



Universiteit  
Leiden  
The Netherlands

## A highly virulent variant of HIV-1 circulating in the Netherlands

Wymant, C.; Bezemer, D.; Blanquart, F.; Ferretti, L.; Gall, A.; Hall, M.; ... ; BEEHIVE Collaboration

### Citation

Wymant, C., Bezemer, D., Blanquart, F., Ferretti, L., Gall, A., Hall, M., ... Fraser, C. (2022). A highly virulent variant of HIV-1 circulating in the Netherlands. *Science*, 375(6580), 540-545.  
doi:10.1126/science.abk1688

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3502598>

**Note:** To cite this publication please use the final published version (if applicable).

## HIV

# A highly virulent variant of HIV-1 circulating in the Netherlands

Chris Wymant<sup>1\*</sup>, Daniela Bezemer<sup>2</sup>, François Blanquart<sup>3,4</sup>, Luca Ferretti<sup>1</sup>, Astrid Gall<sup>5</sup>, Matthew Hall<sup>1</sup>, Tanya Golubchik<sup>1</sup>, Margreet Bakker<sup>6</sup>, Swee Hoe Ong<sup>7</sup>, Lele Zhao<sup>1</sup>, David Bonsall<sup>1,8</sup>, Mariateresa de Cesare<sup>8</sup>, George MacIntyre-Cockett<sup>1,8</sup>, Lucie Abeler-Dörner<sup>1</sup>, Jan Albert<sup>9,10</sup>, Norbert Bannert<sup>11</sup>, Jacques Fellay<sup>12,13,14</sup>, M. Kate Grabowski<sup>15</sup>, Barbara Günsenheimer-Bartmeyer<sup>16</sup>, Huldrych F. Günthard<sup>17,18</sup>, Pia Kivelä<sup>19</sup>, Roger D. Kouyos<sup>17,18</sup>, Oliver Laeyendecker<sup>20</sup>, Laurence Meyer<sup>21</sup>, Kholoud Porter<sup>22</sup>, Matti Ristola<sup>19</sup>, Ard van Sighem<sup>2</sup>, Ben Berkhout<sup>6</sup>, Paul Kellam<sup>23,24</sup>, Marion Cornelissen<sup>6,25</sup>, Peter Reiss<sup>2,26</sup>, Christophe Fraser<sup>1,8\*</sup>, the Netherlands ATHENA HIV Observational Cohort<sup>†</sup>, the BEEHIVE Collaboration<sup>†</sup>

We discovered a highly virulent variant of subtype-B HIV-1 in the Netherlands. One hundred nine individuals with this variant had a 0.54 to 0.74 log<sub>10</sub> increase (i.e., a ~3.5-fold to 5.5-fold increase) in viral load compared with, and exhibited CD4 cell decline twice as fast as, 6604 individuals with other subtype-B strains. Without treatment, advanced HIV–CD4 cell counts below 350 cells per cubic millimeter, with long-term clinical consequences—is expected to be reached, on average, 9 months after diagnosis for individuals in their thirties with this variant. Age, sex, suspected mode of transmission, and place of birth for the aforementioned 109 individuals were typical for HIV-positive people in the Netherlands, which suggests that the increased virulence is attributable to the viral strain. Genetic sequence analysis suggests that this variant arose in the 1990s from de novo mutation, not recombination, with increased transmissibility and an unfamiliar molecular mechanism of virulence.

The risk posed by viruses evolving to greater virulence—i.e., causing greater damage to their hosts—has been extensively studied in theoretical work despite few population-level examples (1–3). The most notable recent example is the B.1.1.7 lineage (Alpha variant) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), for which an increased probability of death has been reported (4–6), as well as increased transmissibility (7, 8). RNA viruses have long been a particular concern, as their error-prone replication results in the greatest known rate of mutation—and thus high potential for adaptation. Greater virulence could benefit a virus if it is not outweighed by reduced opportunity for transmission. These antagonistic selection pressures may result in an intermediate level of virulence being optimal for viral fitness, as observed for HIV (9). Concrete examples of such evolution in action, however, have been elusive. Continued monitoring of HIV virulence is important for global health: 38 million people currently live

with the virus, and it has caused an estimated 33 million deaths (www.unaids.org).

The main (M) group of HIV-1, responsible for the global pandemic, first emerged around 1920 in the area of what is now Kinshasa, Democratic Republic of the Congo (10), and had diversified into subtypes by 1960 (11). The subtypes, and the most common circulating recombinant forms (CRFs) between the subtypes, took different routes for global spread, establishing strong associations with geography (12), ethnicity, and mode of transmission. Differences in virulence between subtypes and CRFs have been reported, though it is challenging to disentangle genotypic effects on virulence from confounding effects while retaining large sample sizes, given the strong associations between viral, host, and epidemiological factors (13). The co-receptor used for cell entry has long been understood to affect virulence (14, 15), and this has been proposed as a mechanism that underlies differences in virulence between subtypes and CRFs (13),

as well as one reported difference within a CRF (16).

HIV-1 virulence is most commonly measured by viral loads (the concentration of viral particles in blood plasma) and CD4 counts (the concentration of CD4<sup>+</sup> T cells in peripheral blood, which tracks immune system damage by the virus). Successful treatment with anti-retroviral drugs suppresses viral load and interrupts the decline in CD4 counts that would otherwise lead to AIDS. Both viral load and rate of CD4 cell decline are heritable properties—that is, these properties are causally affected by viral genetics, leading to correlation between an individual and whomever they infect (17–21). It has therefore been expected that viral load and CD4 cell decline could change with the emergence of a new viral variant. We substantiate that expectation with empirical evidence by reporting a subtype-B variant of HIV-1 with exceptionally high virulence that has been circulating within the Netherlands during the past two decades.

