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

Netherlands ATHENA HIV Observation

2022-02-04

Netherlands ATHENA HIV Observation , BEEHIVE Collaboration , Wymant , C , Bezemer , D , Blanquart , F , Kivelä , P & Ristola , M 2022 , ' A highly virulent variant of HIV-1 circulating in the Netherlands ' , Science , vol. 375 , no. 6580 , pp. 540-+ . https://doi.org/10.1126/science.abk1688

http://hdl.handle.net/10138/351933
https://doi.org/10.1126/science.abk1688

acceptedVersion

Please check all editing changes carefully and respond to the queries that are shown in balloons in the right margins. Inserted text is blue; deleted text is red with strike-through marks. In the reference list, article titles will appear online, but not in print or in the galley proofs. Also, the full reference list (included in this document) will appear online, but not in print. Please mark all changes directly on the galleys and return the galleys to me.

Lauren Kmec
Lead Content Production Editor
Science
Email: lkmec@aaas.org

HIV

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

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

[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

[2]Stichting HIV Monitoring, Amsterdam, ~~the~~ Netherlands.

[3]Centre for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France.

[4]IAME, UMR 1137, INSERM, Université de Paris, Paris, France.

[5]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK.

[6]Laboratory of Experimental Virology, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, ~~the~~ Netherlands.

[7]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK.

[8]Wellcome Centre for Human Genetics, University of Oxford, Oxford, ~~United Kingdom~~**UK**.

[9]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden.

[10]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden.

[11]Division for HIV and Other Retroviruses, **Department of Infectious Diseases,** Robert Koch Institute, Berlin, Germany.

[12]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

[13]Swiss Institute of Bioinformatics, Lausanne, Switzerland.

[14]Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.

[15]Department of Pathology, John Hopkins University, Baltimore, MD, USA.

[16]Department of Infectious Disease Epidemiology, Robert Koch- Institute, Berlin, Germany.

[17]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland.

[18]Institute of Medical Virology, University of Zurich, Zurich, Switzerland.

[19]Department of Infectious Diseases, Helsinki University Hospital, Helsinki, Finland.

[20]Division of Intramural Research, NIAID, NIH, Baltimore, MD, USA.

[21]INSERM CESP U1018, Université Paris Saclay, APHP, Service de Santé Publique, Hôpital de Bicêtre, Le Kremlin-Bicêtre, France.

Commented [LK1]: Correct department name?

[22]Institute for Global Health, University College London, London, UK.

[23]Kymab Ltd., Cambridge, UK.

[24]Department of Infectious Diseases~~,~~ and Department of Medicine, Imperial College London, London, UK.

[25]Molecular Diagnostic Unit, Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, ~~the~~ Netherlands.

[26]Department of Global Health, Amsterdam University Medical Centers, University of Amsterdam and Amsterdam Institute for Global Health and Development, Amsterdam, ~~the~~ Netherlands.
*Corresponding author~~s~~. Email: chris.wymant@bdi.ox.ac.uk (C.W.); christophe.fraser@bdi.ox.ac.uk (C.F.)

†~~Full list of c~~Contributors and affiliations are listed in the ~~S~~supplementary materials~~Text~~.

We discovered a highly virulent variant of subtype-B HIV-1 in the Netherlands. One hundred two~~102~~ individuals with this variant had ~~viral loads~~ a 0.54 to~~–~~ 0.74 $\log_{10}$ increase (i.e., a ~3.47-fold to 5.50-fold increase) in viral load compared with~~copies higher~~, and exhibited CD4 cell decline twice as fast as~~2 times faster~~, ~~than~~ 6~~,~~604 individuals with other subtype-B strains. Without treatment, advanced HIV~~——~~CD4 cell counts below 350 cells per cubic millimeter~~mm~~^{3}, with long-term clinical consequences~~——~~is expected to be reached, on average, 9 months after diagnosis for individuals in their thirties with this variant. ~~The 102 individuals had typical a~~Age, sex, suspected mode of transmission, and place of birth for the aforementioned 102 individuals were typical for HIV-positive people in the Netherlands, which suggests~~ing that~~ the ~~effect~~ increased virulence is attributable to the ~~virus~~viral strain. Genetic sequence analysis suggest~~s~~~~ed~~ 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~~,—~~ thus causing greater damage to their host~~s~~~~,—~~ has been much studied theoretically~~,~~ despite few population-level examples (*1*–*3*). The most notable recent example is the B.1.617.2 lineage (Delta 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 ~~large~~ high adaptive potential. Greater virulence could benefit ~~the~~ a virus if it is not outweighed by reduced opportunity for transmission. These antagonistic selection pressures may result in an intermediate level of virulence that is 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).

Commented [LK2]: Per journal style, a sentence should not begin with a numeral.

Commented [LK3]: "0.54 to 0.74 $\log_{10}$ copies higher" is rather jargony and will probably be confusing for non-specialists. Is the edited text accurate? If not, please revise for clarity.

Commented [LK4]: We spell out units in the abstract.

Commented [LK5]: Edited for clarity – changes okay?

Commented [LK6]: Can this be changed to "has been extensively studied in theory"?

Commented [LK7]: Okay as edited? Or perhaps "high potential for adaptation"?

Commented [LK8]: You are referring to a virus in general here, correct (not to a specific virus)?

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 ~~area around 1920~~ (*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 ~~ying~~ 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 $CD4^+$ T ~~-~~ cells in peripheral blood, which tracks immune system damage by the virus). Successful treatment with antiretroviral 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 ~~: we~~ by report ing a subtype ~~–~~B variant of HIV-1 with exceptionally high virulence ~~,~~ that has been circulating within the Netherlands during the ~~l~~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, ~~first~~ middle column). BEEHIVE is a study of individuals enrolled in eight cohorts across Europe and Uganda ~~,~~ who were selected ~~to~~ because they have well-characterized dates of infection and samples available from early infection, ~~,~~ and for whom whole viral genomes were sequenced. The 17 individuals with the distinct viral variant comprised 15 from the ATHENA cohort in the Netherlands, ~~one~~ 1 from Switzerland, and ~~one~~ 1 from Belgium. See materials and ~~M~~methods for details on the initial discovery.

