

Université de Montréal

**Caractérisation moléculaire et phylogénétique de l'enveloppe du
VIH-1 transmis/fondateur**

Par

Alexis Kafando



Département de microbiologie, infectiologie et immunologie

Faculté de médecine

Thèse présentée

en vue de l'obtention du grade de **Philosophiae Doctor (Ph.D.)**

en Virologie et Immunologie

30 août 2019

© Alexis Kafando, 2019

Université de Montréal

Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

**Caractérisation moléculaire et phylogénétique de l'enveloppe du
VIH-1 transmis/fondateur**

Présentée par : **Alexis Kafando**

a été évaluée par un jury composé des personnes suivantes :

Dre Naglaa Shoukry, B. Pharm, Ph.D.
Présidente-rapporteuse

Dre Cécile Tremblay, MD., FRCPC.
Directrice de recherche

Dr. Hugues Charest, PhD.
Co-directeur de recherche

Dr. Éric A. Cohen, PhD.
Membre du jury

Dre Anne Gatignol, PhD.
Examinatrice externe

Dre Isabelle Boucoiran, MD., MSc.
Représentante de la doyenne

Résumé

Vaincre efficacement l'infection par le VIH nécessite non seulement de raffiner davantage les traitements antirétroviraux, mais aussi de déployer des moyens de prévention à large échelle afin de stopper la propagation de la maladie. Il s'agit en effet de pouvoir développer un vaccin préventif ou thérapeutique à large spectre. L'atteindre de cet objectif requiert une meilleure compréhension des événements précoces médiés par les glycoprotéines d'enveloppe des virus transmis/fondateurs du VIH développés durant la primo-infection. Pendant que les efforts de la communauté scientifique se concentrent à trouver ce vaccin, il est également nécessaire de limiter le taux des nouvelles infections dans la population. Les résultats de nos travaux présentés dans cette thèse s'inscrivent dans cette optique de prévention des infections à VIH par l'étude de la dynamique de transmission de la maladie.

Le premier volet de nos travaux dont les résultats sont consignés dans ce document et publiés dans le journal « **PLOS ONE** » (Chapitre III. Article 1) porte le titre : « **Identification des nouvelles infections par le VIH-1 en utilisant des mesures de la diversité génétique des séquences de l'enveloppe** ». L'objectif de cette étude était d'évaluer et de déterminer quelle méthode moléculaire parmi les mesures de diversité génétique de la séquence de l'enveloppe est capable de définir l'infection par le VIH-1 temporellement (récente versus chronique). Quatre mesures de diversité génétique des séquences virales de courts segments de l'enveloppe du VIH-1 ont été évaluées pour définir l'infection par le VIH-1 temporellement à savoir: le pourcentage de complexité, le pourcentage de diversité, le nombre d'haplotypes et l'entropie de Shannon. Nous avons identifié l'entropie de Shannon comme l'une des meilleures de ces quatre mesures de diversité qui associée à trois courts segments (≤ 100 paires de bases) de la séquence de l'enveloppe du VIH-1 est capable de prédire efficacement la période récente de l'infection par le VIH-1. L'indice de performance de l'entropie de Shannon (Aire sous la courbe (AUC) des trois segments de l'enveloppe du VIH-1 que sont : (1) région constante C2 segment 1 (C2_1), (2) région constante C2 segment 3 (C2_3) et (3) boucle V3 de la gp120 est respectivement de 0.806, 0.805 et 0.812. La capacité de

différencier une infection récente d'une infection ancienne est importante en santé publique. En effet, les personnes nouvellement infectées sont les plus infectieuses et capables de propager rapidement la maladie dans la population. Puisque ces personnes ignorent leur statut sérologique, elles peuvent avoir plus de comportements à risque alors qu'elles ont une charge virale sanguine élevée d'un virus avec une capacité répliquative assez importante (fitness). Pouvoir identifier le plus tôt possible ces individus permettrait de limiter et de prévenir les nouvelles transmissions.

Le deuxième volet de nos travaux consignés dans cette thèse (Chapitre III, Article 2) a été publié dans le journal « **AIDS Research and Human Retroviruses** » avec pour titre : « ***Analyses phylogénétiques des séquences du gène de l'enveloppe du VIH-1 pour l'évaluation et la construction à court terme des réseaux naissants de la transmission de l'infection par le VIH-1*** ». Cette étude a consisté à identifier les liens phylogénétiques ou clusters qui existent entre les séquences virales de l'enveloppe du VIH-1 de patients nouvellement diagnostiqués au Québec. La formation de clusters entre les séquences virales de plusieurs individus pourrait refléter l'appartenance de ces personnes à un même réseau ou chaîne de transmission. Ces personnes pourraient aussi partager des caractéristiques ou facteurs de risques communs au groupe. Les résultats de cette étude ont permis d'identifier 15 chaînes ou réseaux de transmission mineurs (2-5 individus par réseau) de l'infection par le VIH-1 entre ces individus. Nous avons également évalué et comparé la capacité d'un fragment d'une séquence nucléotidique plus large de l'enveloppe du VIH-1 à celui de la boucle V3 beaucoup plus courte à pouvoir prédire et identifier de façon indépendante les individus qui pourraient être inclus dans un réseau ou grappe de transmission. Les résultats de nos analyses ont démontré une concordance modérée en se basant sur le coefficient kappa ($k=0.59$) entre les deux fragments de l'enveloppe virale utilisés. L'utilisation des séquences de la boucle V3 a en effet permis de confirmer 66.70% (10/15) des clusters identifiés avec un fragment plus long de la séquence de l'enveloppe du VIH-1. Le nombre de personnes infectées ayant accès aux tests de génotypage qui utilisent les séquences de la V3 pour déterminer le tropisme viral en pratique clinique courante est limité à l'échelle mondiale. Cependant, les séquences de l'enveloppe virale et particulièrement celles de la

boucle V3 pourraient facilement être intégrées dans un programme de santé publique là où elles sont disponibles en complément de celles du gène *pol* (disponibles dans les laboratoires cliniques dans le cadre des tests de résistance aux traitements antirétroviraux) pour l'identification et la surveillance des réseaux de transmission du VIH-1 dans la population. Nous avons inclus les caractéristiques des sujets (données sociodémographiques, cliniques et facteurs de risque) dans un modèle d'analyse de régression logistique avec comme variable réponse (outcome) la formation de clusters de transmission. Les résultats de l'analyse démontrent que dans la population étudiée, l'âge moyen (<38.8 versus ≥38.8 ans) et le sous-type viral (sous type B versus non-B) sont deux facteurs significativement associés à la formation des clusters ou réseaux. L'identification des chaînes de transmission du VIH-1 dans un échantillon de la population infectée par le VIH-1 au Québec et les facteurs de risques communs à ces personnes devraient permettre de mieux implanter les stratégies de prévention dans les groupes cibles.

Les résultats de la troisième et dernière partie de nos travaux présentée dans cette thèse (Chapitre III, Article 3) ont été publiés dans le Journal « **VIRUSES** » sous le titre : « **Identification de signatures génétiques au niveau de la glycoprotéine de l'enveloppe du VIH-1 associées aux virus transmis/fondateurs et récents de sous-types B** ». Cette étude a consisté à identifier des signatures génétiques au niveau des séquences virales de l'enveloppe des premiers variants du VIH-1 appelés virus transmis/fondateurs ou « Transmitted/Founder viruses ». Ces variants constituent ceux qui sont capables d'établir une infection productive au moment de l'infection. Ils sont en effet sélectionnés parmi une multitude de quasi-espèces virales exposées au cours de la transmission du virus. Ils détiennent ainsi toutes les propriétés phénotypiques et ou fonctionnelles nécessaires à l'établissement de l'infection. L'identification de signatures génétiques observées précocement au niveau de la glycoprotéine de l'enveloppe du VIH-1 au cours de l'infection pourrait informer la conception de nouveaux inhibiteurs d'entrée immunologiques ou chimiques, mais aussi servir de repère pour le design d'un vaccin anti-VIH-1. Nos travaux ont permis d'identifier quelques signatures génétiques, mais une consistant en une mutation/substitution de l'Arginine (R) par une Isoleucine (I) associée au virus TF semble être

particulièrement prononcée. Cette mutation est localisée à la position 841 (R841I) spécifiquement dans le domaine cytoplasmique de la GP41 notamment dans le segment 1 des peptides lytiques de lentivirus (LLP-1). L'isoleucine est sélectionnée (I) à plus de 33% par les virus TF comparativement aux virus d'infections chroniques (9%) et la différence est statistiquement significative, OR=0.2, 95% IC [0.09, 0.44], P= 0.00001. Le domaine cytoplasmique de la GP41 et spécifiquement le LLP-1 est fortement impliqué dans la réplication virale en intervenant dans le trafic intracellulaire et l'incorporation des glycoprotéines de l'enveloppe dans les virions. Une mutation dans ce domaine pourrait être un mécanisme moléculaire (polymorphisme) nécessaire à l'établissement de la transmission et d'échappement à la réponse immunitaire.

Mots-clés: VIH, virus transmis/fondateur, infection récente, signatures génétiques de l'enveloppe, réseau de transmission, analyse phylogénétique.

Abstract

Overcoming HIV infection effectively requires further refinement of antiretroviral therapy as well as intensifying prevention strategies to stop the spread of the disease. It involves the capacity to develop a preventive or therapeutic broad-spectrum vaccine. Reaching this goal requires a better understanding of the acute and early events mediated by HIV founder viruses' envelope glycoproteins during primary infection. While efforts of the scientific community are focused on finding a preventive vaccine, it is also necessary to limit the rate of new infections in the population. This thesis contributes to the efforts focusing on HIV prevention by studying the transmission dynamics of disease transmission and progression.

The first part of my work is detailed in Chapter III and presented in a manuscript published in the journal **"PLOS ONE"** bears the title: **« HIV-1 envelope sequence-based measures for identifying recent infections »** (Article 1). The objective of this study was to evaluate and determine which molecular method, among envelope genetic diversity measures, can qualify the current status of HIV-1 infection. Four genetic diversity measures of short segments of HIV-1 envelope sequences were evaluated for their efficacy to characterize infection recency. They included i) the percent complexity, ii) the percent of diversity, iii) the number of haplotypes and iv), the Shannon entropy. We have identified Shannon entropy as the best diversity measure which can effectively predict the recency of HIV-1 infection when associated to three short segments (less than 100 base pairs) of the HIV-1 envelope gp120 C2 segments 1 and 3 and the V3 loop as envelopes sequenced based diversity measure (s) and segments. The performance index of Shannon entropy for these 3 segments: gp120 C2 segment 1; C2 segment 3; V3 loop was 0,806, 0, 0805, and 0,812 respectively. It could serve as a molecular biomarker for identifying new infections in newly diagnosed HIV-1-positive patients. The ability to differentiate recent infections from established (chronic) infections is important for public health. In fact, newly infected people are the most infectious and able to spread the disease quickly in the population. Since these people are usually unaware of their HIV status, they may engage in risky behaviours. This population generally has a high viral

load with a relatively high-replicating capacity (fitness) that accelerates HIV-1 forward transmission. Being able to identify these individuals as soon as possible could limit and prevent new transmissions. Our work demonstrated that Shannon entropy measuring of these nucleic sequences can identify new infections.

The second part of my work is reported in Chapter III and published in the journal « **AIDS Research and Human Retroviruses** » entitled: « ***A short-term assessment of HIV-1 transmission dynamics among newly diagnosed individuals using envelope sequence-based phylogenetic clustering*** » (Article 2). This study consists of identifying the phylogenetic links or clusters that exist between the viral sequences of the HIV-1 envelope of newly diagnosed patients in Québec. The formation of clusters between the viral sequences of several individuals could reflect the relationship of transmitted viruses between individuals who might share characteristics or common risk factors. The results of this study identified 15 minor transmission clusters (2-5 individuals per cluster) between study participants. We also assessed and compared the ability of a fragment of a larger nucleotide sequence of the HIV-1 envelope to that of the much shorter V3 loop to independently predict and identify individuals that might be included in a transmission cluster. They demonstrated a moderate agreement based on the kappa coefficient ($k= 0.59$) between the two fragments of the viral envelope used. The use of the V3 loop sequences indeed confirmed 66.70% (10/15) of the identified clusters with a longer fragment of the HIV-1 envelope sequence. The number of infected individuals accessing genotyping tests that use V3 sequence to determine viral tropism in routine clinical practice is globally limited. However, the sequences of the viral envelope and particularly those of the V3 loop could easily be integrated into a public health program where they could be available in addition to those of the pol genes (available in clinical laboratories for antiretroviral resistance testing). We also included the characteristics of the subjects (sociodemographic, clinical and risk factors) in a logistic regression analysis model with cluster formation as an outcome). The results of the analysis demonstrate that in the study population, the average age (<38.8 versus >38.8 years) and the viral subtype (subtype B versus non-B) are two factors significantly associated with HIV transmission clustering among participants. Underlying HIV-1 transmission chains among HIV-1+

individuals in Québec and identifying the common risk factors should help the efforts to implement better prevention strategies in the target groups.

The results of the third and last part of this work reported in this thesis (Chapter III, Article 3), were published in « **VIRUSES** » journal entitled: « ***HIV-1 Envelope Glycoprotein Amino Acids Signatures Associated with Clade B Transmitted/Founder and Recent Viruses*** ». It consists in identifying genetic signatures in the viral envelope sequences of the first HIV-1 variants called Transmitted/Founder (TF) viruses selected during the acute stage of infection. These variants can establish a productive infection at the time of infection in the new host. They are indeed selected among a multitude of viral quasi-species exposed during virus transmission. They may possess molecular and phenotypic properties that may govern their functions of the successful establishment of infection. Identifying such early HIV-1 genetic signatures may inform the design of novel immunologic or chemical inhibitors and serve as a benchmark for designing an effective HIV-1 vaccine. This study identified a few genetic signatures among HIV-1 TF viruses' envelope amino acid sequences. One of these signatures consisting of a mutation/substitution of an arginine (R) by an isoleucine (I) associated with TF virus seems to be particularly important. This mutation is located at position 841 (R841I) specifically in the cytoplasmic domain of gp41 in segment 1 of lentivirus lytic peptides (LLP-1). The isoleucine is selected at more than 33% by TF viruses compared to chronic infection viruses (9%). The difference of selected amino acid between the two types of infectious viruses is statistically significant, OR = 0.2, 95% CI [0.09, 0.44], $p = 0.00001$. The cytoplasmic domain of GP41 and specifically the LLP-1 is strongly implicated in viral replication by mediating intracellular trafficking and the envelope incorporation into virions. A mutation in this area may suggest a molecular mechanism (polymorphism) contributing to HIV-1 transmission and/or viral escape from immune response.

Keywords: HIV, transmitted/founder viruses, recent infection, envelope genetic signatures, transmission clusters, phylogenetic analysis.

Table des matières

Résumé	5
Abstract	9
Table des matières	13
Liste des tableaux	17
Liste des figures	19
Liste des abréviations	23
Liste des acides aminés	29
Dédicaces	30
Remerciements	33
CHAPITRE I	37
INTRODUCTION GÉNÉRALE	37
1. Histoire du syndrome d'immunodéficience acquise (SIDA).	39
2. Épidémiologie de l'infection par le VIH.	41
2.1 Épidémiologie dans le monde.....	41
2.2 Épidémiologie du VIH au Canada.....	42
3. Caractéristiques du virus de l'immunodéficience humaine (VIH).....	44
3.1 Classification du VIH.....	44
3.2 Structure du VIH mature.....	45
3.3 Organisation génomique du VIH-1.	46
3.4 Formation des gènes structuraux du VIH-1.	48
3.5 Cycle de réplication du VIH-1.....	52
3.6. Structures et fonctions des glycoprotéines de l'enveloppe du VIH-1.	63
3.6.1 Généralités et fonctions.....	63
3.6.2 Glycoprotéine de surface, SUgp120.	65

3.7 Histoire naturelle de la maladie.....	70
3.8 Origines du VIH.	79
3.9 Diversité génétique du VIH.	81
4. Physiopathologie de l'infection VIH.....	94
4.1 Généralité et mode de transmission de l'infection.	94
4.2 Virus transmis/fondateur ou <i>transmitted/founder</i> » <i>viruses</i>	95
4.3 Virus transmis/fondateur et développement de vaccin anti-VIH.	98
4.5 Réponse de l'hôte face à l'infection par le VIH-1.	102
4.6. Anticorps neutralisants à large spectre contre le VIH.	109
CHAPITRE II	115
HYPOTHÈSES ET OBJECTIFS	115
1. Hypothèses de l'étude.	117
2. Objectifs de l'étude.....	117
2.1 Objectif principal.....	117
2.2 Objectifs secondaires.....	117
2.2.1 Objectif secondaire 1:.....	117
2.2.2 Objectif secondaire 2:.....	118
2.2.3 Objectif secondaire	118
CHAPITRE III	119
RÉSULTATS PAR ARTICLES.....	119
ARTICLE 1: HIV-1 ENVELOPE SEQUENCE-BASED DIVERSITY MEASURES FOR IDENTIFYING RECENT INFECTIONS (333).	121
CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 1.....	123
RÉSUMÉ.....	125
ABSTRACT.....	127
1. INTRODUCTION.....	129
2. MATERIALS AND METHODS	131
3. RESULTS.....	138

4. DISCUSSION.....	153
5. REFERENCES.....	159
ARTICLE 2: A SHORT-TERM ASSESSMENT OF NASCENT HIV-1 TRANSMISSION CLUSTERS AMONG NEWLY DIAGNOSED INDIVIDUALS USING ENVELOPE SEQUENCE-BASED PHYLOGENETIC ANALYSES (667).....	171
CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 2.....	173
RÉSUMÉ.....	173
ABSTRACT.....	177
1. INTRODUCTION.....	179
2. MATERIALS AND METHODS.....	181
3. RESULTS.....	185
4. DISCUSSION.....	198
5. CONCLUSION.....	203
6. ANNEXES.....	205
7. REFERENCES.....	206
ARTICLE 3: HIV-1 ENVELOPE GLYCOPROTEIN AMINO ACIDS SIGNATURES ASSOCIATED WITH CLADE B TRANSMITTED/FOUNDER AND RECENT VIRUSES (668).	216
CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 3.....	217
RÉSUMÉ.....	219
ABSTRACT.....	221
1. INTRODUCTION.....	223
2. MATERIALS AND METHODS.....	225
3. RESULTS.....	231
4. DISCUSSION.....	244
5. CONCLUSIONS.....	249
6. ANNEXES.....	251
7. REFERENCES.....	257
CHAPITRE IV.....	269
DISCUSSION GÉNÉRALE ET PERSPECTIVES DE L'ÉTUDE.....	269

1. Contributions de cette thèse sur le plan scientifique.	270
2. Perspectives	277
CONCLUSIONS	279
ANNEXES GÉNÉRALES	281
Protocole d'amplification de l'enveloppe (gp160) du VIH-1	283
Annexe 1 : Protocole de Reverse Transcriptase PCR (RT-PCR).....	283
Annexe 2 : Protocole d'amplification par RT-PCR nichée de l'enveloppe du VIH-1.....	285
BIBLIOGRAPHIE	287

Liste des tableaux

Tableau I.	Caractéristiques et fonctions des protéines de structures du VIH-1.....	50
Tableau II.	Caractéristiques et fonctions des protéines régulatrices et accessoires du VIH-1.....	51
Tableau III.	Durée des différentes phases de la primo-infection du VIH-1 selon la classification Fiebig.	75
Tableau IV.	(Article 1, S1 Table). Method of calculation of sequence-based diversity measures.....	135
Tableau V.	(Article 1, Table 1): HIV-1 subtype distribution of the sequences analyzed in this study.....	139
Tableau VI.	(Article1, Table 2). Performance of Shannon entropy as a measure for identifying recent HIV-1 infections.....	145
Tableau VII.	(Article 1, Table 3). Performance of 4 sequence-based diversity measures for non-B HIV-1 subtypes by <i>env</i> segments using AUC au ROC analysis.....	151
Tableau VIII.	Tableau: (Annexe 1, S1 table, Article 1): Methode explicative du calcul des quatre mesures de la diversité génétique des séquences virales (Shannon entropy, complexity, diversity (exemple).	160
Tableau IX.	(Article 2, Table 1). Population characteristics.....	187
Tableau X.	(Article2, Table 2). Performance of the <i>env</i> gp120-V3 sequences in determining HIV transmission clustering compared to the use of the full- <i>env</i> fragment.	193
Tableau XI.	(Article 2, Table 3). Summary statistics of the distribution of HIV-1 transmission clusters by factors	194
Tableau XII.	(Article 2, Table 4) Factors associated with HIV-1 transmission clustering using logistic regression analysis.....	196
Tableau XIII.	Table 1. Summary statistics of the important amino acids change among the HIV-1 envelope sequences of chronic and transmitted/founder viruses.	238

Tableau XIV. Table 2. Major amino acid signatures among HIV-1 envelope sequences between chronic and recent viruses..... 243

Tableau XV. (Annexe 1, Table S1, Article 3). Samples description: timeline categories of HIV-1 infection. 252

Tableau XVI. (Annexe 2, Table S2, Article 3). Summary of the HIV-1 subregions and domains locations according HXB2 envelope (gp160) numbering..... 253

Tableau XVII. (Annexe 3, Table S5, Article 3). Descriptive statistics of HIV-1 variable regions characteristics. 255

Liste des figures

Figure I. Représentation de l'évolution du nombre de nouveaux cas de VIH rapportés au Canada, 1996-2018.	43
Figure II. Représentation cartographique de la répartition du taux d'incidence du VIH par provinces et par territoires du Canada en 2018.	44
Figure III. Représentation des principaux composants d'une particule virale (VIH) mature ou virion.	46
Figure IV. Structure et organisation du génome du VIH-1.	48
Figure V. Représentation schématique des grandes étapes du cycle de réplication du VIH-1.	52
Figure VI. Représentation schématique des différentes étapes d'entrée du VIH-1 dans la cellule cible.	54
Figure VII. Représentation schématique des différentes étapes de la transcription inverse de l'ARN.	57
Figure VIII. Représentation schématique des différentes étapes tardives post-entrée du VIH.	61
Figure IX. Schéma de la structure de la glycoprotéine gp160 et de ses dérivés, gp120 et gp41.	64
Figure X. Structure du noyau de la glycoprotéine gp120 de l'enveloppe du VIH-1.	65
Figure XI. Représentation schématique de la structure de la glycoprotéine de surface SUgp120 de l'enveloppe du VIH-1.	67
Figure XII. Représentation de la structure de la TMgp41 de l'enveloppe du VIH-1.	69
Figure XIII. Représentation schématique de l'évolution naturelle de l'infection par le VIH-1 et évolution des marqueurs biologiques.	71
Figure XIV. Histoire naturelle et classification Fiebig de l'infection primaire par le VIH-1.	74
Figure XV. Images des différents chimpanzés et les gorilles ancêtres et principaux réservoirs du VIH.	80

Figure XVI. Liste des espèces de virus simiens ancêtres du VIH-1 humain.....	81
Figure XVII. Mécanisme de recombinaison génétique.....	83
Figure XVIII. Description des propriétés des trois phénotypes de VIH-1 utilisant les cellules Affinofile cells.....	89
Figure XIX. Répartition géographique des différents sous-types viraux et CRF du VIH-1 groupe M, N et O dans le monde.	94
Figure XX. Caractéristiques et mécanismes de sélection du virus transmis/fondateur.	97
Figure XXI. Représentation schématique de la première ligne défense contre le VIH.	105
Figure XXII. Représentation schématique des réponses cellulaires T induites par suite de l'infection par le VIH.	108
Figure XXIII. Représentation des différents anticorps neutralisants à large spectre (bNAbs) dirigés contre les régions conservées de l'enveloppe du VIH-1.....	111
Figure XXIV. Mécanisme d'échappement du VIH-1 aux anticorps neutralisants médié par la formation de bouclier de glycane sur les épitopes de neutralisation au niveau de l'enveloppe virale.....	113
Figure XXV. (Article 1, Figure 1). Schematic figures showing all <i>env</i> segments used for diversity estimates.....	133
Figure XXVI. (Article 1, Figure 2). Number of sequences (one per patient) used in this study.	139
Figure XXVII. (Article 1, Figure 3). Frequency polygons (ggplot2) of percent complexity of <i>env</i> sequences of recent HIV-1 infected individuals compared to chronically infected ones by <i>env</i> segments.....	140
Figure XXVIII. (Article 1, Figure 4). Frequency polygons (ggplot2) of percent diversity of <i>env</i> sequences of recent HIV-1 infected individuals compared to chronically infected ones by <i>env</i> segments.....	141
Figure XXIX. (Article 1, Figure 5). Frequency polygons (ggplot2) of a number of haplotypes of <i>env</i> sequences of recent HIV-1 infected individuals compared to chronically infected ones by <i>env</i> segments.	142

Figure XXX. (Article 1, Figure 6). Frequency polygons (ggplot2) of Shannon entropy index of *env* sequences of recent HIV-1 infected individuals compared to chronically infected ones by *env* segments. 143

Figure XXXI. (Article 1, Figure 7). ROC curves comparing the performance of the 4 sequence-based diversity measures for discriminating recently from chronic HIV-1 infection. 146

Figure XXXII. (Article 1, Figure 8). ROC curves comparing the performance of the 4 sequence-based diversity measures for discriminating recently from chronic HIV-1 infection. 147

Figure XXXIII. (Article 1, Figure 9). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of HIV-1 gp120 conserved subdomains to identify HIV-1 infection recency. 148

Figure XXXIV. (Article 1, Figure 10). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of five HIV-1 *env* gp120 variable loops and one part of the gp41-ectodomain (NHR) to identify HIV infection recency. 149

Figure XXXV. (Article 1, Figure 11). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of HIV gp120-C2-1, gp120-C2_3 and gp120-V3 segments for identifying HIV-1 subtype B infection recency. 150

Figure XXXVI. (Article 2, Figure 1). Flowchart of HIV-1 envelope sequences used in this study. 186

Figure XXXVII. (Article 2, Figure 2). Molecular phylogenetic reconstruction of HIV-1 transmission clusters among newly and chronically HIV-infected individuals in 2015 in Quebec using partial HIV-1 *env*-derived sequences (1070 bp). 189

Figure XXXVIII. (Article 2, Figure 3). Molecular phylogenetic reconstruction of HIV-1 transmission clusters among newly and chronically HIV-infected individuals using *env* gp120 V3 loop-derived sequences (108 bp). 191

Figure XXXIX. (Article 2, Figure 4). Tanglegrams of rooted phylogenetic trees and networks comparing the HIV-1 envelope gp120 loop 3 (a) with the HIV-1 partial envelope length (b) sequence-based clustering. 192

Figure XXXX. (Article 3, Figure 1): Flow chart showing the numbers of HIV-1 positive samples analyzed, and consensus envelope sequences of HIV-1-infected individuals include in the study..... 226

Figure XXXXI. (Article 3, Figure 2): Ridge plot comparing the HIV-1 envelope variable loop number of N-glycosylation sites between TF, RC and chronic (CH) viruses. 233

Figure XXXXII. (Article 3, Figure 3): Ridge plot comparing the clade B HIV-1 envelope variable loop length between CH, RC and TF viruses. 234

Figure XXXXIII. (Article 3, Figure 4). Ridge plot comparing the clade B HIV-1 envelope variable region gp120 loop 3 (V3) net positive between CH, RC and TF viruses. 235

Figure XXXXIV. (Article 3, Figure 5). Genetic signature identified under the HIV-1 envelope GP41 Lentivirus Lytic peptide 1 (LLP-1) associate to clade B HIV-1 TF viruses compared to chronic (CH) using WebLogo..... 236

Figure XXXXV. (Article 3, Figure 6). Genetic signature identified under the HIV-1 envelope signal peptide (SP) associate to clade B HIV-1 chronic compared to TF viruses using WebLogo. 237

Figure XXXXVI. (Article 3, Figure 7). Genetic signature identified under the HIV-1 envelope GP120 V1 loop associate to clade B HIV-1 chronic compared to recent viruses. 239

Figure XXXXVII. (Article 3, Figure 8). Genetic signature identified under the HIV-1 envelope signal peptide (SP) associate to clade B HIV-1 chronic compared to recent viruses. 240

Figure XXXXVIII. (Article 3, Figure 9). Genetic signature identified under the HIV-1 envelope GP41 cytoplasmic tail (CT or CD) loop domain associate to clade B HIV-1 chronic compared to recent viruses..... 241

Figure XXXXIX. (Article 3, Figure 10). Genetic signature identified under the HIV-1 envelope GP41 cytoplasmic tail (CT or CD) between NF-κB activation (NA) and Kennedy Epitope (KE) domains associate to clade B HIV-1 chronic compared to recent viruses. 242

Liste des abréviations

Ab: Antibody

Ac: Anticorps

ADCC: Antibody-dependent cell-mediated cytotoxicity

ADCP : Phagocytose cellulaire dépendante de l'anticorps

ADCVI : Inhibition virale dépendante des anticorps

ADN : Acide désoxyribonucléique

ADNc: ADN complémentaire

APC : Antigen presenting cells ou Cellules présentatrice d'antigènes

APOBEC3G : Polypeptide 3G d'apolipoprotéine B catalytique à l'activité d'édition de L'ARNm (Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G)

ARN: Acide ribonucléique

ARNm: Acide ribonucléique messenger

ARNt (Lys3) : ARN de transfert 3 pour la lysine

ART: Antiretroviral treatment ou traitement antirétroviral

ARVs : Antirétroviraux

ASPC : Agence de la santé publique du Canada

AZT: Azidothymidine ou Zidovudine

BAF: Barrier-to-autointegration factor1

bnAbs: Anticorps neutralisant à large spectre ou broadly neutralizing antibodies

C: Constant domain,

CA: Capside ou capsid protein

CAF: Cell antiviral factor

CCR5: C-C chemokine receptor type 5

CD: Cluster de différenciation,

CD4bs : Site de fixation du CD4 ou CD4 binding site

CD4i: CD4-induced ou CD4 induit

CDC: Centers for Disease Control and Prevention

CDs : Dendritic cells ou cellules dendritiques

CHR: C-terminal heptad repeat,

CMH-I : Complexe majeur d'histocompatibilité de type 1

CPSF6: Cleavage and Polyadenylation Specificity Factor 6"

CRFs: Circulating Recombinant Forms ou Formes recombinantes circulantes

CT: Cytoplasmic tail ou Queue/Domaine cytoplasmique

CTL : Cytotoxic T lymphocyte ou Lymphocytes T cytotoxiques

CTS: Central termination signal

CXCR4: C-X-C chimiokine receptor type 4

CypA: Cyclophiline A

DC-SIGN: Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin

É.-U: États-Unis

ECD : Ectodomaine

EIA: Enzyme immunoassay

ELISA: Enzyme linked immunosorbent assay

env: Envelope

ESCRT: Endosomal sorting complex 1 – 3 required for transport

FcγRIIC: Receptors for the Fc domain of immunoglobulins (Ig)

FDA: Food and Drug Administration

FIs : Fusion inhibitors ou Inhibiteurs de la fusion

FP : Fusion Peptide ou Peptide de fusion (FP)

gag: Group-specific antigen

GP: Glycoprotéin ou Glycoprotéine

GRID: Gay-Related Immune Deficiency

HARSAH : Hommes ayant des rapports sexuels avec d'autres hommes

HLAs: Human Leukocytes antigens

HR1: Helical heptad repeat 1

HTLV-1: Human T-cell leukemia virus type-I

HIV III : *Human T-lymphotropic retrovirus* ou Virus T-lymphotrope humain type III

ICTV: International Committee on the Taxonomy of Viruses

IFN- γ : Interferon gamma

Ig: Immunoglobulin

IL : Interleukine

IN: Intégrase

INI: Integrase Interactor 1

INNRTI : Inhibiteurs non nucléosidiques de la reverse transcriptase

INRTI : Inhibiteurs nucléosidiques de la reverse transcriptase

INs : Inhibiteurs de l'intégrase

INtTI : Inhibiteurs nucléotidiques de la reverse transcriptase

KIF5B: Kinesin-1 motor

KIRs: Killer immunoglobulin-like receptors

LANL: Los Alamos National Laboratory

LAV: Lymphadenopathy-associated virus ou virus associé à la lymphadénopathie

LEDGF: Lens epithelium-derived growth factor

LLP : Peptide lytique de lentivirus ou lentivirus lectic peptides

LTR : Long terminal repeat ou longues séquences de répétitions terminales

MA : Matrix protein ou protéine de la matrice

MAPK: Mitogen-activated protein kinase

MDA5: Melanoma-differentiation-associated gene 5

MIP-1 α/β : macrophage inflammatory protein-1 alpha/beta

MPER: Membrane-proximal external region.

NC: Nucléocapside ou nucleoprotein

Nef : *Negative regulating factor* ou Facteur de régulation négative o

NHR: N-terminal heptad repeat

NK : *Natural Killer* ou cellules tueuses naturelles

NPC : *Nuclear pore complex* ou complexe de pore nucléaire

Nup358: Nucleoporin 358

ONUSIDA : Programme commun des Nations unies sur le VIH/sida

ORF : *Open Reading frame* ou cadres de lecture ouverts

PAMP : Pathogen-associated molecular patterns ou modèles moléculaires associés aux pathogènes

PBMCs : Peripheral blood mononuclear cells ou cellules mononuclées du sang périphérique

PBS: Primer Binding Site

PCR : Polymerase reaction chain ou réaction de polymérisation en chaîne

PDCs: Plasmacytoid dendritic cells

PFD: Pocket-forming domain,

PIC : Preintegration complex ou complexe de pré intégration

pol : Polymerase ou polymérase

PPT: PolyPurine Tract

PR: Protease ou protéase

QNEs: Quaternary neutralizing epitopes

RANTES: Regulated on activation normal T expressed and secreted

RDC : République démocratique du Congo

RE : Réticulum endoplasmique

Rev : RNA splicing regulator ou épissage-régulateur de l'ARN

RIG-I: Retinoic-acid-inducible protein I

RLR: Retinoid acid-inducible gene RIG-like receptors

RNase H : Ribonuclease H ou Ribonucléase H

RRE: Rev response element

RT: Reverse Transcriptase

RTCs : *Reverse transcription complex* ou complexe de transcription inverse

SAMHD1: Sterile Alpha Motif and HD domain 1

SDF-1: Stromal cell-derived factor 1

SERINC 3/5: Serine incorporators 3 and 5

SIDA : Syndrome d'immunodéficience acquise

SIV : *Simian immunodeficiency virus* ou virus de l'immunodéficience simienne **SIVcpzPts**:
 Simian immunodeficiency virus of *Pan troglodytes schweinfurthii*

SIVcpzPtt: Simian immunodeficiency virus of *Pan troglodytes troglodytes* (chimpanzés)

SIVgor: Simian immunodeficiency virus of *Gorillas* (gorilles)

SIVsmm: Simian immunodeficiency virus of *sooty Mangabey* (*singes Mangabey enfumés*)

SNC : Système nerveux central

SP : Signal Peptide ou Peptide signal

SP1 : Spacer peptide 1

SP2 : Spacer peptide 2

ssRNA: Single-stranded RNA virus ou virus ARN simple brin

SUgp120 : Surface glycoprotein (SU) ou Glycoprotéine de surface.

T/F: *Transmitted /Founder virus* ou virus transmis/fondateur

Tat : Transactivator protein ou protéine transactivatrice

TCR : **Tcell receptor** ou récepteur de cellules T

THN (Tetherin): IFN-induced cellular proteins ou BST-2, CD317

Tev: Tat/rev protein

TFIIB: Human transcription factor IIB

TGF-β: Transforming growth factor

Th : T Helper ou cellule CD4 + T auxiliaire

TI: Transcriptase inverse ou reverse transcriptase

TIP47: Tail interacting protein

TLR: Toll-like receptors

TMD : Transmembrane domain ou domaine transmembranaire

TMgp41 : Transmembrane domain glycoprotein, gp41 ou domaine transmembranaire de la glycoprotéine gp41

TNPO3: Cellular transportin 3 protein

Treg: Regulatory T Cells ou Lymphocyte T régulateur

TRIM5α : Tripartite motif-containing protein 5

URFs : Unique Recombinant Forms ou formes recombinantes uniques

V: Variable domain,

Vif : Viral infectivity protein ou Facteur d'infectivité virale

VIH-1 : Virus de l'immunodéficience humaine de type 1.

VIH-2 : Virus de l'immunodéficience humaine de type 2.

VIS : Virus de l'immunodéficience simien

VLP: Virus like protein

Vp R, U, X : Viral protein R, U, X ou Protéine virale R, U, X

Liste des acides aminés

A	Ala	Alanine
C	Cys	Cystéine
D	Asp	Acide aspartique
E	Glu	Acide glutamique
F	Phe	Phénylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Méthionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Sérine
T	Thr	Thréonine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine

Dédicaces

Je dédie ce travail à la mémoire de mon père, feu Kafando Landaogo Pierre, qui a été arraché à notre affection le 28 janvier 2014 au moment où je venais de déposer mes valises au Canada pour le début de cette thèse. Tu as été pour moi un exemple d'humilité et de courage. Merci pour tes conseils, ton exemple, tes motivations et toutes les prières tout au long de ma recherche du financement pour cette thèse. Même dans les moments les plus difficiles et dans le désespoir et malgré ta maladie tu continuais à me soutenir et me motiver. Au fond de toi, tu étais convaincu que je réussirais. Merci. Repose en paix.

Je dédie spécialement ce travail à la mémoire de feu, le professeur Peter Martins NDUMBE qui nous a quittés définitivement le 14 mai 2013. Ex-doyen de la Faculté de médecine et des sciences biomédicales de l'Université de Yaoundé 1, Cameroun; ex-directeur de la recherche et de la publication à l'Organisation mondiale de la santé, région Afrique (OMS/AFRO). Tu as été celui qui m'a inspiré, encouragé et facilité ma réintégration dans les études supérieures. Tu as supervisé et trouvé les financements pour mon mémoire de master à l'université de Yaoundé 1. Tu as été pour moi une source d'inspiration, un exemple d'humilité, de courage et travailleur infatigable doté d'une rigueur scientifique. Au moment d'embarquer pour cette thèse tu m'as conseillé et signé mes premières lettres de recommandation pour me permettre de trouver une inscription et de financement pour ma thèse. Sans tout citer, je te dis infiniment merci et reposez en paix, cher professeur.

Je dédie cette thèse, particulièrement à ma petite famille. Tout d'abord à ma chère épouse Podiama Lady Kapiéko. Tu as accepté de mettre en veille ta propre carrière et tes ambitions professionnelles pour m'accompagner dans cette aventure au Canada. Tu as toujours joué non seulement ton rôle de mère de famille et très souvent de père famille pour le bien-être et le suivi de nos enfants à cause de mes multiples absences. Je te suis infiniment reconnaissant. À mes cher(e)s enfants; Flora, Tégawendé Aimé Césaire, Silvère Williams et Gabriel, merci pour les sacrifices consentis inhérents à mes multiples absences auprès de

vous. Ces absences sont dues à mes études et autres occupations professionnelles dans ces dernières années. Vous auriez voulu avoir votre père tous les soirs à la maison, mais hélas. Je vous dédie et vous confère le mérite de l'aboutissement de ce travail.

Remerciements

L'achèvement de ce programme de doctorat n'a pas été sans difficulté avec d'énormes défis à relever. Me trouver aujourd'hui avec cette thèse représente pour moi une satisfaction et un accomplissement personnel. Il représente aussi pour ma famille et pour tous ceux ou celles qui m'ont accompagné et cru en moi une fierté partagée. Ce diplôme postulé représente une étape et un début pour ma carrière professionnelle et scientifique. Cependant, je n'aurais pu y arriver sans l'appui constant et indéfectible d'immenses personnalités dont je tiens personnellement à remercier et à leur témoigner toute ma gratitude.

Tout d'abord à Dre Cécile L. Tremblay, Professeure titulaire au Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), pour avoir accepté de m'accueillir dans son équipe de recherche en cours de thèse et de diriger cette thèse. Quand je vous ai approchée pour demander de rejoindre votre équipe de recherche vous m'avez accepté en dépit de vos charges et l'immense travail qui devait être fait en un temps restreint. Consciente de la situation, vous aviez immédiatement mis à ma disposition toutes les ressources nécessaires au laboratoire de santé publique et au CRCHUM pour faciliter la conduite de mes travaux le plus efficacement possible. Vous aviez également mis en place une équipe de chercheurs pour m'accompagner. Cette formidable équipe a vraiment été d'un support inestimable qui m'a accompagné tout au long de ma thèse. Je peux vraiment me considérer comme un chanceux qui est tombé dans de bonnes mains à un moment critique. Chanceux également pour vous avoir eu comme directrice de thèse aussi talentueuse, dynamique, pragmatique, travailleuse et méthodique.... Sans votre contribution personnelle, votre acceptation et celles de votre équipe, les travaux présentés dans cette thèse n'auraient pu l'être. Dre Tremblay, merci pour votre vision, votre sens de l'encadrement et votre rigueur scientifique qui ne cesseront de m'inspirer pour la suite de ma carrière professionnelle et académique.

Je tiens également à remercier infiniment le Dr Hugues Charest d'avoir accepté volontiers d'accompagner Dre Cécile Tremblay pour la codirection de cette thèse. Merci, Hugues, pour ta disponibilité et ton encadrement depuis les techniques de base de laboratoire, à l'analyse et à la rédaction de cette thèse. Sans ta disponibilité et ton accompagnement, je n'aurais pu avoir les résultats escomptés à présenter pour cette thèse. Merci pour tes conseils avisés qui ont contribué à l'aboutissement de ce travail.

Je ne saurais terminer sans vraiment remercier grandement tous les membres de cette formidable équipe de recherche qui m'a accompagné dans mes travaux au laboratoire de santé publique du Québec, au CRCHUM et au niveau du département de microbiologie, infectiologie et immunologie de l'UdeM. Vous m'avez individuellement apporté l'appui et l'aide nécessaire en tout moment et tout au long de cette thèse. J'ai nommé personnellement, Mr Éric Fournier, Dre Christine Martineau, Dre Bouchra Sherir, Dre Florence Doualla-Bell, Dr Réjean Dion, Mme Lyne Désautels, Dr Mohamed El-Far, Dre Annie Chamberland, Mr Mohamed Sylla, Dr Mohamed Ndongo Sangaré, Dr Drissa Siam Mr Adama Kazienga.

Je remercie également tout le personnel technique et administratif du laboratoire de biologie moléculaire, de sérodiagnostic et de virologie du LSPQ de l'institut national de santé publique du Québec pour toute l'assistance technique et administrative. Merci à chacune et à chacun de vous pour l'excellent travail d'équipe que nous avons pu réaliser ensemble.

Je profite également de cette tribune pour adresser toutes mes gratitudeux responsables et professeurs du département de microbiologie, infectiologie et immunologie de la faculté de médecine de l'Université de Montréal pour leurs conseils et accompagnement tout au long de ma thèse. Merci à Dr Hugo Soudeyns, directeur du département à Dr Guy Lemay, à Dr Andrès Finzi, à Dr Petronela Ancuta, à Dr Nicolas Chomont, Dr George Szatmari et Mme Youance Lamerchie. Je remercie infiniment le Dr Art FY. Poon de l'Université de Western Ontario et Dr Paul Wankah de l'Université Sherbrooke pour la révision de mes manuscrits et Dre Celine Scornavacca du CNRS/Université Montpellier, France, pour l'appui et conseils pour les analyses bio-informatiques

Je tiens enfin à remercier individuellement tous ceux ou celles qui m'ont aidé depuis mon pays d'origine le Burkina Faso, pour mes recherches de financement et de laboratoire d'accueil pour ma thèse. Je témoigne toute ma gratitude au professeur Nicolas Méda pour son investissement personnel pour permettre de trouver du financement et une inscription pour cette thèse. Votre accompagnement et les recommandations m'ont permis de trouver cette bourse de doctorat auprès de la Banque Islamique de Développement et l'inscription à l'Université de Montréal. Merci pour tous vos conseils, orientations et accompagnement avant et tout au long de ma formation. Je profite de cette occasion aussi pour témoigner vivement toutes mes gratitudes aux professeurs Robert Tinga Guiguemdé, Jean Bosco Ouédraogo, Yves Traoré, Lassana Sangaré, Ramata Ouédraogo, Éléonore Kafando et aux Docteurs Mathias Somé, Bocar Amadou Kouyate, Isaïe Medah, Hervé Hien, Alain Hien, Serges Sawadogo, Onadja Yentéma, etc., pour leurs conseils et les sollicitudes de divers ordres. Votre accompagnement m'a permis d'acquérir le financement et l'inscription pour cette thèse. Soyez-en remerciés.

Ma thèse ne serait pas aussi une réalité et les travaux réalisés ici n'auraient pu l'être sans l'aide de collaborateurs et de partenaires techniques et financiers. Notamment, la Banque Islamique de développement pour m'avoir offert cette bourse de doctorat et pour m'avoir permis d'être accompagné de ma famille pour cette thèse source de stabilité. Je vous témoigne toute ma gratitude. Merci à faculté des études supérieures et postdoctorales de l'Université de Montréal pour la bourse d'exemption pour étudiants étrangers, la bourse de fins d'études doctorales qu'elle m'a octroyées. Merci aux Fonds de Recherche Québec Santé (FRQ/S) et à Génome Canada pour le financement des analyses de laboratoire. Merci au Ministère de l'Éducation et de l'Enseignement supérieur du Québec pour l'octroi d'un prêt et bourse pour l'achèvement de mes études à l'UdeM. Merci au laboratoire de Dre Cécile L. Tremblay du CRCHUM pour la bourse d'appui octroyée. Merci également au (LSPQ/INSQ) pour le cadre et le plateau technique offerts qui m'ont permis de réaliser ces travaux dans de bonnes conditions. Merci au Centre Muraz et le Ministère de la Santé du Burkina Faso pour m'avoir accordé l'autorisation administrative pour entreprendre cette formation de doctorat et l'accompagnement pour l'obtention du financement. Je suis reconnaissant, à tous mes

parents, amis et connaissances pour m'avoir accompagné et permis d'intégrer la communauté burkinabé vivant à Montréal/Québec et cela depuis mon arrivée en mars 2013.

Je rends grâce à Dieu pour tous les bienfaits, les hommes et les femmes qu'il a mis au travers de mon chemin pour l'aboutissement de ce travail.

CHAPITRE I
INTRODUCTION GÉNÉRALE

1. Histoire du syndrome d'immunodéficience acquise (SIDA).

Le syndrome d'immunodéficience acquise (SIDA) est une maladie causée par le virus de l'immunodéficience acquise (VIH). Les premiers cas cliniques de la maladie ont été détectés au début des années 1980 aux États-Unis chez des homosexuels et des utilisateurs de drogues injectables. Ces malades se présentaient avec des symptômes apparentés à une atteinte du système immunitaire. En son début, le nom de « *Gay-Related Immune Deficiency (GRID)* » était attribué à cette pathologie, tel que rapporté par le journal « *the New York Time* » le 11 mai 1982 dans sa rubrique *Sciences*. Parmi les symptômes de la maladie on relevait, entre autres, des adénopathies généralisées, des pneumonies à *Pneumocystis jiroveci* (anciennement nommé *P. carinii*), des rétinites au cytomégalovirus, des méningites à *Cryptococcus neoformans* et des sarcomes de Kaposi (1). Les symptômes décrits se rapportent à des infections opportunistes qui ne devaient normalement pas se manifester chez un sujet immunocompétent.