## Discovery of the highly virulent variant

Within an ongoing study (the BEEHIVE project; www.beehive.ox.ac.uk), we identified a group of 17 individuals with a distinct subtype-B viral variant, whose viral loads in the set-point window of infection (6 to 24 months after a positive test obtained early in the course of infection) were highly elevated (Table 1, middle column). BEEHIVE is a study of individuals enrolled in eight cohorts across Europe and Uganda, who were selected because they have well-characterized dates of infection and samples available from early infection, for whom whole viral genomes were sequenced. The 17 individuals with the distinct viral variant comprised 15 participants in the ATHENA study in the Netherlands, 1 from Switzerland, and 1 from Belgium. See materials and methods for details on the initial discovery.

## Replication of the discovery in Dutch ATHENA data

To replicate the finding and to investigate this viral variant in more detail, we then analyzed data from 6706 participants in ATHENA with

<sup>1</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>2</sup>Stichting HIV Monitoring, Amsterdam, Netherlands. <sup>3</sup>Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France. <sup>4</sup>IAME, UMR 1137, INSERM, Université de Paris, Paris, France. <sup>5</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. <sup>6</sup>Laboratory of Experimental Virology, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. <sup>7</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK. <sup>8</sup>Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>9</sup>Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden. <sup>10</sup>Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden. <sup>11</sup>Division for HIV and Other Retroviruses, Department of Infectious Diseases, Robert Koch Institute, Berlin, Germany. <sup>12</sup>School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. <sup>13</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland. <sup>14</sup>Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland. <sup>15</sup>Department of Pathology, Johns Hopkins University, Baltimore, MD, USA. <sup>16</sup>Department of Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany. <sup>17</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland. <sup>18</sup>Institute of Medical Virology, University of Zurich, Zurich, Switzerland. <sup>19</sup>Department of Infectious Diseases, Helsinki University Hospital, Helsinki, Finland. <sup>20</sup>Division of Intramural Research, NIAID, NIH, Baltimore, MD, USA. <sup>21</sup>INSERM CESP U1018, Université Paris Saclay, APHP, Service de Santé Publique, Hôpital de Bicêtre, Le Kremlin-Bicêtre, France. <sup>22</sup>Institute for Global Health, University College London, London, UK. <sup>23</sup>Kymab Ltd., Cambridge, UK. <sup>24</sup>Department of Infectious Diseases, Faculty of Medicine, Imperial College London, London, UK. <sup>25</sup>Molecular Diagnostic Unit, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands. <sup>26</sup>Department of Global Health, Amsterdam University Medical Centers, University of Amsterdam and Amsterdam Institute for Global Health and Development, Amsterdam, Netherlands. \*Corresponding author. Email: chris.wymant@bdi.ox.ac.uk (C.W.); christophe.fraser@bdi.ox.ac.uk (C.F.) †Contributors and affiliations are listed in the supplementary materials.

subtype-B infections (expanding on the subset of 521 participants in ATHENA who were eligible for inclusion in BEEHIVE). We found 92 additional individuals infected with the viral variant, bringing the total to 109 such individuals in either dataset. When replicating the BEEHIVE test with the ATHENA data (Table 1, right column), we again observed a large rise in viral load in individuals with this viral variant: an increase of 0.54  $\log_{10}$  viral copies/ml (i.e., a ~3.5-fold increase). The effect size was the same in a linear model including age at diagnosis and sex as covariates, and persisted in newly diagnosed individuals over time (Fig. 1A). Henceforth, for brevity, we refer to this viral variant as the “VB variant” (for virulent subtype B), to individuals infected with this variant as “VB individuals,” and to individuals infected with a different strain of HIV as “non-VB individuals.”

### Search for closely related viruses

To test whether the variant was more widely disseminated, we searched publicly available databases for similar HIV viral genotypes. All results had <95% sequence similarity to a representative viral sequence for the variant. Of the 17 VB individuals originally found in BEEHIVE, one was from the Swiss HIV Cohort Study (22) (SHCS). By examining previously published data (23), we found that three other individuals from the SHCS were closely related (a phylogenetic distance below 2.5%). The high coverage of the Swiss HIV Cohort [including 89% of reported new infections from 2009 through 2018, with ~65% of the cohort sequenced (24)] makes it unlikely that

many more VB individuals in Switzerland were undetected. Data to assess viral load or CD4 cell decline for these three individuals were not available, owing to early initiation of treatment.

### More-rapid CD4 cell decline

At the time of diagnosis, CD4 counts for VB individuals were already lower than for non-VB individuals by 73 cells/mm<sup>3</sup> [95% confidence interval (CI): 12 to 134]. These counts subsequently declined faster, by a further 49 cells/mm<sup>3</sup> per year (CI: 20 to 79), in addition to the decline for comparable non-VB individuals [49 cells/mm<sup>3</sup> per year (CI: 46 to 51) for men diagnosed at the age of 30 to 39 years]. The VB variant is therefore associated with a doubling in the rate of CD4 cell decline. These values are averages estimated by using a linear mixed model adjusted for sex and age at diagnosis. Figure 1B illustrates the CD4 count decline that would be expected if disease progression were to continue linearly in the absence of treatment. Initiating treatment at a CD4 count of 350 cells/mm<sup>3</sup>, instead of immediately, was previously shown to substantially increase the subsequent hazard for serious adverse events (25). As seen in Fig. 1B, this stage of CD4 cell decline is expected to be reached in 9 months (CI: 2 to 17) from the time of diagnosis for VB individuals, as opposed to 36 months (CI: 33 to 39) for non-VB individuals, in males diagnosed at the age of 30 to 39 years. It is reached even more quickly in older age groups, for which we found progressively lower CD4 counts at time of diagnosis (table S1). At a CD4 count of 200 cells/

mm<sup>3</sup>, there is a high risk of immediate AIDS-related complications; without treatment this stage of decline would be reached, on average, between 2 and 3 years after diagnosis for VB individuals and between 6 and 7 years after diagnosis for comparable non-VB individuals [the latter being similar to previous reports in Europe (26)].