Replication of the discovery in the Dutch ATHENA cohort

To replicate the finding and to investigate this viral variant in more detail, we then analyzed data from 6 ~~,~~706 individuals in ATHENA with subtype-B infections (expanding on the subset of 521 individuals from 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 ~~in~~ with the ATHENA data (Table 1, ~~second~~ right column), we again observed a large ~~increase~~ rise in viral load in individuals with this viral variant: an increase of 0.54 $\log_{10}$ copies/ml ~~per ml~~ (i.e., a ~3.47-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

Commented [LK9]: It is a bit confusing that "ATHENA cohort" is used in different places to refer to both the patient group and the researcher group. Given that the official name of the collaboration includes the word "Cohort," might the patient group instead be referred to as "ATHENA participants," "individuals in the ATHENA study," or some other term?

Commented [LK10]: Correct?

brevity, we refer to this viral variant as the "VB variant" (for ~~V~~virulent subtype B), to individuals infected **with**~~by~~ this variant as "VB individuals," and to individuals infected with a different strain of HIV as "~~not~~**non**-VB individuals."

Commented [LK11]: Changed to "non-VB individuals" throughout text and figures, as discussed previously via email.

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 ~~less than~~ <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** ~~E~~examining previously published data (*23*), **we found that** three other individuals from the SHCS were ~~found to be~~ 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 ~~roughly~~ ~65% of the cohort sequenced (*24*)~~)~~**]** makes it unlikely that ~~there were~~ many more VB individuals in Switzerland ~~who~~ were ~~missed~~**overlooked**. Data to assess viral load or CD4 cell decline for these three individuals ~~was~~ **were** not available**,** ~~due~~ **owing** to early **initiation of** treatment ~~initiation~~.

Commented [LK12]: As meant?

More ~~-~~rapid CD4 cell decline

Commented [LK13]: Hyphenated for clarity (i.e., to indicate faster decline as opposed to additional and fast decline).

At the time of diagnosis, CD4 counts for VB individuals were already lower **than for non-VB individuals** ~~at time of diagnosis,~~ by 73 cells/**mm**^{3} ~~per mm^3~~ **[**~~(~~**95%** confidence interval (CI)**:** 12 to 134~~)~~**]**. These~~y~~ **counts** subsequently declined faster**,** by a~~n~~ **further**~~additional~~ 49 cells/**mm**^{3} ~~per mm^3~~ per year (CI**:** 20 to 79)**,** ~~on top of~~ **in addition to** the decline for comparable ~~not~~**non**-VB individuals~~,~~ **[**~~which is~~ 49 cells/**mm**^{3} ~~per mm^3~~ per year (CI**:** 46 to 51) for men diagnosed **at the** age~~d~~ **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 adjust**ed**~~ing~~ 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**^{3} ~~per mm^3~~, 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 reached in 9 months (CI**:** 2 to 17) from **the** time of diagnosis for VB individuals, ~~compared~~ **as opposed** to 36 months (CI**:** 33 to 39) for ~~not~~**non**-VB individuals, in males diagnosed **at the** age~~d~~ **of** 30 **to**~~-~~ 39 years. It is reached even more quickly in older age groups, ~~whom~~ **for which** we found ~~to have~~ progressively lower CD4 counts at time of diagnosis (~~Supplementary T~~table S1). At a CD4 count of 200 cells/**mm**^{3} ~~per mm^3~~, there is a high risk of immediate AIDS-related complications; this stage of decline would be reached**,** on average**,** between ~~two~~ **2** and ~~three~~ **3** years after diagnosis for VB individuals~~,~~ and between ~~six~~ **6** and ~~seven~~ **7** years after diagnosis for comparable ~~not~~**non**-VB individuals **[**~~(~~the latter being similar to previous reports in Europe (*26*)~~)~~**]**.

Commented [LK14]: Correct?

Commented [LK15]: 95% CI, correct?

The effect of the VB variant on CD4 cell decline remained after **we** adjust**ed**~~ing~~ 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**^{3} ~~per mm^3~~ per year ~~greater~~ (CI**:** 16 to 72). ~~Comparing~~ **Comparison of** this additional decline with that expected from a +1 increase in $\log_{10}$ viral load, 15 cells/**mm**^{3} ~~per mm^3~~ per year (CI**:** 11 to 18), shows that the variant's effect on CD4 count

Commented [LK16]: Correct?

decline ~~was~~ **is** equivalent to that expected from ~~having a viral load~~ **a** +3.0 **increase in** $\log_{10}$ **viral load**~~copies higher~~. The same analysis of measurements of CD4 percentages (the percent**age** 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** adjust**ed**~~ing~~ for the higher viral load of the variant (~~Supplementary T~~table S2~~,~~ **and** ~~Supplementary Figure~~ **fig.** S1).

Commented [LK17]: As meant?

No difference in CD4 cells after treatment, or in mortality

Measurements of ~~the~~ **treatment** success ~~of treatment~~ include CD4 cell recovery~~,~~ and mortality. CD4 counts and percentages after treatment initiation were similar for VB and ~~not~~**non**-VB individuals, as measured with both linear mixed model~~l~~ing of the CD4 dynamics (~~Supplementary T~~table**s** S3 and S4~~,~~ and ~~Supplementary Figure~~ **fig.** S2) and an individual-matching procedure ~~agnostic of~~ **in which** the dynamics **were unknown**. 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~~p~~ = 0.35, Cox proportional hazards model). Our study had **the ability**~~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 (~~shown in Supplementary Figure~~ **fig.** S3). For example, **al**though the probability of having started treatment was estimated to be similar ~~at~~ 6 months after diagnosis~~,~~ **[**42% (CI**:** 41 to 44%) for ~~not~~**non**-VB individuals compared **with**~~to~~ 46% (CI**:** 35 to 54%) for VB individuals**]**, ~~at 2 years after diagnosis~~ it was different **2 years after diagnosis**~~:~~ **[**65% (CI**:** 64 to 67%) for ~~not~~**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*).~~.~~ ~~t~~**T**his **information** could be relevant ~~should~~ **if** VB or variants like it **are**~~be~~ found in settings with less widespread availability of AIDS care.

Commented [LK18]: Or “of treatment efficacy”?

Commented [LK19]: As meant?

Characteristics of individuals infected with the VB variant

VB individuals were mostly (82%) men who have sex with men, similar to ~~not~~**non**-VB individuals (76%). Age at diagnosis was also similar for VB and ~~not~~**non**-VB individuals (~~Supplementary Figure~~ **fig.** S4). Neither ethnicity nor host genotype data **were**~~was~~ available, but the place of birth was mostly recorded as Western Europe for both groups (71% for ~~not~~**non**-VB individuals, 86% for VB individuals). VB individuals were present in all regions of the Netherlands, but with a different distribution ~~compared~~ **relative** to **that of** ~~not~~**non**-VB individuals (N = 102 versus ~~N =~~ 6604 **individuals**, P~~p~~ < 10^{-7}, simulated Fisher’s exact test)~~:~~. VB individuals were more common in the south (25% of VB individuals versus 6% of ~~not~~**non**-VB individuals) and less common in Amsterdam (20% versus 51%), as shown in ~~Supplementary T~~table S5. ~~Supplementary~~ 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 3.3 years (CI**:** 3.3~~,~~ **to** 4.0) (*27*).

