

Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*

Jörg Kämper¹, Regine Kahmann¹, Michael Bölker², Li-Jun Ma³, Thomas Brefort¹, Barry J. Saville^{4†}, Flora Banuett⁵, James W. Kronstad⁶, Scott E. Gold⁷, Olaf Müller¹, Michael H. Perlin⁸, Han A. B. Wösten⁹, Ronald de Vries⁹, José Ruiz-Herrera¹⁰, Cristina G. Reynaga-Peña¹⁰, Karen Snetselaar¹¹, Michael McCann¹¹, José Pérez-Martín¹², Michael Feldbrügge¹, Christoph W. Basse¹, Gero Steinberg¹, Jose I. Ibeas¹³, William Holloman¹⁴, Plinio Guzman¹⁵, Mark Farman¹⁶, Jason E. Stajich¹⁷, Rafael Sentandreu¹⁸, Juan M. González-Prieto¹⁹, John C. Kennell²⁰, Lazaro Molina¹, Jan Schirawski¹, Artemio Mendoza-Mendoza¹, Doris Greilinger¹, Karin Münch¹, Nicole Rössel¹, Mario Scherer¹, Miroslav Vraneš¹, Oliver Ladendorf¹, Volker Vincon¹, Uta Fuchs¹, Björn Sandrock², Shaowu Meng⁴, Eric C. H. Ho⁴, Matt J. Cahill⁴, Kylie J. Boyce⁶, Jana Klose⁶, Steven J. Klosterman⁷, Heine J. Deelstra⁹, Lucila Ortiz-Castellanos¹⁰, Weixi Li²¹, Patricia Sanchez-Alonso¹⁵, Peter H. Schreier²², Isolde Häuser-Hahn²², Martin Vaupel²², Edda Koopmann²², Gabi Friedrich²², Hartmut Voss²³, Thomas Schlüter²³, Jonathan Margolis²⁴, Darren Platt²⁴, Candace Swimmer²⁴, Andreas Gnirke²⁴, Feng Chen²⁴, Valentina Vysotskaia²⁴, Gertrud Mannhaupt^{1,25}, Ulrich Güldener²⁵, Martin Münsterkötter²⁵, Dirk Haase²⁵, Matthias Oesterheld²⁵, Hans-Werner Mewes^{25,26}, Evan W. Mauceli³, David DeCaprio³, Claire M. Wade³, Jonathan Butler³, Sarah Young³, David B. Jaffe³, Sarah Calvo³, Chad Nusbaum³, James Galagan³ & Bruce W. Birren³

Ustilago maydis is a ubiquitous pathogen of maize and a well-established model organism for the study of plant–microbe interactions¹. This basidiomycete fungus does not use aggressive virulence strategies to kill its host. *U. maydis* belongs to the group of biotrophic parasites (the smuts) that depend on living tissue for proliferation and development². Here we report the genome sequence for a member of this economically important group of biotrophic fungi. The 20.5-million-base *U. maydis* genome assembly contains 6,902 predicted protein-encoding genes and lacks pathogenicity signatures found in the genomes of aggressive pathogenic fungi, for example a battery of cell-wall-degrading enzymes. However, we detected unexpected genomic features responsible for the pathogenicity of this organism. Specifically, we found 12 clusters of genes encoding small secreted proteins with unknown function. A significant fraction of these genes exists in small gene families. Expression analysis showed that most of the genes contained in these clusters are regulated together and induced in infected tissue. Deletion of individual clusters altered the virulence of *U. maydis* in five cases, ranging from a complete

lack of symptoms to hypervirulence. Despite years of research into the mechanism of pathogenicity in *U. maydis*, no ‘true’ virulence factors³ had been previously identified. Thus, the discovery of the secreted protein gene clusters and the functional demonstration of their decisive role in the infection process illuminate previously unknown mechanisms of pathogenicity operating in biotrophic fungi. Genomic analysis is, similarly, likely to open up new avenues for the discovery of virulence determinants in other pathogens.

