EXTENDED PDF FORMAT
SPONSORED BY
Science
AAAS
http://www.sciencemag.org

Science
AAAS

Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism

Pietro D. Spanu, *et al.*

Science **330**, 1543 (2010);
DOI: 10.1126/science.1194573

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this infomation is current as of December 9, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/330/6010/1543.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2010/12/08/330.6010.1543.DC1.html

This article **cites 13 articles**, 5 of which can be accessed free:
http://www.sciencemag.org/content/330/6010/1543.full.html#ref-list-1

This article has been **cited by** 2 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/330/6010/1543.full.html#related-urls

This article appears in the following **subject collections:**
Botany
http://www.sciencemag.org/cgi/collection/botany

Downloaded from www.sciencemag.org on December 9, 2010

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. Copyright 2010 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

P. infestans (*17*) and could modulate expression patterns of genes located in the gene-sparse regions. In addition, histone hypomethylation reduces DNA stability (*18*, *19*) and may have contributed to genome plasticity in the *P. infestans* lineage by regulating transposon activity as well as genomic and expression variability (*20*, *21*). Finally, understanding *P. infestans* genome evolution should prove useful in designing rational strategies for sustainable late blight disease management based on targeting the most evolutionarily stable genes in this lineage.

References and Notes

1. L. P. Kroon, F. T. Bakker, G. B. van den Bosch, P. J. Bonants, W. G. Flier, *Fungal Genet. Biol.* **41**, 766 (2004).
2. J. E. Blair, M. D. Coffey, S. Y. Park, D. M. Geiser, S. Kang, *Fungal Genet. Biol.* **45**, 266 (2008).
3. N. J. Grünwald, W. G. Flier, *Annu. Rev. Phytopathol.* **43**, 171 (2005).
4. B. J. Haas *et al.*, *Nature* **461**, 393 (2009).
5. Materials and methods are available as supporting material on *Science* Online.
6. S. Yoon, Z. Xuan, V. Makarov, K. Ye, J. Sebat, *Genome Res.* **19**, 1586 (2009).
7. Z. Yang, R. Nielsen, *Mol. Biol. Evol.* **17**, 32 (2000).
8. Z. Liu *et al.*, *Mol. Biol. Evol.* **22**, 659 (2005).
9. R. L. Allen *et al.*, *Science* **306**, 1957 (2004).
10. J. Win *et al.*, *Plant Cell* **19**, 2349 (2007).
11. S. Kamoun, *Annu. Rev. Phytopathol.* **44**, 41 (2006).
12. S. K. Oh *et al.*, *Plant Cell* **21**, 2928 (2009).
13. M. Tian, E. Huitema, L. Da Cunha, T. Torto-Alalibo, S. Kamoun, *J. Biol. Chem.* **279**, 26370 (2004).
14. T. Kouzarides, *Curr. Opin. Genet. Dev.* **12**, 198 (2002).
15. Y. Zhang, D. Reinberg, *Genes Dev.* **15**, 2343 (2001).
16. L. I. Elizondo, P. Jafar-Nejad, J. M. Clewing, C. F. Boerkoel, *Curr. Genomics* **10**, 64 (2009).
17. P. van West *et al.*, *Microbiology* **154**, 1482 (2008).
18. A. H. Peters *et al.*, *Cell* **107**, 323 (2001).
19. J. C. Peng, G. H. Karpen, R. S. Hawley, *PLoS Genet.* **5**, e1000435 (2009).
20. D. W. Zeh, J. A. Zeh, Y. Ishida, *Bioessays* **31**, 715 (2009).
21. N. Elango, S. H. Kim, E. Vigoda, S. V. Yi, A. Sidow, *PLOS Comput. Biol.* **4**, e1000015 (2008).

22. We thank the Broad Institute Sequencing Platform and J. Pike for sequencing; G. Kessel and V. Vleeshouwers for biomaterial; and D. Weigel, J. Dangl, and J. Ecker for useful comments. This project was funded by the Gatsby Charitable Foundation, Marie-Curie IEF contract 255104, National Research Initiative of the U.S. Department of Agriculture grant 2006-35600-16623, NSF grant EF-0523670, and research funding program LOEWE of the Ministry of Research, Science and the Arts of Hesse (Germany). Sequences are deposited in GenBank under the submission accession numbers SRA02326–2329 and SRA024355 and in the Short Read Archive under study accession numbers ERP000341–344.