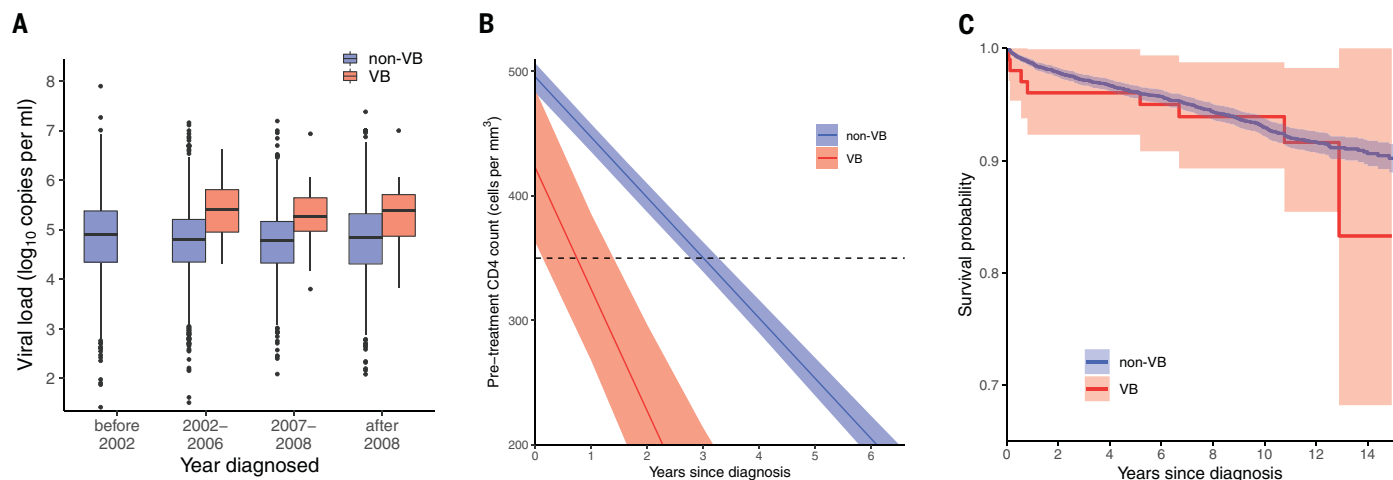
The effect of the VB variant on CD4 cell decline remained after we adjusted for the effect of higher viral load. With this adjustment, VB individuals have a CD4 count at diagnosis as would be expected given their high viral loads, but their subsequent decline in CD4 counts is again twice as fast as for as comparable non-VB individuals with high viral loads—their rate of decline is accelerated by 44 cells/mm<sup>3</sup> per year (CI: 16 to 72). Comparison of this additional decline with that expected from a +1 increase in  $\log_{10}$  viral load, 15 cells/mm<sup>3</sup> per year (CI: 11 to 18), shows that the variant’s effect on CD4 count decline is equivalent to that expected from a +3.0 increase in  $\log_{10}$  viral load. The same analysis of measurements of CD4 percentages (the percentage of all T cells that express CD4) showed that these also declined twice as fast for VB individuals, and again this doubling in speed of decline remained when we adjusted for the higher viral load of the variant (table S2 and fig. S1).

### No difference in CD4 cells after treatment, or in mortality

Measurements of treatment success include CD4 cell recovery and mortality. CD4 counts and percentages after treatment initiation were

**Table 1. Comparison of viral loads between individuals infected with the VB viral variant and other individuals.** When analyzing the viral loads of individuals in the ATHENA study, we first excluded individuals who were in BEEHIVE, so that the test would be independent of the initial finding within the BEEHIVE study. After our statistical tests of viral load, we did not exclude BEEHIVE individuals from the ATHENA data for subsequent analyses. *N*, number of individuals after those without viral load measurements before treatment were excluded; IQR, interquartile range.

Test	Discovery [BEEHIVE dataset (Europe)]	Replication [ATHENA dataset (Netherlands), excluding overlap with BEEHIVE]
Viral load measurements compared	Set-point viral loads for <i>N</i> = 15 VB individuals and <i>N</i> = 2446 individuals with any other HIV-1 strain	Mean pretreatment viral loads for <i>N</i> = 91 VB individuals and <i>N</i> = 5272 individuals with any other subtype-B HIV-1 strain
Mean and IQR of viral load in non-VB individuals, in $\log_{10}$ copies per milliliter	5.10 (IQR: 4.69 to 5.58)	4.79 (IQR: 4.34 to 5.27)
Mean and IQR of viral load in VB individuals, in $\log_{10}$ copies per milliliter	5.84 (IQR: 5.57 to 6.09)	5.33 (IQR: 4.94 to 5.75)
Viral load increase in VB individuals	0.74 $\log_{10}$ viral copies/ml $5 \times 10^{-6}$	0.54 $\log_{10}$ viral copies/ml
<i>P</i> value for increase	(two-tailed <i>t</i> test, significant at a level of $5 \times 10^{-5}$ when Bonferroni-corrected for performing 50 such tests)	$1 \times 10^{-12}$ (one-tailed <i>t</i> test)



