Supporting Information

Vasylyeva et al. 10.1073/pnas.1701447115

Genetic Data

Sequences Sampling. We used 448 HIV-1 genetic sequences sampled in 2012–2015 from patients of 24 AIDS Centers (52% males). Overall, there are 79,000 HIV infected individuals who receive ART in Ukraine (33); 7,000 (10%) of them experience treatment failure and require a drug-resistance test. Due to a limited availability of the tests (~500 per year), physicians decide who gets to be tested, and this decision is based on the current state and medical history of a patient: Patients with multiple changes of drug schemes get a priority. Every region has a quota of tests per year, based on the HIV prevalence and treatment coverage in the region. Unfortunately, even though the number of tests is insufficient to cover the needs of all patients, some physicians also choose to not use their quota. Thus, some test systems remain unused, and the actual coverage is <500 tests per year. All of the available from 2012 to 2015 high-quality (<5% of ambiguous nucleotides) sequences were used in the analysis.

Alignment. We aligned the sequences using MEGA software (Version 7.0) (40) and then manually edited the alignment deleting all of the codon positions associated with drug resistance (RT: 41, 65, 67, 69, 70, 74, 75, 77, 100, 101, 103, 106, 115, 116, 151, 179, 181, 184, 188, 190, 210, 215, 219, 225, 230; PR: 23, 24, 30, 32, 46, 47, 48, 50, 53, 54, 73, 76, 82, 83, 84, 85, 88, and 90) (41).

Global Subtype-A Reference Dataset. To explore the position of the 427 Ukrainian sequences (Ukrainian dataset) within the global subtype-A epidemic, we merged our sequences with a globally representative subtype-A reference dataset which consisted of 2,199 sequences using Clustal W (44). The dataset consisted of all available sequences of minimal length of 300 nucleotides from the *pol* region of HIV genome (nucleotides 6,225–8,795) accessed in 2011 from the HIV sequence database (sequence names are in Dataset S1, Table S1*C*). Sequences came from 36 countries (Afghanistan, Albania, Azerbaijan, Burkina Faso, Burundi, Benin, Democratic Republic of Congo, Chad, Congo, Cameroon, Cuba, Cyprus, Czech Republic, Ethiopia, France, Gabon, Georgia, Ghana, Equatorial Guinea, Kenya, Kazakhstan, Latvia, Mali, Nigeria, Russia, Rwanda, Sudan, Slovenia, Senegal, Togo, Tanzania, Ukraine, Uganda, and Uzbekistan) sampled in 1985–2010. We used RAxML (45) to construct a ML phylogenetic tree of 2,626 sequences (2,199 reference and 427 Ukrainian) (Fig. 1).

Phylogeography

Combined Dataset for Phylogeographical Analysis. We further selected 36 subtype-A HIV-1 reference sequences sampled in 1987–2013 publicly available from the Los Alamos database to improve the molecular clock calibration. Sequences were selected to ensure the best geographical and temporal representation. These reference sequences come from the following countries: Afghanistan, Armenia, Belarus, Benin, Burundi, Cameroon, Cuba, Cyprus, Ethiopia, Georgia, Kazakhstan, Kenya, Kyrgyzstan, Mali, Russian Federation, Rwanda, Senegal, South Africa, Sudan, Tanzania, Uganda, and Zambia. We aligned these reference sequences ($n = 36$) to the Ukrainian dataset ($n = 427$) in a combined dataset using Clustal W. We further constructed a ML phylogenetic tree in RAxML with these 463 sequences (Fig. S1). We have tested the temporal signal of the combined dataset with the Tempest program (55): The observed root-to-tip correlation coefficient was 0.5, which is sufficient signal for our analysis. This combined dataset was further used to produce the posterior tree distribution for the phylogeographic analysis. The ggtree package in R was used to visualize the phylogenetic trees in Fig. 2 and Fig. S1 (56).