Ustilago maydis is a pathogenic basidiomycete fungus that infects maize, one of the world’s major cereal crops. The disease results in stunted plant growth and reduces yield, leading to severe economic losses¹. *U. maydis* is dimorphic (Fig. 1a) and grows in its haploid phase as a saprophytic yeast (Fig. 1c). Sexual development is initiated by the fusion of two haploid cells (Fig. 1d). The resulting dikaryon is filamentous and invades plant cells by means of a specialized infection structure called an appressorium (Fig. 1e). During penetration, the host plasma membrane invaginates and surrounds the invading hypha. An interaction zone develops between plant and fungal

¹Department of Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch Strasse, D-35043 Marburg, Germany. ²Department of Biology, Philipps-University Marburg, Karl-von-Frisch-Strasse 8, D-35032 Marburg, Germany. ³The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁴Department of Biology, University of Toronto at Mississauga, 3359 Mississauga Rd, Mississauga, Ontario, L5L 1C6, Canada. ⁵Department of Biological Sciences, California State University, 1250 Bellflower Boulevard, Long Beach, California 90840, USA. ⁶The Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia, V6T 1Z4, Canada. ⁷Department of Plant Pathology, The University of Georgia, Athens, Georgia 30602-7274, USA. ⁸Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, Kentucky 40292, USA. ⁹Department of Microbiology, Institute of Biomembranes, Utrecht University, Padualaan 8, 3705 SN Utrecht, The Netherlands. ¹⁰Departamento de Ingeniería Genética, Unidad de Biotecnología e Ingeniería Genética de Plantas, Centro de Investigación y de Estudios Avanzados del IPN, 36500 Irapuato, Gto. Mexico. ¹¹Department of Biology, Saint Joseph’s University, 5600 City Ave, Philadelphia, Pennsylvania 19131, USA. ¹²Department of Microbial Biotechnology, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco-UAM, 28049 Madrid, Spain. ¹³Centro Andaluz de Biología del Desarrollo, CSIC/Universidad Pablo de Olavide, Carretera de Utrera Km1, 41013 Sevilla, Spain. ¹⁴Department of Microbiology and Immunology, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA. ¹⁵Department of Plant Genetic Engineering, Cinvestav, Campus-Guanajuato, Km 9.6 Libramiento Norte, 36821 Irapuato, Guanajuato, Mexico. ¹⁶Department of Plant Pathology, University of Kentucky, 1405 Veteran’s Drive, Lexington, Kentucky 40546, USA. ¹⁷Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina 27710, USA. ¹⁸Departamento de Microbiología i Ecología, Facultat de Farmàcia, Universitat de València, Vicent Andrés Estelles s/n, 46100 Burjassot, València, Spain. ¹⁹Departamento de Biotecnología Vegetal II, Centro de Biotecnología Genómica-IPN, Blvd del Maestro s/n, Cd. Reynosa, Tamaulipas, 88710, México. ²⁰Department of Biology, St Louis University, 3507 Laclede Avenue, St Louis, Missouri 63103-2010, USA. ²¹Department of Biology, University of Kentucky, 101 T. H. Morgan Building, Lexington, Kentucky 40546, USA. ²²Bayer CropScience AG, Alfred-Nobel-Strasse 50, D-40789 Monheim, Germany. ²³LION bioscience AG, Waldhofer Strasse 98, D-69123 Heidelberg, Germany. ²⁴Exelixis, Inc., 210 East Grand Avenue, South San Francisco, California 94080, USA. ²⁵Institute for Bioinformatics (MIPS), GSF National Research Center for Environment and Health, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany. ²⁶Genome-oriented Bioinformatics, Technische Universität München, Am Forum 1, D-85354 Freising-Weihenstephan, Germany. †Present address: Forensic Science Program, Trent University, Peterborough, Ontario, K9J 7B8, Canada.

membranes that is characterized by fungal deposits produced by exocytosis⁴ (Fig. 1f). Although hyphae traverse plant cells, there is no apparent host defence response and plant tissue remains alive until late in the infection process. The most characteristic symptom of the disease is large tumours (Fig. 1b), which result from fungus-induced alterations in plant growth. The fungus proliferates and differentiates within tumour tissue (Fig. 1h) and produces masses of black diploid spores (Fig. 1g, i). On germination, spores undergo meiosis and produce the haploid phase⁵.

