

Supplementary Material

Contents

Supplemental Methods

Study design and sample selection	3
Sample size calculations	3
WebLogo analyses	4
Phylogenetic analyses - RAxML trees	4
Structure simulation and modeling of emerging DRMs and treatment-associated <i>pol</i> mutations	4
Structural stability prediction and mutational impact scoring	5
Software scripts and visualization	5
Statistical analysis	6
References	7

Supplemental Figures

Figure S1	9
Figure S2	10
Figure S3	11
Figure S4	12
Figure S5	13
Figure S6	14
Figure S7	15
Figure S8	17
Figure S9	18
Figure S10	19
Figure S11	20
Figure S12	21
Figure S13	22
Figure S14	23

Supplemental Tables

Table S1	24
Table S2	26
Table S3	27
Table S4	28

Supplemental Methods

Study design and sample selection

In an ecological analysis, 8,130 Cameroonian HIV-1 *pol* nucleotide sequences were downloaded from the Los Alamos National Laboratory (LANL) HIV sequence database and studied. The data set was composed of 3,741 RT, 3,585 PR, and 802 IN coding sequences after a multi-step quality check and selection process to eliminate non-functional sequences (sequences with stop codons), clonal sequences, next-generation sequencing reads from same patients, and poor quality sequences. Sequence quality assessment and data selection were made using MEGA v.5.2 [1], Microsoft Excel 2016, FigTree v.1.4.3 [2], program R v.4.0.0 [3], and RStudio v.1.3 [4] with scripts developed based on *seqinr* and *dplyr* packages.

In our goal to assess temporal trends of population outcomes, regional access to cART was the primary variable, and presence/change of polymerase mutations the primary outcome. Thus, the sequences were assigned to periods before (≤ 2003) or after (2004–November 2020) the implementation of cART in Cameroon, i.e., when cART became available to the broader population. For simplification, the former is termed pre-cART and the latter post-cART throughout the paper. In each data set, duplicates were eliminated using the *ElimDupes* tool on the LANL database or data refinement scripts in R and R Studio, as explained above. Given the dominant prevalence of CRF02_AG ($>50\%$ for most times/genomic regions) in Cameroon's epidemic, the sequences were further sub-grouped into CRF02_AG and non-CRF02_AG. Subtype assignment strictly adhered to the provided information of the LANL database entries, i.e., CRF02_AG classification was exclusively given to sequences annotated as CRF02_AG, whereas all other sequence names including 02G and A1G, which indicate relatedness to unique recombinant forms (UFRs), were considered non-CRF02_AG. CRF02_AG sequences from different branches of the phylogenetic tree and from different years before and after cART implementation in Cameroon were checked for recombination breakpoint patterns using the bootscanning methods provided by the REGA HIV-1 subtyping tool v.3.0, implemented in the Stanford University HIV Drug Resistance Database [5]. All sequences for which no subtype information or collection year were provided or exhibited disproportionate phylogenetic placements were excluded from the study. Canonical DRMs were assigned according to the Stanford University HIV Drug Resistance Database [6] as of November 3rd, 2020, which included 41 DRM sites for RT, 33 for PR, and 24 for IN. Susceptibility scoring of mutations to specific antiretroviral drugs and drug classes was based on the scoring provided by the Stanford University HIV Drug Resistance Database. To call mutations per site, we generated CRF02_AG consensus sequences using the Cameroonian data set pre-cART. Compared to the global CRF02_AG consensus sequence provided by LANL, the Cameroonian CRF02_AG consensus sequence had no mismatch in the PR region, and there was one mismatch at position 269 in IN: R globally versus K in Cameroon (HxB2: R). In RT, there were six mismatches: Compared to the global consensus sequence (first letter), the Cameroonian CRF02_AG consensus (last letter) differed at D36E (HxB2: E), R211K (HxB2: R), T322A (HxB2: S), I326V (HxB2: I), M357K (HxB2: M), and R395K (HxB2: K).

Sample size calculations

The study's primary endpoints were statistical comparisons of DRMs and NOPs between pre- and post-cART periods to identify treatment-associated mutations. For PR and IN, full-length

sequences were used that covered the entire 297 bp/99 aa PR region (HxB2 position bp 2,253-2,549, *pol* aa 57-155) or 864 bp/288 aa IN region (HxB2 position bp 4,230-5,093, *pol* aa 716-1,003), respectively. For the larger 1,320 bp/440 aa (HxB2 position bp 2,550-3,869, *pol* aa 156-595) RT region, as of November 3rd, 2020, there were only 93 full RT sequences available for CRF02_AG and 278 for non-CRF02_AG, before quality check and exclusion of non-functional or poor-quality sequences and subgrouping into pre- and post-cART. Since these numbers did not meet the sample size requirements to achieve significant power (see below), we used all Cameroonian RT sequences >100 bp in length, which increased the sample size to achieve significant power. Sequence alignments and site-specific data analysis were performed in the same way for RT as for PR and IN sequences.

Power calculations regarding statistically significant changes in site-specific amino acid (aa) frequencies over time had to consider the disequilibrium in sample numbers between pre- and post-cART periods, which were generally higher in the post-cART era. The highest disequilibrium was obtained for RT CRF02_AG (5.2 ratio post-cART/pre-cART) and the smallest for IN CRF02_AG (1.3 ratio). A ratio of 5.2-times higher post-cART versus pre-cART sequences required 66 versus 343 pre-cART and post-cART sequences, respectively, to achieve 80% power in Fisher exact tests to detect a change from 1 to 8 % mutation prevalence, which was given for all RT and PR data sets. For the IN data sets, sample sizes of 108 versus 140 (ratio 1.3) were needed to achieve 80% power to determine a change from 1 to 8% in site-specific amino acid frequencies in Fisher exact tests. Mann-Whitney comparisons of matched mutation pairs on ≥ 99 amino acid sites achieved $\geq 99.8\%$ power to detect a 0.5 standard deviation (SD) difference of means (5% error). Kruskal Wallis tests were done to compare mutational frequencies among RT, PR, and IN sequences, which achieved $\geq 93.8\%$ power to detect a 0.5 SD difference between two independent groups of ≥ 99 sequences (5% error). Power calculations were done with G*Power v.3.1.9.4.

WebLogo analyses

The WebLogo [7] and AnalyzeAlign tools [8] were used for a comparative amino acid sequence analysis of pre- and post-cART sequences to determine the emergence and quantitative change of DRMs and NOPs.

Phylogenetic analyses - RAxML trees

Phylogenetic trees served the purpose of quality assessment of the sequences, to determine whether pre- and post-cART sequences are evolutionarily interspersed, and/or whether evolutionary changes occurred between pre- to post-cART periods. RT sequences (HIV positions 2,550 to 3,869 according to HxB2 numbering), PR sequences (2,253 to 2,549), and IN sequences (4,230 to 5,093) were aligned with reference and consensus sequences of HIV-1 group M subtypes and CRFs, downloaded from the LANL database [9], using Muscle in MEGA5.2.

Maximum likelihood (ML) trees were generated using the RAxML-HPC v.8.2.12 XSEDE tool on the Cipres Science Gateway [10]. A generalized time-reversible (GTR) substitution matrix (GTRCAT) was applied with a rapid bootstrap analysis (1,000 replicates) and a search for the best-scoring ML tree. Phylogenetic trees were visualized in FigTree v.1.4.3 [2].

Structure simulation and modeling of emerging DRMs and treatment-associated *pol* mutations

Canonical DRMs and novel, treatment-associated mutation sites that exhibited statistically significant changes from the pre- to post-cART period were structurally assessed in published

crystal structures as indicated. The structural models were obtained from the RCSB PDB database and manipulated in UCSF Chimera v1.13.1 [11]. Surface charge predictions were made using the ICM-REBEL tool of ICM-Pro (Molsoft), calculating the accurate electrostatic potential of proteins using a boundary element algorithm and generating a 3D surface skin model colored by potential [12]. Structural simulations and overlays using RT, PR, and IN CRF02_AG consensus sequences in comparison to HxB2 (K03455) were done using the Quaternary Structure Prediction/QSQE toolbox of the SWISS-MODEL server [13] and the MatchMaker tool in Chimera.

Structural stability prediction and mutational impact scoring

The Cartesian_ddg application [14,15] from Rosetta version 2020.28.61328 was used to predict effects of mutations on the stability of the complexes of RT, PR, or IN and drugs. Although mechanisms of drug resistance are still insufficiently explored, it is known that DRMs often alter drug-protein interactions by either stabilizing inhibitor-nonbinding or destabilizing inhibitor-binding conformations. The relevance of DRM-mediated protein destabilization was shown in different inhibitors including antiretroviral drugs, such as HIV protease inhibitors, by widening the active site [16-18]. The Cartesian_ddg protocol is faster than traditional dynamics simulations because only residues near the mutation site can move. Moreover, according to a previous study, it has high accuracy with a correlation coefficient of 0.88 between the predicted and experimental values [19]. Firstly, three complex crystal structures of RT with zidovudine (ID: 3V4I, HIV-1 RT with DNA and AZTTP) [20], nevirapine (ID: 3V81, HIV-1 RT with DNA and Nevirapine) [20], and doravirine (ID: 4NCG, Doravirine) [21], two complex structures of protease with darunavir (ID: 6DGX, HIV-1 protease NL4-3 WT in complex with darunavir) [22] and lopinavir (4L1A, HIV-1 protease Lopinavir complex) [23], and three complex structures of integrase with BI-224436 (ID: 6NUJ, HIV-1 integrase core catalytic domain and allosteric Inhibitor BI-224436 complex) [24], DTG (ID: 6RWN, SIVrcm intasome/dolutegravir complex) [25], and bictegravir (ID: 6RWO, SIVrcm intasome/bictegravir complex) [25] were downloaded from the RCSB PDB database [26]. These structures were prepared before computational mutagenesis by removing solvent and other unrelated structures. Then, the prepared complex structures were energy-minimized and refined using the “relax” command with the default settings. Finally, the lowest-scoring model of the protein-ligand complex after refinement was used as input for the Cartesian ddg protocol. The $\Delta\Delta G$ scores were estimated as a difference in mean scores for ten independent runs for every mutant and wild-type protein-drug complex structure. We defined mutations with a Rosetta $\Delta\Delta G$ between -1 and 1 kcal/mol, >1 kcal/mol, and < -1 kcal/mol, as neutral, destabilizing, and stabilizing, respectively.

Software scripts and visualization

Bar charts with annotations were created using Microsoft Excel 2016, GraphPad Prism v.8 (La Jolla California USA), and Program R/RStudio ggplot2 package. 3D bar charts were created in Excel. Multi-categorical alluvial diagrams were generated using RawGraphs with automatic sorting and 0.5 link opacity [27]. Stream graphs were created in RawGraphs using Silhouette representation (offset) and Basis spline interpolation. Correlograms were generated using the corrplot and RColorBrewer packages in program R/RStudio based on Spearman rank correlation coefficients r and multiplicity-corrected P values (Benjamini-Hochberg method) of pairwise correlations. Chord diagrams summarized all significant correlations in a circular representation where the color of the chord refers to the correlation coefficient r (red: positive correlations; blue: inverse correlations) and the width of the chord to the P value (wider chords for smaller P values/higher significance). The chord diagram exclusively displays mutations involved in significant correlations with other mutations.

Statistical analysis

Statistical testing of mutation/NOP frequencies was performed to quantitatively assess mutation/NOP changes over time and to ease the comparison of these changes across mutation sites. In terms of treatment-associated mutations, calculating statistical significance provided us with a valid measure to narrow down the number of sites per genomic region to the ones with the most pronounced changes over time, which eased visualization of the results.