**Fig. 1. Clinical characteristics of VB individuals.** Those infected with the highly virulent variant (VB individuals) are represented in red; those infected with any other subtype-B virus (non-VB individuals) are shown in blue. **(A)** Box-and-whisker plots of viral load, by year of diagnosis. Diagnosis dates were grouped to produce boundaries that coincide with years and roughly equal numbers of VB individuals (39 in 2002–2006, 35 in 2007–2008, and 27 after 2008; the pattern is robust to other groupings). **(B)** Expected decline in CD4 count in the

absence of treatment. The model was adjusted for sex and age at diagnosis; values shown are for males diagnosed at the age of 30 to 39 years. Shaded regions indicate 95% CIs in the model's prediction of mean values, given the uncertainty in estimation of parameter values (it does not reflect the variability between individuals in each of the two groups, which is much greater). The dashed black line denotes a CD4 count of 350 cells/mm<sup>3</sup> (see text for details). **(C)** Probability of still being alive at a given time after diagnosis.

similar for VB and non-VB individuals, as measured with both linear mixed modeling of the CD4 dynamics (tables S3 and S4 and fig. S2) and an individual-matching procedure. The hazard for death (from any cause) was also similar: VB individuals had a relative hazard of 1.4 (CI: 0.7 to 2.8,  $P = 0.35$ , Cox proportional hazards model). Our study had statistical power to detect only very large differences in mortality, as reflected in the wide CI for relative hazard for death and shown in Fig. 1C. VB individuals had similar CD4 counts and mortality after treatment despite a faster CD4 cell decline before treatment; this could be explained by their tendency to start treatment sooner after diagnosis (fig. S3). For example, although the probability of having started treatment was estimated to be similar 6 months after diagnosis [42% (CI: 41 to 44%) for non-VB individuals compared with 46% (CI: 35 to 54%) for VB individuals], it was different 2 years after diagnosis [65% (CI: 64 to 67%) for non-VB individuals and 93% (CI: 85 to 96%) for VB individuals]. Had VB individuals not started treatment earlier than others, lower CD4 counts at treatment initiation would have been expected, potentially causing increased morbidity and mortality (25). This information could be relevant if VB or variants like it are found in settings with less widespread availability of HIV care.

#### Characteristics of individuals infected with the VB variant

VB individuals were mostly (82%) men who have sex with men, similar to non-VB individ-

uals (76%). Age at diagnosis was also similar for VB and non-VB individuals (fig. S4). Neither ethnicity nor host genotype data were available, but the place of birth was mostly recorded as Western Europe for both groups (71% for non-VB individuals, 86% for VB individuals). VB individuals were present in all regions of the Netherlands, but with a different distribution relative to that of non-VB individuals ( $N = 102$  versus 6604 individuals,  $P < 10^{-7}$ , simulated Fisher's exact test): VB individuals were more common in the south (25% of VB individuals versus 6% of non-VB individuals) and less common in Amsterdam (20% versus 51%), as shown in table S5. Table S6 lists the hospitals included in each region. The average time from infection to diagnosis, for men who have sex with men in this cohort diagnosed in the late 2000s, was previously estimated to be 3.6 years (CI: 3.3 to 4.0) (27).

#### Genotype of the VB variant

Sequence data from the BEEHIVE project are whole-genome data, providing the 17 whole genomes available for the variant; sequence data from ATHENA are partial *pol* gene data only, available for the additional 92 VB individuals. We subtyped the 17 whole genomes for the variant as pure subtype B [with 100% support from two concordant methods (28, 29)], like most HIV-1 in the Netherlands. We predicted co-receptor usage from the 17 whole genomes using two concordant methods (30, 31): one was likely CXCR4-tropic; the other 16 were likely CCR5-tropic. Only one drug-resistance mutation was common for the VB

variant: Met<sup>41</sup>→Leu (M41L), present in 91 of 109 partial *pol* gene sequences. Without other linked resistance mutations, M41L causes only low-level resistance to zidovudine (32, 33). Two of the whole genomes were found to be recombinants between the VB variant and another subtype-B cluster in ATHENA (containing a small amount of sequence from the latter) and were excluded from subsequent sequence analysis. Among whole genomes in BEEHIVE and all whole genomes in the Los Alamos National Laboratory HIV Database (www.hiv.lanl.gov), none appeared to be a candidate for a “recombination parent” of the VB variant—i.e., the many mutations that distinguish the VB variant from any other known virus appear to have arisen de novo, not through recombination.

We compared the consensus sequence for the VB variant with the consensus of all Dutch subtype-B sequences in BEEHIVE, at both the amino acid and the nucleotide level: There were 250 amino acid changes and 509 nucleotide changes, as well as insertions and deletions. These alignments are included as data S1, and the amino acid alignment is illustrated in fig. S5. The distribution of nucleotide changes over the genome is in line with expectations (for example, fewer in the conserved *pol* gene region and more in the variable *env* gene region; see fig. S6). The VB-variant genotype is thus characterized by many mutations spread through the genome, meaning that a single genetic cause for the enhanced virulence cannot be determined from the current data.

