



## The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization

Christina A. Cuomo, *et al.*  
*Science* **317**, 1400 (2007);  
DOI: 10.1126/science.1143708

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of September 6, 2007):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/317/5843/1400>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/317/5843/1400/DC1>

This article **cites 15 articles**, 5 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/317/5843/1400#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

coma overall cannot be distinguished from that resulting from random noise (fig. S1b), which suggests that POAG may be a more complex disease than XFG.

#### References and Notes

1. S. Resnikoff *et al.*, *Bull. World Health Org.* **82**, 844 (2004).
2. P. J. Foster, R. Buhrmann, H. A. Quigley, G. J. Johnson, *Br. J. Ophthalmol.* **86**, 238 (2002).
3. F. Jonasson *et al.*, *Eye* **17**, 747 (2003).
4. A. Ringvold, *Acta Ophthalmol. Scand.* **77**, 371 (1999).
5. S. M. Jeng *et al.*, *J. Glaucoma* **16**, 117 (2007).
6. U. Schlotzer-Schrehardt, G. O. Naumann, *Am. J. Ophthalmol.* **141**, 921 (2006).
7. A. W. Hewitt, J. E. Craig, D. A. Mackey, *Clin. Exp. Ophthalmol.* **34**, 472 (2006).
8. E. M. Stone *et al.*, *Science* **275**, 668 (1997).
9. T. Rezaie *et al.*, *Science* **295**, 1077 (2002).
10. S. Monemi *et al.*, *Hum. Mol. Genet.* **14**, 725 (2005).
11. Materials and methods are available as supporting material on Science Online.
12. B. Devlin, K. Roeder, *Biometrics* **55**, 997 (1999).
13. N. Mantel, W. Haenszel, *J. Natl. Cancer Inst.* **22**, 719 (1959).
14. X. Liu *et al.*, *Nat. Genet.* **36**, 178 (2004).
15. H. A. Lucero, H. M. Kagan, *Cell. Mol. Life Sci.* **63**, 2304 (2006).
16. L. Thomassin *et al.*, *J. Biol. Chem.* **280**, 42848 (2005).
17. R. Ritch, U. Schlotzer-Schrehardt, A. G. Konstas, *Prog. Retin. Eye Res.* **22**, 253 (2003).
18. R. P. Kirwan *et al.*, *Mol. Vis.* **11**, 798 (2005).
19. P. A. Netland, H. Ye, B. W. Streeten, M. R. Hernandez, *Ophthalmology* **102**, 878 (1995).
20. J. D. Pena *et al.*, *Exp. Eye Res.* **67**, 517 (1998).
21. J. Duan *et al.*, *Hum. Mol. Genet.* **12**, 205 (2003).
22. D. Wang, A. D. Johnson, A. C. Papp, D. L. Kroetz, W. Sadée, *Pharmacogenet. Genomics* **15**, 693 (2005).
23. H. Zhang *et al.*, *Hum. Mol. Genet.* **15**, 807 (2006).
24. We thank the participants whose contribution made this study possible. We also thank the nurses at Noatun (deCODE's sample recruitment center) and personnel at the deCODE core facilities for their hardwork and enthusiasm. C. W. was supported by the Swedish Research Council. The authors from deCODE genetics declare competing financial interests.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/1146554/DC1

Materials and Methods

Figs. S1 to S3

Tables S1 to S7

References

15 June 2007, accepted 18 July 2007

Published online 9 August 2007;

10.1126/science.1146554

Include this information when citing this paper.