Variant calling, observations counts, and calculations of site-specific mutation frequencies were done in R/RStudio using scripts developed with `seqinr` and `tidyverse` (`dplyr`, `lubridate`, `magrittr`) packages, exclusively considering non-ambiguous characters and gaps. Data output was collected and summarized in Excel. Fisher exact tests were performed in Program R/RStudio. Multiplicity corrections (Holm-Sidak) and calculations of q values of false discovery rates (FDR; two-stage set-up method of Benjamini, Krieger, and Yekutieli) were done in GraphPad Prism 8. Fisher P values <0.05 were considered significant, and q values <0.01 were considered new discoveries. Statistical comparisons between groups were made using non-parametric Mann-Whitney tests or Kruskal-Wallis tests with Dunn's multiplicity correction (Prism)

References

1. Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28: 2731-9.
2. Rambaut A. (2012). FigTree, v.1.4.3. <http://tree.bio.ed.ac.uk/software/figtree> (accessed on October 15, 2020).
3. R Core Team. R: (2013). A language and environment for statistical computing. <http://www.R-project.org/> (accessed on October 3, 2020).
4. RStudio Team. (2015). RStudio: Integrated Development for R. : RStudio, Inc., Boston, MA. <http://www.rstudio.com/> (accessed on October 3, 2020).
5. Stanford University HIV Drug Resistance Database. (1998). REGA HIV-1 Subtyping Tool - Version 3.0 [updated March 25, 2014]. <http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/> (accessed on October 3, 2020).
6. Stanford University HIV Drug Resistance Database. (2020). HIVdb Program. Genotypic Resistance Interpretation Algorithm. <https://hivdb.stanford.edu/> (accessed on October 3, 2020).
7. Crooks GE, Hon G, Chandonia JM and Brenner SE. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14: 1188-90.
8. Los Alamos National Laboratory (LANL) Database. (2020). AnalyzeAlign. https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html (accessed on October 3, 2020).
9. Los Alamos National Laboratory (LANL). (2020). Los Alamos National Laboratory (LANL) Database. <https://www.hiv.lanl.gov> (accessed on October 15, 2020).
10. A. MM, W. P and T. S. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*. 2010, p. 1-8.
11. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 25: 1605-12.
12. Totrov M and Abagyan R. (2001). Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide. *Biopolymers.* 60: 124-33.
13. Bertoni M, Kiefer F, Biasini M, Bordoli L and Schwede T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific reports.* 7: 10480.
14. Leman JK, Weitzner BD, Lewis SM, Adolfe-Bryfogle J, Alam N, Alford RF, et al. (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods.* 17: 665-80.
15. Park H, Bradley P, Greisen P, Jr., Liu Y, Mulligan VK, Kim DE, et al. (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J Chem Theory Comput.* 12: 6201-12.
16. Barouch-Bentov R and Sauer K. (2011). Mechanisms of drug resistance in kinases. *Expert Opin Investig Drugs.* 20: 153-208.
17. Chang MW and Torbett BE. (2011). Accessory mutations maintain stability in drug-resistant HIV-1 protease. *J Mol Biol.* 410: 756-60.

18. Sheik Amamuddy O, Bishop NT and Tastan Bishop O. (2018). Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics. *Sci Rep.* 8: 17938.
19. Strokach A, Corbi-Verge C and Kim PM. (2019). Predicting changes in protein stability caused by mutation using sequence-and structure-based methods in a CAGI5 blind challenge. *Human mutation.* 40: 1414-23.
20. Das K, Martinez S, Bauman J and Arnold E. (2012). HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol.* 19: 253–259.
21. Côté B, Burch J, Asante-Appiah E, Bayly C, Bédard L, Blouin M. et al. (2014). Discovery of MK-1439, an orally bioavailable non-nucleoside reverse transcriptase inhibitor potent against a wide range of resistant mutant HIV viruses. *Bioorganic Med. Chem. Lett.* 24(3): 917-922.
22. Lockbaum G, Leidner F, Rusere L, Henes M, Kosovrasti K, Nachum G. et al. (2019). Structural Adaptation of Darunavir Analogues against Primary Mutations in HIV-1 Protease. *ACS Infect. Dis.* 5(2): 316-325
23. Liu Z, Yedidi R, Wang Y, Dewdney T, Reiter S, Brunzelle J. et al. (2013). Crystallographic study of multi-drug resistant HIV-1 protease lopinavir complex: Mechanism of drug recognition and resistance. *Biochem. Biophys. Res. Commun.* 437(2): 199-204.
24. Koneru P, Francis A, Deng N, Rebensburg S, Hoyte A, Lindenberger J. et al. (2019). HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. *eLife* 8:e46344.
25. Cook N, Li W, Berta D, Badaoui M, Ballandras-Colas A, Nans A. et al. (2020). Structural basis of second-generation HIV integrase inhibitor action and viral resistance. *Science* 367(6479): 806-810.
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28: 235-42.
27. Mauri M, Elli T, Caviglia G, Uboldi G and Azzi M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. *CHIItaly '17*.

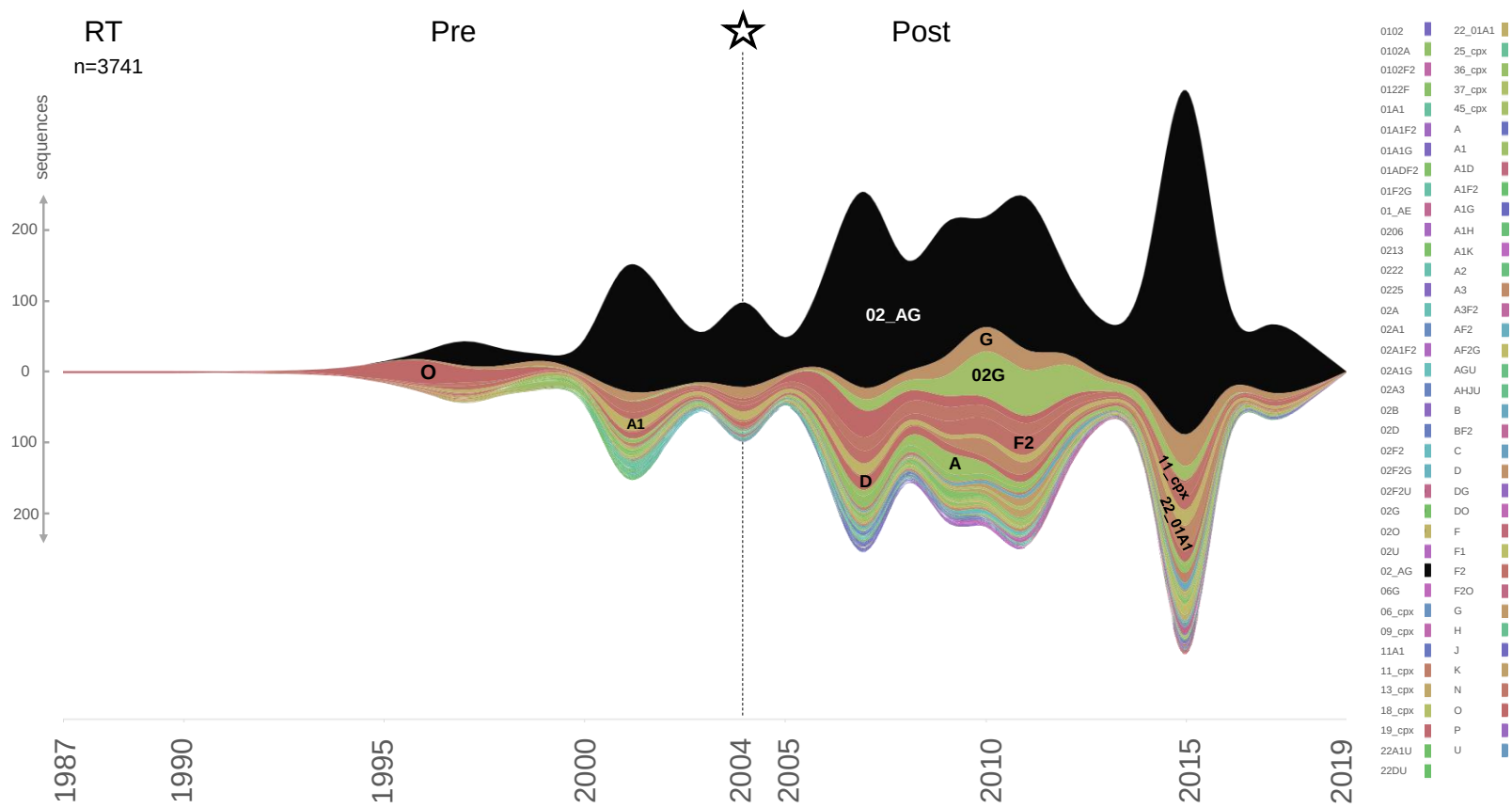


Figure S1. HIV-1 lineage distribution in Cameroon according to HIV-1 *pol* RT sequences from 1987 until 2020.

Streamgraph of lineage distribution of Cameroonian HIV-1 *pol* RT sequences (y-axis). All available RT sequences (HXB2 position bp 2,550-3,869, n=3,741) of at least 100 bp length from the LANL database are shown, as of November 3rd, 2020, after excluding non-functional, poor-quality, duplicate, and clonal sequences, as well as deep sequencing reads. HIV-1 subtypes, recombinant forms, and groups are color-coded according to the legend to the right, and the most prevalent lineages are also annotated in the graph.

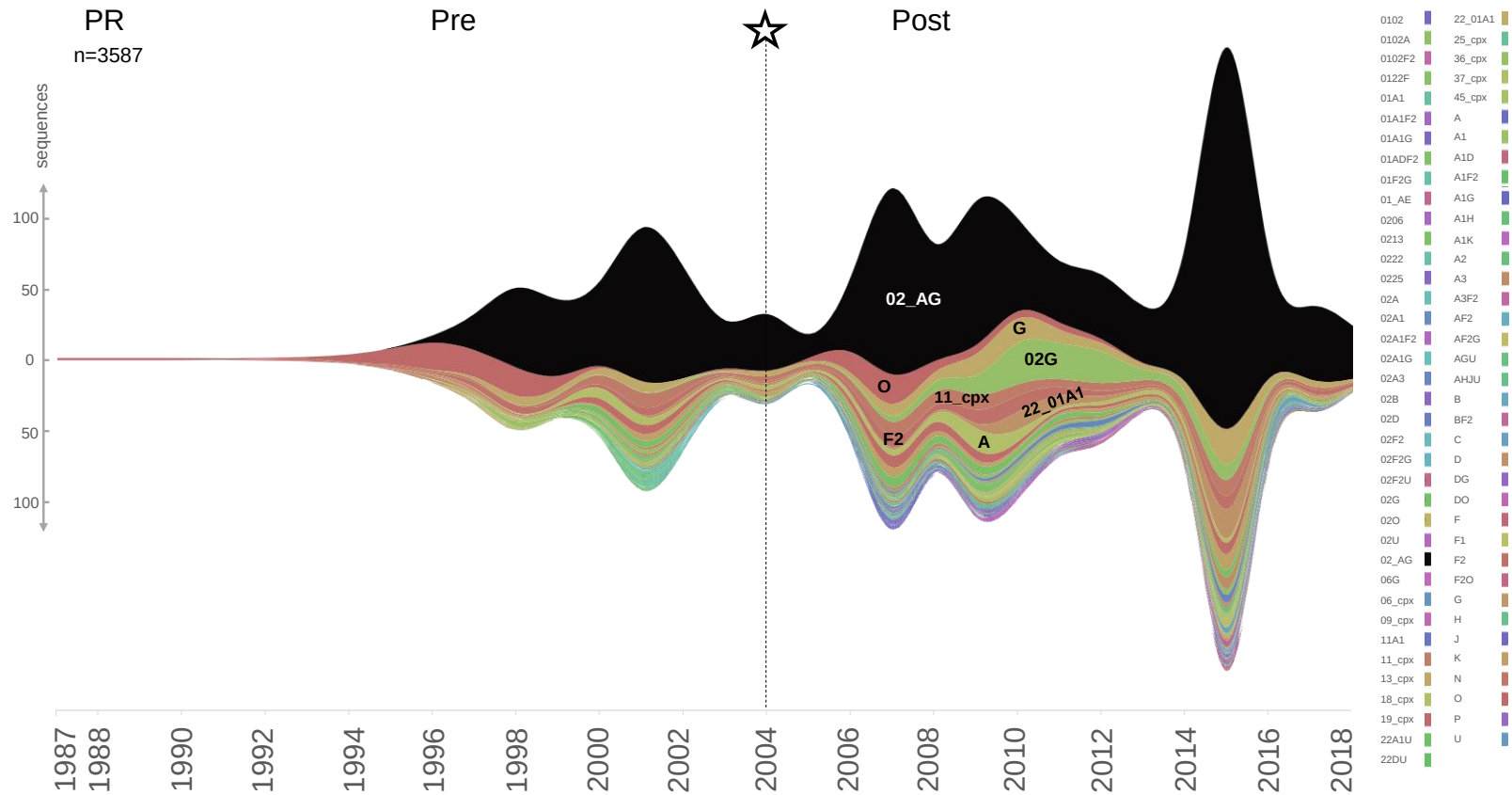


Figure S2. HIV-1 lineage distribution in Cameroon according to HIV-1 *pol* PR sequences from 1987 until 2020.

