Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee

Ricardo Acuña^{a,1}, Beatriz E. Padilla^{a,1}, Claudia P. Flórez-Ramos^a, José D. Rubio^a, Juan C. Herrera^a, Pablo Benavides^a, Sang-Jik Lee^{b,c}, Trevor H. Yeats^b, Ashley N. Egan^{b,d}, Jeffrey J. Doyle^b, and Jocelyn K. C. Rose^{b,2}

^aPlant Breeding, Biotechnology, and Entomology Departments, Cenicafe, A.A. 2427 Manizales, Colombia; ^bDepartment of Plant Biology, Cornell University, Ithaca, NY 14853; ^cBiotechnology Institute, Nongwoo Bio Co., Ltd., Gyeonoggi 469-885, Korea; and ^dDepartment of Biology, East Carolina University, Greenville, NC 27858

Edited by Nancy A. Moran, Yale University, West Haven, CT, and approved February 1, 2012 (received for review December 21, 2011)

Horizontal gene transfer (HGT) involves the nonsexual transmission of genetic material across species boundaries. Although often detected in prokaryotes, examples of HGT involving animals are relatively rare, and any evolutionary advantage conferred to the recipient is typically obscure. We identified a gene (*HhMAN1*) from the coffee berry borer beetle, *Hypothenemus hampei*, a devastating pest of coffee, which shows clear evidence of HGT from bacteria. *HhMAN1* encodes a mannanase, representing a class of glycosyl hydrolases that has not previously been reported in insects. Recombinant *HhMAN1* protein hydrolyzes coffee berry galactomannan, the major storage polysaccharide in this species and the presumed food of *H. hampei*. *HhMAN1* was found to be widespread in a broad biogeographic survey of *H. hampei* accessions, indicating that the HGT event occurred before radiation of the insect from West Africa to Asia and South America. However, the gene was not detected in the closely related species *H. obscurus* (the tropical nut borer or “false berry borer”), which does not colonize coffee beans. Thus, HGT of *HhMAN1* from bacteria represents a likely adaptation to a specific ecological niche and may have been promoted by intensive agricultural practices.

lateral gene transfer | transposable element | transposon

Horizontal gene transfer (HGT), or lateral gene transfer, the process whereby genes move across species boundaries by nonsexual means, is accepted as a significant evolutionary force among prokaryotes, and there are many examples of its importance as a mechanism of ecological adaptation in this domain of life (1, 2). In contrast, other than ancient massive HGT associated with the acquisition of endosymbiotic organelles, there are far fewer reported examples of HGT involving eukaryotes, either between eukaryotic species or from prokaryote to eukaryote (3). Examples of HGT involving animals are even rarer, presumably because the germline is isolated from the somatic cells, and so is less likely to come into contact with foreign DNA (4). Indeed, most of the documented cases have been limited to invertebrates as the recipient and involve endosymbionts, where the association between the bacteria and host cells is particularly intimate and can occur in germ cells (5–8). However, there are reported examples of HGT where the gene donor is not an endosymbiont (3), such as the transfer of genes allowing the synthesis of carotenoids from fungi to aphids (9) and the spider mite *Tetranychus urticae* (10), and large scale transfer of genes to bdelloid rotifers that appear to be derived from bacteria, fungi, and plants (11). Additionally, some cases of close symbiotic associations, such as that seen in coral, do not show evidence of HGT (12), underling its rarity.

There is considerable debate as to whether the apparent relative infrequency of HGT to animals is a reflection of fundamental differences in biology, or whether such events are being overlooked in animals, in part, because of the strategies used for genome sequencing, data filtering, and gene annotation (3, 13). Moreover, there has been considerable controversy regarding the lack of functional validation of genes that are putatively

assigned as being present following HGT, and the underlying evolutionary pressures and biological significance of such genes has often been unclear (14). In prokaryotes, HGT has repeatedly been associated with adaptation to new environments (2), and this trend is also beginning to emerge in eukaryotes. For example, the adoption of a plant–parasite lifestyle in several genera of nematodes was enabled by multiple HGT events (15), and a similar case of HGT facilitating plant parasitism was suggested by genome-scale analyses of oomycetes of the genus *Phytophthora*, which revealed that a substantial portion of the secreted proteomes was a consequence of HGT from fungi (16).

In this study, we obtained clear evidence for functionally important HGT involving an animal through studies of *Hypothenemus hampei*, commonly known as the coffee berry borer. This beetle is a specific pest of coffee (*Coffea* spp.) and is a major threat to coffee production worldwide, causing losses estimated at >500 million US dollars per year and affecting >20 million rural farming families (17). We describe the identification of a mannanase gene in the *H. hampei* genome that is likely to have originated in gut bacteria, as well as features of the associated genomic region that suggest a potential mechanism of gene transfer to the insect. There is a growing awareness that the metabolism of bacterial symbionts can confer important evolutionary advantages to insects (18), whether as endo- or ectosymbionts. This current study further suggests that intestinal bacteria may also provide a source of genetic material that allows rapid adaptations to new ecological niches, in this case, in the context of intensive agricultural production.