Supporting Online Material
www.sciencemag.org/cgi/content/full/330/6010/1540/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S4
References

1 June 2010; accepted 21 October 2010
10.1126/science.1193070

Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism

Pietro D. Spanu,[1]* James C. Abbott,[1]† Joelle Amselem,[2,15]† Timothy A. Burgis,[1]† Darren M. Soanes,[3]† Kurt Stüber,[4]† Emiel Ver Loren van Themaat,[4]† James K. M. Brown,[5]‡ Sarah A. Butcher,[1]‡ Sarah J. Gurr,[6]‡ Marc-Henri Lebrun,[15]‡ Christopher J. Ridout,[5]‡ Paul Schulze-Lefert,[4]‡ Nicholas J. Talbot,[3]‡ Nahal Ahmadinejad,[4] Christian Ametz,[1] Geraint R. Barton,[1] Mariam Benjdia,[4] Przemyslaw Bidzinski,[4] Laurence V. Bindschedler,[7] Maike Both,[1] Marin T. Brewer,[8] Lance Cadle-Davidson,[9,10]‡ Molly M. Cadle-Davidson,[9] Jerome Collemare,[2]§ Rainer Cramer,[7] Omer Frenkel,[8] Dale Godfrey,[11] James Harriman,[9] Claire Hoede,[2] Brian C. King,[8] Sven Klages,[12] Jochen Kleemann,[4] Daniela Knoll,[4] Prasanna S. Koti,[4] Jonathan Kreplak,[2] Francisco J. López-Ruiz,[5] Xunli Lu,[4] Takaki Maekawa,[4] Siraprapa Mahanil,[9] Cristina Micali,[4] Michael G. Milgroom,[8] Giovanni Montana,[1] Sandra Noir,[4]|| Richard J. O'Connell,[4] Simone Oberhaensli,[13] Francis Parlange,[13] Carsten Pedersen,[11] Hadi Quesneville,[2] Richard Reinhardt,[12] Matthias Rott,[4] Soledad Sacristán,[14] Sarah M. Schmidt,[4]¶ Moritz Schön,[4] Pari Skamnioti,[6] Hans Sommer,[4] Amber Stephens,[4] Hiroyuki Takahara,[4] Hans Thordal-Christensen,[11] Marielle Vigouroux,[6] Ralf Weßling,[4] Thomas Wicker,[13] Ralph Panstruga[4]*

Powdery mildews are phytopathogens whose growth and reproduction are entirely dependent on living plant cells. The molecular basis of this life-style, obligate biotrophy, remains unknown. We present the genome analysis of barley powdery mildew, *Blumeria graminis* f.sp. *hordei* (*Blumeria*), as well as a comparison with the analysis of two powdery mildews pathogenic on dicotyledonous plants. These genomes display massive retrotransposon proliferation, genome-size expansion, and gene losses. The missing genes encode enzymes of primary and secondary metabolism, carbohydrate-active enzymes, and transporters, probably reflecting their redundancy in an exclusively biotrophic life-style. Among the 248 candidate effectors of pathogenesis identified in the *Blumeria* genome, very few (less than 10) define a core set conserved in all three mildews, suggesting that most effectors represent species-specific adaptations.

Filamentous eukaryotes such as fungi and oomycetes (stramenopiles) are responsible for many serious plant diseases. Among these pathogens is a group of taxonomically diverse species, collectively termed obligate biotrophs, which only grow and reproduce on living plants. These microorganisms cause rusts, as well as downy and powdery mildews, and form dedicated invasive infection structures (haustoria) for nutrient uptake. Obligate biotrophs are found in two kingdoms (Stramenopila and Fungi) and in both major fungal phyla (Ascomycota and Basidiomycota), indicating that biotrophy is the result of convergent evolution. The ascomycete powdery mildews infect ~10,000 angiosperm species, including many important crops (*1*). They form morphologically complex structures during asexual pathogenesis and produce fruiting bodies (cleistothecia), which develop after sexual reproduction (Fig. 1 and fig. S1).

We sequenced the haploid *Blumeria* genome with the use of Sanger protocols and second-generation methods (table S1) (*2*). We assembled the sequence reads with a combination of the cortex and CABOG (Celera assembler with the best overlap graph) (*3*) assemblers into 15,111 contigs (L_{50}: 18,024 bases; L_{50} is the length of the smallest N_{50} contig, where N_{50} is the minimum number of contigs required to represent 50% of the genome) on 6898 supercontig scaffolds (L_{50}: 2,209,085 bases). The overall assembly size is 119,213,040 nucleotides (table S1). We estimate that the actual genome size is ~120 Mb, corresponding to 140-fold coverage of the *Blumeria* genome. We additionally generated draft genome assemblies (~eightfold coverage each) of two other powdery mildew species, *Erysiphe pisi* [pathogenic on pea (*Pisum sativum*)] and *Golovinomyces orontii* (pathogenic on *Arabidopsis thaliana*). Together with *Blumeria*, these species represent three of the five major tribes of the order Erysiphales, which diverged ~70 million years ago (*4*). We calculated that the genome sizes of the latter two species are ~151 and ~160 Mb, respectively (table