Streamgraph of lineage distribution of Cameroonian HIV-1 *pol* PR sequences (y-axis). All available full-length PR sequences (HXB2 position bp 2,253-2,549, n=3,587) from the LANL database are shown, as of November 3rd, 2020, after excluding non-functional, poor-quality, duplicate, and clonal sequences, as well as deep sequencing reads. HIV-1 subtypes, recombinant forms, and groups are color-coded according to the legend to the right, and the most prevalent lineages are also annotated in the graph.

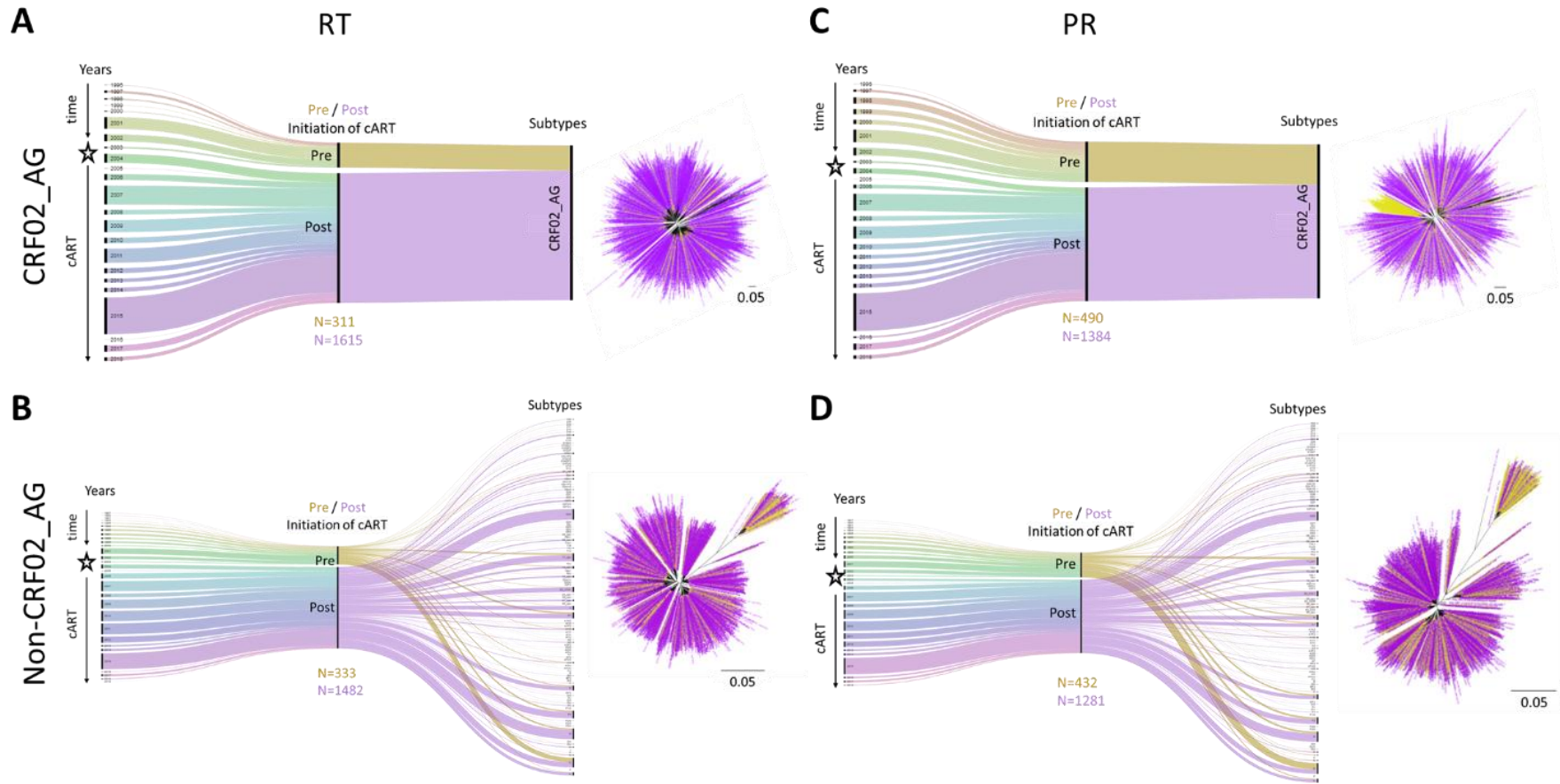
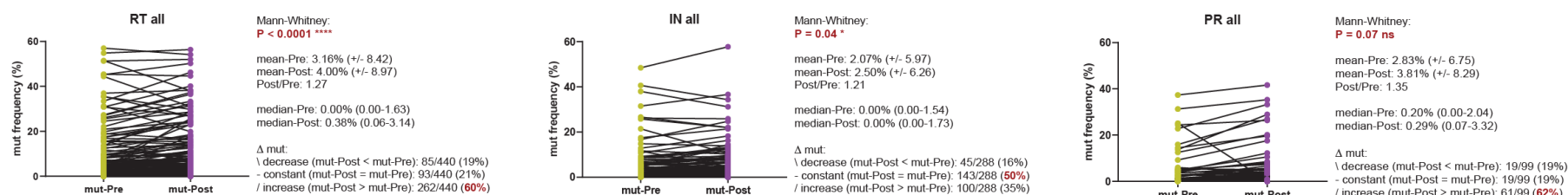


Figure S3. Data segregation and phylogenetic analysis in Cameroonian HIV-1 *pol* RT and PR sequences before and after the regional implementation of cART. HIV-1 RT (A, B) and PR (C, D) sequences from Cameroonian HIV-1-infected individuals (as in Figures S1 and S2) were segregated into CRF02_AG (A, C) and non-CRF02_AG data sets (B, D). The data set composition according to sampling years and lineage distribution is summarized in alluvial diagrams. Asterisks indicate subcategorization breakpoints of sequences collected pre- and post-implementation of cART in Cameroon (2004) along the timeline, and sequences colored in yellow (Pre) and purple (Post), respectively. Pre and Post sample numbers are indicated below the plots. Phylogenetic placement of Pre and Post sequences was analyzed in maximum-likelihood RAxML trees with the same yellow/purple color code. The scale indicates a 5% genetic distance.

A Side-by-side comparison of mutation frequencies Pre-cART vs Post-cART in each group:



B Comparison of Δ -mut frequencies (Post-cART minus Pre-cART) between groups:

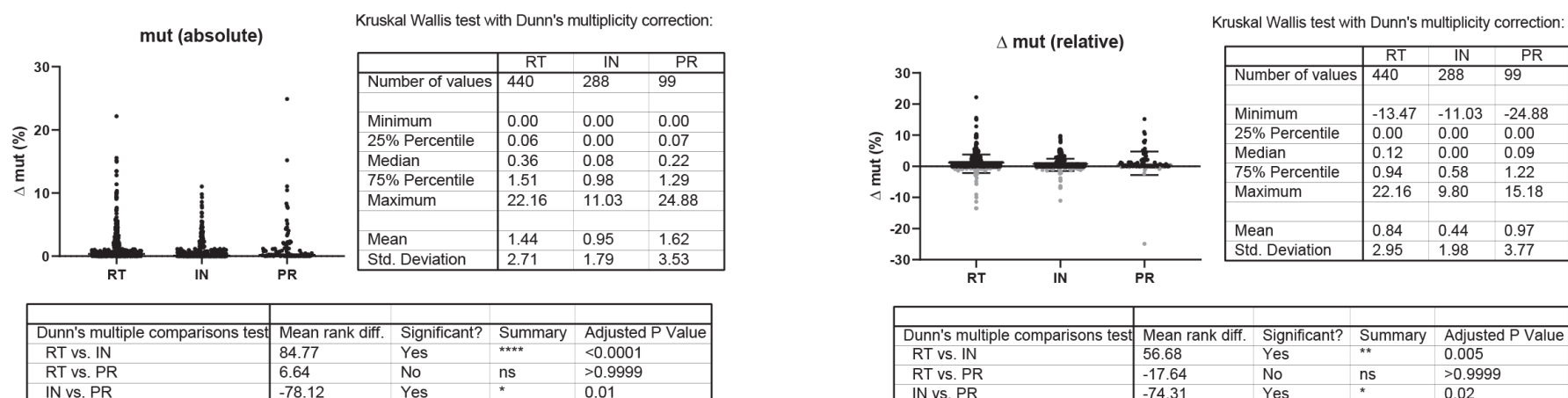


Figure S4. Statistical analysis of mutation frequencies before and after implementing cART in Cameroon and comparison of RT, PR, and IN regions.
(A): Mutation frequencies before and after the regional implementation of cART were compared for each individual amino acid in RT, PR, and IN region of HIV-1 *pol* CRF02_AG sequences. Results of non-parametric Mann-Whitney tests are indicated, including the relative percentage of increasing, decreasing, and constant levels of mutations compared to the respective pre-cART consensus CRF02-AG sequence in Cameroon. (B): Summary of pre- and post-cART mutation differences and comparison between RT, PR, and IN regions using the absolute differences only (left) and including information about the relative change, i.e., increase or decrease (right).

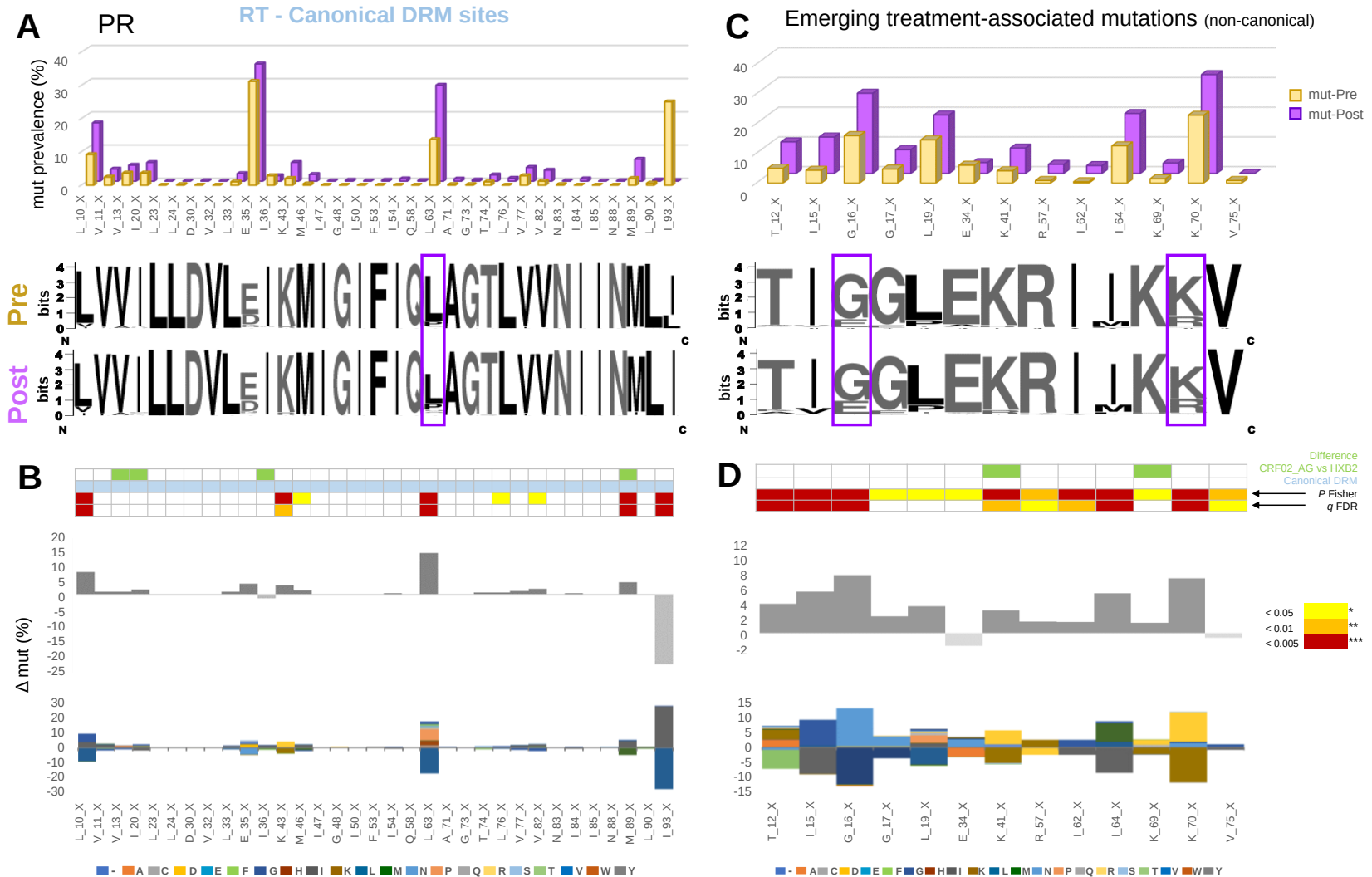


Figure S5. Canonical drug resistance mutations and emerging treatment-associated mutations in Cameroonian HIV-1 CRF02_AG PR sequences before and after the regional implementation of cART.