Genotype of the VB variant

Sequence data from the BEEHIVE project **are**~~is~~ whole-~~-~~genome **data**, providing the 17 whole genomes available for the variant; sequence data from ATHENA ~~is~~ **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**~~using~~ two concordant methods (*28*,

29)~~)~~], like most HIV-1 in the Netherlands. We predicted co-receptor usage from the 17 whole genomes **by** 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^{41}→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~~ing~~ 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: ~~t~~**T**here were 250 amino acid changes and 509 nucleotide changes, as well as insertions and deletions. These alignments are included as ~~Supplementary D~~**d**ata S1, and the amino acid alignment is illustrated in ~~Supplementary Figure~~ **fig.** S5. The distribution of nucleotide changes over the genome is in line with expectations (for example**,** ~~less~~ **fewer** in the conserved *pol* gene region and more in the variable *env* gene region; see ~~Supplementary Figure~~ **fig.** S6). The VB-~~-~~variant genotype is thus ~~characterised~~ **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~~ing~~ the VB variant from the Dutch subtype-B consensus. All **of** the amino~~-~~ acid-~~-~~level changes are listed in ~~Supplementary~~ data S2 with annotations. ~~30 o~~**O**f the **observed** amino acid substitutions ~~observed~~**,** **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 ~~of~~ **that are** similar ~~size~~ to the lineage **in size** (see **materials and** ~~M~~**m**ethods). 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 ~~normalised~~ **normalized** by its overall level of divergence (~~Supplementary Figure~~ **fig.** S11). We also calculated the ratio of rates of non~~-~~synonymous and synonymous changes ~~—(~~d_n/d_s~~dn/ds)—~~ for each gene, for the VB variant**,** and **for** the other 16 Dutch subtype-~~-~~B clades used for comparison. The VB variant had lower d_n/d_s~~dn/ds~~ values than all of the other clades **for**~~in~~ *env*, *pol*, and *tat*, though its values were not extreme~~,~~**;** ~~and~~ for the other genes**,** its d_n/d_s~~dn/ds~~ value was in the range spanned by the other clades (~~Supplementary Figure~~ **fig.** S12). Finally, ~~we noted that~~ at codon position 77 of **the protein** Vpr, the consensus of all Dutch subtype-B sequences in BEEHIVE is ~~G~~**g**lutamine, **whereas**~~while~~ the VB consensus is ~~A~~**a**rginine. Glutamine was previously found to be more common in long-term non~~-~~progressors, and mutation to ~~A~~**a**rginine increased T cell apoptosis in vitro and strongly increased T cell decline in ~~mice~~ **mouse** models (*35*). However, both alleles have been commonly observed in subtype B to date (of 2178 subtype-B Vpr protein sequences **in**~~on~~ the Los Alamos National Laboratory HIV Database, 52% have ~~G~~**g**lutamine~~,~~ **and** 36% have ~~A~~**a**rginine), making it implausible that this mutation alone is the dominant mechanism for the virulence effect **we** observed~~here~~.

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*) using and partial *pol* sequences. ItThe diagram is colored by region, inferred with an ancestral state reconstruction by parsimony (minimizing changes of region). This assigned 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 the phylogeny. All VB-variant sequences date from 2003 onwards; 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, this 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 τ). N_e increased until roughly 2010; after this, there is more uncertainty but a possible downward trend [(which can 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 diagnoses that are VB variant increased until a peak in 2008, and subsequently decreased after that, though again with appreciable uncertainty; absolute numbers of both VB and notnon-VB diagnoses in our dataset have been decreasing since roughly 2008 (Supplementary Figure fig. S7). In a recent analysis of an updated version of the ATHENA dataset (*40*), 33 additional VB individuals were found, which these suggests that VB diagnoses were stable until roughly 2013, and have since been declining since, though albeit with large substantial uncertainty (the dataset is right-censored by several years; Supplementary Figure fig. S7).

We calculated the local branching index (LBI), which is a measure of fitness (*41*). For HIV in a context where in which most individuals start treatment without long delays, the LBI is closely related to transmissibility (see Ssupplementary Ttext). Compared withto that of other transmission clusters, the LBI was higher for the VB variant both in BEEHIVE ($P_p = 2 \times 10^{-7}$) and ATHENA ($P_p < 2 \times 10^{-16}$; Supplementary Figure fig. S8). High pre-treatment transmissibility could 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 any further evolution of viral load within the variant's clade (Ssupplementary Ttext, and Supplementary Figure fig. S9).

The first sampled VB individual

We retrieved and sequenced two additional samples from the VB individual who was diagnosed in 1992, ten 10 years before subsequent diagnoses of other VB individuals, in 1992. Phylogenetic analysis suggested that this individual was infected with a virus that had evolved most of the way, but not entirelyall of the way, toward VB-variant viruses typical of later dates

Commented [LK20]: Correct?

Commented [LK21]: Okay as edited?

(~~S~~supplementary ~~T~~text~~;~~ **and** ~~Supplementary Figure~~ **fig.** S10). This individual was diagnosed in Amsterdam, consistent with the ancestral reconstruction of region. ~~Over~~ **In** the ~~ten~~ **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 **it is plausible that** two or three steps in the transmission chain ~~could easily have been~~ **were also** unsampled. The long phylogenetic branch leading to the 1992 individual could therefore represent between-host evolution~~,—it is~~ not necessarily within-host **evolution** in a single individual.

Commented [LK22]: Correct?

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 ~~characterised~~ **characterized** the variant's genotype and evolutionary history, ~~and~~ **as well as its** association with high viral loads, rapid decline of CD4 cells, and increased transmissibility. We found 107 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 observed association is causal: ~~that t~~**T**he 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 branch **that** connect~~s~~**ing** this clade to other known viruses.

Commented [LK23]: Should this be changed to 102 (the number mentioned elsewhere in the paper)?