# The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization

Christina A. Cuomo,<sup>1</sup> Ulrich Güldener,<sup>2,3</sup> Jin-Rong Xu,<sup>4</sup> Frances Trail,<sup>5</sup> B. Gillian Turgeon,<sup>6</sup> Antonio Di Pietro,<sup>7</sup> Jonathan D. Walton,<sup>5</sup> Li-Jun Ma,<sup>1</sup> Scott E. Baker,<sup>8</sup> Martijn Rep,<sup>9</sup> Gerhard Adam,<sup>10</sup> John Antoniw,<sup>11</sup> Thomas Baldwin,<sup>11</sup> Sarah Calvo,<sup>1</sup> Yueh-Long Chang,<sup>12</sup> David DeCaprio,<sup>1</sup> Liane R. Gale,<sup>12</sup> Sante Gnerre,<sup>1</sup> Rubella S. Goswami,<sup>12</sup> Kim Hammond-Kosack,<sup>11</sup> Linda J. Harris,<sup>13</sup> Karen Hilburn,<sup>14</sup> John C. Kennell,<sup>15</sup> Scott Kroken,<sup>16</sup> Jon K. Magnuson,<sup>8</sup> Gertrud Mannhaupt,<sup>3</sup> Evan Mauceli,<sup>1</sup> Hans-Werner Mewes,<sup>2,3</sup> Rudolf Mitterbauer,<sup>10</sup> Gary Muehlbauer,<sup>12</sup> Martin Münsterkötter,<sup>3</sup> David Nelson,<sup>17</sup> Kerry O'Donnell,<sup>18</sup> Thérèse Ouellet,<sup>13</sup> Weihong Qi,<sup>5</sup> Hadi Quesneville,<sup>19</sup> M. Isabel G. Roncero,<sup>7</sup> Kye-Yong Seong,<sup>12</sup> Igor V. Tetko,<sup>3,21</sup> Martin Urban,<sup>11</sup> Cees Waalwijk,<sup>20</sup> Todd J. Ward,<sup>18</sup> Jiqiang Yao,<sup>4</sup> Bruce W. Birren,<sup>1</sup> H. Corby Kistler<sup>12,14\*</sup>

We sequenced and annotated the genome of the filamentous fungus *Fusarium graminearum*, a major pathogen of cultivated cereals. Very few repetitive sequences were detected, and the process of repeat-induced point mutation, in which duplicated sequences are subject to extensive mutation, may partially account for the reduced repeat content and apparent low number of paralogous (ancestrally duplicated) genes. A second strain of *F. graminearum* contained more than 10,000 single-nucleotide polymorphisms, which were frequently located near telomeres and within other discrete chromosomal segments. Many highly polymorphic regions contained sets of genes implicated in plant-fungus interactions and were unusually divergent, with higher rates of recombination. These regions of genome innovation may result from selection due to interactions of *F. graminearum* with its plant hosts.

*Fusarium*, a genus of plant pathogenic fungi, causes diseases that affect most species of cultivated plants, including root and stem rots, blights, and wilts (1). *F. graminearum*, which causes Fusarium head blight (FHB) disease on wheat and barley, is a leading cause of economic loss in these crops (2). In addition to reducing seed mass and quality, the fungus contaminates grain with toxic metabolites that are a threat to human health (3). *Fusarium* species also can directly infect humans, causing localized necrotic diseases (4) and invasive infection, especially in immunocompromised individuals (5).

The *F. graminearum* genome was whole-genome shotgun sequenced by paired-end se-

quencing of plasmid, Fosmid, and bacterial artificial chromosome (BAC) clones. The resulting assembly totals 36.1 Mb and displays high sequence quality and continuity. Nearly all (99.8%) of the assembly was anchored to the four chromosomes by genetically mapping markers derived from the genome sequence (6), and an initial set of 11,640 genes was predicted (table S1 and SOM text). Functional categories for the predicted genes were inferred by the presence of conserved InterPro domains (7) and were compared with those found in genomes of the related fungi, *Neurospora crassa*, *Magnaporthe grisea* and *Aspergillus nidulans*. The *F. graminearum* genome has greater numbers of genes for several protein categories, including predicted transcrip-

tion factors, hydrolytic enzymes, and transmembrane transporters (Fig. 1 and table S2).

The *F. graminearum* genome has few high-identity duplicated sequences, fewer by at least a factor of 15 than other related fungi, including *Saccharomyces cerevisiae* (table S3 and SOM text). Only a few gene pairs originated from recent duplications (fig. S1 and table S4), and we identified only two small families of transposons (table S5 and SOM text). *F. graminearum* differs from other filamentous fungi because it is homothallic (self-fertile) and rarely out-crosses, which limits the opportunity to acquire new repeats (2). In some ascomycetous fungi, including *F. graminearum*, the lack of repetitive sequence is due to a genome-wide defense system known as repeat-induced point mutation (RIP) (8). RIP identifies duplicated sequences (9) and introduces C:G to T:A transition mutations in both copies during the sexual cycle; this mutational bias was observed in *F. graminearum* transposons (tables S6 and S7 and SOM text).