Sensitivity Analysis. We explored estimates of among-location viral lineage movement under condition of equal representation of locations in the analysis. Specifically, we randomly down-sampled sequences in each location to match the number of sequences in the third smallest location (Odessa; $n = 57$). This resulted in five locations with the same number of sequences (Center, East, Kyiv, Odessa, and South) and two locations with a smaller number of sequences (Crimea and West; $n = 10$ and $n = 22$, respectively), which were unlikely to be significant contributors to virus gene flow, based on the preceding analyses. We did not reduce the number of sequences to the smallest sample size ($n = 10$) to ensure acceptable statistical power.

We conducted 10 replicate analyses with different random subsamples of the dataset. We found that East was the main viral lineage export location in 70% of the analyses (7 of 10 subsamples). In those replicate analyses, where East was the main exporter of the viral lineages, it accounted on average for 91% of migration events. In the other three subsamples, Center, Kyiv, and South were the main exporters each, respectively. For the number of migration events observed in each subsample, see Dataset S1, Table S3.

Epoch Analysis. We compared a time-homogenous model (single epoch), which assumed no major change in viral lineage migration pattern within the dataset, and time-heterogeneous (two epoch), which assumed a change at some point within the sampled dates. For the purposes of this analysis, we specified two epochs for the two-epoch model, before and after December 31, 2013. We were not able to specify the official date of the beginning of war (April 6, 2014) as the transition time, because the dates in our analysis were not resolved at a monthly scale; only sampling year data were available. To define which model fit the data best, we applied path-sampling (PS) (53) and steppingstone sampling (SS) (54) marginal likelihood estimator (MLE) techniques. We assumed that the BF >20 was enough statistical support to favor one of the models.

We first compared the two models as implemented in our primary discrete trait analysis: grouping the data into seven discrete locations. Both PS and SS MLE were higher for the single-epoch model, but with a very low statistical support (BF = 3) (Dataset S1, Table S4). Since we tested the time heterogeneity over a very short period of time (4 and 2 y in each epoch), we decided to reduce the number of locations in the analysis to avoid the overparameterization of the analysis. For this purpose, we created 10 subsets of data where the sequences were grouped into two locations: East and Other. For each of the subsets, we included all of the sequences from East and the same number of sequences subsampled from all of the other locations. Comparison of the two models with two locations resulted in a very strong support for the time-heterogeneous model (BF > 30) for both PS and SS methods. We followed the standard terminology in BF interpretation: The strength of evidence for a particular model is substantial when BF > 3, strong if BF > 10, very strong if BF > 30, and decisive if BF > 100 (23).

Correlation Analysis

IBBS Data. Data on risky injecting behaviors came from the IBBS of PWID ($n = 9,002$) implemented in Ukraine in 2013. IBBS has a cross-sectional survey design and was conducted via respondent-driven sampling. We conducted the secondary analysis of data

collected in 25 regional capital cities of Ukraine (only data corresponding to the 24 administrative units for which genetic sequences were available were used). Only PWID in regional capitals were surveyed, since Ukraine has a highly urbanized HIV epidemic: 77% of all HIV cases in 2013 were registered among the urban population (57). The cities, selected as the study sites for the IBBS, represent all geographical areas of Ukraine. These areas represent varying HIV-prevalence rates among PWID and varying PWID population sizes (58, 59). Sample size was calculated for each study site by combining group size estimation and HIV-prevalence level from the previous IBBS study conducted in 2011 (60). The inclusion criteria for study participants included: (*i*) injection drug use in the last 30 d before the study, (*ii*) 14 y old or older; and (*iii*) permanent resident of the city where IBBS was conducted.