Deferring the initiation of treatment until the development of a CD4 count ~~of~~ ≤350 cells/**mm**^{3} ~~per mm^3~~ (or **the onset of** AIDS)~~,~~ instead of immediately at a CD4 count ~~of~~ ≥500 cells/**mm**^{3} ~~per mm^3 or more,~~ 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**^{3} ~~per mm^3~~ as "~~'~~advanced HIV~~'~~**"** (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**~~to~~ 36 months (CI: 33 to 39) for ~~not~~**non**-VB individuals, in males diagnosed **at the** age~~d~~ **of** 30 **to**~~-~~ 39 years. ~~It~~**Advanced HIV** is reached even more quickly in older age groups~~,~~**;** ~~furthermore,~~ **and** there is considerable variation between individuals **in addition to**~~on top of~~ these expected values. Many individuals could ~~therefore progress to~~ **have** 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, ~~where~~ **in which** diagnosis **often** occurs later in infection, the probability of reaching advanced HIV before diagnosis would be even greater.

Commented [LK24]: Correct?

Commented [LK25]: As meant? Is this synonymous with having a CD4 count at or below 350 cells/mm^3?

Commented [LK26]: As meant? If not, please specify what "on top of" refers to in this context.

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: ~~t~~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 ~~viral load~~ 3.0 $\log_{10}$ **elevation in viral load**~~copies higher~~, though the actual increase is ~~by~~ **a** 0.54 **to**~~–~~ 0.74 $\log_{10}$ **increase**~~copies~~. This means **that** the virulence ~~normalised~~ **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 ~~meta-analysis~~ (*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** ~~when~~ diagnosis ~~occurs~~, 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. ~~With too low~~ **If** infectiousness **is too low**, the pathogen cannot be transmitted when its host contacts other hosts, but ~~with too high~~ **if** virulence **is too high**, ~~its~~ **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 finding**s** may stimulate further interest in whether widespread treatment shifts the balance of the infectious~~–~~~~—~~virulence trade-off toward higher virulence, **thus** promoting the emergence and spread of new virulent variants. Previous model~~l~~ing studies have investigated this **idea** for pathogens generally (*48*)~~,~~ and **for** HIV specifically (*49*, *50*). We discuss some subtleties of the argument in **the** ~~S~~supplementary ~~T~~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~~,~~ ~~not only~~ **in addition to** their relative fitness, and treatment reduces the total onward transmission over the course of one infection**,** regardless **of virulence**~~how virulent it is~~. Put simply, "~~'~~viruses cannot mutate if they cannot replicate~~'~~**"** (anon**ymous**), **and**~~or~~ "~~'~~the best way to stop it changing is to stop it~~'~~**"** (Marc Lipsitch). 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** ~~underlines~~ 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/).

Commented [LK27]: Revised to clarify meaning and reduce jargon. If the edited text does not preserve your intended meaning, please revise further.

Commented [LK28]: Citations are needed for these quotes. Alternatively, please revise so that material is not directly quoted.

References and Notes

1. J. L. Geoghegan, E. C. Holmes, The phylogenomics of evolving virus virulence. *Nat. Rev. Genet.* **19**, 756–769 (2018). doi:10.1038/s41576-018-0055-5 Medline

2. C. E. Cressler, D. V. McLEOD, C. Rozins, J. VAN DEN Hoogen, T. Day, The adaptive evolution of virulence: A review of theoretical predictions and empirical tests. *Parasitology* **143**, 915–930 (2016). doi:10.1017/S003118201500092X Medline

3. J. J. Bull, A. S. Lauring, Theory and empiricism in virulence evolution. *PLOS Pathog.* **10**, e1004387 (2014). doi:10.1371/journal.ppat.1004387 Medline

4. R. Challen, E. Brooks-Pollock, J. M. Read, L. Dyson, K. Tsaneva-Atanasova, L. Danon, Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *BMJ* **372**, n579 (2021). doi:10.1136/bmj.n579 Medline

5. N. G. Davies, C. I. Jarvis, W. J. Edmunds, N. P. Jewell, K. Diaz-Ordaz, R. H. Keogh; CMMID COVID-19 Working Group, Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021). doi:10.1038/s41586-021-03426-1 Medline

6. D. J. Grint, K. Wing, E. Williamson, H. I. McDonald, K. Bhaskaran, D. Evans, S. J. W. Evans, A. J. Walker, G. Hickman, E. Nightingale, A. Schultze, C. T. Rentsch, C. Bates, J. Cockburn, H. J. Curtis, C. E. Morton, S. Bacon, S. Davy, A. Y. S. Wong, A. Mehrkar, L. Tomlinson, I. J. Douglas, R. Mathur, P. Blomquist, B. MacKenna, P. Ingelsby, R. Croker, J. Parry, F. Hester, S. Harper, N. J. DeVito, W. Hulme, J. Tazare, B. Goldacre, L. Smeeth, R. M. Eggo, Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England, 16 November to 5 February. *Euro Surveill.* **26**, (2021). doi:10.2807/1560-7917.ES.2021.26.11.2100256 Medline

7. N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, W. J. Edmunds; CMMID COVID-19 Working Group; COVID-19 Genomics UK (COG-UK) Consortium, Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021). doi:10.1126/science.abg3055 Medline

8. E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson; COVID-19 Genomics UK (COG-UK) consortium, Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021). doi:10.1038/s41586-021-03470-x Medline

9. C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, W. P. Hanage, Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17441–17446 (2007). doi:10.1073/pnas.0708559104 Medline

10. P. M. Sharp, B. H. Hahn, Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **1**, a006841 (2011). doi:10.1101/cshperspect.a006841 Medline

11. M. Worobey, M. Gemmel, D. E. Teuwen, T. Haselkorn, K. Kunstman, M. Bunce, J.-J. Muyembe, J.-M. M. Kabongo, R. M. Kalengayi, E. Van Marck, M. T. P. Gilbert, S. M. Wolinsky, Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* **455**, 661–664 (2008). doi:10.1038/nature07390 Medline

12. J. Hemelaar, R. Elangovan, J. Yun, L. Dickson-Tetteh, I. Fleminger, S. Kirtley, B. Williams, E. Gouws-Williams, P. D. Ghys; WHO–UNAIDS Network for HIV Isolation Characterisation, Global and regional molecular epidemiology of HIV-1, 1990-2015: A systematic review, global survey, and trend analysis. *Lancet Infect. Dis.* **19**, 143–155 (2019). doi:10.1016/S1473-3099(18)30647-9 Medline

13. B. S. Taylor, M. E. Sobieszczyk, F. E. McCutchan, S. M. Hammer, The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **358**, 1590–1602 (2008). doi:10.1056/NEJMra0706737 Medline

14. B. Asjö, L. Morfeldt-Månson, J. Albert, G. Biberfeld, A. Karlsson, K. Lidman, E. M. Fenyö, Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* **2**, 660–662 (1986). doi:10.1016/S0140-6736(86)90169-8 Medline