<sup>1</sup>Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA. <sup>2</sup>Technische Universität München, Freising-Weihenstephan, Germany. <sup>3</sup>Institute for Bioinformatics, GSF National Research Center for Environment and Health, Neuherberg, Germany. <sup>4</sup>Purdue University, West Lafayette, IN 47907, USA. <sup>5</sup>Michigan State University, East Lansing, MI 48824, USA. <sup>6</sup>Cornell University, Ithaca, NY 14853, USA. <sup>7</sup>Universidad de Córdoba, Córdoba, Spain. <sup>8</sup>Pacific Northwest National Laboratory, Richland, WA 99352, USA. <sup>9</sup>University of Amsterdam, Netherlands. <sup>10</sup>BOU, University of Natural Resources and Applied Life Sciences, Vienna, Austria. <sup>11</sup>Rothamsted Research, Harpenden, UK. <sup>12</sup>University of Minnesota, St. Paul, MN 55108, USA. <sup>13</sup>Agriculture and Agri-Food Canada and University of Ottawa, Ottawa, ON, Canada. <sup>14</sup>U.S. Department of Agriculture (USDA) Agricultural Research Service, Cereal Disease Laboratory, St. Paul, MN 55108, USA. <sup>15</sup>St. Louis University, St. Louis, MO 63103, USA. <sup>16</sup>University of Arizona, Tucson, AZ 85721, USA. <sup>17</sup>University of Tennessee, Memphis, TN 38163, USA. <sup>18</sup>USDA ARS, National Center for Agricultural Utilization Research, Peoria, IL 61604, USA. <sup>19</sup>Institut Jacques Monod, Paris, France. <sup>20</sup>Plant Research International, Wageningen, Netherlands. <sup>21</sup>Institute of Bioorganic Chemistry and Photochemistry, National Ukrainian Academy of Sciences, Kiev, Ukraine.

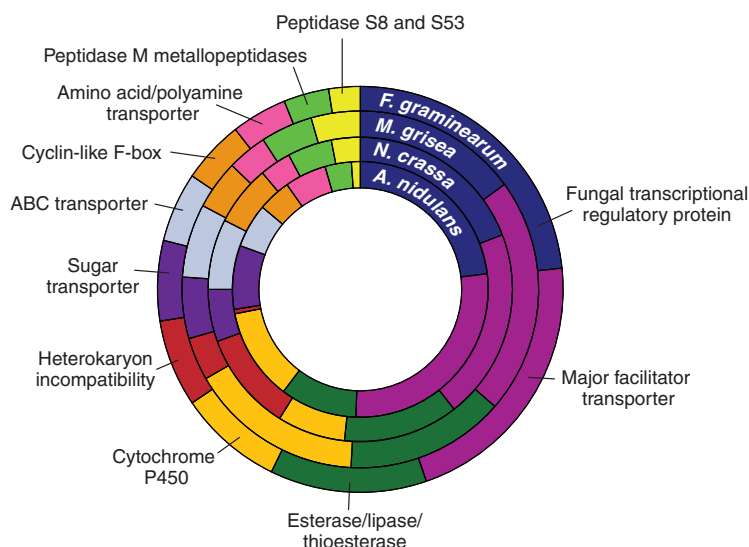
\*To whom correspondence should be addressed. E-mail: hckist@umn.edu

To experimentally confirm the activity of RIP, we examined the stability of transgenically introduced repeats of a hygromycin phosphotransferase (*hph*) gene during sexual and asexual development. Although cultures derived from asexual spores maintained drug resistance, 42% of cultures derived from ascospores, which are formed during the sexual cycle, were sensitive to hygromycin (table S8), and 99% of mutations