[1]Department of Life Sciences, Imperial College London, London, UK. [2]Institute National de la Recherche Agronomique (INRA), Unite de Recherche Genomique Info, Versailles, France. [3]School of Biosciences, University of Exeter, Exeter, UK. [4]Department of Plant Microbe Interactions, Max-Planck Institute for Plant Breeding Research, Cologne, Germany. [5]Department of Disease and Stress Biology, John Innes Centre, Norwich, UK. [6]Department of Plant Sciences, University of Oxford, Oxford, UK. [7]Department of Chemistry, University of Reading, Reading, UK. [8]Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, USA. [9]U.S. Department of Agriculture–Agricultural Research Service Grape Genetics Research Unit, Geneva, NY 14853, USA. [10]Department of Plant Pathology and Plant Microbe Biology, Cornell University, Geneva, NY 14456, USA. [11]Department of Agriculture and Ecology, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark. [12]Max-Planck Institute for Molecular Genetics, Berlin, Germany. [13]Institute of Plant Biology, University of Zürich, Zürich, Switzerland. [14]Centro de Biotecnología y Genómica de Plantas (UPM-INIA) and E.T.S.I. Agrónomos Universidad Politécnica de Madrid, Madrid, Spain. [15]INRA, Unite BIOGER-CPP, Grignon, France.

*To whom correspondence should be addressed. E-mail: p.spanu@imperial.ac.uk (P.D.S.); panstrug@mpiz-koeln.mpg.de (R.P.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
§Present address: Wageningen University, Laboratory of Phytopathology, Wageningen, Netherlands.
||Present address: Institut de Biologie Moléculaire des Plantes–CNRS, Strasbourg, France.
¶Present address: Plant Pathology, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands.

Downloaded from www.sciencemag.org on December 9, 2010

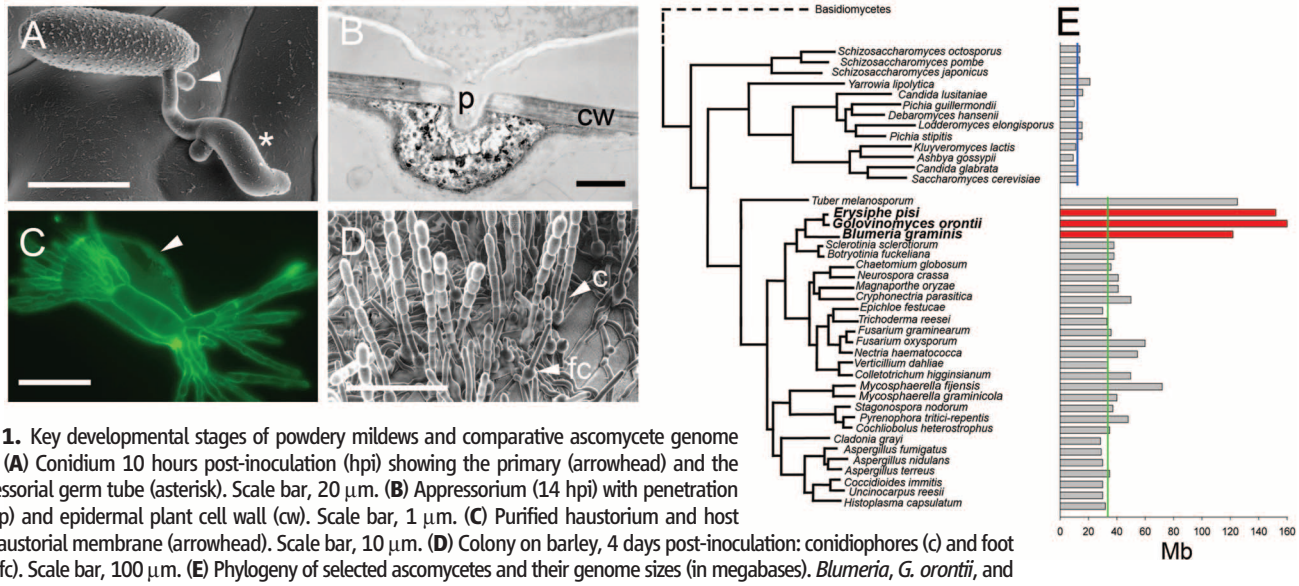

Fig. 1. Key developmental stages of powdery mildews and comparative ascomycete genome sizes. (**A**) Conidium 10 hours post-inoculation (hpi) showing the primary (arrowhead) and the appressorial germ tube (asterisk). Scale bar, 20 μm. (**B**) Appressorium (14 hpi) with penetration peg (p) and epidermal plant cell wall (cw). Scale bar, 1 μm. (**C**) Purified haustorium and host perihaustorial membrane (arrowhead). Scale bar, 10 μm. (**D**) Colony on barley, 4 days post-inoculation: conidiophores (c) and foot cells (fc). Scale bar, 100 μm. (**E**) Phylogeny of selected ascomycetes and their genome sizes (in megabases). *Blumeria*, *G. orontii*, and *E. pisi* are shown in red. The median genome sizes of the hemiascomycetes (blue vertical line, 12.3 Mb) and euascomycetes (green vertical line, 36.7 Mb) are also shown.

