ogies in a $LiFePO_4$ nanorod during the first-order phase transition to $FePO_4$. The anisotropy of Li diffusion and the magnitude of misfit strain (~7% volume strain) have been explicitly included, and the driving force for the transformation is varied via the electrical overpotential (*13*). At a moderate overpotential of 25 mV (see the figure, panel A), stress relaxation causes $FePO_4$ to grow along the [100] longitudinal direction, analogous to that of Huang *et al*., even though Li diffusion is fastest normal to this direction. However, at a higher overpotential of 100 mV, the influence of strain energy is overcome and lateral Li diffusion dominates the transformation morphology (see the figure, panel B).

Thus, interactions among stress, transport anisotropy, and the magnitude of the driving force (among other variables) may influence phase transformation morphology in nanowire electrodes. Other unanswered questions include the effect of electrolyte distribution: Is the one-dimensional transformation in SnO_2 facilitated by having a thin wetted layer of electrolyte, and would results differ for a "flooded" electrolyte battery? And could a competing radial reaction morphology also preserve the nanowire morphology? Regardless, the results presented by Huang *et al*. are testimony to the power of direct observation in electrochemical materials science, and they illustrate a previously unrecognized mode of reaction in battery electrodes. The results should stimulate others to consider analogous experiments and mechanisms in other storage materials, and should contribute to the design of nanoscale electrodes that fully exploit the potential of ultrahigh-capacity storage materials.

References

1. J. Y. Huang *et al*., *Science* **330**, 1515 (2010).
2. P. Poizot, S. Laruelle, S. Grugeon, L. Dupont, J.-M. Tarascon, *Nature* **407**, 496 (2000).
3. H. Li, P. Balaya, J. Maier, *J. Electrochem. Soc.* **151**, A1878 (2004).
4. G. G. Amatucci, N. Pereira, *J. Fluor. Chem.* **128**, 243 (2007).
5. Y. Idota, T. Kubota, A. Matsufuji, Y. Maekawa, T. Miyasaka, *Science* **276**, 1395 (1997).
6. P. Limthongkul, Y.-I. Jang, N. J. Dudney, Y.-M. Chiang, *Acta Mater.* **51**, 1103 (2003).
7. M. N. Obrovac, L. J. Krause, *J. Electrochem. Soc.* **154**, A103 (2007).
8. C. K. Chan *et al*., *Nat. Nanotechnol.* **3**, 31 (2008).
9. G. Chen, X. Song, T. Richardson, *Electrochem. Solid-State Lett.* **9**, A295 (2006).
10. N. Meethong, H.-Y. S. Huang, S. A. Speakman, W. C. Carter, Y.-M. Chiang, *Adv. Funct. Mater.* **17**, 1115 (2007).
11. N. Meethong, Y.-H. Kao, W. C. Carter, Y.-M. Chiang, *Chem. Mater.* **22**, 1088 (2010).
12. M. Tang, W. C. Carter, Y.-M. Chiang, *Annu. Rev. Mater. Res.* **40**, 501 (2010).
13. Y.-H. Kao *et al*., *Chem. Mater.* **22**, 5845 (2010).

10.1126/science.1198591

PLANT SCIENCE

Genome Evolution in Plant Pathogens

Pathogen genes that shut down specific host plant immune responses are highly divergent and have evolved rapidly to accommodate adaptation.

Peter N. Dodds

Food security is of global importance and crop diseases caused by plant pathogens are a major constraint to agriculture worldwide. Many of these pathogens have a similar biotrophic life stage during which they contact host cells and secrete effector proteins that alter plant responses to infection (*1*). In this issue, comparative genomics studies of closely related pathogen species by Raffaele *et al*. on page 1540 (*2*), Baxter *et al*. on page 1549 (*3*), Spanu *et al*. on page 1543 (*4*), and Schirawski *et al*. on page 1546 (*5*) reveal that such effector proteins evolve rapidly and that their diversity contributes to host range and parasite speciation.

Biotrophic infection strategies have evolved independently in diverse lineages of plant pathogens. These include fungus-like parasites (oomycetes) from the kingdom Stramenopila, such as the destructive potato blight pathogen *Phytophthora infestans* (agent of the Irish potato famine), fungi such as powdery mildews (ascomycetes), and rust and smut fungi (basidiomycetes). These pathogens form specialized hyphae (called haustoria) that penetrate the plant cell wall and allow nutrient uptake from host tissue (*6*). These structures also secrete large repertoires of effector proteins that enter host cells and manipulate defense responses and cellular metabolism. Many oomycete effectors require the short amino acid motif RxLR (Arg, any amino acid, Leu, Arg) for entry into plant cells (*7*), independently of other pathogen machinery (*8*, *9*). Some fungal effectors also enter host plant cells (*10*, *11*), although they lack clearly conserved peptide motifs.

Agricultural threats. Many pathogens, such as *Phytophthora infestans*, target major food crops, such as potato (shown). The genes involved in host-pathogen interactions are highly diversified among related *Phytophthora* species that attack different plants.

Raffaele *et al*. compared the genomes of four very closely related *Phytophthora* species that infect quite different host plant species (see the figure). The evolution of these pathogens therefore involved relatively recent shifts in host range, followed by specialization to the new hosts. They found most pathogen genes and genome regions to be highly conserved, but genes involved in host-pathogen interaction appear highly diversified, especially the predicted RxLR-containing effectors. Most of these genes are located in gene-sparse, transposon-rich genome regions, suggesting that these features allow rapid evolution of effector loci after host changes. Genes involved in chromatin modification are also located in these regions and show extensive variation, suggesting that epigenetic regulation of gene expression also contributes to adaptation following host shifts.

Despite having a biotrophic life stage, *Phytophthora* species subsequently kill the infected parts of the plant but continue to feed on the dead plant tissue (and can be cultured on simple medium). By contrast, the related oomycete *Hyaloperonospora arabidopsidis*, which is a pathogen of the model plant *Arabidopsis thaliana*, is exclusively biotrophic and cannot be grown in culture. This pathogen is believed to have evolved from a *Phytophthora*-like hemibiotrophic ancestor. Baxter *et al*. found that its genome contains a unique set of diversified RxLR-containing effectors but has lost many of the hydrolytic

Division of Plant Industry, Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT 2614, Australia. E-mail: peter.dodds@csiro.au

enzymes that *Phytophthora* species use to digest host cell walls, as well as many of the genes that induce host cell death. The reduction in these protein classes is inferred as resulting from selection for "stealth," allowing *H. arabidopsidis* to avoid triggering host defense responses during its extended biotrophic interaction. It is also deficient in metabolic processes that are shared by free-living organisms, such as the lack of nitrate and sulfate assimilation enzymes. Growing exclusively in living plant leaves, *H. arabidopsidis* relies on access to reduced nitrogen and sulfur from host cells.

Likewise, Spanu *et al*. found that the genomes of three species of fungal powdery mildew pathogens (also obligate biotrophs) are deficient in several classes of conserved primary and secondary metabolism genes. These include the nitrate and sulfate assimilation pathways and plant cell wall hydrolytic enzymes, suggesting that the independent evolution to obligate biotrophy that has occurred in the fungal and oomycete lineages involved convergent adaptation to specialize in the exclusively parasitic life-style in plant leaves. The metabolic deficiencies may provide clues for culturing these pathogens in vitro, which so far has proved difficult and hampered research on these organisms. The powdery mildew genomes do not encode RxLR-containing effectors, but encode a unique class of secreted proteins with another conserved amino acid motif, YxC (Tyr, any amino acid, Cys). These genes are highly diverse among the three species, which infect very different host plants, suggesting that most of these effectors are associated with host species–specific adaptation.

In an intriguing twist on these studies, Schirawski *et al*. compared the genome of *Sporisorium reilianum* to that of the related fungal pathogen, *Ustilago maydis* (*12*), both of which infect maize. Most of the predicted secreted effectors are common to both species, but show much higher divergence than the rest of the genome. Thus, even within a host species, selection imposed by the host immune system or selection that targets different host processes can lead to rapid diversification of pathogen effectors. Most effector-encoding genes are located in small clusters in the genomes of the *U. maydis* and *S. reilianum*, and mutational analysis of these clusters has confirmed their important roles in infection (*5*, *12*).

These studies highlight the value of comparative genomics in identifying important virulence genes with host-specific functions. The challenge now is to determine how the effector proteins that these genes encode turn the host cell to their own purposes. What are the specific targets of the effectors in the host and how do they contribute to disease pathogenesis? Understanding the molecular details of the biotrophic lifestyle and plant-microbe coevolution should lead to innovative disease control measures. Some effectors are recognized by components of the plant immune system, a characteristic that is exploited in plant breeding to improve disease resistance of crops. Thus, identifying effectors with important and nonredundant virulence functions that constrain their evolution may allow deployment of more durable resistance sources to safeguard world food production. The recent devastating impact of a new stem rust strain on wheat production in Africa (*13*) emphasizes the critical importance of developing lasting effective strategies to protect agricultural production from disease threats.

References

1. R. Panstruga, P. N. Dodds, *Science* **324**, 748 (2009).
2. S. Raffaele *et al*., *Science* **330**, 1540 (2010).
3. L. Baxter *et al., Science* **330**, 1549 (2010).
4. P. D. Spanu *et al., Science* **330**, 1543 (2010).
5. J. Schirawski *et al., Science* **330**, 1546 (2010).
6. R. T. Voegele, K. Mendgen, *New Phytol.* **159**, 93 (2003).
7. S. C. Whisson *et al*., *Nature* **450**, 115 (2007).
8. D. Dou *et al*., *Plant Cell* **20**, 1930 (2008).
9. S. D. Kale *et al*., *Cell* **142**, 284 (2010).
10. M. Rafiqi *et al*., *Plant Cell* **22**, 2017 (2010).
11. C. H. Khang *et al*., *Plant Cell* **22**, 1388 (2010).
12. J. Kämper *et al*., *Nature* **444**, 97 (2006).
13. R. P. Singh *et al*., *Adv. Agron.* **98**, 271 (2008).

10.1126/science.1200245

MICROBIOLOGY

Sex and Sacrifice

A single genetic locus determines the three sexes of slime mold amoebae.

Richard H. Kessin

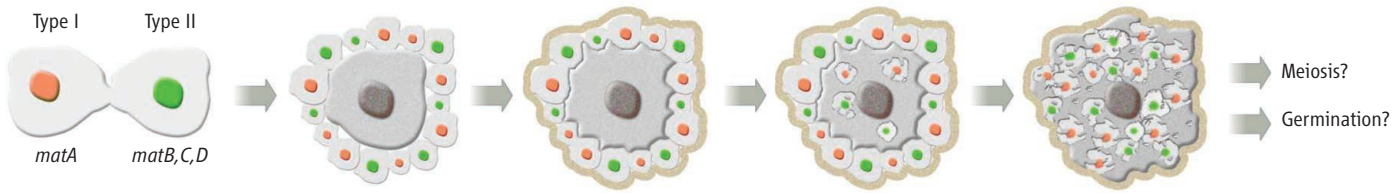
The soil amoeba *Dictyostelium discoideum*, commonly known as a slime mold, has an asexual reproduction cycle that has made it a well-studied model organism. When starved, it creates streams of cells that use chemical cues (chemotaxis) to form aggregates, then migrating "slugs," and finally spore-bearing fruiting bodies. The amoeba also has a less studied sexual cycle and is known for having three sexes, known as mating types I, II, and III. On page 1533 of this issue, Bloomfield *et al*. (*1*) put the sexual part of the slime mold's life cycle on a solid molecular footing.