15. M. Koot, I. P. Keet, A. H. Vos, R. E. de Goede, M. T. Roos, R. A. Coutinho, F. Miedema, P. T. Schellekens, M. Tersmette, Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann. Intern. Med.* **118**, 681–688 (1993). doi:10.7326/0003-4819-118-9-199305010-00004 Medline

16. H. Song, W. Ou, Y. Feng, J. Zhang, F. Li, J. Hu, H. Peng, H. Xing, L. Ma, Q. Tan, D. Li, L. Wang, B. Wu, Y. Shao, Disparate impact on CD4 T cell count by two distinct HIV-1 phylogenetic clusters from the same clade. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 239–244 (2019). doi:10.1073/pnas.1814714116 Medline

17. C. Fraser, K. Lythgoe, G. E. Leventhal, G. Shirreff, T. D. Hollingsworth, S. Alizon, S. Bonhoeffer, Virulence and pathogenesis of HIV-1 infection: An evolutionary perspective. *Science* **343**, 1243727 (2014). doi:10.1126/science.1243727 Medline

18. F. Blanquart, C. Wymant, M. Cornelissen, A. Gall, M. Bakker, D. Bezemer, M. Hall, M. Hillebregt, S. H. Ong, J. Albert, N. Bannert, J. Fellay, K. Fransen, A. J. Gourlay, M. K. Grabowski, B. Gunsenheimer-Bartmeyer, H. F. Günthard, P. Kivelä, R. Kouyos, O. Laeyendecker, K. Liitsola, L. Meyer, K. Porter, M. Ristola, A. van Sighem, G. Vanham, B. Berkhout, P. Kellam, P. Reiss, C. Fraser; BEEHIVE Collaboration, Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLOS Biol.* **15**, e2001855 (2017). doi:10.1371/journal.pbio.2001855 Medline

19. F. Bertels, A. Marzel, G. Leventhal, V. Mitov, J. Fellay, H. F. Günthard, J. Böni, S. Yerly, T. Klimkait, V. Aubert, M. Battegay, A. Rauch, M. Cavassini, A. Calmy, E. Bernasconi, P. Schmid, A. U. Scherrer, V. Müller, S. Bonhoeffer, R. Kouyos, R. R. Regoes; Swiss HIV Cohort Study, Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T-Cell Decline, and Per-Parasite Pathogenicity. *Mol. Biol. Evol.* **35**, 27–37 (2018). doi:10.1093/molbev/msx246 Medline

20. C. Fraser, T. D. Hollingsworth, Interpretation of correlations in setpoint viral load in transmitting couples. *AIDS* **24**, 2596–2597 (2010). doi:10.1097/QAD.0b013e32833e7a64 Medline

21. V. Mitov, T. Stadler, A Practical Guide to Estimating the Heritability of Pathogen Traits. *Mol. Biol. Evol.* **35**, 756–772 (2018). doi:10.1093/molbev/msx328 Medline

22. F. Schoeni-Affolter, B. Ledergerber, M. Rickenbach, C. Rudin, H. F. Günthard, A. Telenti, H. Furrer, S. Yerly, P. Francioli, Cohort profile: The Swiss HIV Cohort study. *Int. J. Epidemiol.* **39**, 1179–1189 (2010). doi:10.1093/ije/dyp321 Medline

23. K. Kusejko, N. Bachmann, S. E. Chaudron, H. Nguyen, D. L. Braun, B. Hampel, M. Battegay, E. Bernasconi, A. Calmy, M. Cavassini, M. Hoffmann, J. Böni, S. Yerly, T. Klimkait, M. Perreau, A. Rauch, H. F. Günthard, R. D. Kouyos; Swiss HIV Cohort Study, A Systematic Phylogenetic Approach to Study the Interaction of HIV-1 With Coinfections, Noncommunicable Diseases, and Opportunistic Diseases. *J. Infect. Dis.* **220**, 244–253 (2019). doi:10.1093/infdis/jiz093 Medline

24. A. U. Scherrer, A. Traytel, D. L. Braun, A. Calmy, M. Battegay, M. Cavassini, H. Furrer, P. Schmid, E. Bernasconi, M. Stoeckle, C. Kahlert, A. Trkola, R. D. Kouyos, P. Tarr, C. Marzolini, G. Wandeler, J. Fellay, H. Bucher, S. Yerly, F. Suter, H. Hirsch, M. Huber, G. Dollenmaier, M. Perreau, G. Martinetti, A. Rauch, H. F. Günthard; Swiss HIV Cohort Study (SHCS), Cohort Profile Update: The Swiss HIV Cohort Study (SHCS). *Int. J. Epidemiol.* 10.1093/ije/dyab141 (2021). doi:10.1093/ije/dyab141 Medline

25. J. D. Lundgren, A. G. Babiker, F. Gordin, S. Emery, B. Grund, S. Sharma, A. Avihingsanon, D. A. Cooper, G. Fätkenheuer, J. M. Llibre, J. M. Molina, P. Munderi, M. Schechter, R. Wood, K. L. Klingman, S. Collins, H. C. Lane, A. N. Phillips, J. D. Neaton; INSIGHT START Study Group, Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N. Engl. J. Med.* **373**, 795–807 (2015). doi:10.1056/NEJMoa1506816 Medline

26. S. Lodi, A. Phillips, G. Touloumi, R. Geskus, L. Meyer, R. Thiébaut, N. Pantazis, J. D. Amo, A. M. Johnson, A. Babiker, K. Porter; CASCADE Collaboration in EuroCoord, Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 cells/mm^3: Assessment of need following changes in treatment guidelines. *Clin. Infect. Dis.* **53**, 817–825 (2011). doi:10.1093/cid/cir494 Medline

27. A. van Sighem, F. Nakagawa, D. De Angelis, C. Quinten, D. Bezemer, E. O. de Coul, M. Egger, F. de Wolf, C. Fraser, A. Phillips, Estimating HIV Incidence, Time to Diagnosis, and the Undiagnosed HIV Epidemic Using Routine Surveillance Data. *Epidemiology* **26**, 653–660 (2015). doi:10.1097/EDE.0000000000000324 Medline

28. A.-C. Pineda-Peña, N. R. Faria, S. Imbrechts, P. Libin, A. B. Abecasis, K. Deforche, A. Gómez-López, R. J. Camacho, T. de Oliveira, A.-M. Vandamme, Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* **19**, 337–348 (2013). doi:10.1016/j.meegid.2013.04.032 Medline