S1). Thus, the genome size of each of the mildews is more than four times larger than the median of other ascomycetes (Fig. 1E). We first annotated the *Blumeria* genome using ab initio gene finders followed by extensive manual curation (table S2). The actual number of curated genes is 5854, which is at the lower end of the range of fungal genomes.

The comparatively low gene number and the inability of the parasite to grow in vitro suggest that the mildew genomes may lack genes typically present in autotrophic ascomycetes. We systematically searched for genes absent in the mildews but present in baker's yeast (*Saccharomyces cerevisiae*) and the phytopathogens *Colletotrichum higginsianum*, *Magnaporthe oryzae*, and *Sclerotinia sclerotiorum* (Fig. 2A). We identified 90 yeast genes by this procedure and 9 additional common ascomycete genes by manual inspection that are missing in the genome assemblies of all three mildews [hereafter referred to as missing ascomycete core genes (MACGs)]. It is unlikely that these gene sets were missed as a result of incomplete genome coverage because (i) the 140× *Blumeria* assembly encompasses >99% of the conserved gene space (table S1), and (ii) these genes are missing in all three assemblies. The MACGs represent a diverse set of metabolic and regulatory proteins, affecting multiple processes and pathways (for example, thiamine biosynthesis), and a considerable subset of MACGs (57 to 77%) also seems to be absent in other obligate biotrophic phytopathogens (fig. S2 and table S3).

The existence of MACGs raises the possibility that their expression may be detrimental to biotrophy. To test this, we determined expression of MACG homologs in *C. higginsianum*, a phytopathogenic fungus that first employs a biotrophic growth mode and later switches to necrotrophic pathogenesis, involving host cell killing. Analysis of the *C. higginsianum* transcriptome revealed that most of the MACG homologs that we tested (26 out of 32) are expressed during the biotrophic