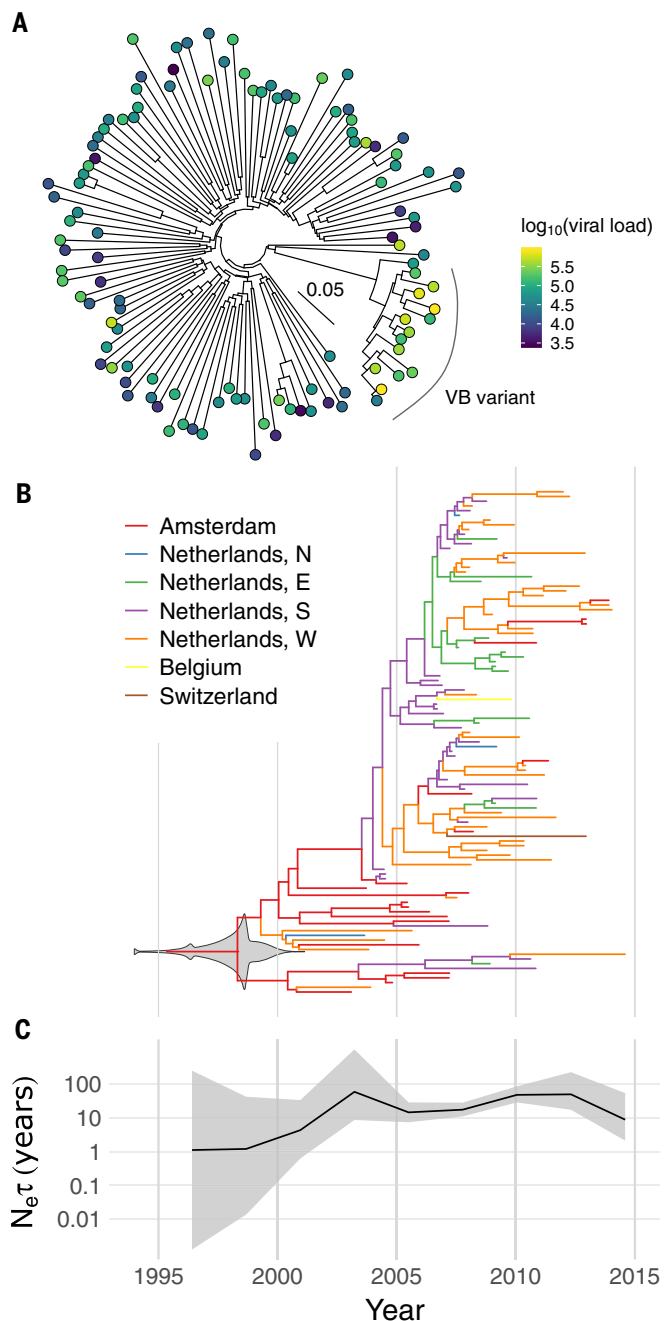
We conducted descriptive analyses of the mutations that distinguish the VB variant from the Dutch subtype-B consensus. All of the amino acid-level changes are listed in data S2 with annotations. Of the observed amino acid substitutions, 30 were previously shown to be positively associated with escape from cytotoxic T lymphocyte (CTL) response for at least one human leukocyte antigen type, and 13 were shown to be negatively associated (34). To provide context for these numbers, within Dutch subtype-B data in BEEHIVE we defined 16 other clades that are similar to the lineage in size (see materials and methods). For each clade, we calculated the amino acid consensus sequence, compared this to the Dutch subtype-B overall consensus, and determined CTL escape mutations. This showed that the number of such mutations for the VB variant is typical when normalized by its overall level of divergence (fig. S11). We also calculated the ratio of rates of nonsynonymous and synonymous changes ( $d_n/d_s$ ) for each gene, for the VB variant and the other 16 Dutch subtype-B clades used for comparison. The VB variant had lower  $d_n/d_s$  values than all of the other clades for *env*, *pol*, and *tat*, though its values were not extreme; for the other genes, its  $d_n/d_s$  value was in the range spanned by the other clades (fig. S12). Finally, at codon position 77 of the protein Vpr, the consensus of all Dutch subtype-B sequences in BEEHIVE is glutamine, whereas the VB consensus is arginine. Glutamine was previously found to be more common in long-term nonprogressors, and mutation to arginine increased T cell apoptosis in vitro and strongly increased T cell decline in mouse models (35). However, both alleles have been commonly observed in subtype B to date (of 2178 subtype-B Vpr protein sequences in the Los Alamos National Laboratory HIV Database, 52% have glutamine and 36% have arginine), making it implausible that this mutation alone is the dominant mechanism for the virulence effect we observed.

### Evolution of the VB variant

The maximum-likelihood phylogeny in Fig. 2A shows the VB variant in the context of background sequences, demonstrating that it is a distinct genetic cluster characterized by high viral loads. The phylogeny was inferred from 15 whole-genome VB-variant sequences and 100 randomly chosen whole-genome subtype-B background sequences from BEEHIVE. Figure 2B shows a dated phylogeny for VB-variant sequences only, estimated by using *BEAST* (36) and partial *pol* sequences. This phylogeny is colored by region, inferred with an ancestral state reconstruction by parsimony (minimizing changes of region). Amsterdam was assigned to the most recent common ancestor in 97% of trees in the posterior, showing that this reconstruction was robust to the uncertainty in

### Fig. 2. Phylogenetic and phylodynamic analysis of the VB variant.

(A) Whole-genome maximum-likelihood phylogeny of 15 VB-variant sequences and 100 background subtype-B sequences. The color of each circle indicates the individual's viral load in  $\log_{10}$  copies per milliliter. The inset scale bar shows the branch length scale in units of substitutions per site. (B) Dated maximum-clade-credibility tree for 107 partial *pol* gene sequences from the VB variant. Colors indicate geographical regions (N, E, S, and W: north, east, south, and west), which are known for the tips of the tree but are otherwise inferred by ancestral state reconstruction. The gray violin plot superimposed on the root node shows the posterior density for its date (i.e., the TMRCA); 1994 contains overflow to earlier dates for clarity. (C) Effective population size ( $N_e$ ) (scaled by the coalescent generation time  $\tau$ ) over time with 95% credibility intervals, with the same time axis as in (B).



the phylogeny. All VB-variant sequences date from 2003 onward; the time of their most recent common ancestor (TMRCA) was estimated as 1998.0 (95% credibility interval: 1995.7 to 2000.1). Trees were visualized by using *ggtree* (37).