Sur le plan biologique, ces patients présentaient une baisse très importante (déplétion) du taux des lymphocytes T CD4+ (*cluster de différenciation*, CD) dans le sang périphérique. Dans la littérature médicale, le « *Centers for Disease Control and Prevention (CDC)* » fut le premier laboratoire à rapporter des cas en 1981 chez des homosexuels des villes de Los Angeles, de San Francisco et de New York (1). D'un groupe restreint de sujet à risque qui contractaient la maladie en son début, le champ de transmission s'est élargi aux autres couches de la société et partout sur la planète pour devenir une véritable pandémie. Outre les transmissions homosexuelles et hétérosexuelles, une description des premiers cas de transmission verticale

de la maladie a par la suite été rapportée (2, 3). Le terme SIDA fut utilisé pour la première fois par le CDC pour désigner cette nouvelle maladie le 24 septembre 1982 (2, 4, 5). Au cours des deux années qui suivirent la description des premiers cas du SIDA, de nombreux chercheurs ont recherché l'agent étiologique causal. Le nom de « *Human T-cell leukemia virus type-I* » (*HTLV-1*) fut l'un des premiers termes proposés pour désigner le virus déjà connu pour donner des leucémies (6, 7), mais son groupe s'était trompé et donc la publication était une erreur. Ce fut finalement le 20 mai 1983 qu'une équipe dirigée par le professeur Luc Montagnier de l'Institut Pasteur de Paris publiait dans la revue *Science* (8) une première description du virus. Elle estimait avoir isolé un nouveau retrovirus qui serait l'agent étiologique responsable du SIDA et différent de HTLV-1 et qu'elle appela : « *Lymphadenopathy-associated virus* » ou « *virus associé à la lymphadénopathie* » (LAV) (8). L'échantillon sur lequel le virus a été isolé provenait d'un patient présentant une lymphadénopathie généralisée d'origine inconnue (8). Finalement, l'équipe conduite par le Dr Gallo de la « *National Cancer Institute* » aux États-Unis a décrit un an après l'isolement d'un autre rétrovirus susceptible d'être l'agent causal du SIDA chez des patients infectés. Elle reconnaitra en effet, un an après, qu'il s'agissait d'un nouveau virus (9, 10). L'équipe le nomma virus T-lymphotrope humain type III (*HTLV-III*) pour « *Human T-lymphotropic virus* ». Mais après séquençage quelques années plus tard, il s'est avéré que le virus isolé du laboratoire de Dr Robert Gallo était très similaire que celui de l'Institut Pasteur et c'était donc une contamination par le virus obtenu de l'Institut Pasteur, il s'agissait en effet du même virus. Cette information suscita une controverse tant scientifique que juridique entre les deux équipes au sujet de la paternité de la découverte du virus responsable du SIDA (11)

portant sur les redevances des brevets.puisque ces publications sont sorties aux mêmes moments. La paternité de la découverte du VIH fut finalement octroyée à l'équipe des Drs. Barré-Sinoussi et Montagnier (et coauteurs) et leur a valu le 6 octobre 2008, le prix Nobel de médecine alors décerné aux professeurs Françoise Barré-Sinoussi et Luc Montagnier pour l'identification en 1983 à l'Institut Pasteur du rétrovirus responsable du SIDA (12). En mai 1986, le virus d'immunodéficience humaine (VIH) fut officiellement déclaré comme le rétrovirus responsable du SIDA par le comité international sur la taxonomie des virus « *the International Committee on the Taxonomy of Viruses* » (ICTV) (13). Un autre rétrovirus apparenté au VIH de type 1, nommé VIH de type 2 (VIH-2) (14) fut découvert plus tard en 1986 chez des sujets atteints de SIDA dans des régions d'Afrique de l'Ouest.

2. Épidémiologie de l'infection par le VIH.

2.1 Épidémiologie dans le monde.

L'infection par le VIH constitue l'un des grands fléaux de santé publique de ces 40 dernières années. De quelques cas rapportés en 1982 et 1983 aux É.-U., l'infection par le VIH est devenue un fléau planétaire. Selon les rapports 2018 du programme commun des Nations unies sur le VIH/sida (ONUSIDA), on estimait que le nombre de personnes vivant avec l'infection au VIH dans le monde en 2018 était d'environ 37.9 millions [32.7 millions–44.0 millions] (15). Le nombre de personnes infectées varie considérablement en fonction des régions géographiques. Selon les statistiques de l'ONUSIDA, environ 25.6 millions d'individus vivent avec le VIH en Afrique subsaharienne ce qui représente 67.5% de la population totale infectée par le VIH à l'échelle mondiale.(15). Malgré la vulgarisation de l'information sur les moyens de prévention contre la maladie, on déplore encore de nouvelles infections. En 2018, on estime le nombre de nouvelles infections à 1.7 million [1.4 million–2.3 millions] représentant environ 5000 nouvelles infections à VIH par jour (15). Malheureusement,

environ 770 000 [570 000–1.1 million] de personnes sont décédées pour des causes liées au VIH dans le monde en 2018 (15). Pour la prise en charge thérapeutique et selon les statistiques de l'ONUSIDA, à la fin de juin 2019, environ 24,5 millions [21,6 millions–25,5 millions] de personnes avaient accès au traitement antirétroviral (15). Les voies de transmission de la maladie dans les pays d'Afrique subsaharienne sont de types hétérosexuels pour la grande majorité tandis que dans les pays nordiques elles sont majoritairement de type homosexuel. Ainsi, 54 % des nouvelles infections par le VIH dans le monde restent imputables aux populations clés représentant plus de risque à contracter la maladie en 2018. Ce risque était de: 1) pour les hommes ayant des rapports sexuels avec d'autres hommes (HARSAH) de (22 fois plus élevé), 2) pour les usagers des drogues injectables avec échanges de seringues contaminées de (22 fois plus élevé), 3) pour les travailleurs du sexe (21 fois plus élevé) et 4) pour les personnes transgenres (12 fois plus élevé) (15). Au regard de ces chiffres, on peut affirmer que la bataille contre l'infection par le VIH n'est pas encore gagnée, mais l'espoir est toujours permis quand on regarde le chemin parcouru et la qualité de vie et de prise en charge des malades et les statistiques de l'infection de nos jours. Pour l'atteinte de l'objectif ambitieux 90-90-90 de l'ONUSIDA, en 2018, il n'était que de 79% [67%– 92%] de personne vivant avec le VIH qui connaissait leur statut, 62% [47%–74%] qui ont l'accès au traitement (objectif de 79%) et 53% [43%– 63%] des personnes vivant avec le VIH qui ont vu leur charge virale supprimée (pour un objectif de 86%).

2.2 Épidémiologie du VIH au Canada.

L'Agence de la santé publique du Canada (ASPC) a rapporté 2 561 cas déclarés de VIH en 2018 comptant dans 70.7% pour les hommes (Figure I) (16). Comparativement à l'année 2017, le taux de diagnostic est croissant et passe de 6.5 à 6.9 pour 100 000 habitants en 2018 (16). La province de la Saskatchewan a rapporté le plus haut taux d'incidence du VIH en 2018, estimé à 14,8 pour 100 000 habitants bien qu'elle ne représente que 6,8 % du nombre total de cas de VIH déclarés dans le pays ((Figure II) (16) . La province du Québec suit avec un taux: 9.1 pour 100 000 habitants, puis le Manitoba: 7.9 pour 100 000 habitants, l'Ontario : 7.0 pour

100 000 habitants, l'Alberta: 5.8 pour 100 000 habitants, la Colombie-Britannique: 4.0 pour 100 000 habitants, les provinces de l'Atlantique: 2.8 pour 100 000 habitants et les territoires: 2.5 pour 100 000 habitants (Figure II) (16). Pour la répartition en nombre absolu des cas rapportés dans le pays en 2018 par provinces et territoires, l'Ontario comptait le plus grand nombre et la proportion la plus élevée (n = 1 003, 39,2 %), suivi du Québec (n = 766, 29,9 %), de l'Alberta (n = 249, 9,7 %) et de la Colombie-Britannique (n = 199, 7,8 %) (16). Les nouvelles infections à VIH rapportées dans l'ensemble du pays demeurent essentiellement masculines (70.7% des cas en 2018) et particulièrement chez les adultes âgés de 30-39 ans (16). Les gbHARSAH (gais, bisexuel et autres hommes ayant des rapports sexuels avec des hommes), les gbHARSAH/UDI (utilisateurs de drogues par injection) et les IDU représentent les populations à risque les plus affectées par le VIH en 2018 respectivement de 41.4,%, 3.4% et 18.3% du nombre total des cas (16, 17).

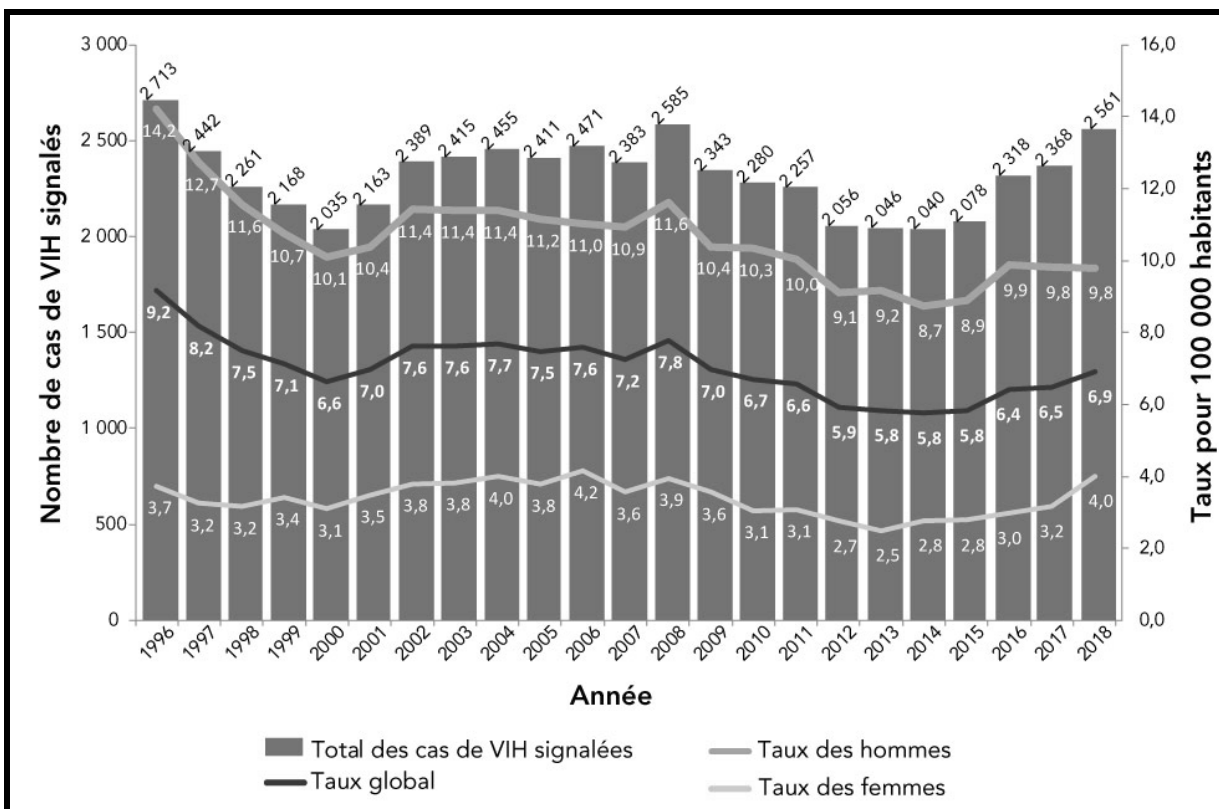


Figure I. Représentation de l'évolution du nombre de nouveaux cas de VIH rapportés au Canada, 1996-2018.

La figure ci-dessus démontre l'évolution du nombre de cas d'infections par le VIH-1 rapportés au Canada de 1996-2016 selon l'Agence de santé publique du Canada. Figure reproduite avec autorisation. Source: Haddad N *et al.* 2019 (16).

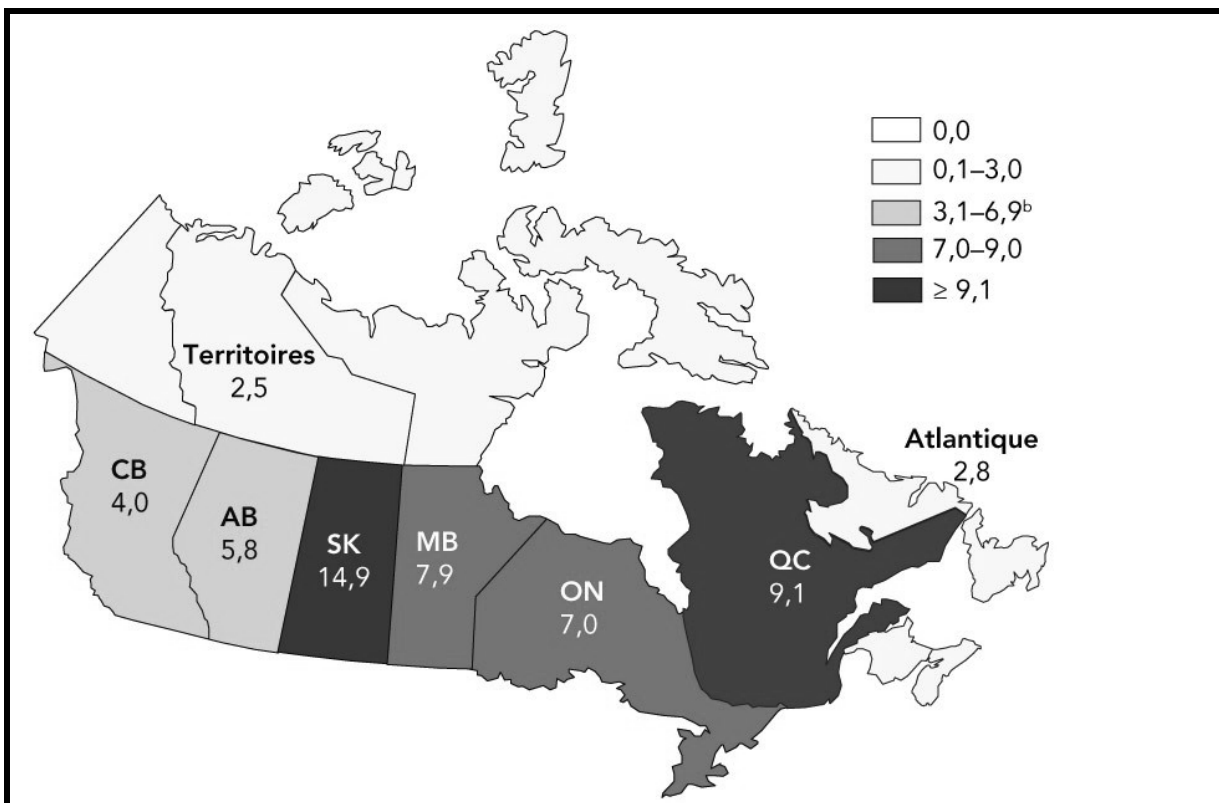


Figure II. Représentation cartographique de la répartition du taux d'incidence du VIH par provinces et par territoires du Canada en 2018.

La figure ci-dessus présente une cartographie du taux d'incidence du VIH par province et territoires du Canada en 2016 selon l'Agence de santé publique du Canada. Carte reproduite avec autorisation. Source : Haddad N *et al.* 2019 (16).

3. Caractéristiques du virus de l'immunodéficience humaine (VIH)

3.1 Classification du VIH.

Sur le plan taxonomique, le VIH appartient à la famille des rétrovirus ou *Rétroviridae* et à la sous-famille des *Orthoretrovirinae* et du groupe VI (18). Il s'agit d'un virus à ARN monocaténaire du genre *Lentivirus* qui se caractérise par sa capacité de rétrotranscription de son matériel génétique d'ARN en ADN grâce à une enzyme, la transcriptase inverse ou

reverse transcriptase (RT) (18). Il est capable de synthétiser son génome viral sous forme d'acide désoxyribonucléique (ADN) bicaténaire à partir de l'acide ribonucléique (ARN) monocaténaire génomique inclus dans la capsid. L'ADN viral transcrit peut intégrer le génome de la cellule cible grâce à l'action d'une enzyme nommée intégrase. C'est un virus doté d'une grande variabilité génétique. On distingue deux grands groupes de virus en se basant sur l'antigénicité : le virus de type 1 (VIH-1) et le type 2 (VIH-2). D'autres virus apparentés du groupe des lentivirus infectent les primates non humains, il s'agit entre autres du virus de l'immunodéficience simien (*VIS*) ou *Simian immunodeficiency virus* (SIV) (19).

3.2 Structure du VIH mature.

Le VIH mature a une taille comprise entre 80 et 120 nm de diamètre. Sa particule virale contient toutes les protéines nécessaires à sa maturation et à sa réplication (Figure III). De façon générale, la particule virale mature contient deux molécules d'ARN génomique monocaténaire de polarité positive, logées dans un noyau recouvert d'une nucléocapside (p7, NC). Une transcriptase inverse (p66/p51, RT) est incorporée à la particule virale et joue un rôle de transcription de l'ARN en ADN. Cette transformation est indispensable à l'intégration du virus au génome cellulaire (2, 8, 20, 21). Le cœur de la particule virale contient une intégrase (p31, IN), une enzyme qui permet l'intégration de l'ADN viral à l'ADN cellulaire. Il contient également des protéines dites accessoires (Vpr, Nef et Vif) (22).

La NC est entourée d'une capsid (p24, CA) contenant la protéase (p11, PR), une autre enzyme dont le rôle est de faciliter l'assemblage des éléments du virus en clivant les précurseurs protéiques Pr55Gag et Pr160Gag-Pol en ses principaux constituants protéiques pendant le processus de réplication virale.

La CA et ses composantes protéiques et ARN sont logés dans une matrice (p17, MA) recouverte d'une bicouche lipidique. Elle est constituée d'un fragment de la membrane de la cellule infectée et des glycoprotéines de surfaces (SU, gp120) et transmembranaires (TM, gp41) de l'enveloppe (2, 20). Les caractéristiques et fonctions des protéines structurales et accessoires du VIH sont représentées dans la figure III et les tableaux I et II respectivement.

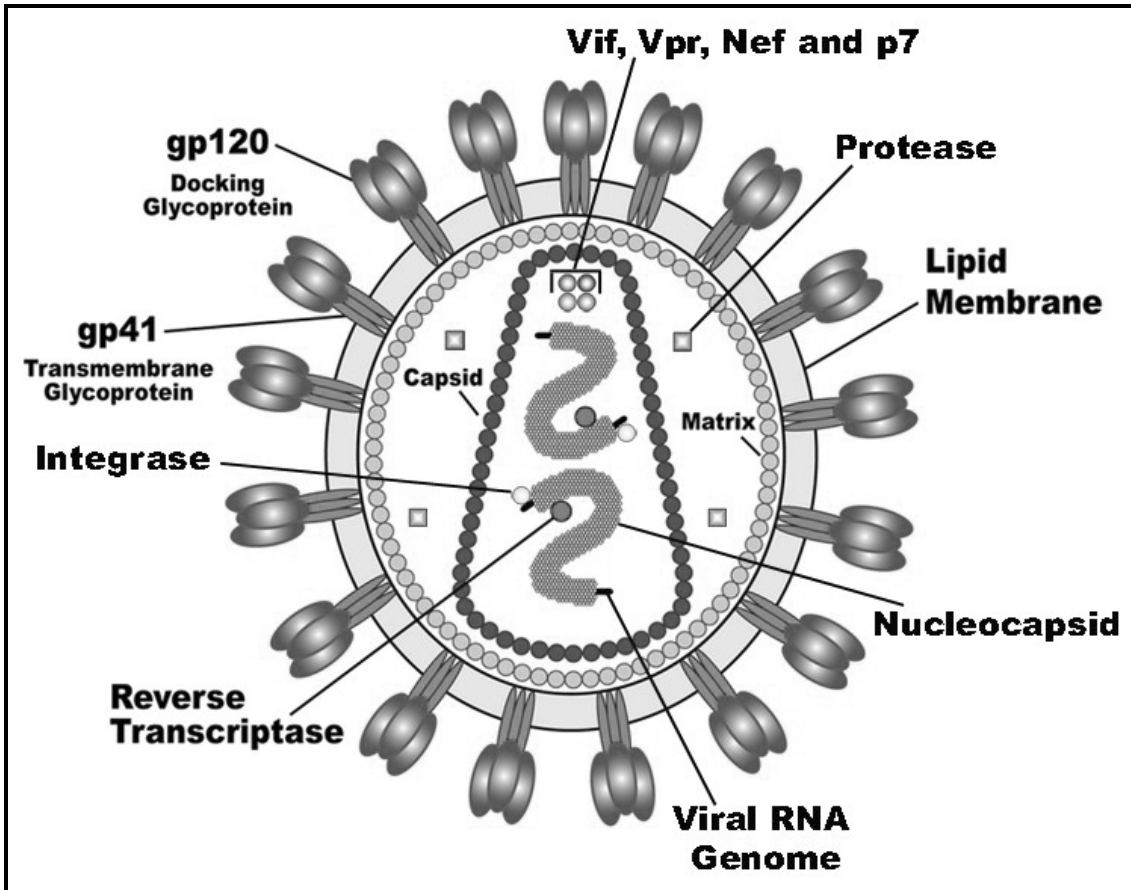


Figure III. Représentation des principaux composants d'une particule virale (VIH) mature ou virion.

La figure ci-dessus présente les principaux composants d'une particule virale. Y sont représentés les principales protéines structurales, régulatrices et les enzymes. Figure reproduite avec autorisation: US National Institute of Health (redrawn by: Carl Henderson, 2005) (23).

3.3 Organisation génomique du VIH-1.

Le génome du virus a une taille d'environ 9700 paires de bases, il contient neuf (9) cadres de lecture ouverts (*open Reading frame, ORF*) (2, 20, 24). L'ADN proviral provient de l'action de la transcriptase inverse qui permet la transcription de l'ARN génomique en ADN bicaténaire. Celui-ci s'intègre au génome de la cellule hôte et devient un provirus. L'intégration de l'ADN viral au génome cellulaire s'effectue grâce à de longues séquences terminales répétées identiques (*long terminal repeats, LTR*) (2, 20). Elles sont localisées aux deux extrémités du

génomique viral, permettent l'intégration de l'ADN viral dans le génome cellulaire et indispensables au contrôle de la transcription des gènes (20, 25, 26). Comme tout rétrovirus, le génome du VIH-1 code pour trois protéines principales. Il s'agit successivement dans le sens 5' vers 3' du génome viral, des gènes : 1) *gag* (group-specific antigen), 2) *pol* (polymérase) et 3) *env* (enveloppe) (2, 20, 24). Chacun de ces trois gènes dits de structure joue un rôle prépondérant pour l'organisation structurale et enzymatique nécessaire à la réplication du VIH (2) (Figure IV). En plus des protéines structurales, le génome code aussi pour: 1) des protéines régulatrices *tat* et *rev* qui sont essentiels à la réplication virale et 2) : des protéines accessoires *Vif*, *Vpr*, *Vpu* et *Nef* qui sont non essentielles à la réplication dans certains contextes cellulaires, mais indispensables pour la pathogénicité (2, 20, 27-30).

Le génome du VIH-2, autre rétrovirus, code pour la *Protéine virale* contrairement à protéine virale pour le VIH-1. Ces différences génétiques pourraient contribuer à la pathogénicité réduite relative au VIH-2 par rapport au VIH-1 (31). Les structures génomiques du virus de l'immunodéficience simien (chimpanzés (*VISScpz*), des gorilles (*VISSgor*)) et du VIH-1 sont presque comparables (20, 32). La différence fondamentale entre ces espèces virales réside dans la présence d'un résidu d'arginine basique (R) à la position 30 (30R) de la protéine p17 (*gag*). Il s'agit là d'une singularité génétique de la séquence protéique du groupe M,N et O du VIH-1 (33). Contrairement au VIH-1, les VIS portent un résidu méthionine hydrophobe (Met) à cette position (30M) (34).

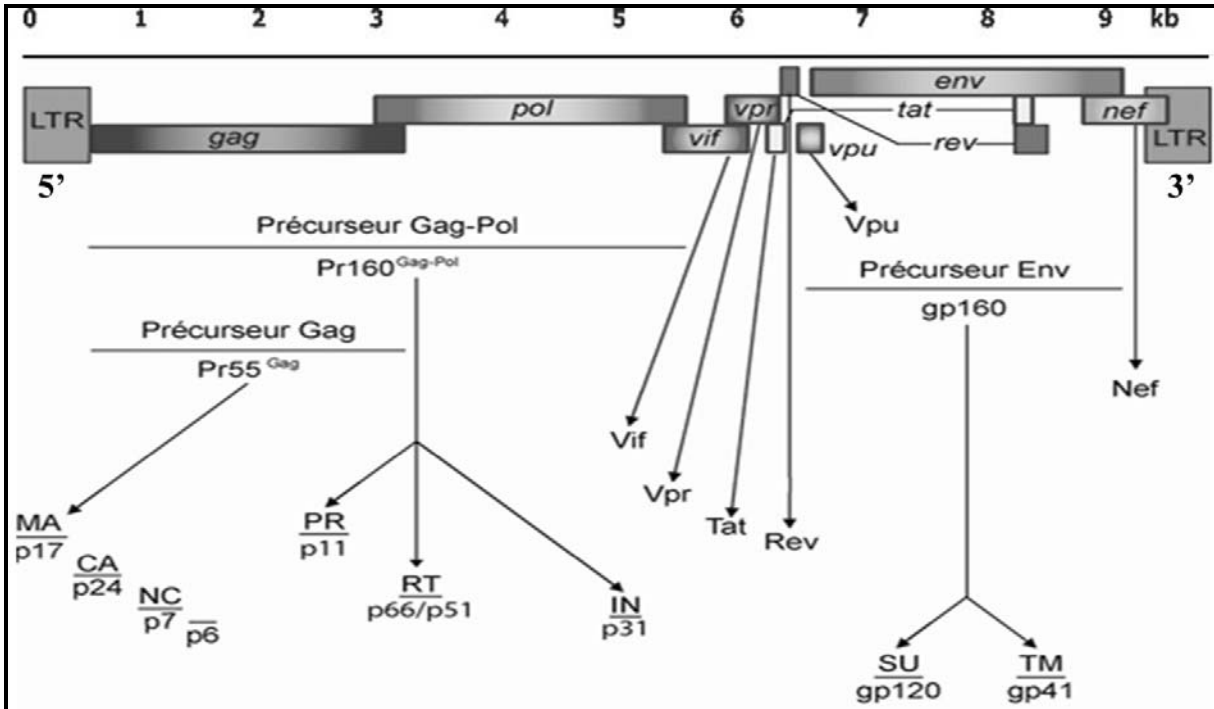


Figure IV. Structure et organisation du génome du VIH-1.

Le schéma ci-dessus illustre les cadres de lecture du génome du VIH-1 et les gènes codants pour les protéines de structures, régulatrices et accessoires flanqués à chaque extrémité par les longues séquences de répétitions terminales (LTR) identiques non traduites en protéines. La taille du génome du VIH-1 est d'environ 9700 paires de bases et approximativement de 9800 paires de bases pour le VIH-2. Abréviations: LTR (long terminal repeat); *gag* (group-specific antigen); *pol* (polymérase); *env* (enveloppe), SU (protéine de surface, gp120), TM (protéine transmembranaire, gp41), MA (matrice, p17), CA (capside, p24), NC (nucléocapside, p7), PR (protéase), IN (intégrase), RT (transcriptase inverse), Vif (Viral infectivity factor), Vpr (viral protein R), Vpu (Viral protein unique), Nef (Négative factor). Schéma reproduit avec autorisation. Source : Gurtler L *et al.* 2016 et Vidricaire G, 2006 (20, 35).

3.4 Formation des gènes structuraux du VIH-1.

1) Le **gène *gag*** code pour un précurseur polyprotéique nommé Pr55Gag. Il sera clivé par la protéase (PR) virale en protéines : matrice (MA ou p17), capsid (CA ou p17), nucléocapsid (NC ou p7), p6, p2 et p1 (2, 20, 36). 2) Le **gène *pol*** code pour les protéines enzymatiques virales, dont la protéase (PR, p12), la reverse transcriptase (RT, p51), la ribonucléase H (RNase H, p15) ou la RT + RNase H (ensemble, p66) et l'intégrase (IN, p32).

Générée par l'autocatalyse de la polyprotéine Pr160Gag-Pol, la protéase (PRp10) est un homodimère appartenant à la famille des protéases à acides aspartiques (37-45). Elle active sous sa forme dimérique par son site actif constitué d'une triade catalytique Asp25-Thr26-Gly27 localisée à l'interface des dimères. Elle joue un rôle important pour le cycle de réplication du VIH-1 et notamment dans la phase de maturation des virions en effectuant une série de 12 clivages permettant la maturation des différents précurseurs *gag*, *gag-pol* et Nef(46-50). Les clivages en effet conduit à la libération des protéines structurales (capside et nucléocapside), d'enzymes virales PR, RT et IN) actives et d'autres facteurs nécessaires pour la réplication virale, la maturation des virions et l'établissement de l'infection(46-49). .3) le gène *env*, la seule protéine présente à la surface des virions est synthétisée comme un précurseur glycoprotéique. Il est premièrement traduit en un polypeptide gp160 dans le réticulum endoplasmique sous forme de trimère. Le trimère est ensuite transporté dans l'appareil de Golgi pour être clivée par une enzyme, de la famille des protéines convertases (furines) en deux glycoprotéines : a) la glycoprotéine de surface SU (gp120) et, b) la glycoprotéine transmembranaire TM (gp41) (2, 20, 51) (cf Chap I. 3.6 pour une description complète). Les caractéristiques et les fonctions des protéines de structures, régulatrices et accessoires du VIH-1 sont résumées dans les tableaux I et II (20).

Tableau I. Caractéristiques et fonctions des protéines de structures du VIH-1.

Gène	Taille*	Protéine	Fonction
<i>gag</i>		Pr55Gag ou Group-specific-antigen	Précurseur des protéines structurales internes
	p24	Capsid protein (CA)	Joue un rôle dans la formation de la capside conique.
	p17	Matrix protein (MA)	Est myristoylée, formant la couche de membrane interne
	p7	Nucleoprotein (NC),	Intervient dans la formation du complexe nucléoprotéine/ARN
	P6		Impliqué dans la libération de particules virales
<i>pol</i>		Pr160GagPol	Précurseur des enzymes virales
	p10	Protease (PR)	Joue un rôle dans le clivage protéolytique de la protéine précurseur Gag (Pr55) et Gag-Pol (Pr160GagPol) et de la libération de protéines structurales et d'enzymes virales.
	p51	Reverse transcriptase (RT)	Enzyme indispensable à la transcription de l'ARN du VIH en ADN proviral
	p15/66	RNase H	Intervenant dans la dégradation de l'ARN viral dans le complexe de réplication ARN/ADN viral.
	p32	Intégrase (IN)	Intervient dans l'intégration de l'ADN proviral dans le génome de la cellule hôte.
<i>env</i>		PrGp160	Précurseur des protéines d'enveloppe de surface, SUG120 et transmembranaire TMgp41 clivé par protéase cellulaire.
	gp120	Surface glycoprotein (SU)	Glycoprotéine de surface faisant saillie au niveau de l'enveloppe virale et permettant au virus de s'attacher à la cellule cible.
	gp41	Transmembrane protein (TM)	Glycoprotéine transmembranaire permettant l'ancrage de la gp120 et la fusion entre les membranes virales et cellulaires

* nombre correspondant à la taille des protéines (p) ou des glycoprotéines (gp) en 1000 Da.
Tableau adapté avec autorisation. Source : Gurtler L et al. 2016 (20).

Tableau II. Caractéristiques et fonctions des protéines régulatrices et accessoires du VIH-1.

Gène	Taille*	Protéine	Fonction
Protéines régulatrices			
Tat	p14	Transactivator protein	Protéine régulatrice essentielle. Active la transcription des gènes et l'élongation des ARNm viraux (20, 52-54).
Rev	p19	RNA splicing regulator	Régule l'exportation de l'ARNm viral non et partiellement épissé hors du noyau (20).
Tev	p26	tat/rev protein	Elle permet la fusion du complexe formé de Tat-env-Rev et régule l'activité de Tat et de Rev dans le noyau (20).
Protéines accessoires			
Nef	p27	Negative regulating factor	Est myristoylée, elle augmente la réplication virale et diminue le nombre de cellules hôtes (CD4) et des molécules d'antigènes des leucocytes humains (HLA) ou complexe majeur d'histocompatibilité (CMH) chez l'humain exprimé en leurs surfaces conduisant à une réduction de la reconnaissance de ces cellules par le système immunitaire (20). Contrecarre les facteurs de restrictions SERINC 3 et 5 (<i>Serine incorporator</i>)(55-59).
Vif	p23	Viral infectivity protein	Essentielle à la production de particules virales infectieuses. Contrecarre les facteurs de restrictions APOBEC3G, 3F et 3H (55, 60-65).
Vpr	p15	Virus protein r	Composant des particules virales interagissant avec la p6 et contribuant au pouvoir infectieux des particules virales. Il intervient sur le cycle cellulaire en activant faiblement la transcription de l'ADN proviral.
Vpu	p16	Virus protein unique	Contribue à la libération efficace de particules virales, contrôle la dégradation des CD4 et module le trafic intracellulaire. Contrecarre le facteur de restriction Tetherin ou BST2 (55, 66-77).
Vpx	p15	Virus protein x,	Interagit avec la p6 dans les particules virales et est impliquée dans les étapes précoces de la réplication virale du VIH-2. Est une composante des particules virales. Contrecarre le facteur de restriction SAMHD1 (55, 78-84).

* nombre correspondant à la taille des protéines (p) ou des glycoprotéines (gp) en 1000 Da.
Tableau adapté avec autorisation. Source: Gurtler L *et al.* 2016 (20, 52-54).

3.5 Cycle de réplication du VIH-1.

Le cycle de réplication du VIH-1 est un processus complexe qui débute par la liaison et la reconnaissance des antigènes viraux par les cellules de l'immunité innée et adaptative de l'hôte au cours de la phase précoce de l'infection. L'interaction entre les cellules de l'hôte et du virus permet le fusionnement entre les membranes cellulaires et virales. Lorsque la particule virale est libérée dans le cytoplasme cellulaire, l'ARN est transcrit en ADN proviral grâce à la transcriptase inverse. Le génome bicaténaire ainsi généré est ensuite inséré dans le génome de la cellule hôte grâce à l'action de l'intégrase devient un provirus. La deuxième phase dite tardive commence par la production d'ARN messager qui produit des protéines qui seront clivées par l'action de la protéase, puis assemblées en virions puis relâchées par la suite hors des cellules. Ces virions vont se transformer en particules virales infectieuses capables d'infecter d'autres cellules dans différents compartiments de l'organisme. La figure V résume l'ensemble des grandes étapes du cycle de réplication du VIH-1.

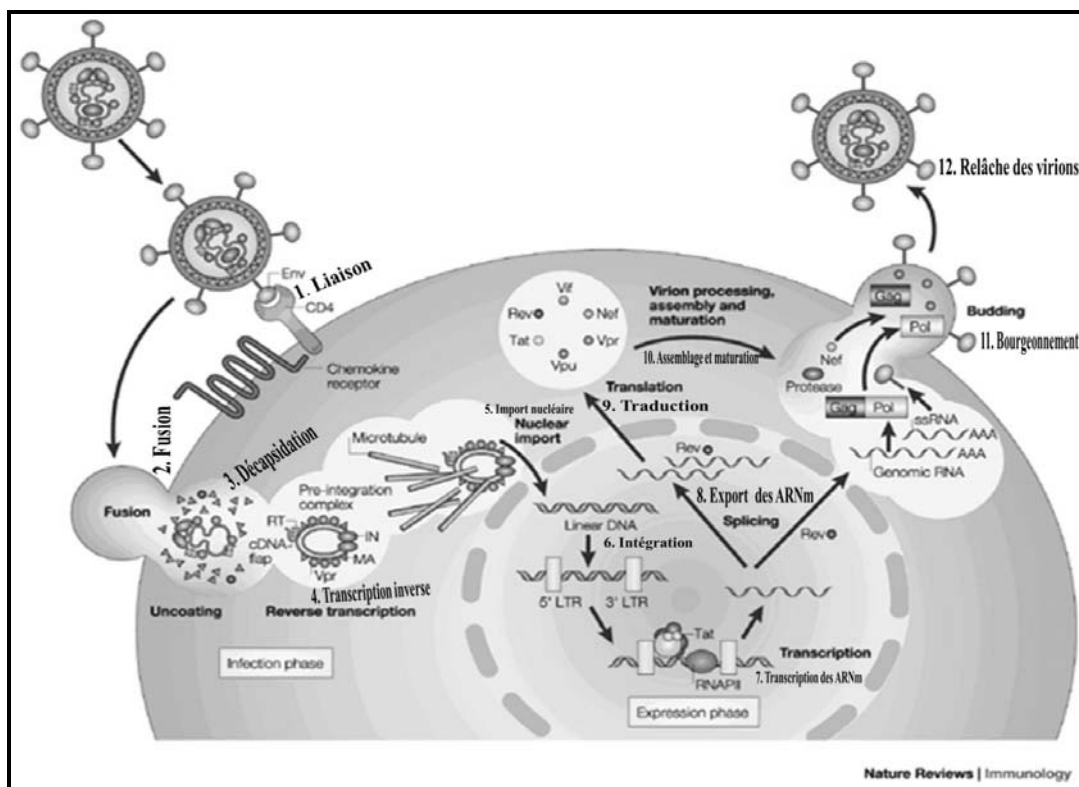


Figure V. Représentation schématique des grandes étapes du cycle de réplication du VIH-1.

Les principales étapes du cycle de réplication du virus sont représentées en 12 points: 1) la liaison, 2) la fusion, 3) la décapsidation, 4) la transcription inverse, 5) l'import nucléaire, 6) l'intégration nucléaire, 7) la transcription des ARN viraux, l'épissage et 8) l'export, 9) la traduction, 10) l'assemblage et la production de nouveaux virions, 11) le bourgeonnement, 12) la libération des virions hors de la cellule suivie de la maturation en particules virales infectieuses. Schéma adapté avec autorisation. Source: Peterlin BM et al. 2003 (85).

3.5.1 Entrée virale.

L'étape initiale d'entrée du virus dans la cellule cible commence par l'interaction entre la glycoprotéine de l'enveloppe (gp120) du virion et son principal récepteur cellulaire, le CD4+ (86-89). Le récepteur CD4 est présent à la surface des lymphocytes T, mais aussi des macrophages (90, 91). L'interaction entre la gp120 et le CD4 induit des changements de conformation au niveau de l'enveloppe virale qui va permettre l'exposition de la boucle V3 de la gp120 (92, 93). Cette boucle exposée restera tout de même protégée par les boucles V1/V2 contre la neutralisation (94, 95) au cours de ce processus (Figure VI). La boucle V3 une fois engagée va interagir avec les principaux corécepteurs du VIH (CCR5 ou CXCR4) (96-99) selon son tropisme (100-102). Une deuxième interaction entre la gp120 et le co-récepteur induit de nouveaux changements de conformation au niveau du trimère d'enveloppe qui conduit à la dislocation de la structure formée et à l'exposition de la super-hélice 1 (*helical heptad repeat 1*, HR1) de l'ectodomaine de la gp41. Ce changement de conformation conduit alors à la présentation de la forme fusogénique de l'enveloppe (103-106) (Figure VI). L'engagement de l'ectodomaine de la gp41 permet l'insertion du peptide de fusion (FP) dans la membrane plasmique de la cellule hôte et favorise la fusion des membranes virales et cellulaires. La fusion membranaire conduit à la libération du cœur viral contenant le matériel génétique (ARN) dans le cytoplasme cellulaire après repliement du paquet de six hélices (*six-helix bundle*) (107, 108) (Figure VI).

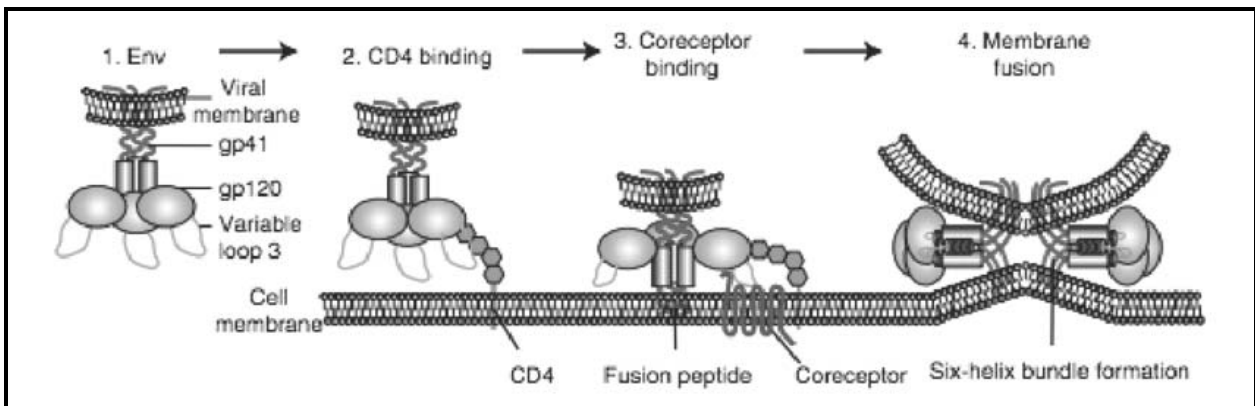


Figure VI. Représentation schématique des différentes étapes d'entrée du VIH-1 dans la cellule cible.

Le schéma décrit le processus (1) d'entrée nécessitant les protéines d'enveloppe comprenant les sous-unités gp120 (bleu) et gp41 (rouge); (2) l'interaction entre la gp120 et le récepteur cellulaire CD4 (noir) qui va induire : (3) un changement de conformation de l'enveloppe permettant le rapprochement de la gp41 avec le corécepteur CCR5 ou CXCR4 (vert) permettant (4) l'insertion du peptide fusion de la gp41 dans la membrane de la cellule. Puis intervient la fusion entre les deux membranes et la libération du matériel génétique dans le cytoplasme cellulaire. Schéma adapté avec autorisation. Source: Si Z *et al.* 2004 (109).

3.5.2 Étapes précoces post-entrée.

Les étapes précoces post-entrée prennent effet après la libération du cœur viral (p24 capsid) dans le cytoplasme, suivi d'une modification structurale de celui-ci (décapsidation) (110, 111). La décapsidation enclenche immédiatement le processus de transcription inverse sous l'action de la transcriptase inverse (RT). Le processus de transcription inverse débute par la formation d'une deuxième copie d'ARN viral. La formation des deux copies permet la synthèse de l'ADN proviral monocaténaire et par la suite en ADN proviral bicaténaire. L'intégrase (IN) catalyse l'intégration de l'ADN viral bicaténaire au génome cellulaire. Les étapes précoces post-entrée du virus dans la cellule hôte et notamment le processus de transcription inverse sont principalement régulées par la capsid. C'est d'ailleurs elle qui contribue à la résistance du VIH face à l'activité antivirale des facteurs de restriction de l'hôte comme TRIM5α RING Domain E3 Ubiquitin Ligase (*Tripartite motif-containing protein 5*) (111-114). Ce composant du cytoplasme de la cellule hôte (115), par son domaine SPRY possédant une activité antivirale se lie aux protéines de la capsid et provoque sa dégradation à travers

des motifs de reconnaissance présents à sa surface nommée (*Complex surface of multimerized CA*) (116, 117). L'interaction entre le domaine SPRY de TRIM5 α et la capsid provoque ainsi son désassemblage rapide et prématuré provoquant une altération de sa structure et prévenant ainsi son action contre le processus de réplication virale (117, 118). Cependant, les fonctions de restriction de TRIM5 α sont spécifiques d'espèce (119). Ceux de l'espèce humaine possèdent moins de fonctions de restriction sur le virus de l'immunodéficience humaine (VIH) et simienne (SIVmac) à cause de la présence d'une mutation (R332P) dans son domaine C terminal (SPRY) (120-122) comparativement à TRIM5 α des primates (119, 120, 123). Parmi ces primates, on distingue notamment les macaques rhésus (*Rhesus monkey*), les singes de l'Ancien Monde (Old World monkeys), les singes verts d'Afrique (*African green monkey*), et des singes-araignées (*spider monkey*) (119, 120, 123). Au niveau de ces espèces, des mutations (H87Q) (124) de la capsid des rétrovirus qui infectent ces primates altèrent les fonctions de restriction de TRIM5 α (125-129). La protéine accessoire Vif joue aussi un rôle important dans le processus de transcription inverse. En dégradant APOBEC3G/3F et en empêchant son incorporation dans les virions, Vif contrecarre l'activité antivirale des APOBEC durant la transcription inverse. Pourtant, les protéines APOBEC sont impliquées dans la transcription inverse de l'ARN VIH en ADN. Elles induisent des hypermutations (cytidine à uridine dans le brin négatif de l'ADNss naissant qui est traduit en G-à-A dans l'ADN complémentaire +). Ce facteur de restriction est développé par le système immunitaire de l'hôte et intervient au cours du processus d'assemblage des protéines virales afin de bloquer le processus de réplication (130, 131). Malheureusement, cette fonction est contrecarrée par la protéine virale Vif qui bloque l'assemblage de la protéine APOBEC dans les nouvelles particules virales.

3.5.2.1 Décapsidation.

La décapsidation est un processus qui conduit à la libération d'un complexe de transcription inverse nommé : *reverse transcription complex (RTCs)*. Ce complexe est constitué de : 1) deux copies d'ARN viral génomique; 2) d'ARN de transfert 3 pour la lysine (ARNt(Lys3) ou (tRNA(Lys3) (132-134) qui intervient comme amorce pour l'initiation de transcription inverse; 3) de la protéine Vpr; et 4) de facteurs cellulaires (135). La transcription inverse intervient

après une phosphorylation de la protéine de la matrice (MA) par un MAPK (*Mitogen-activated protein kinase*) (136, 137) et des protéines cellulaires (CypA (cyclophiline A) et virales (Vif et Nef) (112, 138-140) a lieu environ une heure après la fusion (140, 141) (Figure VII). Cependant, de nouvelles évidences démontrent qu'une décapsidation partielle a lieu dans le cytoplasme et que la totale se ferait au pore nucléaire après transport de la capsidie contenant le RTC et l'ADN proviral (142).

3.5.2.2 Transcription inverse.

La transcriptase inverse (RT) comporte plusieurs activités associées à des domaines bien caractérisés: une activité ADN polymérase ADN négatif (-) et ARN (-) dépendant qui permet la synthèse d'un génome ADN bicaténaire à partir d'un ARN monocaténaire, ainsi qu'une activité d'endonucléase ARN (RNase H) qui permet la dégradation de la matrice de l'ARN et des amorces de transfert pour la lysine, ARNt (Lys3) au cours de la synthèse d'ADN viral (143-145). Comme représenté dans la figure VII, le processus de transcription inverse peut se résumer en 8 principales étapes (146): 1) la synthèse d'un brin d'ADN de polarité négative (ADN-) à partir de l'extrémité 3' de l'ARNt^{3lys} (132-134) qui s'hybride partiellement à la région PBS (Primer Binding Site) de l'ARN génomique, suivi de la synthèse d'un brin d'ADN-stop de polarité négative (-) ssDNA; 2) l'extrémité 5' de l'ARN (hybride ARN/ADN) du génome se dégrade par l'action de la RNase H de la RT conduisant à la translocation de l'ADN-stop à l'extrémité 3' de l'ARN génomique; 3) un premier saut de brin intervient lorsque la séquence R de l'ADN (-) s'hybride avec celle de l'extrémité 3' de la région R de l'ARN viral permettant la poursuite de la synthèse du brin ADN (-); 4) la synthèse du brin ADN (-) se poursuit avec dégradation de l'ARN viral parallèlement à la synthèse de l'ADN, à l'exception de la séquence PPT (*PolyPurine Tract*); 5) l'ARN PPT non dégradé sert d'amorce à la synthèse du brin (+) de l'ADN viral jusqu'à l'extrémité U5; 6) la séquence PBS du brin (+) est synthétisée à partir de l'ARNt ayant servi d'amorce, mais celle-ci se dégrade en même temps par l'action de la RNase H, générant ainsi un deuxième saut de brin nommé : *second-strand transfer* (146); 7) la synthèse du brin d'ADN (+) se poursuit (élongation) et se termine dans la région CTS (*central termination signal*) impliquant le déplacement de 88 ou 98 séquences nucléotidiques nommées DNA *flap* qui se superposent au niveau de l'ADN bicaténaire (147). Les DNA *flap*

interviennent dans le processus d'import nucléaire de l'ADN viral (148-151); 8) la transcription inverse se termine au niveau des extrémités LTR par la formation de deux séquences terminales identiques (U3-R-U5) donnant lieu à un ADN bicaténaire.