Genome analysis was performed on the haploid *U. maydis* strain 521 with the use of whole-genome shotgun (10× coverage) and map-based approaches (2.92× coverage; see Supplementary Methods). The assembly comprises 19.8 million bases (Mb) of the estimated genome size of 20.5 Mb (ref. 6). More than 99% of the assembly is represented in the 24 largest scaffolds (Supplementary Methods), which correspond to 23 identified chromosomes. Only chromosome 4 consists of two large scaffolds separated by the ribosomal DNA repeats (Supplementary Fig. S1). The final assembly contains 251 sequence gaps. Subtelomeric regions were identified for all except two chromosomes (Supplementary Table S1). Of more than 30,000

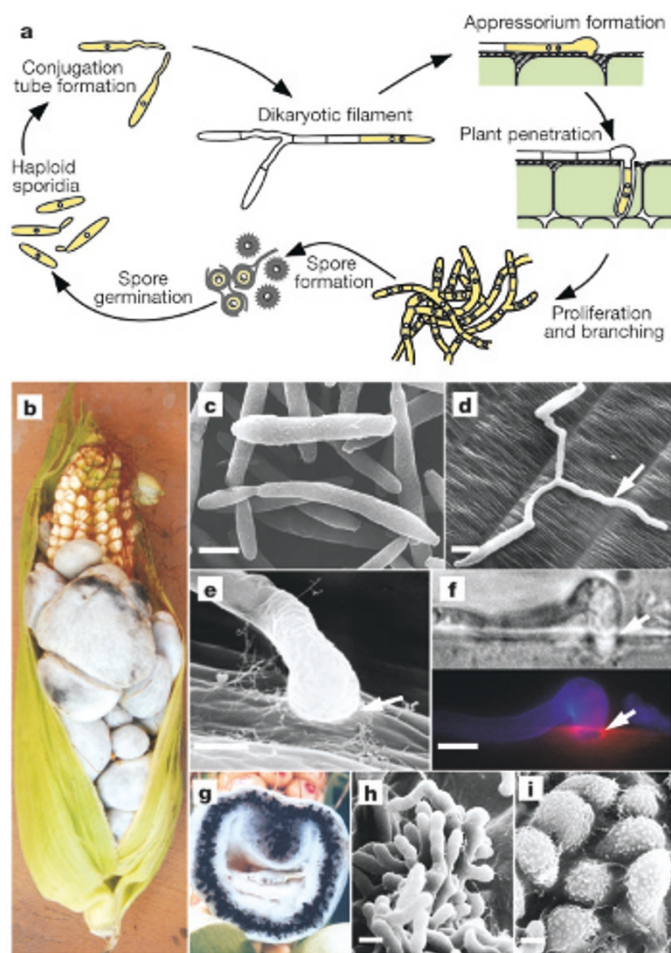


Figure 1 | Life cycle of *U. maydis*. **a**, Developmental stages in the *U. maydis* life cycle. **b**, Tumour formation on maize. **c**, Scanning electron microscopy (SEM) image of haploid sporidia. **d**, SEM image of mated sporidia on plant epidermis; arrow denotes dikaryotic filament. **e**, SEM image of appressorium; arrow marks entry point. **f**, Top, differential interference contrast image of appressorium; bottom, epifluorescence image of fungal cell wall stained with calcofluor (blue) and endocytotic vesicles stained with FM4-64 (red). The bright ring indicates active secretion and endocytosis at the fungus–plant interface; arrows indicate the penetration point. **g**, Black teliospores visible in tumour section. **h**, SEM image of sporogenous hyphae and early stages of spore development. **i**, SEM image of ornamented teliospores. Scale bars, 5 µm.

expressed sequence tags (ESTs; longer than 150 base pairs (bp); Supplementary Methods), 99.6% could be aligned with the genomic sequence, indicating almost complete sequence coverage. Independently, strain FB1—closely related to strain 521 by inbreeding⁷—was shotgun sequenced to 5× coverage, yielding an assembly of 19.3 Mb (a detailed description of the pedigree of strains FB1 and 521 is given in Supplementary Methods). A total of 18.9 Mb of this assembly can be aligned to the strain 521 assembly with 99.97% nucleotide sequence identity.