Why is the sexual cycle of a slime mold of interest at all? The answer, in part, involves the macrocyst, a remarkable structure formed by sexual amoebae that has become an object of frustration and fascination for biologists. Consider this: Two haploid amoebae in the soil meet, and if they have compatible mating types, they fuse (*2*) (see the figure). The resulting zygote forms a so-called giant cell. It begins to eat up amoebae in its neighborhood, secreting cyclic adenosine monophosphate (cAMP), which attracts the chemotactic amoebae, siren-like, to their doom (*3*). (This is the same mechanism used to aggregate cells during fruiting body construction.) The chemotactic cells form an aggregate around the giant cell and synthesize a primary cell wall. The giant cell in the center of the aggregate gradually chews its way to the periphery, killing and consuming hundreds of amoebae (*4*). It is one of the great acts of phagocytosis in biology. The resulting structure—the macrocyst—is now 500 to 1000 times the size of a single cell and has three cellulose walls.

Department of Pathology and Cell Biology, College of Physicians and Surgeons of Columbia University, New York, NY 10032, USA. E-mail: rhk2@columbia.edu

The genetics underlying macrocyst formation, however, were poorly understood. Researchers had problems germinating macrocysts in the laboratory and identifying useful recombinant strains. The strains used in earlier experiments were separate wild isolates and may have evolved incompatibility mechanisms. In contrast, Bloomfield *et al*. produced strains that are genetically identical except at the mating-type locus, which may help in meiotic studies.

Bloomfield *et al*. reasoned that they could locate the mating-type locus of *D. discoideum* by searching for genes present in one mating type but absent (or highly diverged) in the others. They were right. Using DNA microarrays to analyze 10 wild strains, they identified a locus called *matA* that is present in all type I strains and has the coding capacity to make a small protein that would lack homology to any known protein. It has no charac-



Macrocyst formation in *Dictyostelium*. Under certain culture conditions, haploid amoebae with different mating-type alleles (type I and type II shown) fuse to form a zygote called a giant cell. This cell ingests its neighbors and secretes cAMP to attract other amoebae into an aggregate with the giant cell inside. The aggregated cells secrete a primary wall while the giant cell gradually consumes them. Mature macrocysts have three cellulose walls (not shown) and can have a diameter of more than 100 μm.

teristic motifs—nothing to give a clue to its function, except for the fact that it is charged and probably is a cytosolic product. A strain carrying a deletion of *matA* is sterile and will not fuse with type II cells. Restore the gene and it will fuse.

What of the type II locus? Because the genes flanking the *matA* locus are present in strains of all mating types, the authors could recover the type II locus via the polymerase chain reaction (PCR). The type II locus contains three genes: *matB*, which is homologous to the type I *matA* but has diverged; *matC*, which has no homology to other known genes; and *matD*, which encodes a large secreted protein that may be anchored by glycosylphosphatidylinositol (GPI). Type II strains from all over the world have these three genes and they have hardly diverged at all. What happens when the three type II genes are put into the type I knockout? Happily, the amoeba switches to type II, which is to say it mates with type I strains to form macrocysts. The loci are sufficient to determine mating type.

There is a third mating type, type III, which is capable of mating with types I or II but not with itself. Rare homothallic strains also exist and the authors have looked at these as well. Type III strains contain two genes homologous to *matC* and *matD* of the type II strain, and these have been named *matS* and *matT*.

By transforming the *matA* null strain with individual genes, the authors have begun to ask which of these genes is essential for the formation of the macrocyst. Do these macrocysts have the capacity to carry out meiosis and provide researchers with a proper tetrad of recombinants? We will see. For *Dictyostelium*, the identification of such a capability, although arguably decades late for geneticists (more if you work on peas or *Drosophila*), would still be wonderfully useful. Sequencing of single-nucleotide polymorphisms has yielded evidence that wild strains have recombined frequently (*5*). There is a parasexual system of genetic analysis (which involves nonsexual mechanisms for transferring genetic material without meiosis), but it does not have the flexibility or power of meiotic genetics.

The enormous phagocytic capacity of the giant cell offers interesting problems. How do the haploid amoebae suddenly get defined as foreign and targeted for phagocytosis? How does the lysosomal system suddenly ramp up? Does the bacterium *Legionella pneumophila*, known to be capable of derailing phagosome-lysosome fusion in an amoeba (*6*), stand a chance in a giant cell? How does a genome designed for an amoeba handle a cell that is now 1000 times as large? Perhaps with these genes in hand, the formation of macrocysts can be synchronized and the essential genes identified. There is evidence that these genes will overlap with the genes that control fruiting body formation, for which the macrocyst is probably an evolutionary precursor (*7*).

The evolutionary biologists who have recently taken an interest in *Dictyostelium* should prick up their ears. How does a system evolve in which more than 99% of cells perish, as opposed to 20% during the fruiting body version of development? Imagine a population of 999 type I amoebae and one type II amoeba—a single giant cell will form, and it will consume the 998 other amoebae. If the resulting macrocyst then germinates and produces meiotic offspring, the frequency of the type II gene will have gone from 0.001 to 0.5. There could be a balancing of mating types, or perhaps mutant amoebae would be selected that decline the chemotactic call of the giant cell. What was a fragmented literature has now gained a solid footing.

References

1. G. Bloomfield, J. Skelton, A. Ivens, Y. Tanaka, R. R. Kay, *Science* **330**, 1533 (2010).
2. J. C. Blaskovics, K. B. Raper, *Biol. Bull.* **113**, 58 (1957).
3. D. H. O'Day, *Can. J. Microbiol.* **25**, 1416 (1979).
4. M. A. Wallace, K. B. Raper, *J. Gen. Microbiol.* **113**, 327 (1979).
5. J. M. Flowers *et al.*, *PLoS Genet.* **6**, e1001013 (2010).
6. G. P. Otto *et al.*, *Mol. Microbiol.* **51**, 63 (2004).
7. R. H. Kessin, Dictyostelium*: Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge Univ. Press, Cambridge, 2001).

10.1126/science.1199899

GENETICS

First-Class Control of HIV-1

Andrew J. McMichael[1] and E. Yvonne Jones[2]

Genome-wide association studies reveal amino acids of the major histocompatibility complex that associate with the rate of progression to AIDS.

The role of infectious disease in driving human genetic variations (polymorphisms) was first clearly espoused by J. B. S. Haldane in 1949 (*1*). Once the major histocompatibility complex (MHC) was established as the most polymorphic mammalian genetic system, searches began for human MHC [or human leukocyte antigen (HLA)] associations with infectious disease resistance. Such findings have been rare, though, possibly because susceptibility genes have been deselected over evolution. But the appearance of completely new infections caused by viruses, such as HIV-1, has opened opportunities to look at such selection in the MHC as it happens. On page 1551 of this issue, the International HIV Controllers Study (*2*), demonstrates the central role of HLA class I in controlling HIV-1 infection.

Earlier studies had demonstrated the influence of HLA type on slow versus rapid AIDS progression (*3*). Protection is thought to be mediated by strong T cell (CD8 subtype) responses to viral peptides presented

[1]Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OS3 9DS, UK. [2]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. E-mail: andrew.mcmichael@ndm.ox.ac.uk; yvonne@strubi.ox.ac.uk

by N-WASP (Fig. 4B and fig. S7, C to E). The results suggest that nebulin modules and the N-WASP WH2 domains cooperate to nucleate an unbranched actin filament. Then the actin filament might elongate along the nebulin modules from the Z band toward the center of a sarcomere (fig. S8A).

We assessed the requirement of N-WASP for IGF-1–induced actin filament formation in myofibrils and muscle hypertrophy by RNA interference (RNAi) (fig. S9). EGFP–α-actin coexpressed with control small interfering RNA (siRNA) in the fasted mouse muscle was diffusely distributed without IGF-1 stimulation but located to the Z bands and thin filaments within 2 hours after the stimulation (Fig. 4C and fig. S9D). In contrast, EGFP–α-actin coexpressed with siRNA1 or 2 (fig. S11A) remained diffusely distributed after the stimulation. Therefore, N-WASP is indispensable for the recruitment of α-actin to the Z bands and for myofibrillar actin filament formation.

IGF-1 administration to mice caused muscle hypertrophy owing to the increase in myofiber volume. The expression of siRNA1 or 2 reduced the cross-sectional area of the myofibers regardless of IGF-1 administration (Fig. 4D and figs. S10 and S11). Thus, N-WASP plays essential roles in both age-dependent natural hypertrophy and administered IGF-1–induced hypertrophy. N-WASP seems to participate in myofiber hypertrophy by inducing myofibrillar actin filament formation through the nebulin–N-WASP complex. This notion is consistent with the observation that *Neb*-deficient mice develop a muscle atrophy–like phenotype (*15*, *16*).

We elucidated the signaling of IGF-1–induced myofibrillar actin filament formation from the Z bands (fig. S8B) and a mechanism of actin nucleation [supporting online material (SOM) text]. The Neb–N-WASP complex formed by the signaling can explain actin filament formation arising from the Z bands. These findings may provide insights into the mechanisms of muscular diseases, such as nemaline myopathy, caused by *NEB* gene mutations (*17*). The actin filament formation together with myosin filament assembly, which might also induced by IGF-1 signaling, results in myofibrillogenesis required for muscle maturation and hypertrophy.

References and Notes

1. D. J. Glass, *Int. J. Biochem. Cell Biol.* **37**, 1974 (2005).
2. F. Mourkioti, N. Rosenthal, *Trends Immunol.* **26**, 535 (2005).
3. T. Takenawa, S. Suetsugu, *Nat. Rev. Mol. Cell Biol.* **8**, 37 (2007).
4. T. D. Pollard, G. G. Borisy, *Cell* **112**, 453 (2003).
5. M. A. Chesarone, B. L. Goode, *Curr. Opin. Cell Biol.* **21**, 28 (2009).
6. B. Qualmann, M. M. Kessels, *Trends Cell Biol.* **19**, 276 (2009).
7. R. Dominguez, *Crit. Rev. Biochem. Mol. Biol.* **44**, 351 (2009).
8. S. Labeit, B. Kolmerer, *J. Mol. Biol.* **248**, 308 (1995).
9. K. Ojima *et al.*, *J. Cell Biol.* **150**, 553 (2000).
10. A. S. McElhinny, S. T. Kazmierski, S. Labeit, C. C. Gregorio, *Trends Cardiovasc. Med.* **13**, 195 (2003).
11. A. Kontrogianni-Konstantopoulos, M. A. Ackermann, A. L. Bowman, S. V. Yap, R. J. Bloch, *Physiol. Rev.* **89**, 1217 (2009).
12. B. D. Manning, L. C. Cantley, *Cell* **129**, 1261 (2007).
13. E. Paunola, P. K. Mattila, P. Lappalainen, *FEBS Lett.* **513**, 92 (2002).
14. M. Pfuhl, S. J. Winder, M. A. Castiglione Morelli, S. Labeit, A. Pastore, *J. Mol. Biol.* **257**, 367 (1996).
15. C. C. Witt *et al.*, *EMBO J.* **25**, 3843 (2006).
16. M.-L. Bang *et al.*, *J. Cell Biol.* **173**, 905 (2006).
17. D. Sanoudou, A. H. Beggs, *Trends Mol. Med.* **7**, 362 (2001).