29. D. Struck, G. Lawyer, A.-M. Ternes, J.-C. Schmit, D. P. Bercoff, COMET: Adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res.* **42**, e144 (2014). doi:10.1093/nar/gku739 Medline

30. M. A. Jensen, F.-S. Li, A. B. van ’t Wout, D. C. Nickle, D. Shriner, H.-X. He, S. McLaughlin, R. Shankarappa, J. B. Margolick, J. I. Mullins, Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J. Virol.* **77**, 13376–13388 (2003). doi:10.1128/JVI.77.24.13376-13388.2003 Medline

31. M. C. Prosperi, I. Fanti, G. Ulivi, A. Micarelli, A. De Luca, M. Zazzi, Robust supervised and unsupervised statistical learning for HIV type 1 coreceptor usage analysis. *AIDS Res. Hum. Retroviruses* **25**, 305–314 (2009). doi:10.1089/aid.2008.0039 Medline

32. P. Kellam, C. A. Boucher, B. A. Larder, Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1934–1938 (1992). doi:10.1073/pnas.89.5.1934 Medline

33. S.-Y. Rhee, M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, R. W. Shafer, Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **31**, 298–303 (2003). doi:10.1093/nar/gkg100 Medline

34. J. M. Carlson, C. J. Brumme, E. Martin, J. Listgarten, M. A. Brockman, A. Q. Le, C. K. S. Chui, L. A. Cotton, D. J. H. F. Knapp, S. A. Riddler, R. Haubrich, G. Nelson, N. Pfeifer, C. E. Deziel, D. Heckerman, R. Apps, M. Carrington, S. Mallal, P. R. Harrigan, M. John, Z. L. Brumme; International HIV Adaptation Collaborative, Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. *J. Virol.* **86**, 13202–13216 (2012). doi:10.1128/JVI.01998-12 Medline

35. J. J. Lum, O. J. Cohen, Z. Nie, J. G. Weaver, T. S. Gomez, X.-J. Yao, D. Lynch, A. A. Pilon, N. Hawley, J. E. Kim, Z. Chen, M. Montpetit, J. Sanchez-Dardon, E. A. Cohen, A. D. Badley, Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis. *J. Clin. Invest.* **111**, 1547–1554 (2003). doi:10.1172/JCI16233 Medline

36. M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, A. Rambaut, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018). doi:10.1093/ve/vey016 Medline

37. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017). doi:10.1111/2041-210X.12628

38. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013). doi:10.1093/molbev/mss265 Medline

39. E. de Silva, N. M. Ferguson, C. Fraser, Inferring pandemic growth rates from sequence data. *J. R. Soc. Interface* **9**, 1797–1808 (2012). doi:10.1098/rsif.2011.0850 Medline

40. D. Bezemer ~~*et al.*~~, **A. Blenkinsop, M. Hall, A. van Sighem, M. Cornelissen, E. Wessels, J. van Kampen, T. van de Laar, P. Reiss, C. Fraser, O. Ratmann,** ~~In-country acquisition of HIV-1 non-B infection within the Netherlands is frequent, but results in limited onward transmission~~**Many but small HIV-1 non-B transmission chains in the**

Commented [LK29]: Please check the edited reference for accuracy.

Netherlands. *AIDS* **36, 83–94** (~~2021~~**2022**). **doi:10.1097/QAD.0000000000003074 Medline**

41. R. A. Neher, C. A. Russell, B. I. Shraiman, Predicting evolution from the shape of genealogical trees. *eLife* **3**, e03568 (2014). doi:10.7554/eLife.03568 Medline

42. L. Råberg, M. Stjernman, in ~~*Ecological immunology*~~***Ecoimmunology***, G. D**emas**~~.~~, ~~A.~~ R. Nelson, Ed**s**. (**Oxford Univ. Press**~~OUP~~**, 2011**), pp. 548–578.

43. L. Råberg, How to live with the enemy: Understanding tolerance to parasites. *PLOS Biol.* **12**, e1001989 (2014). doi:10.1371/journal.pbio.1001989 Medline

44. N. Pantazis, K. Porter, D. Costagliola, A. De Luca, J. Ghosn, M. Guiguet, A. M. Johnson, A. D. Kelleher, C. Morrison, R. Thiebaut, L. Wittkop, G. Touloumi; CASCADE Collaboration in EuroCoord, Temporal trends in prognostic markers of HIV-1 virulence and transmissibility: An observational cohort study. *Lancet HIV* **1**, e119–e126 (2014). doi:10.1016/S2352-3018(14)00002-2 Medline

45. J. O. Wertheim, A. M. Oster, W. M. Switzer, C. Zhang, N. Panneer, E. Campbell, N. Saduvala, J. A. Johnson, W. Heneine, Natural selection favoring more transmissible HIV detected in United States molecular transmission network. *Nat. Commun.* **10**, 5788 (2019). doi:10.1038/s41467-019-13723-z Medline

46. F. Blanquart, M. K. Grabowski, J. Herbeck, F. Nalugoda, D. Serwadda, M. A. Eller, M. L. Robb, R. Gray, G. Kigozi, O. Laeyendecker, K. A. Lythgoe, G. Nakigozi, T. C. Quinn, S. J. Reynolds, M. J. Wawer, C. Fraser, A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *eLife* **5**, e20492 (2016). doi:10.7554/eLife.20492 Medline

47. J. T. Herbeck, V. Müller, B. S. Maust, B. Ledergerber, C. Torti, S. Di Giambenedetto, L. Gras, H. F. Günthard, L. P. Jacobson, J. I. Mullins, G. S. Gottlieb, Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS* **26**, 193–205 (2012). doi:10.1097/QAD.0b013e32834db418 Medline

48. T. C. Porco, J. O. Lloyd-Smith, K. L. Gross, A. P. Galvani, The effect of treatment on pathogen virulence. *J. Theor. Biol.* **233**, 91–102 (2005). doi:10.1016/j.jtbi.2004.09.009 Medline

49. J. T. Herbeck, J. E. Mittler, G. S. Gottlieb, S. M. Goodreau, J. T. Murphy, A. Cori, M. Pickles, C. Fraser, Evolution of HIV virulence in response to widespread scale up of antiretroviral therapy: A modeling study. *Virus Evol.* **2**, vew028 (2016). doi:10.1093/ve/vew028 Medline

50. H. E. Roberts, P. J. R. Goulder, A. R. McLean, The impact of antiretroviral therapy on population-level virulence evolution of HIV-1. *J. R. Soc. Interface* **12**, 20150888 (2015). doi:10.1098/rsif.2015.0888 Medline

51. **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**.~~The code used for our analyses is available at doi:10.5281/zenodo.5761935~~.