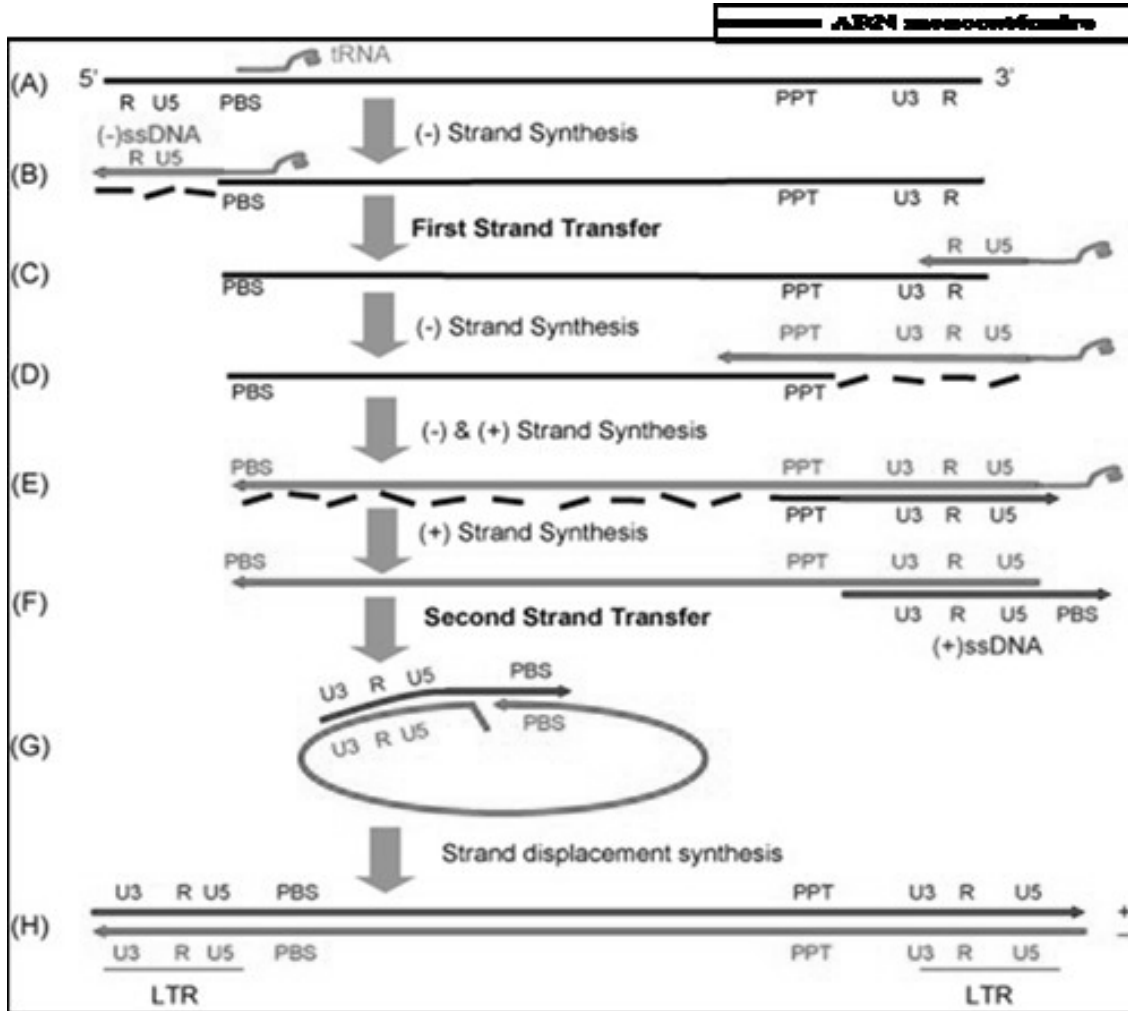


Figure VII. Représentation schématique des différentes étapes de la transcription inverse de l'ARN.

Le processus de transcription résumé dans ce schéma illustre : **(A)** : Initiation de la transcription inverse, début de la synthèse d'un brin ADN de polarité négative) à partir de l'extrémité 3' de l'ARNt3lys (vert); **(B)** : Synthèse d'un brin d'ADN-stop de polarité négative (-) ou ssDNA (rouge); **(C)** : 1^{er} saut de brin, élongation du brin (-) d'ADN; **(D)** : Poursuite de la synthèse du brin ADN (-) avec dégradation de l'ARN viral parallèlement à la synthèse de l'ADN, à l'exception de la séquence PPT (PolyPurine Tract); **(E)** : L'ARN PPT non dégradé sert d'amorce à la synthèse du brin (+) d'ADN viral jusqu'à l'extrémité U5 en rouge; **(F)** : Synthèse de la séquence PBS (Primer Binding Site) du brin (+) à partir de l'ARNt suivi du deuxième saut de brin (second-strand transfer);

(G) : Poursuite de la synthèse du brin d'ADN (+) (élongation) qui se termine dans la région CTS (*central termination signal*); **(H)** Achèvement de la synthèse complète de l'ADN viral bicaténaire au niveau des extrémités LTR et génération de deux séquences terminales (U3-R-U5) identiques Schéma adapté avec autorisation. Source: Basu VP et al. 2008 (146).

3.5.2.3 Intégration de l'ADN viral dans le génome cellulaire.

La décapsidation est immédiatement suivie de la formation d'un complexe de préintégration (PIC). Il s'agit d'un filament d'une taille de 9,7 kb constitués d'un ADN bicaténaire, des facteurs cellulaires (*Barrier-to-auto-integration factor1*, BAF), *Cleavage and Polyadenylation Specificity Factor 6*". (CPSF6), *cellular transportin 3 protein* (TNPO3), *kinesin-1 motor* (KIF5B) et *nuclear pore* (Nup358), et des facteurs viraux (Vpr, MA, CA et IN) (152-155). Ces éléments du PIC permettent l'insertion du génome viral dans le génome cellulaire à travers le noyau grâce à un complexe de pore nucléaire (*Nuclear pore complex*, NPC) et qui s'effectue sans dommage pour l'enveloppe (142, 156, 157). Une fois intégré au génome cellulaire, le provirus est transcrit en ARN messager (ARNm) permettant la production de protéines virales et d'ARN génomique qui seront traduits, épissés et assemblés pour donner naissance à de nouveaux virions. Lors de l'import du PIC au niveau du noyau, la capsid est mise à contribution. Celle-ci interagit avec les facteurs cellulaires requis pour le transport nucléaire et l'infectiosité du VIH-1(112, 158-160). Le processus d'intégration de l'ADN viral au génome cellulaire est catalysé par l'intégrase et nécessite aussi l'implication de cofacteurs de l'hôte comme : *Lens epithelium-derived growth factor* (LEDGF), *Integrase Interactor 1* (INI1), *Barrier to autointegration factor* (BAF) et de la capsid, (CA) (154, 155, 161-166). Ces cofacteurs déterminent le site d'intégration de l'ADN viral, mais aussi de la réparation des dommages causés au niveau du site d'insertion par les protéines de l'ADN viral, les enzymes (polymérase, nucléase et ligase) et les facteurs cellulaires, mettant ainsi fin au processus d'intégration (157, 167-169). La réparation des dommages cellulaires clôt le processus d'intégration.

3.5.3 Étapes tardives post-entrée.

Une fois l'ADN viral intégré dans l'ADN du génome de la cellule hôte, il devient une de ses composantes (ADN proviral). À partir de ce moment le contrôle de certains mécanismes de synthèse de la cellule hôte permettent l'expression du provirus de la même façon que l'expression des ARNs et des protéines cellulaires. La transcription des ARN messager (ARNm) viraux intervient sous contrôle du promoteur dans la région LTR localisée à l'extrémité 5' du génome viral. La transcription de l'ARN viral se fait par l'ARN polymérase II (RNAPII) qui interagit avec de multiples facteurs dont *Human transcription factor IIB* (TFIIB) pour déterminer le site d'initiation de la transcription (170, 171).

D'une taille de 86 à 101 acides aminés, Tat, une protéine régulatrice essentielle, joue un rôle majeur dans la transcription de l'ARN viral (172, 173). Elle régule le taux de synthèse des ARN viraux et des protéines virales au cours de l'infection et se trouve dans les noyaux et les nucléoles des cellules infectées par le VIH (174, 175). Encodé par deux exons distincts, Tat est une protéine de liaison à l'ARN qui renforce l'activité de l'ARN polymérase (ARN POL-II) en cours de synthèse par augmentation de la production des ARNs viraux en se liant à une tige-boucle nommée *Trans-activating response element* (TAR) localisée dans la terminaison 5' des ARN du VIH (175-177). La protéine Tat intervient spécifiquement dans la phase d'élongation de la transcription virale formant un complexe transcription efficace qui favorise la production complète des transcrits (178, 179). Tat recrute pour ce faire les protéines cellulaires cycline T1(180) et la protéine kinase-9 cycline-dépendante (Cdk9) (181, 182) qui constituent le facteur d'élongation de la transcription b (pTEF-b) (183, 184) pour ainsi catalyser l'hyperphosphorylation du domaine C terminal de l'ARN Pol-II augmentant d'une façon importante (au moins 1000 fois) son activité (177, 179, 185, 186).

La transcription produit l'ARN génomique multiépissé qui va être exporté de façon indépendante hors du noyau par les mécanismes de synthèse de la cellule hôte. L'ARN multiépissé est traduit par les mécanismes de réductase cellulaire (protéines Rev, Nef et Tat). Tat et Rev sont importées dans le noyau où Tat va augmenter fortement l'élongation de la transcription par le recrutement du facteur pTEFb (183, 184). L'exportation des ARNm mono ou non épissés hors du noyau est régulée par le taux de protéines Rev qui se lie à la protéine

pour permettre l'export des ARNs qui possèdent RRE (*Rev response element*) (187-189). Les principales étapes tardives post-entrée sont résumées dans la figure VIII et comprennent 6 grandes étapes: 1) la transcription des ARN viraux et l'export nucléaire; 2) la traduction et la production de protéines virales; 3) l'assemblage; 4) le bourgeonnement des nouveaux virions; 5) la relâche des virions hors de la cellule infectée, 6) la maturation des virions et leur transformation en particules virales infectieuses grâce aux composants viraux et cellulaires (190). Ces différentes étapes sont fondamentales et indispensables pour la finalisation du cycle de réplication du VIH (191).

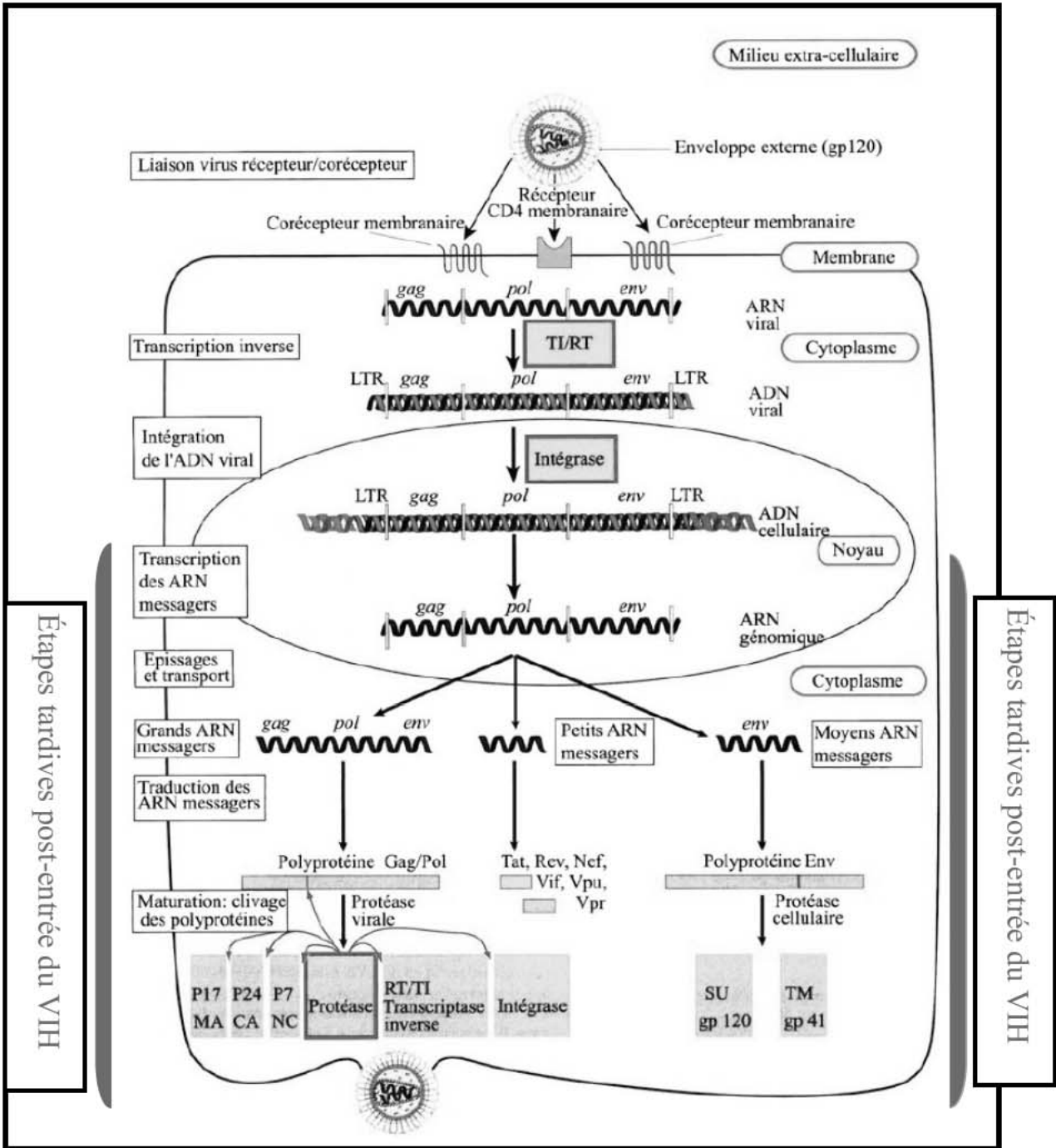


Figure VIII. Représentation schématique des différentes étapes tardives post-entrée du VIH.

Figure adaptée avec autorisation, traitée de Virologie médicale. J.M. Hureau, éditions Estem, 2003.

3.5.3.1 Production de protéines virales, traduction et assemblage.

Les ARNm relâchés hors du noyau vont être traduits en protéines virales accessoires, enzymatiques ou structurelles dont certaines seront incorporées dans de nouvelles particules virales (22, 192). L'encapsidation de l'ARN génomique et l'implication des protéines Vif, Vpr et Nef déclenchent le processus d'assemblage des nouvelles particules virales (virions). Puis suivra la libération des nouveaux virions par bourgeonnement hors de la membrane cytoplasmique de la cellule infectée (193, 194) sous l'influence du sous-domaine p6 du précurseur *gag* (195, 196).

3.5.3.2 Production et libération des nouveaux virions.

La production et la libération de nouveaux virions nécessitent des éléments de la machinerie cellulaire. Plus particulièrement du facteur ESCRT (*endosomal sorting complex 1-3 required for transport*) qui contribue à la rupture des membranes cellulaires et virales afin de favoriser la libération des particules hors de la cellule infectée (197-206). La polyprotéine structurale virale *gag* joue un rôle central dans le processus d'assemblage et de libération des nouveaux virions (22, 207). Elle comprend plusieurs domaines qui sont impliqués dans la liaison à la membrane plasmique, la multimérisation, la liaison à l'acide nucléique et l'interaction avec les composantes cellulaires impliquées dans la libération des nouveaux virions au niveau des bourgeons de la cellule infectée. Le gène *gag* dirige aussi la formation des particules pseudo-virales nommée *virus like protein* (VLP) en l'absence d'autres protéines virales (191).

3.5.3.3 Maturation et génération de particules virales infectieuses.

La maturation du virion a lieu au même moment ou après le bourgeonnement au niveau de la membrane plasmique (208). Elle est médiée par le clivage des protéases et des polyprotéines Gag et Gag-Pro-Pol sur une dizaine de sites différents conduisant à la production des protéines virales (MA, CA, NC, p6, PR, RT et IN) (190, 209). Les virus matures seront alors capables d'infecter la membrane plasmique des nouvelles cellules hôtes. L'initiation et la finalisation du cycle de réplication nécessitent en effet deux copies d'ARN (+), de nombreuses protéines virales et des facteurs cellulaires, des enzymes et des molécules de synthèse. Il s'agit de la polyprotéine de Gag, les enzymes de la protéase (PR), de

la reverse transcriptase (RT) et de l'intégrase (IN), de l'ARN de transfert (tRNA^{Lys}) et trois (3) molécules de synthèse de l'ADNc (22). L'incorporation des protéines d'*env* confère le pouvoir infectieux à la particule virale (85, 191).

3.6. Structures et fonctions des glycoprotéines de l'enveloppe du VIH-1.

3.6.1 Généralités et fonctions.

L'enveloppe du VIH-1 est la première à interagir avec les récepteurs (lymphocytes TCD4+, macrophages) et corécepteurs (CCR5 et CXCR4) de la cellule hôte. Après avoir subi de nombreux changements de conformation elle permet la fusion entre les membranes cellulaires et virales conduisant à l'entrée du noyau viral (capside) dans le cytoplasme cellulaire (210, 211) (Figure IX). L'entrée du virus dans la cellule cible peut se faire par deux mécanismes : soit par fusion directe avec la membrane plasmique, indépendamment du pH, ou bien soit par endocytose médiée par la clathrine qui est une protéine structurale, constituant l'enveloppe de certaines vésicules. Ensuite survient la fusion avec la membrane endosomale, même à faible pH (212-214). L'enveloppe du VIH-1 se présente comme un complexe trimérique constituant un spicule à la surface du virus (107, 215, 216). Elle dérive d'un précurseur protéolytique fortement glycosylé et oligomérisé dans le réticulum endoplasmique appelé glycoprotéine gp160. Elle est soumise à un processus de glycosylation dans le complexe de Golgi puis clivée par des furines ou des analogues furines-protéases cellulaires pour donner la glycoprotéine de surface SU gp120 et la glycoprotéine transmembranaire TMgp41 (217-221). Le traitement protéolytique de la gp160 est indispensable pour l'activation de l'activité fusogène de l'enveloppe donc essentiel pour l'infectiosité virale (210). Elle est d'une taille d'environ 856 acides aminés (217, 218). Les différentes structures dérivées de la synthèse de la gp160 sont résumées dans la figure IX.

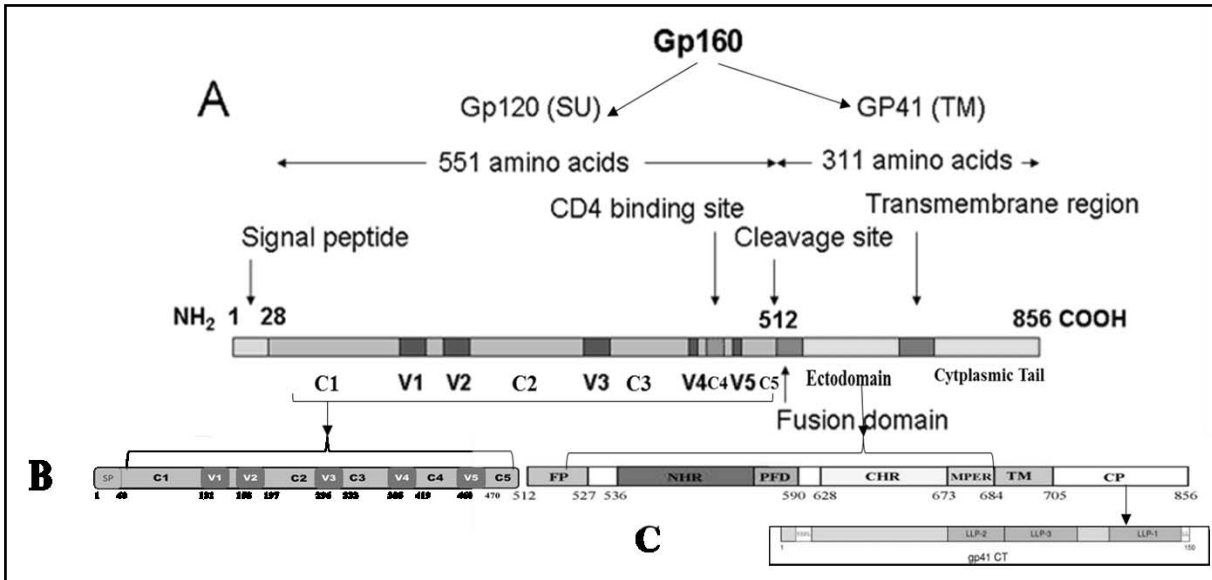


Figure IX. Schéma de la structure de la glycoprotéine gp160 et de ses dérivés, gp120 et gp41.

Le schéma illustre l'organisation linéaire des différentes glycoprotéines et les principaux sous-domaines qui les composent et leurs localisations dans l'enveloppe du VIH en référence de la numérotation de la gp160 (enveloppe) de la souche VIH-1 HXB2. (A) Structure du domaine linéaire de la gp160 et de ses dérivés : gp120 (la protéine de surface) et gp41 (la protéine de fusion transmembranaire ; (B) localisation des sous-domaines de la gp120 et de la gp41 (C) les sous-domaines de la queue cytoplasmique de la gp41. Abréviations : SP : peptide signal (*signal peptide*) ; C : Constant domain ; V : *variable domain* ; FP : peptide de fusion (*fusion Peptide*) ; NHR : *N-terminal heptad repeat* ; PFD : *pocket-forming domain* ; CHR : *C-terminal heptad repeat* ; CP : domaine/queue cytoplasmique (*cytoplasmic tail*) ; TM : domaine transmembranaire (*transmembrane domain*) ; MPER : *membrane-proximal external region*. Schéma adapté avec autorisation. Sources: Checkley MA, et al. 2011 ; Hunt R. 2016 ; Yu F, et al. 2013 (210, 222, 223).

Résumé de la localisation des sous-domaines de l'enveloppe en référence à la numérotation de la séquence d'acides aminés de la gp160 du VIH-1 HXB2 (résidus 1- 856) de la souche HXB2 (224): env GP120 (résidus, 1-511), SP (1-30), C1 (31-130), V1V2 (131-157+158-196), C2 (197-295), V3 (296-331), C3 (332-384), V4 (C385-C418), C4 (419-459), V5 (460-469) et la C5 (470-511). GP41 (345 acides aminés, résidus (512-856), ectodomaine (résidus 512-677) : peptide de fusion (résidus 512-527), N-terminal heptad repeat (NHR, résidus 546-581), C-terminal heptad repeat (CHR, résidus 623-659), MPER (résidus 673-684) ; domaine transmembranaire (TM, résidus 678-689) ; queue cytoplasmique (CP, résidus 700-856) : LLP-2 (résidus 768 - 788), LLP-3 (résidus 789 - 815), et LLP-1 (résidus 828 - 856) (225, 226).

3.6.2 Glycoprotéine de surface, SUGp120.

3.6.2.1 Structure et fonctions.

Dérivée du clivage de la gp160, la gp120 est une structure moléculaire d'environ 511 acides aminés incluant les séquences du peptide signal (*signal peptide*, SP). Elle comprend un domaine dit interne et un autre dit externe, reliés entre eux par un feuillet β interdomaine nommé, *bridging sheet*. Le premier est plus proche de la gp41 et des extrémités N et C-terminales de la gp120. Le deuxième domaine est beaucoup plus exposé à la surface de l'enveloppe virale et regroupe la plupart des sites de glycosylation. Entre les deux domaines et le feuillet β se trouve le site de liaison au CD4, une structure assez conservée par de nombreuses souches virales. L'interaction de la gp120 à la surface du récepteur CD4 initie la liaison du virus. Cette interaction va engendrer des changements de conformation en son sein (227) provoquant le déplacement des boucles V1/V2. Un repositionnement du domaine interne formant le feuillet β interdomaine va initier la sortie du site de liaison du corécepteur (CCR5 ou CXCR4) (228) grâce à la boucle V3 et des résidus du sous-domaine conservé C4 (229). L'insertion du peptide de fusion hydrophobe dans la membrane cellulaire dérive donc de la formation du complexe gp120-CD4-CCR5/CXCR4 (Figure X) (229).

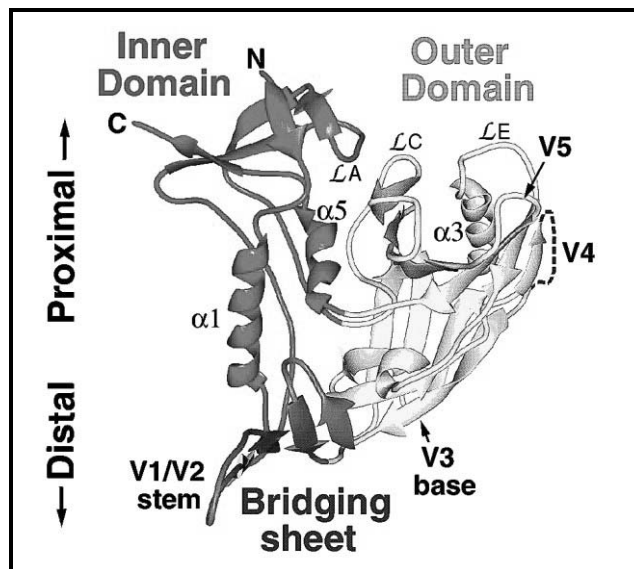


Figure X. Structure du noyau de la glycoprotéine gp120 de l'enveloppe du VIH-1.

Sur la figure sont illustrés le domaine interne en rouge, le domaine externe en jaune et le feuillet en bleu de la glycoprotéine GP120 de l'enveloppe du VIH-1 Figure adaptée avec autorisation. Source: Wyatt R et al.1998 (230, 231).

La glycoprotéine gp120 est constituée de cinq sous-domaines et comprend des résidus cystéine (C) hautement conservée (C1-C5) qui forment des ponts disulfures très importants pour la formation et le maintien de la structure tertiaire de l'enveloppe (232-234). Ces domaines conservés interagissent avec 5 sous-régions hypervariables et fortement glycosylées appelées boucles V1-V5 (2, 210, 232-235) (Figure XI). Les boucles variables sont indispensables au maintien de l'intégrité structurelle et l'entrée du virus dans la cellule hôte. Elles contiennent également les déterminants de l'interaction avec les récepteurs et corécepteurs du VIH et de la résistance à la neutralisation par les anticorps neutralisants à large spectre (bnAbs) (86, 96, 236). Plus spécifiquement, les boucles V1/V2 participent à la stabilisation de la gp120 au sein du trimère de l'enveloppe (229). Elles influencent, entre autres, la liaison du corécepteur, participent à la résistance à la neutralisation au stade précoce (en masquant les épitopes de neutralisation) et à la stabilisation du trimère de l'enveloppe (96). La charge positive nette de la loop V3 module l'utilisation du corécepteur associé au tropisme (CCR5 ou CXCR4) (96). Les domaines V4-V5 contribuent à la résistance à la neutralisation en masquant les épitopes de neutralisation. La V5 participe également à la préparation du site de liaison du CD4 (CD4bs) (96). La gp120 comprend également 9 ponts disulfures et entre 20 à 35 sites de N-glycosylation, qui contribuent par ailleurs pour la moitié de sa masse moléculaire (232, 237, 238). Elle comprend aussi des sites de O-glycosylation (239). Les N-glycosylations servent de bouclier contre l'action des anticorps neutralisants. Cependant, le rôle fonctionnel des O-glycosylations demeure ambigu (240). La glycosylation O est caractérisée par la présence d'un résidu thréonine (Thr) hautement conservée localisée près de l'extrémité C-terminal de la gp120. Elle a été décrite comme fortement impliquée dans l'accroissement de l'infectiosité virale (240). Cependant, des études récentes démontrent clairement que même en absence de O-glycosylation, le VIH est capable d'infecter ses cibles naturelles (240). Cependant, les auteurs émettent certaines réserves, car

des facteurs enzymatiques et/ou cellulaires pourraient modifier ou affecter le niveau d'implication des glycosylations O in vivo au cours de l'infection naturelle (240).

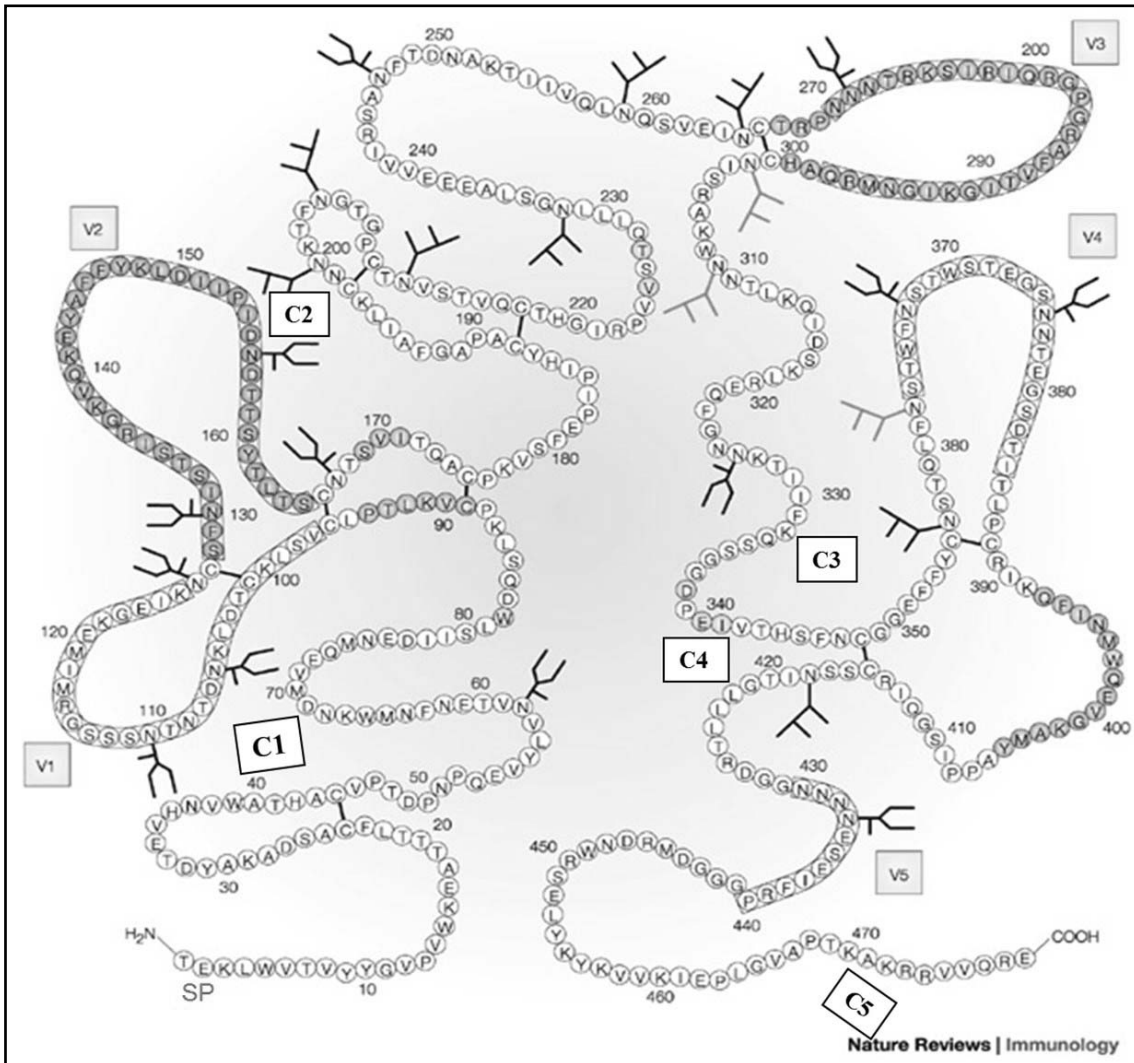


Figure XI. Représentation schématique de la structure de la glycoprotéine de surface SUgp120 de l'enveloppe du VIH-1.

La figure représente l'organisation des domaines de la gp120, les 5 régions conservées (C1-C5), les 5 domaines variables (V1-V5) et le peptide signalé (SP). Figure reproduite avec autorisation de : Gallaher WR et al. 1989 et Zolla-Pazner S et al. 2004 (241, 242).

3.6.3 Glycoprotéine transmembranaire, TMgp41.

3.6.3.1 Structure et fonctions.

La glycoprotéine transmembranaire, TMgp41 dérive du clivage de la gp160. Elle contient les domaines impliqués dans la formation de la structure trimérique de l'enveloppe et de fusion de l'enveloppe virale dans la membrane de la cellule hôte (107, 215, 243). D'une taille de 345 acides aminés (aa), elle comprend trois domaines majeurs (Figure IX et XII) :

1) **un domaine extracellulaire nommé ectodomaine (ECD)** d'une taille de 180 acides aminés, quatre sites de N-glycosylation et un pont disulfure. Ce dernier est impliqué dans la formation d'une petite boucle; 2) **un domaine transmembranaire (TMD)** composé d'une vingtaine d'acides aminés. Il contribue à l'ancrage de la glycoprotéine transmembranaire gp41 dans la bicouche phospholipidique de la membrane plasmique (241); 3) **Une queue ou domaine cytoplasmique (C-terminal cytoplasmic tail, CT)**; d'une taille assez longue de 150 acides aminés. Elle régule le trafic intracellulaire, facilite l'assemblage des protéines virales et module l'expression des glycoprotéines de surface de l'enveloppe du VIH-1 (244, 245). La queue cytoplasmique est constituée de trois régions organisées en hélice : la première est localisée en C-terminal et connue sous le nom de LLP1 (*Lentivirus lytic peptide 1*) et la deuxième et troisième situées au niveau central sont connues sous les noms de LLP2 et LLP3 (246) (Figure IX). Ces régions en association avec la membrane plasmique, déstructure la bicouche lipidique de cette dernière pour ainsi faciliter l'incorporation des glycoprotéines d'enveloppe du VIH-1 (247-249). Les domaines transmembranaires, TMD et cytoplasmiques, CTs contiennent des motifs ou des facteurs impliqués dans la réplication virale (2, 243). Le domaine extracellulaire (ECD) comprend les déterminants majeurs de la fusion, dont un peptide de fusion (16 aa), une région hydrophobe formant une structure α hélicoïdale torsadée (*α -helical coiled coils*) connue sous les noms d'*heptad-repeat régions 1 (HR1) et 2 (HR2)* et de la région externe de la gp41 proximale de la membrane (MPER, *membrane-proximal external region*).

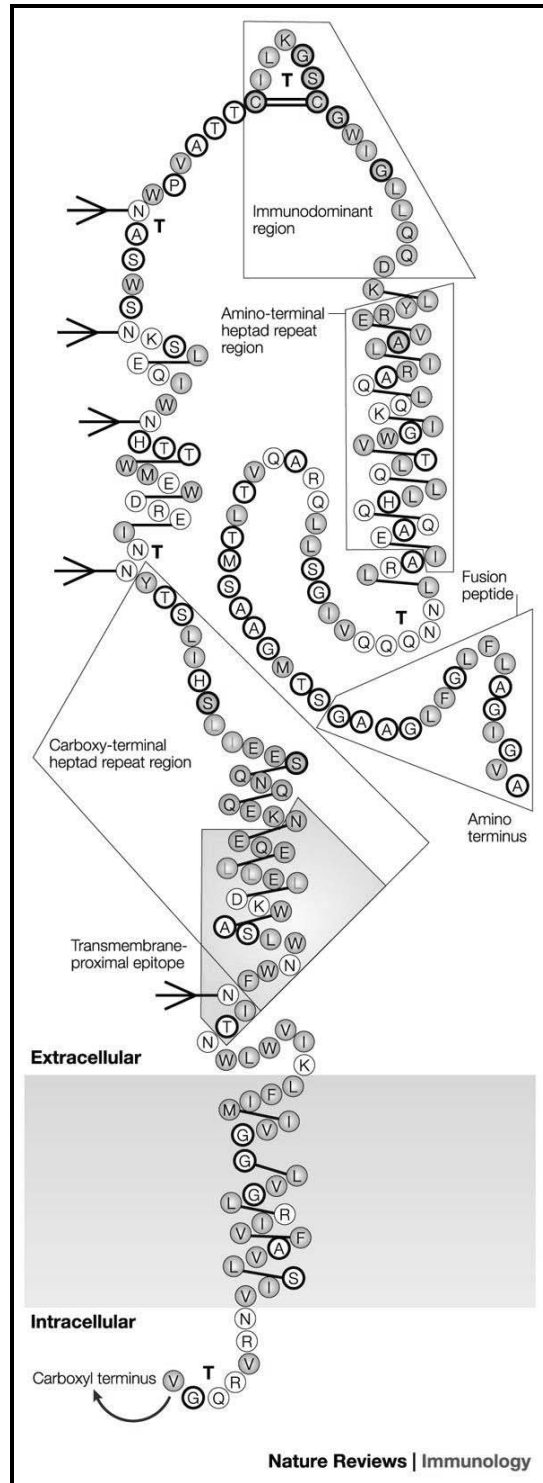


Figure XII. Représentation de la structure de la TMp41 de l'enveloppe du VIH-1.

La figure ci-dessus illustre les principaux domaines de la gp41. Figure reproduite avec autorisation. Sources : Gallaher WR *et al.* 1989 ; Zolla-Pazner S *et al.* 2004 (241, 242).

Du point de vue fonctionnel, la gp41 dirige l'incorporation de l'enveloppe dans les virions à la suite de son interaction avec la matrice par l'intermédiaire de la protéine TIP47 (*Tail interacting protein 47 kDa*) (245, 250). Les régions HR1 et HR2 sont liées par un pont disulfure dans une boucle hydrophile. L'interaction entre elles enclenche le processus de fusion de l'enveloppe sur la membrane cellulaire. Elles sont également responsables de la formation du trimère de la protéine d'enveloppe. Normalement enfoui dans le complexe quaternaire de la gp120/gp41, le peptide de fusion s'expose et pénètre la membrane cellulaire causant sa déstructuration suite à la liaison de la gp120 au récepteur CD4 et au corécepteur CCR5 ou CXCR4 (251). Ce processus induit la formation d'un faisceau de six hélices qui stabilise le pore préalablement formé par l'insertion du peptide de fusion dans la membrane cellulaire, entraînant ainsi le rapprochement entre les membranes virales et cellulaires pour initier la fusion (215, 243, 252-255). La formation du faisceau favorise la fusion complète du peptide et permet la libération du génome viral dans le cytoplasme de la cellule (252).

3.7 Histoire naturelle de la maladie.

3.7.1 Généralités

L'infection par le VIH est d'évolution chronique et est incurable jusqu'à présent. Néanmoins, les traitements antirétroviraux sont disponibles et ont la capacité de réduire la charge virale plasmatique chez un individu infecté jusqu'à un niveau indétectable empêchant l'évolution clinique de la maladie. Il n'existe pas de vaccin préventif et/ou thérapeutique permettant de prévenir la transmission en dépit des moyens de prévention traditionnels (relations sexuelles protégées, prophylaxie préexposition, PrEP). L'évolution naturelle de la maladie se déroule principalement en 4 phases basées sur les manifestations cliniques : une primo-infection, incluant une phase aigüe et asymptomatique, et une phase chronique comprenant une phase symptomatique et une phase SIDA (Figure XIII).

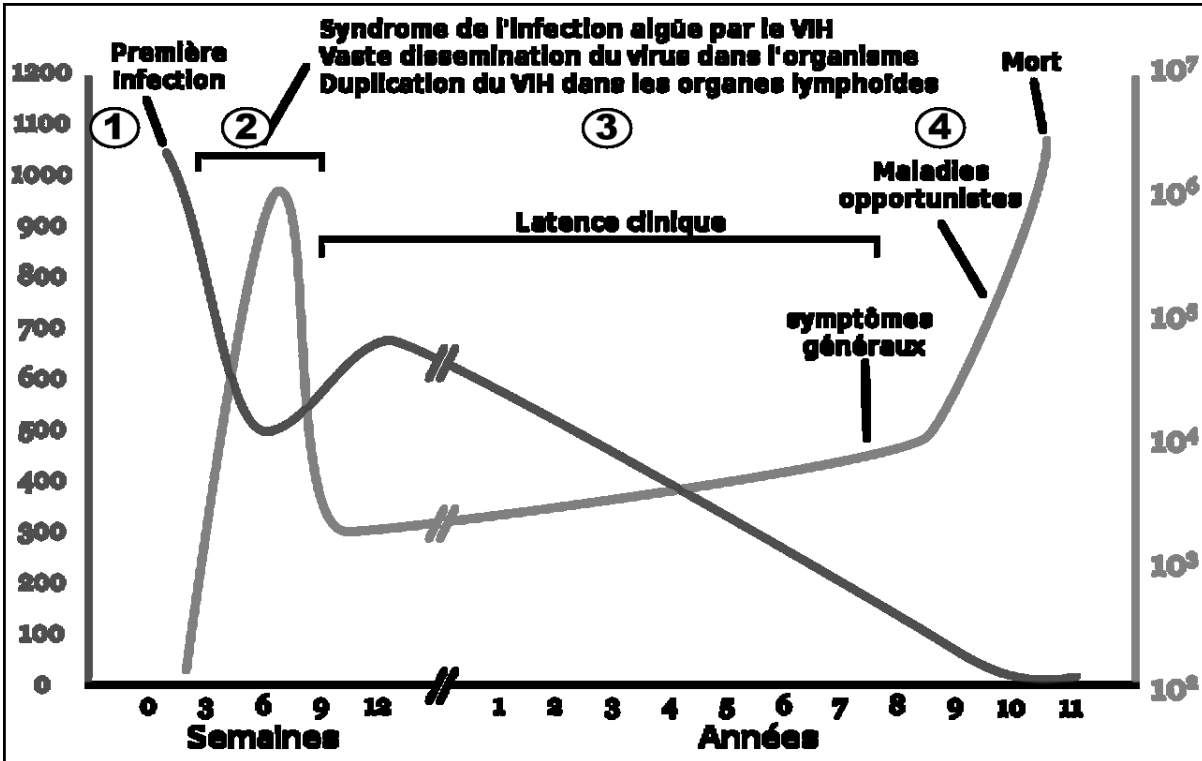


Figure XIII. Représentation schématique de l'évolution naturelle de l'infection par le VIH-1 et évolution des marqueurs biologiques.

Le diagramme décrit l'évolution des deux principaux marqueurs biologiques (Taux de lymphocytes TCD4+ et charge virale plasmatique) au cours de l'évolution naturelle de la maladie. Le nombre de lymphocytes TCD4+/mm³ est en bleu et le nombre de copies d'ARN viral par ml en rouge représenté respectivement à gauche et droite de l'axe des Y. La durée de chaque phase est représentée sur l'axe des X. Les différentes phases de l'infection sont notées de 1 à 4. Diagramme reproduit avec autorisation. Source : Celum CL *et al.* 1991 (256).

Cependant, certains sujets naïfs de thérapies antirétrovirales et bien que porteurs de la maladie et présentant une répllication active du virus dans le sang, ne manifestent aucun symptôme clinique associé à l'infection, ni d'immunodépression sur le plan biologique (stabilité du nombre de CD4 ≥ 500 cellules/ μ l) pendant une longue période (7-10 ans) (257-260). Ces sujets sont dits non-progresseurs à long terme (NPLT) ou *long-term nonprogressors* ou porteurs asymptomatiques à long terme (ALT) (257-259). En Europe et aux États-Unis, la proportion de ces sujets varie entre 5 et 8% des personnes infectées (258, 261, 262), et autour de 9% rapportées dans certaines études en Afrique (263). Certains sujets en revanche bien qu'ils soient infectés par le VIH et naïfs de traitement antirétroviral arrivent à contrôler

leur virémie (charge virale indétectable) pendant plus de 12 mois après trois séries d'évaluation (1, 9 et 12 mois) (264) avec maintien du niveau des CD4 élevé (265, 266). Ces sujets sont dits contrôleurs élités ou *HIV controllers* (HICs) (265, 266). La prévalence des sujets infectés par le VIH-1 dits contrôleurs élités est généralement faible parmi les personnes infectées dans des cohortes publiées (<1%) (263, 266, 267).

Les sujets non-progresseurs à long terme présentent plusieurs caractéristiques qui militeraient en faveur de la non-progression de l'infection par le VIH (268). Certains présentent naturellement une délétion de 32 nucléotides au niveau du CCR5 (CCR5- δ 32) qui rend le corécepteur du VIH non fonctionnel procurant ainsi une protection naturelle contre le VIH présent chez certaines populations (269-272). En effet; les Européens présentant cette rare mutation sont généralement protégés contre la peste. Cette approche a été utilisée pour une transplantation allogénique d'un patient hétérozygote (CCR5 VIH+) par de cellules souches hématopoïétiques d'un donneur compatible et homozygote pour la mutation CCR5 delta 32. Cette allogreffe a permis de contrôler complètement la réplication virale chez le patient receveur préalablement infecté par le VIH et aussi atteint d'une leucémie myéloïde aiguë (*patient de Berlin, Timothy Ray Brown*). Il s'agit du premier cas de guérison clinique du VIH documenté à ce jour.(273, 274). Un deuxième cas de guérison de guérison du VIH-1 par transplantation de cellules souches hématopoïétiques a été rapporté (*patient de Londres*) (275).

D'autres mutations (CCR2-64I, SDF1-3'A, et RANTES-28 G) sont également associées à la non-progression des maladies chez un petit nombre de sujets non-progresseurs à long terme (269-272). Un niveau élevé d'ARN messager (mRNA) du facteur de restriction (APOBEC3G) a été retrouvé chez les patients VIH+ NPLT comparativement aux sujets non infectés ou infectés, mais symptomatiques (269, 272, 276, 277). Ces observations ont été retrouvées dans une cohorte de sept patients VIH+ NPLT assemblée par le Dr. Spyros A. Kalams au : « Vanderbilt-Meharry Developmental Center for AIDS Research (CFAR), HIV Immunopathogenesis Core » Nashville, USA (276). La non-progression de la maladie chez les sujets NPLT serait qu'ils ont été infectés par de virus défectifs donc dépourvus de capacités de réplication (fitness) (269, 272). En dehors de la présence d'APOBEC3G, une délétion dans

la région où le gène Nef et la séquence terminale longue répétée (LTR) se chevauchent dans le génome du VIH-1 constitue un autre cas de transmission par un virus defectif (*Sydney Blood Bank Cohort*) (278). Les sujets NPLT présenteraient également un niveau élevé d'anticorps neutralisants à large spectre et/ou de réponses T cytotoxiques, mais aussi de facteurs génétiques particuliers (allèles codant les HLA-B27 et HLA-B57) (269). Chez les enfants âgés de 1 à 2 ans, une proportion des cellules T CD8+HLA-DR+ (<5%) est le résultat d'une baisse de l'activation des cellules T et de l'inflammation prédisant la non-progression de l'infection (269, 272). Un niveau élevé de certains autoanticorps circulants dirigés contre un épitope conformationnel du domaine extramembranaire du CCR5 entraînant une régulation du corecepteur est associé aussi à la non-progression de l'infection par le VIH (269, 272). Les sujets dits élites contrôleurs (HICs) sont capables de contrôler leur virémie pendant de longues périodes. Ils expriment quant à eux fortement le facteur de transcription T-bet, de la perforine et du granzyme B (269, 272).

3.7.2 Primo-infection.

La classification Fiebig résume la primo-infection en 6 phases (279, 280). Chacune des phases à une durée définie correspondant à la période de détection de marqueurs biologiques usuels (Figure XIV et Tableau III) (279, 280).

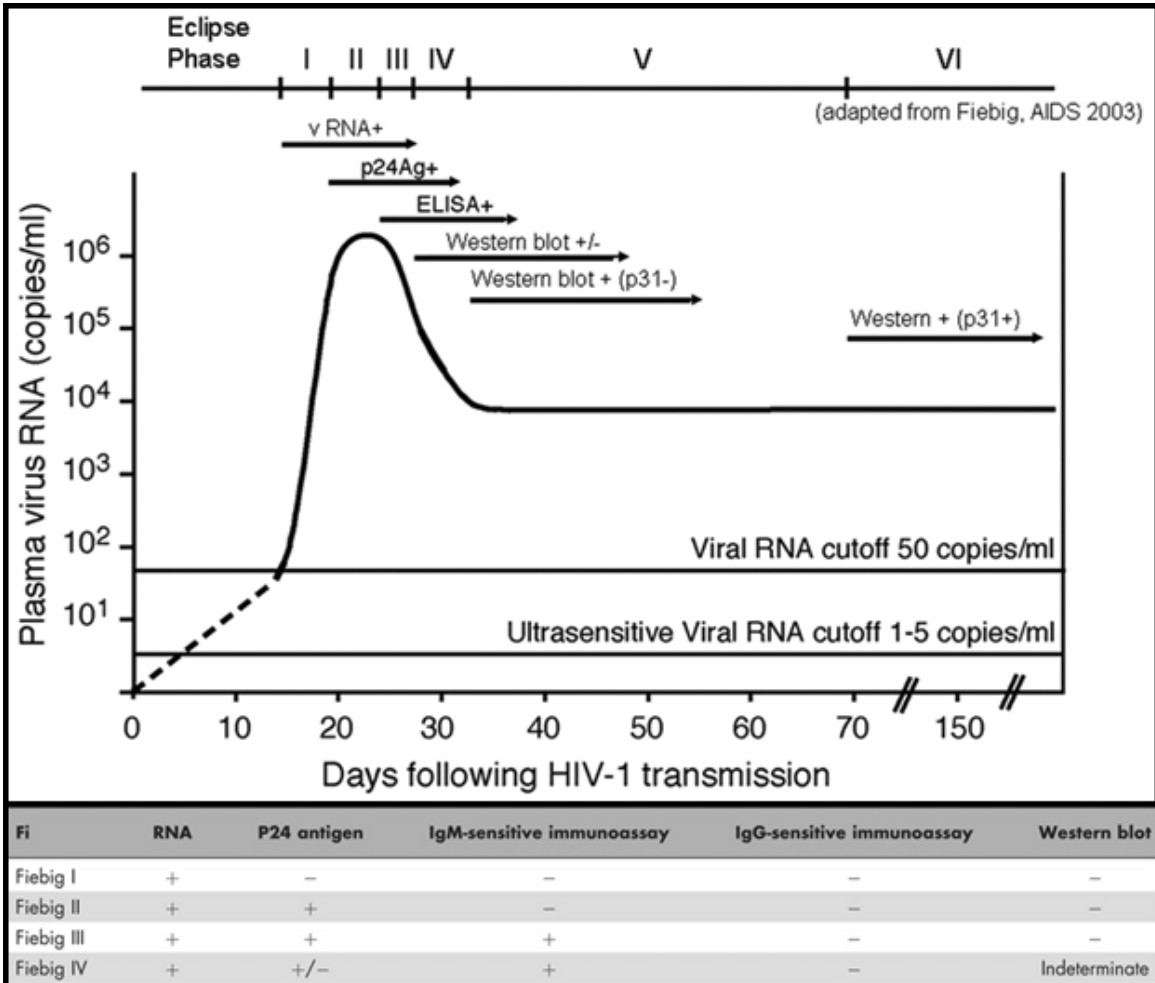


Figure XIV. Histoire naturelle et classification Fiebig de l'infection primaire par le VIH-1.