### Phylodynamics of the VB variant

The effective population size ( $N_e$ ) of a pathogen is indicative of the number of infectious people. For the VB variant,  $N_e$  was estimated by using a skygrid demographic model (38) in *BEAST* and is shown in Fig. 2C (scaled by the coalescent generation time  $\tau$ ).  $N_e$  increased until roughly 2010; after this, there is more uncertainty but a possible downward trend

[which may be an artefact of  $N_e$  inference methods in the recent past (39)]. The proportion of VB-variant cases among all new subtype-B cases increased until a peak in 2008 and subsequently decreased, though again with appreciable uncertainty [absolute numbers of both VB and non-VB diagnoses in our dataset have been decreasing since roughly 2008, and the data are right-censored by several years (fig. S7)]. In a recent analysis of an updated version of the ATHENA dataset (40), 33 additional VB individuals were found, which suggests that VB diagnoses were stable until roughly 2013 and have since been declining, still with appreciable uncertainty.

We calculated the local branching index (LBI), which is a measure of fitness (41). For HIV in a context in which most individuals start treatment without long delays, the LBI is closely related to transmissibility (see supplementary text). Compared with that of other transmission clusters, the LBI was higher for the VB variant both in BEEHIVE ( $P = 2 \times 10^{-7}$ ) and ATHENA ( $P < 2 \times 10^{-16}$ ; fig. S8). High pretreatment transmissibility may explain why the VB variant grew to be the 10th largest of 1783 clusters in the full ATHENA tree.

### Tree imbalance and evolution within the VB-variant clade

We found nothing unusual in the extent to which the VB variant's phylogeny is imbalanced, nor did we detect any indication of further evolution of viral load within the variant's clade (supplementary text and fig. S9).

### The first sampled VB individual

We retrieved and sequenced two additional samples from the VB individual who was diagnosed in 1992, 10 years before subsequent diagnoses of other VB individuals. Phylogenetic analysis suggested that this individual was infected with a virus that had evolved most of the way, but not entirely, toward VB-variant viruses typical of later dates (supplementary text and fig. S10). This individual was diagnosed in Amsterdam, consistent with the aforementioned ancestral reconstruction of region. In the 10 years before this first VB diagnosis, the proportion of individuals diagnosed in the Netherlands for whom a viral sequence was available was roughly one-third. The proportion of those diagnosed or undiagnosed would be smaller still. This means that the infector of the 1992 individual was most likely not sampled, and indeed two or three steps in the transmission chain could have been unsampled. The long phylogenetic branch leading to the 1992 individual could therefore represent between-host evolution, not necessarily within-host evolution in a single individual.

### Discussion

Previous studies of the heritability of viral load and CD4 cell decline led us to expect that these properties could change with the emergence of a new variant of HIV-1. We provided strong evidence for this, discovering a virulent subtype-B variant (the VB variant) that has been circulating in the Netherlands since the late 1990s. We characterized the variant's genotype and evolutionary history, as well as its association with high viral loads, rapid decline of CD4 cells, and increased transmissibility. We found 109 individuals with the variant (VB individuals) whose age, sex, suspected mode of transmission, and region of birth are all typical for people living with HIV in the Netherlands. This suggests that the ob-

served association is causal: The increased virulence is a property of the virus rather than a confounding property of individuals in this transmission cluster. An absence of viral load evolution inside the clade of VB variants suggests that the increased virulence is a property of the whole clade and not a subset of it—i.e., that the virulence evolution occurred on the long phylogenetic branch that connects this clade to other known viruses.

Deferring the initiation of treatment until the measurement of a CD4 count's decline to  $\leq 350$  cells/mm<sup>3</sup> or the onset of AIDS, instead of immediate treatment initiation upon diagnosis, was previously shown to increase the subsequent hazard of serious AIDS-related events by a factor of 3.6 (CI: 2.0 to 6.7) and of any serious event (including death) by a factor of 2.4 (CI: 1.6 to 3.3) (25). This long-lasting immunological damage justifies WHO's classification of 350 CD4 cells/mm<sup>3</sup> as “advanced HIV” ([www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf](http://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf)). Without treatment, advanced HIV is expected to be reached in only 9 months (CI: 2 to 17) from the time of diagnosis for VB individuals, compared with 36 months (CI: 33 to 39) for non-VB individuals, in males diagnosed at the age of 30 to 39 years. Advanced HIV is reached even more quickly in older age groups, and there is considerable variation between individuals around these expected values. Many individuals could therefore progress to advanced HIV by the time they are diagnosed, with a poorer prognosis expected thereafter in spite of treatment. In practice, there is still substantial variation in the delay from becoming infected to starting treatment, making the VB variant a concern even in the high-awareness and highly monitored context of the Dutch HIV-1 epidemic. In contexts with less awareness and monitoring, in which diagnosis often occurs later in infection, the probability of reaching advanced HIV before diagnosis would be even greater.

Future *in vitro* investigations could more firmly establish the role of the viral genotype, and reveal an as-yet-unknown virulence mechanism at the molecular or cellular level. A higher replicative capacity of the virus might be observed, given the increased viral loads seen here. However, it is likely that there will be more to the virulence mechanism: The VB variant doubles the rate of CD4 cell decline, measured with both counts and T cell percentages, even after adjusting for its higher viral load. This rate is equivalent to the acceleration of CD4 degradation that would be expected from a 3.0 log<sub>10</sub> increase in viral load, though we observed a 0.54 to 0.74 log<sub>10</sub> increase. This means that the virulence normalized by the amount of virus—the “per-parasite pathogenicity” (42, 43), which for HIV is heritable (19)—is much higher for the VB variant. Using two aforementioned methods, we predicted

that, of the 17 whole genomes available, 16 use only the R5 co-receptor for cell entry, which is typical for subtype-B viruses in early infection (13). This finding suggests that the underlying virulence mechanism is distinct from the well-known effect of cell tropism (14, 15).

Previous studies have reported population-wide increases (44, 45) and decreases (46) in virulence over time. Mixed results between individual studies [see (47) for a meta-analysis] can be attributed to differences in epidemic context (such as the dominant subtypes), statistical power, and observational biases over time. Temporal virulence trends could also be due to changing confounders, such as a shift in which subpopulations are most affected, the stage of infection at time of diagnosis, or coinfections. We expand on these studies by resolving a change in virulence to an individual viral variant.

