



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

Chris Wymant, Daniela Bezemer, François Blanquart, Luca Ferretti, Astrid Gall, Matthew Hall, Tanya Golubchik, Margreet Bakker, Swee Hoe Ong, Lele Zhao, et al.

### ► To cite this version:

Chris Wymant, Daniela Bezemer, François Blanquart, Luca Ferretti, Astrid Gall, et al.. A highly virulent variant of HIV-1 circulating in the Netherlands. *Science*, 2022, 375 (6580), pp.540-545. 10.1126/science.abk1688 . hal-03835688

HAL Id: hal-03835688

<https://hal.science/hal-03835688>

Submitted on 18 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

**One sentence summary:** A new variant of subtype-B HIV-1 has evolved with viral loads 0.54 – 0.74 log<sub>10</sub> copies higher, CD4 cell decline twice as fast, and increased transmissibility.

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>, Mariska Hillebregt<sup>2</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 Gunsenheimer-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,\*</sup>, on behalf of the Netherlands ATHENA HIV Observational Cohort<sup>†</sup> and the BEEHIVE Collaboration<sup>†</sup>

\*corresponding authors: chris.wymant@bdi.ox.ac.uk and christophe.fraser@bdi.ox.ac.uk

† Full list of contributors in Supplementary Text.

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
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, 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
22. Institute for Global Health, University College London, London, UK
23. Kymab Ltd, Cambridge, UK

24. Department of Infectious Diseases, 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

**Abstract.** We discovered a highly virulent variant of subtype-B HIV-1 in the Netherlands. 102 individuals with this variant had viral loads  $0.54 - 0.74 \log_{10}$  copies higher, and CD4 cell decline 2 times faster, than 6,604 individuals with other subtype-B strains. Without treatment, advanced HIV – CD4 cell counts below 350 cells per mm<sup>3</sup>, 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 age, sex, suspected mode of transmission and place of birth, suggesting the effect is attributable to the virus. Genetic sequence analysis suggested 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 – causing greater damage to their host – has been much studied theoretically, despite few population-level examples (1-3). The most notable recent example is the Delta variant of 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 adaptive potential. Greater virulence could benefit the virus if it is not outweighed by reduced opportunity for transmission. These antagonistic selection pressures may result in an intermediate level of virulence 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](http://www.unaids.org)).

The main (M) group of HIV-1, responsible for the global pandemic, first emerged in the Kinshasa 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 coreceptor used for cell entry has long been understood to affect virulence (14, 15), and this has been proposed as a mechanism underlying 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: 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 report a subtype B variant of HIV-1 with exceptionally high virulence, that has been circulating within the Netherlands during the last two decades.

**Discovery of the highly virulent variant.** Within an ongoing study (the BEEHIVE project, [www.beehive.ox.ac.uk](http://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 (6-24 months after a positive test obtained early in infection) were highly elevated (Table 1 first column). BEEHIVE is a study of individuals enrolled in eight cohorts across Europe and Uganda, who were selected to have well-characterized date 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 from the ATHENA cohort in the Netherlands, one from Switzerland, and one from Belgium. See 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 analysed 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 the ATHENA data (Table 1 second column), we again observed a large increase in viral load in individuals with this viral variant:  $0.54 \log_{10}$  copies per ml. 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 (Figure 1a). Henceforth, for brevity, we refer to this viral variant as the VB variant (for Virulent subtype B), to individuals infected by this variant as VB individuals, and to individuals infected with a different strain of HIV as not-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 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). Examining previously published data(23), 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 2009-2018, with roughly 65% of the cohort sequenced(24)) makes it unlikely that there were many more VB individuals in Switzerland who were missed. Data to assess viral load or CD4 cell decline for these three individuals was not available due to early treatment initiation.