(A): Comparison of site-specific frequencies at canonical drug resistance mutation (mut) sites in Cameroonian HIV-1 *pol* PR before (yellow, Pre) and after (purple, Post) regional implementation of cART. The CRF02_AG consensus sequence (derived from pre-cART data) served as a reference to call mut variants per site. For each site,

the dominant (consensus) amino acid is indicated, followed by the position in RT. X indicates any mutation/minority variant. Below the bar chart, weblogs of amino acid occurrences per site are indicated from both Pre and Post datasets. Sites at which mutations increased more than 10% from pre- to post-cART period are boxed. **(B)**: Same selection of all canonical DRM sites in PR (as in **(A)**). On the y-axis, the difference in mut percentage (Δ mut) between Post and Pre is indicated for each site, with increasing mut frequencies from Pre to Post shown as positive values (dark gray bars) and decreasing frequencies shown as negative values (light gray bars). The mirror bar chart below indicates all amino acid (aa) changes according to the aa color code at the bottom. The 4-row color strips on top indicate differences between CRF02_AG consensus sequences and HXB2 (green), sites of canonical drug resistance mut (DRM) sites (blue), and statistically significant differences between Pre and Post in Fisher Exact tests (P values) and false discovery rates (FDR, q values), according to the legend to the right. **(C, D)**: Same analysis as in **(A, B)** for all emerging treatment-associated mutation sites, i.e., all RT sites other than the canonical DRM sites with a significant change in mutation frequencies according to $P < 0.05$ in Fisher Exact tests.

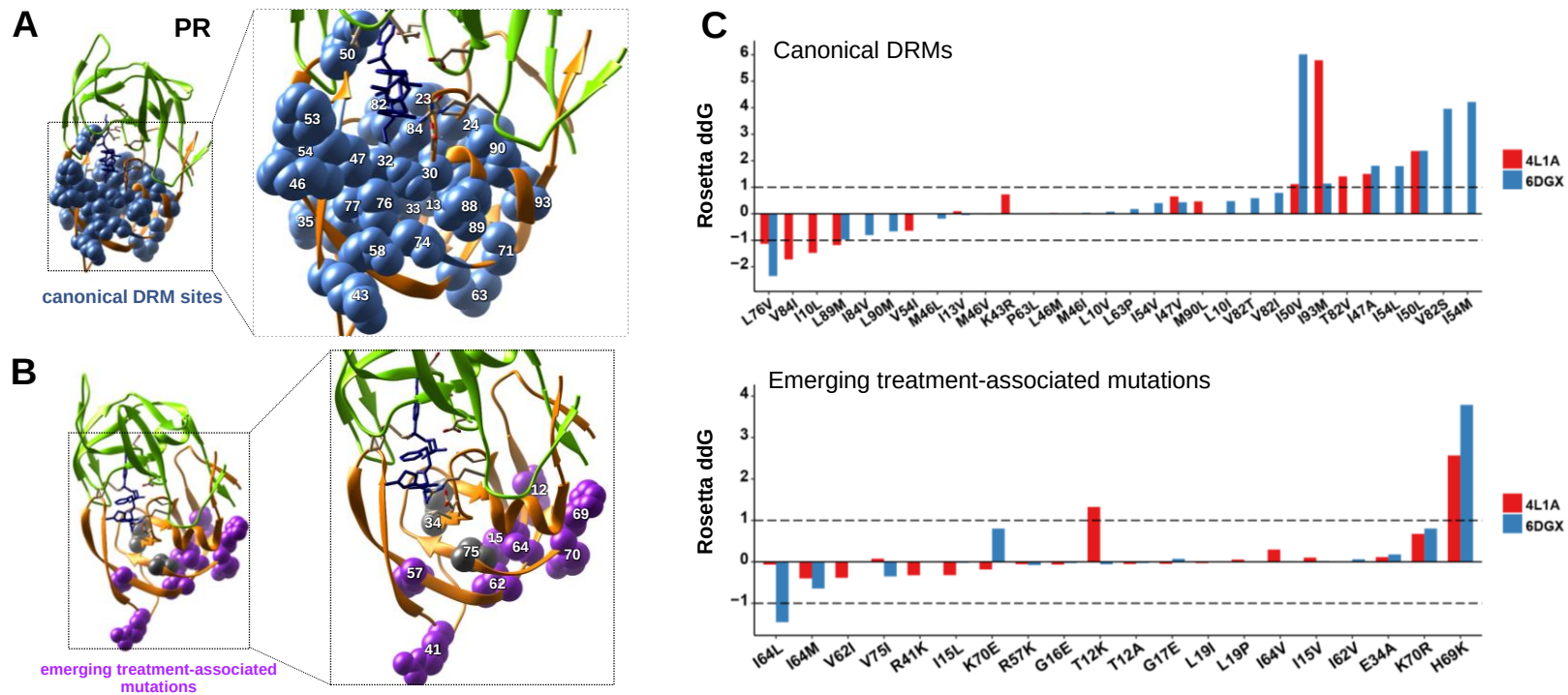


Figure S6. Structural analysis of emerging DRMs and treatment-associated mutations in CRF02_AG pol PR.

(A, B): Sites of significantly increasing canonical drug resistance mutations (DRMs) **(A)** and emerging treatment-associated mutations **(B)**, as identified in **Figure S5**, are projected onto a PR structure (PDB 4L1A). Detailed views of the drug-binding regions with annotated aa sites are shown in boxes to the right. In **(A)**, the DRM residues

are displayed as blue or gray spheres, according to a significant increase or decrease from pre- to post-cART periods, respectively ($P < 0.05$, according to **figure S5B**). Accordingly, in **(B)**, the treatment-associated mutation residues are displayed as magenta or gray spheres, according to a significant increase or decrease from pre- to post-cART, respectively ($P < 0.05$, according to **Figure S5D**). The structure is shown with bound antiretroviral drug (Lopinavir, dark blue) that demarcates the active center of the PR protein. **C**: The effect of prominent PR DRMs and all significantly emerging CRF02_AG treatment-associated mutations on two different published PR protein structures were analyzed with the Cartesian ddg application (Rosetta). ddG values >1 and <1 are characteristic for destabilizing and stabilizing mutations, respectively.

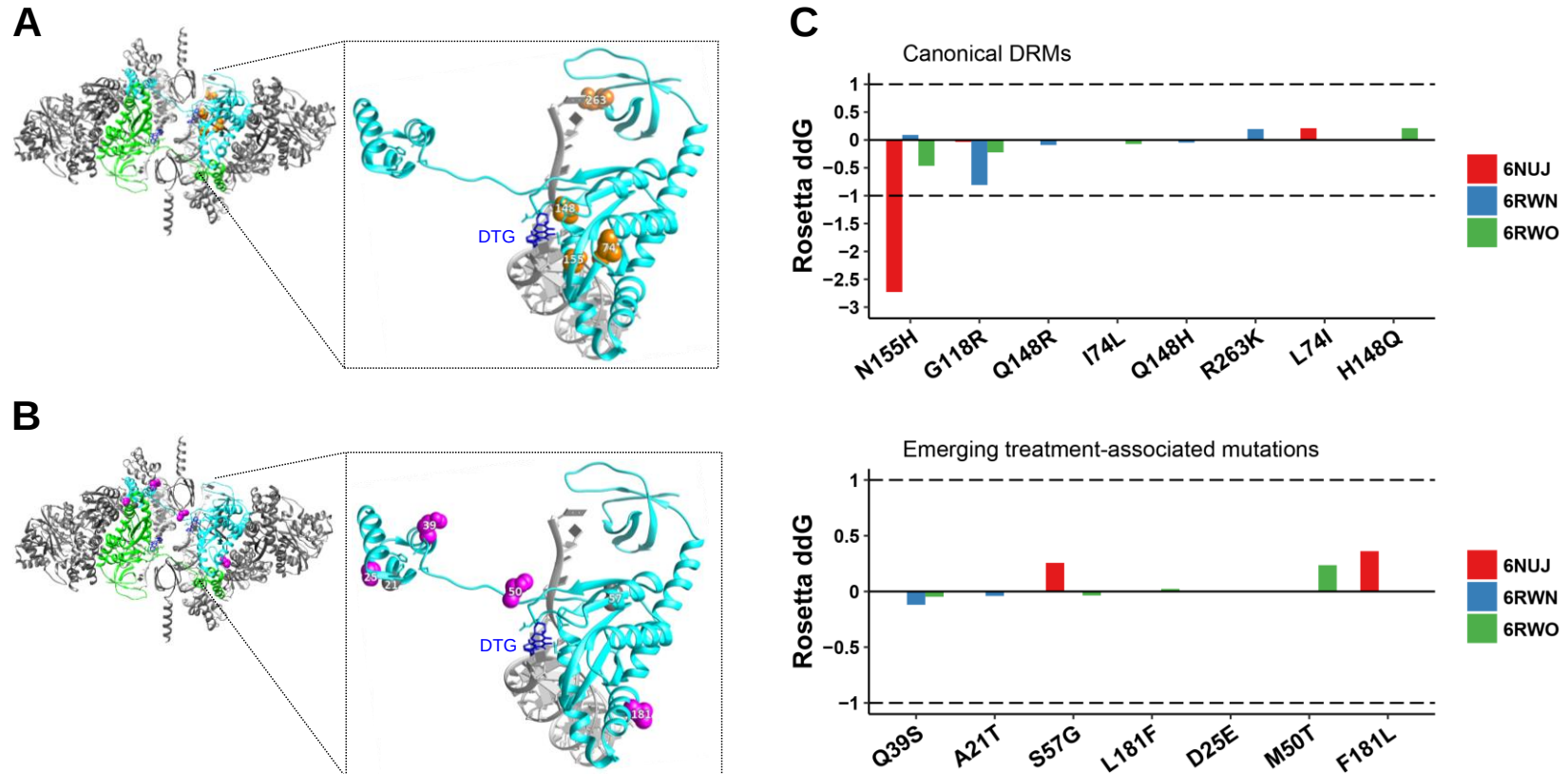


Figure S7. Structural analysis of emerging DRMs and treatment-associated mutations in CRF02_AG pol IN.

(A, B): Sites of known canonical drug resistance mutations (DRMs) **(A)** and emerging treatment-associated mutations **(B)**, the latter identified in **Figure 5**, are projected onto an IN structure (PDB 6RWN). Two IN monomers within the full integrasome-DNA complex are highlighted in green and cyan. Detailed views of an IN monomer with the drug-binding site and annotated aa sites shown in boxes to the right. Dolutegravir (DTG) is shown in dark blue indicating the active center of the IN protein. DRM residues are shown in orange, and treatment-associated mutation residues in magenta or gray, according to a significant increase or decrease from pre- to post-

cART, respectively ($P < 0.05$ according to **Figure 5**). (C): The effect of prominent IN canonical DRMs and all significantly emerging CRF02_AG treatment-associated mutations on three different published IN protein structures were analyzed with the Cartesian ddg application (Rosetta). ddG values >1 and <1 are characteristic for destabilizing and stabilizing mutations, respectively.