The *U. maydis* genome (20.5 Mb)⁶ is rather small in comparison with genomes of other plant pathogenic fungi (see the Broad Institute's Fungal Genome Initiative (FGI) Candidate Genome website, <http://www.broad.mit.edu/annotation/fungi/fgi/candidates.html>). The MIPS *Ustilago maydis* database (MUMDB; <http://mips.gsf.de/genre/proj/ustilago/>) currently lists 6,902 gene models, whereas in the phytopathogenic ascomycetes, gene numbers are considerably higher (12,841 in *Magnaporthe grisea*, 16,597 in *Stagonospora nodorum* and 11,640 in *Fusarium graminearum*; see the FGI website). The small number of genes is partly reflected in the absence of significant expansions of gene families (Supplementary Table S2). The small number of introns and their short mean length qualify as additional explanations for the small genome size of *U. maydis*. The average number of introns per gene is 0.46, with 70% of genes containing no intron. The related basidiomycetes *Cryptococcus neoformans*, *Coprinus cinereus* and *Phanerochaete chrysosporium* contain an average of 5.3, 4.5 and 2.6 introns per gene, respectively⁸. One outstanding example is the highly conserved *tor1* kinase gene, which lacks introns in *U. maydis* but contains 23 in *C. neoformans*. Apparently, the *U. maydis* genome has been shaped by massive, lineage-specific intron loss, as has been observed in the ascomycetous yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*⁹. Intron loss has been proposed to occur by the recombination of reverse-transcribed transcripts with the genomic copy⁹. The small number of introns observed in *U. maydis* might therefore be a consequence of the highly efficient homologous recombination system^{10,11}.

The genome of *U. maydis* does not show signs of large-scale duplication events (Supplementary Fig. S2) and is largely devoid of repetitive DNA elements. Only 1.1% of the assembly consists of mostly non-functional, transposon-derived sequences. This is considerably lower than in most fungi (Supplementary Fig. S2, and Supplementary Tables S3 and S4). Surprisingly, we could not detect any of the known components of the RNA interference pathway, which are thought to have a function in restricting mobile elements¹², and there is no genomic evidence for gene inactivation by repeat-induced point mutation (RIP)¹³. However, it has recently been shown that the expression of heterologous genes in *U. maydis* often results in premature termination of transcription¹⁴. On these grounds, we speculate that *U. maydis* uses this novel mechanism to restrict the activity of invading genes.

Similarly to other fungi and plants^{8,15,16}, centromeres in *U. maydis* seem to coincide with retroelement (HobS)-containing regions that occur once on each chromosome (Supplementary Fig. S1). All known DNA fragments (ARS elements) that allow autonomous replication of plasmids in *U. maydis*¹⁷ match such regions (Supplementary Fig. S1 and Supplementary Table S4). Apparently, the maintenance of plasmids in *U. maydis* requires a centromeric region in addition to an origin of DNA replication, in a similar manner to the situation in *Yarrowia lipolytica*¹⁸. Within the *U. maydis* ARS elements we detected a perfectly conserved 11-bp sequence (ATTCACGATTC) that is strongly over-represented in the genome (5,236 versus 10 expected). Because 96% of these elements are located in intergenic regions, we postulate that this motif defines the origin of replication. In *S. pombe*, functional replication origins occur in comparable numbers and are restricted to intergenic regions. However, these AT-rich regions lack a conserved consensus sequence¹⁹. Even within the basidiomycetes, *U. maydis* is unique in exhibiting such a conserved motif.

Remarkably, *U. maydis* possesses only few genes known to be involved in pathogenesis in other fungi. For example, the cereal pathogens *M. grisea*, *F. graminearum* and *Cochliobolus heterostrophus* contain large numbers (15–25) of genes encoding polyketide synthases^{20,21}, enzymes involved in the production of small bioactive compounds such as antibiotics or mycotoxins. In contrast, *U. maydis* contains only three. Genes encoding polysaccharide hydrolases, polysaccharide lyases and pectin esterases are considered to be signatures of necrotrophic fungi that use such enzymes to degrade living and dead plant tissue. *U. maydis* contains only 33 such hydrolytic enzymes, in contrast with 138 and 103 for *M. grisea* and *F. graminearum*, respectively (Supplementary Table S5). The minimal set of hydrolytic enzymes found in *U. maydis* seems perfectly in line with its biotrophic lifestyle, in which damage to the host should be minimized and the release of cell wall fragments, which often trigger plant defence responses, has to be avoided².