**More rapid CD4 cell decline.** CD4 counts for VB individuals were already lower at time of diagnosis, by 73 cells per mm<sup>3</sup> (confidence interval (CI) 12 to 134). They subsequently declined faster by an additional 49 cells per mm<sup>3</sup> per year (CI 20 to 79) on top of the decline for comparable not-VB individuals, which is 49 cells per mm<sup>3</sup> per year (CI 46 to 51) for men diagnosed aged 30-39 years. The VB variant is therefore associated with a doubling in the rate of CD4 cell decline. These values are averages estimated using a linear mixed model adjusting 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 per mm<sup>3</sup>, instead of immediately, was previously shown to substantially increase the subsequent hazard for serious adverse events (25). As seen in Figure 1b this stage of CD4 cell decline is reached in 9 months (CI 2 to 17) from time of diagnosis for VB individuals, compared to 36 months (CI 33 to 39) for not-VB individuals, in males diagnosed aged 30-39 years. It is reached even more quickly in older age groups, whom we found to have progressively

lower CD4 counts at time of diagnosis (Supplementary Table S1). At a CD4 count of 200 cells per mm<sup>3</sup>, there is a high risk of immediate AIDS-related complications; this stage of decline would be reached on average between two and three years after diagnosis for VB individuals, and between six and seven years after diagnosis for comparable not-VB individuals (the latter being similar to previous reports in Europe (26)).

The effect of the VB variant on CD4 cell decline remained after adjusting 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 comparable non-VB individuals with high viral loads: their rate of decline is 44 cells per mm<sup>3</sup> per year greater (CI 16 to 72). Comparing this additional decline with that expected from a +1 increase in log<sub>10</sub> viral load, 15 cells per mm<sup>3</sup> per year (CI 11 to 18), shows that the variant's effect on CD4 count decline was equivalent to that expected from having a viral load +3.0 log<sub>10</sub> copies higher. The same analysis of measurements of CD4 percentages (the percent 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 adjusting for the higher viral load of the variant (Supplementary Table S2, Supplementary Figure S1).

**No difference in CD4 cells after treatment, or in mortality.** Measurements of the success of treatment include CD4 cell recovery, and mortality. CD4 counts and percentages after treatment initiation were similar for VB and not-VB individuals, as measured with both linear mixed modelling of the CD4 dynamics (Supplementary Table S3 and S4, and Supplementary Figure S2) and an individual-matching procedure agnostic of the dynamics. 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 power to detect only very large differences in mortality, as reflected in the wide CI for relative hazard for death and shown in Figure 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 S3). For example, though the probability of having started treatment was estimated to be similar at 6 months after diagnosis, 42% (CI 41 to 44%) for not-VB individuals compared to 46% (CI 35 to 54%) for VB individuals, at 2 years after diagnosis it was different: 65% (CI 64 to 67%) for not-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 could be relevant should VB or variants like it be found in settings with less widespread availability of AIDS care.

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

**Genotype of the VB variant.** Sequence data from the BEEHIVE project is whole genome, providing the 17 whole genomes available for the variant; sequence data from ATHENA is partial *pol*-gene only, available for the additional 92 VB individuals. We subtyped the 17 whole genomes for the variant as pure subtype B (with 100% support using two concordant methods (28, 29)), like most HIV-1 in the Netherlands. We predicted coreceptor usage from the 17 whole genomes using two concordant methods (30, 31): one was likely CXCR4-tropic, the other 16 likely CCR5-tropic. Only one drug resistance mutation was common for the VB variant: 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](http://www.hiv.lanl.gov)), none appeared to be a candidate for a ‘recombination parent’ of the VB variant, i.e. the many mutations distinguishing 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 Supplementary Data S1, and the amino acid alignment is illustrated in Supplementary Figure S5. The distribution of nucleotide changes over the genome is in line with

expectations (for example less in the conserved *pol* gene region and more in the variable *env* gene region; see Supplementary Figure S6). The VB variant genotype is thus characterised by many mutations spread through the genome, meaning a single genetic cause for the enhanced virulence cannot be determined from the current data.

