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

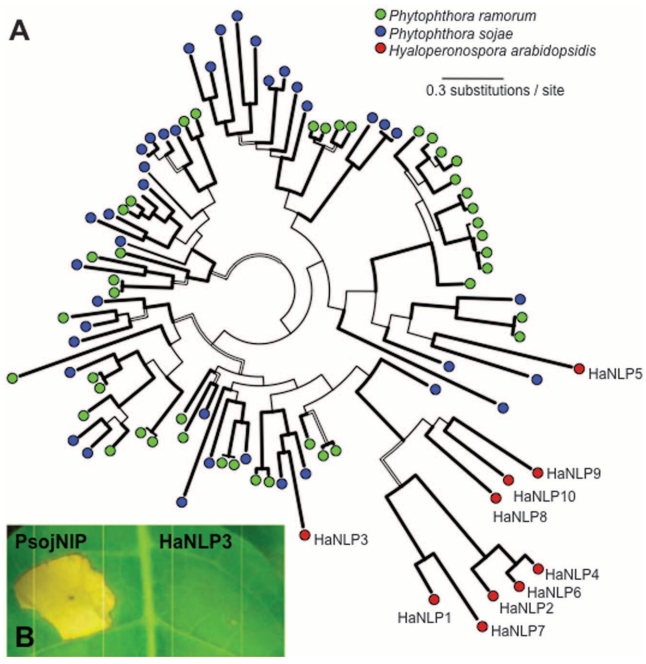

Fig. 1. Diversity, evolutionary history, and functional analysis of oomycete necrosis and NLPs. (**A**) Phylogeny of oomycete NLPs. A consensus tree from the Bayesian inference is shown. Thick lines indicate high support in minimum evolution (>90), maximum likelihood (>90), and Bayesian inference (>0.95). 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.

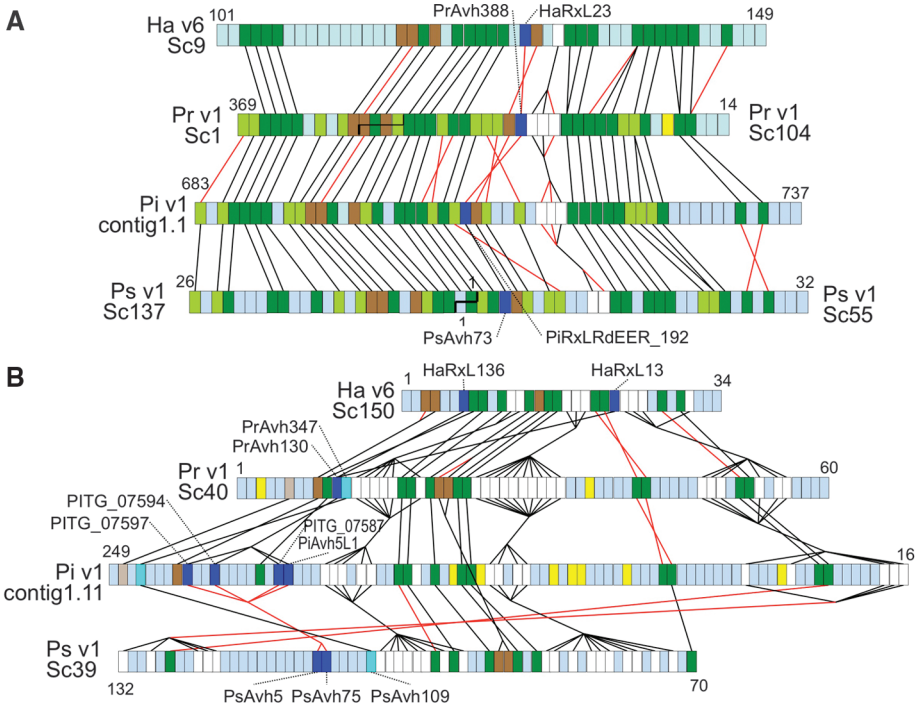

Fig. 2. Synteny of conserved RXLR effectors. (**A**) Region around HaRxL23, spanning scaffold_9:467737-739923 (v6) and supercontig16:325445-40488 (v8.3.2) (**B**) Region around HaRxL13 and HaRxL136, spanning scaffold_150:3503-183330 (v6) and supercontig35:456072-268293 (v8.3.2). Colored boxes show order 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.)