The basic theory of an infectiousness–virulence trade-off is that infectiousness and virulence are linked (for example, by how fast a pathogen replicates in its host) and that selection pressures favor intermediate values rather than extreme ones. If infectiousness is too low, the pathogen cannot be transmitted when its host contacts other hosts, but if virulence is too high, the host becomes too ill to have such contacts. In the case of HIV, the implication of this theory is that we would not expect highly virulent viruses to spread widely through a population in the absence of widespread treatment, because their hosts would progress to AIDS very quickly, limiting the opportunities for transmission (9). Most of the evolution that gave rise to the VB variant occurred before 1992, before effective combination treatment was available. However, our findings may stimulate further interest in whether widespread treatment shifts the balance of the infectiousness–virulence trade-off toward higher virulence, thus promoting the emergence and spread of new virulent variants. Previous modeling studies have investigated this idea for pathogens generally (48) and for HIV specifically (49, 50). We discuss subtleties of the argument in the supplementary text, but our conclusion is that widespread treatment is helpful to prevent new virulent variants, not harmful. The absolute fitness of viral variants must be considered in addition to their relative fitness, and treatment reduces the total onward transmission over the course of one infection, regardless of virulence. Put simply, “viruses cannot mutate if they cannot replicate” (anonymous), and “the best way to stop it changing is to stop it” [Marc Lipsitch (51)]. Early treatment also prevents CD4 cell decline from leading to later morbidity and mortality; thus clinical, epidemiological, and evolutionary considerations are aligned. Our discovery of a highly virulent and transmissible viral variant therefore emphasizes the importance of access

to frequent testing for at-risk individuals and of adherence to recommendations for immediate treatment initiation for every person living with HIV ([www.who.int/hiv/pub/arv/](http://www.who.int/hiv/pub/arv/)).

#### REFERENCES AND NOTES

- J. L. Geoghegan, E. C. Holmes, *Nat. Rev. Genet.* **19**, 756–769 (2018).
- C. E. Cressler, D. V. McLeod, C. Rozins, J. van den Hoogen, T. Day, *Parasitology* **143**, 915–930 (2016).
- J. J. Bull, A. S. Luring, *PLOS Pathog.* **10**, e1004387 (2014).
- R. Challen *et al.*, *BMJ* **372**, n579 (2021).
- N. G. Davies *et al.*, *Nature* **593**, 266–269 (2021).
- D. J. Grint *et al.*, *Euro Surveill.* **26**, (2021).
- N. G. Davies *et al.*, *Science* **372**, eabg3055 (2021).
- E. Volz *et al.*, *Nature* **593**, 266–269 (2021).
- C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, W. P. Hanage, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17441–17446 (2007).
- P. M. Sharp, B. H. Hahn, *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011).
- M. Worobey *et al.*, *Nature* **455**, 661–664 (2008).
- J. Hemelaar *et al.*, *Lancet Infect. Dis.* **19**, 143–155 (2019).
- B. S. Taylor, M. E. Sobieszczyk, F. E. McCutchan, S. M. Hammer, *N. Engl. J. Med.* **358**, 1590–1602 (2008).
- B. Asjö *et al.*, *Lancet* **2**, 660–662 (1986).
- M. Koot *et al.*, *Ann. Intern. Med.* **118**, 681–688 (1993).
- H. Song *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 239–244 (2019).
- C. Fraser *et al.*, *Science* **343**, 1243727 (2014).
- F. Blanquart *et al.*, *PLOS Biol.* **15**, e2001855 (2017).
- F. Bertels *et al.*, *Mol. Biol. Evol.* **35**, 27–37 (2018).
- C. Fraser, T. D. Hollingsworth, *AIDS* **24**, 2596–2597 (2010).
- V. Mitov, T. Stadler, *Mol. Biol. Evol.* **35**, 756–772 (2018).
- F. Schoeni-Affolter *et al.*, *Int. J. Epidemiol.* **39**, 1179–1189 (2010).
- K. Kusejko *et al.*, *J. Infect. Dis.* **220**, 244–253 (2019).
- A. U. Scherrer *et al.*, *Int. J. Epidemiol.* [10.1093/ije/dyab141](https://doi.org/10.1093/ije/dyab141) (2021).
- J. D. Lundgren *et al.*, *N. Engl. J. Med.* **373**, 795–807 (2015).
- S. Lodi *et al.*, *Clin. Infect. Dis.* **53**, 817–825 (2011).
- A. van Sighem *et al.*, *Epidemiology* **26**, 653–660 (2015).
- A.-C. Pineda-Peña *et al.*, *Infect. Genet. Evol.* **19**, 337–348 (2013).
- D. Struck, G. Lawyer, A.-M. Ternes, J.-C. Schmit, D. P. Bercoff, *Nucleic Acids Res.* **42**, e144 (2014).
- M. A. Jensen *et al.*, *J. Virol.* **77**, 13376–13388 (2003).
- M. C. Prosperi *et al.*, *AIDS Res. Hum. Retroviruses* **25**, 305–314 (2009).
- P. Kellam, C. A. Boucher, B. A. Larder, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1934–1938 (1992).
- S.-Y. Rhee *et al.*, *Nucleic Acids Res.* **31**, 298–303 (2003).
- J. M. Carlson *et al.*, *J. Virol.* **86**, 13202–13216 (2012).
- J. J. Lum *et al.*, *J. Clin. Invest.* **111**, 1547–1554 (2003).
- M. A. Suchard *et al.*, *Virus Evol.* **4**, vey016 (2018).
- G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, *Methods Ecol. Evol.* **8**, 28–36 (2017).
- M. S. Gill *et al.*, *Mol. Biol. Evol.* **30**, 713–724 (2013).
- E. de Silva, N. M. Ferguson, C. Fraser, *J. R. Soc. Interface* **9**, 1797–1808 (2012).
- D. Bezemer *et al.*, *AIDS* **36**, 83–94 (2022).
- R. A. Neher, C. A. Russell, B. I. Shraiman, *eLife* **3**, e03568 (2014).
- L. Råberg, M. Stjernman, in *Ecoimmunology*, G. Demas, R. Nelson, Eds. (Oxford Univ. Press, 2011), pp. 548–578.
- L. Råberg, *PLOS Biol.* **12**, e1001989 (2014).
- N. Pantazis *et al.*, *Lancet HIV* **1**, e119–e126 (2014).
- J. O. Wertheim *et al.*, *Nat. Commun.* **10**, 5788 (2019).
- F. Blanquart *et al.*, *eLife* **5**, e20492 (2016).
- J. T. Herbeck *et al.*, *AIDS* **26**, 193–205 (2012).
- T. C. Porco, J. O. Lloyd-Smith, K. L. Gross, A. P. Galvani, *J. Theor. Biol.* **233**, 91–102 (2005).
- J. T. Herbeck *et al.*, *Virus Evol.* **2**, vew028 (2016).
- H. E. Roberts, P. J. R. Goulder, A. R. McLean, *J. R. Soc. Interface* **12**, 20150888 (2015).
- Marc Lipsitch (@mlipsitch), Twitter, 15 February 2020, <https://twitter.com/mlipsitch/status/1228734360716791815>.
- C. Wymant, C. Fraser, F. Blanquart, Analysis code for “A highly virulent variant of HIV-1 circulating in the Netherlands”, Zenodo, version 1 (2021); <https://doi.org/10.5281/zenodo.5761935>.