We conducted descriptive analyses of the mutations distinguishing the VB variant from the Dutch subtype-B consensus. All the amino-acid-level changes are listed in Supplementary data S2 with annotations. 30 of the amino acid substitutions observed 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 similar size to the lineage (see 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 normalised by its overall level of divergence (Supplementary Figure S11). We also calculated the ratio of rates of non-synonymous and synonymous changes – dn/ds – for each gene, for the VB variant and the other 16 Dutch subtype B clades used for comparison. The VB variant had lower dn/ds values than all of the other clades in *env*, *pol* and *tat*, though its values were not extreme, and for the other genes its dn/ds value was in the range spanned by the other clades (Supplementary Figure S12). Finally, we noted that at codon position 77 of Vpr, the consensus of all Dutch subtype-B sequences in BEEHIVE is Glutamine, while the VB consensus is Arginine. Glutamine was previously found to be more common in long-term non-progressors, and mutation to Arginine increased T cell apoptosis in vitro and strongly increased T cell decline in mice models (35). However, both alleles have been commonly observed in subtype B to date

(of 2178 subtype-B Vpr protein sequences on the Los Alamos National Laboratory HIV Database, 52% have Glutamine, 36% have Arginine), making it implausible that this mutation alone is the dominant mechanism for the virulence effect observed here.

**Evolution of the VB variant.** The maximum-likelihood phylogeny in Figure 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 using *BEAST* (36) using partial *pol* sequences. It is coloured by region, inferred with an ancestral state reconstruction by parsimony (minimizing changes of region). This assigned Amsterdam 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 using *ggtree* (37).

**Phyldynamics 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 was estimated using a skygrid demographic model (38) in *BEAST*, and is shown in Figure 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 can be an artefact of  $N_e$  inference methods in the recent past (39)). The proportion of all new subtype-B diagnoses that are VB variant increased until a peak in 2008, and decreased after that, though again with appreciable uncertainty; absolute numbers of both VB and not-VB diagnoses in our dataset have been decreasing since roughly 2008

(Supplementary Figure S7). In a recent analysis of an updated version of the ATHENA dataset (40), 33 additional VB individuals were found; these suggest that VB diagnoses were stable until roughly 2013, and have been declining since, though with large uncertainty (the dataset is right-censored by several years; Supplementary Figure S7).

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

**The first sampled VB individual.** We retrieved and sequenced two additional samples from the VB individual who was diagnosed ten years before subsequent diagnoses of VB individuals, in 1992. Phylogenetic analysis suggested that this individual was infected with a virus that had evolved most of the way, but not all of the way, toward VB variant viruses typical of later dates (Supplementary Text; Supplementary Figure S10). This individual was diagnosed in Amsterdam, consistent with the ancestral reconstruction of region. Over the ten 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 two or three steps in the transmission chain could easily have been unsampled. The long phylogenetic branch leading to the 1992 individual could therefore represent between-host evolution – it is not necessarily within-host 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 characterised the variant’s genotype and evolutionary history, and 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 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 branch connecting this clade to other known viruses.

Deferring the initiation of treatment until the development of a CD4 count of 350 cells per mm<sup>3</sup> or AIDS, instead of immediately at a CD4 count of 500 cells per mm<sup>3</sup> 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 per mm<sup>3</sup> as

‘advanced HIV’ (<https://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 time of diagnosis for VB individuals, compared to 36 months (CI 33 to 39) for not-VB individuals, in males diagnosed aged 30-39 years. It is reached even more quickly in older age groups; furthermore, there is considerable variation between individuals on top of 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, where diagnosis 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 is equivalent to the acceleration of CD4 degradation that would be expected from a viral load  $3.0 \log_{10}$  copies higher, though the actual increase is by  $0.54 - 0.74 \log_{10}$  copies. This means the virulence normalised 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 methods we predicted that, of the 17 whole genomes available, 16 use only the R5 coreceptor for cell entry, which is typical for subtype-B viruses in early infection (13).

This suggests 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)) 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 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 favour intermediate values rather than extreme ones. With too low infectiousness, the pathogen cannot be transmitted when its host contacts other hosts, but with too high virulence, its 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 may stimulate further interest in whether widespread treatment shifts the balance of the infectiousness-virulence trade-off towards higher virulence, promoting the emergence and spread of new virulent variants. Previous modelling studies have investigated this for pathogens generally (48), and HIV specifically (49, 50). We discuss some subtleties of the argument in

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, not only their relative fitness, and treatment reduces the total onward transmission over the course of one infection regardless how virulent it is. Put simply, *viruses cannot mutate if they cannot replicate* (anon), 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 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/](http://www.who.int/hiv/pub/arv/)).