Commented [LK30]: Please check the edited reference for accuracy.

Commented [LK31]: Please check the edited reference for accuracy.

52. T. S. Boender, C. Smit, A. V. Sighem, D. Bezemer, C. J. Ester, S. Zaheri, F. W. N. M. Wit, P. Reiss; ATHENA national observational HIV cohort, AIDS Therapy Evaluation in the Netherlands (ATHENA) national observational HIV cohort: Cohort profile. *BMJ Open* **8**, e022516 (2018). doi:10.1136/bmjopen-2018-022516 Medline

53. D. Bezemer, F. de Wolf, M. C. Boerlijst, A. van Sighem, T. D. Hollingsworth, C. Fraser, 27 years of the HIV epidemic amongst men having sex with men in the Netherlands: An in depth mathematical model-based analysis. *Epidemics* **2**, 66–79 (2010). doi:10.1016/j.epidem.2010.04.001 Medline

54. D. Bezemer, S. Jurriaans, M. Prins, L. van der Hoek, J. M. Prins, F. de Wolf, B. Berkhout, R. Coutinho, N. K. T. Back, Declining trend in transmission of drug-resistant HIV-1 in Amsterdam. *AIDS* **18**, 1571–1577 (2004). doi:10.1097/01.aids.0000131357.52457.33 Medline

55. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011). doi:10.1038/msb.2011.75 Medline

56. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019). doi:10.1093/bioinformatics/bty633 Medline

57. M. Cornelissen, A. Gall, M. Vink, F. Zorgdrager, Š. Binter, S. Edwards, S. Jurriaans, M. Bakker, S. H. Ong, L. Gras, A. van Sighem, D. Bezemer, F. de Wolf, P. Reiss, P. Kellam, B. Berkhout, C. Fraser, A. C. van der Kuyl; BEEHIVE Consortium, From clinical sample to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing. *Virus Res.* **239**, 10–16 (2017). doi:10.1016/j.virusres.2016.08.004 Medline

58. A. Gall, B. Ferns, C. Morris, S. Watson, M. Cotten, M. Robinson, N. Berry, D. Pillay, P. Kellam, Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* **50**, 3838–3844 (2012). doi:10.1128/JCM.01516-12 Medline

59. M. Hunt, A. Gall, S. H. Ong, J. Brener, B. Ferns, P. Goulder, E. Nastouli, J. A. Keane, P. Kellam, T. D. Otto, IVA: Accurate de novo assembly of RNA virus genomes. *Bioinformatics* **31**, 2374–2376 (2015). doi:10.1093/bioinformatics/btv120 Medline

60. C. Wymant, F. Blanquart, T. Golubchik, A. Gall, M. Bakker, D. Bezemer, N. J. Croucher, M. Hall, M. Hillebregt, S. H. Ong, O. Ratmann, J. Albert, N. Bannert, J. Fellay, K. Fransen, A. Gourlay, M. K. Grabowski, B. Gunsenheimer-Bartmeyer, H. F. Günthard, P. Kivelä, R. Kouyos, O. Laeyendecker, K. Liitsola, L. Meyer, K. Porter, M. Ristola, A. van Sighem, B. Berkhout, M. Cornelissen, P. Kellam, P. Reiss, C. Fraser; BEEHIVE Collaboration, Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus Evol.* **4**, vey007 (2018). doi:10.1093/ve/vey007 Medline

61. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi:10.1093/bioinformatics/btu033 Medline

62. A. M. Wensing, V. Calvez, H. F. Günthard, V. A. Johnson, R. Paredes, D. Pillay, R. W. Shafer, D. D. Richman, 2015 Update of the Drug Resistance Mutations in HIV-1. *Top. Antivir. Med.* **23**, 132–141 (2015). Medline

63. C. Wymant, M. Hall, O. Ratmann, D. Bonsall, T. Golubchik, M. de Cesare, A. Gall, M. Cornelissen, C. Fraser; STOP-HCV Consortium, The Maela Pneumococcal Collaboration, The BEEHIVE Collaboration, PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Mol. Biol. Evol.* **35**, 719–733 (2018). doi:10.1093/molbev/msx304 Medline

64. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015). doi:10.18637/jss.v067.i01

65. L. Harrison, D. T. Dunn, H. Green, A. J. Copas, Modelling the association between patient characteristics and the change over time in a disease measure using observational cohort data. *Stat. Med.* **28**, 3260–3275 (2009). doi:10.1002/sim.3725 Medline

66. J. C. Ho, G. T. Ng, M. Renaud, A. F. Y. Poon, sierra-local: A lightweight standalone application for ~~secure HIV-1~~ drug resistance prediction. ***J. Open Source Softw.* 4, 1186 (2019)**~~bioRxiv 393207 (2018)~~**. doi:10.21105/joss.01186**

67. H. Song, E. E. Giorgi, V. V. Ganusov, F. Cai, G. Athreya, H. Yoon, O. Carja, B. Hora, P. Hraber, E. Romero-Severson, C. Jiang, X. Li, S. Wang, H. Li, J. F. Salazar-Gonzalez, M. G. Salazar, N. Goonetilleke, B. F. Keele, D. C. Montefiori, M. S. Cohen, G. M. Shaw, B. H. Hahn, A. J. McMichael, B. F. Haynes, B. Korber, T. Bhattacharya, F. Gao, Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nat. Commun.* **9**, 1928 (2018). doi:10.1038/s41467-018-04217-5 Medline

68. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi:10.1016/S0022-2836(05)80360-2 Medline

69. B. Korber, in *Computational and Evolutionary Analysis of HIV Molecular Sequences*, A. G. Rodrigo, G. H. Learn, Eds. (Kluwer Academic Publishers, 2000), pp. 55–72.

70. M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986). Medline

71. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline

72. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLOS Biol.* **4**, e88 (2006). doi:10.1371/journal.pbio.0040088 Medline

73. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018). doi:10.1093/sysbio/syy032 Medline

74. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018). doi:10.1093/ve/vex042 Medline

75. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi:10.1093/molbev/msu300 Medline

Commented [LK32]: This reference has been updated to reflect that this paper is now published in a journal - okay?