La figure résume les 6 phases de la primo-infection de l'infection par le VIH-1 selon la classification Fiebig et les marqueurs biologiques correspondants. Figure adaptée avec autorisation. Sources: Cohen MS *et al.* 2010 (279, 280) et Robb, Merlin L et Ananworanich, Jintanat, 2016 (281).

Tableau III. Durée des différentes phases de la primo-infection du VIH-1 selon la classification Fiebig.

Stage	Defining finding and/or marker	Duration, mean (range), days	
		Individual phase	Cumulative duration
Eclipse	...	10 (7–21)	10 (7–21)
I	vRNA positive	7 (5–10)	17 (13–28)
II	p24 antigen positive	5 (4–8)	22 (18–34)
III	ELISA positive	3 (2–5)	25 (22–37)
IV	Western blot positive or negative	6 (4–8)	31 (27–43)
V	Western blot positive, p31 antigen negative	70 (40–122)	101 (71–154)
VI	Western blot positive, p31 antigen positive	Open-ended	...

NOTE. ELISA, enzyme-linked immunoassay; vRNA, viral RNA.

Le tableau ci-dessus résume la durée en temps des différents marqueurs biologiques au cours des différentes phases de la primo-infection par le VIH-1. Tableau reproduit avec autorisation.: Cohen MS et al. 2010 (279).

4.4.1.1 Phase aiguë.

D'une durée de 7 à 28 jours, elle couvre la période qui suit le début réel de l'infection aussi appelée « éclipse » et le début de la réplication virale (256). Durant cette période, la personne nouvellement infectée ne présente aucun symptôme de la maladie en dehors de quelques-uns (fièvre, maux de gorge, douleurs musculaires, fatigue, gonflement des ganglions lymphatiques et éruptions cutanées), qui passent le plus souvent inaperçus. Ces symptômes sont la conséquence du début d'une augmentation de la virémie au cours de la première semaine. Elle doublerait chaque 6 à 10 heures et chaque cellule cible infectée est capable d'en infecter une vingtaine d'autres (282, 283). Sur le plan biologique, le virus se multiplie rapidement à l'intérieur de la cellule hôte présentant le récepteur et le ou les corécepteurs appropriés (284) au niveau du foyer infectieux. Il se propage par la suite à un grand nombre de cellules cibles *in situ* ou dans d'autres organes et tissus de l'organisme. La phase aiguë de la primo-infection constitue la période propice pour l'établissement rapide

des réservoirs viraux du VIH (285). C'est également au cours de cette période que la probabilité de transmission de la maladie à une tierce personne est la plus élevée compte tenu du niveau élevé de la charge virale et de la capacité de réplication élevée du virus (286). Le nombre de particules virales infectieuses dans le sang au cours de cette période se situerait entre 10^6 /ml à plus de 10^8 /ml (285). Cette période marque aussi le début de la séroconversion caractérisée par le début d'apparitions des anticorps anti-VIH-1 (279).

3.7.3 Phase asymptomatique.

D'une durée de 18 et 34 jours (256), la phase asymptomatique couvrant la primo-infection se caractérise par un accroissement continu de la réplication/propagation virale (hausse importante de la charge virale plasmatique suivie d'une baisse). La baisse de la virémie est accompagnée d'une augmentation des lymphocytes T cytotoxiques ou CTL (CD8+) correspondant au début d'une réponse immunitaire cellulaire de l'hôte (287-290). Cependant, cette réponse n'est pas assez efficace pour pouvoir contrôler de façon permanente l'infection. Cette période se caractérise particulièrement par une baisse accrue du nombre de lymphocytes TCD4+ sans l'apparition de manifestations cliniques. En outre, le virus continue de se multiplier, la taille des réservoirs viraux s'agrandit et la personne infectée continue d'être hautement contagieuse. Sur le plan biologique, les marqueurs biologiques comme l'antigène p24 et la charge virale plasmatique et les anticorps anti-VIH (279) peuvent être détectés dans le sang (291).

3.7.4 Diagnostic du VIH au cours de la primo-infection.

De nos jours, plusieurs tests et méthodes biologiques sont utilisés pour le diagnostic de l'infection par le VIH au cours de la primo-infection (292-294). Parmi ces méthodes, on distingue des méthodes immunosérologiques qui consistent à mettre en évidence les anticorps et/ou les antigènes du VIH (293, 294) et des méthodes moléculaires (diversité génétique) (292, 294, 295). Les tests sérologiques et la recherche de l'antigène p24 sont les plus couramment utilisés au Québec. Parmi les tests sérologiques « *l'EIA, Enzyme immunoassay* » (296, 297), un test de troisième génération de détection des anticorps anti-VIH est beaucoup utilisé. L'antigénémie p24 compte aussi parmi les plus utilisés dans les

laboratoires cliniques et permet de détecter la présence de l'antigène p24 du virus le plus tôt après l'infection (295, 298, 299), mais sa fenêtre de détection assez réduite (14-22 jours) constitue un inconvénient majeur (279) (Tableau III). Des tests de quatrième génération (300) permettent maintenant de détecter simultanément les anticorps anti-VIH et l'antigène p24 (EIA/p24). D'autres stratégies de détection du VIH se basent sur l'index d'avidité (AI) des anticorps anti-VIH, qui est proportionnel à la durée de l'infection par le VIH. Une personne nouvellement infectée par le VIH présente un AI faible : <1.5 ODN (LAg) et 0.8 ODN (BED), 20 S/C (LS Vitros), 60% AI (Vitros avidity) et 40% (BioRad Avidity) comparativement à une infection chronique (301). Ces méthodes combinent aussi les résultats de deux ou plusieurs tests d'AI différents dans un algorithme « *Multi Assay Algorithm, MAA* » de détection des infections récentes avec pour objectifs de réduire le taux des faux récents et d'accroître l'efficacité de la méthode diagnostique (293). Dans cette dynamique et pour optimiser les méthodes d'identification des infections récentes, une étude visant à évaluer la performance d'un algorithme de détection des infections récentes du VIH-1 incluant deux tests d'avidité : « (CDC, modified Bio-Rad-Avidity assay » suivi du « Sedia-LAg-Avidity assay » (293) est en cours d'optimisation au laboratoire de santé publique du Québec. L'objectif de cette étude est de pouvoir intégrer cet algorithme dans le programme de surveillance des récentes infections à VIH dans cette province canadienne. Bien que peu utilisés en routine, les tests moléculaires se basent généralement sur l'évaluation de la diversité des séquences virales du génome du VIH-1 pour identifier les nouvelles infections (292, 294, 302). Ils mettent à profit la corrélation qui existe entre le niveau de diversité des séquences virales de certaines régions du génome du VIH (enveloppe par exemple) et le stade/durée de l'infection (Fiebig stages) (303). Cette approche permet de déterminer les infections récentes (diversité génétique moins élevée par rapport aux chroniques (diversité génétique élevée) (109, 292, 304, 305). En l'absence de test ou méthode de référence « *gold standard* » désignée pour la détection des nouvelles infections par le VIH, chacune des méthodes identifiées est utilisée pour définir le stade de l'infection par le VIH au regard de la cinétique d'apparition des marqueurs biologiques avec leurs avantages et inconvénients (294, 302).

3.7.5 Phase chronique.

3.7.5.1 Phase symptomatique.

Cette période est caractérisée par le début des symptômes de la maladie et peut durer jusqu'à 12 ans sans traitement (20). Elle se manifeste par des symptômes systémiques comme : la fatigue chronique, sueurs nocturnes, fièvre, diarrhée, perte de poids importante, etc. Ces manifestations sont la conséquence d'un affaiblissement du système immunitaire (baisse progressive des lymphocytes T CD4+ et d'une réplication virale continue). Outre les symptômes cliniques évocateurs de la maladie, l'infection peut être diagnostiquée dans le sang par la mise en évidence de marqueurs biologiques usuels (306) comme : les anticorps anti-VIH-1, les antigènes du VIH et la charge virale.

3.7.5.2 Phase SIDA.

Elle survient plusieurs années après le début de l'infection (entre 2 et 25 ans) (20) et se manifeste par une défaillance généralisée du système immunitaire (immunodéficiences). Cette immunodéficiences est marquée par une baisse considérable des lymphocytes T CD4+ (< 200 cellules/ μ l) précédée d'une augmentation de la charge virale plasmatique (307). L'augmentation de la charge virale est synonyme d'une multiplication accrue du virus dans d'autres compartiments de l'organisme que les organes lymphoïdes (308). L'immunodéficiences conduit à la réémergence de particules virales hautement infectieuses (309-311) et à des symptômes de tout genre comme les infections opportunistes généralisées et la réactivation d'infections ou de maladies chroniques: pneumonie à *Pneumocystis jiroveci*, toxoplasmose, infection à cytomégalo virus, sarcome de Kaposi). La mort surviendra si aucun traitement antirétroviral n'est initié pour rétablir l'équilibre du système immunitaire.

3.7.6 Algorithme de dépistage et de confirmation du VIH au Québec.

Le dépistage et la confirmation du sérologique de l'infection par le VIH au Québec se fait principalement à trois niveaux : dans les hôpitaux et services de santé, au laboratoire de santé publique du Québec (LSPQ) et au Laboratoire national des services de référence du VIH du Canada (LNSRV). Dans les hôpitaux et services de santé, les tests de dépistage (recherche

des anticorps et antigènes du VIH) sont effectués. Les spécimens présentant des résultats positifs répétés sont envoyés au LSPQ pour la confirmation de la présence d'anticorps contre le VIH ou pour la détection de l'antigène p24 (312). Chaque province du Canada et d'autres pays ont des algorithmes qui pourraient être différents et les adaptent selon la disponibilité des épreuves homologuées disponibles. La détection de l'acide nucléique par amplification génique est souvent utilisée sous certaines conditions et dans certains pays pour la détection de l'ARN viral ou proviral (295, 306, 313).

3.8 Origines du VIH.

Après l'identification du VIH comme l'agent étiologique du SIDA, d'intenses recherches ont été menées afin de mieux connaître son origine et son réservoir. De nombreuses années de recherche épidémiologiques et phylogénétiques sur des échantillons rétrospectifs ont permis de constater que le VIH aurait été introduit dans la population humaine au début de siècle dernier (314-319). Les analyses ont par ailleurs permis d'identifier comme principaux réservoirs des ancêtres du VIH-1 des virus simiens infectant des chimpanzés d'Afrique centrale et des gorilles de l'ouest et pour le VIH-2 des VIS infectant des singes, les mangabeys enfumés d'Afrique de l'Ouest (314-319) (Figure XV). Les VIH-1 et 2 seraient le résultat de plusieurs transmissions interespèces des virus simiens à l'homme (318, 320-322). Le virus *SIVcpzPtt* (*Pan troglodytes troglodytes*), infectant des chimpanzés vivant dans le sud et l'ouest du Cameroun en Afrique centrale, serait l'ancêtre des groupes M et N du VIH-1 (318, 322, 323). Le groupe M serait présent dans la population humaine depuis les années 1908, IC 95% [1884-1924] notamment en République démocratique du Congo (RDC) et au Cameroun (324). Le groupe N, identifié en 1998 au Cameroun, aurait été présent dans la population humaine depuis l'année 1963, IC 95% [1948-1977] (325). Le virus *SIVgor* (*Gorilla gorilla*) présent chez le gorille de l'ouest de l'Afrique centrale a été décrit comme étant l'ancêtre du VIH-1 des groupes O et P (317, 318, 322). Le groupe O identifié en 1990 aurait déjà été présent dans la population humaine notamment au Cameroun entre les années 1905 et 1920, IC 95% [1890-1940] (325). Le VIH-1 du groupe P décrit en 2009 chez une patiente d'origine camerounaise immigrée en Belgique serait probablement apparu dans la

population humaine entre les années 1845 et 1989 (319, 322, 326-328). Quant au VIH-2, il résulterait d'au moins 8 transmissions indépendantes (correspondant à ses 8 sérotypes) du virus du singe, *sooty Mangabey (SIVsmm)* à l'homme. Il aurait pour réservoir le mangabey enfumé (*Cercocebus atys*) d'Afrique de l'Ouest qui se serait transmis à l'homme (318, 322, 323, 329) (Figures XVI). Cette diversification des sources de transmission du VIH à l'homme expliquerait, du moins en partie, les diverses propriétés génétiques, phénotypiques et cliniques observées, et la difficulté à concevoir un traitement ou un vaccin universel.

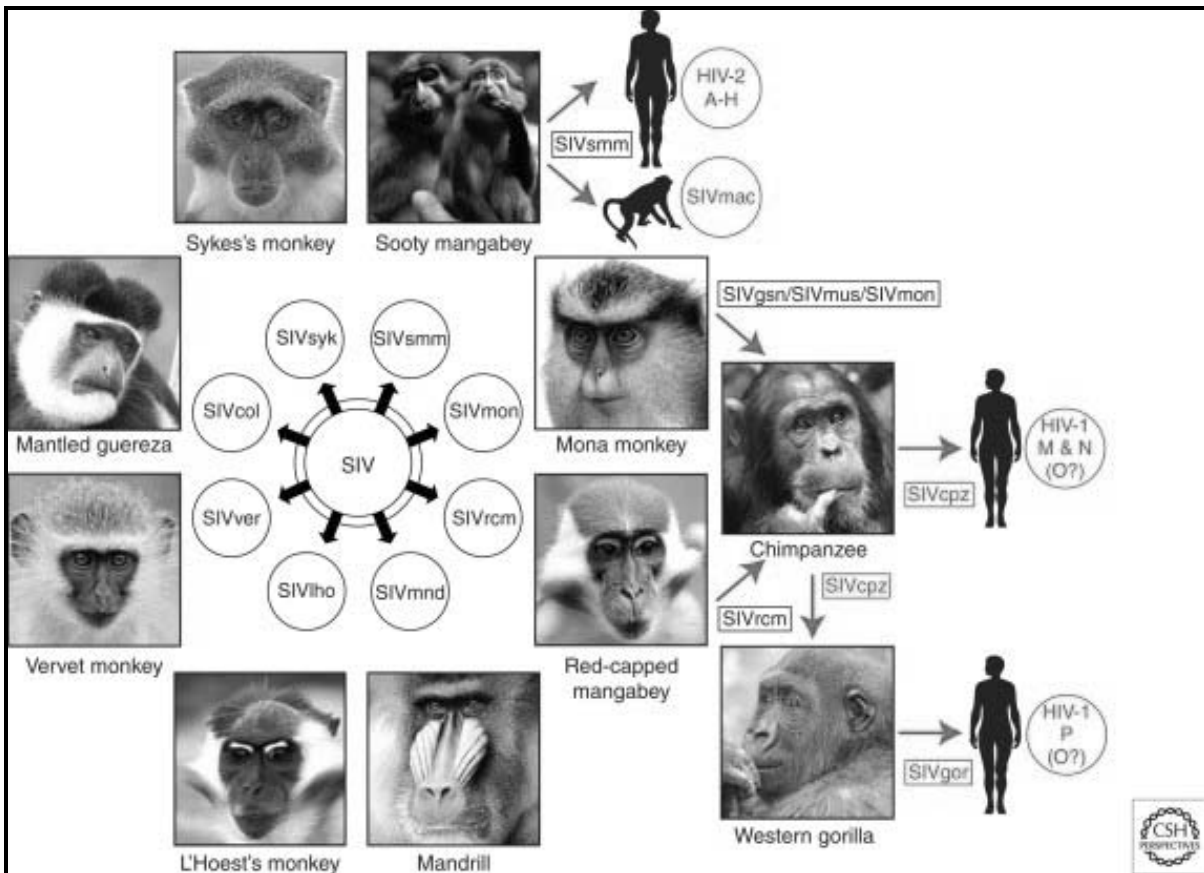


Figure XV. Images des différents chimpanzés et les gorilles ancêtres et principaux réservoirs du VIH.

Les images représentent les singes de l'Ancien Monde qui sont naturellement infectés avec plus de 40 lentivirus différents appelés virus de l'immunodéficience simienne (SIV). Plusieurs de ces SIVs ont franchi la barrière d'espèce (spécificité pour les grands singes) pour se retrouver chez l'être humain où il y'a eu la génération de nouveaux agents pathogènes. Des transmissions interespèces connues et la descendance de virus sont surlignés en rouge.

Images reproduites avec autorisation de *Cold Spring Harbor Laboratory Press*. Sources : Sharp PM Hahn BH. 2011(318).

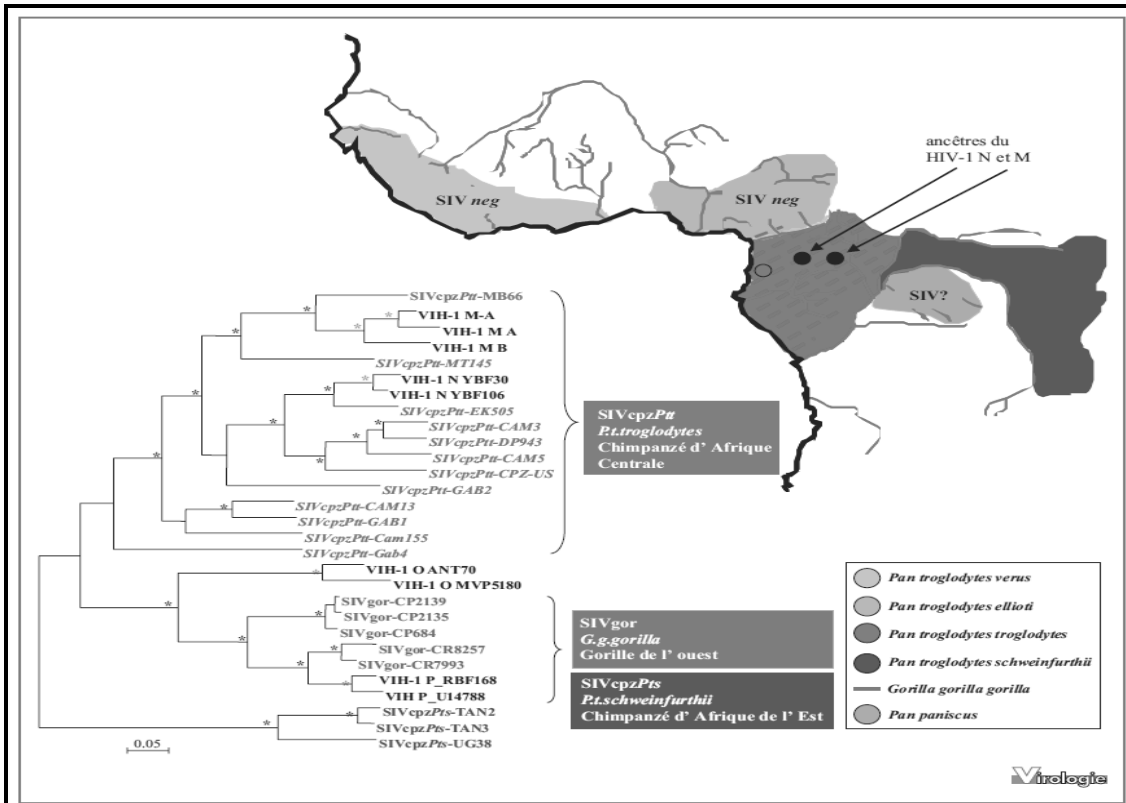


Figure XVI. Liste des espèces de virus simiens ancêtres du VIH-1 humain.

(1) L'arbre phylogénétique représente les relations évolutives entre les différentes lignées du VIH-1 des groupes M, N, O, P (en noir), les SIVcpzPts des *Pan troglodytes schweinfurthii* (en bleu), les SIVcpzPtt des *Pan troglodytes troglodytes* (en rouge), et les SIVgor des gorilles (*Gorilla gorilla*) (en vert). L'arbre a été construit avec la méthode du maximum de vraisemblances à partir d'un alignement protéique de séquences (410 pb) de la gp41 de l'enveloppe virale de VIS/VIH. (2) La carte représente les répartitions géographiques des espèces concernées: *G. g. gorilla* (traits verts) et les quatre sous-espèces de chimpanzés (*Pan troglodytes*). Les points noirs et verts sur la carte indiquent respectivement les réservoirs des ancêtres des différents groupes du VIH-1 dans les populations de chimpanzés et de gorilles sauvages. Figure reproduite avec autorisation: Peeters M et Chaix M-L, 2013 (322).

3.9 Diversité génétique du VIH.

Le VIH-1, comme tous les autres rétrovirus (*Oncovirus*, responsable de tumeurs comme les lymphomes et des leucémies et les *Spumavirus*, non pathogènes), se caractérise par sa grande diversité génétique se traduisant par une diversité antigénique. Tous les gènes du VIH-1 sont variables. Cependant, le gène *env* par rapport à *gag* et *pol* est le plus variable estimé à 40% (*env*) contre 8–10% (*pol/gag*) (330).

Durant la phase asymptomatique de la primo-infection, on estime que la diversité de la population virale du gène *env* accroît continuellement de l'ordre de 1% par année avant de se stabiliser ou de diminuer au cours de la maladie (331). L'extrême variabilité génomique (taux de mutation), la capacité de réplication virale rapide et les recombinaisons génétiques constituent les trois principales causes de la diversité génétique (332).

3.9.1 Principales causes de la diversité génétique.

3.9.1.1 Taux de mutation élevé induit par les erreurs de la transcriptase inverse et des facteurs de l'hôte.

Les nombreuses et fréquentes mutations introduites dans le génome viral par les erreurs de l'enzyme de la transcriptase inverse du virus [(\approx 0.2 erreur par génome et par cycle de réplication viral (333)] favorisent la diversité génétique du VIH-1 (331, 332, 334). Certains facteurs de l'hôte comme APOBEC3G par exemple dans son action de restriction insèrent généralement des codons d'arrêt susceptibles d'induire des substitutions d'acides aminés dans la synthèse de l'ADN du nouveau virion (331, 334, 335). Le taux des mutations induites par les erreurs de transcription inverse est de 1 substitution pour 1700 nucléotides polymérisés et augmente de 1 sur 70 dans certaines régions du génome viral particulièrement dans certaines les zones chaudes (336, 337). Les premières études *in vitro* montraient que les taux d'erreurs de mutations variaient de 5×10^{-4} à $6,7 \times 10^{-4}$ (338, 339). D'autres études *in vivo* démontrent que les taux de mutations pour un seul cycle de réplication seraient de 5% inférieurs à ceux des études *in vitro* estimées à $3,5 \times 10^{-5}$ à $5,4 \times 10^{-5}$ erreurs de substitution par paire de bases (340-342). Des études plus récentes évaluent ce taux d'erreurs de mutation à $1,4 \times 10^{-5}$ par paire de bases, et par cycle de réplication (343).

3.9.1.2 Recombinaison génétique.

Les recombinaisons génétiques sont le résultat des croisements intergénomiques (co-infection) d'une cellule par deux virions génétiquement distincts. Elles constituent également des facteurs contribuant à la diversité génétique du VIH-1 (331, 332, 334). La recombinaison génétique peut se définir comme un phénomène d'échange de matériel génétique entre deux virus génétiquement distincts cohabitant dans la même capsid d'une même cellule

cible et dont le résultat du croisement donne naissance à un ARN fille portant les caractéristiques génétiques des deux virus (344-346). Cependant, et selon le principe de Hardy-Weinberg (347), toutes les co-infections ne conduisent pas nécessairement une recombinaison génétique. Les génomes génétiquement distincts sont regroupés dans la descendance virale de la façon suivante : dans 50% des cas les virions contiennent une seule copie de chaque génome viral distinct (hétérozygote), dans 25% des cas ils contiennent deux copies identiques de la première souche (homozygote) et dans 25% ils contiennent des copies identiques de la deuxième la souche (homozygote). Seul le virion hétérozygote peut recombiner la forme virale dans les cycles ultérieurs de la réplication virale (Figure XVII). La recombinaison génétique peut subvenir du croisement de virus du même sous type viral (intra-sous-type) ou du croisement de virus de sous-types différents (inter-sous-types). Ce phénomène justifie ainsi les nombreuses formes recombinantes uniques (URF, *Unique recombinant form*), et des formes recombinantes circulantes (CRF, *circulating recombinant form*) (346). L’Afrique subsaharienne est la région du monde où l’on retrouve la plus grande diversité génétique du VIH-1 avec un grand nombre de sérotypes circulants (348, 349) contrairement à l’Europe et aux pays nord-américains (350). Selon les études, la fréquence de ces recombinaisons génétiques serait de l’ordre de 2,8 évènements par génome viral et par cycle de réplication (351).

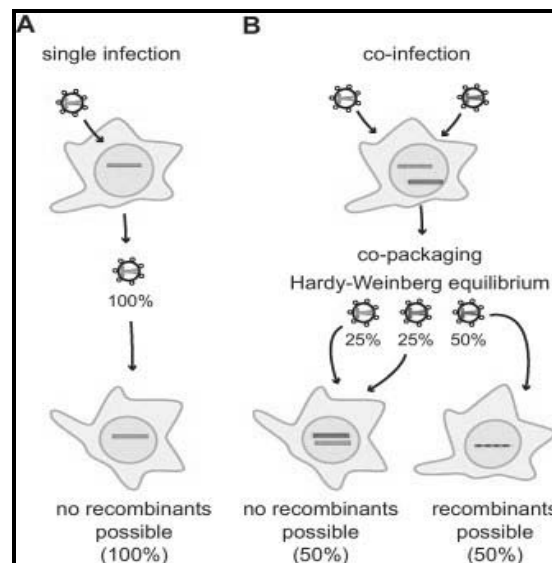


Figure XVII. Mécanisme de recombinaison génétique.

La figure ci-dessus illustre une co-infection d'une cellule hôte par deux souches virales génétiquement distinctes. Cette co-infection est nécessaire à la génération d'un virus recombinant (souche 1 en bleue, souche 2 : en rouge.). En (A) l'infection d'une cellule cible par un seul virus, n'infère pas forcément une recombinaison génétique. Cependant (B) l'infection d'une cellule par deux souches virales génétiquement différentes n'induit pas forcément une recombinaison génétique. Selon la théorie de Hardy–Weinberg (347) après la co-infection il y'a une probabilité de 50% que des virions générés contiennent de l'information génétique de chacun des variants viraux (1 et 2), ils sont dits hétérozygotes. Dans 25% des cas ils posséderont de l'information génétique identique soit du variant 1 ou du variant 2. Ils sont dits homozygotes. Seuls les virions hétérozygotes pourront former une recombinaison génétique au cours de réplication virale. Figure reproduite avec autorisation. Source : Smyth, R. P. *et al.*, 2012 (346).

3.9.1.3 Capacité élevée de la réplication du VIH-1.

La capacité élevée de la réplication virale (*fitness*) dans les cellules infectées constitue aussi des facteurs contributifs à l'essor de la diversité génétique du VIH-1 (331, 332, 334). Elle résulte de la pression de sélection exercée par le système immunitaire et/ou les traitements antirétroviraux. On estime que la pression de sélection exercée par le système immunitaire induit environ $\approx 2,5$ changements par mois dans le gène *env* (352). En effet, le nombre de virions produits par jour chez un individu infecté est très élevé, soit environ de 10^{10} à 10^{12} (353, 354). Les taux de réplication virale élevés contribuent à l'accroissement de la probabilité de génération de nombreuses quasi-espèces virales et des mutations au sein d'un même individu ou d'un groupe d'individus. Il est établi qu'en début d'infection par le VIH, la diversité génétique des séquences virales est généralement faible chez la personne nouvellement infectée, mais elle augmente au cours de la progression de la maladie (chronicité) (355). Ainsi, le niveau de diversité génétique des séquences de certaines régions du génome viral est fortement corrélé avec le stade d'évolution de la maladie (303). Cette propriété est exploitée dans certains cas comme biomarqueur moléculaire pour différencier les nouvelles infections des infections chroniques chez les personnes nouvellement diagnostiquées positives au VIH-1 (292, 304, 355, 356) (cf Chap. III, article 1 (292)).

3.9.2 Conséquences de la diversité génétique du VIH-1.

La grande diversité génétique du génome du VIH-1 confère sa complexité tant à l'échelle individuelle qu'à l'échelle collective. Cette diversification génétique constitue un atout pour le virus pour l'établissement de l'infection, et un obstacle majeur pour la prise en charge clinique à cause des conséquences biologiques, cliniques et épidémiologiques (357-360) qui y sont associées.

3.9.2.1 Conséquences biologiques.

a. La détermination du phénotype viral : tropisme. L'enveloppe est une protéine présente à la surface de la membrane lipidique de la particule virale (VIH-1) (361). Elle constitue la première composante du virus à interagir avec les protéines de surface des principales cellules hôtes (lymphocytes TCD4+, macrophages, monocytes). Ces protéines de surfaces comprennent : le CD4 (principal récepteur du virus), mais aussi des corecepteurs (CCR5 et CXCR4) (361). Le mode d'interaction des protéines d'enveloppe du VIH-1 avec les principales cellules cibles change en fonction de facteurs environnementaux (localisation) de la cellule, mais aussi du stade de l'infection qui est susceptible d'induire des changements dans l'évolution de ces cellules(362, 363). L'utilisation d'un des corecepteurs (CCR5 ou CXCR4) par le virus pour entrer dans la cellule cible (lymphocyte TCD4) définit le phénotype viral. Ce processus définit comme le tropisme viral est directement affecté par la diversité et la variabilité génomique du VIH-1 (360), mais aussi par la densité de la principale protéine de surfaces (CD4) à la surface de la cellule. Il a été démontré que certaines cellules myéloïdes (monocytes et macrophages) expriment de faibles densités du récepteur CD4 à leurs surfaces par rapport à la densité des CD4 présente à la surface des lymphocytes T CD4+ (élevée) (364). Le tropisme viral est un mécanisme qui favorise la sélection d'une souche virale particulière qui évolue en fonction du stade de l'infection (362, 363). Au regard des paramètres ci-dessous énumérés, il est établi actuellement qu'il existe trois types de VIH-1 définis selon leurs tropismes et non deux comme indiqué dans les premières littératures. Ces trois types de VIH-1 sont: 1) les virus R5 *T cell-tropic*, 2) les virus X4 *T cell-tropic* et, 3) les virus R5 macrophages tropic ou *M-tropic* (365) (Figure XVIII) .

Les virus R5 T cell-tropic sont majoritairement présent dans le sang en début d'infection (365). Ils sont constitués par les virus transmis/fondateurs (*Transmitted/founder viruses*) et les virus latents réactivés (365) . Ils utilisent pour une entrée efficacement dans le lymphocyte TCD4+, le corécepteur CCR5 associé à une haute densité de CD4 à la surface de la cellule cible afin d'établir l'infection (365).

Les virus X4 T cell-tropic et M-tropic quant à eux sont présents dans la phase chronique de la maladie. Ils utilisent le corécepteur CXCR4 associé à une faible densité de CD4 exprimée à la surface des cellules myéloïdes (monocytes et macrophages) cibles pour entrer efficacement dans la cellule afin d'établir l'infection(365). Cependant, il est démontré que les vrais virus M-tropic sont essentiellement ceux retrouvés dans les liquides cérébro-spinaux du cerveau des personnes infectées par le VIH-1 et souffrant de démence (366). Ces virus sont rarement observés dans le sang là où se trouve principalement des virus X4 T cell-tropic ou R5 T cell-tropic (365-367).

La détermination du phénotype ou tropisme viral constitue un outil important pour déterminer la souche virale présente chez l'individu infecté. Elle permet de juger de la possibilité d'utiliser un antagoniste du CCR5 comme inhibiteur d'entrée pour le traitement antirétroviral du sujet infecté (362, 363). Elle peut se faire par des méthodes phénotypiques et génotypiques (368-374).

Les méthodes phénotypiques

Les méthodes phénotypiques sont des tests fonctionnels. Actuellement de l'utilisation de la lignée cellulaire « *Affinofile cell line* » (375) est la mieux indiquée. Les cellules Affinofile peuvent être induites pour exprimer des densités variables de CD4 non seulement à la surface des lymphocytes T CD4+(CD4 élevée), mais aussi à la surface des macrophages et des monocytes (CD4 faibles) (375). Ces cellules Affinofile expriment simultanément le CD4 et le CCR5 sous le contrôle de promoteurs inductibles où l'induction de chacun est titrable partant d'un niveau d'expression faible à un niveau d'expression élevé (375). Des études réalisées sur Affinofile cell Line démontre clairement que les variants VIH-1 T *cell-tropic* utilisant soit les corecepteurs CXCR4 et CCR5 ne peuvent pas infecter efficacement les cellules exprimant une faible densité de CD4. Ces virus requièrent nécessairement un niveau élevé de CD4 comme

chez les lymphocytes TCD4+ pour établir une entrée efficace dans la cellule (375). Pourtant, les virus M-tropic sont capables d'entrée dans les cellules cibles exprimant de faibles niveaux de CD4 à leurs surfaces comme ce qui est généralement retrouvé à la surface des macrophages tout en étant capables d'infecter des cellules exprimant des niveaux élevés de CD4 (365, 375) (Figure XVIII). Contrairement aux méthodes récentes utilisant les cellules affino-file, les premières méthodes phénotypiques utilisaient des virus recombinants ou pseudotypés (isolats viraux) provenant de culture sur du PBMCs (*Peripheral blood mononuclear cell*) d'échantillons de patients infectés et de lignées cellulaires) (371). Ces méthodes ne considèrent pas tous les quasi-espèces viraux présents chez le sujet infecté (quasi-espèces virales) (375), mais seulement ceux ayant pu se développer dans ces cellules. Ces virus pourraient être ceux qui ont pu survivre à cause d'une adaptation de ces isolats aux conditions de culture et du milieu et qui auraient pu induire une modification du phénotype d'entrée. En outre, tous les virus se développent dans les PBMCs contenant des lymphocytes T CD4+, mais pas pour les macrophages. Pourtant, certains virus sont définis inadéquatement par ces méthodes comme ayant un tropisme pour les cellules T, mais pas pour tous.

La plupart des virus non-syncytium-inducing (NSI) ne se développent pas bien dans les macrophages pourtant, ils sont tous désignés comme virus M-tropic. Les vrais virus M-tropic sont responsables de démences chez les patients infectés et toujours présents dans le cerveau. Ces derniers utilisent de faibles densités de CD4 beaucoup plus efficacement que les virus R5 T cell tropic pour infecter leurs cellules cibles (367, 376-381) de façon similaire pourtant ils sont désignés différemment. En fin, la plupart des isolats de VIH-1 incluant pratiquement tous les isolats transmis (355) n'infectent pas efficacement les macrophages (CCR5+) et toutes les autres cellules exprimant de faibles densités de CD4 à leurs surfaces. Pourtant par ces premières méthodes tous les virus R5 seront tous désignés virus *M-tropic* (365).

Les méthodes génotypiques

Les méthodes génotypiques sont des tests moléculaires. Le principe des méthodes génotypiques consiste à mettre en évidence la présence de résidus d'arginine et de lysine positivement chargés aux positions 11 et 24 ou 25 de la séquence de la boucle V3.

Actuellement, la méthode d'amplification de génomes uniques ou (single genome amplification, SGA) une technique permettant d'amplifier et de cloner le gène env à partir d'échantillons de patients infectés par le VIH-1 constitue la technique appropriée pour générer des clones d'enveloppe du virus pour la détermination du phénotype viral (382-385). Le SGA repose sur une amplification par PCR nichée de l'ADNc obtenu par transcription inverse après une étape de dilution. IL permet d'amplifier une séquence complet du gène env (386) sans risque de recombinaison génétique in vitro dans la mesure où l'ADNc est issu d'un seul clone et sert de matrice pour chacune des PCR (382-386). Cette technique permet aussi de réduire le risque de ré-échantillonnage dans la mesure où la population virale résultante après dilution comprend un seul clone. Contrairement à la PCR classique, le SGA permet de réduire l'amplification artificielle d'un variant minoritaire au sein d'une population hétérogène (383, 387, 388). Le tropisme viral est déterminé en soumettant les séquences virales de l'enveloppe incluant les régions C2V3 à l'algorithme Geno2Pheno coreceptor (389). Les méthodes génotypiques au regard de leur facilité d'exécution et des coûts réduits sont les plus utilisées (390). Des séquences d'enveloppes virales obtenues après SGA d'échantillons sanguins de patients infectés par le VIH-1 et soumises à l'algorithme Geno2Pheno coreceptor démontrent qu'au stade chronique de la maladie, les taux de faux positif (False positive rate, FPR) de 10% sont des virus R5 (R5 T cell-tropic), des valeurs <2% sont des virus X4 (X4 T cell-tropic) et les taux FPR compris entre 2% et 10% sont majoritairement des virus R5 (R5 T cell-tropic) (391). NB. Les vrais virus R5 macrophages tropic ou M-tropic se trouvent rarement dans le sang, mais plutôt dans les liquides cérébro-spinaux du cerveau.

La boucle V3, par sa composition et sa charge positive nette, constitue ainsi une sous-région de l'enveloppe du VIH-1 jouant un rôle majeur dans la détermination du tropisme et toute mutation même mineure induite en son sein favorise la sélection ou le changement de phénotype viral (392-395). On note cependant que les boucles V1/V2 et C4 de la gp120 peuvent être impliquées dans la détermination du tropisme/phénotype viral (396).

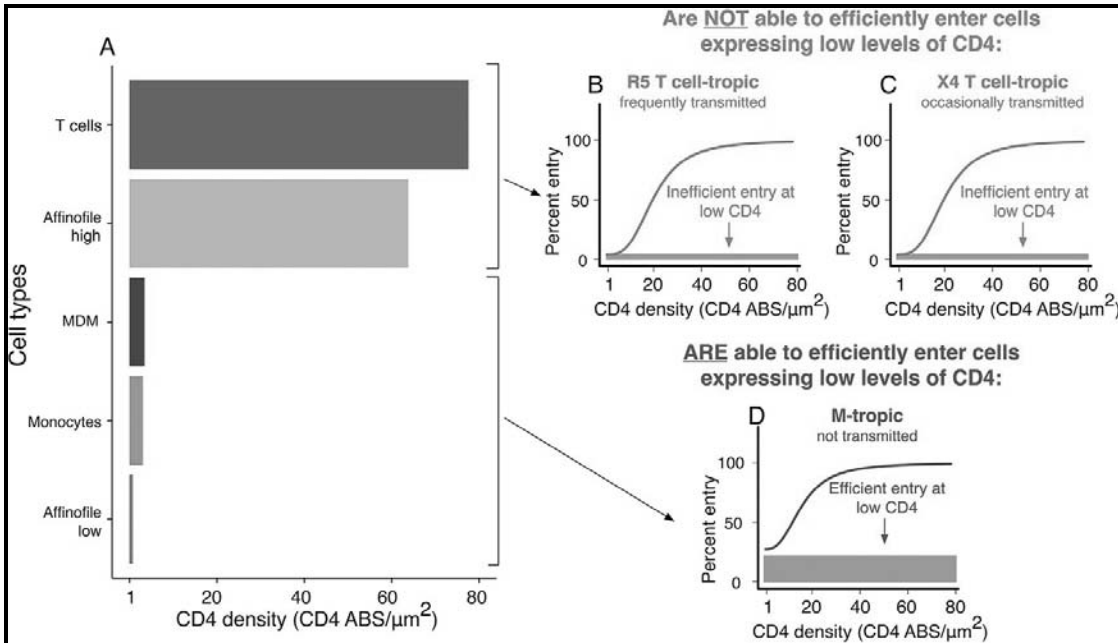


Figure XVIII : Description des propriétés des trois phénotypes de VIH-1 utilisant les cellules Affinofile cells.

(A) Les cellules Affinofile (375) sont induites pour exprimer des densités variables de CD4 allant du niveau élevé sur les lymphocytes TCD4+ à faibles sur les macrophages et les monocytes. Ce réactif a été utilisé pour évaluer la capacité des protéines d'*env* à médier l'entrée dans les cellules TCD4+ exprimant des niveaux variables de CD4 (367). Cette approche démontre que les isolats viraux utilisant le CXCR4 (B) et CCR5 (C) sont incapables d'infecter les cellules exprimant de faibles niveaux de CD4, mais plutôt pour celles exprimant des densités élevées cas des lymphocytes T. Les protéines d'*env* des virus M-tropic sont capables d'infecter les cellules cibles exprimant de faibles niveaux de CD4, cas des macrophages (D). Figures reproduites avec autorisation. Source : Joseph, SB, Swanstrom, R. 2018 (365).

b. La diversité génétique affecte les longues séquences terminales répétées (LTR) et d'autres facteurs de l'hôte (NF-KB et CypA).

Les longues séquences terminales répétées (LTR) du génome du VIH régulent la transcription et la réplication en tant que promoteur de la protéine Tat (360). Elles présentent les mêmes fonctions pour tous les sous-types de VIH-1, mais diffèrent dans leurs structures, leurs fonctions de base et leurs réponses aux cytokines et aux facteurs de transcription (360, 397) en fonction du sous-type. Des études montrent également que les sous-types du groupe M du contiennent deux sites de fixation pour NF-KB (Nuclear factor-Kappa B), un facteur de transcription tandis que certains sous-types C en possèdent un de plus. Celui-ci est impliqué

dans la réplication virale en présence de TNF- α et l'activation de la réponse immunitaire au stade chronique de l'infection (398). La diversité génétique confère également une différence de la cyclophiline A (CypA), une protéine cellulaire impliquée dans la stabilisation de la capsid du VIH-1 en interagissant directement avec *gag*, se lie à la capsid mature, restreint l'activité du facteur de restriction (TRIM5 α) et confère une augmentation de l'infectivité (399-402). La cyclophiline A est requise pour la réplication des groupes M du VIH-1 et VIS contrairement à celle du groupe O (403).

3.9.2.2 Conséquences cliniques et épidémiologiques de la diversité génétique.

L'extrême variabilité génétique du génome du VIH-1 complique l'action des mécanismes de défense immunitaire et engendre de nombreuses difficultés pour la prise en charge des personnes vivant avec le VIH : les tests diagnostiques, de résistance aux traitements antirétroviraux et pour la conception de candidat-vaccin à large spectre.

a. La diversité génétique du VIH-1 favorise l'échappement viral aux mécanismes de défense du système immunitaire de l'hôte.

La diversité génétique du VIH-1 constitue l'un des principaux mécanismes d'échappement aux réponses immunes médiées par l'organisme. Des réponses cytotoxiques (CTL) sont développées au cours de la phase précoce de l'infection. Ces réponses cytotoxiques restent cependant insuffisantes et limitées dans le temps (404-406) à cause des mutations de résistances induites précocement au niveau des épitopes codants pour des réponses T cytotoxiques (CTL) CD8+ (382, 407-411). Par le biais des nombreuses mutations qui s'opèrent au niveau de ces épitopes cibles des CTL, les virus deviennent méconnaissables et échappent aux réponses T cytotoxiques (cf chap. I.4). La diversité antigénique des nombreuses quasi-espèces virales cause (412-414) une activation généralisée du système immunitaire (415) qui provoque une déplétion des lymphocytes T CD4+ et l'évolution de la maladie. Cette altération fonctionnelle des lymphocytes T CD4+, chefs d'orchestre du système, abroge toute coopération cellulaire provoquant une progression rapide vers la maladie (SIDA). Les recombinaisons génétiques résultant de la co-infection (416-419) et les mutations silencieuses (*synonymous substitutions*) au niveau de l'enveloppe (420) sont des facteurs qui contribuent également à l'évasion immunitaire et à la progression de la maladie

vers le stade SIDA. Des anticorps neutralisants sont également produits au cours des 2 à 5 semaines suivant la transmission du VIH-1 (421). Cependant, cette réponse humorale est incapable de contrôler l'infection en raison de la flexibilité conformationnelle et des nombreuses mutations qui s'opèrent continuellement au niveau de l'enveloppe virale (422, 423) (cf chap. I-3.6).

b. La diversité génétique contribue à la résistance aux traitements antirétroviraux.

La variabilité génétique du VIH-1 a lieu au niveau de tout le génome (*gag*, *pol*, *env*), mais varie en fonction du sous-type viral. Les nombreuses mutations qui sont générées au niveau des gènes affectent ou induisent des résistances aux différents traitements antirétroviraux. Ces derniers ont la capacité de contrôler la réplication virale, mais ne peuvent pas éliminer la maladie (424-427). Tout changement même d'un nucléotide au niveau d'un codon particulier du génome viral peut se traduire par des changements dans l'acide aminé (428) avec pour corollaire l'induction d'une résistance à une molécule antirétrovirale. Ainsi, la fréquence des mutations de résistance associées aux traitements antirétroviraux est directement liée aux nombres de modifications dans le génome viral nécessaire pour l'acquisition d'un acide aminé associé à une résistance particulière (428). Par exemple, au niveau des sous-types B, C et F, une substitution de la valine (V) par l'Alanine (A) en position 82 (V82A) du gène Pol confère une résistance aux inhibiteurs de la protéase (426, 429, 430). Cependant, pour les sous-types G, une substitution de l'isoleucine par l'alanine (I82A) à cette même localisation confère la même résistance (426, 430). Des mutations de résistance aux inhibiteurs de la reverse transcriptase (Q151M et L210W) sont également propres de certains sous-types viraux (426, 430, 431). Le taux de réplication virale élevé introduit aussi des mutations de résistance naturelle pour certains traitements antirétroviraux chez des personnes infectées naïves (432).

c. La diversité génétique affecte le développement des tests de diagnostics universels pour l'infection par le VIH.

La diversité génétique du VIH associée aux multiples sous-types viraux et les formes recombinantes circulantes ont entravé la mise en place d'un test diagnostique universel

(360). Cette situation impose une évaluation permanente et le développement de nouveaux tests diagnostiques pour prendre en compte les nombreux groupes et sous-types de VIH circulants. La majorité des premières générations de tests diagnostiques mettant en évidence des anticorps du VIH ont été développées dans les pays développés et prenaient en compte des épitopes du VIH-1 de sous-type B (groupe M) dans leur conception. Elles ne permettaient pas de détecter les autres groupes comme le O (360, 433, 434). La problématique des tests diagnostiques est progressivement résolue avec le développement de nouveaux tests sérologiques (435, 436) classés en générations. Les épreuves de 3^{ème} générations détectent les anticorps IgM et IgG du VIH entre 20 et 30 jours après transmission du VIH, mais se positive tardivement pour les variants non B. (435, 437-440). Les épreuves de 4^{ème} générations sont des tests combinés, capables de détecter simultanément entre 15-20 jours après transmission les anticorps anti-VIH-1 et 2 et l'Ag p24, mais avec une performance variable pour les sous types non-B ou des sous types B présentant des mutations dans le domaine transmembranaire de la gp41 (440-444). Les tests de 6^{ème} génération sont capables de détecter et d'identifier simultanément les anticorps du VIH-1 (groupes M et O), VIH-2 et l'Ag. P24 du VIH-1 (*BioPlex 2200 HIV Ag-Ab assay*) (445). Ils constituent une alternative pour les tests de 4^{ème} génération.

d. La diversité génétique constitue un des facteurs clés des échecs de développement d'un vaccin anti-VIH efficace.