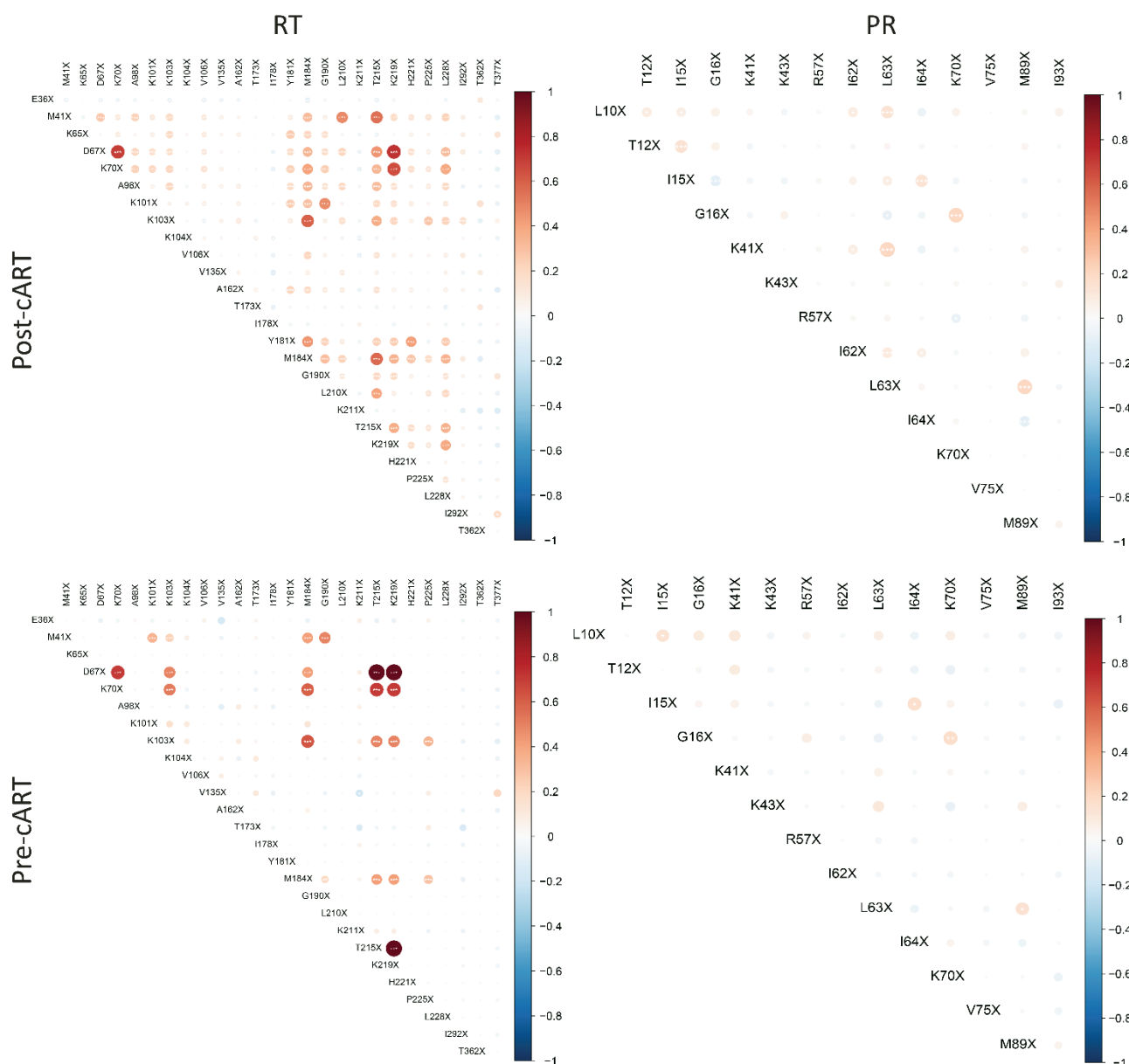


Figure S8. Pairwise correlations between emerging mutations in CRF02_AGRT and PR.

Correlograms of pairwise correlations between indicated mutations in HIV-1 reverse transcriptase (RT) or protease (PR) sequences, collected before (pre-cART) or after (post-cART) implementation of combinational antiretroviral treatment in Cameroon (2004). Non-parametric spearman rank tests were performed. Circles are sized and colored according to the correlation coefficient (r), and asterisks within the circles indicate significance levels after multiplicity adjustment with the Benjamini-Hochberg method. * $P<0.05$, ** $P<0.01$, and *** $P<0.005$.

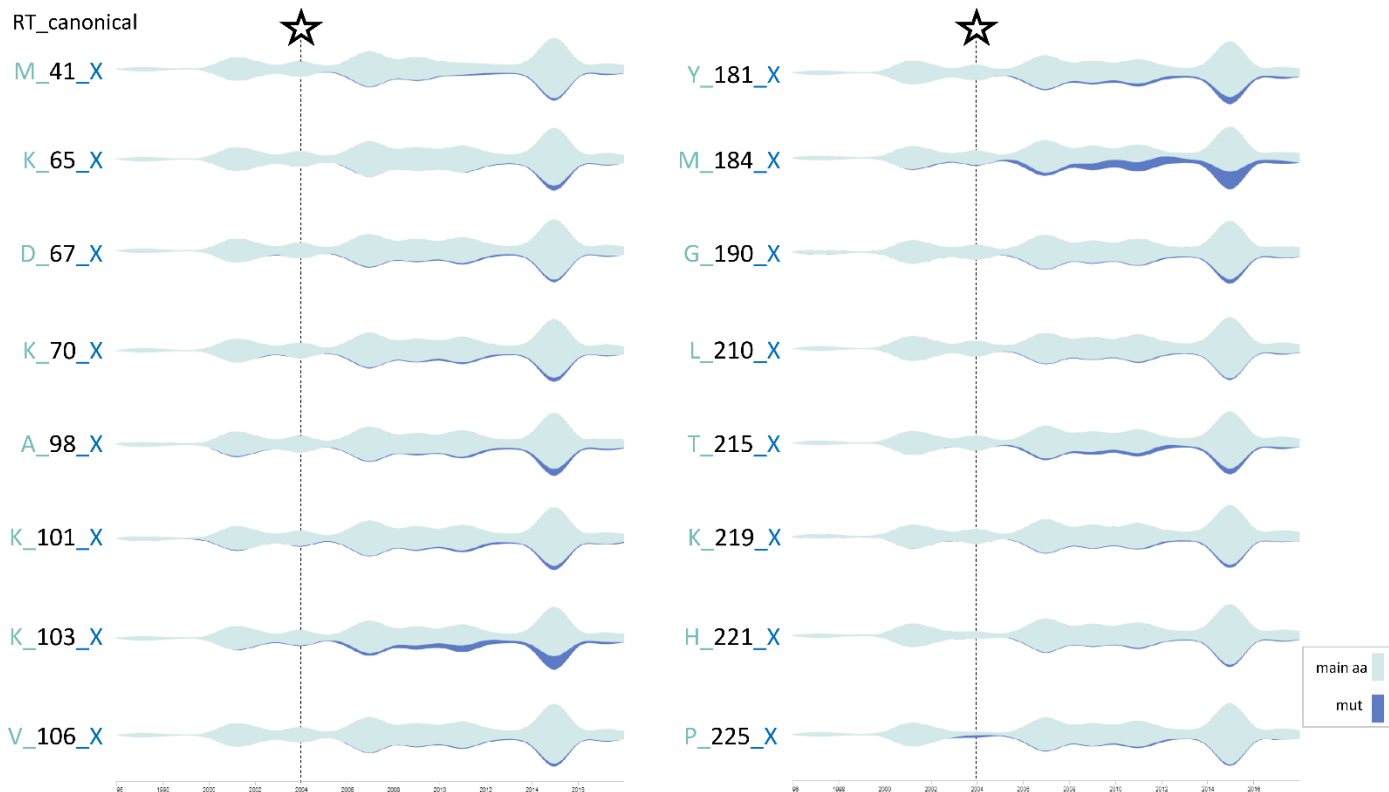


Figure S9. Time series analysis of canonical drug resistance mutation frequencies in HIV-1 CRF02_AGpol RT. Streamgraphs in silhouette mode display the amount of study sequences on the y-axis along the timeline on the x-axis. The gray-green color indicates the absence of mutations and presence of the dominant aa residue pre-cART, as listed to the left of each streamgraph (first position). The number indicates the aa position in RT; X indicates any mutation/minority variant. Blue color indicates mutation/minority frequencies, according to significant increase from pre- to post-cART ($q < 0.05$; see **Figure 4**). The star and the dashed line mark the time point of cART implementation in Cameroon.

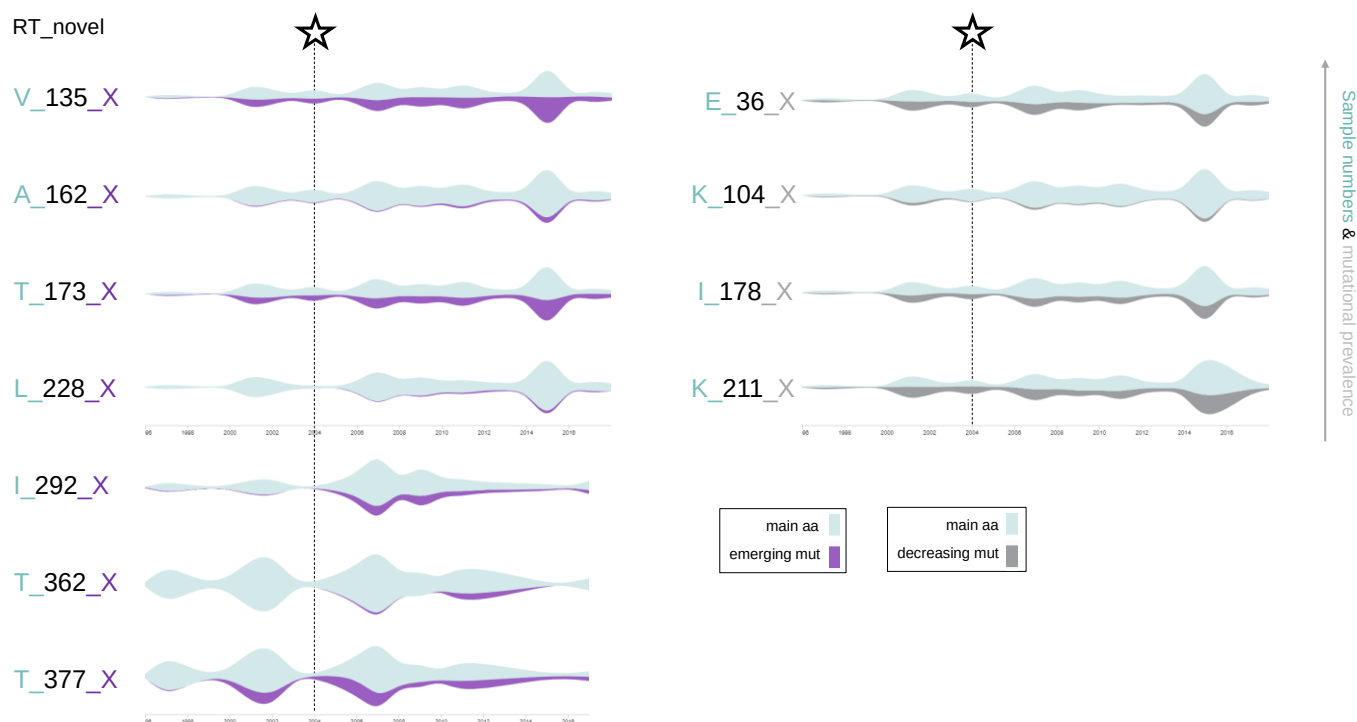


Figure S10. Time series analysis of treatment-associated mutation frequencies in HIV-1 CRF02_AG pol RT.

Streamgraphs in silhouette mode display the amount of study sequences on the y-axis along the timeline on the x-axis. The gray-green color indicates the absence of mutations and presence of the dominant aa residue, as listed to the left of each streamgraph (first position). The number indicates the aa position in RT. X indicates any mutation/minority variant. Purple and gray colors indicate mutation/minority frequencies, according to a significant increase or decrease from pre- to post-cART, respectively ($q < 0.05$; see **Figure 5**). The star and the dashed line mark the time point of cART implementation in Cameroon.