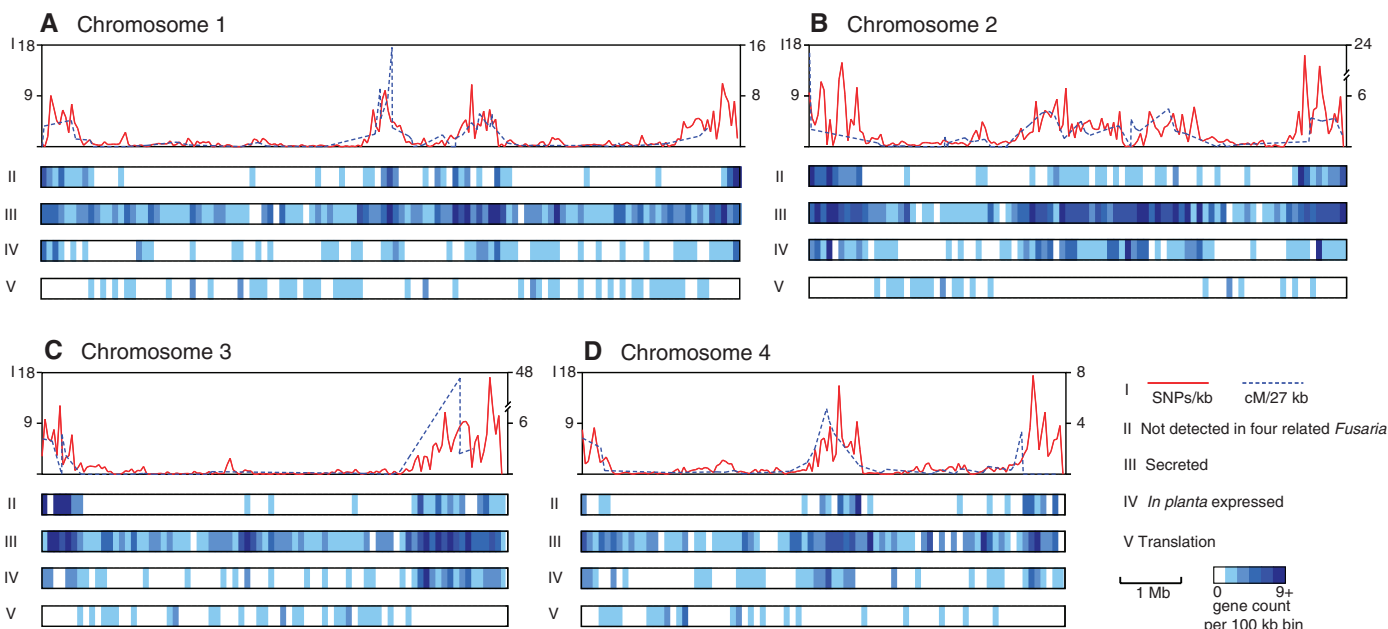
found in the *hph* gene were C to T point mutations, most at CpA sites (tables S9 and S10), indicating that repetitive sequences are frequently mutated by a RIP-like process during meiosis. Ascospores are important for FHB infection (2) and, being homothallic, *F. graminearum* may undergo meiosis and produce ascospores with greater frequency because it is unconstrained by the necessity for finding a compatible mate.

An increased opportunity for meiosis may impact the genome by increasing the frequency of RIP, allowing it to play a more central role.

We compared the assembly of strain PH-1 with ~0.4-fold coverage of whole-genome shotgun sequence from a second strain of *F. graminearum*, GZ3639. We identified 10,495 single-nucleotide polymorphisms (SNPs) between the two strains and mapped SNP positions along each chromosome. Because the GZ3639 data does not cover every base in the genome, SNP densities were normalized to the number of high-quality base alignments (7). For this normalized data set, SNP densities ranged from 0 to 17.5 SNPs per kb. We found that the distribution of SNPs is biased, because 25% of SNPs are found within 5% of the genome sequence, and 50% of SNPs are within 13% of the genome sequence. Regions exhibiting high SNP densities were clustered along each chromosome (Fig. 2). In particular, all telomere proximal regions displayed very high densities of SNPs and contained the vast majority of the highest SNP density windows. In addition to chromosome ends, three chromosomes were found to have one or two large interstitial regions of high SNP density. Whereas telomeres are well established in many species as sites of sequence variation and rearrangement (10–12), the presence of discrete interstitial regions of high diversity in addition to the subtelomeres is striking. One simple explanation is that these sites may reflect ancestral chromosome fusion. Although *F. graminearum* has four chromosomes, related fungi with similar



**Fig. 1.** Functional classification of *F. graminearum* proteins and comparison to other fungi. Displayed Interpro gene categories were found to have at least 50% greater gene abundance in *F. graminearum* than in *N. crassa*. Representative gene numbers for *M. grisea* and *A. nidulans* are also shown. Each circle displays the relative fraction of genes represented in each of the categories for each genome. For a full list of Interpro categories, see [mips.gsf.de/genre/proj/fusarium/Search/Catalogs/searchIprComp.html](http://mips.gsf.de/genre/proj/fusarium/Search/Catalogs/searchIprComp.html).