Epidemiological Data. We accessed per-region statistics from publicly available sources as follows:

(*i*) The number of IDP was available from the website of the CEDOS analytical center (www.cedos.org.ua/). This website summarizes data delivered by the United Nations High Commissioner for Refugees and the Ukrainian Ministry of Social Policy (61);
(*ii*) HIV prevalence estimates (cases per 100,000 population) in the general population were available from the website of the Ministry of Health of Ukraine (62);
(*iii*) The number of HIV-positive IDP registered with an AIDS center at the place of their current residence was available from the National AIDS Center website (ucdc.gov.ua/); and
(*iv*) The population size of every region as of 2016 was available from the website of the State Statistics Service of Ukraine (63).

Construction of Behavioral and Epidemiological Variables. For each of the seven locations, we calculated values of a set of behavioral and epidemiological variables. For every variable, we present summary values of this variable for each of the administrative regions in the respective location. As mentioned in the main text, for continuous variables, we used mean and median values, and for the categorical variables we used a proportion of positive answers. The list of explanatory variables includes:

(*i*) Injecting practices that might increase the HIV transmission risk ("Risky injecting practices"):

- Median frequency of injections per week in regions that are part of a location (i.e., for East, median duration of drug use for PWID in Donetsk and Lugansk);

- Median number of PWID that PWID used drugs with at the same place;

- Proportion of PWID who sometimes gave used syringes to other injectors.

(*ii*) Injecting practices that might protect against HIV transmission ("Safe injecting practices"):

- Proportion of PWID who used sterile syringe at the last injection.

(*iii*) Sexual practices that might increase the HIV transmission risk ("Risky sexual practices"):

- Mean number of casual sexual partners per injector;

- Mean number of commercial partners per injector;

- Mean number of all sexual partners (overall number of sexual partners);

- A proportion of PWID who participated in group sex.

(*iv*) Sexual practices that might protect against HIV ("Safe sexual practices"):

- Proportion of PWID who always used condoms with casual partners;

- Proportion of PWID who always used condoms with commercial partners;

- Proportion of PWID who used condoms at the last sexual intercourse.

(*v*) Regional level epidemiological characteristics:

- HIV prevalence in PWID as measured in the IBBS-2013;

- Number of IDP relocated to and registered in a location;

- Number of HIV-positive IDP relocated to and registered in a location;

- Cumulative number of HIV-diagnosed people in a location (HIV prevalence in general population) as of 2013;

- Population size (number of residents) in a location (combined number of residents of all of the regions in this location) as of 2013.

The behavioral variables were constructed as following:

(*i*) For continuous variables, we looked at mean and median values per location. For example, mean number of commercial partners per an injector. The list of all of the variables is given above and in Dataset S1, Table S5.
(*ii*) For categorical variables, we looked at the percent of respondents who answered "Yes" or "Always" to a question about a specific behavior. For example, the proportion of PWID who answered "Always" to the question "How often did you use someone else's used syringe?"

Finally, behavioral data were aggregated to the same geographic locations as those used in the phylogeographic analysis. Thus, "a median age of PWID" for East is a median for the combined responses of PWID in Donetsk and Lugansk.

Limitations. While there might be a further variability in behaviors within a location, we were not able to conduct a more detailed statistical analysis. It was necessary to group sequences into seven locations because of the small number of sequences from some administrative regions. We then had to request the aggregated behavioral data that correspond to the locations from the phylogeographical analysis. This possibly resulted in lessened statistical power to detect significant correlations between the virus movement and behaviors and limited our ability to control for potential bias.

Drug Resistance

We performed a basic drug-resistance analysis on our samples, but did not observe a change in the frequency of drug-resistance mutations since the beginning of the war, apart from a small decrease in L90 mutation (Fig. S2 and Dataset S1, Table S6*B*). We found almost no difference in the number of drug-resistance mutations in the sequence from 2012 to 2013 and those from 2014 to 2015 (Fig. S2 and Dataset S1, Table S6*B*) for all of the classes of drugs used in ART in Ukraine in those years: nucleoside reverse-transcriptase (RT) inhibitors [azidothymidine, tenofovir desoproxil fumarate (TDF), lamivudine, emtricitabine, and abacavir], nonnucleoside RT inhibitors (efavirenz and nevirapine), and protease inhibitors (lopinavir/ritonavir and darunavir/ritonavir). We observed an increase (albeit nonsignificant) in the proportion of sequences with a K65R mutation since the initiation of war, which also coincides with an increased percentage of TDF-based regimens prescribed in the general HIV population (Dataset S1, Table S6*A*). K65R is a signature mutation for TDF resistance associated with early virological failure (64). However, if we compare the prevalence of this mutation between the beginning (2012) and the end (2015)