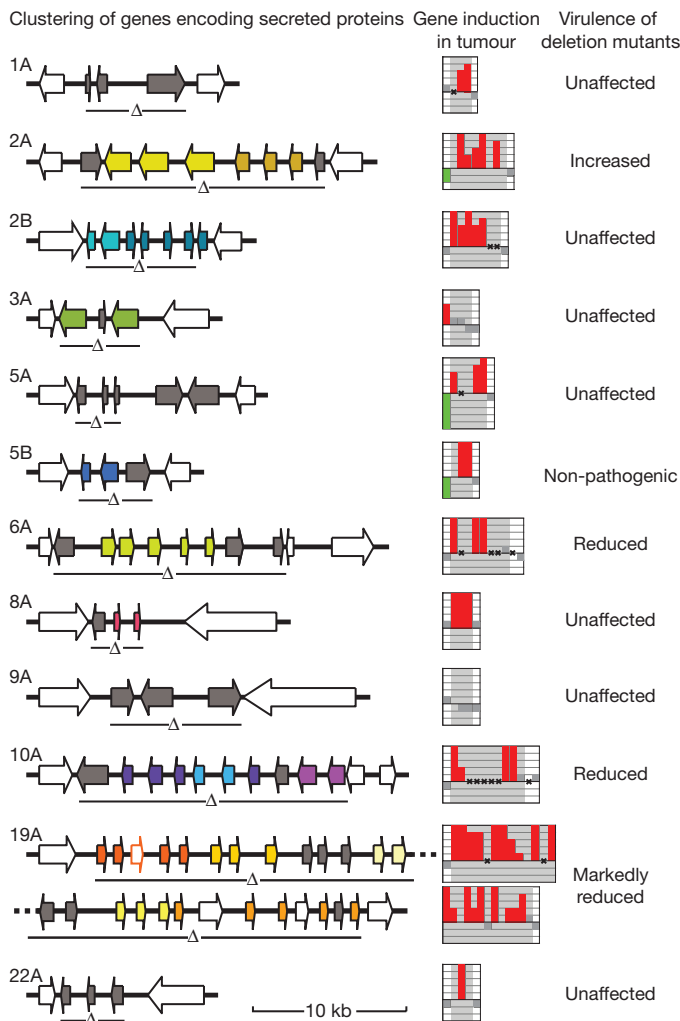


Figure 2 | Gene clusters for secreted proteins. Left, schematic representation of gene clusters and flanking genes. Genes are depicted as arrows, filling reflects the fact that encoded proteins are predicted to be secreted. Colours indicate protein families, unrelated cluster genes are shaded in grey. Regions deleted in cluster mutants are indicated (Δ). Middle, DNA-array analysis of cluster gene expression in tumour tissue. Significant changes in gene expression (compared with axenic culture) were categorized (fold changes on the y-axis: less than 3, 3 to less than 5, 5 to less than 10, 10 to less than 20, at least 20), with red columns indicating induction, green columns repression, and grey columns insignificant changes. Gaps indicate undetectable transcript levels; genes not present on the array are marked with a cross. Light grey areas demarcate clustered genes. See Supplementary Table S6 and Supplementary Fig. S3 for further details.

The paucity of secreted cell-wall-degrading enzymes stands in sharp contrast to the large number of secreted proteins with unknown function. Of 426 proteins predicted to be secreted, 298 (70%) cannot be ascribed a function, and of these almost two-thirds (193) are specific for *U. maydis*. Of all genes encoding secreted proteins, 18.6% are arranged in 12 gene clusters (Fig. 2). The clusters are scattered all over the genome and comprise 3–26 genes. Eight of the 12 clusters contain groups of two to five related genes in tandem arrays, indicating that they might have arisen by duplication. DNA-array analysis revealed that the expression of most clustered genes is induced in tumour tissue, whereas that of flanking genes is not (Fig. 2, and Supplementary Table S6). Although seven of the clustered genes that were induced in tumour tissue were also upregulated by the central regulator of pathogenic development, the bE/bW heterodimer, they were not induced by pheromone stimulation, cyclic AMP signalling, changing of nitrogen or carbon sources, iron depletion, or oxidative stress (data not shown). The specific upregulation of many cluster genes in tumour tissue indicates a possible concerted function during pathogenic development.