18. We thank N. Watanabe and T. Itoh for comments. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and by the Research Grants (17A-10 and 20B-13) for Nervous and Mental Disorders from the Ministry of Health, Labor, and Welfare of Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6010/1536/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
References
Movies S1 to S10

14 September 2010; accepted 1 November 2010
10.1126/science.1197767

Genome Evolution Following Host Jumps in the Irish Potato Famine Pathogen Lineage

Sylvain Raffaele,[1]* Rhys A. Farrer,[1]*† Liliana M. Cano,[1]* David J. Studholme,[1]‡ Daniel MacLean,[1] Marco Thines,[1,2,3] Rays H. Y. Jiang,[4] Michael C. Zody,[4] Sridhara G. Kunjeti,[5] Nicole M. Donofrio,[5] Blake C. Meyers,[5] Chad Nusbaum,[4] Sophien Kamoun[1]§

Many plant pathogens, including those in the lineage of the Irish potato famine organism *Phytophthora infestans*, evolve by host jumps followed by specialization. However, how host jumps affect genome evolution remains largely unknown. To determine the patterns of sequence variation in the *P. infestans* lineage, we resequenced six genomes of four sister species. This revealed uneven evolutionary rates across genomes with genes in repeat-rich regions showing higher rates of structural polymorphisms and positive selection. These loci are enriched in genes induced in planta, implicating host adaptation in genome evolution. Unexpectedly, genes involved in epigenetic processes formed another class of rapidly evolving residents of the gene-sparse regions. These results demonstrate that dynamic repeat-rich genome compartments underpin accelerated gene evolution following host jumps in this pathogen lineage.

Phytophthora infestans is an economically important specialized pathogen that causes the destructive late blight disease on *Solanum* plants, including potato and tomato. In central Mexico, *P. infestans* naturally co-occurs with two extremely closely related species, *Phytophthora ipomoeae* and *Phytophthora mirabilis*, that specifically infect plants as diverse as morning glory (*Ipomoea longipedunculata*) and four-o'clock (*Mirabilis jalapa*), respectively. Elsewhere in North America, a fourth related species, *Phytophthora phaseoli*, is a pathogen of lima beans (*Phaseolus lunatus*). Altogether these four *Phytophthora* species form a very tight clade of pathogen species that share ~99.9% identity in their ribosomal DNA internal transcribed spacer regions (*1*). Phylogenetic inferences clearly indicate that species in this *Phytophthora* clade 1c [nomenclature of (*2*)] evolved through host jumps followed by adaptive specialization on plants belonging to four different botanical families (*2*, *3*). Adaptation to these host plants most likely involves mutations in the hundreds of disease effector genes that populate gene-poor and repeat-rich regions of the 240–megabase pair genome of *P. infestans* (*4*). However, comparative genome analyses of specialized sister species of plant pathogens have not been reported, and the full extent to which host adaptation affects genome evolution remains unknown.

To determine patterns of sequence variation in a phylogenetically defined species cluster of host-specific plant pathogens, we generated Illumina reads for six genomes representing the four clade 1c species. We included the previously sequenced *P. infestans* strain T30-4 (*4*) to optimize bioinformatic parameters (figs. S1 to S3) (*5*). To estimate gene copy number variation (CNV) in the five resequenced genomes relative to T30-4, we used average read depth per gene and GC content correction (*5*) (fig. S4). After GC content correction (*6*), average read depth provided a good estimate of gene copy number in T30-4 (fig. S5). In the other genomes, we detected 3975 CNV events in coding genes, among which there are 1046 deletion events (Fig. 1).

[1]The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK. [2]Biodiversity and Climate Research Centre BiK-F, Senckenberganlage 25, D-60325 Frankfurt (Main), Germany. [3]Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, Johann Wolfgang Goethe University, Siesmayerstraße 70, D-60323 Frankfurt (Main), Germany. [4]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02141, USA. [5]University of Delaware, Newark, DE 19716, USA.

*These authors contributed equally to this work.
†Present address: Imperial College London, London SW7 2AZ, UK.
‡Present address: Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK.
§To whom correspondence should be addressed. E-mail: sophien.kamoun@tsl.ac.uk

In total, we identified 746,744 nonredundant coding sequence single-nucleotide polymorphisms (SNPs) in the resequenced strains (Fig. 1). We calculated rates of synonymous (dS) and nonsynonymous (dN) substitutions for every gene (*5*, *7*). Average dS divergence rates relative to *P. infestans* T30-4 were consistent with previously reported species phylogeny (Fig. 1) (*2*). We detected a total of 2572 genes (14.2% of the whole genome) with dN/dS ratios >1 indicative of positive selection in the clade 1c strains, with the highest number in *P. mirabilis* (1004 genes) (fig. S6). A high proportion of genes annotated as effector genes show signatures of positive selection (300 out of 796) (fig. S6). This supports previous observations that effector genes are under strong positive selection in oomycetes (*8*–*10*).

Haas *et al*. (*4*) reported that the *P. infestans* genome experienced a repeat-driven expansion relative to distantly related *Phytophthora* spp. and shows an unusual discontinuous distribution of gene density. Disease effector genes localize to expanded, repeat-rich and gene-sparse regions of the genome, in contrast to core ortholog genes, which occupy repeat-poor and gene-dense regions (*4*). We exploited our sequence data to determine the extent to which genomic regions with distinct architecture evolved at different rates. We used statistical tests (table S1) and random sampling

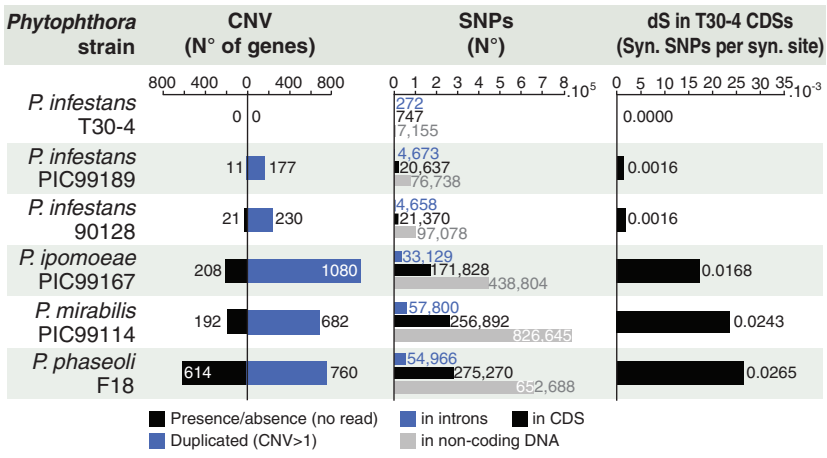


Fig. 1. Summary of genome sequences obtained for *Phytophthora* clade 1c species. Six strains representing four species were analyzed. *P. infestans* T30-4 previously sequenced by Haas *et al*. (*4*) was included for quality control. CDS, coding sequence; CNV, copy number variation; SNP, single-nucleotide polymorphism; syn., synonymous.

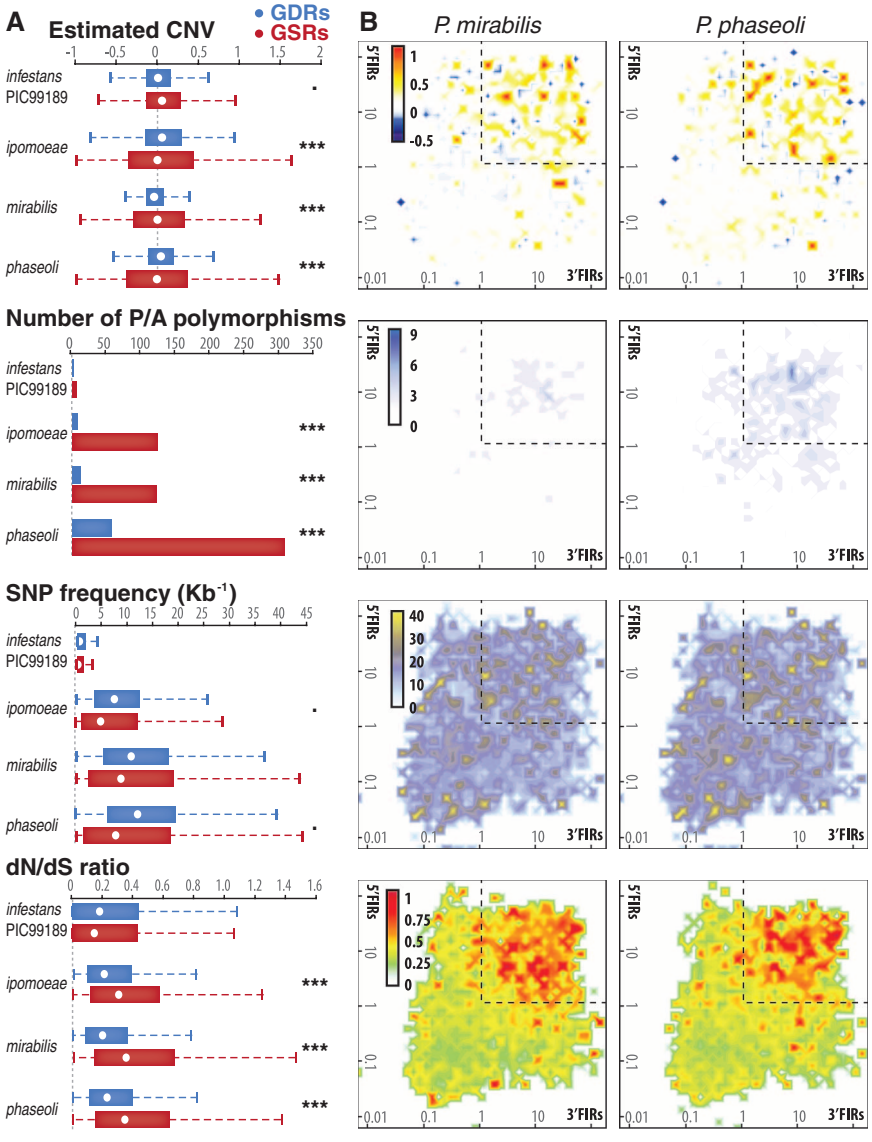


Fig. 2. The two-speed genome of *P. infestans*. (**A**) Distribution of copy number variation (CNV), presence/absence (P/A) and single-nucleotide polymorphisms (SNP), and dN/dS in genes from gene-dense regions (GDRs) and gene-sparse regions (GSRs). Statistical significance was assessed by unpaired *t* test assuming unequal variance (CNV, dN/dS); assuming equal variance (SNP frequency); or by Fisher's exact test (P/A) ($\bullet P < 0.1$; $***P < 10^{-4}$). Whiskers show first value outside 1.5 times the interquartile range. (**B**) Distribution of polymorphism in *P. mirabilis* and *P. phaseoli* according to local gene density (measured as length of 5′ and 3′ flanking intergenic regions, FIRs). The number of genes (P/A polymorphisms) or average values (CNV, SNP, dN/dS) associated with genes in each bin are shown as a color-coded heat map.

(table S2) to determine the significance of differences in CNV, presence/absence polymorphisms, SNP frequency, and dN/dS values in genes located in gene-dense versus gene-sparse regions (*5*) (fig. S7 and table S3). Although averages of gene copy numbers were similar in both regions, significantly higher frequency of CNV and gain/loss were observed in genes located in the repeat-rich regions (Fig. 2A and fig. S7). Notably, presence/absence polymorphisms were 13 times as abundant in the gene-sparse compared to the gene-dense regions. In addition, even though SNP frequency was similar across the genomes, average dN/dS was significantly higher in gene-sparse regions, indicating more genes with signatures of positive selection (Fig. 2A). Indeed, 23% of the genes in the gene-sparse regions showed dN/dS > 1 in at least one of the resequenced genomes compared to only 11.5% of genes in the gene-dense regions. In total, 44.6% of the genes in the gene-sparse regions showed signatures of rapid evolution (deletion, duplication, or dN/dS > 1) compared to only 14.7% of the remaining genes. The uneven distribution in gene density in the *P. infestans* genome can be visualized with plots of two-dimensional bins of 5′ and 3′ flanking intergenic region (FIR) lengths (*4*). We adapted the plots to illustrate the relationships between gene density and polymorphism and confirmed the increased rates in the gene-sparse regions (Fig. 2B and fig. S8). We conclude that different regions of the examined genomes evolved at markedly different rates, with the gene-sparse, repeat-rich regions experiencing accelerated rates of evolution.