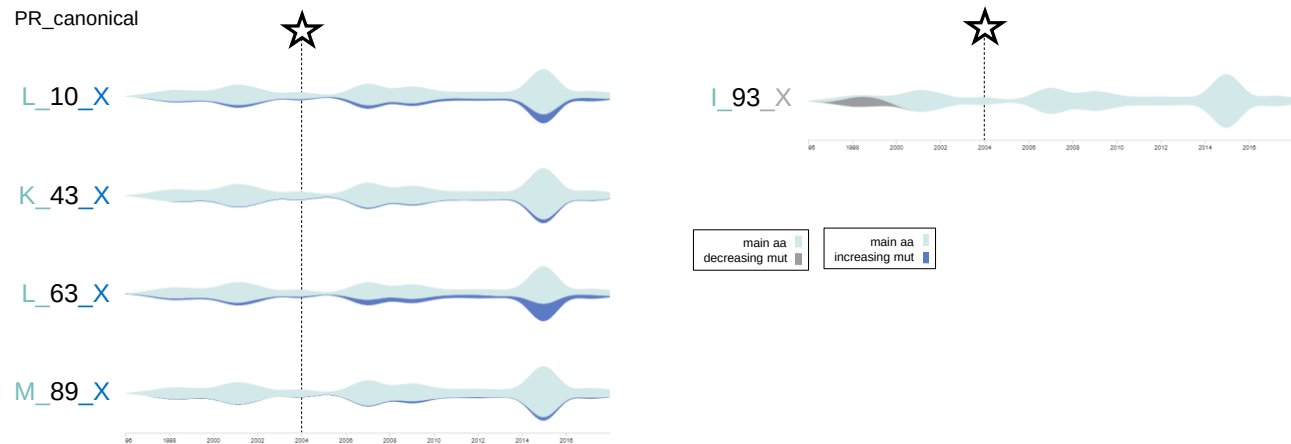


Figure S11. Time series analysis of canonical drug resistance mutation frequencies in HIV-1 CRF02_AG *pol* PR. Streamgraphs in silhouette mode display the amount of study sequences on the y-axis along the timeline on the x-axis. The gray-green color indicates the absence of mutations and presence of the dominant aa residue pre-cART, as listed to the left of each streamgraph (first position). The number indicates the aa position in PR; X indicates any mutation/minority variant. Blue (left) and gray (right) colors indicate mutation/minority frequencies, according to significant increase or decrease from pre- to post-cART, respectively ($q < 0.05$; see **Figure S5**). The star and the dashed line mark the time point of cART implementation in Cameroon.

PR_novel

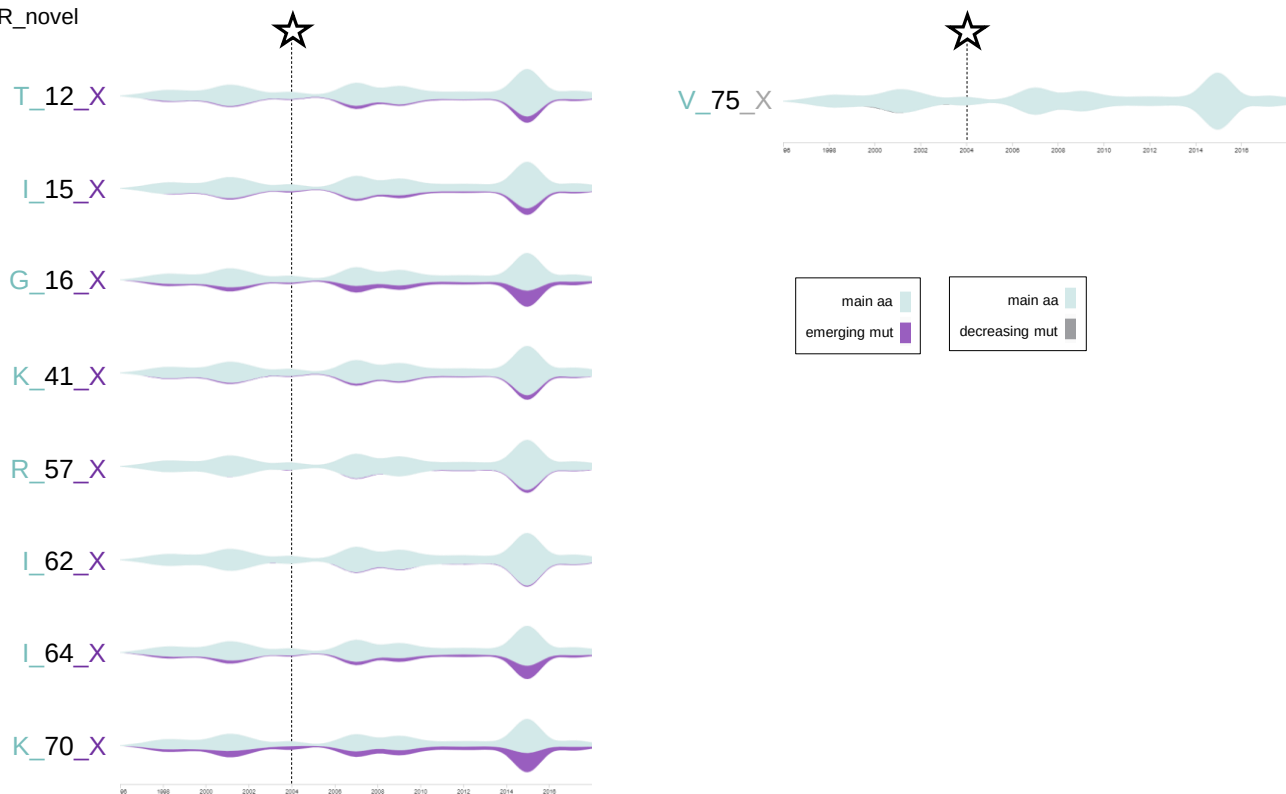


Figure S12. Time series analysis of treatment-associated mutation prevalence in HIV-1 CRF02_AG pol PR.

Streamgraphs in silhouette mode display the amount of study sequences on the y-axis along the timeline on the x-axis. The gray-green color indicates the absence of mutations and presence of the dominant aa residue, as listed to the left of each streamgraph (first position). The number indicates the aa position in PR. X indicates any mutation/minority variant. Purple (left) and gray (right) colors indicate mutation/minority frequencies, according to significant increase or decrease from pre- to post-cART, respectively ($q < 0.05$; see **Figure S5**). The star and the dashed line mark the time point of cART implementation in Cameroon.

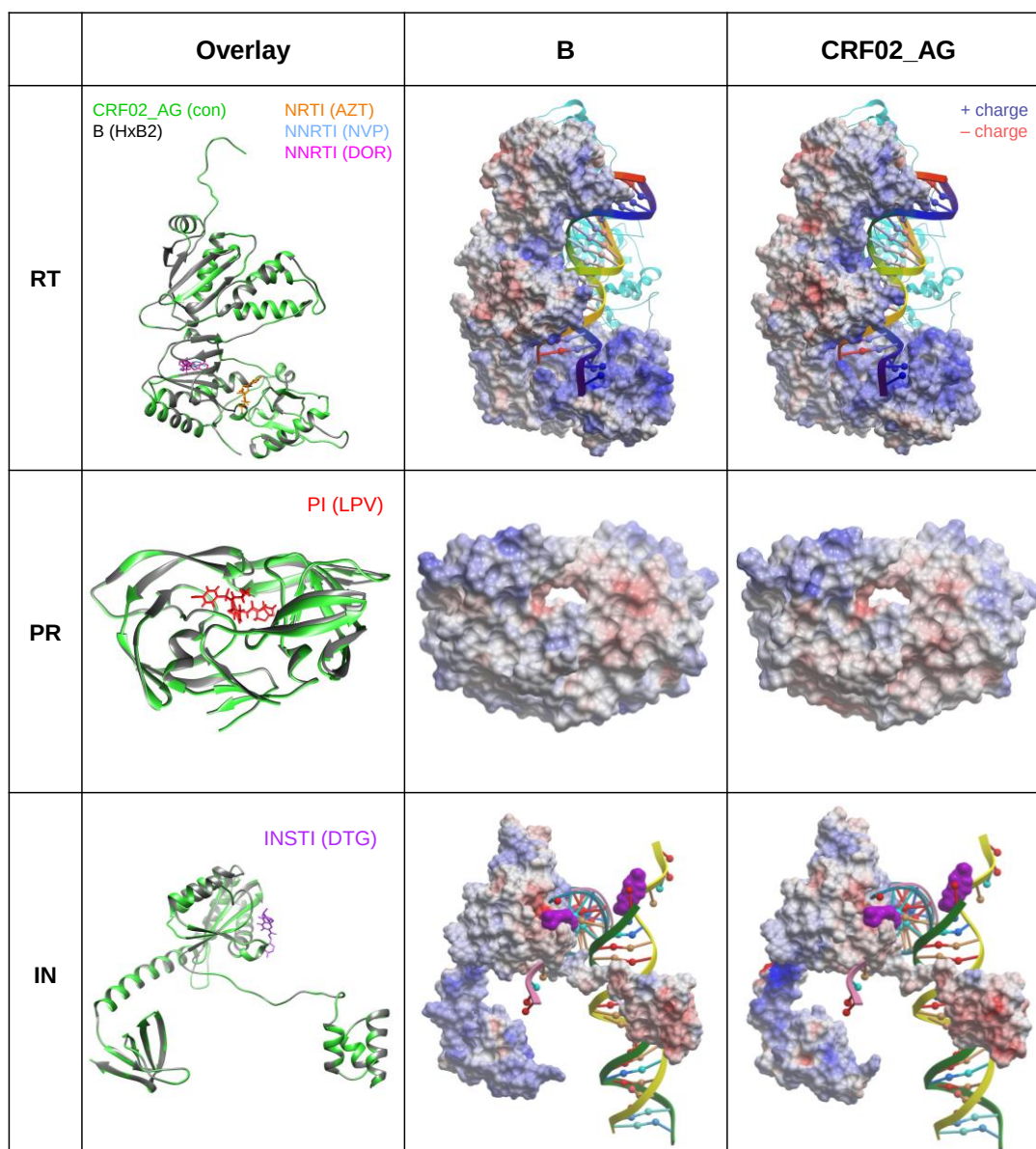


Figure S13. Structural comparison of HIV-1 RT, PR, and IN in the context of clades B and CRF02_AG.

Left: Structural overlays of subtype B HxB2 (green; accession number K03455) and CRF02_AG consensus (gray; derived from pre-cART data sets) RT, PR, and IN structures. Structural simulations were done on the SWISS-MODEL server using pdb 3v4i (RT), 4l1a (PR), and 6rwn (IN) template structures. Structural overlays were performed using Matchmaker in Chimera. Monomer RT, dimer PR, and monomer IN structures are shown. Middle and right: Surface charges were determined using ICM-Pro and displayed in blue and red gradients according to positive and negative charges, respectively. RT and IN proteins are complexed with DNA. AZT: azidothymidine, con: consensus, DOR: doravirine, DTG: dolutegravir, INSTI: integrase strand transfer inhibitor, LPV: lopinavir, NVP: nevirapine, NRTI: nucleoside/nucleotide reverse transcriptase inhibitor, NNRTI: Non-NRTI, PR: protease, PI: protease inhibitor, RT: reverse transcriptase

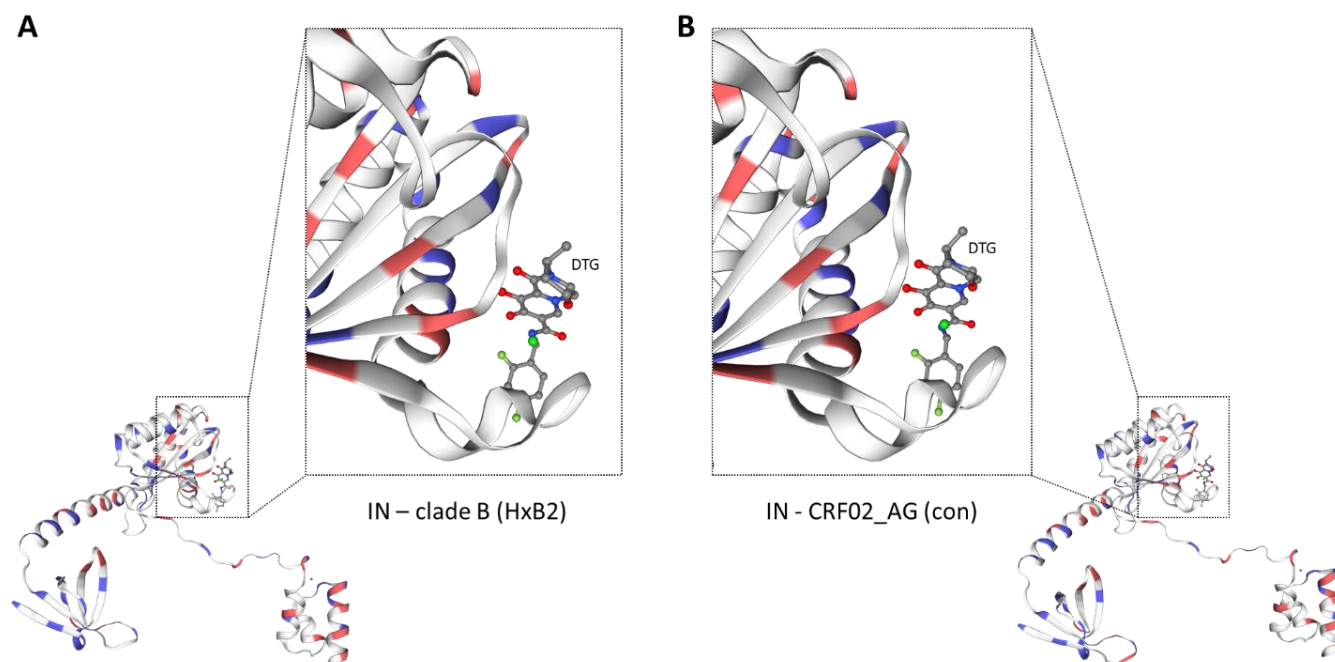


Figure S14. Structural comparison of integrase drug binding sites in clades B and CRF02_AG.