To test this assertion, we constructed deletion mutants for all 12 clusters in strain SG200. Mutants were not affected in their growth on minimal medium, showed no morphological alterations and were indistinguishable from strain SG200 in their ability to produce filaments (not shown). Mutants were syringe-injected into maize seedlings. In five cases, deletions caused clear disease-associated phenotypes (Fig. 2, Supplementary Table S6 and Supplementary Fig. S3). Linkage of the observed phenotype to the respective deletion was directly confirmed for cluster 5B by complementation (Supplementary Fig. S4). For the other four clusters, which were significantly larger, complementation attempts were not successful for technical reasons. We therefore generated three independent mutants in each case, which all displayed the same virulence phenotype (Supplementary Fig. S3). Mutants 6A and 10A were still able to infect plants, but the incidence and size of tumours were reduced. Two mutants, 5B and 19A, arrested growth at distinct stages of biotrophic development. Mutant 19A was able to penetrate and to grow inside the plant tissue, but failed to induce large tumours and was defective in spore formation. It is conceivable that some of the proteins encoded by this cluster have tumour-inducing activity. Alternatively, they may suppress plant defence reactions or reprogram the metabolism of the host to allocate resources to the fungal parasite. Deletion of cluster 5B resulted in growth arrest early during penetration of the epidermis, which could indicate a specific need of these proteins during the establishment of a functional interface between the fungus and the host cell. Finally, mutants deleted for cluster 2A showed increased virulence, as judged by the incidence and size of tumours. This hyper-virulence phenotype indicates that the respective proteins might attenuate fungal proliferation. For biotrophic fungi it may be important to prevent the premature development of disease symptoms, because this could affect plant growth so severely that the fungus might not be able to complete its life cycle. Seven cluster mutants were not affected in virulence. For four of these clusters, related genes were found elsewhere in the genome (Supplementary Table S6); the number of clusters with crucial functions for disease progression might therefore actually be even higher.

Our results have shown that secreted protein effectors are essential for fungal proliferation inside the plant host. How these novel effectors exert their function is currently unknown. We envisage that some of the proteins are translocated into plant cells, as has recently been observed in rust and oomycete plant pathogens^{22–24}. These pathogens develop specialized infection structures (haustoria), which are implicated in the exchange of nutrients and proteins². *U. maydis* lacks such structures, but during intracellular growth of the infecting hyphae an extended interaction zone is established, which may be the site at which protein translocation into the host cell takes place.

The genome sequence of the plant pathogenic fungus *U. maydis* has provided unexpected insights into the peculiarities of a biotrophic fungal pathogen. It is apparent that plant cell wall degradation by the fungus is minimized, whereas the secretion of novel protein effectors has a decisive function during infection. This strategy—to live in ‘pretend harmony’ with its host—may be shared not only with other obligate biotrophic pathogens but also with plant-growth-promoting mycorrhizal fungi. The availability of the genome sequence of *U. maydis*, combined with its genetic tractability, therefore offers an excellent opportunity to unravel the molecular secrets of fungal biotrophy. The identification of ‘biotrophy clusters’ in *U. maydis* is likely to have the same inspiring impact on understanding fungal disease strategies as the way in which the discovery of bacterial pathogenicity islands has shaped our present view of bacterial infection strategies.

METHODS

Strains. *Ustilago maydis* 521 (DSMZ number 14603; Supplementary Methods) and FB1 (ref. 7) were used as DNA donors for sequencing. FB2 (ref. 7) was used as the mating partner of FB1 for plant infections. The haploid pathogenic strain SG200 was used for generating deletion mutants and is described in Supplementary Methods.

Genome sequencing and annotation. Strain 521 was sequenced by a combinatorial approach relying on a mapped bacterial artificial chromosome library (2.9× coverage) and a whole-genome shotgun approach (10× coverage). Strain FB1 was sequenced by a whole-genome shotgun approach (5× coverage) (see Supplementary Methods). Predicted protein-encoding genes (see the FGI website) for *U. maydis* were refined manually, including sequence information from more than 4,100 EST clusters (Supplementary Methods) and automatically annotated with the MIPS PEDANT suite²⁵. Data can be accessed on the MUMDB website.

Prediction of secreted proteins. For the prediction of amino-terminal secretion signals, SignalP 3.0 (ref. 26) was used. A total of 750 proteins were predicted to carry a signal peptide both by the hidden Markov and the neural network algorithms. These candidates were analysed with the integral prediction score of ProtComp 6.0 (<http://www.softberry.com>), yielding 426 candidate secreted proteins. In addition, TARGETP²⁷ was used to predict protein localization.