To gain insights into the functional basis of the uneven evolutionary rates detected in the gene-sparse versus gene-dense regions of the clade 1c species, we plotted genome-wide microarray expression data on the FIR length maps (fig. S9) (*4*). Gene-dense regions were enriched in genes induced in sporangia, the asexual spores that are produced by all *Phytophthora* species. In marked contrast, distribution patterns of genes induced during preinfection and infection stages indicate enrichment in genes located in gene-sparse loci (fig. S9). χ^2 tests revealed that the relationships between gene density (FIR length) and patterns of gene expression are significant (fig. S9 and table S3). We conclude that the gene-sparse, repeat-rich regions are highly enriched in genes induced in planta, therefore implicating host adaptation in genome evolution.

To assign biological functions to genes with accelerated rates of evolution that populate the gene-sparse, repeat-rich regions, we performed Markov clustering on the predicted proteome of *P. infestans* and implemented gene ontology mapping. Protein families (tribes) significantly enriched or deficient in genes that locate to gene-sparse regions or are rapidly evolving were identified with Fisher's exact test. In total, 811 tribes with five or more proteins were generated (44% of proteome) (figs. S10 and S11). Of these, 163 tribes were statistically enriched in genes from gene-sparse regions (Fig. 3A and fig. S12), 123 tribes were enriched in fast-evolving genes (fig. S12), and 65 tribes were enriched for both criteria (Fig. 3B and fig. S12). As expected, several of these tribes (19 out of 65) consist of effector families (*4*, *11*–*13*) (table S4). Other notable tribes include genes encoding various enzymes such as cell wall hydrolases and proteins related to epigenetic maintenance (Fig. 3B and table S4). Surprisingly, tribes annotated as histone and ribosomal RNA (rRNA) methyltransferases were particularly rich in genes located in gene-sparse regions and exhibiting presence/absence polymorphisms (table S4 and figs. S13 and S14). Several genes encoding DOT1-like and SET domain histone methyltransferases and SpoU-like rRNA methyltransferases are exceptional among genes involved in epigenetic maintenance for residing largely in gene-sparse regions and showing high rates of polymorphism (fig. S15).

Our study demonstrates that highly dynamic genome compartments enriched in noncoding sequences underpin accelerated gene evolution following host jumps. Gene-sparse regions that drive the extremely uneven architecture of the *P. infestans* genome are highly enriched in plant-induced genes, particularly effectors, therefore implicating host adaptation as a driving force of genome evolution in this lineage. In addition, we unexpectedly identified several genes involved in epigenetic processes, notably histone methyltransferases, as rapidly evolving residents of the gene-sparse regions. Histone methylation indirectly modulates gene expression in various eukaryotes (*14*, *15*) and could underlie concerted and heritable gene induction patterns through long-range remodeling of chromatin structure (*16*). Histone acetylation and methylation are thought to be key regulators of gene expression in

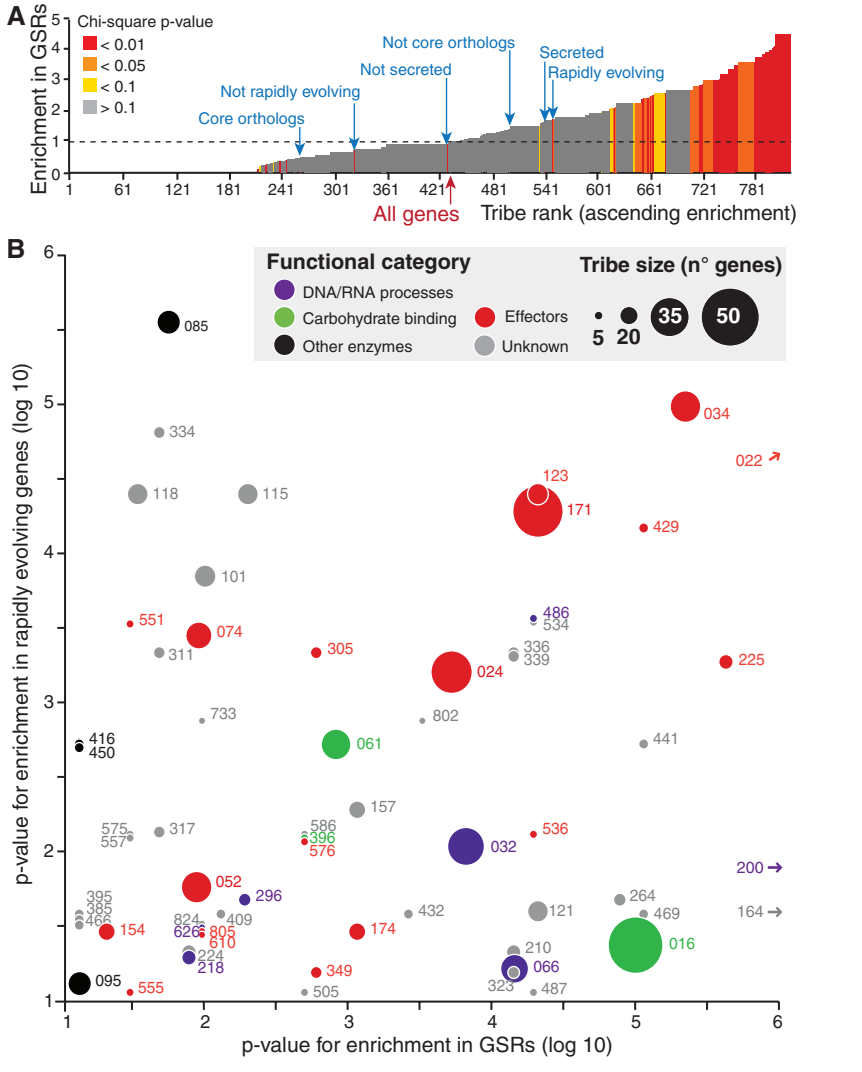


Fig. 3. Enrichment of *P. infestans* families (tribes) in genes residing in gene-sparse regions and rapidly evolving genes. (**A**) The 811 *P. infestans* tribes with five or more genes (*x* axis) ranked on the basis of ascending enrichment in GSR genes (*y* axis). *P* value of a χ^2 test for significance of enrichment is indicated. Additional gene categories (core/not core orthologs, secreted/not secreted, and rapidly/not rapidly evolving) are shown for reference. (**B**) *P* values of χ^2 tests for tribe enrichment in GSR genes (*x* axis) and rapidly evolving genes (*y* axis). Tribes with *P* values < 0.1 ($\log_{10}$) are shown. Bubble sizes are proportional to the number of genes in tribes. Bubble colors indicate functional categories. Numbers refer to tribe identifiers as listed in table S4.

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.

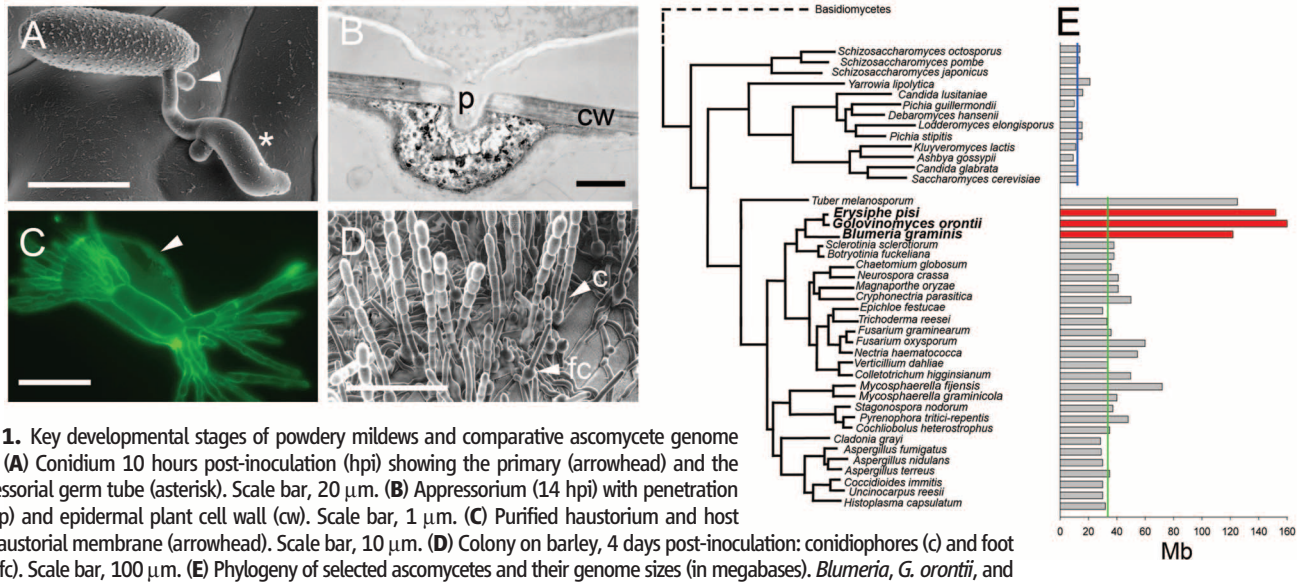


Fig. 1. Key developmental stages of powdery mildews and comparative ascomycete genome sizes. (**A**) Conidium 10 hours post-inoculation (hpi) showing the primary (arrowhead) and the appressorial germ tube (asterisk). Scale bar, 20 μm. (**B**) Appressorium (14 hpi) with penetration peg (p) and epidermal 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

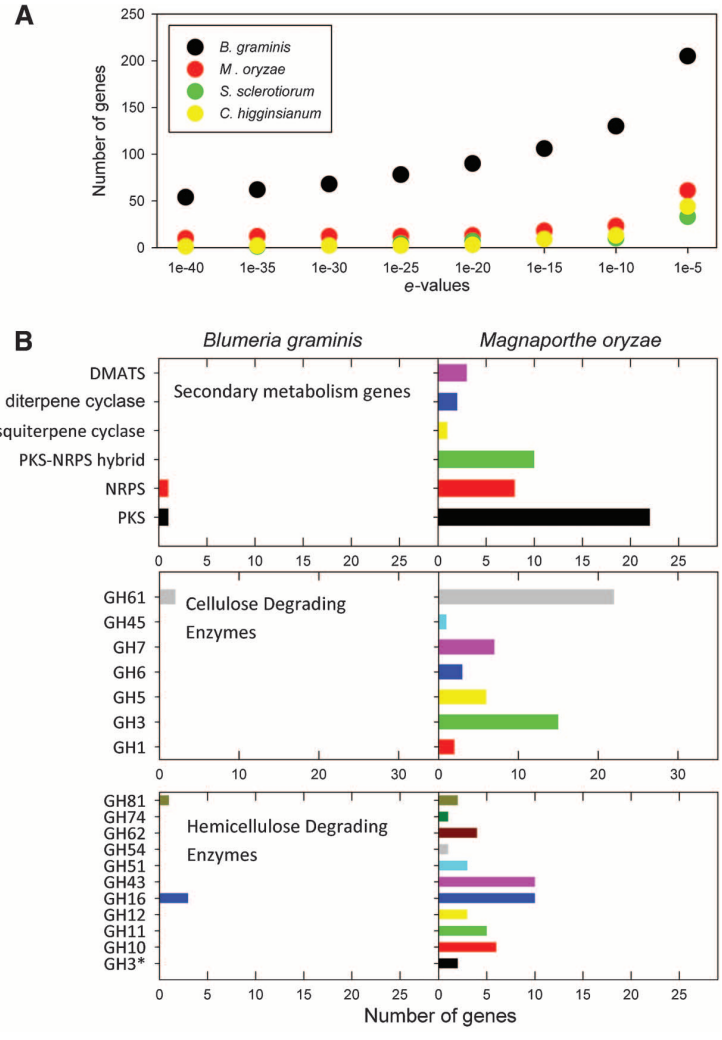


Fig. 2. Gene losses in powdery mildews. (**A**) Number of missing *S. cerevisiae* proteins in one fungus compared with the three others as a function of TBLASTN *e*-values. (**B**) Number of genes devoted to secondary metabolism and genes encoding cellulose- or hemicellulose-degrading enzymes in *Blumeria*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

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*).