## References

1. J. L. Geoghegan, E. C. Holmes, The phylogenomics of evolving virus virulence. *Nature Reviews Genetics.* **19**, 756–769 (2018).
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).
3. J. J. Bull, A. S. Lauring, Theory and Empiricism in Virulence Evolution. *PLOS Pathogens.* **10**, 1–3 (2014).

4. R. Challen *et al.*, Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ*. **372** (2021), doi:10.1136/bmj.n579.
5. N. G. Davies *et al.*, Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* (2021).
6. D. J. Grint *et al.*, Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England, 16 November to 5 February. *Eurosurveillance*. **26** (2021), doi:<https://doi.org/10.2807/1560-7917.ES.2021.26.11.2100256>.
7. N. G. Davies *et al.*, Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* (2021), doi:10.1126/science.abg3055.
8. E. Volz *et al.*, Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. **593**, 266–269 (2021).
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. *Proceedings of the National Academy of Sciences*. **104**, 17441–17446 (2007).
10. P. M. Sharp, B. H. Hahn, Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine*. **1** (2011).
11. M. Worobey *et al.*, Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. **455**, 661–664 (2008).
12. J. Hemelaar *et al.*, Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *The Lancet Infectious Diseases*. **19**, 143–155 (2019).
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).
14. B. Åsjö, REPLICATIVE CAPACITY OF HUMAN IMMUNODEFICIENCY VIRUS FROM PATIENTS WITH VARYING SEVERITY OF HIV INFECTION. *The Lancet*. **328**, 660–662 (1986).
15. M. Koot *et al.*, 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).
16. H. Song *et al.*, Disparate impact on CD4 T cell count by two distinct HIV-1 phylogenetic clusters from the same clade. *Proc Natl Acad Sci USA*. **116**, 239 (2019).
17. C. Fraser *et al.*, Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. *Science*. **343** (2014), doi:10.1126/science.1243727.
18. F. Blanquart *et al.*, Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biol*. **15**, e2001855 (2017).

19. F. Bertels *et al.*, Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T-Cell Decline, and Per-Parasite Pathogenicity. *Molecular Biology and Evolution*. **35**, 27–37 (2017).
20. C. Fraser, T. D. Hollingsworth, Interpretation of correlations in setpoint viral load in transmitting couples. *AIDS*. **24** N2 - (2010).
21. V. Mitov, T. Stadler, A Practical Guide to Estimating the Heritability of Pathogen Traits. *Molecular Biology and Evolution*. **35**, 756–772 (2018).
22. T. S. H. C. Study *et al.*, Cohort Profile: The Swiss HIV Cohort Study. *International Journal of Epidemiology*. **39**, 1179–1189 (2009).
23. K. Kusejko *et al.*, 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).
24. A. U. Scherrer *et al.*, Cohort Profile Update: The Swiss HIV Cohort Study (SHCS). *International Journal of Epidemiology* (2021).
25. The INSIGHT START Study Group, Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N Engl J Med*. **373**, 795–807 (2015).
26. S. Lodi *et al.*, Time From Human Immunodeficiency Virus Seroconversion to Reaching CD4+ Cell Count Thresholds <math>\leq</math>200, <math>\leq</math>350, and <math>\leq</math>500 Cells/mm<sup>3</sup>: Assessment of Need Following Changes in Treatment Guidelines. *Clinical Infectious Diseases*. **53**, 817–825 (2011).
27. A. van Sighem *et al.*, Estimating HIV Incidence, Time to Diagnosis, and the Undiagnosed HIV Epidemic Using Routine Surveillance Data. *Epidemiology KW*. **26** (2015).
28. A.-C. Pineda-Peña *et al.*, Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infection, Genetics and Evolution*. **19**, 337–348 (2013).
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 Research*. **42**, e144 (2014).
30. M. A. Jensen *et al.*, 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. *Journal of Virology*. **77**, 13376–13388 (2003).
31. M. Prosperi *et al.*, Robust Supervised and Unsupervised Statistical Learning for HIV Type 1 Coreceptor Usage Analysis. *AIDS Research and Human Retroviruses*, 305–314 (2009).