Structure and charge simulation of integrase drug binding sites in clades B (HxB2; accession number K03455) and CRF02_AG (consensus sequence from pre-cART data set). Structural simulations of integrase monomers were done using the SWISS-MODEL server against the pdb 6rwn template IN structure with bound dolutegravir (DTG). Blue and red parts of the ribbon protein structures indicate positively and negatively charged amino acid stretches, respectively. DTG is shown in ball and stick representation.

(A): Canonical drug resistance mutation (DRM) statistics. **(B):** Treatment-associated mutation statistics. Abbreviations and explanations: Aa: amino acid; CM: Cameroon; con: consensus; mut: mutation prevalence based on CRF02_AG con as reference; n: case number/coverage per site; pre-cART/post-cART: pre/post implementation of combinational antiretroviral therapy in Cameroon; Ref: reference amino acid positions and residues; statistical interpretation: * <0.05, ** <0.01, *** <0.005, empty cells: non-significant.

Mutations refer to HxB2 reference sequence.

RT	Mutations		RT with Zidovudine (3v4i)				RT with Nevirapine (3v81)				RT with Doravirine (4ncg)			
type	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect
canonical DRM	K65R	K65R	-2633.81	-2633.52	0.28	neutral	-2499.26	-2498.63	0.63	neutral	-2690.24	-2689.63	0.60	neutral
	K103N	K103N	-2635.77	-2636.64	-0.87	neutral	-2507.05	-2503.77	3.28	destabilizing	-2695.19	-2693.73	1.46	destabilizing
	Y181C	Y181C	-2640.46	-2633.38	7.08	destabilizing	-2507.74	-2503.80	3.95	destabilizing	-2696.30	-2694.11	2.19	destabilizing
	M184V	M184V	-2636.01	-2637.84	-1.83	stabilizing	-2500.39	-2495.98	4.41	destabilizing	-2692.46	-2691.28	1.18	destabilizing
	T215F	T215F	-2634.12	-2636.40	-2.28	stabilizing	-2500.10	-2502.53	-2.43	stabilizing	-2693.67	-2694.88	-1.22	stabilizing
	T215Y	T215Y	-2634.12	-2636.69	-2.57	stabilizing	-2500.10	-2502.81	-2.71	stabilizing	-2693.70	-2694.74	-1.03	stabilizing
type	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect
emerging TAM	P4H	P4H	-2634.65	-2634.62	0.03	neutral	-2501.40	-2501.84	-0.43	neutral	-2690.26	-2688.72	1.54	destabilizing
	P4Q	P4Q	-2634.66	-2635.11	-0.45	neutral	-2501.40	-2502.63	-1.23	stabilizing	-2690.26	-2690.36	-0.10	neutral
	P4S	P4S	-2634.67	-2632.40	2.27	destabilizing	-2501.39	-2501.11	0.29	neutral	-2690.26	-2688.42	1.84	destabilizing
	P4T	P4T	-2634.58	-2633.12	1.46	destabilizing	-2501.39	-2502.28	-0.88	neutral	-2690.26	-2689.99	0.27	neutral
	E6A	E6A	-2630.36	-2628.52	1.83	destabilizing	-2499.55	-2498.14	1.41	destabilizing	-2688.88	-2689.29	-0.41	neutral
	E6D	E6D	-2630.36	-2630.84	-0.48	neutral	-2499.55	-2500.77	-1.22	stabilizing	-2689.23	-2688.50	0.73	neutral
	E6K	E6K	-2630.36	-2627.74	2.62	destabilizing	-2499.55	-2497.32	2.22	destabilizing	-2689.29	-2688.01	1.28	destabilizing
	E6N	E6N	-2630.36	-2627.76	2.60	destabilizing	-2499.55	-2497.52	2.03	destabilizing	-2689.29	-2688.36	0.92	neutral
	K30I	K30I	-2634.12	-2636.21	-2.09	stabilizing	-2501.23	-2508.80	-7.57	stabilizing	-2689.22	-2691.18	-1.96	stabilizing
	K30N	K30N	-2633.91	-2632.22	1.69	destabilizing	-2501.24	-2501.07	0.17	neutral	-2689.20	-2690.30	-1.10	stabilizing
	K30Q	K30Q	-2633.78	-2636.14	-2.36	stabilizing	-2501.23	-2502.27	-1.03	stabilizing	-2689.20	-2696.92	-7.71	stabilizing
	K30R	K30R	-2634.21	-2635.44	-1.23	stabilizing	-2501.23	-2503.07	-1.83	stabilizing	-2689.43	-2689.95	-0.52	neutral
	I31L	I31L	-2633.42	-2627.32	6.10	destabilizing	-2501.34	-2491.94	9.41	destabilizing	-2691.19	-2686.61	4.58	destabilizing
	I31R	I31R	-2633.33	-2623.21	10.12	destabilizing	-2501.34	-2496.62	4.72	destabilizing	-2691.19	-2683.96	7.23	destabilizing
	I31T	I31T	-2633.50	-2625.70	7.80	destabilizing	-2501.34	-2493.87	7.47	destabilizing	-2691.19	-2683.61	7.58	destabilizing
	I31V	I31V	-2633.59	-2631.75	1.84	destabilizing	-2501.34	-2499.35	1.99	destabilizing	-2691.18	-2689.65	1.53	destabilizing
	E36D	E36D	-2632.94	-2632.74	0.21	neutral	-2500.91	-2500.92	-0.01	neutral	-2690.62	-2691.42	-0.80	neutral
	S48E	S48E	-2636.86	-2635.25	1.61	destabilizing	-2507.28	-2504.97	2.31	destabilizing	-2689.73	-2690.79	-1.06	stabilizing
	S48T	S48T	-2636.87	-2639.35	-2.48	stabilizing	-2507.35	-2509.22	-1.87	stabilizing	-2689.71	-2692.51	-2.79	stabilizing
	V60I	V60I	-2637.94	-2632.01	5.94	destabilizing	-2501.91	-2501.28	0.63	neutral	-2692.04	-2688.08	3.97	destabilizing
	K104R	K104R	-2634.37	-2635.30	-0.93	neutral	-2506.38	-2506.89	-0.51	neutral	-2693.99	-2693.57	0.42	neutral
	V135I	I135V	-4.21	-3.98	0.23	neutral	-34.34	-34.34	0.00	neutral	-351.95	-351.85	0.10	neutral
	A162C	S162A	-3.99	-3.90	0.09	neutral	-34.35	-34.08	0.27	neutral	-350.75	-351.16	-0.41	neutral
	A162D	S162A	-3.98	-3.90	0.08	neutral	-34.34	-34.19	0.15	neutral	-350.72	-351.16	-0.44	neutral
	K166R	K166R	-2634.80	-2634.49	0.31	neutral	-2500.80	-2499.86	0.94	neutral	-2692.26	-2695.64	-3.39	stabilizing
	T173A	K173T	-4.16	-4.16	0.00	neutral	-34.06	-34.11	-0.05	neutral	-352.84	-352.86	-0.02	neutral
	T173I	K173T	-4.16	-4.16	0.00	neutral	-34.07	-34.11	-0.04	neutral	-352.84	-352.86	-0.02	neutral
	I178L	I178L	-2635.26	-2631.10	4.15	destabilizing	-2504.59	-2500.75	3.84	destabilizing	-2693.84	-2690.13	3.70	destabilizing
	I178M	I178M	-2635.31	-2635.32	0.00	neutral	-2504.59	-2504.68	-0.09	neutral	-2693.84	-2693.62	0.21	neutral
	E203D	E203D	-2632.30	-2629.50	2.80	destabilizing	-2501.94	-2500.15	1.78	destabilizing	-2691.81	-2689.41	2.40	destabilizing
	E203K	E203K	-2632.30	-2630.05	2.25	destabilizing	-2501.94	-2498.94	3.00	destabilizing	-2691.81	-2689.05	2.76	destabilizing
	E203Q	E203Q	-2632.30	-2631.03	1.27	destabilizing	-2501.94	-2501.36	0.58	neutral	-2691.81	-2690.59	1.22	destabilizing
	K211R	R211K	-4.03	-3.89	0.14	neutral	-34.33	-34.34	-0.01	neutral	-352.12	-352.17	-0.05	neutral
	F214L	L214F	-3.95	-3.93	0.02	neutral	-34.33	-34.49	-0.15	neutral	-351.38	-350.61	0.77	neutral
	H221Y	H221Y	-3.87	-4.42	-0.56	neutral	-32.25	-32.36	-0.12	neutral	-353.07	-353.05	0.02	neutral
	E224D	E224D	-2631.77	-2631.32	0.44	neutral	-2499.06	-2498.68	0.38	neutral	-2691.09	-2690.84	0.26	neutral
	E224G	E224G	-2631.77	-2630.31	1.46	destabilizing	-2499.06	-2497.52	1.54	destabilizing	-2690.86	-2691.45	-0.59	neutral
	L228H	L228H	-2634.35	-2632.38	1.98	destabilizing	-2500.60	-2497.77	2.83	destabilizing	-2690.25	-2688.06	2.19	destabilizing
	L228R	L228R	-2634.40	-2632.35	2.05	destabilizing	-2500.60	-2497.78	2.82	destabilizing	-2689.91	-2688.17	1.74	destabilizing
	W229K	W229K	-2640.17	-2622.60	17.57	destabilizing	-2507.49	-2487.68	19.81	destabilizing	-2694.77	-2678.93	15.84	destabilizing
	W229R	W229R	-2640.18	-2626.58	13.59	destabilizing	-2507.52	-2491.36	16.16	destabilizing	-2694.69	-2680.86	13.83	destabilizing
	Q245E	V245Q	-4.12	-4.14	-0.02	neutral	-34.34	-34.33	0.01	neutral	-353.98	-353.97	0.00	neutral
	Q245K	V245Q	-4.12	-4.14	-0.02	neutral	-34.34	-34.33	0.01	neutral	-353.97	-353.98	-0.01	neutral
	A288T	A288T	-2629.95	-2630.24	-0.29	neutral	-2498.94	-2499.61	-0.67	neutral	-2689.59	-2689.42	0.18	neutral
	D291E	E291D	-4.20	-4.20	0.00	neutral	-34.34	-34.34	0.00	neutral	-353.97	-353.97	0.00	neutral
	V292I	V292I	-4.13	-4.13	0.01	neutral	-34.34	-34.34	-0.01	neutral	-353.97	-353.66	0.31	neutral
	V326I	I326V	-4.17	-4.15	0.02	neutral	-34.34	-34.34	0.00	neutral	-353.97	-353.97	0.00	neutral
	V326M	I326V	-4.17	-4.15	0.02	neutral	-34.34	-34.34	0.01	neutral	-353.98	-353.98	0.00	neutral
	D335G	G335D	-4.30	-4.30	0.00	neutral	-34.10	-34.35	-0.25	neutral	-353.95	-355.06	-1.11	stabilizing
	T362N	T362N	-2632.83	-2630.05	2.78	destabilizing	-2503.07	-2497.98	5.09	destabilizing	-2694.08	-2688.85	5.24	destabilizing
	T362P	T362P	-2632.84	-2625.54	7.29	destabilizing	-2503.00	-2495.19	7.81	destabilizing	-2694.08	-2687.59	6.49	destabilizing
	T362S	T362S	-2632.77	-2629.09	3.69	destabilizing	-2503.02	-2498.94	4.07	destabilizing	-2694.08	-2689.94	4.15	destabilizing
	A376T	T376A	-4.13	-4.14	0.00	neutral	-34.34	-34.34	-0.01	neutral	-354.04	-353.99	0.05	neutral
	A376S	T376A	-4.13	-4.14	0.00	neutral	-34.34	-34.34	-0.01	neutral	-353.97	-353.96	0.01	neutral
	A376V	T376A	-4.13	-4.14	0.00	neutral	-34.34	-34.34	0.00	neutral	-353.94	-353.97	-0.04	neutral
	T377I	T377I	-2632.04	-2633.26	-1.23	stabilizing	-2503.59	-2504.90	-1.31	stabilizing	-2692.20	-2693.60	-1.40	stabilizing
	T377L	T377L	-2632.04	-2632.49	-0.45	neutral	-2503.68	-2504.67	-0.99	neutral	-2692.16	-2692.94	-0.78	neutral
	T377M	T377M	-2632.04	-2632.12	-0.08	neutral	-2504.07	-2503.97	0.09	neutral	-2692.24	-2692.54	-0.30	neutral
	T377R	T377R	-2632.04	-2630.66	1.37	destabilizing	-2504.19	-2502.93	1.26	destabilizing	-2692.11	-2691.11	1.01	destabilizing
	T377V	T377V	-2632.04	-2633.43	-1.40	stabilizing	-2504.14	-2505.40	-1.26	stabilizing	-2692.11	-2693.84	-1.73	stabilizing
	K395R	K395R	-2631.38	-2630.65	0.73	neutral	-2501.71	-2502.73	-1.03	stabilizing	-2694.53	-2694.86	-0.33	neutral