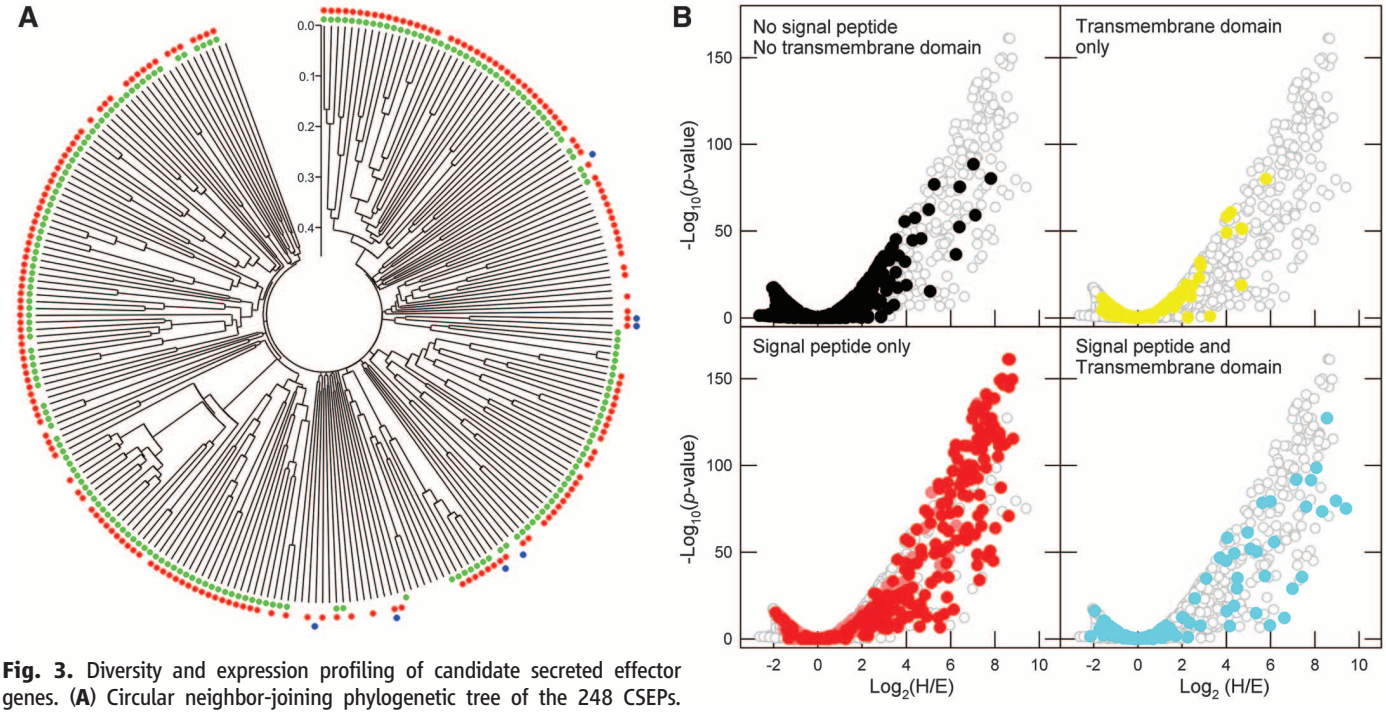


Fig. 3. Diversity and expression profiling of candidate secreted effector genes. (**A**) Circular neighbor-joining phylogenetic tree of the 248 CSEPs. Scale, amino acid substitutions per site; green, CSEPs harboring the YxC motif; blue, CSEPs conserved in *E. pisi* and/or *G. orontii*; red, CSEPs predominantly 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.

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

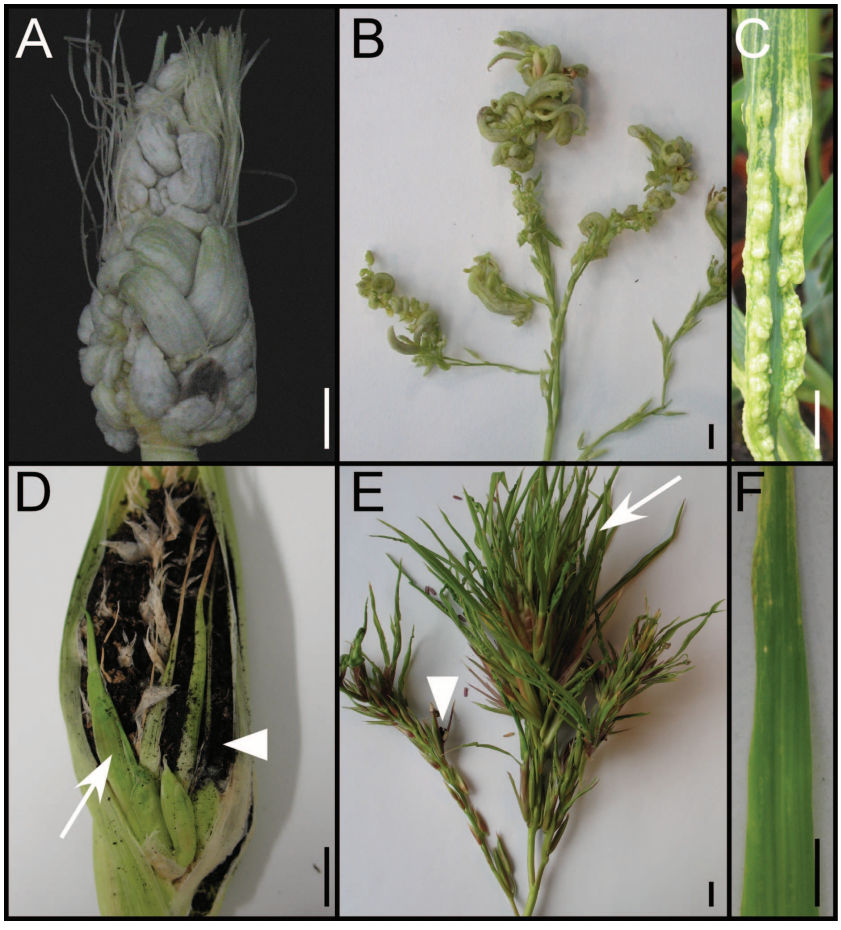


Fig. 1. Infection symptoms of *U. maydis* and *S. reilianum*. Tumor formation on ear (**A**), tassel (**B**), and leaf (**C**) after inflorescence infection [(A) and (B)] or seedling infection (C) by *U. maydis*. Maize seedling infection with *S. reilianum* leads to formation of spores (arrowheads) and leaflike structures (arrows) in ear (**D**) and tassel (**E**), whereas inoculated leaves show mild chlorosis but no tumors (**F**). Scale bars indicate 1 cm.

respective genes. Instead, we detected between one and four variants of a conserved 10–base pair (bp) sequence highly overrepresented in intergenic regions of *U. maydis* (*3*) that is largely absent from the genome of *S. reilianum* (table S4). We speculate that the genes encoding components of the RNAi machinery were lost in *U. maydis* by rare homologous recombination events involving the 10-bp motifs. To investigate whether the generation of small regulatory RNAs could explain the differences in symptoms of *U. maydis* and *S. reilianum*, we deleted the putative dicer gene *sr16838* in a solopathogenic strain of *S. reilianum* (*8*). Infection experiments revealed that *sr16838* deletion mutants were affected in neither virulence nor symptom development (fig. S2).

To detect regions of high sequence divergence, we compared the genomes of *U. maydis* and *S. reilianum* gene by gene (*8*, *10*) and identified regions encoding genes with low sequence conservation ("divergence clusters") in a conserved genomic context (*8*). This analysis revealed the presence of 43 divergence clusters (table S5) (*10*). Seventy-one percent of the genes in divergence clusters occurred in both organisms, whereas 19% were *S. reilianum*–specific and 10% were *U. maydis*–specific. Sixty-one percent of the genes in divergence clusters are predicted to encode secreted proteins. In contrast, of all *S. reilianum* and *U. maydis* genes less than 12% encode potentially secreted proteins. Ninety-four of the genes in divergence clusters code for proteins without functional annotation (table S5). It is notable that 442 of the 494 putative effectors detected in *U. maydis* are conserved in *S. reilianum*, with amino acid identities ranging from 20 to 98% (*10*). In addition, 445 genes are present only in *U. maydis*, whereas 372 exist only in *S. reilianum*, and of these about 15% encode secreted proteins.

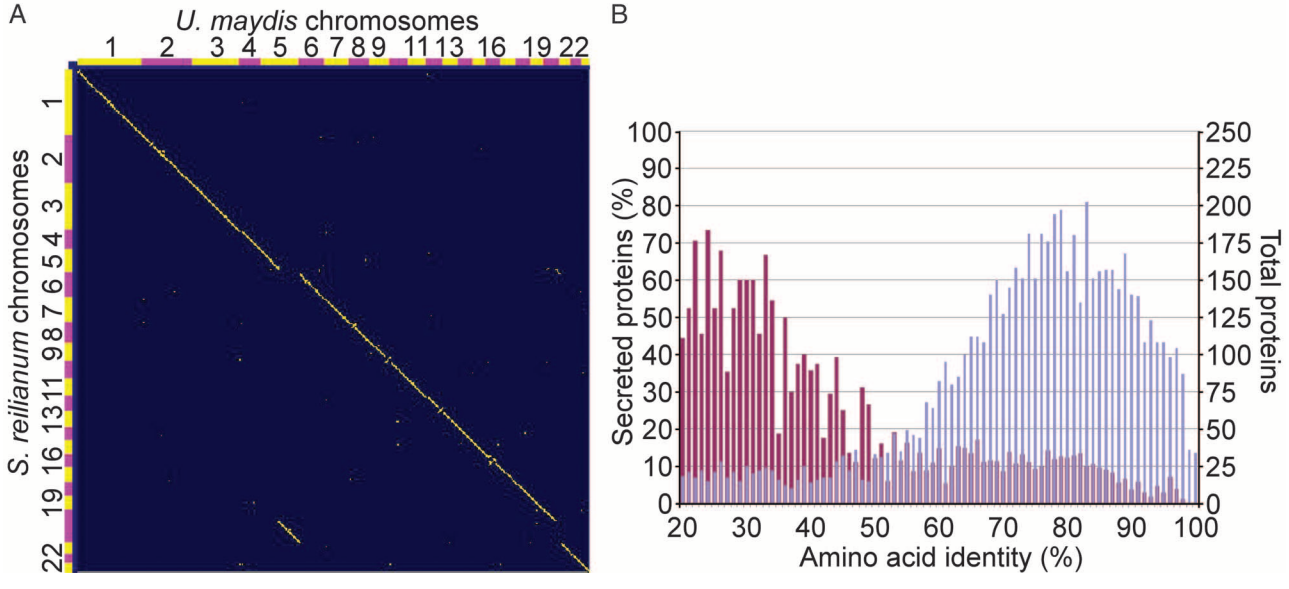


Fig. 2. Comparison of *U. maydis* and *S. reilianum* genomes. (**A**) Synteny (diagonal lines) of protein-encoding genes on the 23 chromosomes of *U. maydis* compared to those on the 23 chromosomes of *S. reilianum*. (**B**) Distribution of amino acid identities of all protein-encoding genes occurring in both genomes (right axis, blue bars). The percentage of proteins with a predicted secretion signal is given for each amino acid identity value (left axis, red bars).