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 USA*. **89**, 1934 (1992).
33. S.-Y. Rhee *et al.*, Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Research*. **31**, 298–303 (2003).
34. J. M. Carlson *et al.*, Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. *Journal of Virology*. **86**, 13202–13216 (2012).
35. J. J. Lum *et al.*, Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis. *J Clin Invest*. **111**, 1547–1554 (2003).
36. M. A. Suchard *et al.*, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*. **4** (2018).
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 in Ecology and Evolution*. **8**, 28–36 (2017).
38. M. S. Gill *et al.*, Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci. *Molecular Biology and Evolution*. **30**, 713–724 (2012).
39. E. de Silva, N. M. Ferguson, C. Fraser, Inferring pandemic growth rates from sequence data. *Journal of The Royal Society Interface*. **9**, 1797–1808 (2012).
40. D. Bezemer *et al.*, In-country acquisition of HIV-1 non-B infection within the Netherlands is frequent, but results in limited onward transmission. *AIDS* (2021).
41. R. A. Neher, C. A. Russell, B. I. Shraiman, Predicting evolution from the shape of genealogical trees. *eLife*. **3**, e03568 (2014).
42. L. Råberg, M. Stjernman, in *Ecological immunology*, G. D. A. R. Nelson, Ed. (OUP), pp. 548–578.
43. L. Råberg, How to Live with the Enemy: Understanding Tolerance to Parasites. *PLOS Biology*. **12**, e1001989 EP – (2014).
44. N. Pantazis *et al.*, Temporal trends in prognostic markers of HIV-1 virulence and transmissibility: an observational cohort study. *The Lancet HIV*. **1**, e119–e126 (2014).
45. J. O. Wertheim *et al.*, Natural selection favoring more transmissible HIV detected in United States molecular transmission network. *Nature Communications*. **10**, 5788 (2019).
46. F. Blanquart *et al.*, A transmission-virulence evolutionary trade-off explains attenuation of HIV-1 in Uganda. *eLife*. **5**, e20492 (2016).

47. J. T. Herbeck *et al.*, Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS*. **26** (2012).
48. T. C. Porco, J. O. Lloyd-Smith, K. L. Gross, A. P. Galvani, The effect of treatment on pathogen virulence. *Journal of Theoretical Biology*. **233**, 91–102 (2005).
49. J. T. Herbeck *et al.*, Evolution of HIV virulence in response to widespread scale up of antiretroviral therapy: a modeling study. *Virus Evolution*. **2** (2016), doi:10.1093/ve/vew028.
50. H. E. Roberts, P. J. R. Goulder, A. R. McLean, The impact of antiretroviral therapy on population-level virulence evolution of HIV-1. *Journal of The Royal Society Interface*. **12**, 20150888 (2015).
51. The code used for our analyses is available at doi:10.5281/zenodo.5761935.

## References appearing only in Supplementary Material

52. T. S. Boender *et al.*, AIDS Therapy Evaluation in the Netherlands (ATHENA) national observational HIV cohort: cohort profile. *BMJ Open*. **8**, e022516 (2018).
53. D. Bezemer *et al.*, 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).
54. D. Bezemer *et al.*, Declining trend in transmission of drug-resistant HIV-1 in Amsterdam. *AIDS*. **18** (2004).
55. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*. **7**, 539 (2011).
56. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. **35**, 526–528 (2018).
57. M. Cornelissen *et al.*, From clinical sample to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing. *Virus Research*. **239**, 10–16 (2016).
58. A. Gall *et al.*, Universal Amplification, Next-Generation Sequencing, and Assembly of HIV-1 Genomes. *Journal of Clinical Microbiology*. **50**, 3838–3844 (2012).
59. M. Hunt *et al.*, IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics* (2015), doi:10.1093/bioinformatics/btv120.