#### ACKNOWLEDGMENTS

We thank K. Fransen and G. Vanham for help with the Belgian data, O. Ratmann for help in identifying the Dutch clusters, K. Kusejko for testing for additional VB individuals in the SHCS, B. Foley for help with genome sharing, B. Dearlove and L. Thomson for help with software, and J. Herbeck and three other reviewers for helpful suggestions. **Funding:** This study was funded by ERC Advanced Grant PBDR-339251 and a Li Ka Shing Foundation grant, both awarded to C.F. The ATHENA Cohort is managed by Stichting HIV Monitoring and supported by a grant from the Dutch Ministry of Health, Welfare and Sport through the Centre for Infectious Disease Control of the National Institute for Public Health and the Environment. **Author contributions:** Funding acquisition: C.F. Conceptualization: C.W. and C.F. Data generation and management: all authors. Investigation: C.W., F.B., D.B., L.F., P.R., and C.F. Writing, original draft: C.W. and C.F. Writing, review and editing: all authors. **Competing interests:** P.K. is an employee of Kymab, a Sanofi company. H.F.G. reports grants from the Swiss National Science Foundation, National Institutes of Health (NIH), and the Swiss HIV Cohort Study; unrestricted research grants from Gilead Sciences, Roche, and the Yvonne Jacob Foundation; and personal fees from consulting or advisory boards or data safety monitoring boards for Merck, Gilead Sciences, Viiv Healthcare, Mepha, and Sandoz. H.F.G.'s institution received money for participation in the following clinical COVID-19 studies: 540-7773/5774 (Gilead), TICO (ACTIV-3, INSIGHT/NIH), and the Morningsky study (Roche). **Data and materials availability:** Code illustrating the analysis of the source clinical data, and of the genomic

distribution and annotation of VB-variant mutations, is openly available at GitHub ([https://github.com/ChrisHIV/hiv\\_vb\\_variant](https://github.com/ChrisHIV/hiv_vb_variant)); a version has also been deposited at Zenodo (52). The 17 VB-variant whole genomes are publicly available at GenBank with accession numbers MT458931 to MT458935 and MW689459 to MW689470; the two putative recombinants have accession numbers MW689465 and MW689466. Data on viral loads, pretreatment CD4 counts, and mortality are provided as data S3. Requests for further data access can be made by submission of a concept sheet to the corresponding authors; these will be reviewed on a case-by-case basis, given that the data underlying this study contain sensitive and potentially identifying information. Once submitted, the proposed research and/or analysis will undergo review by the BEEHIVE Data Access Committee, which includes representatives of the ATHENA Cohort, for evaluation of scientific value, relevance to the study, design and feasibility, statistical power, and overlap with existing projects. If the proposed analysis is for verification and/or replication, data will then be made available. If the proposed research is for novel science, upon completion of the review, feedback will be provided to the proposer(s). In some circumstances, a revision of the concept may be requested. If the concept is approved for implementation, a writing group will be established that will consist of the proposers (up to three persons that were centrally involved in the development of the concept) and members of the BEEHIVE Collaboration and ATHENA Cohort (or other appointed cohort representatives). All persons involved in the process of reviewing these research concepts are bound by confidentiality. **Ethics statement:** At initiation, the ATHENA Cohort was approved by the institutional review board of all participating institutions. People beginning HIV care receive written material about participation in the ATHENA study and are informed by their treating physician of the purpose of data collection, after which they can consent verbally or opt out. Data are pseudonymized before being provided to investigators and may be used for scientific purposes. A designated quality management coordinator safeguards compliance with the European General Data Protection Regulation. Additional written informed consent was obtained for ATHENA participants enrolled in BEEHIVE for whole-genome sequencing.

#### SUPPLEMENTARY MATERIALS

[science.org/doi/10.1126/science.abk1688](https://science.org/doi/10.1126/science.abk1688)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S12  
Tables S1 to S7  
References (53–85)  
MDAR Reproducibility Checklist  
Data S1 to S4

25 June 2021; accepted 4 January 2022  
10.1126/science.abk1688