Among the divergence gene clusters identified in this study were 7 of the 12 previously described *U. maydis* effector gene clusters (*3*), and these included all clusters whose deletion affected virulence in seedling infections (*3*) (fig. S3). In contrast, four of the gene clusters of secreted proteins without a deletion phenotype in *U. maydis* seedling infections (*3*) did not classify as divergence clusters (fig. S4). To test whether any of the newly identified divergence gene clusters also harbor virulence factors, we individually deleted six randomly chosen divergence clusters in *U. maydis* (Fig. 3). In seedling infection assays, loss of three divergence gene clusters (15-12, 5-21, and 20-15) attenuated virulence, whereas deletion of two clusters (5-18 and 11-16) did not affect virulence (Fig. 3 and fig. S5). The absence of a virulence phenotype likely reflects redundancy because in these cases potential paralogs exist elsewhere in the genome (table S6). In cluster 8-12, we detected the *mig1* gene that is highly induced during plant colonization and encodes an effector with similarity to apoplastic fungal avirulence proteins (*13*, *14*). Cluster 8-12 contains three additional genes, of which two encode proteins with similarity to Mig1. In *S. reilianum*, the *mig1* gene family is expanded to eight members residing in a single cluster (Fig. 3 and fig. S6). Whereas the deletion of only *mig1* did not affect virulence (*13*), cluster 8-12 deletion caused hypervirulence (fig. S5). Hypervirulence could result from an active attenuation of fungal proliferation by respective effectors (*3*). However, given the conserved features between avirulence proteins and Mig1 effectors, we now propose that genes whose deletion leads to hypervirulence encode weak avirulence proteins that trigger defense responses in plants expressing a cognate resistance protein, resulting in an attenuation of fungal growth.

Although most of the deleted genes in the four divergence clusters with an effect on virulence encode putatively secreted proteins, cluster 15-12 encodes only proteins without identifiable secretion signals, suggesting that additional mechanisms for virulence modulation exist in *U. maydis*. With respect to the origin of these clusters, we do not detect hallmarks for horizontal gene transfer like an altered GC content or an association with repetitive elements (fig. S7). Therefore, because many divergence regions contain members of small gene families (table S4) we propose that the majority of divergence clusters have been generated by local gene duplications followed by strong natural selection resulting from interaction with different host molecules.

Our studies demonstrate the power of comparative genomics of closely related species for the identification of new virulence genes. The *U. maydis* and *S. reilianum* pathosystems present a unique example of differentiation of two closely related pathogens parasitizing the same host. We have recognized that the *U. maydis* and *S. reilianum* genomes comprise conserved effector genes as expected for pathogens infecting the same host. However, although the two pathogens are both recognized and challenged by the maize immune system, they also possess strongly differentiated effectors, suggesting that they are targeting different host molecules. We speculate that their different infection strategies lead to the interaction between different host-pathogen molecules and thereby the evolution of differentiated sets of effector proteins, although we cannot exclude contributions of the species-specific genes. The assertion that closely related pathogens interact with and affect different host targets suggests a high variety in pathogen targets within the same host. It also suggests a temporal and spatial difference in the composition of different host proteins, which can drive the evolution of different sets of effectors in pathogens with different infection strategies.

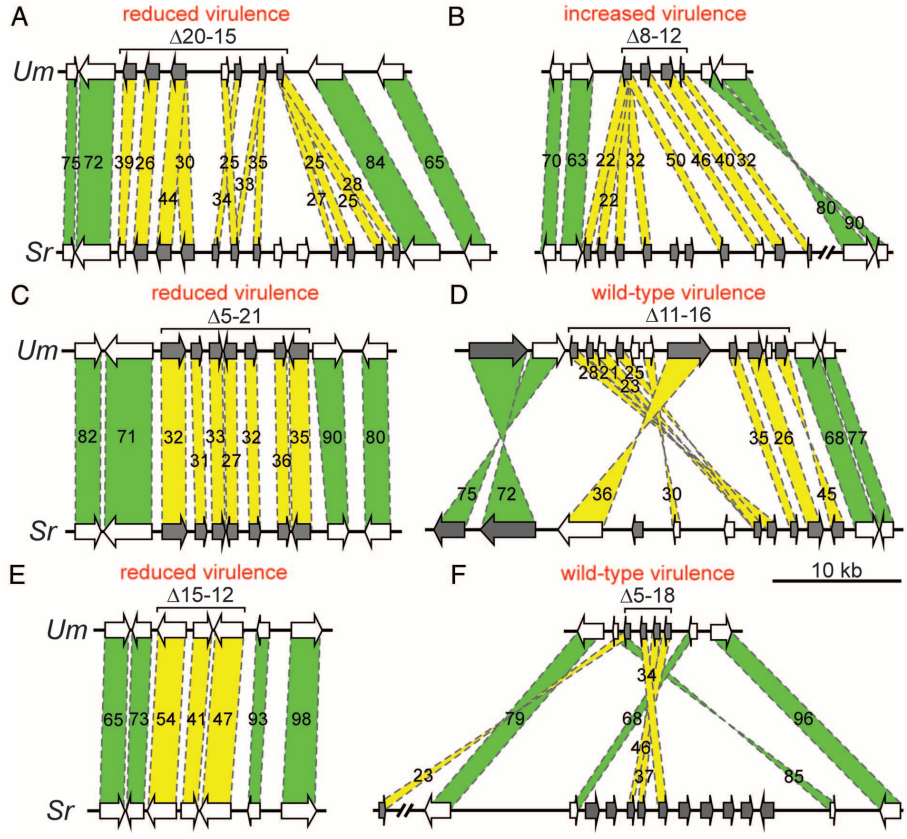


Fig. 3. Gene-by-gene comparisons of divergence clusters between *U. maydis* (*Um*) and *S. reilianum* (*Sr*) deleted in this study. (**A**) Cluster 15-12, (**B**) cluster 20-15, (**C**) cluster 11-16, (**D**) cluster 8-12, (**E**) cluster 5-21, and (**F**) cluster 5-18. Genes encoding putatively secreted proteins are shaded in gray. Bars connecting syntenic homologs are color-coded (green, high; yellow, weak conservation), and numbers give amino acid identities. Brackets denote regions deleted in *U. maydis* mutants. Virulence phenotypes of the respective mutants are indicated; the corresponding scores are found in fig. S5.

References and Notes

1. M. Stoll, D. Begerow, F. Oberwinkler, *Mycol. Res.* **109**, 342 (2005).
2. C. Martinez, C. Roux, R. Dargent, *Phytopathology* **89**, 247 (1999).
3. J. Kämper *et al.*, *Nature* **444**, 97 (2006).
4. T. Brefort *et al.*, *Annu. Rev. Phytopathol.* **47**, 423 (2009).
5. G. Doehlemann *et al.*, *PLoS Pathog.* **5**, e1000290 (2009).
6. D. S. Skibbe, G. Doehlemann, J. Fernandes, V. Walbot, *Science* **328**, 89 (2010).
7. B. J. Haas *et al.*, *Nature* **461**, 393 (2009).
8. Materials and methods are available as supporting material on *Science* Online.
9. MIPS, *Sporisorium reilianum* data base (MSRDB), http://mips.helmholtz-muenchen.de/genre/proj/sporisorium.
10. http://mips.helmholtz-muenchen.de/genre/export/sites/default/sporisorium/Download/Genome_comparison.xls
11. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
12. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Nature* **404**, 245 (2000).
13. C. W. Basse, S. Stumpferl, R. Kahmann, *Mol. Cell. Biol.* **20**, 329 (2000).
14. I. Stergiopoulos, P. J. de Wit, *Annu. Rev. Phytopathol.* **47**, 233 (2009).

15. We are grateful to the Max Planck Society for funding genome sequencing and annotation. T.B. was supported through the Deutsche Forschungsgemeinschaft via SFB593. A.M.M. received a Humboldt fellowship. H.G. was funded by the International Max Planck Research School for Environmental, Cellular, and Molecular Microbiology. R.K. acknowledges support through the LOEWE program of the State of Hesse. We thank U. Bonas and G. Gottschalk for critical comments on the manuscript. The sequence was submitted to European Molecular Biology Laboratory database under accession numbers FQ311430 to FQ311474.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6010/1546/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S6
References

19 July 2010; accepted 25 October 2010
10.1126/science.1195330

Signatures of Adaptation to Obligate Biotrophy in the *Hyaloperonospora arabidopsidis* Genome

Laura Baxter,[1]*† Sucheta Tripathy,[2]* Naveed Ishaque,[3]‡ Nico Boot,[4] Adriana Cabral,[4] Eric Kemen,[3] Marco Thines,[3,5,6] Audrey Ah-Fong,[7] Ryan Anderson,[8] Wole Badejoko,[1] Peter Bittner-Eddy,[1]† Jeffrey L. Boore,[9] Marcus C. Chibucos,[2]† Mary Coates,[1] Paramvir Dehal,[10] Kim Delehaunty,[11] Suomeng Dong,[12,13] Polly Downton,[1]† Bernard Dumas,[14,15] Georgina Fabro,[3] Catrina Fronick,[11] Susan I. Fuerstenberg,[9] Lucinda Fulton,[11] Elodie Gaulin,[14,15] Francine Govers,[16] Linda Hughes,[1] Sean Humphray,[17] Rays H. Y. Jiang,[16,18] Howard Judelson,[7] Sophien Kamoun,[3] Kim Kyung,[11] Harold Meijer,[16] Patrick Minx,[11] Paul Morris,[19] Joanne Nelson,[11] Vipa Phuntumart,[19] Dinah Qutob,[12] Anne Rehmany,[1] Alejandra Rougon-Cardoso,[3]† Peter Ryden,[1]† Trudy Torto-Alalibo,[2] David Studholme,[3]† Yuanchao Wang,[13] Joe Win,[3] Jo Wood,[17] Sandra W. Clifton,[11] Jane Rogers,[17]† Guido Van den Ackerveken,[4] Jonathan D. G. Jones,[3]†‡ John M. McDowell,[8]‡§ Jim Beynon,[1]‡ Brett M. Tyler[2,8]‡

Many oomycete and fungal plant pathogens are obligate biotrophs, which extract nutrients only from living plant tissue and cannot grow apart from their hosts. Although these pathogens cause substantial crop losses, little is known about the molecular basis or evolution of obligate biotrophy. Here, we report the genome sequence of the oomycete *Hyaloperonospora arabidopsidis* (*Hpa*), an obligate biotroph and natural pathogen of *Arabidopsis thaliana*. In comparison with genomes of related, hemibiotrophic *Phytophthora* species, the *Hpa* genome exhibits dramatic reductions in genes encoding (i) RXLR effectors and other secreted pathogenicity proteins, (ii) enzymes for assimilation of inorganic nitrogen and sulfur, and (iii) proteins associated with zoospore formation and motility. These attributes comprise a genomic signature of evolution toward obligate biotrophy.

