Supplementary methods

Phylogenetic Analysis

We aligned the sequences by using the BioEdit tool and manually corrected the alignment according to the encoded reading frame. Duplicate sequences from the same province were discarded. If several sequences from the same patient might be available in the database, we kept only the oldest one. Long branches tree were re-confirmed for their genotype and those found miscatalogued were eliminated from our study.

Definition of Recent Infection

The ambiguity score is a measure of the viral nucleotide diversity determined using bulk sequencing, which provides an estimate of the infection duration. Sequences in which ≤0.5% of the nucleotides are ambiguous indicate that the TDR test was performed on a recently infected patient. However, because diversity may be low in long-term HIV infections, we excluded patients with a score ≤0.5% and a CD4 cell count of <200 cells per μL to reduce false positives.

What Hukou system is in China

The Hukou system is a basic system of household registration in China. It officially identifies a person as a resident of an area and includes identifying information such as name, parents, spouse, and date of birth. The Hukou system is of critical importance to people in China. Someone without Hukou is regarded as an illegal resident. The Hukou is analogous to the US Permanent Resident Card, but it has a more important meaning to Chinese people. Hukou is the system of household registration used in China. The system itself is more properly called "Huji", and has origins in ancient China. A household registration record officially identifies a person as a resident of an area and includes identifying information such as name, parents, spouse, and date of birth. Due to its connection to social programs, which assigns benefits based on agricultural and non-agricultural residency status (often referred to as rural and urban), the Hukou system is sometimes likened to a form of caste system. Until very recently, each citizen was classified in an agricultural or non-agricultural Hukou and further categorized by location of origin. Those residents who held non-agricultural (i.e. urban) Hukou status received benefits not available to their rural counterparts. Internal migration was also tightly controlled, and only in the past few decades have these restrictions been loosened. While this system has played a major role in China's rapid economic growth, Hukou has also promoted and aggravated social stratification [1]. Therefore, the Hukou was the most appropriate variable to define the geographical origin of the individuals in this study.

1 Hukou system; http://wikipedia.moesalih.com/Hukou. (accessed Dec 3, 2018).