Results

Identification of *HhMAN1*. The coffee berry borer (Fig. 1A) reproduces and feeds exclusively inside coffee berries (Fig. 1B), which are primarily composed of carbohydrates (60%), proteins (10%), lipids (12%), organic acids, and caffeine (1%) (19). The feeding habits of *H. hampei* larvae suggest that they possess a hydrolytic system that is capable of enzymatically digesting the polysaccharides of coffee seeds (commonly referred to as coffee beans), which, in turn, are absorbed by the insect midgut during feeding. However, this process is poorly understood. To identify proteins that are secreted into the *H. hampei* midgut that may be involved in coffee berry infestation and digestion, we conducted a secretome functional screen using a yeast-secretion trap (20) with cDNA derived from *H. hampei* midgut. Among the cDNAs

Author contributions: R.A., P.B., S.-J.L., A.N.E., J.J.D., and J.K.C.R. designed research; R.A., B.E.P., C.P.F.-R., J.D.R., J.C.H., P.B., S.-J.L., and A.N.E. performed research; R.A., B.E.P., C.P.F.-R., J.D.R., J.C.H., P.B., S.-J.L., T.H.Y., A.N.E., J.J.D., and J.K.C.R. analyzed data; and R.A., J.C.H., S.-J.L., T.H.Y., A.N.E., J.J.D., and J.K.C.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. [GQ375156](#) and [GQ375157](#)).

¹R.A. and B.E.P. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: jr286@cornell.edu.

Fig. 1. The mannanase gene *HhMAN1* from the coffee berry borer (*Hypothenemus hampei*). (A) An adult *H. hampei* burrows into and lays its eggs in coffee berries (B). (C) Best maximum likelihood (ML) phylogenetic analysis of *HhMAN1* gene (red) with homologs from fungi (blue), bacteria (black), animals (purple), and plants (green). GenBank accession numbers are shown adjacent to each gene and bootstrap values are indicated on key branches. Where relevant, bootstrap values are noted with maximum likelihood followed by maximum parsimony (MP). Nodes with only one bootstrap value are from ML analyses only and were not supported in the MP bootstrap analysis.

recovered, we identified one (*HhMAN1*) that is predicted to encode a secreted protein with homology to mannanase enzymes of glycosyl hydrolase family 5 (Carbohydrate-Active enZymes Database; www.cazy.org). This gene was selected for further characterization not only because galactomannan is the major storage polysaccharide of coffee beans (21) but also because there was no precedent for an insect mannanase in sequence databases.

We subsequently determined the full-length sequence by 3' RACE and performed phylogenetic analyses of taxonomically diverse mannanase genes, including examples from both prokaryotes and eukaryotes (Fig. 1C). Whereas mannanases from plants, fungi, and animals formed a distinct eukaryotic clade, *HhMAN1* was most closely related to prokaryotic mannanases.

Phylogenetic analyses strongly supported *HhMAN1* as grouping with the *Bacillus* clade (Fig. 1C); trees in which *HhMAN1* was constrained to group with animal sequences had significantly lower likelihoods. This indicated that *HhMAN1* is an example of HGT and has a bacterial origin before transfer into the *H. hampei* genome.

***HhMAN1* Has Some Hallmarks of a Eukaryotic Gene and Is Integrated in the *H. hampei* Genome.** Despite showing strong overall sequence similarity to bacterial mannanase genes, sequence analysis of the *HhMAN1* cDNA and genomic region revealed that *HhMAN1* has adopted several features indicative of a eukaryotic context. A predicted polyadenylation signal (AATAAA) is present 34 bp downstream of the *HhMAN1* ORF, and the *HhMAN1* cDNA

Fig. 3. Detection of *HhMAN1* in genomic DNA from a geographically broad range of *H. hampei* accessions. The arrows indicate the diagnostic amplified band. A positive control (C+, *HhMAN1* cDNA plasmid) and a negative control (C–, empty vector) are shown. Molecular mass markers in base pairs (bp) are indicated.

Discussion

Although HGT in prokaryotes has long been established as a central mechanism of evolutionary adaptation, its significance in animals is often debated, and there are relatively few examples. Furthermore, whereas the mechanisms of HGT in prokaryotes are well known (2, 27), the mechanisms of eukaryotic HGT remain elusive, and the functional significance of HGT is rarely evident. Here, we report that *HhMAN1*, which encodes a secreted mannanase that is active against the primary seed storage polysaccharide of coffee, is the result of HGT between bacteria and an insect. This transfer likely confers an adaptive advantage because *H. hampei* uses the coffee berry as a specific host.