L'extrême variabilité génétique au niveau individuel et entre les différents sous-types du VIH implique une très grande diversité d'épitopes antigéniques. Cette diversité antigénique entrave les efforts de conception d'un vaccin anti-VIH à large spectre (360, 446). La diversité génétique a des conséquences sur toutes les stratégies de conception d'un vaccin anti-VIH-1. Un candidat vaccin spécifique d'isolats (interclades) pourrait induire des réponses immunitaires croisées contre les différents sous-types viraux circulants. Mais l'efficacité d'une telle approche se trouve affectée en raison de la grande diversité entre séquences virales des protéines de l'enveloppe qui est estimée à 30% (446). Pour contourner ces obstacles, d'autres approches seraient de concevoir des vaccins spécifiques du sous-type viral prédominant dans une région géographique, ce qui limiterait le phénomène de variabilité

(446). D'autres stratégies visent à utiliser les séquences consensus ou de virus ayant un ancêtre commun dans la conception vaccinale (360, 446). En somme, la diversité des antigènes viraux circulants au sein d'un même individu, d'un même pays ou dans le monde affecte la plupart des réponses immunes susceptibles d'être induites par un vaccin. L'infection par le VIH induit la production d'anticorps neutralisants (447) dirigés contre certaines régions conservées au niveau de l'enveloppe virale (211). Mais la flexibilité conformationnelle de ces épitopes limite l'efficacité de la liaison avec les anticorps (448, 449). Aussi, l'hypervariabilité associée à la très forte glycosylation et la formation de bouclier de résidus de glycans au niveau de l'enveloppe virale limitent la production d'anticorps neutralisants à larges spectres au stade précoce de l'infection conduisant déjà à l'évasion immunitaire (422) (cf chap.I.3.7). La diversité génétique du VIH limite également le développement d'un vaccin capable d'induire des réponses T cytotoxiques. En effet, les mutations de résistance, même une seule, qui interviennent au niveau des épitopes cibles (région *gag*) affectent significativement le développement d'un vaccin à potentiel cytotoxique (450, 451).

e. Diversité génétique contribue à l'expansion géographique des différents sous-types viraux. La pluralité des différents sous-types et formes recombinantes circulantes du VIH-1 dans le monde entre sous-types (de 25 à 35%) ou intra sous-types (de 15 et 20%) (452) contribue à l'émergence continue de nouvelles souches. Le VIH-1 est ainsi subdivisé en groupes, en sous-types ou en clades, et en de nombreuses formes recombinantes (CRF, *circulating recombinant form*). Les formes recombinantes proviennent de croisements intergénomiques entre des sous-types de VIH-1 infectant plusieurs personnes (CRF) ou de croisement intergénomique de deux ou de plusieurs sous-types viraux chez un même patient (les formes recombinantes uniques, URFs). On distingue le groupe M (*major*) qui est le plus prévalent dans le monde suivi des groupes N (non-M), O (*outlier*) et P, mais avec des prévalences très limitées (327, 328, 453, 454). Le groupe M se subdivise en 9 sous-types dont : A, B, C, D, F, G, H, J et K et les sous sous-types: A1 à A6 et F1 à F2 et plus de 96 CRFs et URFs caractérisées à ce jour (20, 322, 455-457). Le VIH-2 se subdivise en 9 groupes dont les A, B, C, G, H, I circulant principalement en Côte D'Ivoire et au Ghana (458, 459) et les D, E et F circulant en Sierra

Leone et au Liberia (318, 349, 460-462). La ressemblance (identité) des séquences du génome complet du VIH-1 et du VIH-2 est de l'ordre 48% en général. Mais elle est de 60% si on considère le gène *gag* (463) et assez faible (35-40%) si on considère les séquences de l'enveloppe (464, 465). La figure XIX donne un aperçu de la distribution géographique des différents sous types et formes recombinantes du VIH-1 dans le monde en 2014.

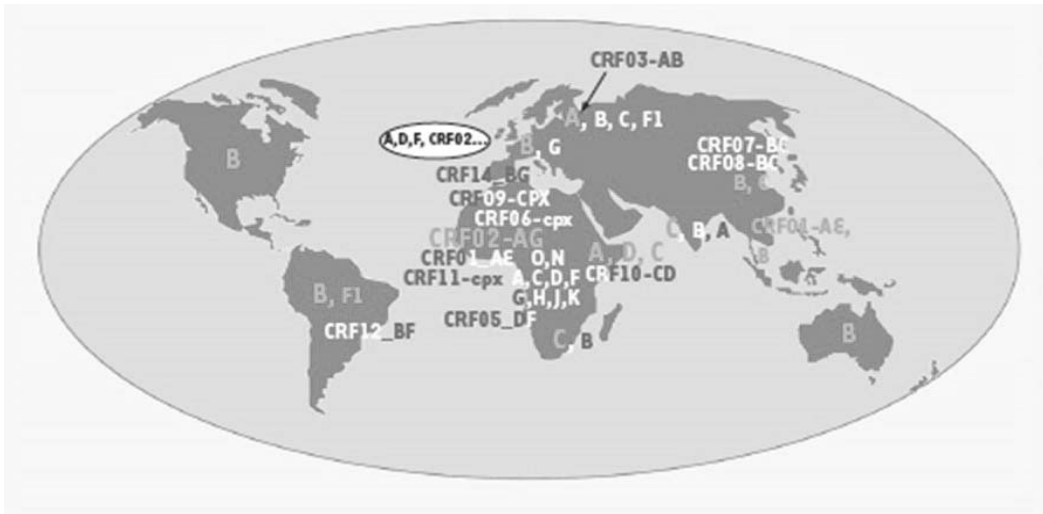


Figure XIX. Répartition géographique des différents sous-types viraux et CRF du VIH-1 groupe M, N et O dans le monde.

La figure illustre la distribution géographique des principaux sous-types viraux et formes recombinantes dans le monde, mais la diversité s'accroît continuellement. Figure reproduite avec autorisation. Source: Peeters M et al. 2014 et 2008 (348, 349).

4. Physiopathologie de l'infection VIH.

4.1 Généralité et mode de transmission de l'infection.

Le VIH se transmet par contact avec des fluides biologiques contenant le virus provenant d'une personne préalablement infectée. Les principaux modes de transmission sont; 1) les relations sexuelles non protégées (hétérosexuelles ou homosexuelles); 2) l'injection avec du matériel contaminé; 3) la transmission verticale *in utero*, pendant l'accouchement ou au cours de l'allaitement; 4) la transfusion de sang contaminé. *Il faut noter que tout contact avec un fluide biologique contaminé ne conduit pas forcément à une transmission de la maladie.* Selon de récentes études, le risque de transmission du VIH estimé par acte entre un sujet porteur du VIH et celui non infecté diffère selon le risque (466, 467). Ainsi, le risque de

transmission du VIH estimé par acte pour 10 000 expositions est élevé en cas de transfusion sanguine. Elle constitue le plus grand risque qui est estimé par acte à 9250 (IC 95% 8900–9610) pour 10 000 expositions. Le deuxième grand risque est la transmission de la mère à l'enfant qui est estimé à 2260 (IC 95% 1700–2900) pour 10 000 expositions (466-469). Le risque pour les rapports sexuels anaux réceptifs (HARSAH) est 138 (IC à 95%: 102–186) pour 100 000 expositions, suivi des échanges d'aiguille contaminé par les usagers de drogues injectables [63 (IC 95% 41–92)] et des blessures percutanées par piqûre d'aiguilles contaminées [23 (IC 95% 0–46)] (468-471). Le risque pour les rapports hétérosexuels (pénis-vagin) est estimé à 4 (IC à 95% 1–14) pour le donneur et de 8 (IC 95% 6–11) pour le receveur. Le risque est de 11 (IC à 95% 4–28) s'il s'agit d'une relation sexuelle pénis anale (472-474). Le risque de transmission par la voie de relation sexe oral est assez faible (IC 95% 0–4) (467).

L'efficacité de la transmission de l'infection par le VIH dépend aussi des facteurs liés au receveur comme des ulcérations génitales ou des muqueuses, mais aussi du donneur, comme le niveau élevé de la charge virale et le stade de l'infection (récence ou chronique). Elle dépend aussi de la présence dans les liquides de transmission, de variants du virus qui ont une grande capacité d'établir efficacement l'infection (355, 475). Ces variants sont appelés *transmitted/founder virus* (Figure XXII). Environ 80-90% des transmissions de VIH sont attribuées à ce seul variant du virus ou *transmis/fondateur*. Cependant, selon le mode de transmission de la maladie (acquisition directe ou indirecte) plusieurs variants du VIH pourraient établir l'infection (355, 476-478). On estime que : environ 80% des transmissions hétérosexuelles (Het) (303, 479), 75% des HARSAH (475, 480, 481), 40–80% des utilisateurs de drogues par injection (IVIDU)(481-483) et 70% des transmissions mère-enfant (TME) sont dues à un seul variant du VIH-1(484). *Pour plus de détails sur ce sujet, voir chapitre IV, Article 3 de cette thèse.*

4.2 Virus transmis/fondateur ou « transmitted/founder viruses ».

4.2.1 Caractérisation et mode de sélection.

Par définition, un virus transmis/fondateur (303, 485-488) est un variant unique du VIH qui, parmi un ensemble de virions atteignant la surface des tissus lymphoïdes sous-muqueux ou

locorégionaux au cours de la transmission, aura la capacité d'infecter efficacement la cellule cible, de se répliquer, de se propager à l'ensemble des autres cellules pour conduire à une infection systémique avec des conséquences cliniques manifestes (355, 478, 486, 488) (Figure XX). Au cours de la transmission du VIH, des milliers de variants du virus sont exposés à la surface des muqueuses du receveur. Parmi ces nombreux virions qui ont pu se maintenir et survivre dans les fluides biologiques du donneur, les virus T/F seront capables de franchir toutes les barrières mécaniques, physiologiques et immunologiques pour établir l'infection chez le receveur (355, 475, 478, 485). Une fois présent dans les fluides biologiques du donneur, le virus TF sera confronté à plusieurs facteurs qui vont jouer des rôles de facilitation et dans certains cas d'inhibition pour permettre sa dissémination depuis le point d'entrée. Le mucus, le pH, la qualité de l'épithélium, la disponibilité des cellules cibles, des cellules innées et adaptatives, des chimiokines/cytokines inhibitrices et des cellules présentatrices d'antigènes sont des facteurs facilitateurs ou inhibiteurs (355, 475). Comme représenté dans la figure XX, les virus T/F se caractérisent (355) par : (a) leur résistance aux anticorps autologues présents dans les liquides de transmission (sperme ou le mucus cervico-vaginal) du donneur, (b) ils sont faiblement glycosylés leur permettant de ne pas se lier par des lectines présentes dans le liquide de transmission du donneur; (c) ils sont faiblement recouverts de glycanes ce qui facilite leur migration à travers le liquide de transmission du donneur; (d) ils possèdent des protéines d'enveloppe bien adaptées à la pénétration cellulaire de la sous-muqueuse génitale du receveur; (e) contrairement aux autres variants qui ont un phénotype à faible capacité de réplication puisqu'ils infectent préférentiellement les cellules cibles au repos (macrophages et lymphocytes T CD4+ au repos) qui sont faiblement activées donc incapables de faciliter une expansion virale rapidement.

Les virus T/F sont fortement transmissibles et se répliquent rapidement, car leur phénotype a une grande capacité à infecter préférentiellement des cellules cibles qui l'amplifieront rapidement (les lymphocytes T CD4+ activés). Le taux de transmission (R_0) du virus T/F est supérieur à 1. Ce taux leur confère la capacité d'établir efficacement l'infection et de transmettre le VIH, contrairement aux virus dont le R_0 est <1 . Le R_0 (*risque résiduel*) est une estimation du nombre de cellules cibles qui seront infectées à partir d'une seule cellule

infectée. S'il est <1 , le processus de réplication reste inachevé entraînant un échec de la transmission, s'il est >1 , le processus de réplication virale va jusqu'à son terme et l'infection est établie efficacement.

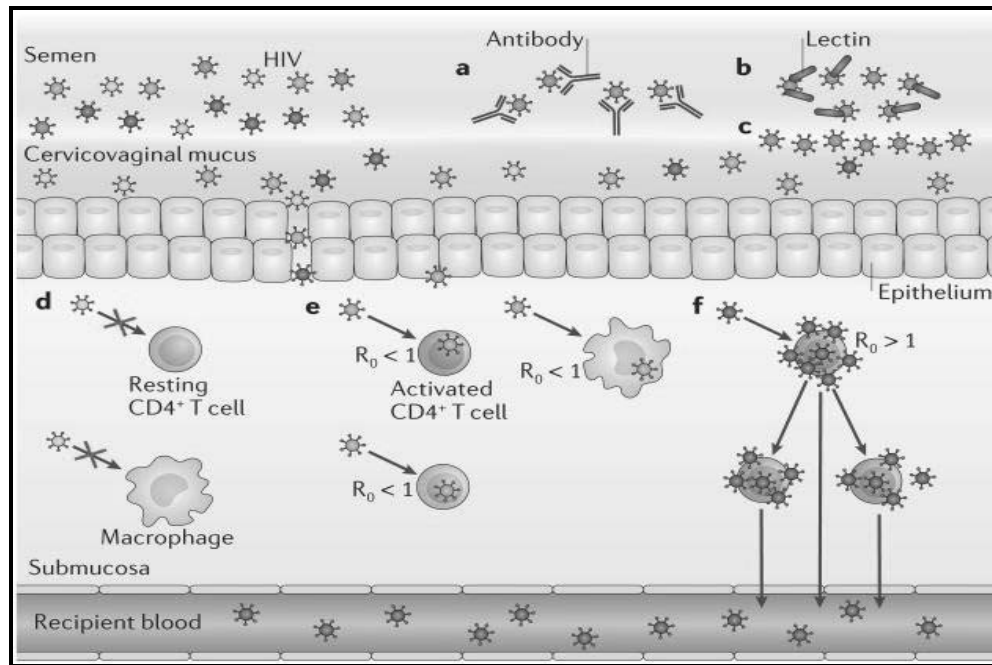


Figure XX. Caractéristiques et mécanismes de sélection du virus transmis/fondateur.

La figure ci-dessus décrit les modes de sélection et les caractéristiques d'un virus transmis/fondateur. Elle montre que parmi la multitude des variants exposés au cours de la transmission (voie des muqueuses), seulement un seul sera transmis au receveur. Les virus transmis/fondateurs infectent préférentiellement les cellules T activées où ils se répliquent rapidement. Il est aussi capable de résister à certains facteurs biologiques du donneur ainsi que du receveur. Figure reproduite avec autorisation. Source: Joseph SB et al. 2015 (355).

4.2.2 Propriétés génotypiques et phénotypiques des virus transmis fondateurs.

Les virus T/F présentent généralement un tropisme pour le récepteur à C-C chimiokine de type 5 (CCR5), plutôt que le récepteur C-X-C chimiokine de type 4 (CXCR4) (212, 303, 489-491). Il infecte préférentiellement les lymphocytes T CD4 (303, 492-494). Les propriétés phénotypiques suivantes sont propres au virus T/F (495) : 1) une forte capacité d'interaction et de liaison aux cellules dendritiques (CDs) qui en tant que cellules présentatrices d'antigènes (APC) contribuent à l'expansion et à la dissémination virale dans d'autres

compartiments de l'organisme (495), facilitant ainsi le transfert du virus aux cellules T CD4+ naïves (495) ; 2) contiennent plus de particules d'enveloppe à leur surface ce qui accroîtrait leur pouvoir infectieux (495) ; 3) résistent à l'interféron α , une cytokine capable d'inhiber la multiplication virale en rendant les cellules cibles résistantes à l'infection ou indirectement en activant les cellules NK (495) ; 4) leur enveloppe présente une grande capacité de liaison à l' $\alpha 4\beta 7$ qui est un récepteur d'intégrine présent à la surface des lymphocytes T CD4 + activés au niveau du site de transmission sexuelle (496, 497). La liaison du virus à ce récepteur d'intégrine facilite plus tard son expansion dans l'intestin (478).

Les virus T/F présentent aussi des propriétés moléculaires comme une réduction de la taille des sous-domaines gp120 V1/V2, V3 et V4 associés à de faibles nombres de sites de N-glycosylation et une charge positive nette assez faible des séquences virales de la boucle V3 comparativement aux virus d'infections chroniques (303, 477, 494, 498-502). Les propriétés ainsi décrites sont associées aux sous- types A, C, D et G du VIH-1. Mais les résultats contradictoires pour le sous-type B du VIH-1 pourraient être en lien avec les modes de transmission de la maladie qui est généralement de type anal (501, 502). D'autres signatures génétiques particulières associées aux virus T/F ont aussi été décrites (503). Il s'agit, entre autres, de la présence à la position 12 du peptide signal d'un résidu histidine (503), et d'une réduction du nombre de sites de N-glycosylation aux positions 413-415 de la boucle V4 de la gp120 (503) (cf chap. IV. Article 3).

4.3 Virus transmis/fondateur et développement de vaccin anti-VIH.

À ce jour, aucun moyen de prévention efficace et définitive n'est disponible pour prévenir la transmission de la maladie en dehors des méthodes traditionnelles comme le port du condom (504) et la prophylaxie préexposition (504-506). Développer un vaccin préventif et/ou thérapeutique contre le VIH-1 demeure un défi majeur à relever (8). Parmi la multitude des essais vaccinaux développés au cours des dernières décennies (507), six ont pu atteindre un niveau de développement clinique assez avancé (Vax004, Vax003, HVTN 502/STEP Trial, HVTN 503 (Phambili trial), RV144, HVTN 505) (508, 509).

Cependant, un seul, le RV144 dont l'essai s'est conduit entre les années 2003 et 2009 en Thaïlande a pu conférer jusqu'à 31,2% d'efficacité à 42 mois et 60% à 12 mois (510-512). Cette protection partielle induite par RV144 est médiée par une cytotoxicité cellulaire dépendante de l'anticorps (ADCC : *antibody-dependent cellular cytotoxicity*) (513).

Le vaccin RV144 a été testé dans un essai clinique randomisé, multicentrique et à double aveugle avec un contrôle placebo utilisant deux vaccins combinés incluant ALVAC-HIV (vCP1521) de Sanofi Pasteur et de « AIDS VAX B/E » de « Global Solutions for Infectious Diseases » (510). L'ALVAC-HIV (vCP1521) est un vecteur viral (canarypox) qui code pour les gènes *gag*, *pol*, et le domaine transmembranaire de la gp41 d'une souche du VIH-1 de clade B, mais aussi du fragment gp120 de l'enveloppe d'un isolat de CRF01-AE. AIDS VAX B/E contient une forme recombinante monomérique de la gp120 d'un isolat de CRF01 et de clade B. L'ALVAC-HIV était administré à 0, 1, 3 et 6 mois. AIDS VAX B/E plus un adjuvant (Alun) a été ajouté au 3^{ème} et 6^{ème} mois (510).

Deux corrélats de protection ont été induits par ce vaccin RV144. Le premier est une corrélation entre la reconnaissance des protéines de l'enveloppe du VIH-1 par des immunoglobulines de type A (IgA) et le risque accru d'être infecté par le VIH-1. Le deuxième est une corrélation entre la reconnaissance d'épitopes cibles au niveau des boucles V1/V2 de la gp120 par des IgG3 et la protection contre l'acquisition de la maladie (514, 515). Les IgA semblent avoir contribué à l'acquisition de la maladie en inhibant la réponse cytotoxique dépendante de l'anticorps (ADCC), la neutralisation par les anticorps et la capacité des cellules tueuses naturelles (NK) à induire la lyse des cellules T CD4+ infectées (513, 516-518). Des études récentes démontrent également que pour le vaccin RV144: 1) l'interféron de type I activateur des propriétés antivirales de *Interferon Regulatory Factor 7* (IRF7) et les gènes de stimulation de l'interféron de type II (*Type II interferon-stimulated genes*) qui sont impliqués dans la présentation antigénique sont associés avec une réduction du risque d'acquisition du VIH-1 (519) chez les sujets vaccinés non infectés; 2) une analyse des échantillons de l'essai vaccinal RV144 identifie: IRF7 comme un médiateur de protection et l'activation de *Mammalian target of rapamycin* (mTORC1) comme un corrélat du risque d'acquisition du VIH-1 (519). Malgré cette avancée, il reste encore nécessaire d'étudier les

caractéristiques génotypiques et phénotypiques de l'enveloppe des virus transmis/fondateurs (T/F) afin de mieux outillé la conception d'un vaccin efficace et de large spectre (281).

En effet, comme discuté précédemment au point 4.2, la population virale au cours de la primo-infection est quasi-homogène et se caractérise par des signatures génétiques particulières qui constituent des facteurs indispensables à l'établissement de la transmission du virus, à l'accroissement et au maintien de la capacité de réplication de ces premiers variants viraux (T/F) (281, 303, 520-524). Lesquels facteurs permettent aux virus T/F de dominer ou de contourner les mécanismes défenses développés par le système immunitaire. Ainsi, pour concevoir un vaccin efficace contre le VIH-1, il faudrait prendre en compte non seulement les caractéristiques des variants mineurs (524), des différents sous-types du VIH-1, des voies de transmission et les caractéristiques des cellules T CD8+ (525-527). Ces facteurs agissent précocement et sont susceptibles d'affecter l'efficacité du vaccin (cf chap. IV, article 3). Prendre en compte les propriétés génotypiques et phénotypiques des virus T/F ne déroge pas aux stratégies développées dans les vingtaines d'essais cliniques en cours de développement dans le monde à ce jour (509). Ces stratégies se résument à : 1) développer un vaccin qui va induire des réponses avec des anticorps neutralisants (bNAbs) spécifiques à la boucle V2 de l'enveloppe pour catalyser sur les acquis obtenus avec le vaccin RV144 dans d'autres contextes épidémiologiques (Afrique du Sud) (528, 529) et/ou capable d'induire des réponses anticorps non neutralisantes polyfonctionnelles (509, 530-533); 2) développer un vaccin à base d'anticorps neutralisants à large spectre qui sera déjà prêt à être administré de façon passive (529, 534-538); 3) développer un vaccin capable d'induire des cellules T mémoires effectrices au niveau des sites d'infections des muqueuses (539, 540).

Des résultats par ailleurs encourageants ont déjà été obtenus avec un nouveau vaccin en développement (essai clinique NCT02315703 (APPROACH trial)). Ce vaccin obéit dans sa conception aux stratégies énoncées (*induction de réponses anticorps polyfonctionnelles*) et les premiers résultats déjà publiés le 06 juillet 2018 dans le journal « The Lancet », (541). Ces résultats donnent 100% des réponses anticorps spécifiques de l'enveloppe du VIH-1 et

d'anticorps dépendant de la phagocytose cellulaire (80%) à 52 semaines et de réponses T (83%) à 50 semaines chez l'humain.

4.4 Traitement de l'infection par le VIH-1.

L'une des plus grandes avancées scientifiques dans la lutte contre l'infection par le VIH aura été la découverte des médicaments antirétroviraux (ARVs). Le premier ARV à être autorisé par la Food and Drug Administration (FDA) est la zidovudine (AZT), le 19 mars 1987. Les traitements antirétroviraux ne guérissent cependant pas la personne infectée de l'infection. Ils parviennent néanmoins à contrôler le taux de réplication virale en réduisant le niveau de la charge virale plasmatique jusqu'à un seuil indétectable. Cette réduction de la virémie permet ainsi d'améliorer non seulement la qualité de vie des personnes infectées, mais aussi contribue à réduire le taux de mortalité et les conséquences sociales qui y sont associées. En date du 12 avril 2018, environ 24 médicaments antirétroviraux et 16 combinaisons thérapeutiques étaient approuvés par la FDA (542). L'ensemble de ces médicaments est divisé principalement en six classes en fonction de leurs mécanismes d'actions sur le plan moléculaires (interrompre une étape du cycle de réplication virale) et de leurs profils de résistance (543). Les principales classes de traitements'antirétroviraux sont : 1) les inhibiteurs nucléosidiques et non nucléosidiques de la reverse transcriptase (INRTI/ (INtTI) qui ont pour objectifs de bloquer l'action de la transcriptase inverse (TI), l'enzyme impliquée dans la transcription de l'ARN viral en ADN du cycle de réplication ; 2) les inhibiteurs de l'intégrase (INs) qui bloquent l'action de l'intégrase, cette enzyme étant requise pour l'intégration du génome du virus dans celui de la cellule hôte; 3) les inhibiteurs de la fusion (FIs) qui bloquent la liaison des protéines de l'enveloppe virale aux récepteurs CD4 présents à la surface des lymphocytes T ; 4) les antagonistes du corécepteur (CCR5) qui ont pour rôle de bloquer la liaison du virus au Corécepteur CCR5 présent à la surface de certaines cellules immunitaires (macrophages, monocytes) ; 5) les inhibiteurs de la protéase permet de scinder les protéines précurseurs *gag* et *pol* pour favoriser la production de particules virales désorganisées sur le plan structurel et fonctionnel et donc non infectieux et 6) les inhibiteurs post-attachement

du virus dont l'objectif est de bloquer l'expression des récepteurs CD4 à la surface de certaines cellules des lymphocytes T.

L'un des plus grands obstacles à la guérison complète de l'infection par le VIH-1 par les traitements antirétroviraux demeure les réservoirs viraux qui se constituent très tôt après la transmission (424, 425, 544, 545). Ces réservoirs sont généralement inaccessibles par les traitements antirétroviraux disponibles et le VIH est capable de se réactiver à tout moment après un arrêt ou échec thérapeutique(546, 547). Le virus se cache dans des sanctuaires anatomiques ou tissulaires servant de réservoirs de réplication du VIH comme : 1) le tissu lymphoïde, 2) le tractus gastro-intestinal, 3) le cerveau et le système nerveux central (SNC), 4) les voies génitales, 5) le sperme et 6) les poumons (548). Les principaux réservoirs cellulaires du VIH-1 (548) sont : 1) les ganglions lymphatiques ils sont constitués par l'ADN proviral (provirus) intégré dans les lymphocytes T activés, les cellules dendritiques, les lymphocytes T mémoires, les lymphocytes T au repos, les monocytes-macrophages. La présence de provirus dans une cellule T CD4+ au repos est typiquement associée à la latence, 2) le système nerveux central où l'ADN proviral du VIH-1 est intégré dans les macrophages périvasculaires et microgliaux, 3) le sang où l'ADN proviral du VIH-1 est intégré dans les mastocytes, les cellules tueuses naturelles ou NK, les lymphocytes T activés ou au repos.

4.5 Réponse de l'hôte face à l'infection par le VIH-1.

4.5.1 Immunité innée.

Au cours de l'infection par le VIH-1, la première ligne de défense est constituée par les barrières mécaniques et cellulaires présentes au niveau du site de l'infection. Les barrières cellulaires sont constituées par des cellules épithéliales, du système du complément, et des cellules présentant des propriétés phagocytaires (*cellules tueuses naturelles ou NK*) et de présentation d'antigènes (APC : *granulocytes, les macrophages et les cellules dendritiques ou CD*s) (549, 550).

Les barrières mécaniques comprennent les muqueuses et les sous-muqueuses de l'hôte notamment l'épithélium. C'est ce dernier qui contient et active les cellules de l'immunité innée (*cellules dendritiques, macrophages, monocytes, polynucléaires neutrophiles et*

cellules NK) à travers une cascade de signalisation via les TLR (*Toll-like receptors*). Le TLR est un récepteur membranaire membre de la famille des récepteurs de reconnaissance de motifs (*Pattern recognition receptors*, PRR). Les PRR jouent un rôle central dans l'immunité innée par la reconnaissance des structures conservées de façon évolutive sur le VIH, appelées modèles moléculaires associés aux pathogènes (*pathogen-associated molecular patterns*, PAMP) qui sont toujours différents des cellules du soi (549-553). La reconnaissance des virus est principalement médiée par le TLR 9 pour les virus à ADN, les TLRs 7&8 pour les virus à ARN monocaténaire (ssRNA) et le TLRs 3 pour les virus à ARN bicaténaire (dsRNA) (554-558). Un autre récepteur d'adhésion de surface cellulaire, nommée DC-SIGN, joue aussi un rôle dans la médiation des interactions entre les CD4 et les cellules T en facilitant l'infection-trans des cellules T par les CD4 (559). De plus, le VIH-1 exploite l'interaction entre DC-SIGN et le facteur de signalisation de l'immunité innée médiée par TLR8 pour établir une infection productive dans les DCs (560). Outre les récepteurs de surface décrits, d'autres PRR notamment des hélicases jouent des rôles importants dans la détection des ARNs Intracytoplasmiques (561, 562). Il s'agit principalement du retinoïd *acid-inducible gene* RIG-like receptors (RLR) : *retinoic-acid-inducible protein 1* (RIG-I) et *melanoma-differentiation-associated gene 5* (MDA5). Le recrutement de molécules inhibitrices comme : SDF-1 ou *Stromal cell-derived factor 1* (CXCL12) qui accroît l'entrée des virus X4-tropic HIV-1 dans les lymphocytes TCD4+ mémoires et favorisant ainsi la sélection de variant X4 et aussi retardant leurs propagations jusqu'au stade chronique de l'infection par le VIH -1 (563, 564). MIP-1 α/β (*macrophage inflammatory protein-1 alpha/beta* (CCL3/4) et RANTES (*regulated on activation normal T expressed and secreted*) sont également des molécules inhibitrices qui bloquent la liaison du virus sur le corécepteur CCR5 présent à la surface des monocytes macrophages (565).

La réplication initiale du virus a lieu à l'intérieur des cellules T CD4+ au niveau du site d'infection (muqueuses) (566). Le début de cette réplication conduit au recrutement des cellules dendritiques qui vont induire la production de cytokines inflammatoires. Celles-ci vont non seulement jouer une activité antivirale directe au niveau du site d'infection, mais indirectement en activant d'autres cellules de l'immunité innée ayant une activité antivirale

(NK) (567). Les cellules NK jouent un rôle antiviral direct (*activation ou inhibition*) à travers la reconnaissance de motifs de détresse cellulaire. Ces motifs se caractérisent par la présence du récepteur appelé « *killer immunoglobulin-like receptors (KIRs)* (568-571) à la surface des cellules infectées par le VIH. La liaison entre le récepteur et son ligand présent sur le NK va conduire à la destruction de la cellule infectée par cytolysse ou par apoptose (567, 572). L'expression à la surface de plusieurs cellules cibles par les cellules de l'immunité innée, des facteurs inflammatoires et de l'interféron gamma (IFN- γ) contribue au contrôle de l'infection. Ils créent par la même occasion des conditions favorables à la transmission cellule-cellule du virus, mais également de canaux de transport et de dissémination du virus dans d'autres compartiments de l'organisme (573, 574) (Figure XXI).

L'interféron alpha (INF α) ou interféron de type I) est également produit par les cellules dendritiques plasmacytoïdes (pDCs). En début d'infection par le VIH-1, les taux plasmatiques d'INF α sont augmentés dès la première semaine d'infection (575). L'INF α peut induire des effets bénéfiques en bloquant l'acquisition du virus notamment au niveau des muqueuses ou en intervenant aux différents stades du cycle de réplication pour limiter la réplication virale une fois l'infection déjà établie par plusieurs mécanismes (576) par : 1) stimulation des cellules T CD8 pour réguler positivement les molécules du CMH qui en se liant aux récepteurs des inhibiteurs des cellules NK leur rendent, résistant à la cytotoxicité induite par les cellules NK, 2) faciliter la prolifération et la survie des cellules NK et activation de la cytotoxicité des cellules NK, contribuant ainsi à l'éradication des cellules T CD4 infectées par le VIH, 3) suppression des fonctions effectrices des lymphocytes T CD4 et inhibition de leurs proliférations, 4), induction de facteurs de restriction limitant la réplication du VIH à diverses étapes du cycle de réplication (entrée du VIH, la transcription inverse, l'entrée nucléaire, l'intégration, la transcription et le bourgeonnement). Ces différents facteurs de restriction sont TRIM5 α , SAMHD1, APOBEC3, MX2 (MX Dynamin-Like GTPase 2), Schlafen 11 et tetherin/BST2 (277, 577-585). En outre, il peut jouer un rôle délétère en raison des réponses proinflammatoires (579) qui conduisent dans le cadre de l'infection par le VIH à : une exacerbation de l'immunité systémique induisant une restauration médiocre des cellules TCD4+, le recrutement de cellules TCD4+ cibles, l'épuisement immunitaire et une progression

rapide de l'infection par le VIH vers le SIDA. La production d'interféron de type I induit l'apoptose des lymphocytes TCD4+ et limite les réponses des lymphocytes TCD4+ et CD8+ spécifiques (586). Cependant, la résistance apparente des virus transmis/fondateurs à l'INF α au cours de transmission (522, 587, 588) ne permet pas de bloquer la transmission virale, l'établissement rapide de réservoirs viraux et la progression de la maladie (576).

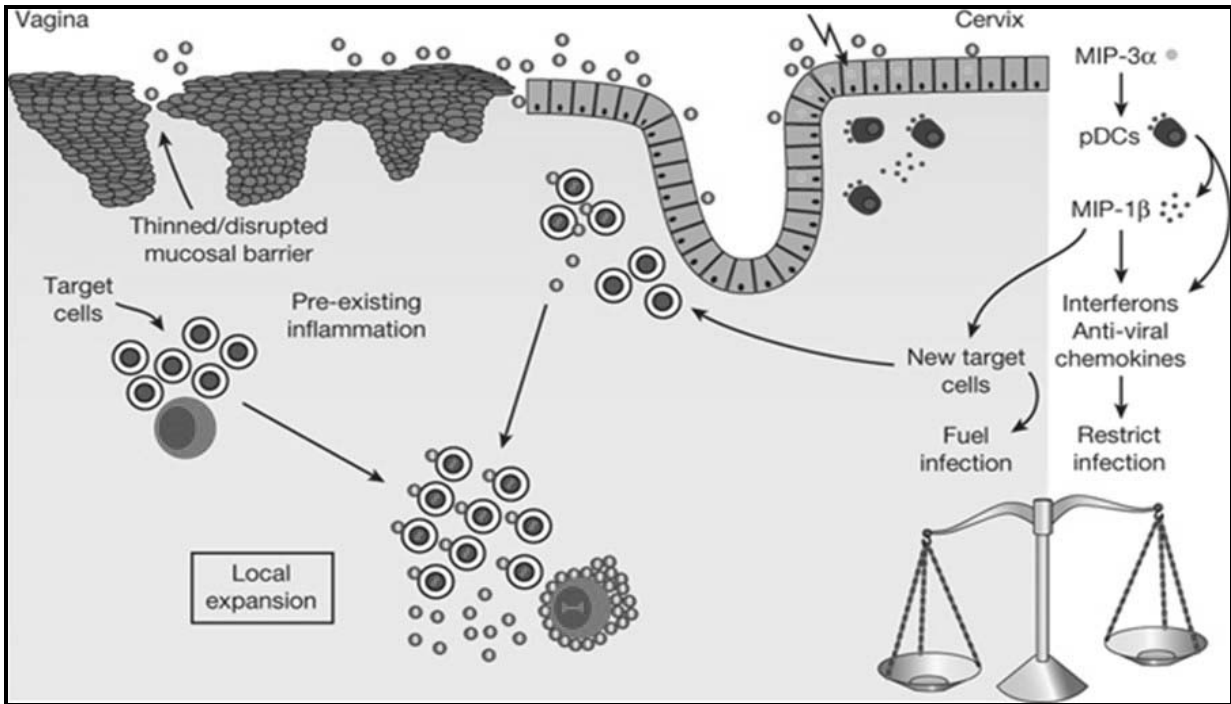


Figure XXI. Représentation schématique de la première ligne de défense contre le VIH.

La figure ci-dessus représente les principaux mécanismes de défense de première ligne contre le VIH. Ces facteurs sont : les facteurs inflammatoires et de signalisation épithéliale des muqueuses et des cellules de l'immunité innée et des cellules cibles qui sont présentes au niveau des lignes de front des muqueuses. La présence du virus dans l'épithélium endocervical déclenche la signalisation (flèche dentelée) qui augmente l'expression de MIP-3 α dans l'épithélium favorisant le recrutement des cellules dendritiques (pDCs). Celles-ci vont à leur tour via les chimiokines telles que le MIP-1 β activer les cellules T CD4 qui vont alimenter l'expansion virale locale. Les interférons produits par les pDCs et des chimiokines vont tenter d'inhiber la réplication virale, mais le virus par un certain nombre de mécanismes va contourner cette barrière et continuer son expansion locale, sa dissémination et l'établissement systémique de l'infection. Une inflammation vaginale ou locale préexistante facilite également l'infection. Figure reproduite avec autorisation. Source: Haase AT et al. 2010 (566).

4.5.2 Immunité adaptative et infection par le VIH-1.

L'immunité cellulaire ou adaptative est déclenchée après l'établissement de l'infection et l'ensemble des mécanismes de persistance à long terme (réservoirs viraux). L'immunité adaptative est médiée par les lymphocytes T CD8+ cytotoxiques et lymphocytes T CD4+.

4.5.2.1 Immunité adaptative médiée par les lymphocytes T CD8+.

Les réponses immunes T CD8+ spécifiques au VIH interviennent rapidement après l'immunité innée. Elles jouent un rôle déterminant dans le contrôle de l'infection au stade précoce qui se traduit non seulement par une baisse rapide du niveau de la charge virale plasmatique (288), mais aussi, par une expansion rapide du niveau des cellules TCD8+ (589-592), une prolifération et une différenciation des cellules T CD8+ naïves en cellules T CD8+ mémoires ou cytotoxiques (CTL) (593, 594). Le VIH n'a aucun tropisme pour ces cellules T CD8+. Le niveau des cellules T CD8+ naïves au cours de la progression de la maladie démontre une atteinte du processus ou de progéniteur lymphoïde T. Les cellules T CD8+ arrivent à contrôler l'infection virale au début de l'infection (288) en exerçant une activité cytolytique (lyse des cellules T CD4+ infectées) par plusieurs mécanismes: une liaison avec les peptides viraux présentés par les HLAs (CMH-I) à la surface des cellules infectées formant un complexe qui va entraîner l'induction d'une réponse cytolytique (destruction de la cellule cible infectée). Cette propriété de cytolyse est médiée par la libération de perforine qui dégrade la cellule infectée en créant des trous à travers lesquels des granzymes vont s'introduire dans la cellule pour la détruire. Ce processus a lieu avant que le processus de génération de nouveaux virions soit complété. Elles peuvent également induire la mort programmée (apoptose) de la cellule cible infectée par l'expression des facteurs solubles. Les facteurs solubles impliqués dans l'apoptose sont : le Fas ligand, présent à sa surface de la cellule CD8+, en se fixant à son récepteur Fas (CD95) présent à la surface de la cellule cible induit l'apoptose de la cellule infectée (288, 595, 596); l'induction d'une activité antivirale médiée par l'interféron gamma (INF- γ) qui rend les cellules cibles assez résistantes à une infection productive grâce à l'activation induite par le récepteur, l'induction de la production de chimiokines telles que MIP-1 α/β et RANTES qui va se lier au CCR5 et provoquer sa dégradation/internalisation privant le virus de son principal corécepteur (597), minimisant ainsi le nombre de cellules

cibles qui seront infectées. Enfin, les cellules T CD8⁺ induisent l'apoptose des cellules infectées par le VIH par la sécrétion du CAF (*cell antiviral factor*) qui est un facteur possédant une activité antivirale (598) (inhibition de la transcription virale) (599). En dépit de ces différentes propriétés, l'efficacité de l'immunité adaptative médiée par les lymphocytes T CD8⁺ pour contrôler l'infection est incomplète et limitée dans le temps (589-592). Elle évolue en réponse aux mutations du virus dont les fréquences contribuent à rendre les cellules cibles infectées méconnaissables par les lymphocytes TCD8⁺ (600, 601). Malgré la forte implication des cellules T CD8⁺ dans le contrôle de la réplication virale, il n'en demeure pas moins que son activité à long terme est entravée par l'émergence des virus mutants résistants (600, 601) comme décrits plus haut. Cela conduit à une perte de la fonctionnalité : réduction des fonctions lytiques (602, 603), de prolifération (604), d'expression des molécules de signalisation (605) médiée par l'activation des récepteurs de cellules T (TCR) et l'épuisement lymphocytaire (606-609). En effet, l'efficacité des cellules T CD8⁺ spécifiques au VIH est en partie limitée par les effets de la stimulation immunitaire chronique sur la fonction des cellules T CD8⁺. Ces molécules co-inhibitrices qui altèrent la fonction antivirale des lymphocytes T CD8⁺ spécifiques du VIH comprennent : PD-1 (*programmed death 1*), TIM-3, (*T cell Ig and mucin domain 3*), CD160, le récepteur de cellule tueuse naturelle (NK cell receptor 2B4), LAG-3, (*lymphocyte activation gene 3*) et CTLA-4 (*cytotoxic T lymphocyte antigen 4*) (608, 610-619). En outre, une expression simultanée de plusieurs récepteurs de molécules co-inhibitrices peut entraîner une exacerbation de l'altération immunitaire fonctionnelle des Cellules TCD8⁺ (620). Cependant, le contrôle de l'épuisement des cellules TCD8⁺ pourrait fournir des outils pour l'élaboration des vaccins préventifs contre le VIH-1(621). Il s'agira dans les recherches vaccinales futures de : bloquer les récepteurs des molécules co-inhibitrices par exemple PD-L1 (*PD ligand 1*) pour PD-1 afin de restaurer les fonctions effectrices des réponses TCD8⁺ ou en combinant cette approche avec des thérapies à base d'anticorps neutralisants (bNAb), d'antirétroviraux, de thérapies cellulaires, du blocage des voies de régulation et l'exploitation d'autres fonctions cellules immunitaires (509, 608, 610, 621). La figure ci-dessous décrit les mécanismes d'activation des réponses T à la suite de l'infection par le VIH (Figure XXII).

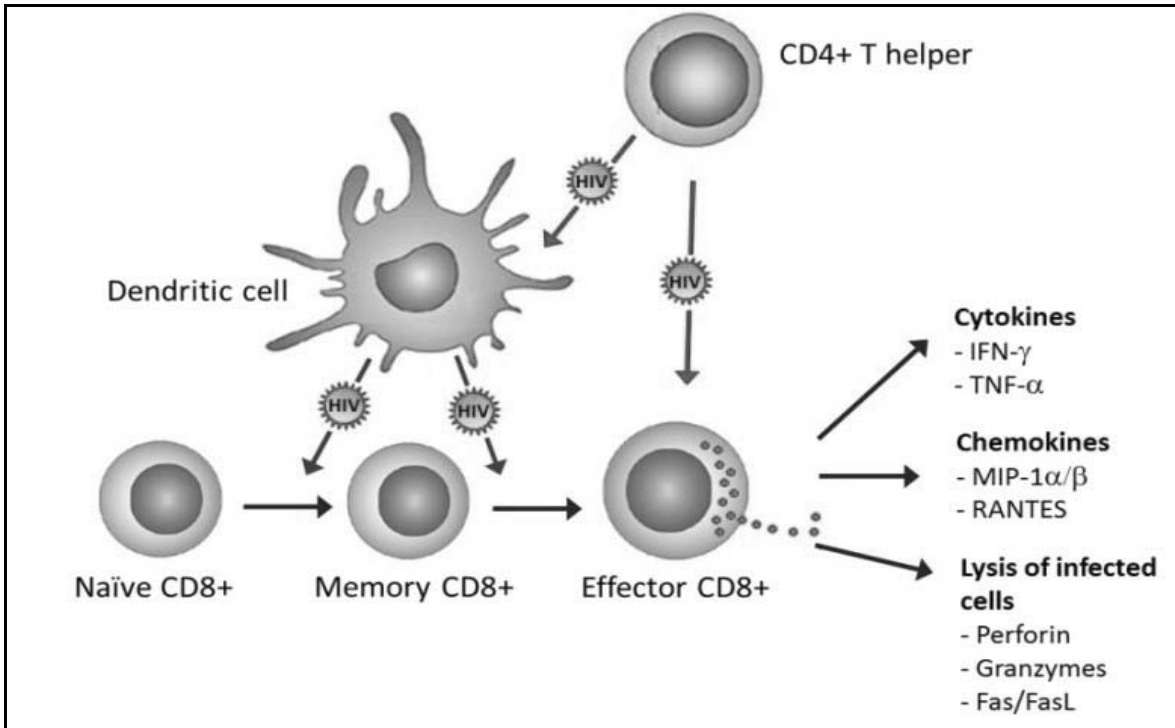


Figure XXII. Représentation schématique des réponses cellulaires T induites par suite de l'infection par le VIH.

Le schéma illustre les principales chimiokines et des cytokines produites par les cellules T CD8 + effectrices spécifiques du VIH. Les cellules T auxiliaires CD4+ aident à stimuler les cellules dendritiques et des cellules T CD8 + pour maintenir une réponse de T CD8+ mémoire. Le VIH inhibe généralement la fonction de coopération cellulaire entre les lymphocytes T CD4 + et T CD8+ afin de favoriser son évaison. Figure reproduite avec autorisation. Source: Winni DH et al. 2012 (622).

4.5.2.2 Immunité humorale médiée par les lymphocytes B.

Les réponses anticorps médiées par les lymphocytes (immunité humorale) par suite de l'infection par le VIH sont caractérisées par des taux d'anticorps élevés, mais non neutralisants pour la plupart (623). Malgré la pression que les lymphocytes B exercent, l'échec de réponse humorale est dû à l'émergence rapide de virus résistants aux anticorps sitôt après l'infection et à l'inaccessibilité des épitopes de neutralisation au niveau du trimère de l'enveloppe virale (624). Les nombreux sites de glycosylation présents sur les séquences de l'enveloppe virale contribuent aussi à cette résistance contre la neutralisation (623).

4.5.2.3 Immunité adaptative médiée par les lymphocytes TCD4+.

Les réponses immunitaires induites à la suite de l'infection par le VIH sont insuffisantes et inefficaces pour contrôler totalement la maladie. Cette dysfonctionnalité immunologique est causée par la destruction ou une altération des cellules T (625) par le virus. Elle se traduit par une déplétion progressive et précoce des lymphocytes T CD4+ entraînant une apoptose anormale de ces cellules (625) et une inflammation persistante qui conduit à long terme à l'immunosénescence cellulaire (625). La raison principale de l'inefficacité des cellules T CD4+ est l'échappement immunitaire relié à la diversité génétique (mutations fréquentes) du virus induit par les nombreuses erreurs de transcription inverse non correctives ($3,5 \times 10^5$ erreurs par base et par cycle de réplication) (503). L'expression de l'interféron induit une réponse contre les antigènes viraux. Cependant, la prolifération des cellules T est rarement observée chez une personne infectée non traitée à l'exception des sujets non-progresseurs à long terme. Cette défaillance des fonctions de restauration cellulaire est la conséquence de la capacité du virus à provoquer la destruction massive et précoce des cellules T. Ceci entraîne non seulement un épuisement cellulaire et/ou un dysfonctionnement des lymphocytes T CD4+ régulateurs (Treg), mais aussi des organes lymphoïdes qui constituent les sites de maturation et de production de cellules T (626). Cette perte de fonctionnalité est la conséquence de l'activation constante du système immunitaire qui entraîne une surproduction thymique (625). L'atteinte précoce des lymphocytes TCD4+ en tant que chef d'orchestre du système immunitaire abroge toutes possibilités de coopération cellulaire nécessaire à la mise en place d'une réponse immunitaire spécifique et efficace contre le VIH.

4.6. Anticorps neutralisants à large spectre contre le VIH.

4.6.1 Définition et rôle.

Un anticorps neutralisant à large spectre (bNAbs), dans le cadre de l'infection par le VIH, se définit comme un type d'anticorps qui se développe généralement dans les 2 à 4 années après l'infection, mais dans de rares cas pourrait se développer avant la première année (627). Ces anticorps à large spectre sont munis d'une capacité fonctionnelle d'inhibition de l'infection des cellules cibles d'un ou de plusieurs variants de VIH-1 (clades). Cette propriété

fonctionnelle des bNAbs se traduit par une reconnaissance des épitopes cibles conservés, présents à la surface de l'enveloppe du VIH. Par saturation du site antigénique, ils empêchent la fixation du virus à la surface cellulaire et/ou sa libération dans le cytoplasme (628-633) et empêchent la transmission cellule-cellule du virus (634). Par ailleurs, les anticorps neutralisants à large spectre peuvent jouer un rôle fonctionnel d'inhibition de l'infection par le VIH-1 par reconnaissance directe des cellules cibles infectées grâce à des motifs de détresse présents à la surface de l'enveloppe virale. La reconnaissance de ces motifs va conduire à l'induction d'une cytotoxicité cellulaire dépendant de l'anticorps (ADCC) par activation des cellules tueuses naturelles (*Natural killer*, NK) (635). Dès lors, comprendre les mécanismes d'action des anticorps neutralisants permet d'évaluer les réponses immunitaires spécifiques induites chez les personnes infectées par le VIH, d'évaluer l'efficacité des anticorps monoclonaux, d'étudier leurs rôles dans la diversité virale et d'évaluer l'immunogénicité des candidats vaccins anti-VIH en cours de conception (636). Les bNAbs peuvent être considérés et utilisés en thérapeutique anti-VIH comme des inhibiteurs d'entrée (634, 637). La liste des meilleurs anticorps neutralisants à large spectre est représentée dans la figure XXIII. Une liste complète des bNAbs identifiés à ce jour est disponible et régulièrement mise à jour dans le site du Los Alamos National Laboratoire (LANL) (224, 638) à l'adresse ci-dessous:

https://www.hiv.lanl.gov/content/immunology/tables/ab_best_neutralizing_summary.html

4.6.2 Localisation des principaux épitopes de neutralisation au sein du trimère de l'enveloppe du VIH-1.

Le trimère d'enveloppe du VIH-1 comprend également des régions conservées et des structures exposées constituant des cibles de la majorité des anticorps neutralisants (bnAbs) contre le VIH-1. Les meilleurs bnAbs identifiés sont dirigés principalement contre des épitopes présents au niveau du site de liaison au CD4 (*CD4 binding site*), des glycanes associées aux domaines C2, C3, V3, V4, et de la gp120, de la structure quaternaire formée par les loops V1/V2 ou *Quaternary neutralizing epitopes* (QNEs), mais aussi de la région externe de la gp41 (ectodomaine) proximale à la membrane virale (MPER, *Membrane-proximal external region*) (211, 508, 639-643). La figure XXIII ci-dessous illustre la localisation des

principaux épitopes de neutralisation du trimère de l'enveloppe du VIH-1. Le tableau résume l'ensemble des anticorps neutralisants à large spectre reconnus pour le moment et leurs épitopes cibles.