**Fig. 2.** Correlation of high SNP regions with recombination and gene sets. The four *F. graminearum* chromosomes are shown proportionally to size. Panels (A) to (D) represent chromosomes 1 to 4, respectively. (I) Distribution of SNPs (red lines) and recombination rate (blue dashed lines) for 50 kb non-overlapping windows across each chromosome (x axis). SNPs are plotted on the left y axis as the number of SNPs per kb of

high-quality aligned bases (7). Recombination rate [cM/27 kb (6)] is plotted on the right y axis. (II to V) Gene counts for the following gene sets (7) are shown with relative density shading for tiled 100-kb windows: (II) Genes found in *F. graminearum* but absent from four related *Fusarium* species. (III) Secreted proteins. (IV) Genes expressed specifically in planta. (V) Proteins involved in translation.

genome size, including *F. verticillioides*, *F. oxysporum*, and *F. solani*, have many more, ranging from 9 to >17. All closely related species in the *F. graminearum* species complex, as well as *F. culmorum*, also have four chromosomes, which indicates that if chromosome fusion occurred, it was not a recent event.

The regions of highest SNP density were significantly correlated with the regions of highest recombination ( $0.55, P = 1.2 \times 10^{-13}$ ), similar to correlations of SNP distribution or nucleotide diversity with recombination frequency observed in humans and *Drosophila* (13, 14). Additionally, regions of high SNP density have significantly lower G+C content than the rest of the genome ( $-0.43, P = 1.1 \times 10^{-8}$ ). The low G+C content of internal regions further supports the idea that these regions may represent ancestral telomeres.

To determine whether high diversity SNP regions evolved recently, we examined the sequence divergence of genes in these regions. We compared *F. graminearum* coding regions to those resulting from a low coverage (4X) assembly of *F. verticillioides* (7). Comparing the best matches for *F. graminearum* proteins from high and low SNP density regions (top and bottom quartiles) to the *F. verticillioides* assembly revealed that proteins from the highest SNP density regions have fewer putative orthologs compared with the rest of the genome and that these orthologs share lower identity (7). Although variation in the local mutation rate is expected to produce a correlation between polymorphism and divergence, more polymorphisms were found in high SNP regions than predicted on the basis of divergence (table S13), and the ratio of synonymous to nonsynonymous polymorphisms is higher than that of less diverse regions ( $\chi^2$  value =  $3.7 \times 10^{-7}$ ) (table S14).

Blast analysis (7) identified 704 genes as specific to *F. graminearum*, and these show significant enrichment in the high-density SNP regions ( $P = 4.5 \times 10^{-15}$ ). We also compared *F. graminearum* with the closely related *F. asiaticum*, *F. boothii*, *F. culmorum*, and *F. pseudograminearum* using genomic DNA hybridizations to a *F. graminearum* microarray (15) and identified 382 genes that are *F. graminearum* specific. These genes were overrepresented (by a factor of 2.7) in the high-density SNP regions ( $P = 3.4 \times 10^{-34}$ ). These data further demonstrate that genomic regions exhibiting the highest intraspecific variability also exhibit the highest interspecific variability.

*F. graminearum* genes specifically expressed during plant infection—including predicted secreted proteins, major facilitator transporters, amino acid transporters, and cytochrome P450s—are all overrepresented in high SNP density regions (Fig. 2, table S14, and SOM text). Conversely, genes predicted to be highly conserved, such as nuclear encoded mitochondrial genes or genes involved in translation, are underrepresented in regions of high diversity (table S15 and SOM text).

Comparison of gene expression of *F. graminearum* infection on barley and under varied nutritional culture conditions (7, 15) identified 408 genes as exclusively expressed during barley infection. These genes are highly enriched in the high-SNP-density regions ( $P = 7.4 \times 10^{-15}$ ), and 31% are predicted to be secreted, representing enrichment by a factor of 3 over the genome as a whole (table S14 and SOM text). Four of these genes have similarity to known virulence factors, and another 32 genes are predicted plant cell-wall degrading enzymes (table S16). Among these enzymes are xylanases, which degrade xylan, the major hemicellulose portion of monocot cell walls, pectate lyases, which cleave pectin, another essential component of plant cell walls and cutinases, enzymes that hydrolyze cutin polyesters that coat all outer plant surfaces. Such enzymes may function in the penetration and maceration of plant tissues and for the acquisition of nutrients from plant polymers (16) and may be involved as effector molecules that trigger host-plant defense responses (17). The high genetic diversity of this group of genes suggests that the fungus has a great capacity for adaptability and genetic change during its interaction with even this single host species.