The oomycete *Hyaloperonospora arabidopsidis* (*Hpa*, formerly *Peronospora parasitica* or *Hyaloperonospora parasitica*) is a natural pathogen of *Arabidopsis thaliana* and a model for dissection of *A. thaliana* pathogen response networks (*1*, *2*). *Hpa* belongs to a group of "downy mildew" pathogens, comprising more than 800 species that cause disease on hundreds of plant species (*3*). Downy mildew pathogens are related to other destructive oomycete plant pathogens (e.g., *Phytophthora* species) (*4*, *5*). Oomycetes belong to the kingdom Stramenopila, which includes brown algae and diatoms. Although oomycetes and fungi share morphological and ecological similarities, they evolved independently to colonize plants.

Hpa hyphae grow between plant cells and establish feeding structures called haustoria, which have also evolved in fungal pathogens (*2*, *6*). Downy mildews are obligately biotrophic and cannot be cultured apart from their hosts. In contrast, *Phytophthora* species are hemibiotrophic; an initial phase of biotrophic growth is followed by a necrotrophic phase. Molecular phylogenies show that downy mildews arose from a paraphyletic, *Phytophthora*-like, hemibiotrophic ancestor (*4*, *5*, *7*). Thus, insight into the genomic basis and evolution of obligate biotrophy can be obtained through comparison of the *Hpa* genome to the recently sequenced genomes of *Phytophthora* species (*8*, *9*).

Genome analysis was performed on the *Hpa* Emoy2 isolate (*1*) using DNA from asexual spores (*10*). Sanger shotgun sequencing at 9.5-fold coverage, combined with 97 sequenced bacterial artificial chromosome inserts, yielded an assembly of 77.8 Mb. Illumina sequencing at 46-fold coverage yielded 3.8 Mb of additional sequence, which was integrated into the Sanger assembly to form the 81.6-Mb final version (v8.3.2). Forty-two percent of the *Hpa* genome is composed of repetitive elements (table S1). Analysis of Sanger and Illumina read depth suggests that v8.3 contains ~12.7 Mb composed of tandem repeats compressed into reduced copies in the assembly, indicative of a genome size of around 100 Mb (fig. S1). The CEGMA (Core Eukaryotic Genes Mapping Approach) pipeline, combined with manual examination, identified *Hpa* orthologs of 95% of the 248 conserved single-copy eukaryotic genes (fig. S2). Moreover, 94% of 31,759 expressed sequence tag (EST) reads aligned to the assembly, indicating that the draft genome assembly encompasses a very high percentage of the *Hpa* gene space.

A total of 14,543 genes were computationally predicted in v8.3, of which 80% are supported by ESTs and/or Illumina cDNA tags. This predicted gene content is similar to *P. ramorum* (65 Mb, 15,743 genes) and lower than *P. sojae* (95 Mb, 19,027 genes) (*9*) or *P. infestans* (240 MB,

[1]School of Life Sciences, Warwick University, Wellesbourne, CV35 9EF, UK. [2]Virginia Bioinformatics Institute, Virginia Polytechnic Institute, Blacksburg, VA 24061, USA. [3]Sainsbury Laboratory, John Innes Centre, Norwich NR4 7UH, UK. [4]Plant-Microbe Interactions, Department of Biology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, Netherlands. [5]Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt (Main), Germany. [6]Johann Wolfgang Goethe University, Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, Siesmayerstr. 70, D-60323 Frankfurt (Main), Germany. [7]Department of Plant Pathology and Microbiology, University of California, Riverside, CA 92521, USA. [8]Department of Plant Pathology, Physiology and Weed Science, Virginia Polytechnic Institute, Blacksburg, VA 24061, USA. [9]Department of Integrative Biology, University of California, Berkeley, USA. [10]Lawrence Berkeley National Laboratories, Berkeley, CA 94720, USA. [11]Genome Sequencing Centre, Washington University School of Medicine, St. Louis, MO 63110, USA. [12]Agriculture and Agri-Food Canada, London, Ontario N5V 4T3, Canada. [13]Department of Plant Pathology, Nanjing Agricultural University, Nanjing 210095, China. [14]Université de Toulouse, UPS, Surfaces Cellulaires et Signalisation chez les Végétaux, 24 chemin de Borde Rouge, BP42617, Auzeville, F-31326 Castanet-Tolosan, France. [15]CNRS, Surfaces Cellulaires et Signalisation chez les Végétaux, 24 chemin de Borde Rouge, BP42617, Auzeville, F-31326 Castanet-Tolosan, France. [16]Laboratory of Phytopathology, Wageningen University, and Centre for BioSystems Genomics, NL-1-6708 PB Wageningen, Netherlands. [17]Sanger, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [18]The Broad Institute of MIT and Harvard, Cambridge, MA 02141–2023, USA. [19]Department of Biological Sciences, Bowling Green State University, Bowling Green, OH 43403–0212, USA.

*These authors contributed equally to this work.
†Present addresses and other affiliations are listed in the supporting online material.
‡These authors contributed equally to this work.
§To whom correspondence should be addressed. E-mail: johnmcd@vt.edu

Table 1. Copy numbers of annotated *Hpa* genes for hydrolases, PAMPS, and effectors, compared with *Phytophthora* genomes.

Gene product	*H. arabidopsidis*	*P. sojae*	*P. ramorum*
Extracellular proteases	18	47	48
Glycosyl hydrolases	>60	125	114
Endoglucanases (EGL12)	3	10	8
Polygalacturonases	3	25	16
Pectin methyl esterases	3	19	15
Cutinases	2	16	4
Chitinases	1	5	2
Elicitins	1	18	17
Elicitin-like	14	39	31
CBEL and CBEL-like	2	13	15
RXLR	134	396	374
NLP	10	29	40
Crinklers	20	40	8
PPAT12/24-like	8	0	0

17,887 genes) (*8*). A total of 6882 predicted genes in *Hpa* had no identifiable ortholog in sequenced *Phytophthora* species or similarity to known proteins, and as such represent potentially lineage-specific genes. Some of these genes may play roles that are specific to biotrophy. For example, a novel family of secreted small, cysteine-rich proteins exists in *Hpa* (PPAT12/24-like) (Table 1) (*11*).

Pathogenicity genes were compared among *Hpa* and *Phytophthora* species, revealing that families encoding host-targeted, degradative enzymes (secreted proteinases, cell wall–degrading enzymes) are reduced in *Hpa* (Table 1). Two notable examples are the family 12 endoglucanases (*EGL12*) and pectin methyl esterases (*Pect*). Phylogenetic analyses delineated several EGL12 and Pect gene clades containing genes from *P. sojae* and *P. ramorum* but not *Hpa* (figs. S6 and S7). Because *Hpa* and *P. sojae* likely share a sister group relationship relative to *P. ramorum* (*4*, *5*), it is probable that a number of *EGL12* and *Pect* genes were lost from *Hpa* after divergence of the lineage leading to *Hpa* and *P. sojae*. Hydrolytic enzymes that target the host cell wall can release cell wall fragments that elicit host defenses. It is conceivable that in evolving a biotrophic lifestyle, *Hpa* has lost most of the secreted hydrolytic enzymes that were present in a hemibiotrophic ancestor.

Similarly, gene families encoding necrosis and ethylene-inducing (Nep1)–like proteins (NLPs) are significantly reduced in *Hpa*, compared with *P. sojae* and *P. ramorum*. NLPs in *Phytophthora* and *Pythium* can trigger plant cell death and defenses (*12*), which have been implicated in the transition from biotrophy to necrotrophy. Only three of the 13 oomycete NLP clades contain genes from *Hpa*. However, one clade contains an expanded family that is specific to *Hpa* (Fig. 1A). All 10 *HaNLP* genes are supported by transcriptional data. Of these, HaNLP3 is most closely related to the PsojNIP and PiNPP1.1 proteins, but it did not induce necrosis in *Nicotiana tabacum* (Fig. 1B). These results suggest that downy mildew NLP genes may have evolved a different function than in *Phytophthora*. Copy number reduction was also evident for genes encoding known pathogen-associated molecular patterns (PAMPs) such as sterol-binding elicitins (*13*) and carbohydrate-binding CBEL (cellulose-binding, elicitor, lectin-like) genes (*14*) (Table 1). These examples further suggest that selection for "stealth" (avoidance of host defenses) was a major force during downy mildew evolution.

Phytophthora genomes encode hundreds of potential effector proteins (*9*, *15*, *16*) with RXLR cell entry motifs (*16*–*18*) that likely function to suppress host defenses (*19*, *20*). The Hpa genome contained 134 high-confidence effector gene candidates (HaRxL genes), including the known effector genes *Atr1* and *Atr13* (*21*, *22*), significantly fewer than in the *Phytophthora* genomes (*9*, *15*). Single-nucleotide polymorphisms arising from heterozygosity in v8.3 occurred at a rate 5 times as high in RXLR effector candidates [1 per ~500 base pairs (bp)] than in other genes (1 per ~2500 bp). Only 36% of the high-confidence *Hpa* effectors had significant matches in any *Phytophthora* genome (sequence similarity >30%), consistent with strong divergent selection on

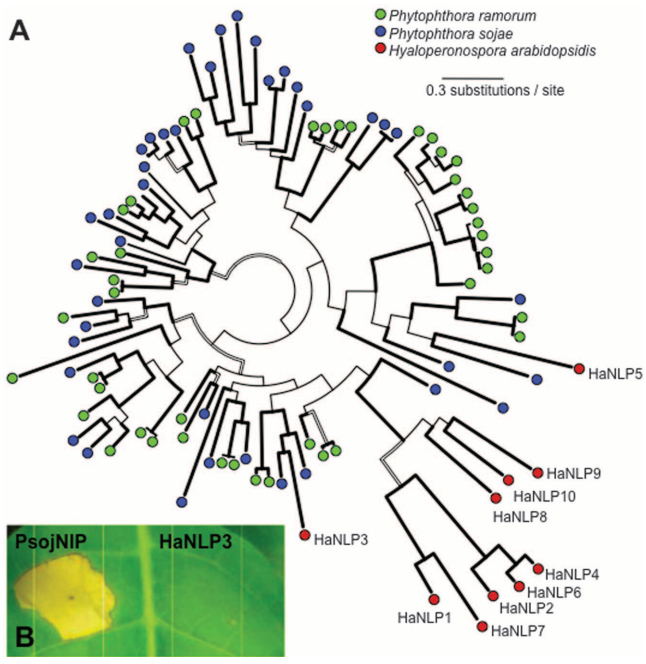


Fig. 1. Diversity, evolutionary history, and functional analysis of oomycete necrosis and NLPs. (**A**) Phylogeny of oomycete NLPs. A consensus tree from the Bayesian inference is shown. Thick lines indicate high support in minimum evolution (>90), maximum likelihood (>90), and Bayesian inference (>0.95). Hollow lines indicate branches highly supported in at least two analyses. Branches with high support in less than two analyses are represented by thin lines. (**B**) An *Hpa* NLP ortholog does not induce necrosis in plant leaves. NLP genes were transiently expressed in *Nicotiana tabacum* by agroinfiltration.