76. E. M. Volz, S. D. W. Frost, Scalable relaxed clock phylogenetic dating. *Virus Evol.* **3**, vex025 (2017). doi:10.1093/ve/vex025

77. B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly, J. A. Mumford, E. C. Holmes, Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332 (2004). doi:10.1126/science.1090727 Medline

78. B. L. Dearlove, S. D. W. Frost, Measuring Asymmetry in Time-Stamped Phylogenies. *PLOS Comput. Biol.* **11**, e1004312 (2015). doi:10.1371/journal.pcbi.1004312 Medline

79. S. Alizon, C. Fraser, Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology* **10**, 49 (2013). doi:10.1186/1742-4690-10-49 Medline

80. D. Bonsall, T. Golubchik, M. de Cesare, M. Limbada, B. Kosloff, G. MacIntyre-Cockett, M. Hall, C. Wymant, M. A. Ansari, L. Abeler-Dörner, A. Schaap, A. Brown, E. Barnes, E. Piwowar-Manning, S. Eshleman, E. Wilson, L. Emel, R. Hayes, S. Fidler, H. Ayles, R. Bowden, C. Fraser; HPTN 071 (PopART) Team, A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *J. Clin. Microbiol.* **58**, e00382-20 (2020). doi:10.1128/JCM.00382-20 Medline

81. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010). doi:10.1371/journal.pone.0009490 Medline

82. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017). doi:10.1038/nmeth.4285 Medline

83. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016). doi:10.1093/ve/vew007 Medline

84. M. Hernán, J. Robins, *Causal Inference: What If* (~~Boca Raton:~~ Chapman & Hall/CRC **Press**, 2020).

Acknowledgments: We thank ~~Katrien~~ **K.** Fransen and ~~Guido~~ **G.** Vanham for help with the Belgian data, ~~Oliver~~ **O.** Ratmann for help in identifying the Dutch clusters, ~~Katharina~~ **K.** Kusejko for testing for additional VB individuals in the SHCS, ~~Brian~~ **B.** Foley for help with genome sharing, ~~Bethany~~ **B.** Dearlove and ~~Laura~~ **L.** Thomson for help with software, and ~~Joshua~~ **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 ~~c~~**C**ohort 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. ~~Conceptualisation~~**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)**; a** version **has also been** deposited at **Zenodo** (*51*). 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, pre-treatment 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 **C**ohort, for evaluation of **the** 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 **C**ollaboration and ATHENA **C**ohort (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 **C**ohort 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 individuals enrolled in BEEHIVE for whole-genome sequencing.

Commented [LK33]: As meant? We avoid the "slash" construction because it is often ambiguous, although "and/or" is acceptable.

Commented [LK34]: As meant?

Commented [LK35]: Okay to change "ATHENA cohort" to "ATHENA study" here, given that this refers to the group for patients (and, by extension, the study itself) rather than the formal research collaboration?

Supplementary Materials

science.org/doi/10.1126/science.abk1688

Materials and Methods

Supplementary Text

Figs. S1 to S12

Tables S1 to S7

References (*52*–*84*)

MDAR Reproducibility Checklist

Data S1 to S4

25 June 2021; accepted 4 January 2022

Fig. 1. Clinical characteristics of "VB" individuals. Those infected with the highly virulent variant, (VB individuals) are represented shown in red, ; and "not-VB" individuals, those infected with any other subtype-B virus, (non-VB individuals) are shown in blue. (**A**) a bBox-and-whisker plots of viral load, by year of diagnosis. Diagnosis dates were grouped to give produce boundaries that coincideing with years and roughly equal numbers of VB individuals (39, 35, and 27 in the second, third, and fourth groups, respectively; the pattern is robust to other groupings). (**B**) the eExpected decline in CD4 count in the absence of treatment. The model was adjusted for Ssex and age at diagnosis have been adjusted for; values shown are for males diagnosed at the aged of 30 to 39 years. The sShaded ing regions indicates 95% CIsconfidence intervals in the model's prediction of the 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 horizontal dashed black line denotesshows a CD4 count of 350 cells/mm^3 per mm^3, (discussed in the see text for details). (**C**) The pProbability of still being alive at a given time after diagnosis.

Fig. 2. Phylogenetic and phylodynamic analysis of the VB variant. (**A**) A wWhole-genome maximum-likelihood phylogeny of 15 VB- variant sequences and 100 background subtype-B sequences. The color of each circle at the tips indicates the individual's viral loads in $\log_{10}$ copies per milliliterml. The inset scale bar shows the branch length scale in units of substitutions per site. (**B**) A dDated maximum-clade-credibility tree for 107 partial *pol* gene sequences from the VB variant. Colors indicates geographical regions the region of the Netherlands (N, E, S, and W: abbreviating north, east, south, and west), which areis known for the branch tips, and 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**) The eEffective population size (N_e) (scaled by the coalescent generation time τ) over time with 95% credibility intervals, with the same time axis as in (B)the panel above.

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, for so that the test to 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.

Commented [LK36]: Do "the second, third, and fourth groups" refer to "the 2002–2006, 2007–2008, and after-2008 groups"? It would be helpful to specify this for clarity.

Commented [LK37]: As meant?

Commented [LK38]: Should this be "102" instead?

Commented [LK39]: Revised because other countries are represented with specific colors as well (not just the Netherlands).

Commented [LK40]: Changes have been made for consistency with journal style and format for tables.

~~†The number of individuals~~ *N*, ~~indicated is~~ **number of individuals** after ~~excluding~~ 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]**
~~Dataset~~	~~BEEHIVE (Europe)~~	~~ATHENA (Netherlands), excluding overlap with BEEHIVE*~~
Viral load measurements compared~~†~~	Set-point viral loads for *N* = 15 VB individuals and *N* = 2446 individuals with any other HIV-1 strain	Mean pre-treatment log viral loads for *N* = 91 VB individuals and *N* = 5272 individuals with any other subtype-B HIV-1 strain
Mean and ~~inter-quartile range (~~IQR~~)~~ of viral load in ~~not~~**non-**VB individuals, in $\log_{10}$ copies per **milliliter**~~ml~~	5.10 (IQR**:** 4.69 **to**~~–~~ 5.58)	4.79 (IQR**:** 4.34 **to**~~–~~ 5.27)
Mean and ~~inter-quartile range (~~IQR~~)~~ of viral load in VB individuals, in $\log_{10}$ copies per **milliliter**~~ml~~	5.84 (IQR**:** 5.57 **to**~~–~~ 6.09)	5.33 (IQR**:** 4.94 **to**~~–~~ 5.75)
Viral load increase in VB individuals	0.74 $\log_{10}$ copies**/ml** ~~per ml~~	0.54 $\log_{10}$ copies**/ml** ~~per ml~~
***P*~~p~~** value for increase	5×10^{-6} (two-tailed *t* test, significant at a level of 5×10^{-5} when Bonferroni-corrected for performing 50 such tests)	1×10^{-12} (one-tailed *t* test)