DNA-array analysis. For DNA-array analysis, custom-designed Affymetrix chips were used. Probe sets were designed on the basis of the map-based sequencing assembly. For each predicted gene, 33 perfect match and 33 corresponding mismatch probes were designed, covering a region of 800 bp at the 3′ ends. The *U. maydis* DNA arrays address about 6,200 genes. Probe sets for the individual genes are shown on the MUMDB website. For DNA-array analysis, RNA was extracted from strain FB1 grown to an $A_{600\text{ nm}}$ of 0.5 at 28 °C in liquid array medium (AM), which consisted of 6.25% (v/v) salt solution²⁸, 1% (v/v) vitamin solution²⁸, 30 mM L-glutamine, 1% (w/v) glucose, pH 7.0 (filter sterilized). For RNA from tumour tissue, 7-day-old corn plants were infected with a mixture of strains FB1 and FB2, as described previously²⁹. Tumours were harvested at day 13 after infection. Total RNA was extracted from tumour tissue and axenic cultures with the use of the Trizol method in accordance with the manufacturer’s instructions (Invitrogen). RNA was purified with RNeasy MinElute columns (Qiagen), and RNA integrity was checked on an Agilent Bioanalyser 2100. DNA-array analyses were performed in accordance with the standard Affymetrix protocol in at least two biological replicates. Data analysis was performed with the Affymetrix Micro Array Suite 5.1 software package.

Mutant generation and analysis. Cluster mutants were generated in strain SG200 by gene replacement with PCR-generated constructs as described¹¹, or by subcloning the PCR-derived constructs first. In the latter case, both border fragments were sequenced and shown not to carry mutations. For each cluster, at least two independent mutants were generated and assayed repeatedly for virulence on 7-day-old seedlings of Early Golden Bantam, with a minimum of 40 plants per mutant. Symptom development was scored 12 days after infection. Details of the infection procedure and rating of symptoms are given in Supplementary Fig. S3. Fungal development was monitored by staining with calcofluor and chlorazole black E as described²⁹. Cluster mutant 5B, which is nonpathogenic, was complemented by transformation with a DNA fragment comprising the entire cluster region, ligated to a carboxin resistance cassette. Details are given in Supplementary Fig. S4.

Received 18 August; accepted 15 September 2006.

1. Martinez-Espinoza, A. D., Garcia-Pedrajas, M. D. & Gold, S. E. The Ustilaginales as plant pests and model systems. *Fungal Genet. Biol.* **35**, 1–20 (2002).