of our observation period, then the increase is statistically significant (χ^2 test: $P = 0.03$): K65 was observed in 51 of 152 patients (34%) tested in 2015 and in 2 of 22 patients (9%) in 2012.

Based on the recent scale-up of TDF in the Ukrainian HIV population, we would expect the time to treatment failure in our sample to be ∼12–24 mo, comparable to the time observed in the early days of highly active ART (65). The increase in patients with an absence of primary resistance mutations within the sampling period was insignificant (Dataset S1, Table S6*B*), suggesting that cases with severe adherence issues were uniformly distributed before and after war.

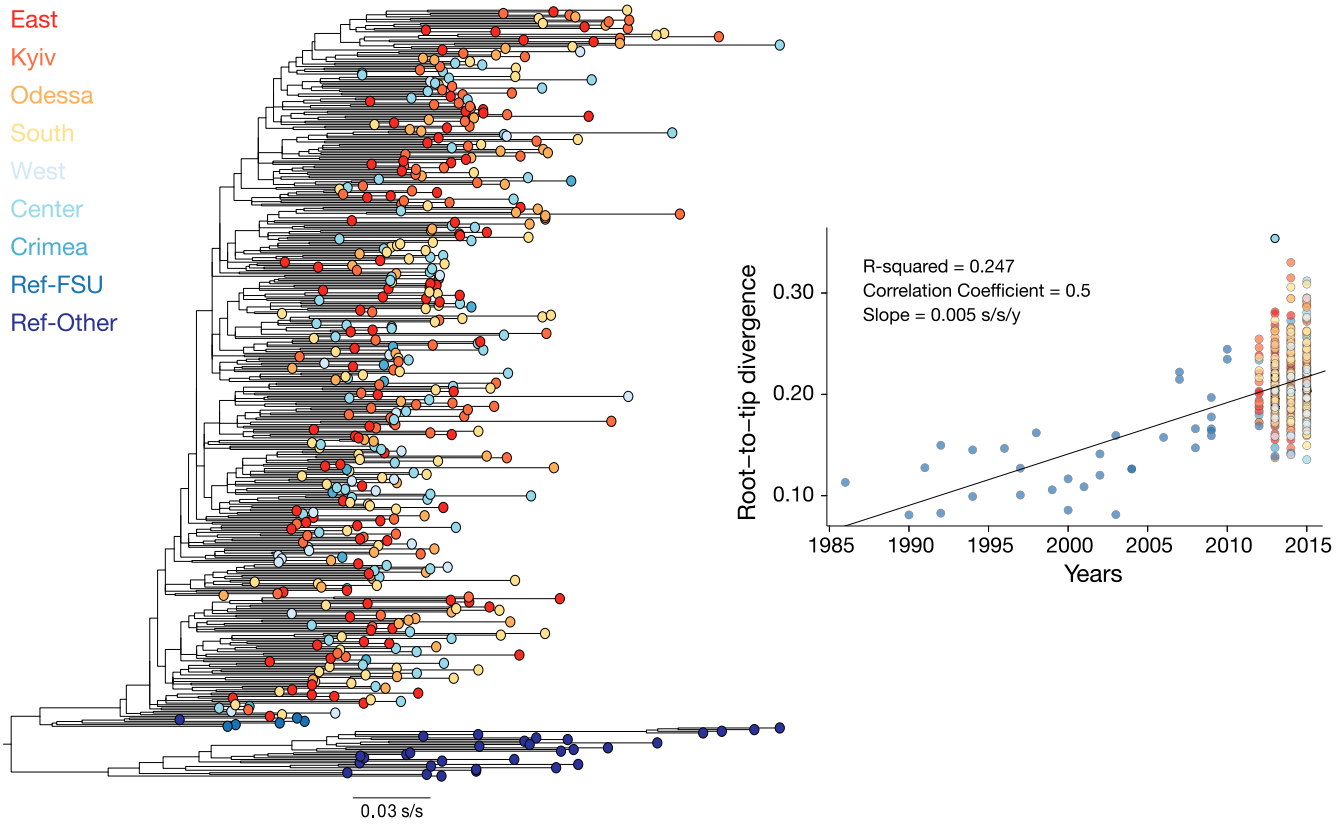

Fig. S1. ML phylogenetic tree reconstructed in RaXML from the combined dataset ($n = 463$). The combined dataset consisted of the Ukrainian dataset ($n = 427$) and the reference sequences ($n = 36$); reference sequences were included to improve the temporal signal. Results of the Tempest analysis that indicate the strength of the signal in the combined dataset are presented in *Right*.

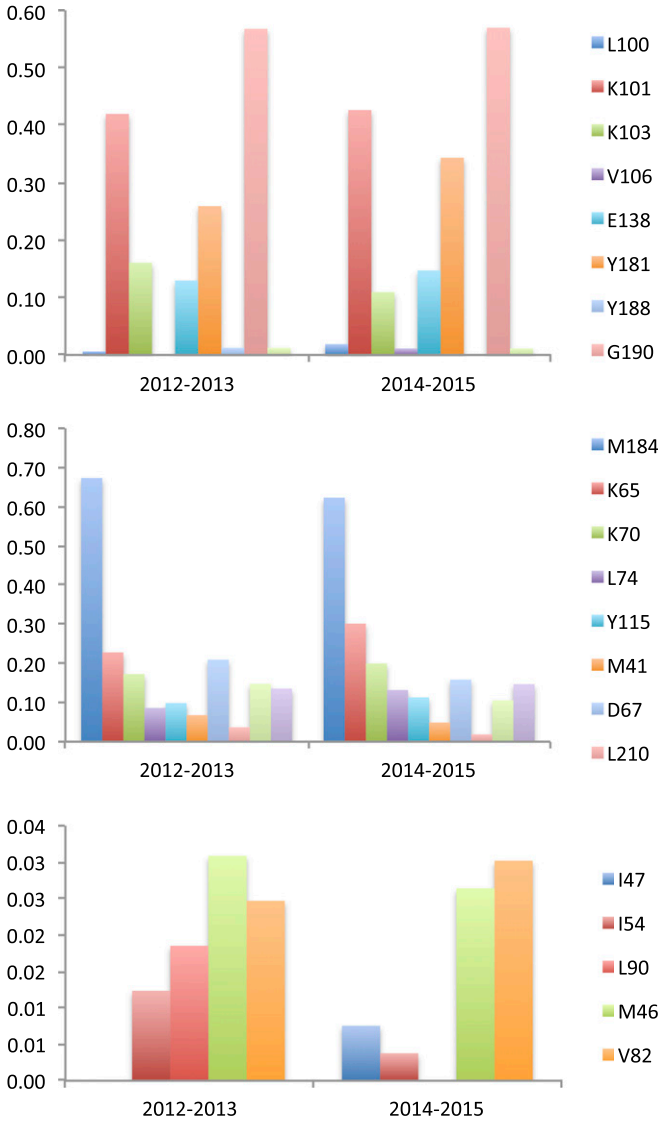

Fig. S2. Proportion of patients with drug-resistance-associated mutations in patients before and after the initiation of war.

Other Supporting Information Files

Dataset S1 (XLSX)