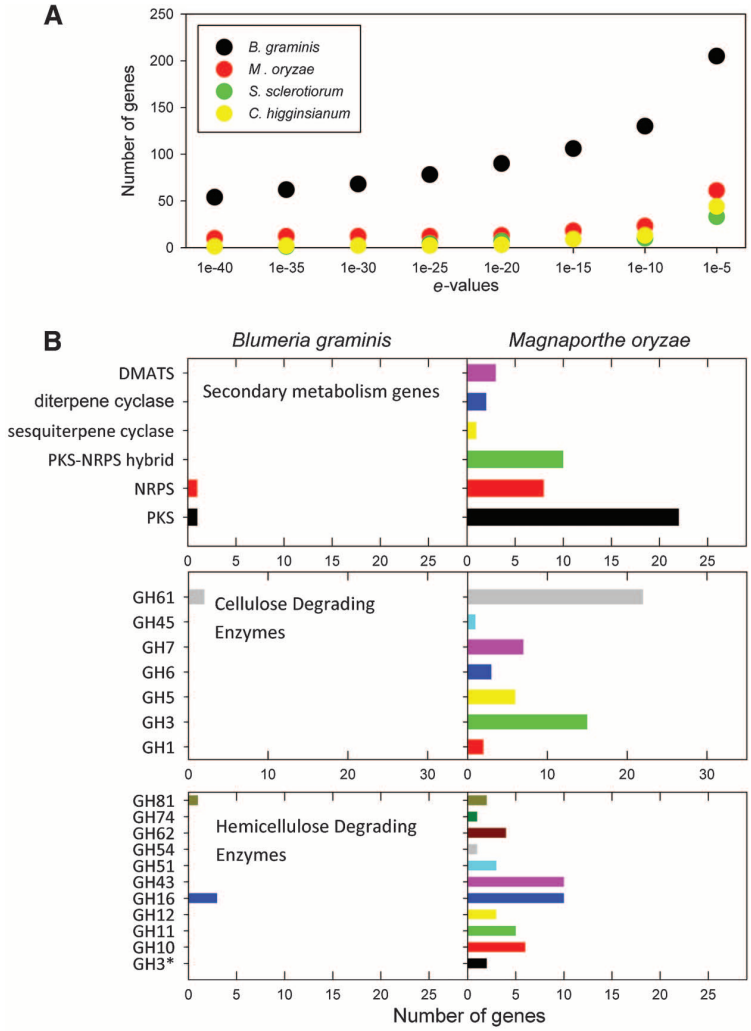

Fig. 2. Gene losses in powdery mildews. (**A**) Number of missing *S. cerevisiae* proteins in one fungus compared with the three others as a function of TBLASTN *e*-values. (**B**) Number of genes devoted to secondary metabolism and genes encoding cellulose- or hemicellulose-degrading enzymes in *Blumeria* (left) and *M. oryzae* (right). DMATS, dimethylallyl diphosphate tryptophan synthase; GH1, glycosyl hydrolase.

Downloaded from www.sciencemag.org on December 9, 2010

stage (fig. S3). Their expression is, therefore, unlikely to be detrimental to biotrophic growth.

Although the vast majority of genes encoding enzymes of primary metabolism are retained, notable exceptions include anaerobic fermentation, biosynthesis of glycerol from glycolytic intermediates, and inorganic nitrogen (nitrate) assimilation. These deficiencies are consistent with an exclusively aerobic, parasitic life-style on aerial plant organs, the production of solutes for the generation of osmotic pressure during plant cell wall penetration from triacyl glycerol breakdown, and the assimilation of organic nitrogen in the form of host-derived amino acids.

Filamentous fungi generally produce an array of secondary metabolites, some of which are involved in pathogenesis (*5*). Key enzymes that catalyze their biosynthesis include polyketide synthases (PKSs), modular nonribosomal peptide synthetases (NRPSs), terpene cyclases, and dimethylallyl diphosphate tryptophan synthases (*6*). *Blumeria* encodes only two such proteins (one PKS and one NRPS), the lowest number known in fungi (Fig. 2B and fig. S4). We hypothesize that *Blumeria* synthesizes only one iron siderophore and one simple polyketide, possibly the pigment observed on cleistothecia (fig. S1, G and H). Similar trends are observed in other biotrophs, such as the basidiomycete *Ustilago maydis* and the plant symbiotic fungus *Tuber melanosporum*. Therefore, it appears that biotrophy is associated with a convergent loss of secondary metabolic enzymes. We also noted a marked reduction in genes encoding specific subfamilies of transporters (fig. S5), which typically function in secretion of toxins into the host and extrusion of host defense compounds in necrotrophic fungi (*7*).

Unlike other known plant pathogenic fungi, *Blumeria* has an extremely reduced set of carbohydrate active enzymes devoted to plant cell wall depolymerization (Fig. 2B and fig. S6) (*8*). We found no canonical cellulose-, xylan-, or pectin-degrading enzymes. Other biotrophic phytopathogens, such as *U. maydis* and *Puccinia graminis*, also possess reduced enzyme systems for degradation of the plant cell wall, but both species have predicted cellulases and xylanases. An example of structural proteins lacking in the mildews are the hydrophobins, a class of cell wall proteins that are typically present in fungi (*9*).

We found a massive proliferation of transposable elements (TEs) (table S4). In *Blumeria*, where TEs account for 64% of the genome size, the most abundant families comprise non–long terminal repeat (LTR) retrotransposons lacking LTRs (fig. S7). TEs were evenly distributed throughout the *Blumeria* genome, with no evidence of clustering of particular TEs (fig. S8). Protein-coding genes are typically in small clusters (2 to 10 genes) interspersed between extended stretches of TEs. In all three powdery mildew genomes, genes required for repeat-induced point mutations (RIPs) are absent, whereas all components known to be necessary for mitotic and meiotic silencing are present (table S5). Thus, dysfunctionality of the RIP pathway has probably contributed to genome-size inflation, and extensive retrotransposition (rather than gradual pseudogenization) may account for the observed gene losses and reshuffling. An example of the latter is the otherwise-conserved mating type (*MAT*) locus (*10*). In all other ascomycetes, *MAT* genes are flanked by a conserved cluster of functionally unrelated genes. Although the microsynteny of these genes is retained in the *Blumeria* genome (fig. S9), *MAT* was found on a different supercontig, indicating physical separation.