60. C. Wymant *et al.*, Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus Evolution*. **4** (2018).
61. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
62. A. M. Wensing *et al.*, 2015 Update of the Drug Resistance Mutations in HIV-1. *Top Antivir Med*. **23**, 132–141 (2015).
63. C. Wymant *et al.*, PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Molecular Biology and Evolution*, msx304 (2017).
64. D. Bates, M. Mächler, Ben Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, Articles*. **67**, 1–48 (2015).
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).
66. J. C. Ho, G. T. Ng, M. Renaud, A. F. Poon, sierra-local: A lightweight standalone application for secure HIV-1 drug resistance prediction. *bioRxiv*, 393207 (2018).
67. H. Song *et al.*, Tracking HIV-1 recombination to resolve its contribution to HIV-1 evolution in natural infection. *Nature Communications*. **9**, 1928 (2018).
68. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology*. **215**, 403–410 (1990).
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. *Molecular Biology and Evolution*. **3**, 418–426 (1986).
71. K. Katoh, K. Misawa, K. I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. **30**, 3059–3066 (2002).
72. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology*. **4** (2006), doi:10.1371/journal.pbio.0040088.
73. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*. **67**, 901–904 (2018).
74. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution*. **4** (2018).

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. *Molecular Biology and Evolution*. **32**, 268–274 (2014).
76. E. M. Volz, S. D. W. Frost, Scalable relaxed clock phylogenetic dating. *Virus Evolution*. **3** (2017).
77. B. T. Grenfell *et al.*, Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*. **303**, 327–332 (2004).
78. B. L. Dearlove, S. D. W. Frost, Measuring Asymmetry in Time-Stamped Phylogenies. *PLoS Comput Biol*. **11**, e1004312 (2015).
79. S. Alizon, C. Fraser, Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology*. **10**, 49 (2013).
80. D. Bonsall *et al.*, A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *Journal of Clinical Microbiology*. **58** (2020), doi:10.1128/JCM.00382-20.
81. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*. **5**, e9490 (2010).
82. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. **14**, 587–589 (2017).
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 Evolution*. **2**, vew007 (2016).
84. M. Hernán, J. Robins, *Causal Inference: What If* (Boca Raton: Chapman & Hall/CRC, 2020).

## Acknowledgements

We thank Katrien Fransen and Guido Vanham for help with the Belgian data, Oliver Ratmann for help in identifying the Dutch clusters, Katharina Kusejko for testing for additional VB individuals in the SHCS, Brian Foley for help with genome sharing, Bethany Dearlove and Laura Thomson for help with software, and Joshua 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 CF. **Author contributions:** Funding acquisition: CF. Conceptualisation: CW and CF. Data generation and management: all authors. Investigation: CW, FB, DB, LF, PR and CF. Writing, original draft: CW and CF. Writing, review and editing: all authors. **Competing interests:** PK is an employee of Kymab a Sanofi company. HFG 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 Yvonne Jacob Foundation, personal fees from consulting or advisory boards or data safety monitoring boards for Merck, Gilead Sciences, ViiV Healthcare, Mepha, and Sandoz. HFG'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 [https://github.com/ChrisHIV/hiv\\_vb\\_variant](https://github.com/ChrisHIV/hiv_vb_variant), version deposited at (51). The 17 VB variant whole genomes are publicly available on GenBank with accession numbers MT458931-MT458935 and MW689459 - MW689470; the two putative recombinants are MW689465 and MW689466. Data on viral loads, pre-treatment CD4 counts and mortality are provided as Supplementary 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 contains sensitive and potentially identifying information. Once submitted the proposed research/analysis will undergo review by the BEEHIVE Data Access Committee, which includes representatives of the ATHENA cohort, 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/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 consisting 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 approval:** At initiation, the ATHENA cohort was approved by the institutional review board of all participating centres. People entering HIV care receive written material about participation in the ATHENA cohort and are being informed by their treating physician of the purpose of collection of data, after which they can consent verbally or elect to opt-out. Data are pseudonymised 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 those ATHENA individuals enrolled in BEEHIVE for whole-genome sequencing.