Phylogenetic analysis of *HhMAN1* clearly indicates that the sequence is more closely related to eubacterial than eukaryotic genes, and, previously, there have been no mannanase sequences identified in insects. Because symbiotic microorganisms can be the source of polysaccharide hydrolytic activity required for plant biomass degradation (28), it was important to exclude the possibility of *HhMAN1* actually being encoded by the midgut microbiota of *H. hampei*. PCR amplification of genomic DNA derived from multiple organs and developmental stages of *H. hampei* established that the sequence was ubiquitously present and, thus, not a contaminant of midgut-derived material. The *HhMAN1* cDNA and genomic region revealed that *HhMAN1* has adopted several features indicative of a eukaryotic context, perhaps the most intriguing of which is that the *HhMAN1* gene is flanked by two eukaryotic transposons of the hAT and Tc1/mariner superfamilies. Not only does this provide further evidence that *HhMAN1* occurs in the genome of *H. hampei* and not that of a symbiotic microbe, but it also suggests a potential mechanism for HGT in animals. Genome recombination, mobilization, and reorganization are associated with transposable elements and HGT of the mariner transposon between several

eukaryotic species has been reported previously (29), although, in this case, there was no observation of HGT of neighboring nontransposon genes. Such an association is suggested by our results, as well as a study of bdelloid rotifers that identified extensive examples of HGT in telomeric regions rich in retrotransposons and DNA transposons, including those of the hAT and Tc1/mariner superfamilies (11). We propose that transposon-mediated incorporation of foreign DNA could be one component of the poorly understood mechanism of HGT involving animals.

It has been suggested that HGT is particularly prevalent in the contexts of parasitic or pathogenic relationships or in the adaptation to specific environmental conditions (30). In support of this idea, recent examples involving nematodes (15), oomycetes (16), the spider mite (10), and the case we describe here of *HhMAN1* share a common feature in that they all involve the acquisition of secreted plant cell wall or other glycan-degrading enzymes by HGT that is associated with the adoption of a new plant host. It seems very likely that acquisition of a mannanase activity by *H. hampei* would confer a strong adaptive advantage to the insect because it adopted a lifestyle that is dependent on a host tissue that is rich in mannans.

Thus, *HhMAN1* appears to be a relatively recent case of HGT that would have been essential for adaptation to a new ecological niche by an important agricultural pest. Our results further suggest that intensive agriculture, and particularly monoculture, may provide a powerful selective pressure for HGT and that intestinal bacteria may provide a rich source of genes that can be acquired by eukaryotic hosts.

Materials and Methods

Insect Material. Tissue was dissected from larvae or adults, as indicated, and immediately placed in a chilled microcentrifuge tube with 100 μ L of 5 mM phenylmethylsulfonyl fluoride or 0.1% RNA Later reagent (Qiagen) for protein or RNA isolation, respectively, and stored at -80°C .

20. Lee SJ, et al. (2006) A functional screen to characterize the secretomes of eukaryotic pathogens and their hosts in planta. *Mol Plant Microbe Interact* 19: 1368–1377.
21. Bradbury AGW, Halliday J (1990) Chemical structures of green coffee bean polysaccharides. *J Agric Food Chem* 38:389–392.
22. Nakagawa S, Niimura Y, Miura K, Gojobori T (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc Natl Acad Sci USA* 107:6382–6387.
23. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
24. Villegas C (1998) Manejo de insectos y ácaros asociados a las estructuras reproductivas de la macadamia. *CENICAFÉ. Avances Técnicos* 250:1–8.
25. Waller JM, Bigger M, Hillocks RJ (2007) *Coffee Pests, Diseases and Their Management*, eds Waller JM, Bigger M, Hillocks RJ (CABI, Wallingford, UK), pp 325–335.
26. Benavides P, Vega FE, Romero-Severson J, Bustillo AE, Stuart JJ (2005) Biodiversity and biogeography of an important inland pest of coffee, coffee berry borer (Coleoptera: Curculionidae: Scolytinae). *Ann Entomol Soc Am* 98:359–366.
27. Zhaxybayeva O, Doolittle WF (2011) Lateral gene transfer. *Curr Biol* 21:R242–R246.
28. Watanabe H, Tokuda G (2010) Cellulolytic systems in insects. *Annu Rev Entomol* 55: 609–632.
29. Bartolomé C, Bello X, Maside X (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol* 10:R22.
30. Keeling PJ (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin Genet Dev* 19:613–619.
31. Krishnamurthy N, Brown D, Sjölander K (2007) FlowerPower: Clustering proteins into domain architecture classes for phylogenomic inference of protein function. *BMC Evol Biol* 7(Suppl 1):S12.
32. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
33. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
34. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–772.
35. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320.
36. Reeves JH (1992) Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. *J Mol Evol* 35:17–31.
37. Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* 10:1396–1401.
38. Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M (1994) Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: Instability of a tree based on a single gene. *J Mol Evol* 39:519–527.
39. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer Associates, Sunderland, MA), Version 4.
40. Stamatakis A (2006) RAXML-VI-HP: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
41. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116.
42. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254.
43. Bernfeld P (1955) Aminoacylases. α and β . *Methods Enzymol* 1:149–158.