In addition to >1350 paralog copies of the previously described atypical avirulence genes *AVRk1* and *AVRa10* (*11*), we predicted 248 *Blumeria* proteins with a signal peptide (SP) but lacking any transmembrane domain and BLAST (Basic Local Alignment Search Tool) hit outside the mildews, thus representing candidates for secreted effector proteins (CSEPs) (*12*). The CSEPs have distinctive features (table S6) and show great sequence diversity with few members grouping in small families (Fig. 3A). We noted no obvious clustering of *CSEPs* within the *Blumeria* contigs. Approximately 80% harbor a recently identified N-terminal tripeptide motif, termed "YxC," (*13*), that typically occurs within the first 30 amino acids after the predicted SP cleavage site. Searches in the *E. pisi* and *G. orontii* genomes revealed that the vast majority of *CSEPs* are confined to *Blumeria* (Fig. 3A and table S6). Thus, powdery mildew genomes preferentially harbor species- and/or tribe-specific innovations, which possibly evolved in the context of cospeciation with their plant hosts (*11*). Upon comparison of global gene expression in haustoria (Fig. 1C) and epiphytic structures (Fig. 1D), we observed preferential expression of the majority of the *CSEPs* (79%) in haustoria (Fig. 3B), suggesting they have specific functions in biotrophic pathogenesis (*14*).

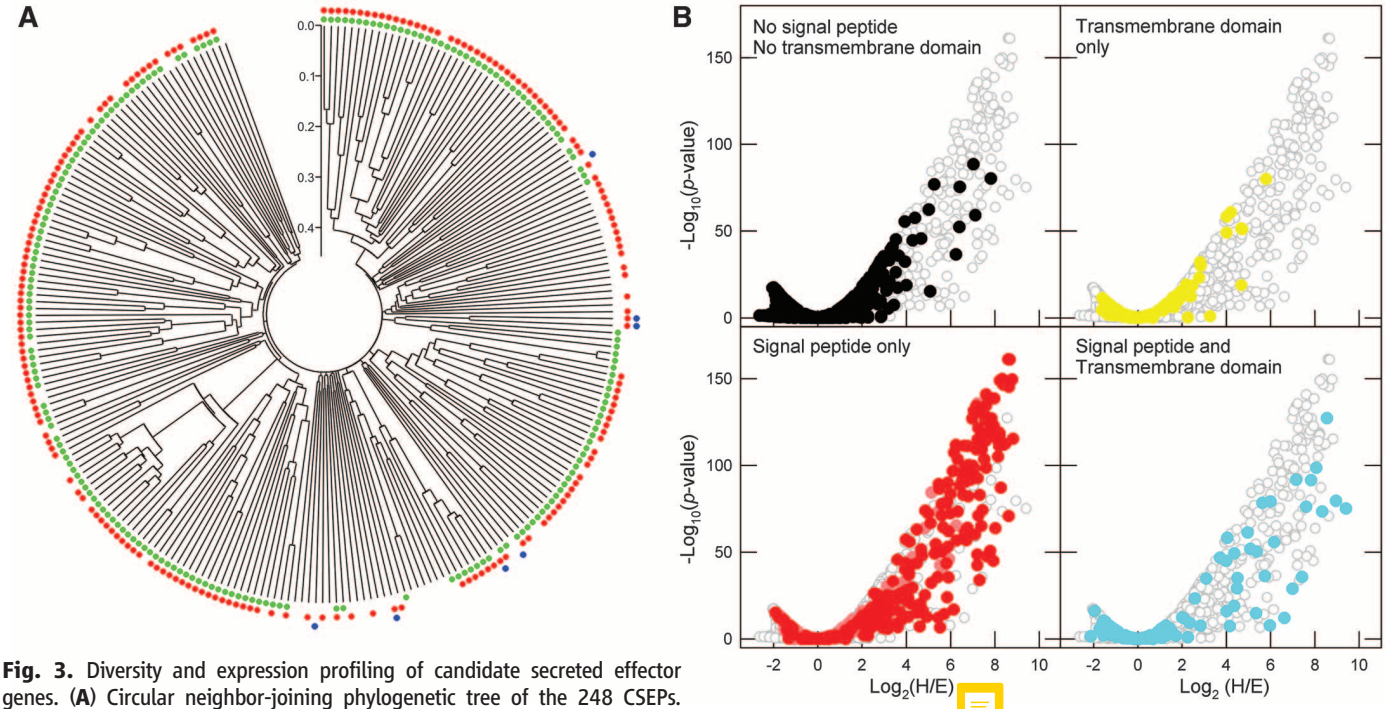

Fig. 3. Diversity and expression profiling of candidate secreted effector genes. (**A**) Circular neighbor-joining phylogenetic tree of the 248 CSEPs. Scale, amino acid substitutions per site; green, CSEPs harboring the YxC motif; blue, CSEPs conserved in *E. pisi* and/or *G. orontii*; red, CSEPs predominantly expressed in haustoria. (**B**) Global gene expression in haustoria (H) versus epiphytic structures (E). Relative abundance of each gene plotted versus the *p*-value as a measure of the statistical significance. White circles, all gene models; black, no SP and no TM domain; yellow, TM only; red, CSEPs; pink, genes with BLASTP hits in the National Center for Biotechnology Information nr database and SP only; light blue, both SP and TM domains.