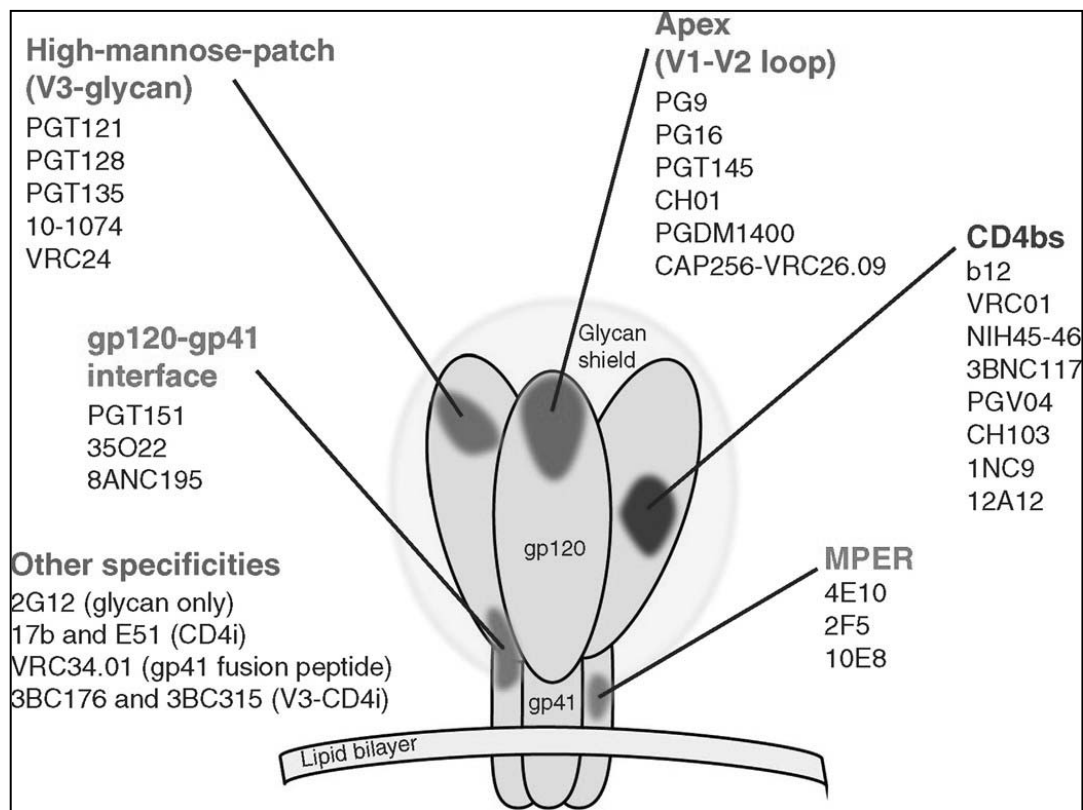


Figure XXIII. Représentation des différents anticorps neutralisants à large spectre (bNAbs) dirigés contre les régions conservées de l'enveloppe du VIH-1.

La figure ci-dessus illustre les principaux épitopes de neutralisation au sein du trimère de l'enveloppe du VIH-1 et les bNAbs qui sont associés. Y sont représentés : le site de fixation du CD4 (CD4bs), de la structure quaternaire formée par la gp120 V1-V2, du MPER, de l'interface gp120-gp41, de la glycane associée à la gp120 V3. Figure reproduite avec autorisation. Source : Fuchs, Sebastian P et al. 2016 (447, 644).

4.6.3 Mécanismes de résistance du VIH-1 à la réponse neutralisante.

Le VIH développe un certain nombre de mécanismes au niveau des glycoprotéines de son enveloppe pour résister à la neutralisation induite par les bNAbs produits par le système immunitaire (231, 623, 645-653). Parmi ces nombreux mécanismes de résistance, on peut citer entre autres :1) une très grande variabilité génétique des séquences virales de certaines

régions de l'enveloppe, dont les domaines hypervariables (V1-V5) de la gp120. L'hypervariabilité de ces domaines contribue à masquer les épitopes de neutralisation au niveau du spicule de l'enveloppe; 2) une flexibilité conformationnelle développée par certaines régions du virus où les N-glycanes sont impliqués, contribuant à réduire l'accessibilité de ses différents sites fonctionnels aux bnAbs (653); 3) un remodelage permanent de la distribution des N-glycanes sur l'enveloppe du virus pour créer ainsi des contraintes stériques restreignant ainsi l'accessibilité de la plupart des épitopes de neutralisation (Figure XXIV) (654). Ces *N-glycans* se dressent comme des boucliers protégeant le squelette protéique du virus contre les bnAbs (653). Ils protègent certaines régions de la gp120 contre leur dégradation protéolytique en réduisant le nombre d'épitopes T *helper* reconnu (655). Aussi, le VIH s'adapte constamment à la pression exercée par les bnAbs en modulant le nombre de *N-glycans* à la surface de la gp120 pour mieux masquer le squelette protéique contre la neutralisation (Figure XXIV) (653) et enfin; 4) le très haut degré de glycosylation (N ou O) contribue à cacher ou à masquer les sites vulnérables de la gp120 contre la neutralisation ou accroître le potentiel infectieux du virion (229). La N-glycosylation correspond à la fixation de structures glycosidiques sur l'asparagine d'un séquon NXS/T ou X représente n'importe quel acide aminé sauf la proline (653). La O-glycosylation correspond à la fixation d'un oligosaccharide sur un résidu sérine ou thréonine (653); enfin, 5) la modulation de la charge positive globale de la boucle V3 de la SUgp120 contribue non seulement à déterminer le tropisme viral, mais aussi à la résistance du virus contre la neutralisation (charge +3 V3) (656). La N-glycosylation de l'enveloppe du VIH-1 est beaucoup étudié principalement pour son implication dans l'antigénicité, l'immunogénicité et la neutralisation par des anticorps dirigés contre le VIH-1 (423, 657-661). Cependant, certains bNAbs (2G12, 8ANC195) reconnaissent des glycanes et non pas des régions masquées par ceux-ci (447, 644, 662-666).

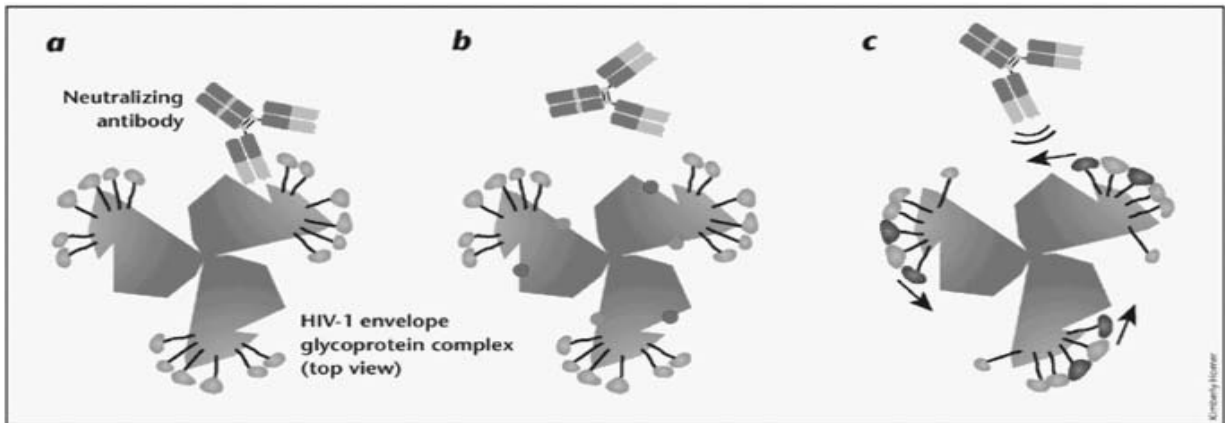


Figure XXIV. Mécanisme d'échappement du VIH-1 aux anticorps neutralisants médié par la formation de bouclier de glycane sur les épitopes de neutralisation au niveau de l'enveloppe virale.

Sont illustrés dans la figure ci-dessus : (a) le mécanisme de liaison des anticorps neutralisants au complexe trimérique des glycoprotéines d'enveloppes virales (points bleus, gp120); (b) les mécanismes d'échappement viral à la neutralisation impliquant des mutations aléatoires au niveau de l'enveloppe sont représentés par des points rouge et orange. Ces mutations modifient l'épitope neutralisant (points rouges) et diminuent la liaison de l'anticorps ; (c) les mécanismes d'échappement viral à la neutralisation par la formation de bouclier glycane selon Wei *et al.* (623) induit des blocages stériques au niveau de l'enveloppe qui empêche les anticorps neutralisants de se lier à leurs épitopes apparentés (points violets). Les ajouts ou le repositionnement des glycanes sont indiqués en violet. Figure reproduite avec autorisation. Source: Mascola JR, Montefiori DC. 2003 (654).

Le VIH-1 se caractérise par une diversité des protéines externes et internes dues aux nombreuses erreurs de la transcriptase inverse. Les protéines externes constituées par les glycoprotéines de l'enveloppe sont les plus exposées, les plus variables interagissant avec les récepteurs et corécepteurs cellulaires pour établir l'infection par le VIH-1. Elles sont les seules protéines virales exprimées à la surface des virions et constituent ainsi les cibles des anticorps neutralisants. Elles contiennent des régions hypervariables fortement glycosylées et variables en tailles servant de mécanismes d'échappement à la réponse immunitaire développée par le système immunitaire de l'hôte contre l'infection par le VIH-1. Les virus T/F sont les variants responsables de la transmission du VIH-1. Il serait important d'étudier les séquences des protéines d'enveloppe de ces virus pour mieux comprendre les événements

moléculaires développés au cours de la transmission pour permettre d'identifier d'immunogènes capables d'induire des réponses anticorps neutralisants de large spectre contre le VIH-1.

CHAPITRE II

HYPOTHÈSES ET OBJECTIFS

1. Hypothèses de l'étude.

La transmission du VIH-1 à un nouvel hôte requiert généralement un seul variant du virus nommé *virus transmis/fondateur* pour « *transmitted/founder* » *virus* (TF). Ce virus est sélectionné parmi une multitude de virus exposés à la surface des muqueuses du donneur et du receveur et a la capacité d'établir efficacement l'infection. L'enveloppe est la première région du génome viral à interagir avec les récepteurs du système immunitaire de l'hôte. Puisque le virus T/F arrive à déjouer l'ensemble des mécanismes de résistance de l'organisme pour établir l'infection, nous fondons l'hypothèse qu'il possède des caractéristiques génotypiques et/ou phénotypiques spécifiques. Mieux caractériser ces propriétés moléculaires de l'enveloppe des virus fondateurs contribuera à mieux informer pour la conception de vaccin contre le VIH-1. Au cours de la primo-infection, les séquences virales de l'enveloppe des virus transmis fondateurs sont homogènes, puis elles se diversifient en fonction de l'évolution de la maladie. La mise en place d'une méthode moléculaire pour le suivi de la diversification des séquences de l'enveloppe chez les sujets infectés par le virus VIH-1 constituera un biomarqueur pour la détection des nouvelles infections, la prédiction de l'évolution de la maladie et l'identification des réseaux de transmission de la maladie dans la population.

2. Objectifs de l'étude.

2.1 Objectif principal.

Établir une caractérisation moléculaire et phylogénétique des séquences de l'enveloppe de virus transmis/fondateurs du VIH-1.

2.2 Objectifs secondaires.

2.2.1 Objectif secondaire : Identifier des signatures génétiques particulières au niveau de l'enveloppe de virus transmis/fondateur susceptibles de mieux informer les événements précoces de la primo-infection par le VIH-1.

2.2.2 Objectif secondaire 2 : Identifier une mesure de diversité génétique de la séquence de l'enveloppe virale capable de mieux prédire les nouvelles infections à VIH-1.

2.2.3 Objectif secondaire 3:

Identifier les réseaux (*cluster*) précoces de transmission de l'infection par le VIH-1 par des analyses phylogénétiques utilisant deux fragments de séquences de l'enveloppe du VIH-1

CHAPITRE III
RÉSULTATS PAR ARTICLE

ARTICLE 1

HIV-1 ENVELOPE SEQUENCE-BASED DIVERSITY MEASURES FOR IDENTIFYING RECENT INFECTIONS (292).

Alexis Kafando¹, Eric Fournier², Bouchra Serhir², Christine Martineau², Florence DouallaBell^{2,3}, Mohamed Ndongo Sangaré⁴, Mohamed Sylla⁵, Annie Chamberland⁵, Mohamed El-Far⁵, Hugues Charest^{1,2}, Cécile L. Tremblay^{1,2,5*}

Publié dans le journal : PLoS One. 2017 Dec 28;12(12). doi: 10.1371/journal.pone.0189999.

¹ Département de microbiologie, infectiologie et immunologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada,

² Laboratoire de santé publique du Québec, Institut national de santé publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada,

³ Département of medicine, division of expérimental medicine, McGill University, Montréal, Québec, Canada,

⁴ Département de médecine sociale et préventive, École de santé publique, université de Montréal, Montréal, Québec, Canada,

⁵ Centre de recherche du centre hospitalier de l'Université de Montréal, Montréal, Québec, Canada.

§ Corresponding author: Cécile L. Tremblay, MD, FRCPC

Département de Microbiologie, infectiologie et immunologie

Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM)

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

E-mail: [REDACTED]

CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 1

M. Kafando a contribué à l'article 1 par la conception de l'étude selon les directives du Dre Cécile Tremblay. Il a passé en revue les procédures expérimentales de l'étude sous la supervision du Dr Hugues Charest, Dre Christine Martineau et du Dre Annie Chamberland. M Kafando a réalisé toutes les expériences de laboratoire passant par l'identification des échantillons disponibles dans la serothèque du Laboratoire de santé publique du Québec (sous le contrôle du Dre Bouchra Sherir), à l'extraction des ARN du VIH, aux analyses de biologie moléculaire (PCR), aux séquençages (NGS) et à l'assemblage des séquences (supervision M Éric Fournier). M Kafando a réalisé les analyses bio-informatiques et une partie des analyses statistiques (supervision Dr Mohamed Ndongo Sangaré). Il a également assuré la synthèse de tous les résultats de l'étude et rédigé le premier manuscrit qu'il a soumis à l'ensemble de l'équipe de recherche pour les révisions et commentaires. Il a par la suite rédigé le manuscrit final, identifié le journal en concertation avec l'équipe et procédé à sa soumission. Il a également préparé les séquences de l'enveloppe du VIH-1 de tous les sujets de l'étude qu'il a soumis dans la base de données de Genbank en collaboration avec Éric Fournier.

M. Kafando a interagi avec les éditeurs du journal et les reviseurs de l'article et a apporté toutes les corrections nécessaires au manuscrit en collaboration avec les coauteurs. M Kafando a rédigé la version finale du manuscrit soumise au journal qui a été publiée.

En somme, M. Kafando a participé à la conception de l'étude, à la réalisation des expériences de laboratoire et a été l'interface entre les différents coauteurs, les éditeurs et les reviseurs pour la rédaction du manuscrit.

Les contributions de chacun et de chacune des coauteurs sont décrites plus en détail dans l'article.

RÉSUMÉ

Contexte : Identifier les nouvelles infections à VIH-1 est crucial pour une meilleure évaluation de l'incidence de la maladie dans la population et contribue à une optimisation des efforts de prévention en matière de santé publique. Nous avons évalué et comparé la performance de quatre mesures de diversité génétiques des séquences virales de l'enveloppe du VIH -1 dans leurs capacités à bien identifier les nouvelles infections au VIH-1. Ces quatre mesures comprennent le pourcentage de diversité, le pourcentage de complexité, l'entropie de Shannon et le nombre d'haplotypes ciblant treize segments génétiques de l'enveloppe VIH-1.

Méthodes : Au total, 597 échantillons provenant de patients infectés par le VIH en 2013 et 2015 ont été sélectionnés et analysés. Mais à l'issue des analyses d'amplification par réactions de polymérisation en chaîne (PCR), du séquençage (nouvelles générations utilisant la technologie MiSeq d'Illumina) et du traitement bio-informatique des séquences virales, un total de 249 séquences majoritaires représentant 134 infections récentes et 115 d'infections chroniques ont été retenues. Les séquences des régions codantes pour les domaines V1-C5 de la glycoprotéine gp120 et une portion de la gp41 réparties pour les treize segments génétiques ont été analysées. La capacité des quatre mesures de diversité génétiques à pouvoir identifier correctement les infections récentes au VIH-1 a été évaluée en utilisant des courbes de distribution de fréquence, de la médiane et l'interquartile et l'aire sous la courbe (AUC).

Résultats : Les mesures de la médiane et l'interquartile et l'appréciation des courbes de distribution de fréquence associées aux 4 mesures de diversité ont permis de démontrer que le pourcentage de diversité, le nombre d'haplotypes et l'entropie de Shannon présentaient un potentiel statistiquement significatif pour la différenciation des infections chroniques et récentes ($p < 0,0001$). En utilisant l'AUC, seule la mesure de l'entropie de Shannon dans trois segments de l'enveloppe du VIH-1 permettait d'identifier avec précision les infections récentes à un niveau satisfaisant. Il s'agit en effet, des segments gp120 C2_1 (AUC = 0,806), gp120 C2_3 (AUC = 0,805) et la gp120 V3 (AUC = 0,812).

Conclusions : Nos résultats indiquent ainsi clairement que la mesure de l'entropie de Shannon des trois segments de l'enveloppe virale identifiés pourrait servir de biomarqueur pour le diagnostic des infections récentes pour le VIH-1.

ABSTRACT

Background: Identifying recent HIV-1 infections is crucial for monitoring HIV-1 incidence and optimizing public health prevention efforts. To identify recent HIV-1 infections, we evaluated and compared the performance of 4 sequence-based diversity measures including percent diversity, percent complexity, Shannon entropy and number of haplotypes targeting 13 genetic segments within the *env* gene of HIV-1.

Methods: A total of 597 diagnostic samples obtained in 2013 and 2015 from recently and chronically HIV-1 infected individuals were selected. From the selected samples, 249 (134 from recent versus 115 from chronic infections) *env* coding regions, including V1-C5 of gp120 and the gp41 ectodomain of HIV-1, were successfully amplified and sequenced by next generation sequencing (NGS) using the Illumina MiSeq platform. The ability of the four sequence-based diversity measures to correctly identify recent HIV infections was evaluated using the frequency distribution curves, median and interquartile range and area under the curve (AUC) of the receiver operating characteristic (ROC).

Results: Comparing the median and interquartile range and evaluating the frequency distribution curves associated with the 4 sequence-based diversity measures, we observed that the percent diversity, number of haplotypes and Shannon entropy demonstrated significant potential to discriminate recently from chronic infections ($p < 0.0001$). Using the AUC of ROC analysis, only the Shannon entropy measure within three HIV-1 *env* segments could accurately identify recent infections at a satisfactory level. The *env* segments were gp120 C2_1 (AUC = 0.806), gp120 C2_3 (AUC = 0.805) and gp120 V3 (AUC = 0.812).

Conclusion: Our results clearly indicate that the Shannon entropy measure represents a useful tool for predicting HIV-1 infection recency.

1. INTRODUCTION

At the end of 2014, the Public Health Agency of Canada (PHAC) estimated that approximately 75,500 people were living with HIV/AIDS in Canada, of whom 21% were unaware of their status [1]. Some efforts are made nowadays to increase diagnosis and linkage to care to newly diagnosed persons. But it is important to differentiate individuals who were recently infected from those with chronic infection as it may have a different impact on the epidemic and its control. It is estimated that the probability of HIV transmission is 26 times higher during the first 3 months of infection [2, 3] due to the high viral load in newly infected individuals [2-7]. Hence, identifying recently infected individuals is not only a key measure for better estimating HIV-1 incidence within the general population [4, 6, 8-10], but also is a valuable tool for monitoring HIV-1 epidemics and optimizing prevention efforts [3-6] to reduce HIV-1 forward transmission [2, 11-13]. Several strategies have been developed to identify recent infections, each exhibiting a varying efficacy [4, 14-18]. The traditional epidemiological method consists of prospectively following-up HIV infection rates within cohorts of HIV-1 negative individuals presenting with a high risk of HIV infection [4, 18]. This approach requires complicated logistics, is expensive and leads to outcome results that are not representative of the situation prevailing in the general population [18-22]. Serology-based methods consist of evaluating biomarkers, such as the presence of antigens and specific antibodies, and their modulation in time [23-35]. A multi-assay based serological algorithm based on two commercially available avidity assays [24] was recently developed in our laboratory. It has been shown to provide good discriminatory power to identify individuals infected within 136 days mean duration of recent infection (MDRI), with an estimated false recency rate of 3.3% [24]. This algorithm was used in the present study to classify clinical specimens as recently infected individuals (MDRI < 136 days) or chronically infected individuals (>136 days) [24]. Finally, a variety of molecular-based assays monitoring the HIV-1 viral genetic diversity throughout disease progression have been described, including: 1) The High-Resolution Melting Assay (HRM), which evaluates the melting temperatures of HIV amplicons to estimate the number of HIV-1 quasi-species present in a

given individual specimen [36-38]; 2) the number of ambiguous nucleotides (mixed bases) [17, 39], for which DNA sequences are usually provided by first generation sequencing; 3) the Hamming Distance (HD), which measures points of variation between two sequences of equal length [40, 41] using first generation sequencing; and 4) sequence-based diversity measurements as assessed by next generation sequencing (NGS) [42, 43], which is able to detect minor variants/mutations at low rates[44]. NGS is a powerful tool for evaluating HIV-1 sequence-based diversity [45] and was previously shown to be more accurate at detecting recent infections than any other molecular-based method [46]. In this study, HIV-1 *env* gene sequences, rather than those from *pol* or *gag*, were analyzed, as they are known to evolve more rapidly than other HIV-1 gene sequences [47-49]. The *env* diversity has already been shown to correlate with the HIV-1 Fiebig stages [50] but not in comparison with *pol* and *gag*. The HIV-1 envelope is a complex trimeric glycoprotein located on the viral surface and composed of the gp120 and gp41 subunits [47-49]. The gp120 subunit is subdivided into five conserved sub-domains (C1-C5) and five hyper-variable glycosylated loops (V1-V5) [51-54]. The gp41 subunit consists of an ectodomain (ECD), transmembrane domain (TM), and long cytoplasmic domain (CP) [55]. Each HIV-1 *env* subdomain or region plays a specific role in pathogenesis [52, 54, and 56] and is differentially impacted by selective pressure. The present study evaluated the capacity to predict HIV-infection recency using four sequence-based diversity measures including the percent diversity, percent complexity, Shannon entropy and number of haplotypes, screening 13 HIV-1 *env* segments throughout the gp120 V1-C5 and the gp41 ectodomains.

2. MATERIALS AND METHODS

Patients and specimens

HIV positive samples. In the province of Quebec (Canada), all serum samples that are repeatedly reactive using a screening HIV-1,2-enzyme immunoassay (EIA) are submitted to the provincial reference microbiology laboratory (*Laboratoire de Santé publique du Québec, LSPQ*) for confirmation mainly via a HIV-1 Western blot (WB) and/or HIV-1 p24 EIA. Western blot positive samples are submitted to a multi-assay algorithm (MAA) that combines a Centers for Disease Control and Prevention (CDC) modified Bio-Rad-Avidity assay followed by the Sedia-LAg-Avidity assay [24]. This MAA previously demonstrated good performance for identifying recent HIV-1 infections, showing a false recent rate (FRR) of 3.3% for a mean duration of recent infection (MDRI) of 136 days [24]. After the WB and EIA assays, residual sample volumes are stored at -20°C. For this study, recent infection samples were defined as follows: WB Negative or indeterminate, positive for HIV-1 p24 (within 22 days of infection) or positive according to WB but determined by MAA to be recent (within 136 days of infection). Established infection (chronic) samples were those that were positive according to WB and determined by MAA to be long-standing (> 136 days of infection). A total of 164 recent (including 26 p24 antigen positives) and 154 chronic infection samples collected in 2013 as well as 117 recent (including 28 p24 antigen positives) and 162 chronic samples collected in 2015 were evaluated.

Amplification and Sequencing

Total nucleic acids were extracted from 100 µl of serum using an automated BioRobot MDx extraction platform using the QIAamp Virus BioRobot MDx Kit (QIAGEN, Valencia, CA, USA). HIV-1 RNA was amplified using the Superscript III One-Step RT-PCR system with Platinum 1 Taq DNA polymerase (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) and, the primers *env* up forward (5'-GTTTCTTTTAGGCATCTCCTATGGCAGGAAGAAG-3', HXB2 positions 5957–5983) and *env*-lo reverse (5'-GTTTCTTCCAGTCCCCCTTTTCTTTTAAAAG-3', HXB2 positions 9063–9088) [57]. The amplification conditions were as follows: 53°C for 30

minutes (for reverse transcription) and 94°C for 2 minutes for Taq DNA polymerase activation, followed by 40 cycles at 94°C for 15 s, 55°C for 30 s, and 68°C for 4 min. Nested amplification was performed using the Expand™ High Fidelity PCR System kit (Roche Diagnostics, Indianapolis, USA) as described by the manufacturer. The primers E60F forward (5'-TAATCAGTTTATGGGATCAAAGC-3', HXB2 nucleotides positions 6547–6569) [58] and E55R reverse (5'-GCCCCAGACTGTGAGTTGCAACAGATG-3', HXB2 nucleotides positions 7940–7914) [59] were used. The amplification conditions were: 94°C for 2 min, followed by 45 cycles at 94 for 15 s, 55°C for 30 s, and 68°C for 2 min. PCR products were visualized by agarose gel electrophoresis and purified using the QIAquick 96 PCR Purification Kit from QIAGEN (QIAGEN, Valencia, CA, USA). The nested RT-PCR generated ≈1400 bp of the *env* gene. For next generation sequencing (NGS), one nanogram (1 ng) of DNA quantified using the Quant-iT™ PicoGreen1 dsDNA Assay kit (Life technologies, Oregon, USA) was used for library preparation using the Nextera XT DNA library preparation kit from Illumina (Illumina, San Diego, CA) following the manufacturer's protocol. DNA sequencing was performed on a MiSeq instrument (Illumina, San Diego, CA, USA) using MiSeq1 Reagent Kits v3 (600 cycles) following the manufacturers protocols.

Sequence data processing and genetic diversity calculation

The quality of the NGS runs was evaluated using the Illumina Sequencing Analysis Viewer V1.10.2 Software and the FastQC application (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sequencing depth and coverage were available under Coverage.txt and ComputeGP120Coverage. Sh in OneDrive HIV_A_kafando project following these links respectively: <https://onedrive.live.com/?authkey=%21AB4CmrTlu182Xw8&cid=709AAE8E69A7368F&id=709AAE8E69A7368F%21361&parId=709AAE8E69A7368F%21351&o=OneUp>. A species with a coverage less than 100x were excluded in final statistical analyses. Sequences were de novo assembled using Iterative Virus Assembler (IVA) [60] to generate a consensus. The HIV-1 *env* subdomains gp120-V1 to C5 and a part of the gp41 ectodomain (first 158pb) were analyzed separately. The gp120-C2 and C3 subdomains were subdivided into 3 and 2 segments for

subsequent analyses to compare DNA sequences of sizes like the other regions as shown in Fig 1.

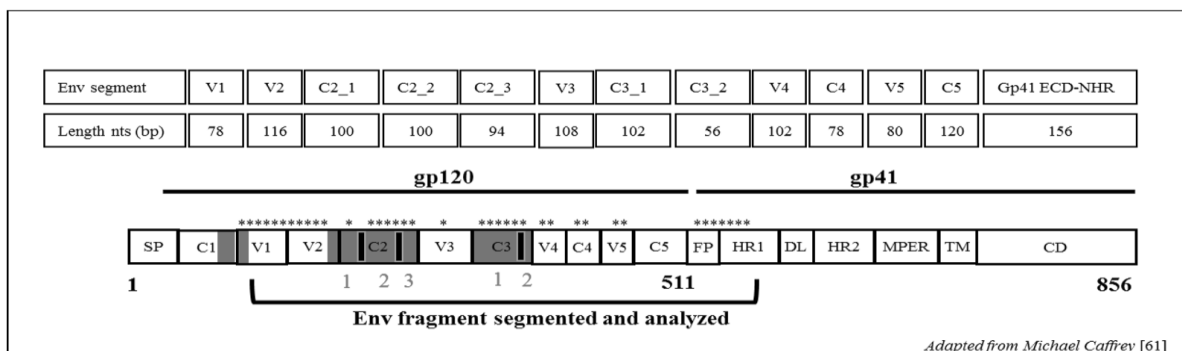


Figure XXV. (Article 1, Figure 1). Schematic figures showing all *env* segments used for diversity estimates.

Segments length corresponds to that of strain HXB2 of HIV-1 nucleotides positions. Segments used are denoted by asterisks. *env* domain abbreviations: SP, signal peptide; C1–C5, conserved domains 1 to 5; V1–V5, variable domains 1 to 5; FP, fusion peptide; HR1, heptad repeat 1 (NHR); DL, disulfide loop; HR2, heptad repeat 2 (CHR); MPER, membrane proximal ectodomain region; TM, transmembrane domain; CD, cytoplasmic domain. Image were friendly adapted from Michael Caffrey [61]; Trends in Microbiology, Volume 19, Issue 4, Pages 191–197 (April 2011) 10.1016/j.tim.2011.02.001. <https://doi.org/10.1371/journal.pone.0189999.g001>.

To map subdomains, consensus sequences were aligned with the HXB2 *env* reference sequence (*GenBank accession number K03455.1-HIVHXB2CG env nucleotides positions 6225-8795*) using Clustal W in MEGA7.0 (www.megasoftware.net) [62]. The *env* subdomain length delimitations followed the HXB2 complete genome numbering were as follows: gp120 V1 (6615-6692≈78 pb), V2(6696-6812≈116 pb), C2_segment 1 (6813-6913≈100 pb), C2_segment 2 (6914-7014≈100 pb), C2_segment 3 (7015-7109≈94 pb), V3 (7110-7217≈108 pb), C3_segment 1 (7218-7320≈102 pb), C3_segment 2 (7321-7376≈56 pb),V4 (7377-7478≈102 pb), C4 (7479-7556≈78 pb), V5 (7557-7637≈80 pb), C5 (7638-7757≈120 pb) and gp41-ectodomain (7758-7915≈158pb). Intra-patient genetic diversity was evaluated for each subdomain/segment using an in-house coded Python pipeline. SMALT (<http://www.sanger.ac.uk/science/tools/smalt-0>) was used to map the reads against their respective consensus sequence, and SAM tools (Sequence Alignment/Map) [63] were used for analysis of the mapping file generated by SMALT. Bioconductor packages (<https://bioconductor.org/>) [64] were used for the genetic diversity calculation. More details

about the specific packages and the python codes used for diversity estimates are available and DOIs to access are below: <https://github.com/EricFournier3/> HIV variants and <https://1drv.ms/f/s!Ao82p2mOrppwgl8eApq05btfNI8P>.

The four sequence-based diversity measures were calculated as previously described [36, 42] as shows in S1 Table (Table IV from thesis, Example (Annexe 1, french). Briefly, the percent diversity was evaluated as the average pairwise genetic distance between two sequences [42], the percent complexity was expressed as the number of distinct variants divided by the total number of reads x 100 [42], and the Shannon entropy index (S) was calculated using a formula that accounts for both the number of distinct reads and their proportional representation in the dataset [42, 43]. The number of haplotypes strictly included the number of distinct quasi-species or variants present in at least 1% or more in the viral population [42]. The frequency distribution curves (ggplot2, graphs) of the percent diversity, percent complexity, Shannon entropy and number of haplotypes for recent compared to chronic sequences (statistical analysis) were generated using the statistical software R [65].

Tableau IV. (Article 1, S1 Table). Method of calculation of sequence-based diversity measures.

Description of sequence-based diversity measures	
Diversity measure	Method of calculation
% diversity	MEGA 7.0 (www.megasoftware.net) was used to analyze gap-stripped sequence alignments. This value is reported as a percentage.
% complexity	$[(\text{no. of distinct variants})/(\text{total no. of reads})] \times 100$
Shannon entropy (S)	$S = - \left(\frac{1}{\log N} \right) \sum_{i=1}^n p_i \log p_i$
	where p_i is the proportion of reads consisting of each unique sequence pattern, N is the total no. of sequences, and n is the no. of distinct sequences. <i>Consider the proportion and the number of distinct reads in the same sequence.</i>
Number of haplotype	Consider the number of distinct variant in the same specimen.

Table reproduced with authorization from. Cousins MM *et al.* J Clin Microbiol. 2012 [42].

Determination of HIV Subtypes

We used two HIV subtyping tools to determine a consensus HIV subtype. The Rega HIV

Subtyping Tool V3 [66] (<http://regatools.med.kuleuven.be/typing/v3/hiv/typingtool>) and, Confirmation with the NCBI HIV Subtyping tool [67] (<https://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi>).

Determination of the sequence-based diversity measure performance

The performance of individuals or combined sequence-based diversity measures for discriminating recently from chronic infections was evaluated using the area under the curve (AUC) of the receiver operating characteristics (ROC) [68]. The best value for the AUC is 1.0, which represents 100% sensitivity and 100% specificity at distinguishing recently from chronic infections. Interpretations of the AUC values for the sequence-based diversity of recent versus chronic HIV-1 infected individuals in our study followed the grading guidelines that were previously described by *D.G Kleinbaum and M. Klein (2012)* [68]. Briefly, scores from 0.90-1.0 were excellent discrimination (scored A), scores from 0.80-0.90 good discrimination (scored B), scores from 0.70±0.80 fair discrimination (scored C), scores from 0.60±0.70 poor discrimination (scored D) and scores from 0.50±0.60 failed discrimination (scored F). AUC of ROC analysis was also used to identify the optimal cut-off values that would distinguish recently from chronic infections with high accuracy (high sensitivity and high specificity) as previously described [69]. The following online link will help understanding calculation methods and identification of the optimal or best cut-off values: [https://ncss-wpengine.netdna-ssl.com/wpcontent/themes/ncss/pdf/Procedures/NCSS/One ROC Curve and Cutoff Analysis.pdf](https://ncss-wpengine.netdna-ssl.com/wpcontent/themes/ncss/pdf/Procedures/NCSS/One_ROC_Curve_and_Cutoff_Analysis.pdf).

The best cut-off values of measures must have the highest accuracy, which corresponds to its capacity to correctly classify the highest true positive rate (TPR) or sensitivity and highest true negative rate (TNR) or specificity. Here, the TPR represents the recent HIV-1 infected individuals who were correctly classified, and the TNR represents the chronic HIV-1 infected individuals who were also correctly classified by the same test. Statistical analyzes summary statistics (mean, median and interquartile range) were used to estimate the intra and inter-patient envelope genetic diversity. The student t-test was used to compare the diversity measures between sequences from recent and chronic infections. Analyses were performed

using Epi Info™ 7 (www.cdc.gov/epiinfo) and IBM SPSS Statistics software. P-values below 0.05 were considered statistically significant.

Ethics Statement

A retrospective patient's samples were used in this study and were obtained from the LSPQ serobank collection. They were collected for routine diagnostic purposes in 2013 and 2015. All samples were anonymized before we accessed them for the study. No nominal information of patients were used for analysis and data management. Written informed consent was obtained from individuals in the primary HIV infection (PHI) cohort of Quebec included in study. Ethical clearance was obtained from the « *Le Comité d'éthique de la recherche (CER) du Centre hospitalier de l'Université de Montréal* » (CHUM), Montréal, Canada.

Nucleotide sequence accession number

The Miseq d'Illumina sequencing data obtained in this study (n=249) were deposited and available in the GenBank Sequence Database (NCBI) under GenBank accession KY946451 to KY946713 as reported in S1 Dataset.

3. RESULTS

A total of 597 specimens from individuals newly diagnosed HIV-1+ and sampled in years 2013 and 2015 were collected in this study. All, except for p24 positive samples, were subjected to an avidity Multi-assay algorithm (MAA) to assess infection recency by serological tests. From those, 281 were categorized as recent infections and 316 as chronic. These specimens were not successfully passed the PCR amplification and sequencing process. The success rate of the nested RT-PCR step was 46% (276/597), and 97% of the latter were successfully sequenced (n = 263) as presented in Fig 2. At the sequence data management and processing, the very short ones or containing gaps estimates to 5% (14/263) after alignment with HXB2 *env* reference sequence were excluded. Finally, n = 249 sequences that corresponds to one per patient were included in this current study (Fig 2 and S1 Dataset).

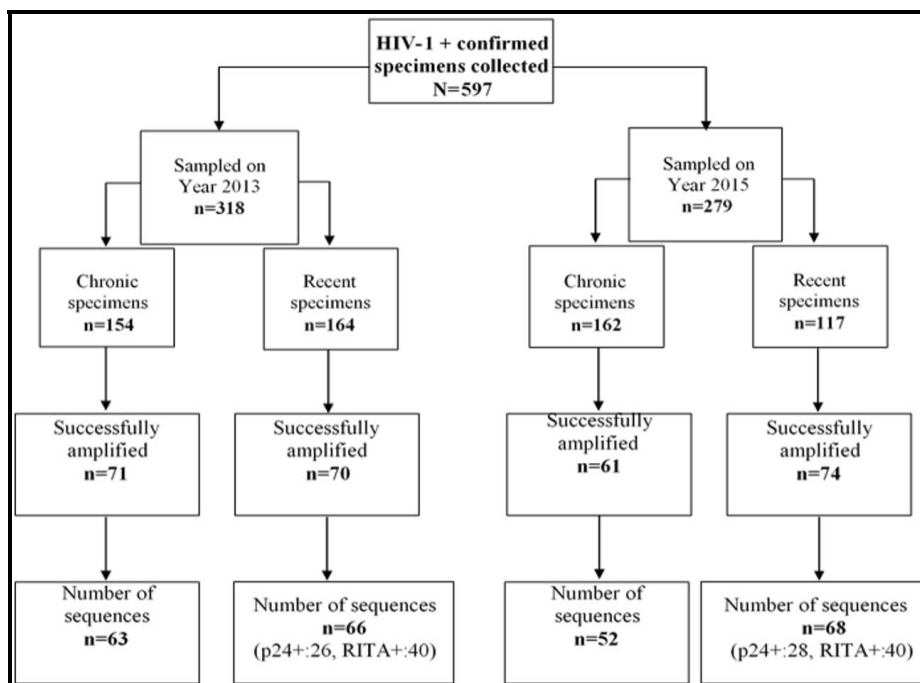


Figure XXVI. (Article 1, Figure 2). Number of sequences (one per patient) used in this study.

N=249 derived from 134 recently versus 115 chronically HIV-1 infected individual's sequence data were included in the study.

For HIV-1 subtyping, seventy-seven per cent (77%) of the sequenced samples (n = 192/249) were clade B subtypes and 30% for non-B as presented in Table 1 and reported in S1 Dataset.

Tableau V. (Article 1, Table 1): HIV-1 subtype distribution of the sequences analyzed in this study.

HIV-1 Subtype	Number of sequences for recently HIV-1 infected individuals		Number of sequences for chronically HIV-1 infected individuals		Total of number of sequences	
	No	%	No	%	No	%
A1	11	8,21%	12	10,43%	23	9,24%
B	106	79,10%	86	74,78%	192	77,11%
C	3	2,24%	6	5,22%	9	3,61%
CRF 01_AE	7	5,22%	0	0,00%	7	2,81%
CRF 02_AG	1	0,75%	3	2,61%	4	1,61%
CRF11	0	0,00%	1	0,87%	1	0,40%
D	5	3,73%	2	1,74%	7	2,81%
F1	1	0,75%	3	2,61%	4	1,61%
G	0	0,00%	2	1,74%	2	0,80%
TOTAL	134	100%	115	100%	249	100%

<https://doi.org/10.1371/journal.pone.0189999.t001>

The HIV-1 *env* diversity in specimens from recent and chronic infections was examined using four distinct sequence-based diversity measures including percent diversity, percent complexity, Shannon entropy and number of haplotypes. To study the profile of diversity values that are associated with recent versus chronic specimens, frequency distribution curves (ggplot2 of R) were generated for each of the 4 sequence-based diversity measures and across all the HIV-1 *env* segments selected (Figs 3-6).

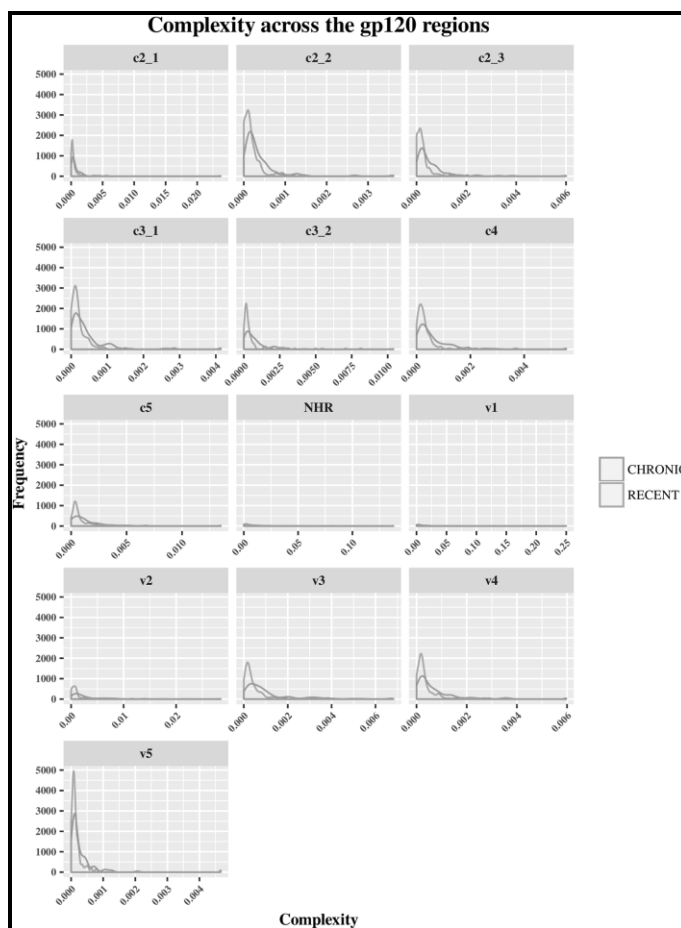


Figure XXVII. (Article 1, Figure 3). Frequency polygons (ggplot2) of percent complexity of *env* sequences of recent HIV-1 infected individuals compared to chronically infected ones by *env* segments.

The Y axis represents the density of observations (frequency) and the X axis the percent complexity distribution as sequence-based diversity measure. The blue color represents plot and distribution for recent HIV-1 infected population and the red color plot and distribution for chronic infected ones.

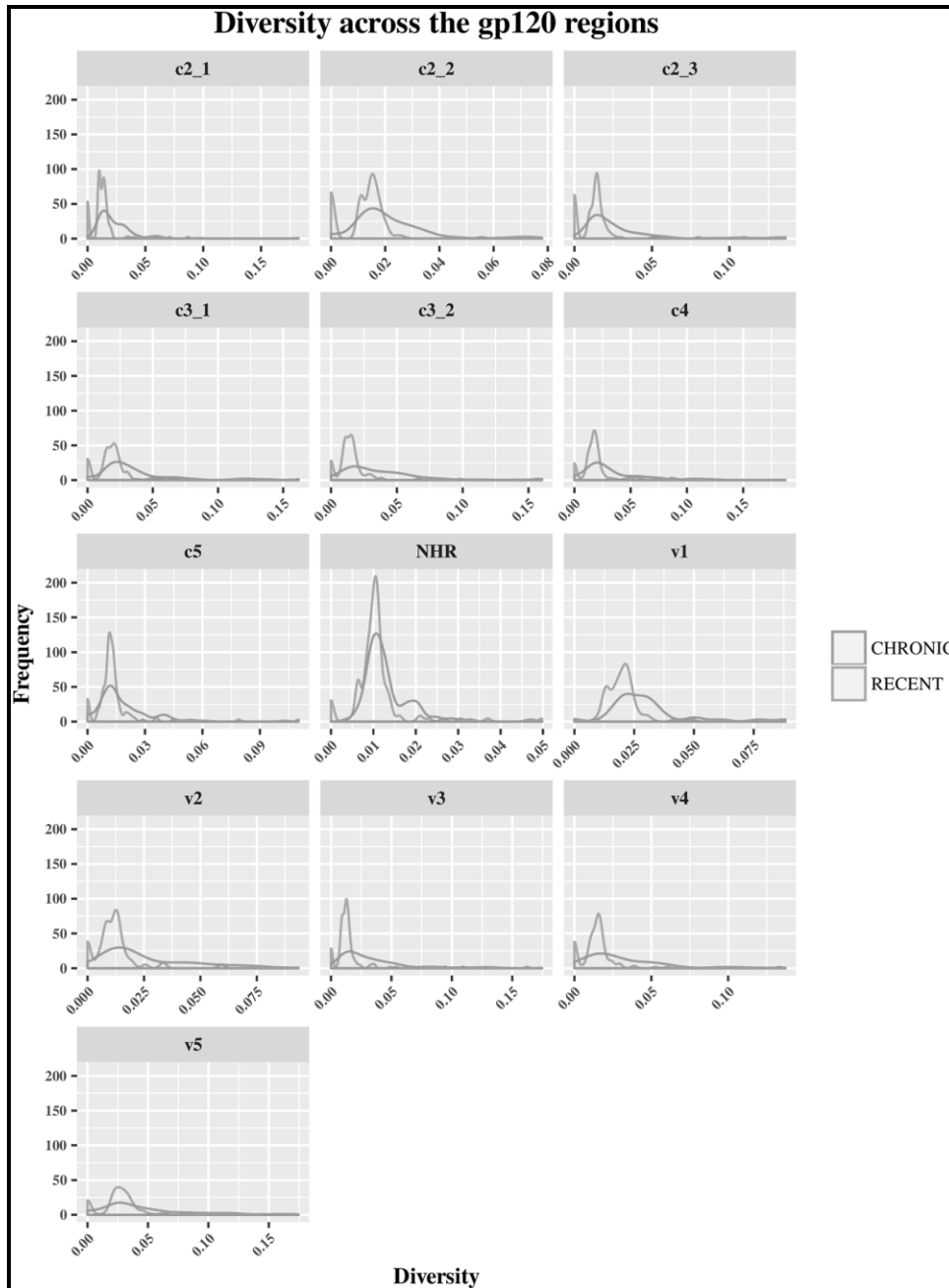


Figure XXVIII. (Article 1, Figure 4). Frequency polygons (ggplot2) of percent diversity of *env* sequences of recent HIV-1 infected individuals compared to chronically infected ones by *env* segments.

The Y axis represents the density of observations (frequency) and the X axis the percent diversity distribution as sequence-based diversity measure. The blue color represents plot and distribution for recent HIV-1 infected population and the red color plot and distribution for chronic infected ones.