The completed genome of *F. graminearum* allowed us to identify distinct regions of high diversity. We found that these regions are enriched for infection-related genes, which may allow the fungus to adapt rapidly to changing environments or hosts. Recognition of these high-diversity areas of the genome focuses the direction of future work toward those regions that may have the greatest potential in elucidating the dynamics of host pathogen interactions.

#### References and Notes

1. J. F. Leslie, B. A. Summerell, *The Fusarium Laboratory Manual* (Blackwell, Ames, Iowa, 2006).
2. R. S. Goswami, H. C. Kistler, *Mol. Plant Pathol.* **5**, 515 (2004).

3. J. W. Bennett, M. Kilch, *Clin. Microbiol. Rev.* **16**, 497 (2003).
4. D. C. Chang *et al.*, *JAMA* **296**, 953 (2006).
5. M. C. Dignani, E. Anaissie, *Clin. Microbiol. Infect.* **10**, (Suppl 1), 67 (2004).
6. L. R. Gale *et al.*, *Genetics* **171**, 985 (2005).
7. Materials and methods are available as supporting material on Science Online.
8. E. U. Selker, E. B. Cambareri, B. C. Jensen, K. R. Haack, *Cell* **51**, 741 (1987).
9. M. K. Watters, T. A. Randall, B. S. Margolin, E. U. Selker, D. R. Stadler, *Genetics* **153**, 705 (1999).
10. H. C. Mefford, B. J. Trask, *Nat. Rev. Genet.* **3**, 91 (2002).
11. M. J. Gardner *et al.*, *Nature* **419**, 498 (2002).
12. E. A. Winzeler *et al.*, *Genetics* **163**, 79 (2003).
13. D. J. Begun, C. F. Aquadro, *Nature* **356**, 519 (1992).
14. A. Kong *et al.*, *Nat. Genet.* **31**, 241 (2002).
15. U. Guldener *et al.*, *Fungal Genet. Biol.* **43**, 316 (2006).
16. P. S. Solomon, K.-C. Tan, R. P. Oliver, *Mol. Plant Pathol.* **4**, 203 (2003).
17. J. D. Walton, *Plant Physiol.* **104**, 1113 (1994).
18. We thank the Genome Sequencing Platform at the Broad Institute; A. Rokas, M. Borowsky, and G. Fink for critical reading of the manuscript; D. Neafsey, M.-J. Daboussi, and P. Tiffin for helpful discussions; S. Gale, H. Khalili, L. Gaffney, and D. Park for technical assistance; and Syngenta for sharing their 0.5X sequence from *F. graminearum* and their 4X sequence from *F. verticillioides*. Supported by the National Research Initiative of the USDA Cooperative State Research, Education, and Extension Service. F.G.D.B. was supported by the Austrian genome program GEN-AU, the Impuls- und Vernetzungsfonds der Helmholtz-Gemeinschaft, and a grant from the German Federal Ministry of Education and Research. Sequences have been deposited in GenBank: *F. graminearum* assembly AACM00000000, Broad *F. graminearum* annotation XM\_380177-XM\_391816, *F. verticillioides* assembly AAIMO1000000, SNP sequences (ss# 73405725 to 73416217), and hygromycin Loss/gain clones DV998659 to DV998664.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5843/1400/DC1  
Materials and Methods  
SOM Text  
Figs. S1 to S3  
Tables S1 to S16  
References

12 April 2007; accepted 18 July 2007  
10.1126/science.1143708

## The Perception of Rational, Goal-Directed Action in Nonhuman Primates

Justin N. Wood,<sup>1\*</sup> David D. Glynn,<sup>1</sup> Brenda C. Phillips,<sup>4</sup> Marc D. Hauser<sup>1,2,3</sup>

Humans are capable of making inferences about other individuals' intentions and goals by evaluating their actions in relation to the constraints imposed by the environment. This capacity enables humans to go beyond the surface appearance of behavior to draw inferences about an individual's mental states. Presently unclear is whether this capacity is uniquely human or is shared with other animals. We show that cotton-top tamarins, rhesus macaques, and chimpanzees all make spontaneous inferences about a human experimenter's goal by attending to the environmental constraints that guide rational action. These findings rule out simple associative accounts of action perception and show that our capacity to infer rational, goal-directed action likely arose at least as far back as the New World monkeys, some 40 million years ago.

A central characteristic of human action perception is the capacity to read beneath the surface appearance of behavior. When someone acts, we make inferences about

their goals and intentions by referencing each action against a backdrop of environmental constraints. In humans, this capacity appears around the first year of life. For instance, Gergely