## Supplementary Material

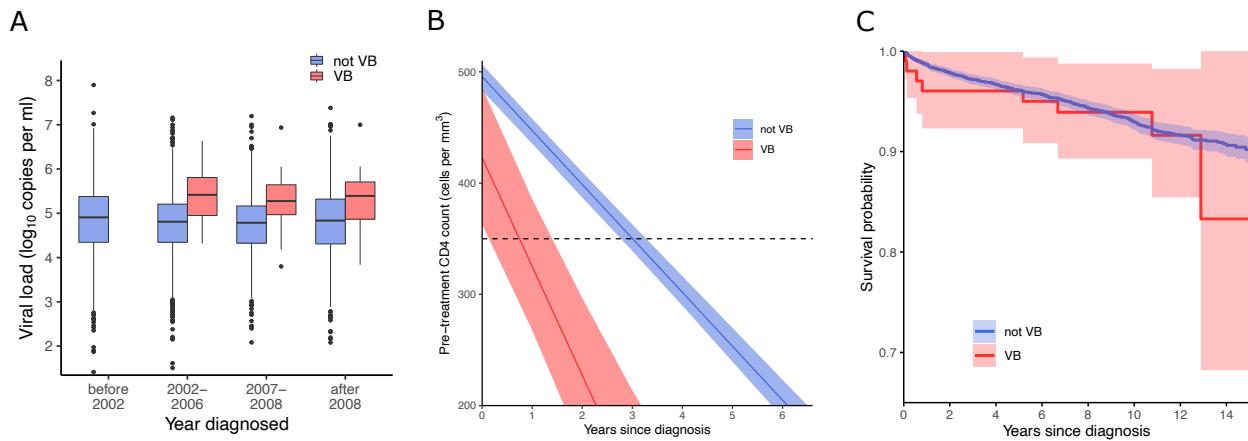
Materials and Methods

Supplementary Text

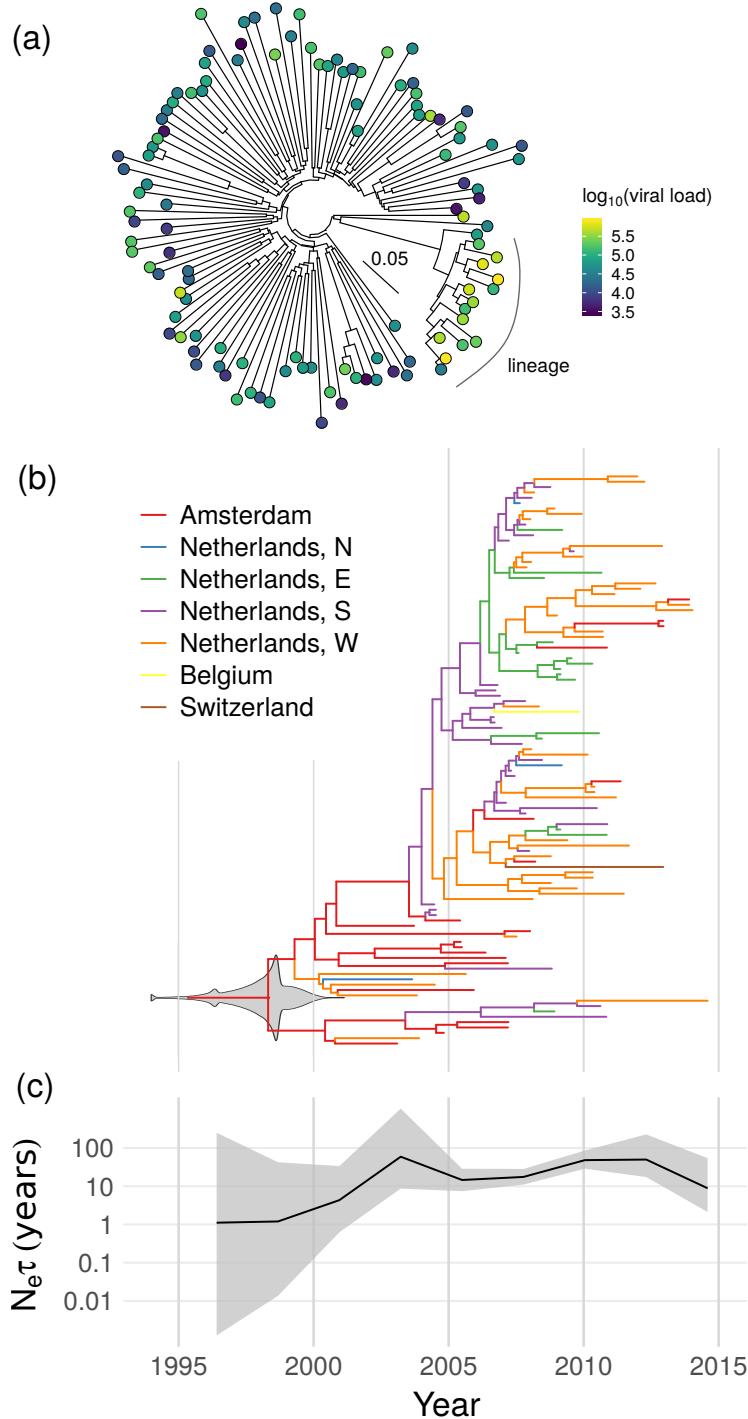
Figs. S1 to S8

Tables S1 to S7

Data S1 to S4



**Caption for Figure 1: clinical characteristics of “VB” individuals (those infected with the highly virulent variant, shown in red) and “not VB” individuals (those infected with any other subtype-B virus, shown in blue).** (a): a box-whisker plot of viral load, by year of diagnosis. Diagnosis dates were grouped to give boundaries coinciding 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 expected decline in CD4 count in the absence of treatment. Sex and age at diagnosis have been adjusted for; values shown are for males diagnosed aged 30-39 years. The shading indicates 95% confidence intervals in the model’s prediction of the mean values, given 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 shows a CD4 count of 350 cells per  $\text{mm}^3$ , discussed in the text. (c) The probability of still being alive at a given time after diagnosis.



**Caption for Figure 2: phylogenetic and phylodynamic analysis of the VB variant.** (a) A whole-genome maximum-likelihood phylogeny of 15 VB variant sequences and 100 background subtype-B sequences. The colour of each circle at the tips indicates the individual's viral loads in

$\log_{10}$  copies per ml. The inset scale bar shows the branch length scale in units of substitutions per site. (b) A dated maximum-clade-credibility tree for 107 partial *pol*-gene sequences from the VB variant. Colour indicates the region of the Netherlands (N, E, S, W abbreviating north, east, south and west), which is known for the tips, and