Table S2. Impact of HIV-1 CRF02_AG RT mutations on protein stability.

Heatmap of $\Delta\Delta G$ values of canonical drug resistance mutations (DRMs) and emerging treatment-associated mutations, as observed in Cameroonian CRF02_AG RT datasets. $\Delta\Delta G$ values were determined using three different co-crystal structures of RT with inhibitors. Stabilizing and destabilizing mutations are highlighted in blue and red according to $\Delta\Delta G < 1$ and > 1 , respectively. Yellow highlighted cells indicate adjusted calculations due to amino acid divergence between CRF02_AG and studied crystal structures.

PR		PR with Lopinavir (411a)						PR with Darunavir (6dgx)					
type	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect		MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect
canonical DRM	L10I	I10L	-209.26	-210.73	-1.48	stabilizing		L10I	L10I	-227.83	-227.35	0.48	neutral
	L10V	I10L	-209.32	-210.76	-1.44	stabilizing		L10V	L10V	-227.84	-227.76	0.08	neutral
	V13I	I13V	-211.51	-211.41	0.10	neutral		V13I	I13V	-232.77	-232.82	-0.05	neutral
	V13A	I13V	-211.50	-211.50	0.00	neutral		V13A	I13V	-232.78	-232.68	0.10	neutral
	K43R	K43R	-219.26	-218.53	0.73	neutral		K43R	K43R	-241.14	-241.15	-0.01	neutral
	M46L	L46M	-216.84	-216.81	0.03	neutral		M46I	M46I	-239.68	-239.64	0.04	neutral
	M46I	L46M	-216.86	-216.81	0.05	neutral		M46L	M46L	-239.68	-239.86	-0.18	neutral
	M46V	L46M	-216.87	-216.81	0.06	neutral		M46V	M46V	-239.68	-239.70	-0.02	neutral
	I47A	I47A	-212.81	-211.30	1.50	destabilizing		I47A	I47A	-231.85	-230.04	1.81	destabilizing
	I47V	I47V	-212.95	-212.29	0.65	neutral		I47V	I47V	-231.89	-231.46	0.43	neutral
	I50L	I50L	-214.54	-212.17	2.37	destabilizing		I50L	I50L	-229.39	-227.01	2.38	destabilizing
	I50V	I50V	-214.55	-213.43	1.12	destabilizing		I50V	I50V	-229.46	-223.45	6.02	destabilizing
	I54V	V54I	-217.05	-217.69	-0.64	neutral		I54L	I54L	-240.24	-238.45	1.80	destabilizing
	I54L	V54I	-217.06	-217.69	-0.63	neutral		I54M	I54M	-240.25	-236.03	4.22	destabilizing
	I54M	V54I	-217.05	-217.68	-0.63	neutral		I54V	I54V	-240.24	-239.84	0.40	neutral
	L63P	P63L	-221.51	-221.51	0.00	neutral		L63P	L63P	-246.01	-245.83	0.18	neutral
	L76V	L76V	-215.29	-216.42	-1.13	stabilizing		L76V	L76V	-234.43	-236.78	-2.35	stabilizing
	V82T	T82V	-211.36	-209.95	1.41	destabilizing		V82I	V82I	-228.22	-227.43	0.79	neutral
	V82V	T82V	-211.31	-209.93	1.38	destabilizing		V82S	V82S	-228.77	-224.81	3.96	destabilizing
	V82S	T82V	-211.20	-210.19	1.01	destabilizing		V82T	V82T	-228.25	-227.66	0.59	neutral
	I84V	V84I	-210.14	-211.87	-1.72	stabilizing		I84V	V84I	-228.82	-229.62	-0.80	neutral
emerging TAM	M89I	L89M	-208.85	-210.03	-1.18	stabilizing		M89I	L89M	-227.85	-228.82	-0.97	neutral
	L90M	M90L	-204.73	-204.27	0.47	neutral		L90M	L90M	-225.07	-225.73	-0.66	neutral
	I93M	I93M	-205.79	-200.00	5.79	destabilizing		I93M	I93M	-229.38	-228.24	1.14	destabilizing
	T12A	T12A	-214.00	-214.05	-0.05	neutral		T12A	T12A	-237.48	-237.51	-0.03	neutral
	T12K	T12K	-214.30	-212.97	1.32	destabilizing		T12K	T12K	-237.37	-237.43	-0.06	neutral
	I15L	I15L	-218.29	-218.61	-0.32	neutral		I15L	I15L	-241.05	-241.07	-0.02	neutral
	I15V	I15V	-218.19	-218.09	0.10	neutral		I15V	I15V	-241.06	-241.04	0.02	neutral
	G16E	G16E	-222.65	-222.71	-0.07	neutral		G16E	G16E	-246.67	-246.70	-0.03	neutral
	G17E	G17E	-222.59	-222.64	-0.05	neutral		G17E	G17E	-246.66	-246.58	0.07	neutral
	L19I	L19I	-222.47	-222.50	-0.03	neutral		L19I	L19I	-246.61	-246.61	0.00	neutral
	L19P	L19P	-222.44	-222.39	0.05	neutral		L19P	L19P	-246.64	-246.64	0.00	neutral
	E34A	E34A	-217.07	-216.95	0.11	neutral		E34A	E34A	-239.24	-239.06	0.18	neutral
	K41R	R41K	-222.64	-222.96	-0.32	neutral		K41R	R41K	-246.67	-246.66	0.01	neutral
	K41N	R41K	-222.54	-222.64	-0.10	neutral		K41N	R41K	-246.67	-246.67	0.00	neutral
	R57K	R57K	-219.04	-219.10	-0.06	neutral		R57K	R57K	-240.49	-240.56	-0.08	neutral
	I62V	V62I	-217.30	-217.68	-0.38	neutral		I62V	I62V	-238.14	-238.08	0.06	neutral
	I64L	I64L	-215.06	-215.13	-0.07	neutral		I64L	I64L	-236.39	-237.85	-1.46	stabilizing
	I64M	I64M	-214.83	-215.23	-0.40	neutral		I64M	I64M	-236.40	-237.05	-0.64	neutral
	I64V	I64V	-215.04	-214.74	0.30	neutral		I64V	I64V	-236.39	-236.38	0.01	neutral
	K69Q	H69K	-208.45	-205.89	2.57	destabilizing		K69Q	H69K	-231.03	-227.24	3.79	destabilizing
	K69R	H69K	-208.44	-205.88	2.56	destabilizing		K69R	H69K	-231.04	-227.08	3.96	destabilizing
	K70E	K70E	-219.05	-219.23	-0.18	neutral		K70E	K70E	-242.66	-241.86	0.80	neutral
	K70R	K70R	-219.72	-219.04	0.67	neutral		K70R	K70R	-242.64	-241.83	0.80	neutral
	V75I	V75I	-215.28	-215.20	0.08	neutral		V75I	V75I	-235.61	-235.96	-0.35	neutral

Table S3. Impact of HIV-1 CRF02_AG PR mutations on protein stability.

Heatmap of $\Delta\Delta G$ values of canonical drug resistance mutations (DRMs) and emerging treatment-associated mutations, as observed in Cameroonian CRF02_AG PR datasets. $\Delta\Delta G$ values were determined using two different co-crystal structures of PR with inhibitors. Stabilizing and destabilizing mutations are highlighted in blue and red according to $\Delta\Delta G < 1$ and > 1 , respectively. Yellow highlighted cells indicate adjusted calculations due to amino acid divergence between CRF02_AG and studied crystal structures.

IN	IN with Dolutegravir (6rwn)						IN with Bictegravir (6rwo)						IN with BI-224436 (6nuj)					
type	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect
canonical DRM	L74I	I74L	-13.86	-13.86	-0.01	neutral	L74I	I74L	-11.83	-11.91	-0.07	neutral	L74I	L74I	-131.02	-130.81	0.21	neutral
	G118R	G118R	-13.32	-14.12	-0.81	neutral	G118R	G118R	-11.41	-11.63	-0.22	neutral	G118R	G118R	-135.89	-135.92	-0.03	neutral
	Q148H	Q148H	-13.46	-13.51	-0.05	neutral	Q148H	H148Q	-11.44	-11.23	0.21	neutral	N155H	N155H	-130.71	-133.44	-2.73	stabilizing
	Q148R	Q148R	-13.46	-13.55	-0.09	neutral	Q148R		-11.43	-10.72	0.71	neutral						
	N155H	N155H	-13.92	-13.83	0.09	neutral	N155H	N155H	-11.77	-12.24	-0.46	neutral						
	R263K	R263K	-13.67	-13.48	0.20	neutral	R263K	R263K	-12.02	-12.01	0.02	neutral						
type	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect	MUT (CRF02_AG)	MUT (studied)	Mean_ΔG (WT)	Mean_ΔG (MUT)	ΔΔG (kcal/mol)	Effect
emerging TAM	A21T	A21T	-13.64	-13.68	-0.04	neutral	A21T	A21T	-11.99	-11.99	0.01	neutral	S57G	S57G	-130.37	-130.11	0.26	neutral
	D25E	D25E	-13.63	-13.64	0.00	neutral	D25E	D25E	-12.00	-12.00	0.01	neutral	F181L	F181L	-128.44	-128.08	0.36	neutral
	S39C	Q39S	-13.52	-13.64	-0.12	neutral	S39C	Q39S	-11.95	-12.00	-0.05	neutral						
	M50T	M50T	-13.33	-13.33	0.00	neutral	M50T	M50T	-11.72	-11.48	0.24	neutral						
	S57G	S57G	-13.65	-13.64	0.01	neutral	S57G	S57G	-11.96	-12.00	-0.04	neutral						
	F181L	L181F	-13.63	-13.64	-0.01	neutral	F181L	L181F	-12.00	-11.98	0.02	neutral						

Table S4. Impact of HIV-1 CRF02_AG IN mutations on protein stability

Heatmap of $\Delta\Delta G$ values of canonical drug resistance mutations (DRMs) and emerging treatment-associated mutations, as observed in Cameroonian CRF02_AG IN datasets. $\Delta\Delta G$ values were determined using three different co-crystal structures of IN with inhibitors. Stabilizing and destabilizing mutations are highlighted in blue and red according to $\Delta\Delta G < 1$ and > 1 , respectively. Yellow highlighted cells indicate adjusted calculations due to amino acid divergence between CRF02_AG and studied crystal structures.