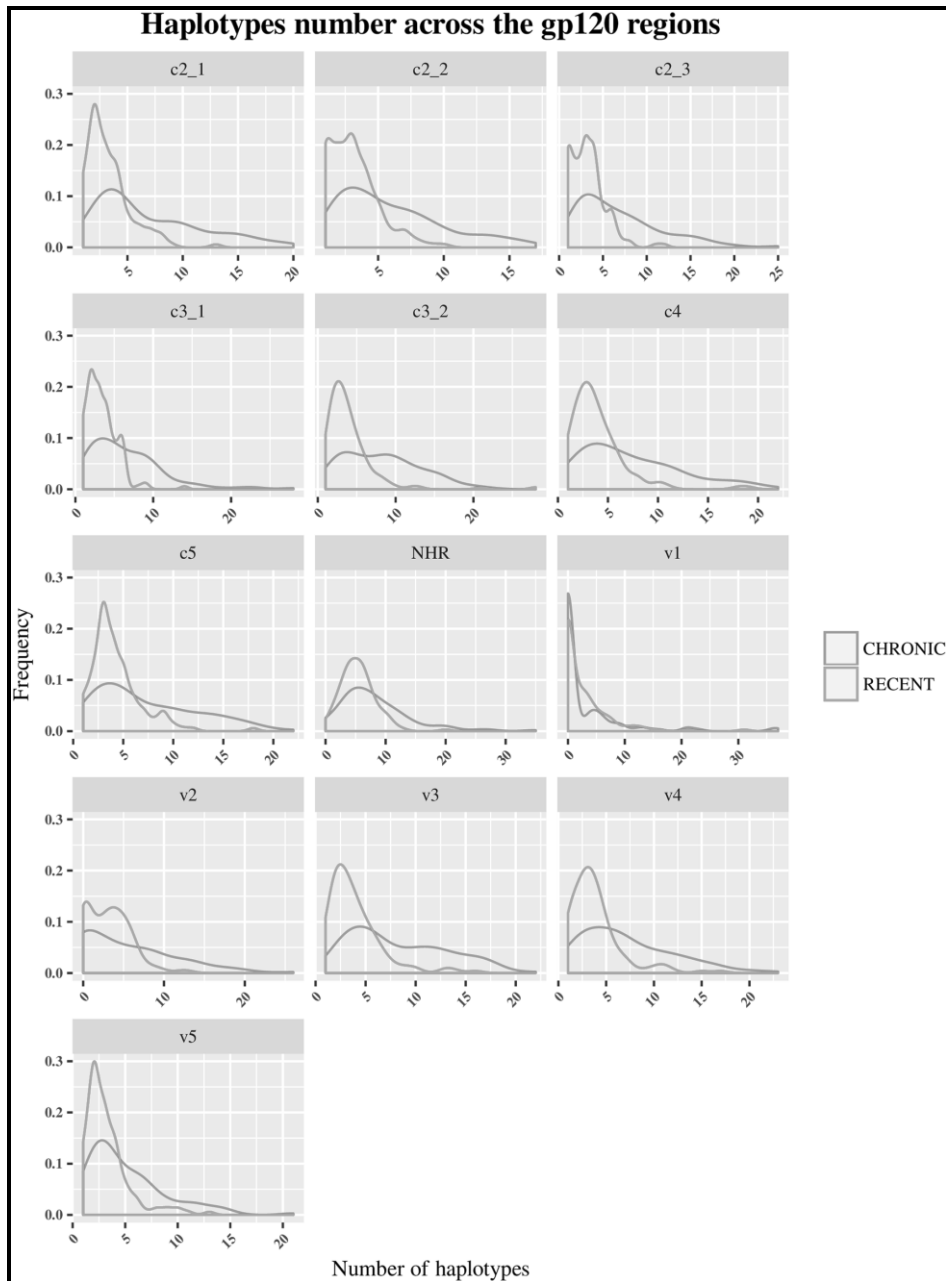


Figure XXIX. (Article 1, Figure 5). Frequency polygons (ggplot2) of a number of haplotypes of *env* sequences of recent HIV-1 infected individuals compared to chronically infected ones by *env* segments.

The Y axis represents the density of observations (frequency) and the X axis the number of haplotype distribution as sequence-based diversity measure. The blue color represents plot and distribution for recent HIV-1 infected population and the red color plot and distribution for chronic infected ones.

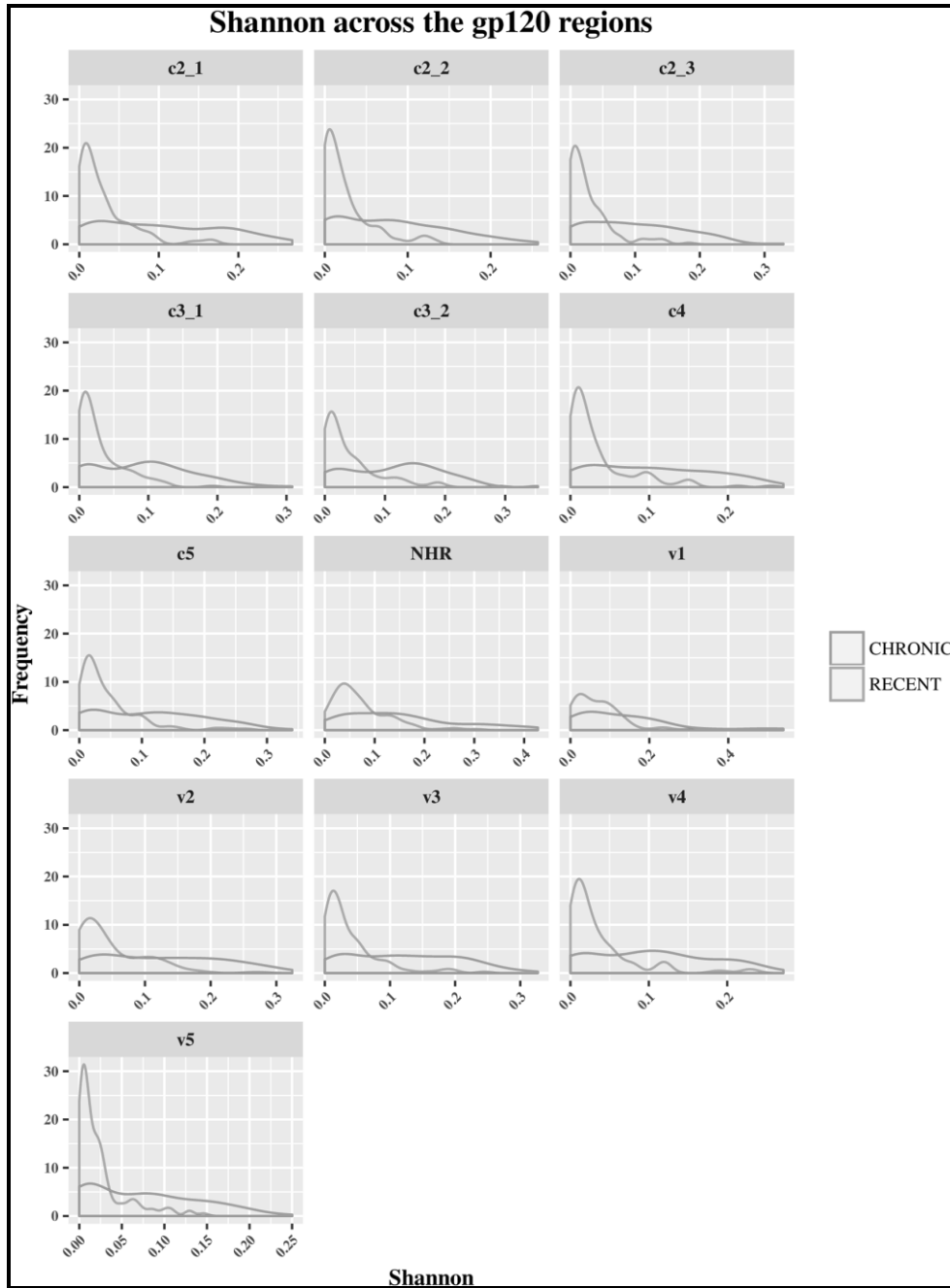


Figure XXX. (Article 1, Figure 6). Frequency polygons (ggplot2) of Shannon entropy index of *env* sequences of recent HIV-1 infected individuals compared to chronically infected ones by *env* segments.

The Y axis represents the density of observations (frequency) and the X axis the Shannon entropy index distribution as sequence-based diversity measure. The blue color represents plot and distribution for recent HIV-1 infected population and the red color plot and distribution for chronic infected ones.

These frequencies of diversity distribution curves for each measure comparing recent to chronic HIV-1 infected individuals by *env* segments are also available online following this link: <https://1drv.ms/f/s!Ao82p2mOrppwgl8eApq05btfNI8P>. As shown in Fig 3, the percent complexity frequency distribution curves from recent versus chronic infections were overlapped at the same low complexity level. These results showed that this diversity measure did not allow for a clear discrimination between the two HIV-1 infected populations. The medians and means analyses of diversity seemed to confirm these observations for all the HIV-1 *env* segments analyzed (S2, S3 and S4 Tables). But, a statistical analysis (student t-test) confirms this view for only *env* gp120 C3_2, V1 and V5 segments ($P>0.05$), the others *env* segments were statistically significant ($P<0.05$) as summarized in S4 Table. Conversely, the percent diversity (Fig 4), number of haplotypes (Fig 5) and Shannon entropy (Fig 6) distribution curves of recently infected individual's sequences peaked at lower diversity values compared to the curves associated with chronic infections, which were more widely distributed and shifted towards higher diversity values for all the *env* segments analyzed. These observations were indicative of the good discriminatory power of these 3 diversity measures for all the *env* segments tested. The summary statistics (mean, medians and IQR observations seemed to confirm the differences between recent and chronic HIV-1 infected populations as shown in S2, S3 and S4 Tables. Also, the statistical analysis using student t test confirms and demonstrates a significant difference between recent and chronic sequence diversity ($P<0.05$) for these 3 measures for any *env* segments analyzed as shown in S4 Table. The area under the curve (AUC) of receiver operating characteristics (ROC) analysis was used to compare the performance of each sequence-based diversity measure in their ability to identify HIV-1 infection recency based on analysis of the 13 segments of HIV-1 *env* (Figs 7 and 8). Using the percent complexity measure, we observed that both the gp120-V2 and V3 loop segments showed fair discrimination (AUC= 0.7) as opposed to the other eleven *env* segments tested, which presented poor discrimination (AUC \leq 0.6). Using the percent diversity measure, fair discrimination was observed for 11 *env* segments (AUC= 0.7), while the 3 *env* segments, gp120- C4, C5 and V5, exhibited poor discrimination (AUC=0.6). The number of haplotype measures exhibited fair discrimination for 10 *env* segments (AUC= 0.7) and poor

discrimination for 3 *env* segments, gp120 C5, V1 and V5 (AUC=0.6). However, the Shannon entropy showed good discrimination power for 3 *env* segments, gp120 C2_1, C2_3 and V3 (AUC=0.8); fair discriminatory power (AUC = 0.7) for 9 *env* segments, gp120 C2_1, C3_1, C3_2, C4, C5, V2, V4, V5 and gp41-ectodomain; and poor discriminatory power (AUC = 0.6) for the *env* gp120 V1 segment. The Shannon entropy was the only single sequence-based diversity measure for which a significantly good discriminatory power was observed. It is therefore identified as the best performing diversity measure for the HIV-1 *env* segments analyzed. More specifically, the gp120 C2_1, C2_3 and V3 HIV-1 *env* segments appeared to be the most predictive for identifying HIV-1 recency (Table 2 and Figs 7-11).

Tableau VI. (Article1, Table 2). Performance of Shannon entropy as a measure for identifying recent HIV-1 infections

HIV-1 <i>env</i> segment	Diversity measure	Optimal Cut-off value	HIV-1 Subtype	AUC	AUC (95% CI)	TPR (Sn)	TNR (Sp)	N recent/ chronic
GP120-C2_1	Shannon entropy	0.054	Multiple	0.806	[0.752–0.861]	82%	69%	134/115
GP120-C2_3	Shannon entropy	0.059	Multiple	0.805	[0.749–0.862]	90%	67%	134/115
GP120-V3	Shannon entropy	0.083	Multiple	0.812	[0.758–0.866]	87%	66%	134/115
GP120-C2_1	Shannon entropy	0.054	B	0.791	[0.726–0.856]	80%	69%	106/86
GP120-C2_3	Shannon entropy	0.06	B	0.810	[0.747–0.873]	91%	66%	106/86
GP120-V3	Shannon entropy	0.083	B	0.831	[0.773–0.889]	90%	66%	106/86
GP120-V3	Shannon entropy	0.083	Multiple:WB+/MAA+specimens as recent	0.801	[0.740–0.863]	89%	66%	80/115
GP120-V3	Shannon entropy	0.097	Multiple: p24+ specimens as recent	0.827	[0.763–0.891]	93%	60%	54/115
GP120-C2_1	Shannon entropy	0.06	Multiple: p24+ specimens as recent	0.850	[0.794–0.905]	93%	66%	54/115
GP12-C2_3	Shannon entropy	0.048	Multiple: p24+ specimens as recent	0.844	[0.786–0.902]	91%	71%	54/115

TPR (true positive rate) = sensitivity: recent HIV-1+ specimens correctly classified,
TNR (true negative rate) = specificity: chronic HIV-1+ specimens correctly classified. The optimal cutoff value indicates the proportion of patients correctly classified and represented the highest TPR (recency) + TNR (chronic) or (Sensitivity+ Specificity).

<https://doi.org/10.1371/journal.pone.0189999.t002>

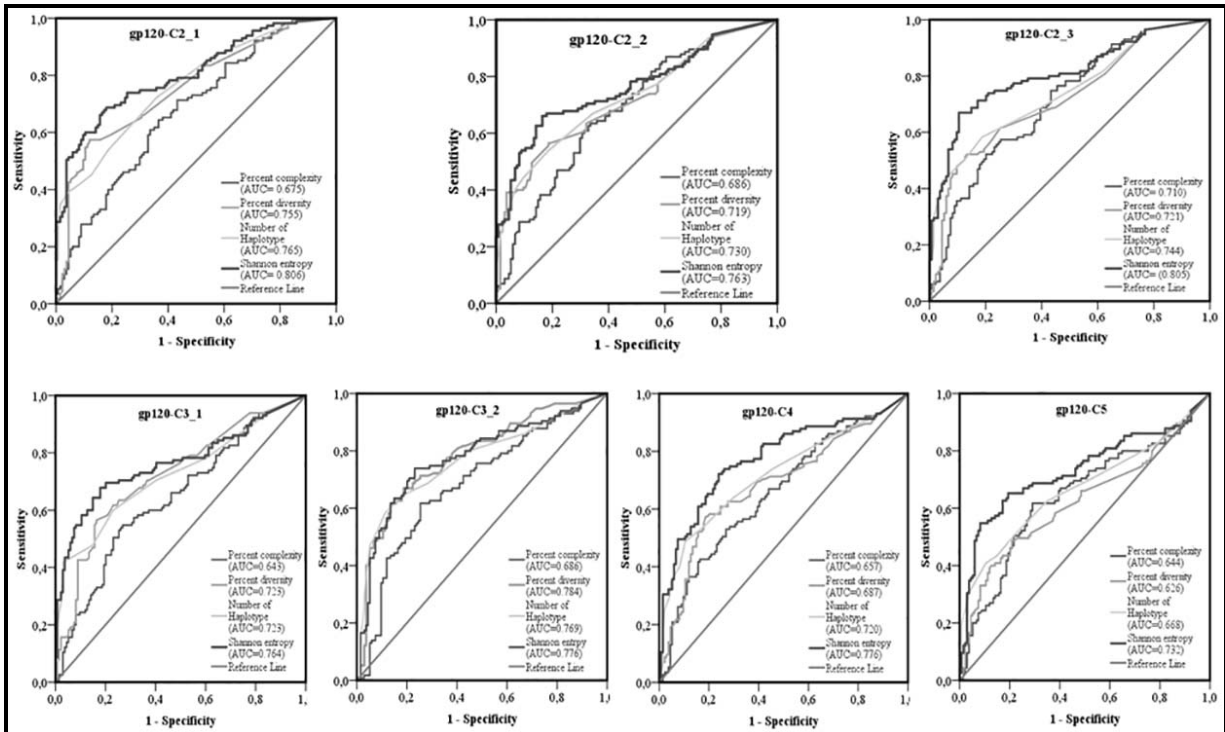


Figure XXXI. (Article 1, Figure 7). ROC curves comparing the performance of the 4 sequence-based diversity measures for discriminating recently from chronic HIV-1 infection.

Four selected HIV-1 gp120 conserved subdomains (C2, C3, C4 and C5) subdivided on seven segments were analyzed, 3 segments on the gp120-C2 region (C2_1; C2_2 and C2_3), 2 segments on the gp120-C3 region (C3_1 and C3_2), 1 segment on gp120-C4 and 1 segment on gp120-C5. The Y axis represents the proportion of sequences from true recent HIV-1 infected individuals (sensitivity), and the X axis the proportion of recent HIV-1 infected individuals who were incorrectly classified (1-specificity). ROC = receiver operating characteristics. AUC (area under the curve) values between 0.8 and 1 were considered performance measures.

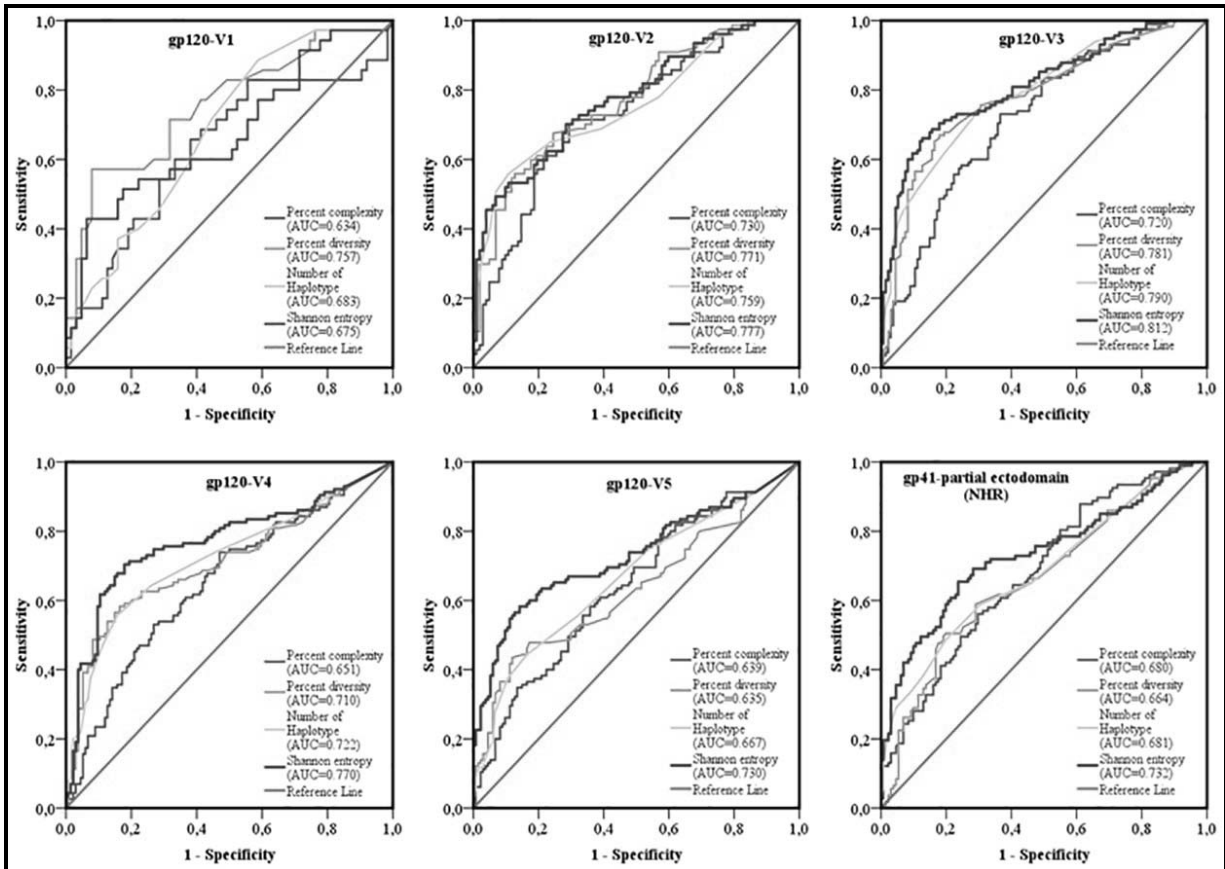


Figure XXXII. (Article 1, Figure 8). ROC curves comparing the performance of the 4 sequence-based diversity measures for discriminating recently from chronic HIV-1 infection.

Five HIV-1 gp120 variable loops and one part of gp41 ectodomain (NHR). Five segments represented each of the HIV-1 gp120 variable loop as well as 1 segment of the gp41- NHR ectodomain were analyzed: gp120-V1 loop, gp120-V2 loop, gp120-V3 loop, gp120-V4 loop, gp120-V5 loop and part of the gp41-NHR ectodomain. The Y axis represents the proportion of sequences from true recent HIV-1 infected individuals (sensitivity), and the X axis represents the proportion of recent HIV- 1 infected individuals who were incorrectly classified (1-specificity). ROC = receiver operating characteristics. NHR = N-terminal heptad repeat. AUC values between 0.8 and 1 were considered performance measures.

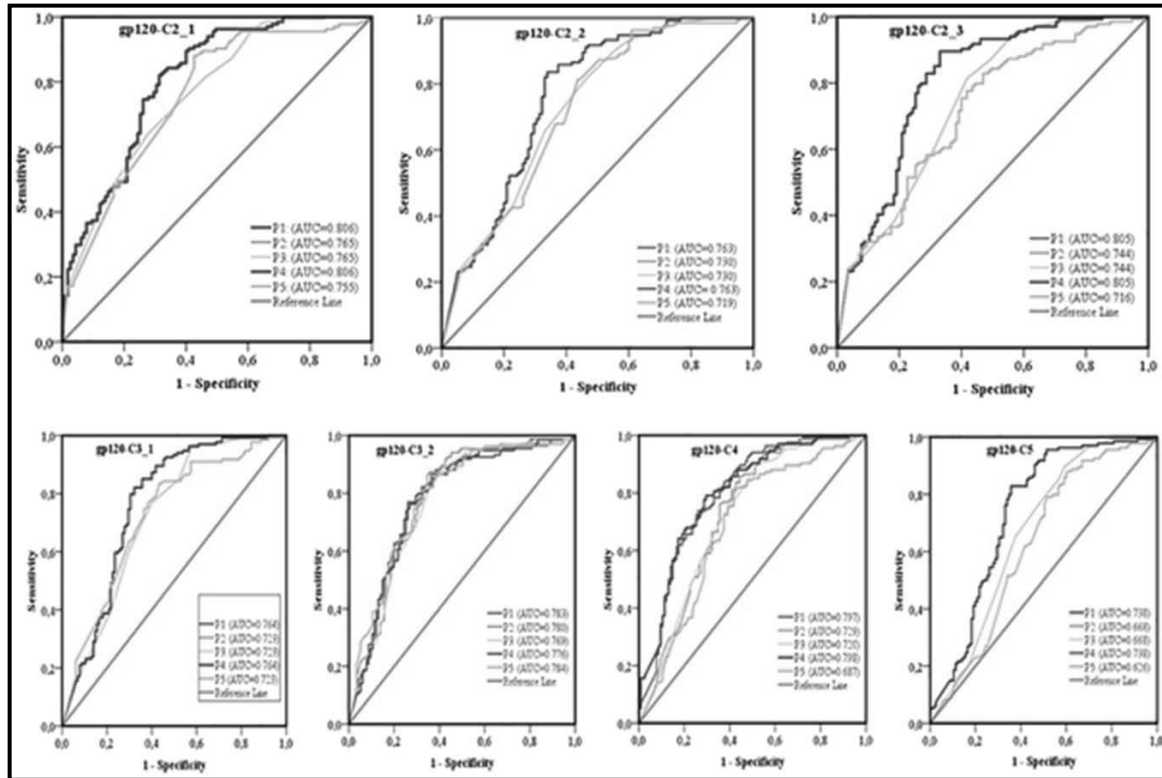


Figure XXXIII. (Article 1, Figure 9). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of HIV-1 gp120 conserved subdomains to identify HIV-1 infection recency.

Five combinations of sequence-based diversity measures were analyzed. Shannon entropy + percent diversity + percent complexity: P1; percent diversity+ number of haplotypes+ percent complexity: P2; number of haplotypes+ percent complexity: P3; Shannon entropy+ percent complexity: P4 and percent diversity+ percent complexity: P5. Seven HIV-1 *env* segments were considered: gp120-C2_1; gp120-C2_2; gp120-C2_3; gp120-C3_1; gp120-C3_2; gp120-C4 and gp120-C5. ROC = receiver operating characteristics; AUC = area under the curve. AUC values between 0.8 and 1 were considered performance measures.

<https://doi.org/10.1371/journal.pone.0189999.g009>

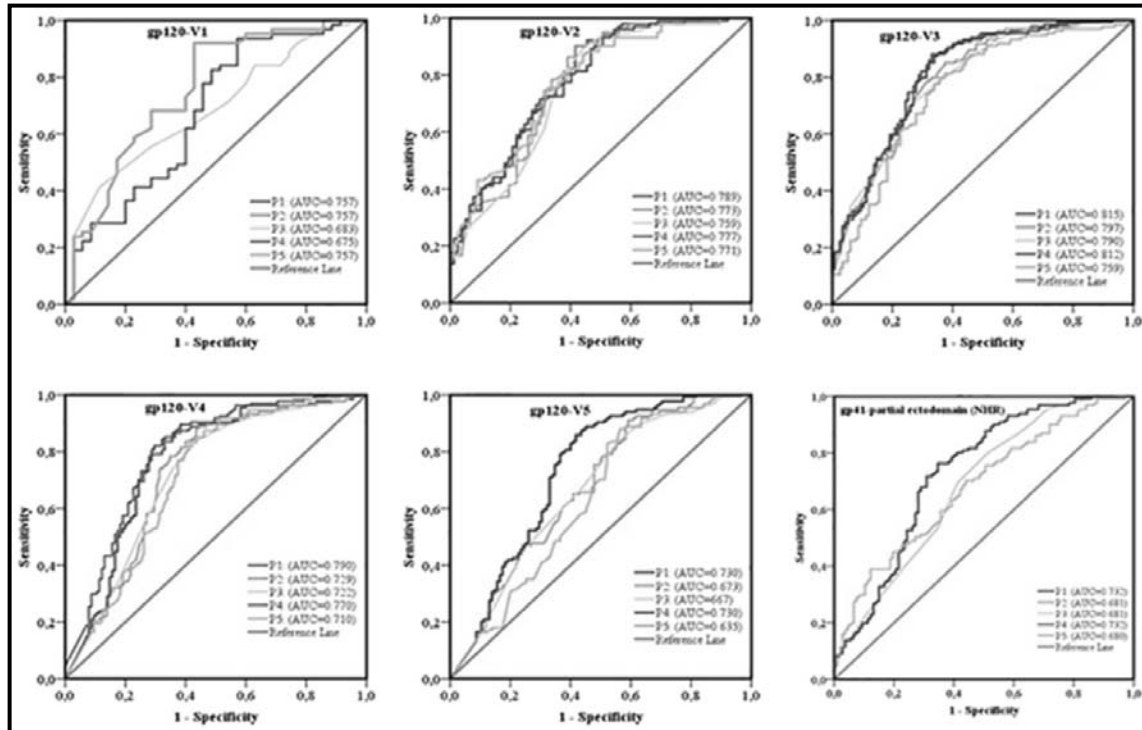


Figure XXXIV. (Article 1, Figure 10). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of five HIV-1 *env* gp120 variable loops and one part of the gp41-ectodomain (NHR) to identify HIV infection recency.

Five combinations of sequence-based diversity measures were analyzed. Shannon entropy + percent diversity + percent complexity: P1; percent diversity+ number of haplotypes+ percent complexity: P2; number of haplotypes+ percent complexity: P3; Shannon entropy+ percent complexity: P4 and percent diversity+ percent complexity: P5. Six HIV-1 *env* segments were considered: gp120-V1 loop; gp120-V2 loop; gp120-V3 loop; gp120-V4 loop; gp120-V5 loop and, gp41-NHR (partial ectodomain). NHR = N-terminal heptad repeat. ROC = receiver operating characteristics; AUC = area under the curve. AUC values between 0.8 and 1 were considered performance measures.

<https://doi.org/10.1371/journal.pone.0189999.g010>

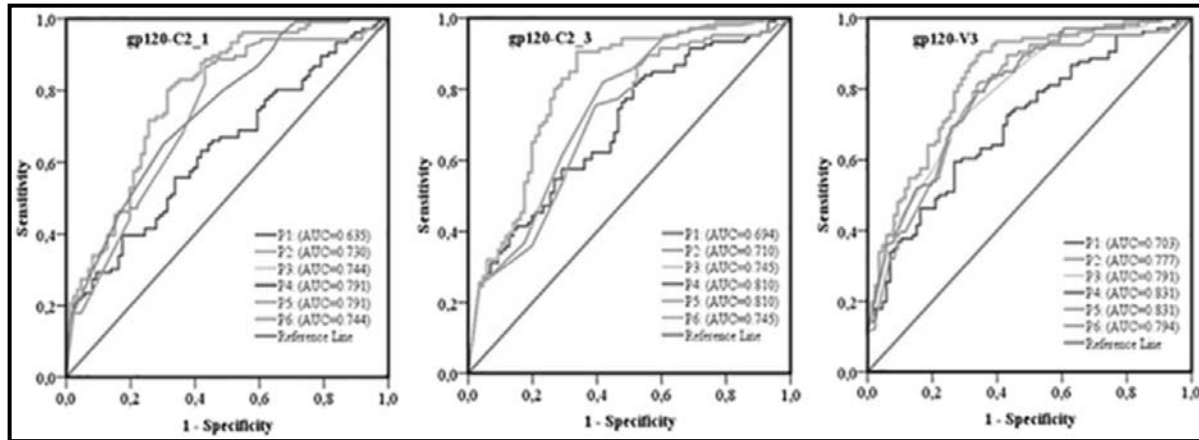


Figure XXXV. (Article 1, Figure 11). ROC curves comparing the predictive performance of different combinations of sequence-based diversity measures of HIV gp120-C2-1, gp120-C2_3 and gp120-V3 segments for identifying HIV-1 subtype B infection recency.

Five combinations of sequence-based diversity measures were analyzed: P1, percent complexity; P2, percent diversity; P3, number of haplotypes; P4, Shannon entropy; P5, Shannon entropy+ percent diversity and P6, Number of haplotypes+ percent diversity. Three HIV-1 *env* segments were considered: gp120-C2_1, gp120-C2_3 and gp120-V3. ROC = receiver operating characteristics; AUC = area under the curve. AUC values between 0.8 and 1 were considered performance measures. <https://doi.org/10.1371/journal.pone.0189999.g011>.

To increase the discriminatory power of our assays, we combined the 4 sequence-based diversity measures and used logistic regression analyses to identify the combination that performed best. As presented in Figs 9 and 10, only combinations including Shannon entropy, such as Shannon entropy + percent diversity + percent complexity, or Shannon entropy + percent diversity or Shannon entropy + percent complexity of gp120 V3 (AUC = 0.815), gp120 C2_1 (AUC = 0.806) and gp120 C2_3 (AUC = 0.805), presented a performance equivalent to that of Shannon entropy alone for the same respective *env* segments, gp120 V3 (AUC = 0.812), gp120 C2_1 (0.806) and gp120 C2_3 (0.805). The other sequence-based diversity measure combinations for any of the 13 *env* segments analyzed showed a fair to poor discriminatory power (Figs 9 and 10). The *env* gp120 C2 and C3 subregions as too long were previously segmented on 3 and 2 segments respectively to respect the objective of less than 100bp as sliding windows for our analyses. As shown in Table 3, the combining multiple segments didn't increase the discriminatory effect. For example, when combining gp120-C2

sub region including C2_1+C2_2+C2_3 segments the area under the curve (AUC) of ROC of Shannon entropy for example were: 0,790 CI95% [0,757–0,823] and less than this previous value for the 3 other measures. This combination predicts moderate discriminatory effects compared to C2_1 and C2_3 alone which adequately predicted HIV recency with respective AUC = 0.806 and AUC = 0.805. Combining *env* gp120 C3 (C3_1+C3_2 segments) or considering *env* gp120-V1C5 also does not perform well either (AUC<0.8).

Tableau VII. (Article 1, Table 3). Performance of 4 sequence-based diversity measures for non-B HIV-1 subtypes by *env* segments using AUC au ROC analysis.

Env segment	Diversity measure	AUC	95% CI	Optimal Cut-off	TPR (Recent)	TNR (Chronic)	N: recent/chronic)
GP120 C2_1	Percent complexity	0.805	[0,693–0,918]	0,0004	75%	72%	28/29
	Percent diversity	0.842	[0,743–0,941]	0,0231	100%	52%	28/29
	Shannon entropy index	0.844	[0,741–0,946]	0,0895	96%	66%	28/29
	Nb_haplotypes	0.826	[0,720–0,931]	8	96%	55%	28/29
GP120 C3_2	Percent complexity	0.759	[0,631–0,887]	0,0004	93%	55%	28/29
	Percent diversity	0.821	[0,705–0,938]	0,0216	93%	72%	28/29
	Shannon entropy index	0.795	[0,669–0,921]	0,0774	89%	72%	28/29
	Nb_haplotypes	0.849	[0,740–0,959]	7	93%	72%	28/29
GP120-C3	Percent complexity	0.718	[0,623–0,812]	0,0002	38%	62%	28/29
	Percent diversity	0.803	[0,720–0,886]	0,0222	88%	72%	28/29
	Shannon entropy index	0.767	[0,677–0,857]	0,0776	86%	64%	28/29
	Nb_haplotypes	0.812	[0,728–0,896]	7	95%	62%	28/29
GP120 C4	Percent complexity	0.698	[0,557–0,839]	0,0003	89%	55%	28/29
	Percent diversity	0.819	[0,702–0,936]	0,0235	96%	66%	28/29
	Shannon entropy index	0.752	[0,622–0,883]	0,1214	100%	48%	28/29
	Nb_haplotypes	0.756	[0,625–0,887]	7	96%	55%	28/29
GP120 V2	Percent complexity	0.762	[0,621–0,903]	0,0012	79%	67%	28/29
	Percent diversity	0.845	[0,732–0,959]	0,0174	92%	67%	28/29
	Shannon entropy index	0.766	[0,622–0,909]	0,1338	92%	62%	28/29
	Nb_haplotypes	0.761	[0,615–0,907]	8	92%	57%	28/29

TPR (true positive rate) = sensitivity: recent HIV-1+ specimens correctly classified. TNR (true negative rate) = specificity: chronic HIV-1+ specimens correctly classified. The optimal cutoff value indicates the highest accuracy (proportion of patients correctly classified) represented the highest TPR (recency) + TNR (chronic) or (Sensitivity+ Specificity).

<https://doi.org/10.1371/journal.pone.0189999.t003>

Considering only HIV-1 *env* sequences from B subtype specimens, which were the most prevalent in Canada and most represented in our study population (Table 1), the AUC values were slightly increased for two *env* segments, gp120 C2_3 (AUC = 0.810 for subtype B alone compared to 0.805 for all subtypes) and gp120 V3 (AUC = 0.831 for subtype B compare 0.812 for all subtypes), as shown Fig 11 and Table 2.

The performance of sequence-based diversity measures and identifying the most predictable *env* segments were also evaluated for non-B subtypes. As shown in Table 3, the percent complexity, percent diversity, Shannon entropy and number of haplotypes perform better for gp120 C2_1 segment, respectively with an AUC = 0.805, 0.842, 0.844 and 0.826 of each measure. Also, the percent diversity (AUC = 0.821) and number of haplotypes (AUC = 0.849) performed well for *env* gp120 C3_2 segment. Finally, the Percent diversity performed well for gp120 C4 (AUC = 0,819) and gp120 V2 (AUC = 0.84) segments/sub regions. For non-B subtypes, percent diversity seemed to perform well in several *env* segments analyzed (C2_1, C3_2, C4 and V2). But, the lower sequences data, least than 30%, n = 55 (28 recent versus 29 chronic) used in the current study limited the statistical conclusion as well as their performances.

Taking into consideration the Shannon entropy index (S) as the best sequence-based diversity measure and *env* gp120-C2_1, gp120-C2-3 and gp120-V3 as the more predictive *env* subregions/segments, we identified the optimal cut-off values.

Indeed, for all HIV-1 subtypes, the best Shannon entropy index (S) cut-off values were as follows: (S) = 0.054 for gp120-C2_1, (S) = 0.059 for gp120-C2_3 and (S) = 0.083 for gp120-V3 (Table 2). Using these cut-off values, the related sensitivity, which determines the true positive rate (recent HIV-1 infected specimens correctly classified), versus specificity, which determines the true negative rate (chronic HIV-1 infected specimens correctly classified), was 82% versus 69%, 90% versus 67% and 87% versus 66% for the *env* segments gp120-C2_1, gp120-C2_3 and gp120-V3, respectively.

Moreover, if only sequences from subtype B specimens were considered (Table 2), the sensitivity (recent HIV-1 infected individuals correctly classified) versus specificity (chronic HIV-1 infected individuals correctly classified) of the Shannon entropy index (S) was, respectively to 91% versus 66%, with a cut-off of (S) = 0.059 for *env* gp120 C2_3 and 90% versus 66%, respectively, with a cut-off of (S) = 0.082 for gp120 V3.

Further analyses excluding p24 positive samples for recency sequences and including only recent infection as determined by MAA and Western Blot positivity, showed that only the Shannon entropy measure of the gp120-V3 segment presented good discriminatory power

(AUC = 0.801). This measurement presented 89% of sensitivity to identify recent specimens and 66% of specificity for identifying chronic specimens at a cut-off of (S) = 0.0803 (Table 2). On the other hand, when comparing only sequences obtained from p24 positive samples (acute infection), the performance (AUC) and accuracy (sensitivity and specificity) slightly increased for Shannon entropy index (S) measures for three *env* segments (gp120-V3, gp120-C2-1 and gp120-C2-3) (Table 2). Indeed, the Shannon entropy AUC for gp120-V3 was 0.827, which represents a sensitivity of 93% and specificity of 60% at a cut-off value of (S) = 0.097 (Table 2). For gp120-C2_1, the AUC was 0.850, which represents a sensitivity of 93% and specificity of 66% at a cut-off of (S) = 0.060 (Table 2). Finally, for gp120-C2_3, the AUC was 0.844, representing a sensitivity of 91% and specificity of 71% at a cut-off of (S) = 0.048 to correctly identify HIV-1 recency as resumed in Table 2.

4. DISCUSSION

In this study, we assessed the performance of 4 sequence-based diversity measures including percent complexity, percent diversity, Shannon entropy and number of haplotypes used either as independent markers or in combinations to predict HIV-1 infection recency. Our analyses focused on 10 subdomains/sub-regions of the HIV-1 envelope gene between gp120-V1 and gp120-C5 and the gp41-ectodomain. These sub-regions or domains are segmented into 13 fragments of 94 to 158 bp. Because they are too long, the gp120-C2 and gp120-C3 sub-regions are fragmented into 3 and 2 segments, respectively. The choice of these *env* sub-regions was strictly guided by an objective to include all of the gp120 variable regions [70]. Combining multiples segments of them, gp120 C2 (C2_1+C2_2+C2_3) or gp120 C3 (C3_1+C3_2) did not increase discriminatory power of recent HIV-1 infections from chronic ones based on sequences diversities.

We observed that the Shannon entropy measure, which considers the number of reads and proportional representation of each read in individual specimens [42, 43], when calculated for *env* gp120-V3, gp120-C2 segment 1 and gp120-C2 segment 3, can correctly distinguish between recently infected and chronically infected individuals with good performance (AUC \geq 0.8). The fragment lengths of these *env* segments were previously described (i.e., 116

bp for V3, 100 bp for C2_1 and 94 bp for C2_3). These *env* segment sizes suggested that a short fragment of the HIV-1 *env* gene can be useful for predicting HIV-1 recency. Combining Shannon entropy with other measures, such as the percent diversity and/or percent complexity and/or percent complexity within the *env* subdomains, did not markedly increase its predictive value compared to Shannon entropy alone (Figs 9 and 10). This suggests that the single Shannon entropy index as measure performs better than combining with any others diversity measures. It is therefore suggested that the Shannon entropy index (S) within 3 *env* segments (gp120-C2_1, gp120-C2_3 and gp120-V3) as well as HIV-1 subtype B, could be used in public health programs to monitor newly acquired HIV-1 infections in multiple HIV-1 subtype circulating areas.

The utility of viral sequence diversity measurements to determine HIV-1 recency has already been demonstrated [39, 41, 46, and 70]. Analysis of segmented regions of the HIV-1 genome to identify the most predictive genomic regions for infection recency has been previously described for *gag* [70]. In this previous study, Wu et al., 2015 used a longitudinal subtype C sequence and compared 5 *gag* fragments of 50 bp, 100 bp, 150 bp, 200 bp and 250 bp. They observed that the most predictive regions for recency were those with higher mutation rates, such as *gag* p17 and p2/p7/p1/p6, compared to more conserved regions, such as *gag* p24 [49]. Furthermore, data used for the latter study were derived from first generation sequencing, which probably underestimates viral diversity since minor variants need to represent more than 20% of the total population to be detected using this technique [45]. The NGS approach used in our current study has been shown to be more sensitive and may offer the possibility of detecting minor HIV-1 variants/quasi-species that are present in less than 1% of the viral population sampled [42, 43]. We have decided to screen the HIV-1 envelope sequence diversity (the gp120 and gp41 regions) as this gene include the most variable regions of the HIV genome. It is therefore more representative of viral diversification over time as they undergo constant selective pressure from the immune system [71, 72]. Here, when using the median calculation and the frequency distribution curves, we showed that recently infected individuals presented significantly less *env* sequence-based diversity than chronically infected ones. Our data confirmed previous observations indicating that

sequences from recently infected populations are more homogeneous than those from chronically infected populations [50]. These differences were clearly observed by the Shannon entropy, percent diversity and number of haplotypes measures, while the percent complexity was not clearly different between the two groups for 3 *env* segments (S2, S3 and S4 Tables). These results confirm previous results by Cousins et al., 2012 [42], who analyzed mostly subtype D *env* gp41 fragments, while our specimen collection contained predominantly subtype B and evaluated 13 shorts segments of HIV-1 *env* gp120 and gp41 (from 94bp to 158bp). The analysis of shorts segments of the *env* gene is interesting from a technical standpoint since amplification and library preparation from shorts DNA fragments during sequencing is time-saving (i.e., no fragmentation step required) and can be achieved at a lower cost on Illumina MiSeq platforms.

The performance of the sequence-based diversity measure using the AUC of ROC analysis in our current study globally supports previous investigations by Moyo et al., 2016 [73], However this study was conducted on the *gag* and *env* regions from SGA of HIV-1 subtype C using the pairwise genetic distance (PwD) or percent diversity as measure of *env* gp120-V1C5 fragment. The authors determinates an AUC of 0.83 at 130 days of infection [73], which is considered to be good discriminatory power.

Comparatively to Moyo et al., 2016 approaches, the current study determines moderate discriminatory power of *env* gp120-V1C5 with Shannon entropy index of (AUC = 0.765 [0.747–0,784] and for Percent diversity (AUC = 0,704 [0,684–0,723]. Comparatively to our results, sequences data used (HIV-1 C subtype compared to predominant B subtype) may probably impact performance of sequences-based diversity estimating. In the current study, HIV-1 C subtype represented less than 4% (n = 9) of study population so that, we cannot make performant statistical analyses and address objective comparison with Moyo et al., 2016 study. However, our finding may contribute to knowledges with identification of very shorts predictable *env* segments of B and non-B subtypes populations. Also, we have identified a best sequenced-based diversity measure (Shannon entropy) which performed well for accurate identifying of HIV-1 recency. However, it would be subjective to address a comparison using the same algorithm with similar segmented regions if we must consider the

HIV-1 subtype on these published data and compare accuracy of both approaches. Futures studies using the same sequences data would be necessary.

In our current study, we found that only the Shannon entropy index presented good discriminatory power for three *env* segments (gp120-C2-1, gp120-C2_3 and gp120-V3), while the percent diversity measure presented fair or poor discrimination for a few *env* segments. These differences could be linked to the fact that, in our study, shorts segments of the *env* gene were analyzed and that our specimen collection was mostly composed of HIV-1 subtype B. Nevertheless, by 130 days of infection, the sensitivity (true recent infected individuals correctly classified) of *Moyo et al., 2016* [73] study was 79.37% and established (specificity) 72.57% at PwD cut-off of 0.005. This sensitivity (79.37%) was less than that found in our study for Shannon entropy, which provided a sensitivity of 87% and specificity of 66% in gp120 V3, 82% versus 69% in gp120 C2_1 and 90% versus 67% in gp120 C2_3.

These differences indicate that the Shannon entropy index performs better for the identification of HIV-1 recency regarding the highest proportion of recently HIV-1 infected individuals correctly identified comparatively to the chronically infected ones.

Yang et al., 2012 also used the PwD to identify recently HIV-1 subtype B and CRF07_BC infected individuals using the *env* gp120-C2V5 region. In this previous study, authors found an AUC = 0.97 at a sensitivity of 90 to 95% versus specificity of 78.8% in population of (n = 160 for recent versus 264 chronic infected individuals) at a PwD cut-off value of 0.24 by 150–350 days of infection [74]. We observed the similar performance with true recent HIV-1 infected individuals correctly classified (sensitivity) versus chronic ones (specificity) of 90% versus 66% in gp120 V3 and 91% versus 66% in gp120_C2_3 for the Shannon entropy measure.

The HIV-1 recency power increased for acutely infected individuals' samples (p24+ WB-) compared to chronically infected individuals' samples over that of recently infected individuals < 136 days (WB+ and MAA determination) [24] compared to chronically infected individuals, as shown in Table 2. This is consistent with *env* gene diversification (Fiebig stage) following HIV-1 transmission, as shown by *Keele et al., 2008* [50], and confirms the greater HIV-1 *env* sequence homogeneity and diversity increasing in the acute to recent stage of

infection and the highest *env* sequence diversity (heterogeneity) in chronic/late stage of infection.

In summary, our current study shows that the Shannon entropy of HIV-1 *env* gp120-V3 and gp120 C2 segments 1 and 3 correctly predicts recent HIV-1 infections with performant accuracy. Importantly, HIV-1 *env* gp120-V3 was shown to be the best predictor of HIV-1 recency for the B and non-B subtypes and percent diversity for non-B alone. Sequencing of the V3 loop is often performed to determine HIV-1 co-receptor tropism [75] allowing combination with this method to obtain recency data.

Therefore, we suggest that targeted sequencing of short *env* segments can be useful for determining HIV-1 recency with more sensitivity than sequencing the entire *env* gene and may represent an option that minimizes both cost and time factors compared to full-length HIV-1 envelope amplification and sequencing, which constitute a serious limitation for the use of sequence-based diversity for HIV-1 recency identification.

Supporting information

S1 Dataset. Viral sequences data qualifiers.

<https://doi.org/10.1371/journal.pone.0189999.s001>

S1 Table. Sequence based diversity measures calculation methods.

<https://doi.org/10.1371/journal.pone.0189999.s002>

S2 Table. Sequence-based diversity measures expressed as the median/IQR and calculated from NGS of 7 segments representing 4 selected HIV-1 *env* GP120 conserved subdomains.

<https://doi.org/10.1371/journal.pone.0189999.s003>

S3 Table. Sequence-based diversity measures expressed as the median/IQR and calculated from NGS of 5 segments representing 5 selected HIV-1 *env* GP120 variable loops and 1 segment for a part of the gp41ectodomain.

<https://doi.org/10.1371/journal.pone.0189999.s004>

S4 Table. Sequence-based diversity measures expressed as the mean with student t-test results comparing recent versus chronic HIV infected sequences by 13 *env* segments.

<https://doi.org/10.1371/journal.pone.0189999.s005>

Acknowledgments

We are grateful to all the technicians of the molecular biology, serodiagnostic and virology units of the provincial public health reference laboratory (LSPQ) of the “*Institut national de santé publique du Québec*”, Canada, for their technical support. Special thanks to Lyne Désautels, Sylvie-Nancy Beaulac and Lise Leblanc from these units.

We are also grateful to Dr Réjean Dion from the LSPQ, Montréal; and Dr Drissa Sia from the “*Université du Québec en Outaouais*”, Ontario, Canada, for the statistical analysis supporting. We thank Drs Andrés Finzi, Petronela Ancuta, Nicolas Chomont and Guy Lemay, from the department of microbiology, infectiology and immunology of the University of Montreal for their help to conceive this research project. Special thanks to all the HIV-1 infected patients of Quebec from whom serum samples were used in the present study.

Author Contributions

Conceptualization: Alexis Kafando, Bouchra Serhir, Christine Martineau, Florence Doualla-Bell, Mohamed El-Far, Hugues Charest, Cécile L. Tremblay.

Data curation: Alexis Kafando, Bouchra Serhir, Mohamed Sylla, Annie Chamberland.

Formal analysis: Alexis Kafando, Éric Fournier, Mohamed Sylla.

