., M.E.H., B.W.D., D.E.C., T.G.L., X.G., S.M., T.J.H., and J.W. are listed as inventors on a provisional patent application filed by Wisconsin Alumni Research Foundation that covers the findings reported herein.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1228746/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S3
References (*33*–*48*)

13 August 2012; accepted 26 September 2012
Published online 11 October 2012;
10.1126/science.1228746

An Exon Splice Enhancer Primes IGF2:IGF2R Binding Site Structure and Function Evolution

Christopher Williams,[1]* Hans-Jürgen Hoppe,[2]* Dellel Rezgui,[2] Madeleine Strickland,[1] Briony E. Forbes,[3] Frank Grutzner,[3] Susana Frago,[2] Rosamund Z. Ellis,[1] Pakorn Wattana-Amorn,[1] Stuart N. Prince,[2] Oliver J. Zaccheo,[2] Catherine M. Nolan,[4] Andrew J. Mungall,[5] E. Yvonne Jones,[6] Matthew P. Crump,[1]† A. Bassim Hassan[2]†

Placental development and genomic imprinting coevolved with parental conflict over resource distribution to mammalian offspring. The imprinted genes *IGF2* and *IGF2R* code for the growth promoter insulin-like growth factor 2 (IGF2) and its inhibitor, mannose 6-phosphate (M6P)/IGF2 receptor (IGF2R), respectively. M6P/IGF2R of birds and fish do not recognize IGF2. In monotremes, which lack imprinting, IGF2 specifically bound M6P/IGF2R via a hydrophobic CD loop. We show that the DNA coding the CD loop in monotremes functions as an exon splice enhancer (ESE) and that structural evolution of binding site loops (AB, HI, FG) improved therian IGF2 affinity. We propose that ESE evolution led to the fortuitous acquisition of IGF2 binding by M6P/IGF2R that drew *IGF2R* into parental conflict; subsequent imprinting may then have accelerated affinity maturation.

The sequence of molecular evolutionary events that established placental viviparity, genomic imprinting, and parental conflict in mammals remain poorly understood (*1*). Genomic imprinting occurs when expression of one allele of a diploid gene is silenced depending on the parent of origin, either from the father or the mother. Parental conflict over the distribution of resources to offspring has been supported by the observation of reciprocal imprinting of genes coding for the growth promoter insulin-like growth factor 2 (IGF2) and the cation-independent mannose 6-phosphate/IGF2 receptor (M6P/IGF2R or IGF2R) (*2*). *IGF2* and *IGF2R* are 2 of the ~80 genes imprinted in mammals and two of the five genes (with *INS*, *MEST*/*PEG1*, and *PEG10*) imprinted in marsupials. So far, no evidence supports the existence of imprinting in monotremes despite the presence of a yolk sac placenta (*3*, *4*). On the basis of functional data, IGF2R transports M6P-modified acid hydrolases to pre-lysosomes (*5*). Of the 15 extracellular domains of IGF2R, domain 11 binds IGF2 in therians and internalizes the ligand for degradation, whereas M6P ligands bind to domains 3, 5, and 9 (*5*). *Igf2* rescues placental-dependent embryonic lethality associated with laboratory-created murine parthenogenesis, implicating IGF2 supply as a regulator of placental development (*6*). Disruption of the maternal *Igf2r* allele results in *Igf2*-dependent overgrowth and fatality, supporting that IGF2R antagonizes the function of IGF2 (*7*, *8*). The structure of the unbound human domain 11 shows that the IGF2 binding site is composed of defined loops (AB, CD, FG, and HI; Fig. 1A and fig. S1) but how this domain 11 evolved to bind IGF2 and the relation to imprinting coevolution remains unknown (*9*–*12*).

We established a high-resolution structure of the human IGF2R:IGF2 complex and then compared this to other phylogenetically informative vertebrates. We adopted a nuclear magnetic resonance (NMR) approach because the side chain amino acid interactions across the binding interface were not resolved in our 4.1 Å resolution cocrystal structures (*9*). Wild-type (WT) human domain 11 and IGF2 failed to form a stable association in initial NMR studies. However, we identified an AB loop mutant (domain $11^{E1544K,\ K1545S,\ L1547V\ or\ clone\ E4}$) with an increased affinity for IGF2 that formed a tight complex [binding affinity (K_D) = 15 nM versus 46 to 64 nM for WT domain 11; Table 1 and ^{1}H-^{15}N correlation spectra, figs. S2 and S3] (*12*). We solved the solution structure of this 24.2-kD complex (IGF2: domain 11^{E4}) with NMR structural and quality statistics in table S1. When free IGF2 (Fig. 1A) binds to the single domain 11^{E4}

[1]Department of Organic and Biological Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. [2]Cancer Research UK Tumour Growth Control Group, Oxford Molecular Pathology Institute, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK. [3]School of Molecular and Biomedical Science, University of Adelaide, Adelaide 5005, Australia. [4]School of Biology and Environmental Science, University College Dublin, Belfield, Dublin, Ireland. [5]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, Canada. [6]Cancer Research UK Receptor Structure Research Group, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

*These authors contributed equally to this work.
†These authors contributed equally to this work. To whom correspondence should be addressed. E-mail: bass.hassan@path.ox.ac.uk (A.B.H.); matt.crump@bristol.ac.uk (M.P.C.)