2. Mendgen, K. & Hahn, M. Plant infection and the establishment of fungal biotrophy. *Trends Plant Sci.* **7**, 352–356 (2002).
3. Wassenaar, T. M. & Gaastra, W. Bacterial virulence: can we draw the line? *FEMS Microbiol. Lett.* **201**, 1–7 (2001).
4. Bauer, R., Oberwinkler, F. & Vanky, K. Ultrastructural markers and systematics in smut fungi and allied taxa. *Can. J. Bot.* **75**, 1273–1314 (1997).
5. Banuett, F. Genetics of *Ustilago maydis*, a fungal pathogen that induces tumors in maize. *Annu. Rev. Genet.* **29**, 179–208 (1995).
6. Meksem, K. et al. A bacterial artificial chromosome based physical map of the *Ustilago maydis* genome. *Genome* **48**, 207–216 (2005).
7. Banuett, F. & Herskowitz, I. Different *a* alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proc. Natl Acad. Sci. USA* **86**, 5878–5882 (1989).
8. Loftus, B. J. et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321–1324 (2005).
9. Roy, S. W. & Gilbert, W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Rev. Genet.* **7**, 211–221 (2006).
10. Holliday, R. Early studies on recombination and DNA repair in *Ustilago maydis*. *DNA Repair (Amst.)* **3**, 671–682 (2004).
11. Kämper, J. A PCR-based system for highly efficient generation of gene replacement mutants in *Ustilago maydis*. *Mol. Genet. Genomics* **271**, 103–110 (2004).
12. Vastenhouw, N. L. & Plasterk, R. H. RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet.* **20**, 314–319 (2004).
13. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417–423 (2004).
14. Zarnack, K. et al. Tetracycline-regulated gene expression in the pathogen *Ustilago maydis*. *Fungal Genet. Biol.* **43**, 727–738 (2006).
15. Cambareri, E. B., Aisner, R. & Carbon, J. Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. *Mol. Cell. Biol.* **18**, 5465–5477 (1998).
16. Houben, A. & Schubert, I. DNA and proteins of plant centromeres. *Curr. Opin. Plant Biol.* **6**, 554–560 (2003).
17. Tsukuda, T., Carleton, S., Fotheringham, S. & Holloman, W. K. Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis*. *Mol. Cell. Biol.* **8**, 3703–3709 (1988).
18. Vernis, L. et al. Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. *J. Mol. Biol.* **305**, 203–217 (2001).
19. Dai, J., Chuang, R. Y. & Kelly, T. J. DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc. Natl Acad. Sci. USA* **102**, 337–342 (2005).
20. Dean, R. A. et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
21. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl Acad. Sci. USA* **100**, 15670–15675 (2003).
22. Kemen, E. et al. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* **18**, 1130–1139 (2005).
23. Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A. & Ellis, J. G. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755–768 (2004).
24. Birch, P. R., Rehmany, A. P., Pritchard, L., Kamoun, S. & Beynon, J. L. Trafficking arms: oomycete effectors enter host plant cells. *Trends Microbiol.* **14**, 8–11 (2006).
25. Mewes, H. W. et al. MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
26. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
27. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
28. Holliday, R. in *Handbook of Genetics* (ed. King, R. C.) 575–595 (Plenum, New York, 1974).
29. Brachmann, A., Schirawski, J., Müller, P. & Kahmann, R. An unusual MAP kinase is required for efficient penetration of the plant surface by *Ustilago maydis*. *EMBO J.* **22**, 2199–2210 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.K., M. B. and R.K. thank G. Sawers and U. Kämper for critical reading of the manuscript. The genome sequencing of *Ustilago maydis* strain 521 is part of the fungal genome initiative and was funded by National Human Genome Research Institute (USA) and BayerCropScience AG (Germany). F.B. was supported by a grant from the National Institutes of Health (USA). J.K. and R.K. thank the German Ministry of Education and Science (BMBF) for financing the DNA array setup and the Max Planck Society for their support of the manual genome annotation. F.B. was supported by a grant from the National Institutes of Health, B.J.S. was supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Foundation for Innovation, J.W.K. received funding from the Natural Sciences and Engineering Research Council of Canada,

J.R.-H. received funding from CONACYT, México, A.M.-M. was supported by a fellowship from the Humboldt Foundation, and L.M. was supported by an EU grant.

Author Contributions All authors were involved in planning and executing the genome sequencing project. B.W.B., J.G., L.-J.M., E.W.M., D.D., C.M.W., J.B., S.Y., D.B.J., S.C., C.N., E.K., G.F., P.H.S., I.H.-H., M. Vaupel, H.V., T.S., J.M., D.P., C.S., A.G., F.C. and V. Vysotskaia contributed to the three independent sequencing projects; M.M., G.M., U.G., D.H., M.O. and H.-W.M. were responsible for gene model refinement, database design and database maintenance; G.M., J. Kämper, R.K., G.S., M. Feldbrügge, J.S., C.W.B., U.F., M.B., B.S., B.J.S., M.J.C., E.C.H.H., S.M., F.B., J.W.K., K.J.B., J. Klose, S.E.G., S.J.K., M.H.P., H.A.B.W., R.deV., H.J.D., J.R.-H., C.G.R.-P., L.O.-C., M.McC., K.S., J.P.-M., J.I.I., W.H., P.G., P.S.-A., M. Farman, J.E.S., R.S., J.M.G.-P., J.C.K., W.L. and D.H. were involved in functional annotation and

interpretation; T.B., O.M., L.M., A.M.-M., D.G., K.M., N.R., V. Vincon, M. Vraneš, M.S. and O.L. performed experiments. J. Kämper, R.K. and M.B. wrote and edited the paper with input from L.-J.M., J.G., F.B., J.W.K., B.J.S. and S.E.G. Individual contributions of authors can be found as Supplementary Notes.

Author Information The whole-genome shotgun data have been deposited at GenBank under the project accession number AACP00000000. Individual, automatically generated gene models have been deposited under accession numbers XM_751055 to XM_757567. Supplementary Table S8 lists all gene models that have been altered during manual reannotation. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.K. (kahmann@mpi-marburg.mpg.de), M.B. (boelker@staff.uni-marburg.de) or J.K. (kaemper@mpi-marburg.mpg.de).