otherwise inferred by ancestral state reconstruction. The grey 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 effective population size  $N_e$  (scaled by the coalescent generation time  $\tau$ ) over time with 95% credibility intervals, with the same time axis as the panel above.

Test	Discovery	Replication
<b>Dataset</b>	BEEHIVE (Europe)	ATHENA (Netherlands), excluding overlap with BEEHIVE*
<b>Viral load measurements compared<sup>†</sup></b>	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
<b>Mean and inter-quartile range (IQR) of viral load in</b>	5.10 (IQR 4.69 – 5.58)	4.79 (IQR 4.34 – 5.27)

<b>not-VB individuals, in <math>\log_{10}</math> copies per ml</b>		
<b>Mean and inter-quartile range (IQR) of viral load in VB individuals, in <math>\log_{10}</math> copies per ml</b>	5.84 (IQR 5.57 – 6.09)	5.33 (IQR 4.94 – 5.75)
<b>Viral load increase in VB individuals</b>	0.74 $\log_{10}$ copies per ml	0.54 $\log_{10}$ copies per ml
<b>p value for increase</b>	$5 \times 10^{-6}$ (two-tailed t-test, significant at a level of $5 \times 10^{-5}$ when Bonferroni-corrected for performing 50 such tests)	$1 \times 10^{-12}$ (one-tailed t-test)

**Table 1: Comparison of viral loads between individuals infected with the VB viral variant, and other individuals.** \*When analysing the viral loads of individuals in ATHENA, we first excluded individuals who were in BEEHIVE, for the test to be independent of the initial finding within BEEHIVE. After our statistical tests of viral load, we did not exclude BEEHIVE individuals from the ATHENA data for subsequent analyses. †The number of individuals  $N$  indicated is after excluding those without viral load measurements before treatment.