Downloaded from www.sciencemag.org on December 9, 2010

We detected common genomic hallmarks in the powdery mildews associated with obligate biotrophy. These include gene losses and extensive gene reshuffling correlated with expansion in (retro-)transposon number and genome size. Together, these hallmarks may represent a tradeoff between advantages of increased genetic variation independent of sexual recombination and irreversible deletion of genes dispensable for biotrophy. Hence, their evolution provides a notable example of Dollo's law of evolutionary irreversibility (*15*). This may explain why powdery mildews and possibly other biotrophic parasites became obligate.

References and Notes

1. D. A. Glawe, *Annu. Rev. Phytopathol.* **46**, 27 (2008).
2. Materials and methods are available as supporting material on *Science* Online.
3. J. R. Miller *et al.*, *Bioinformatics* **24**, 2818 (2008).
4. S. Takamatsu, *Mycoscience* **45**, 147 (2004).
5. N. Möbius, C. Hertweck, *Curr. Opin. Plant Biol.* **12**, 390 (2009).
6. J. Collemare, A. Billard, H. U. Böhnert, M. H. Lebrun, *Mycol. Res.* **112**, 207 (2008).
7. J. Glazebrook, *Annu. Rev. Phytopathol.* **43**, 205 (2005).
8. B. L. Cantarel *et al.*, *Nucleic Acids Res.* **37** (database issue), D233 (2009).
9. H. A. B. Wösten, *Annu. Rev. Microbiol.* **55**, 625 (2001).
10. R. Debuchy, B. Turgeon, in *The Mycota I: Growth, Differentiation and Sexuality*, U. Kües, R. Fischer, Eds. (Springer, Berlin, 2006), vol. I, pp. 293–323.
11. S. Sacristán *et al.*, *PLoS ONE* **4**, e7463 (2009).
12. R. Panstruga, P. N. Dodds, *Science* **324**, 748 (2009).
13. D. Godfrey *et al.*, *BMC Genomics* **11**, 317 (2010).
14. A.-M. Catanzariti, P. N. Dodds, G. J. Lawrence, M. A. Ayliffe, J. G. Ellis, *Plant Cell* **18**, 243 (2006).
15. C. R. Marshall, E. C. Raff, R. A. Raff, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12283 (1994).

16. We thank B. Thomas, T. Carver, and D. Grunewald for providing scanning electron micrographs. This work was supported by funds from the Biotechnology and Biological Sciences Research Council [grant BB/E0009831/1], INRA, the BioExploit European Union Framework 6 project, Deutsche Forschungsgemeinschaft (priority program SPP1212), the Leverhulme Trust, and the Max Planck Society. GenBank project identification numbers are as follows: *Blumeria*, 28821; *E. pisi*, 50315; and *G. orontii*, 50317.

Supporting Online Material
www.sciencemag.org/cgi/content/full/330/6010/1543/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
Tables S1 to S6
References

2 July 2010; accepted 1 November 2010
10.1126/science.1194573

Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison

Jan Schirawski,[1,2]* Gertrud Mannhaupt,[1,3] Karin Münch,[1] Thomas Brefort,[1]† Kerstin Schipper,[1] Gunther Doehlemann,[1] Maurizio Di Stasio,[1] Nicole Rössel,[1] Artemio Mendoza-Mendoza,[1]‡ Doris Pester,[1]§ Olaf Müller,[1]|| Britta Winterberg,[1]¶ Elmar Meyer,[1] Hassan Ghareeb,[1]* Theresa Wollenberg,[1]* Martin Münsterkötter,[3] Philip Wong,[3] Mathias Walter,[3] Eva Stukenbrock,[1] Ulrich Güldener,[3] Regine Kahmann[1]#

Biotrophic pathogens, such as the related maize pathogenic fungi *Ustilago maydis* and *Sporisorium reilianum*, establish an intimate relationship with their hosts by secreting protein effectors. Because secreted effectors interacting with plant proteins should rapidly evolve, we identified variable genomic regions by sequencing the genome of *S. reilianum* and comparing it with the *U. maydis* genome. We detected 43 regions of low sequence conservation in otherwise well-conserved syntenic genomes. These regions primarily encode secreted effectors and include previously identified virulence clusters. By deletion analysis in *U. maydis*, we demonstrate a role in virulence for four previously unknown diversity regions. This highlights the power of comparative genomics of closely related species for identification of virulence determinants.