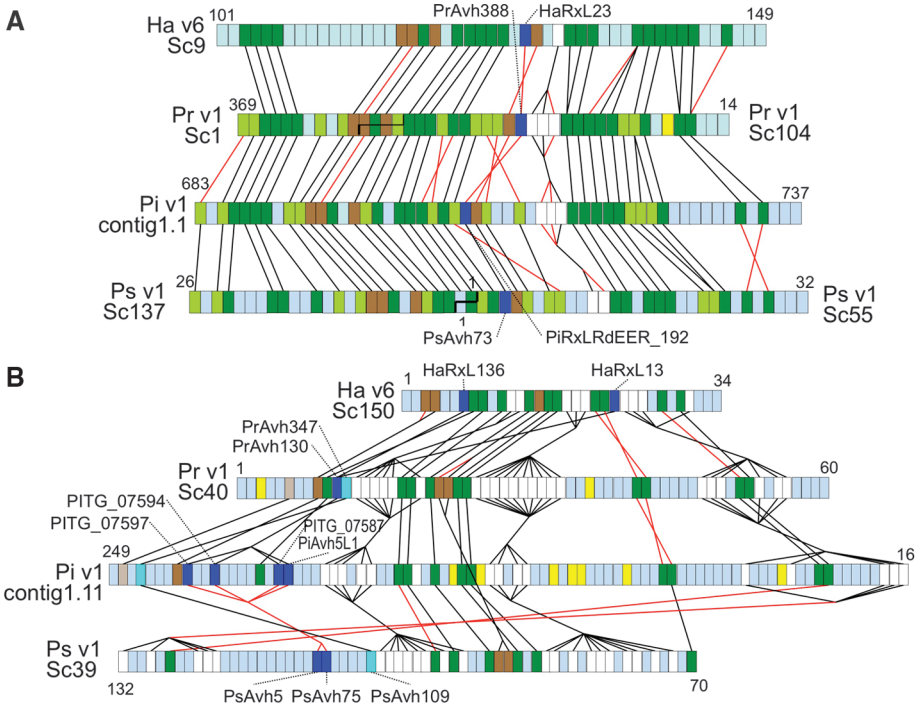


Fig. 2. Synteny of conserved RXLR effectors. (**A**) Region around HaRxL23, spanning scaffold_9:467737-739923 (v6) and supercontig16:325445-40488 (v8.3.2) (**B**) Region around HaRxL13 and HaRxL136, spanning scaffold_150:3503-183330 (v6) and supercontig35:456072-268293 (v8.3.2). Colored boxes show order of gene models. Noncoding DNA is not represented. Dark green, orthologs; light green, orthologs found only in *Phytophthora*; dark brown, syntenic paralogs; light brown, syntenic orthologs found only in *Phytophthora*; white, syntenic gene families; dark blue, syntenic conserved RXLR effectors; cyan, syntenic RXLR effectors conserved only in *Phytophthora*; yellow, RXLR effectors not syntenic or conserved; blue-gray, other genes not conserved or syntenic. Black lines join syntenic genes with the same orientation; red lines join genes with reversed orientations. Staggered black lines in (A) show scaffold joins predicted from the synteny analysis. HaRxL23, HaRxL13, and HaRxL136 have 47, 38, and 40% amino acid identity, respectively, with their most similar *Phytophthora* ortholog within the normally hypervariable C terminus.

RXLR effector genes (*15*, *23*). Moreover, *Hpa* effectors generally were not located in syntenic locations relative to *Phytophthora* genomes, except for three families of effectors, which have unusually high levels of sequence conservation (Fig. 2).

As obligate biotrophs, downy mildews may have lost some metabolic pathways. We identified several potential metabolic defects in *Hpa* compared with *P. sojae* and *P. ramorum* (fig. S9). For example, genes for nitrate and nitrite reductases, a nitrate transporter, and sulfite reductase were missing (fig. S10 and table S3), which is also a feature of the genomes of obligately parasitic powdery mildew fungi (*24*). *Hpa* also lacks genes required for synthesis of arachidonic acid and polyamine oxidases.

Flagellated zoospores are produced by many oomycetes (*25*). Contrastingly, several downy mildew lineages germinate by extending infective germ tubes from nonmotile conidiospores, although evidence exists for a rare zoosporic stage in some otherwise conidial downy mildews (*26*, *27*). To conclusively determine whether spore motility has been lost from the *Hpa* lineage, we searched the *Hpa* genome for 90 flagella-associated genes using *Chlamydomonas* sequences and their *Phytophthora* orthologs (*28*). No matches were detected in *Hpa* for any of these. Similarly, many *Phytophthora* adhesion-related genes are reduced in number or absent from *Hpa*, consistent with the lack of adherent cysts that normally develop from zoospores during infection.

Analysis of *Hpa* gene space revealed genomic signatures of major alterations in pathogenic strategy, metabolism, and development that occurred during the evolution of obligate biotrophy from a facultative, hemibiotrophic ancestor. Interestingly, some features of *Hpa* gene space (large numbers of secreted effectors, reduction in degradative enzymes, and loss of N and S assimilation) are mirrored in genomes of biotrophic fungi (*24*, *29*, *30*). These similarities indicate that convergent adaptations occurred during the independent evolution of biotrophy in fungal and oomycete lineages.

References and Notes

1. E. B. Holub, *Eur. J. Plant Pathol.* **122**, 91 (2008).
2. M. E. Coates, J. L. Beynon, *Annu. Rev. Phytopathol.* **48**, 329 (2010).
3. J. Clark, P. Spencer-Phillips, in *Encyclopedia of Microbiology* (Academic Press, 2000), vol. 2, pp. 117–129.
4. X. Giresse, S. Ahmed, S. Richard-Cervera, F. Delmotte, *J. Phytopathol.* **158**, 321 (2010).
5. M. Göker, H. Voglmayr, A. Riethmüller, F. Oberwinkler, *Fungal Genet. Biol.* **44**, 105 (2007).
6. R. Panstruga, P. N. Dodds, *Science* **324**, 748 (2009).
7. M. Thines, *PLoS ONE* **4**, e4790 (2009).
8. B. J. Haas *et al.*, *Nature* **461**, 393 (2009).
9. B. M. Tyler *et al.*, *Science* **313**, 1261 (2006).
10. Materials and methods are available as supporting material on *Science* Online.
11. P. D. Bittner-Eddy, R. L. Allen, A. P. Rehmany, P. Birch, J. L. Beynon, *Mol. Plant Pathol.* **4**, 501 (2003).
12. M. Gijzen, T. Nürnberger, *Phytochemistry* **67**, 1800 (2006).
13. S. Kamoun, *Annu. Rev. Phytopathol.* **44**, 41 (2006).
14. E. Gaulin *et al.*, *Plant Cell* **18**, 1766 (2006).
15. R. H. Jiang, S. Tripathy, F. Govers, B. M. Tyler, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4874 (2008).
16. S. C. Whisson *et al.*, *Nature* **450**, 115 (2007).
17. D. Dou *et al.*, *Plant Cell* **20**, 1930 (2008).
18. S. Kale *et al.*, *Cell* **142**, 284 (2010).
19. D. Dou *et al.*, *Plant Cell* **20**, 1118 (2008).
20. K. H. Sohn, R. Lei, A. Nemri, J. D. Jones, *Plant Cell* **19**, 4077 (2007).
21. R. L. Allen *et al.*, *Science* **306**, 1957 (2004).
22. A. P. Rehmany *et al.*, *Plant Cell* **17**, 1839 (2005).
23. J. Win *et al.*, *Plant Cell* **19**, 2349 (2007).
24. P. Spanu, *Science* **330**, 1543 (2010).
25. A. R. Hardham, G. J. Hyde, *Adv. Bot. Res.* **24**, 353 (1997).
26. D. G. Milbrath, *J. Agric. Sci.* **23**, 989 (1923).
27. V. Skalicky, *Preslia* **38**, 117 (1966).
28. G. J. Pazour, N. Agrin, J. Leszyk, G. B. Witman, *J. Cell Biol.* **170**, 103 (2005).
29. J. Kämper *et al.*, *Nature* **444**, 97 (2006).
30. F. Martin *et al.*, *Nature* **452**, 88 (2008).

31. We thank E. Holub for providing the Emoy2 isolate, D. Greenshields and N. Bruce for technical assistance, A. Heck and M. Slijper for analysis of secreted Hpa proteins, R. Hubley for creating repeat modeller libraries, and participants in the 2007 Annotation Jamboree and in the 2008 and 2009 Oomycete Bioinformatics Training Workshops for sequence annotations. This research was supported by grants EF-0412213, IOS-0744875, IOS-0924861, and MCB-0639226 from the U.S. NSF and 2004-35600-15055 and 2007-35319-18100 from the U.S. Department of Agriculture National Institute of Food and Agriculture to B.M.T. and J.M.M.; Biotechnology and Biological Sciences Research Council (BBSRC) BB/C509123/1, BB/E024815/1, and Engineering and Physical Sciences Research Council/BBSRC Systems Biology DTC student EP/F500025/1 to J.B.; Gatsby GAT2545 and BBSRC BB/F0161901, BB/E024882/1, and BBSRC CASE studentship T12144 to J.D.G.J. Other support is detailed in the supporting online material. Genome browsers are maintained at the Virginia Bioinformatics Institute (vmd.vbi.vt.edu) and the Sainsbury Laboratories (gbrowse2.tsl.ac.uk/cgi-bin/gb2/gbrowse/hpa_emoy2_publication).

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6010/1549/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S3
References

16 July 2010; accepted 25 October 2010
10.1126/science.1195203

The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation

The International HIV Controllers Study*†

Infectious and inflammatory diseases have repeatedly shown strong genetic associations within the major histocompatibility complex (MHC); however, the basis for these associations remains elusive. To define host genetic effects on the outcome of a chronic viral infection, we performed genome-wide association analysis in a multiethnic cohort of HIV-1 controllers and progressors, and we analyzed the effects of individual amino acids within the classical human leukocyte antigen (HLA) proteins. We identified >300 genome-wide significant single-nucleotide polymorphisms (SNPs) within the MHC and none elsewhere. Specific amino acids in the *HLA-B* peptide binding groove, as well as an independent *HLA-C* effect, explain the SNP associations and reconcile both protective and risk *HLA* alleles. These results implicate the nature of the HLA–viral peptide interaction as the major factor modulating durable control of HIV infection.

HIV infection is characterized by acute viremia, often in excess of 5 million viral particles per milliliter of plasma, followed by an average 100-fold or greater decline to a relatively stable plasma virus load set point (*1*). In the absence of antiretroviral therapy, the level of viremia is associated with the rate of $CD4^+$ T cell decline and progression to AIDS. There is substantial interperson variability in the virus load set point, with most individuals having stable levels exceeding 10,000 RNA copies/ml. Yet a small number of people demonstrate sustained ability to control HIV replication without therapy. Such individuals, referred to as HIV controllers, typically maintain stable $CD4^+$ cell counts, do not develop clinical disease, and are less likely to transmit HIV to others (*2*).

To determine the genetic basis for this rare phenomenon, we established a multinational consortium (www.hivcontrollers.org) to recruit HIV-1 controllers, who are defined by at least three measurements of plasma virus load (VL) < 2000 RNA copies/ml over at least a 12-month period in the absence of antiviral therapy. We performed a genome-wide association study (GWAS) in the HIV controllers (median VL, CD4 count, and disease duration of 241 copies/ml, 699 cells/mm^3, and 10 years, respectively) and treatment-naïve chronically infected individuals with advanced disease (median VL and CD4 count of 61,698 copies/ml and 224 cells/mm^3, respectively) enrolled in antiviral treatment studies led by the AIDS Clinical Trials Group. After quality control and imputation on the basis of HapMap

*All authors with their contributions and affiliations appear at the end of this paper.
†To whom correspondence should be addressed. E-mail: bwalker@partners.org (B.D.W.); pdebakker@rics.bwh.harvard.edu (P.I.W.d.B.)