Smut fungi are biotrophic pathogens causing disease in a number of agriculturally important crop plants. *Ustilago maydis* and the related fungus *Sporisorium reilianum* both parasitize maize (*1*, *2*). Their life cycle leading to the infectious form is similar (*2*, *3*); however, shortly after infection *U. maydis* locally induces tumors on all aerial parts of the plant, whereas *S. reilianum* spreads systemically and causes symptoms in male and female inflorescences only (Fig. 1). Both *S. reilianum* and *U. maydis* establish an intimate communication with their host through secreted protein effectors that enable biotrophic development (*3*, *4*). Effector proteins like *U. maydis* Pep1 can suppress plant defense responses (*5*). Additional effector genes were identified in the genome as genes encoding *U. maydis*–specific secreted proteins, most of which are up-regulated during host colonization (*3*). Many of these effector genes are clustered, and deletion of five of these clusters affected virulence in seedlings (*3*). Some cluster genes are induced in specific plant organs, and respective cluster mutants show altered virulence depending on the host tissue infected (*6*). In plant parasitic oomycetes, genes for effector proteins are under diversifying selection and occur in highly flexible genomic regions (*7*). In accordance with this emerging picture of plant-pathogen communication via rapidly evolving effector proteins, we hypothesized that virulence-associated *U. maydis* genes might be identified as genomic regions with high sequence variability in closely related smut species.

To identify regions of high diversity in the *U. maydis* genome, we sequenced the genome of *S. reilianum* strain SRZ2 (*8*). The *S. reilianum* genome assembly covers 97% of the 18.7-Mb genome (*9*). As in *U. maydis* (*3*), the genome is organized in 23 chromosomes, to which 6648 gene models could be assigned after manual annotation. The genomes of *U. maydis* and *S. reilianum* exhibit a remarkable degree of synteny (Fig. 2A) (*10*) despite an average amino acid identity of predicted proteins of only 74.2% (Fig. 2B). Interestingly, some chromosome ends are extended by up to 20 genes in *U. maydis* compared with ends in *S. reilianum*. About 90% of these chromosome end–associated genes do not carry any functional annotation and no enrichment for secreted effectors is evident (fig. S1), whereas the others likely encode enzymes for secondary metabolism (table S1). Because orthologs of these genes are lacking in *S. reilianum*, their presence is likely dispensable for virulence. Compared with an average amino acid identity of 76% for nonsecreted proteins, secreted proteins in both organisms display an average identity of only 62% and are enriched among the weakly conserved proteins (Fig. 2B). This suggests that genes coding for secreted proteins are subject to more rapid evolution.

Manual sequence comparisons of predicted gene models of *S. reilianum* and *U. maydis* led to a reannotation of more than 300 gene models of *U. maydis* (table S2). The *S. reilianum* genome has a 5.7% higher GC content than the *U. maydis* genome and a 5% higher coding potential (table S2). More than 99% of all InterPro (www.ebi.ac.uk/interpro/) domains are equally or close to equally represented in the two genomes, suggesting that the biosynthetic repertoire of both species is comparable. However, *S. reilianum* contains three putative RNA-dependent RNA polymerase genes (table S3). A search for other components (*11*, *12*) of a putative RNA interference (RNAi) machinery in *S. reilianum* identified homologs of *dicer* and *argonaute* (table S3). These genes all lie in highly syntenic regions (*10*); however, the corresponding intergenic regions in *U. maydis* lack traces of the

[1]Department of Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch Straße 10, 35043 Marburg, Germany. [2]Department for Molecular Biology of Plant-Microbe Interactions, Albrecht-von-Haller Institute for Plant Sciences, Georg-August-Universität Göttingen, Untere Karspüle 2, 37073 Göttingen, Germany. [3]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.

*Present address: Department for Molecular Biology of Plant-Microbe Interactions, Albrecht-von-Haller Institute for Plant Sciences, Georg-August-Universität Göttingen, Untere Karspüle 2, 37073 Göttingen, Germany.
†Present address: Febit Biomed GmbH, Im Neuenheimer Feld 519, 69120 Heidelberg, Germany.
‡Present address: Bio-Protection Research Centre, Post Office Box 84, Lincoln University, Lincoln 7647, New Zealand.
§Present address: Institute for Plant Health, Austrian Agency for Health and Food Safety (AGES), Spargelfeldstraße 191, 1226 Wien, Austria.
||Present address: Roy J. Carver Center for Comparative Genomics, University of Iowa, Department of Biological Sciences, 303 Biology Building, Iowa City, IA 52242, USA.
¶Present address: Research School of Biology, Building 41, Linnaeus Way, Australian National University, Canberra, ACT 0200, Australia.
#To whom correspondence should be addressed. E-mail: Kahmann@mpi-marburg.mpg.de

Downloaded from www.sciencemag.org on December 9, 2010