Funding acquisition: Florence Doualla-Bell, Hugues Charest, and Cécile L. Tremblay.

Investigation: Alexis Kafando, Éric Fournier.

Methodology: Alexis Kafando, Éric Fournier, Bouchra Serhir, Christine Martineau, Mohamed Sylla, Annie Chamberland.

Project administration: Florence Doualla-Bell.

Resources: Hugues Charest, Cécile L. Tremblay.

Software: Alexis Kafando, Éric Fournier, Mohamed Ndongo Sangaré.

Supervision: Bouchra Serhir, Christine Martineau, Florence Doualla-Bell, Mohamed El-Far, Hugues Charest, Cécile L. Tremblay.

Validation: Bouchra Serhir, Christine Martineau, Mohamed El-Far, Hugues Charest, Cécile L. Tremblay.

Visualization: Alexis Kafando, Mohamed El-Far, Hugues Charest, Cécile L. Tremblay.

Writing original draft: Alexis Kafando, Éric Fournier, Bouchra Serhir, Christine Martineau, Florence Doualla-Bell, Mohamed Ndong Sangaré, Mohamed Sylla, Annie Chamberland, Hugues Charest, Cécile L. Tremblay.

Writing reviews & editing: Alexis Kafando, Florence Doualla-Bell, Hugues Charest, and Cécile L. Tremblay.

5. ANNEXE (ARTICLE 1)

Tableau VIII. Tableau: (Annexe 1, S1 table, Article 1): Methode explicative du calcul des quatre mesures de la diversité génétique des séquences virales (Shannon entropy, complexity, diversity (exemple)).

<p>Exemple, on a trois séquences nucléotidiques différentes ayant les proportions suivantes;</p> <p>Seq1 : x10 AAACTGGGAAT Seq2 : x15 AAACGGGGAAT Seq3 : x20 AGACTGGGGAT</p> <p>=></p> <p>Nombre de séquences uniques = 3 Total = 45</p>
<p>1. Calcul du Shannon entropie (fonction def ComputeShannon())</p> <p>Probabilité de Seq1 = 10/45 Probabilité de Seq2 = 15/45 Probabilité de Seq3 = 20/45</p> <p>↓</p> <p>Shannon de Seq1 = Probabilité de Seq1 x log (Probabilité de Seq1) Shannon de Seq2 = Probabilité de Seq2 x log (Probabilité de Seq2) Shannon de Seq3 = Probabilité de Seq3 x log (Probabilité de Seq3)</p> <p>↓</p> <p>Somme des Shannon = Shannon de Seq1 + Shannon de Seq2 + Shannon de Seq3</p> <p>↓</p> <p>Shannon = $-(1/\log(\text{Total})) \times \text{Sommes des Shannon}$</p>
<p>2. Calcul de la Complexity (fonction def ComputeComplexity())</p> <p>Complexity = Nombre de séquences uniques / Total</p>
<p>3. Calcul de la Diversity (fonction def ComputeDiversity())</p> <p>On crée un fichier mutlifasta avec nos trois séquences uniques =></p> <pre>>Seq1 AAACTGGGAAT >Seq2 AAACGGGGAAT >Seq3 AGACTGGGGAT</pre> <p>↓</p> <p>On aligne les séquences du fichier multifasta avec la fonction msa (librarie R msa) du scriptComputeDiversity.R (https://github.com/EricFournier3/HIVvariant/blob/master/ComputeDiversity.R) => un alignement multiple (msa)</p> <p>↓</p> <p>On calcul la diversité nucléotidique du msa avec la fonction nuc.div (libraire R pegas; du script ComputeDiversity.R.</p>
<p>4. Calcul du nombre d'haplotype= Nombre de séquences uniques</p>

Les calculs de diversité (Shannon entropy, complexity, diversity) sont effectués avec le script `pythongp120variantcomputer.py` : (<https://github.com/EricFournier3/HIVvariant/blob/master/GP120VariantComputer.py>)

6. REFERENCES

1. Public Health Agency of Canada. HIV and AIDS in Canada, surveillance report to December 31, 2014. <http://healthycanadians.gc.ca/publications/diseases-conditions-maladies-affections/hiv-aids-surveillance-2014-vih-sida/index-eng.php?page=6> .
2. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stages of infection. *Journal of Infectious Diseases*. 2008; 198(5):687–93.
3. Van Sighem A, Nakagawa F, De Angelis D, Quinten C, Bezemer D, de Coul EO, et al. Estimating HIV Incidence, Time to Diagnosis, and the Undiagnosed HIV Epidemic Using Routine Surveillance Data. *Epidemiology*. 2015; 26(5):653–60.
4. Mastro TD. Determining HIV Incidence in Populations: Moving in the Right Direction. *Journal of Infectious Diseases*. 2013; 207(2):204–6.
1. Smith MK, Rutstein SE, Powers KA, Fidler S, Miller WC, Eron JJ Jr., et al. The Detection and Management of Early HIV Infection: A Clinical and Public Health Emergency. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2013; 63:S187–S99.
2. Sharma UK, Schito M, Welte A, Rousseau C, Fitzgibbon J, Keele B, et al. Workshop summary: Novel biomarkers for HIV incidence assay development. *AIDS research and human retroviruses*. 2012;28(6):532–9.
3. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li CJ, Wabwire-Mangen F, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N Engl J Med*. 2000;342(13):921–9.
4. Hall HI, Green TA, Wolitski RJ, Holtgrave DR, Rhodes P, Lehman JS, et al. Estimated Future HIV Prevalence, Incidence, and Potential Infections Averted in the United States: A Multiple Scenario Analysis. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2010;55(2):271–6.

5. Gray RH, Li X, Kigozi G, Serwadda D, Nalugoda F, Watya S, et al. The impact of male circumcision on HIV incidence and cost per infection prevented: a stochastic simulation model from Rakai, Uganda. *Aids*. 2007;21(7):845–50.
6. Fiamma A, Lissouba P, Amy OE, Singh B, Laeyendecker O, Quinn TC, et al. Can HIV incidence testing be used for evaluating HIV intervention programs? A reanalysis of the Orange Farm male circumcision trial (ANRS-1265). *Bmc Infectious Diseases*. 2010;10.
7. Miller WC, Rosenberg NE, Rutstein SE, Powers KA. Role of acute and early HIV infection in the sexual transmission of HIV. *Curr Opin HIV AIDS*. 2010;5(4):277–82.
8. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis*. 2007;195(7):951–9.
9. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, Weinberg J, et al. Transmission Clustering Drives the Onward Spread of the HIV Epidemic Among Men Who Have Sex with Men in Quebec. *Journal of Infectious Diseases*. 2011;204(7):1115–9.
10. Aghaizu A, Murphy G, Tosswill J, De Angelis D, Charlett A, Gill ON, et al. Recent infection testing algorithms (RITA) applied to new HIV diagnoses in England, Wales and Northern Ireland, 2009 to 2011. *Eurosurveillance*. 2014;19(2):5–11.
11. Murphy G, Parry JV. Assays for the detection of recent infections with human immunodeficiency virus type 1. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*. 2008;13(36).
12. Kim AA, Parekh BS, Umuro M, Galgalo T, Bunnell R, Makokha E, et al. Identifying Risk Factors for Recent HIV Infection in Kenya Using a Recent Infection Testing Algorithm: Results from a Nationally Representative Population-Based Survey. *Plos One*. 2016;11(5).

13. Moyo S, Wilkinson E, Novitsky V, Vandormael A, Gaseitsiwe S, Essex M, et al. Identifying Recent HIV Infections: From Serological Assays to Genomics. *Viruses-Basel*. 2015;7(10):5508–24.
14. Truong HHM, Kellogg T, Louie B, Klausner J, Dilley J, McFarland W. Recent HIV-1 Infection Detection: Comparison of Incidence Estimates Derived by Laboratory Assays and Repeat Testing Data. *Aids-Journal of Acquired Immune Deficiency Syndromes*. 2009;51(4):502–5.
15. Brookmeyer R. Measuring the HIV/AIDS Epidemic: Approaches and Challenges. *Epidemiologic Reviews*. 2010;32(1):26–37.
16. Barnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell M-L. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. *Aids*. 2008;22(1):139–44.
17. Kumwenda N, Hoffman I, Chirenje M, Kelly C, Coletti A, Ristow A, et al. HIV incidence among women of reproductive age in Malawi and Zimbabwe. *Sexually Transmitted Diseases*. 2006;33(11):646–51.
18. Ghys PD, Kufa E, George MV, Mod URGE. Measuring trends in prevalence and incidence of HIV infection in countries with generalized epidemics. *Sexually Transmitted Infections*. 2006;82:152–16.
19. Rosenberg NE, Pilcher CD, Busch MP, Cohen MS. How can we better identify early HIV infections? *Current Opinion in Hiv and Aids*. 2015;10(1):61–8.
20. Serhir B, Hamel D, Doualla-Bell F, Routy JP, Beaulac S-N, Legault M, et al. Performance of Bio-Rad and Limiting Antigen Avidity Assays in Detecting Recent HIV Infections Using the Quebec Primary HIV-1 Infection Cohort. *Plos One*. 2016;11(5).

21. Duong YT, Qiu M, De AK, Jackson K, Dobbs T, Kim AA, et al. Detection of Recent HIV-1 Infection Using a New Limiting-Antigen Avidity Assay: Potential for HIV-1 Incidence Estimates and Avidity Maturation Studies. *Plos One*. 2012;7(3).
22. Le Vu S, Meyer L, Cazein F, Pillonel J, Semaille C, Barin F, et al. Performance of an immunoassay at detecting recent infection among reported HIV diagnoses. *Aids*. 2009;23(13):1771–7.
23. McWalter TA, Welte A. A Comparison of Biomarker Based Incidence Estimators. *Plos One*. 2009;4(10).
24. Morgado MG, Bastos FI. Estimates of HIV-1 incidence based on serological methods: a brief methodological review. *Cadernos De Saude Publica*. 2011;27:S7–S18.
25. Mullis CE, Munshaw S, Grabowski MK, Eshleman SH, Serwadda D, Brookmeyer R, et al. Differential Specificity of HIV Incidence Assays in HIV Subtypes A and D-Infected Individuals from Rakai, Uganda. *Aids Research and Human Retroviruses*. 2013;29(8):1146–50.
26. Xu J, Wang H, Jiang Y, Ding G, Jia M, Wang G, et al. Application of the BED capture enzyme immunoassay for HIV incidence estimation among female sex workers in Kaiyuan City, China, 2006–2007. *International Journal of Infectious Diseases*. 2010;14(7):E608–E12.
27. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, Green T, et al. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence. *Aids Research and Human Retroviruses*. 2002;18(4):295–307.
28. Suligoi B, Massi M, Galli C, Sciandra M, Di Sora F, Pezzotti P, et al. Identifying recent HIV infections using the avidity index and an automated enzyme immunoassay. *Jaids- Journal of Acquired Immune Deficiency Syndromes*. 2003;32(4):424–8.

29. Salustiano DM, de Lima KO, Salustiano Cavalcanti AM, Diaz RS, Lacerda HR. Comparison among the BED capture enzyme immunoassay test and AxSYM avidity index assay for determining recent HIV infection and incidence in two Voluntary Counselling and Testing Centres in Northeast Brazil. *Brazilian Journal of Infectious Diseases*. 2014;18(4):449–53.
30. Duong YT, Kasantjee R, Welte A, Morgan M, De A, Dobbs T, et al. Recalibration of the Limiting Antigen Avidity EIA to Determine Mean Duration of Recent Infection in Divergent HIV-1 Subtypes. *Plos One*. 2015;10(2):15.
31. Draenert R, Verrill CL, Tang Y, Allen TM, Wurcel AG, Boczanowski M, et al. Persistent recognition of autologous virus by high-avidity CD8 T cells in chronic, progressive human immunodeficiency virus type 1 infection. *Journal of Virology*. 2004;78(2):630–41.
32. Cousins MM, Swan D, Magaret CA, Hoover DR, Eshleman SH. Analysis of HIV Using a High-Resolution Melting (HRM) Diversity Assay: Automation of HRM Data Analysis Enhances the Utility of the Assay for Analysis of HIV Incidence. *Plos One*. 2012;7(12).
33. Cousins MM, Laeyendecker O, Beauchamp G, Brookmeyer R, Towler WI, Hudelson SE, et al. Use of a High-Resolution Melting (HRM) Assay to Compare *gag*, *pol*, and *env* Diversity in Adults with Different Stages of HIV Infection. *Plos One*. 2011;6(11).
34. Towler WI, James MM, Ray SC, Wang L, Donnell D, Mwatha A, et al. Analysis of HIV Diversity Using a High-Resolution Melting Assay. *Aids Research and Human Retroviruses*. 2010;26(8):913–8.
35. Kouyos RD, von Wyl V, Yerly S, Boeni J, Rieder P, Joos B, et al. Ambiguous Nucleotide Calls From Population-based sequencing of HIV-1 are a Marker for Viral Diversity and the Age of Infection. *Clinical Infectious Diseases*. 2011;52(4):532–9.

36. Allam O, Samarani S, Ahmad A. Hammering out HIV-1 incidence with Hamming distance. *Aids*. 2011;25(16):2047–8.
37. Park SY, Love TMT, Nelson J, Thurston SW, Perelson AS, Lee HY. Designing a genome-based HIV incidence assay with high sensitivity and specificity. *Aids*. 2011;25(16):F13–F9.
38. Cousins MM, Ou S-S, Wawer MJ, Munshaw S, Swan D, Margaret CA, et al. Comparison of a High-Resolution Melting Assay to Next-Generation Sequencing for Analysis of HIV Diversity. *Journal of Clinical Microbiology*. 2012;50(9):3054–9.
39. Wu JW, Patterson-Lomba O, Novitsky V, Pagano M. A Generalized Entropy Measure of Within-Host Viral Diversity for Identifying Recent HIV-1 Infections. *Medicine*. 2015;94(42).
40. Bimber BN, Dudley DM, Lauck M, Becker EA, Chin EN, Lank SM, et al. Whole-Genome Characterization of Human and Simian Immunodeficiency Virus Intra-host Diversity by Ultradeep Pyrosequencing. *J Virol*. 2010;84(22):12087–92.
41. Mohamed S, Penaranda G, Gonzalez D, Camus C, Khiri H, Boulme R, et al. Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure. *Aids*. 2014;28(9):1315–24.
42. Xia X-Y, Ge M, Hsi JH, He X, Ruan Y-H, Wang Z-X, et al. High-Accuracy Identification of Incident HIV-1 Infections Using a Sequence Clustering Based Diversity Measure. *Plos One*. 2014;9(6).
43. Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science*. 1998;280(5371):1884–8.
44. Moulard M, Hallenberger S, Garten W, Klenk HD. Processing and routing of HIV glycoproteins by furin to the cell surface. *Virus Res*. 1999;60(1):55–65.

45. Haim H, Salas I, Sodroski J. Proteolytic Processing of the Human Immunodeficiency Virus Envelope Glycoprotein Precursor Decreases Conformational Flexibility. *J Virol.* 2013;87(3):1884–9.
46. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America.* 2008;105(21):7552–7.
47. Hartley O, Klasse PJ, Sattentau QJ, Moore JP. V3: HIV's switch-hitter. *AIDS Res Hum Retroviruses.* 2005;21(2):171–89.
48. Checkley MA, Lutttge BG, Freed EO. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J Mol Biol.* 2011;410(4):582–608.
49. Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H, et al. identification and characterization of conserved and variable regions in the envelope gene of HTLV-III LAV, the retrovirus of AIDS. *Cell.* 1986;45(5):637–48.
50. Freed EO. HIV-1 replication. *Somatic cell and molecular genetics.* 2001;26(1–6):13–33.
51. Yu F, Lu L, Du LY, Zhu XJ, Debnath AK, Jiang SB. Approaches for Identification of HIV-1 Entry Inhibitors Targeting gp41 Pocket. *Viruses.* 2013;5(1):127–49.
52. Douglas NW, Munro GH, Daniels RS. HIV/SIV glycoproteins: Structure-function relationships. *Journal of Molecular Biology.* 1997;273(1):122–49.
53. Asin-Milan O, Wei Y, Sylla M, Vaisheva F, Chamberland A, Tremblay CL. Performance of a clonal-based HIV-1 tropism phenotypic assay. *J Virol Methods.* 2014;204:53–61.
54. Sanders-Buell E, Saliminen S.O., McCutchan F.E., 1995. In: *Sequencing primers for HIV1.* Los Alamos National Laboratory. 1995;pp. 1521–1527.

55. Cleghorn FR, Jack N, Carr JK, Edwards J, Mahabir B, Sill A, et al. A distinctive clade BHIV type 1 is heterosexually transmitted in Trinidad and Tobago. *Proc Natl Acad Sci U S A*. 2000;97(19):10532–7.
56. Hunt M, Gall A, Ong SH, Brener J, Ferns B, Goulder P, et al. IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics*. 2015;31(14):2374–6.
57. Caffrey M. HIV envelope: challenges and opportunities for development of entry inhibitors. *Trends in Microbiology*. 2011;19(4):191–7.
58. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016;33(7):1870–4.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
60. Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. *Bioinformatics*. 2015;31(24):3997–9.
61. Wickham H. *ggplot2 Elegant Graphics for Data Analysis*. Second Edition ed: Springer; 2016. 268 p.
62. Pineda-Pena AC, Faria NR, Imbrechts S, Libin P, Abecasis AB, Deforche K, 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. *Infect Genet Evol*. 2013;19:337–48.
63. Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T. A web-based genotyping resource for viral sequences. *Nucleic Acids Res*. 2004;32:W654–W9.
64. Kleinbaum DG, Klein M. *Survival Analysis A Self-Learning Text*, Third Edition. 3, editor. Verlag New York: Springer; 2012. XV, 700 p.

65. NCSS. One ROC Curve and Cutoff Analysis: NCSS.com; 2017 [cited 2017 2017-01-01]. Chapter 546: [<https://www.ncss.com/software/ncss/roc-curves-ncss/>] .
66. Wu JLW, Patterson-Lomba O, Novitsky V, Pagano M. A Generalized Entropy Measure of Within-Host Viral Diversity for Identifying Recent HIV-1 Infections. *Medicine*. 2015;94(42).
67. Watkins BA, Reitz MS, Wilson CA, Aldrich K, Davis AE, Robertguroff M. Immune escape by Human-immunodeficiency-virus type-1 from neutralizing antibodies—evidence for multiple pathways. *J Virol*. 1993;67(12):7493–500.
68. Hamoudi M, Simon-Loriere E, Gasser R, Negroni M. Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. *Retrovirology*. 2013;10.
69. Moyo S, Vandormael A, Wilkinson E, Engelbrecht S, Gaseitsiwe S, Kotokwe KP, et al. Analysis of Viral Diversity in Relation to the Recency of HIV-1C Infection in Botswana. *Plos One*. 2016;11(8).
70. Yang J, Xia X, He X, Yang S, Ruan Y, Zhao Q, et al. A new pattern-based method for identifying recent HIV-1 infections from the viral sequence. *Science China-Life Sciences*. 2012;55(4):328–35.
71. Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, et al. Use of Four Next-Generation Sequencing Platforms to Determine HIV-1 Coreceptor Tropism. *PLoS One*. 2012;7(11).

ARTICLE 2

**A SHORT-TERM ASSESSMENT OF NASCENT HIV-1
TRANSMISSION CLUSTERS AMONG NEWLY DIAGNOSED
INDIVIDUALS USING ENVELOPE SEQUENCE-BASED
PHYLOGENETIC ANALYSES (667)**

Alexis Kafando,¹ Bouchra Serhir,² Florence Doualla-Bell,² Eric Fournier,² Mohamed Ndongo Sangaré,⁵ Christine Martineau,² Mohamed Sylla,³ Annie Chamberland,³ Mohamed El-Far,³ Hugues Charest,^{1,2} Cécile L. Tremblay^{1,2,3§}.

Publié dans le journal: AIDS research and human retroviruses. 2019 Oct; 35(10):906-919. doi: 10.1089/AID.2019.0142

¹ Département de microbiologie, infectiologie et immunologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada; ² Laboratoire de santé publique du Québec, Institut national de santé publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada; ³ Centre de recherche du Centre hospitalier de l'Université de Montréal, Montréal, Québec, Canada; ⁵ Département de médecine sociale et préventive, École de santé publique, Université de Montréal, Montréal, Québec, Canada.

§ Corresponding author: Cécile L. Tremblay, MD, FRCPC.

Département de Microbiologie, infectiologie et immunologie

Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM)

██

██

Tel.: ██ FAX: ██

E-mail: ██

CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 2

M. Kafando Alexis a contribué à l'article 2 pour la conception de l'étude selon les directives du Dre Cécile Tremblay. Il a passé en revue les procédures expérimentales de l'étude sous la supervision du Dr Hugues Charest, Dre Christine Martineau et du Dre Annie Chamberland. M Kafando a réalisé toutes les expériences de laboratoire passant par l'identification des échantillons disponibles dans la serothèque du Laboratoire de santé publique du Québec (sous le contrôle du Dre Bouchra Sherir), à l'extraction des ARN du VIH, aux analyses de biologie moléculaire (PCR), aux séquençages (NGS) et à l'assemblage des séquences (supervision M Éric Fournier). M Kafando a réalisé les analyses bio-informatiques et une partie des analyses statistiques (supervision Dr Mohamed Ndongo Sangaré). Il a également assuré la synthèse de tous les résultats de l'étude et rédigé le premier manuscrit qu'il a soumis à l'ensemble de l'équipe de recherche pour les révisions et commentaires. Il a par la suite rédigé le manuscrit final, identifié le journal en concertation avec l'équipe et procédé à sa soumission. Il a également préparé les séquences de l'enveloppe du VIH-1 de tous les sujets de l'étude qu'il a soumis dans la base de données de Genbank en collaboration avec Éric Fournier.

M Kafando a interagi avec les éditeurs du journal et les reviseurs de l'article et a apporté toutes les corrections nécessaires au manuscrit en collaboration avec les coauteurs. M Kafando a rédigé la version finale du manuscrit soumise au journal qui a été publiée.

En somme, M. Kafando a participé à la conception de l'étude, à la réalisation des expériences de laboratoire et a été l'interface entre les différents coauteurs, les éditeurs et les reviseurs pour la rédaction du manuscrit.

Les contributions de chacun et de chacune des coauteurs sont décrites plus en détail dans l'article.

RÉSUMÉ

Contexte : L'identification des réseaux de transmission du VIH-1 à l'aide d'analyses phylogénétiques peut fournir des informations sur la chaîne de transmission virale entre les individus et contribuerait à améliorer les stratégies de prévention. Nous avons comparé et évalué l'usage d'un fragment partiel de l'enveloppe du VIH-1 d'une taille de 1070 paires de bases avec celui de la boucle V3 d'une taille de 108 paires de bases en tant qu'outils indépendants capables de prédire et d'identifier convenablement les réseaux de transmission ou cluster.

Méthodes : Les acides ribonucléiques (ARN) viraux extraits de sérums issus de patients infectés par le VIH-1 et nouvellement diagnostiqués en 2015 et comprenant des cas récents (n=106) et des cas chroniques (n=156) ont été amplifiés et séquencés. Des analyses phylogénétiques ont porté sur des séquences nucléotidiques de l'enveloppe du VIH-1 et servi à l'évaluation et l'identification des réseaux de transmission au sein de cette population. Les réseaux de transmission entre paires de séquences ont été construits à l'aide du logiciel ClusterPickerGUI_1.2.3 en considérant des distances génétiques de $\leq 10\%$ entre elles. Des analyses de régression logistique ont été utilisées pour examiner la relation entre les facteurs démographiques qui sont probablement associés à la formation de ces réseaux de transmission du VIH-1.

Résultats : Quarante-vingt-dix-huit séquences consensuelles de l'enveloppe soit une séquence par sujet infecté par le VIH-1 ont été finalement obtenues et soumises aux analyses phylogénétiques. En considérant le fragment partiel de la séquence de l'enveloppe du VIH-1 de la taille de 1070 pb, 42 séquences individuelles se sont regroupées en 15 petits réseaux de transmission. Cependant, en considérant le fragment constitué de la boucle V3 uniquement d'une taille de 108 pb, seulement 10 des 15 réseaux de transmission ont pu être reproduits. L'agrément de concordance entre les deux fragments de la séquence de l'enveloppe du VIH-1 à pouvoir identifier correctement les mêmes réseaux de transmission est statistiquement significatif, mais modéré. Le coefficient Kappa (K) de Cohen est de 0.59, $p < 0.00001$.

L'âge moyen (<38,8 ans) et le sous-type du VIH-1 (B) ont été identifiés comme deux facteurs démographiques significativement associés à la formation des réseaux de transmission du VIH-1 dans cette population. Le rapport de cotes (Odds ratio, OR) est successivement de: 0.25, IC95% [0.04-0.66], P = 0,002 pour l'âge et de 0.17, IC95% [0.10 à 0.6], p = 0.011 pour le sous type viral associé à la formation des réseaux de transmission.

Conclusion : La présente étude confirme que le fragment partiel de la séquence de l'enveloppe virale d'une taille de 1070 pb est un meilleur prédicteur pour l'identification des réseaux de transmission du VIH-1. Cependant, le fragment constitué des séquences de la boucle 3 peut être utile pour le dépistage des réseaux de transmission naissante. Elle pourrait en effet être plus facilement intégrable dans les programmes de surveillance.

Mots-clés: réseaux de transmission du VIH-1; grappes, séquences du gène d'enveloppe; distance génétique par paire.

ABSTRACT

Background: The identification of transmission clusters (TCs) of HIV-1 using phylogenetic analyses can provide insights into viral transmission network and help improve prevention strategies. We compared the use of partial HIV-1 envelope fragment of 1070 bp with its loop 3 (108 bp) to determine their utility in inferring HIV-1 transmission clustering.

Methods: Serum samples of recently (n=106) and chronically (n=156) HIV-1 infected patients with status confirmed were sequenced. HIV-1 envelope nucleotide-based phylogenetic analyses were used to infer HIV-1 transmission clusters. Those were constructed using ClusterPickerGUI_1.2.3 considering a pairwise genetic distance of $\leq 10\%$ threshold. Logistic regression analyses were used to examine the relationship between the demographic factors that were likely associated with HIV-1 clustering.

Results: Ninety-eight distinct consensus envelope sequences subjected to phylogenetic analyses. Using a partial envelope fragment sequence, 42 sequences were grouped into 15 distinct small transmission clusters while the V3 loop reproduces 10 clusters. The agreement between the partial envelope and the V3 loop fragment was significantly moderate with a Cohen's Kappa (κ) coefficient of 0.59, $p < 0.00001$. The mean age (< 38.8 years) and HIV-1 B subtype are two factors identified that were significantly associated with HIV-1 transmission clustering in the cohort, OR= 0.25, 95% CI [0.04-0.66], $P=0.002$ and OR: 0.17 95% CI [0.10-0.61, $P=0.011$, respectively.

Conclusion: The current study confirms that a partial fragment of the HIV-1 envelope sequence is a better predictor of transmission clustering. However, the loop 3 segment may be useful in screening purposes and may be more amenable to integration in surveillance programs.

Keywords: HIV-1 transmission networks; clusters; envelope gene sequences; pairwise genetic distance

1. INTRODUCTION

In 2015, a total of 609 HIV cases were reported to the public health laboratory in Quebec (LSPQ), including 299 new diagnoses¹. To better characterize transmission patterns^{2,3}, we evaluated molecular epidemiology methods using phylogenetic analyses of HIV genome sequences obtained from infected individuals. Pairwise genetic distances derived from *env*, *gag* and *pol* nucleotide sequences were used to assess the relationship between sequences and therefore define study HIV transmission dynamics^{4,5}. Combining sequence-based clustering with risk factors and sociodemographic, clinical and geographical parameters may help identify patterns of transmission in the HIV-infected populations⁶⁻¹¹. The *env* gene, encoding glycoproteins gp120 and gp41, is the most variable region of the HIV-1 genome¹²⁻¹⁴. GP120 contains 5 hypervariable regions (V1 to V5) and 5 conserved regions (C1-C5)¹⁵⁻¹⁸. The gp41 region consists of 3 domains, the ectodomain (ECD), the transmembrane domain (TM) and the long cytoplasmic domain (CP)¹⁹. Each subregion and domain of the *env* plays a key role in HIV pathogenesis^{16,18,20}. In the literature, *pol* gene-derived sequences are mostly used for phylogenetic analysis to infer HIV-1 transmission clustering²¹. For example, Hassan AS *et al.* AIDS. 2017, reported that around 98% of 105 publish articles from 2005 to 2016 related to HIV-1 transmission clustering used mostly HIV-1 *pol* sequences as genetic regions. The main reason is that *Pol* sequences are readily available in clinical laboratories performing drug resistance testing in most countries. However, little is known about the potential of HIV-1 envelope-derived sequences for this application^{22,23}. It has been shown that within-host *env* genetic diversity is significantly associated with HIV-1 Fiebig stage and disease progression^{24,25}. The HIV-1 evolution in a new recipient may recover some ancestral features of infected donor 26, such as the genetic distance between donor and recipient^{22,27}. Such characteristics may be used to establish the closely transmitted network (cluster) between HIV-infected individuals.

The first objective of this study was to assess the potential of HIV *env*-derived sequences (1070 bp) to determine HIV-1 transmission clustering.

The second objective was to investigate whether shorter portions of the *env* gene, such as the *env* V3 loop-derived sequences (108 bp), could be a valuable tool to achieve the same meanwhile reducing technical, time and cost constraints^{28,29} associated with partial but long length *env* sequencing. The third objective was to identify risk factors associated with HIV transmission clustering.

2. MATERIALS AND METHODS

Patients and specimens

Serum samples that were reactive in Quebec diagnostic laboratories using a screening HIV-1/2 enzyme immunoassay (EIA) was submitted to the LSPQ for confirmation by Western blot (WB) and/or p24 EIA. Positive p24 antigen samples were de facto classified as recent infections. Samples from newly confirmed HIV-1 individuals by WB were then submitted to a recent infection testing algorithms (RITA) based on antibody avidity. The latter combines a Centers for Disease Control and Prevention (CDC)-modified Bio-Rad Avidity Assay and Sedia-LAg-Avidity assay³⁰. Recent infections were defined as a sample that was positive for HIV-1 p24 antigen or positive for HIV-1 antibodies by Western blot but classified as “recent” by RITA testing (≤ 136 days of infection). Established infections were defined as samples positive by WB and classified as long standing by RITA testing (> 136 days of infection).

Based on these criteria, we selected 262 newly diagnosed HIV samples that included recent ($n=106$) and long-standing ($n=156$) infections collected in 2015. The risk factors and clinical and epidemiological parameter data reported for each sample were extracted from the HIV provincial surveillance program.

Amplification and Sequencing of Partial Envelope Fragment

Amplification and sequencing protocols were conducted as previously described by Kafando et al.³¹. Briefly, total nucleic acids were extracted from 100 μ l of serum using an automated BioRobot MDx extraction platform. HIV-1 RNA was amplified using the Superscript III One-Step RT-PCR system with platinum[®] Taq DNA polymerase (Invitrogen, Thermo Fisher Scientific, and Carlsbad, CA, USA) with primers *env* up forward (5'-GTTTCTTTTAGGCATCTCCTATGGCAGGAAGAAG-3', HXB2 positions 5957-5983) and *env*-lo reverse (5'-GTTTCTTCCAGTCCCCCTTTTCTTTTAAAAG-3', HXB2 positions 9063–9088)³². Nested amplification was performed using the Expand[™] High Fidelity PCR System Kit (Roche Diagnostics, Indianapolis, USA). Primers E60F forward (5'-TAATCAGTTTATGGGATCAAAGC-3', HXB2 positions 6547-6569)³³ and E55R reverse (5'-GCCCCAGACTGTGAGTTGCAACAGATG-3,

HXB2 positions 7940-7914)³⁴, were used, generating PCR products covering \approx 1400 bp of the *env* gene. Amplification conditions, library preparation and Illumina MiSeq next generation sequencing (NGS) were conducted as previously described³¹. Following quality control with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), raw sequence reads were de novo assembled using Iterative Virus Assembler (IVA)³⁵ to generate a consensus sequence.

Sequence data processing, phylogenetic analysis and HIV transmission cluster reconstruction.

All consensus *env* nucleic acid sequences were aligned with Molecular Evolutionary Genetics Analysis Using (MEGA7) software (www.megasoftware.net) under ClustalW method³⁶. All aligned sequences were submitted to MAFFT multiple sequence alignment software version 7.37 to verify the reliability of the alignments. The human immunodeficiency virus type 1 K03455.1 (HXB2) *env* nucleotide sequence (nt) positions (6225 to 8795) was included in the alignment to serve as a reference. Matching sequences, excluding gaps with equal lengths (HXB2 nt positions 6831 to 7900 \approx 1070 bp) in this case, were selected for phylogenetic analyses. The *env* partial fragment analyzed included the gp120-C2 to C5 subregions (HXB2 nt positions 6813 to 7757) and the gp41 partial ectodomain (HXB2 nt positions 7758 to 7915). The HIV-1 *env* gp120-V3 loop sequences (HXB2 genome nt positions 7110 to 7217 \approx 108 bp) were also analyzed separately. Phylogenetic trees were constructed in MEGA7 using the maximum likelihood (ML) algorithm, and their reliability was estimated from 1000 bootstraps replicates. Transmission clusters were evaluated among sequences that grouped around common proximal nodes with \geq 99% bootstraps, as supported by a previous study¹¹. In the current study, in the absence of the gold standard for HIV-1 envelope sequence-based clustering, we considered a pairwise genetic distance of \leq 10% as a threshold, as suggested by Novitsky et al.²². The HIV-1 transmission clustering was defined as two or more HIV-1-infected individual genomic sequences whose branches were grouped under a genetic distance threshold in the phylogenetic tree²¹. The interpretation of the extent of the relevant transmission clustering depended on the genetic distances threshold, the time that it was established, the number sequence dataset, the sampling densities, the node bootstrap

and the HIV-1 genomics region concern ^{21, 22}. We considered a high pairwise genetic distance (PWD) ($\leq 10\%$) to define clustering for the HIV-1 envelope because it presented the highest variability compared to *gag* and *pol* (PWD $\leq 1.5\%$). The extent of the HIV transmission cluster was classified as unique (1 member), small (2-4 members), or large (5-60 members) in the transmission chain ⁶. We used ClusterPickerGUI_1.2.3 (<http://hiv.bio.ed.ac.uk/software.html>) to construct cluster trees ³⁸ and Figtree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) to view them. We used Dendroscope version 3.5.10 to build a Tanglegram of connected taxa between rooted phylogenetics trees and networks ^{39, 40}. The NCBI subtyping tool ⁴¹ was used to determine HIV viral subtypes, which was confirmed by the REGA HIV-1 subtyping tool version 3 ⁴².

Statistical analysis.

We used bivariate and multivariate analyses to determine the independence of associations between exposed variables (epidemiologic and clinical factors) and outcomes (clustering). Categorical variables were compared using a chi-square or Fisher's exact test, and continuous variables were compared using a two-sample Student's t-test or Wilcoxon rank sum test. All variables with $P < 0.25$ as a cutoff point from bivariate analysis were included in the multivariate analysis using logistic regression. At this point, $P < 0.05$ ^{43, 44} using the Wald test was considered to be statistically significant, and selected variables were included in the final regression model. Statistics analyses were performed in SPSS version 24 and STATA version 14. The determination of agreement between *env* gp120-V3 sequences clustering determining with a partial *env* fragment as a gold standard to correctly identify individuals in a cluster tree was assessed using Cohen's Kappa coefficient. The estimated parameters were sensitivity, specificity, positive predictive value, and negative predictive value. Stata IC version 14 was used as a statistical package, and a p value less than 5% was considered statistically significant.

Ethics statement. All work was conducted in accordance with the Declaration of Helsinki in terms of informed consent. All samples were anonymized before we accessed them for this study. No nominal information was used for analysis or data management. Ethical approval

was given and renewed yearly by our institutional Review Board (IRB): the « Comité d'éthique et de la recherche des Centres hospitaliers affiliés à l'Université de Montréal ».

3. RESULTS

Amplification and DNA sequencing of HIV-1 *env* were attempted on 262 specimens and was successful for 39% of them (n=102) (Figure 1). We think that the samples qualities related to the conditions of storage, the technical procedures beginning of HIV-1 viral RNA extraction, associate with viral loads of <20,000 copies/ml for most of them, the PCR enzymes and the specificity of the HIV-1 envelope known to be difficult to amplify or the sequence length^{28,29} constitute various factors that may have possibly influenced the amplification success rate. Of the included HIV-1 *env* sequences (n=102), 47% (n=48) and 35% (n=36) were classified as chronic and recent HIV infections, respectively, using the RITA algorithm (26), whereas 18% (n=18) were found to be recent as confirmed by EIA-p24 testing (Figure 1).

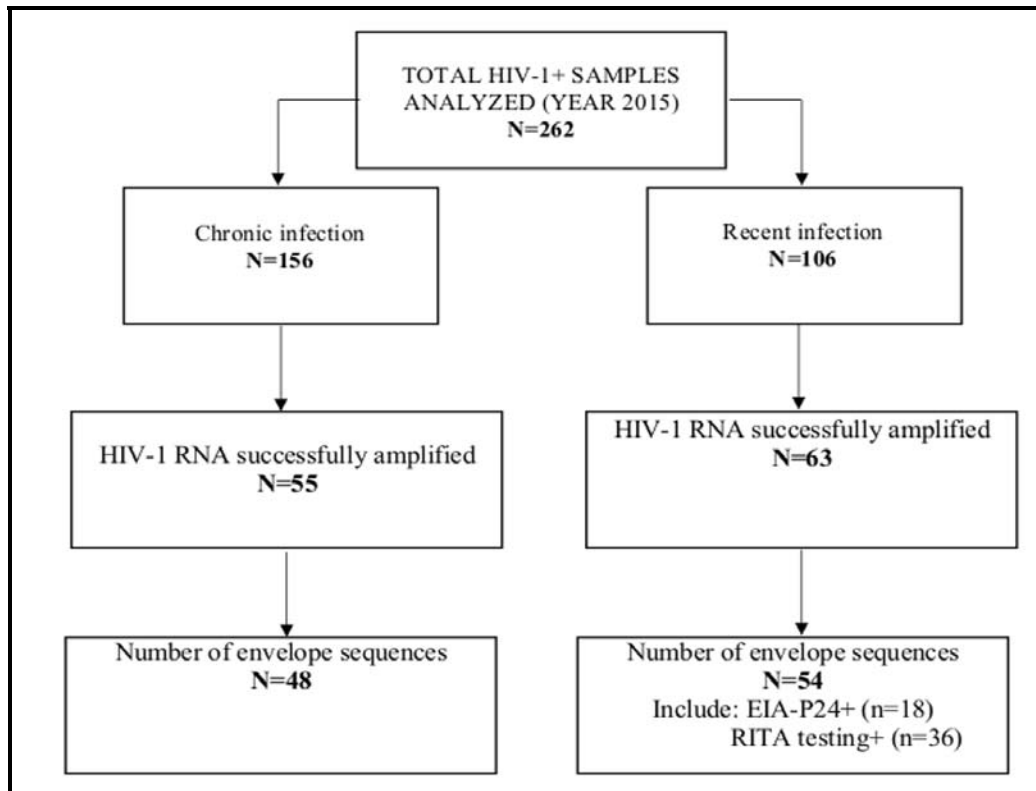


Figure XXXVI. (Article 2, Figure 1). Flowchart of HIV-1 envelope sequences used in this study.

The figure presents the total number of specimens sampled (n=262) and the final sequences obtained after amplification, sequencing and sequence analysis processes (n=102).

The demographic characteristics are described in Table 1. Briefly, 34% of individuals were asymptomatic, whereas 13% of individuals presented acute HIV infection symptoms, and 11% had reached the AIDS stage (Table 1). The mean CD4 count in total population was 375 cell/mm³, range [5-1150]. The mean HIV-1 viral load in total population was 4.94 log₁₀ copies/ml, range [1.60-8.20], and the mean age was 38.8 years, range [18-78]. Of the 102 individuals, 79.4% were infected with HIV-1 subtypes B, and 20.6% with HIV-1 non-B subtypes. Age and baseline CD4+ T-cell counts were significantly different between the recent and chronic infections using the Wilcoxon rank sum test. The mean age of recent infected individuals was 34.15, range [18-78], and for chronic infections, it was 44.03, range [21-58], P=0.0002. The mean baseline CD4+ T-cell count was 521, range [12-1150] in recent infections and 186, range [5-698] in chronic infections.

Tableau IX. (Article 2, Table 1). Population characteristics

Variable characteristics	Recent	Chronic	Total
Gender (No. (% of individuals))			
Female	1 (1.90)	13 (28.20)	14 (14.14)
Male	52 (98.10)	33 (71.80)	85 (85.86)
Risk factor for HIV acquisition (No. (% of individuals))			
MSM	36 (85.71)	19 (48.72)	55 (67.90)
Heterosexual	0	3 (7.69)	3 (3.70)
Relation with at-risk heterosexual	2 (4.76)	2 (5.13)	4 (4.94)
Endemic country	1 (2.38)	13 (33.33)	14 (17.28)
MSM / IDU	3 (7.14)	0	3 (3.70)
IDU	0	2 (5.13)	2 (2.47)
Place of birth (No. (% of individuals))			
Canadian	27 (65.85)	19 (48.72)	46 (57.50)
Caribbean	3 (7.32)	6 (15.38)	9 (11.25)
Europe	5 (12.20)	1 (2.56)	6 (7.50)
North Africa and Middle East	1 (2.44)	0	1 (1.25)
Sub-Saharan Africa	2 (4.83)	10 (25.64)	12 (15)
Latin, Central and South America	4.88	2 (5.13)	4 (5)
Asia	1 (2.44)	0	1 (1.25)
Aboriginal (First Nations)	0	1 (2.56)	1 (1.25)
HIV subtype (No. (% of individuals))			
B	45 (83.30)	36 (75.00)	81 (79.40)
Non-B	9 (17.70)	12 (25.00)	21 (20.60)
Age (No. [mean; range])	31 [18 -78]	46.5 [21-77]	37 [18-78]
15-24 years	12	2	14
25-34 years	23	11	34
35-44 years	6	7	13
45-54 years	7	19	26
55-64 years	4	5	9
>64 years	1	2	3
Baseline CD4+ T-cell count (cells/mm³) No. of individuals [median; range]	42 [458.5; 12-1150]	36 [175; 5-698]	78 [375; 5-1150]
>500	19 [710; 506-1150]	3 [609; 580-698]	22 [700; 506-1150]
351-500	11 [415; 390-490]	7 [420; 360-485]	18 [417.5; 360-490]
201-350	8 [250; 217-320]	5 [302; 220-340]	13 [252; 217-340]
≤ 200	4 [185; 12-200]	21 [59; 5-196]	25 [63; 5-200]
HIV viral load (log₁₀ copy/mL) No. of individuals [median; range]	42 [4.94; 3.43-8.20]	38 [4.76; 1.60 - 6.34]	80 [4.90; 1.60-8.20]
≥ 5	20 [5.53; 5-8.20]	15 [5.60; 5.15 - 6.34]	35 [5.60; 5 - 8.20]
< 5	22 [4.46; 3.43-96]	23 [4.25; 1.60 - 4.98]	45 [4.27; 1.60-4.98]

The table presents the repartition of demographic, epidemiologic, and clinical and risk factors of the study population by infection status (recent or chronic).

HIV-1 partial envelope fragment-length sequence-based clustering using phylogenetic analysis

From the 102 sequences analyzed, four (4) were excluded because of insufficient *env* length. Finally, 98 sequences with satisfactory and similar lengths were included in the phylogenetic analysis. Of these 98 sequences, 57.14% (n=56) did not form clusters, whereas 42.85% (n=42) were part of 16 small clusters ranging from 2 to 5 individuals using the partial *env* fragment length (1070 bp) at a distance cutoff of 0.1 (Figure 2). The HIV-1 subtype B envelope sequence represented 93.3% of clusters identified (n=14/15) and were labelled clust1 to 15, except for cluster 4 that was constituted by HXB2 and BAL reference sequences. The non-B HIV-1 subtypes, which represented 7.14% and 6.7% (n=1/15), which was labelled clust16, as shown in Figure 2. Sequences from recent HIV-1 infected individuals were mostly represented in clusters; 50% were (n=27/54) involved, whereas 31.21% (n=15/48) were comprised of chronic infections (Figure 2). The inclusion or non-inclusion in clusters for acutely and chronically HIV-1 infected individuals envelope sequences was statistically significant using the chi-square test as follows: odds ratio (OR): 0.4, P=0.02.

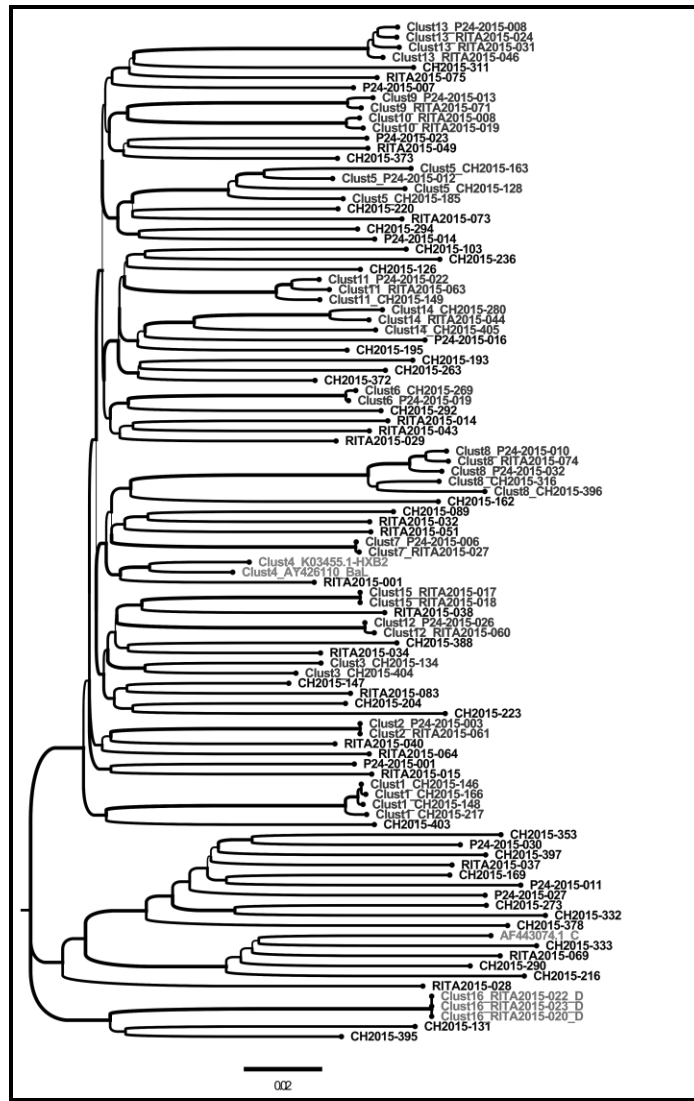


Figure XXXVII. (Article 2, Figure 2). Molecular phylogenetic reconstruction of HIV-1 transmission clusters among newly and chronically HIV-infected individuals in 2015 in Quebec using partial HIV-1 *env*-derived sequences (1070 bp).

Note: The cluster tree contains 101 sequences from HIV-1-infected individuals, including 98 from the study cohort and three (3) reference sequences introduced in the analyses as controls (AF443074.1_C, AY426110_BaL and K03455.1-HXB2). A total of 42 HIV-1 individuals' sequences are grouped into 16 distinct small HIV-1 transmission clusters. These clusters are labelled Clust1 to Clust16 and are depicted using different colors. The cohort-derived clusters are shown in blue for subtype B (Clust1 to Clust15, except Clust4) and orange for subtype D (Clust16). The control sequence names are shown in red (Clust4). Sequences that did not form a cluster are shown in black. Each depicted cluster satisfied $\geq 99\%$ support on bootstrap resampling of 1000 replicates and 10% of pairwise distance 22. Abbreviations: RITA: Recent

infection status determined by recent infection testing algorithm³⁰; p24: recent infection status determined by p24 antigen positivity and CH: chronic, HIV-1 infection sequences.
HIV-1 envelope gp120 loop 3 (V3) sequence-based clustering using phylogenetic analysis

Phylogenetic analyses using only the V3 loop (108 bp)-derived sequences were also performed to identify HIV-1 transmission clusters (TCs). The env loop 3 sequence-based clustering reproduced 66.66% (n=10/15) of the TCs observed with the partial env fragment length (Figure 3). Five clusters labelled (Clust3, 5, 11, 13, 14) shown in Figure 2 were not observed when using the env V3 sequence-based clustering (Figure 3). However, it identified an additional cluster (clust6) (Figure 3) that was not detected by the partial env fragment length analyses.

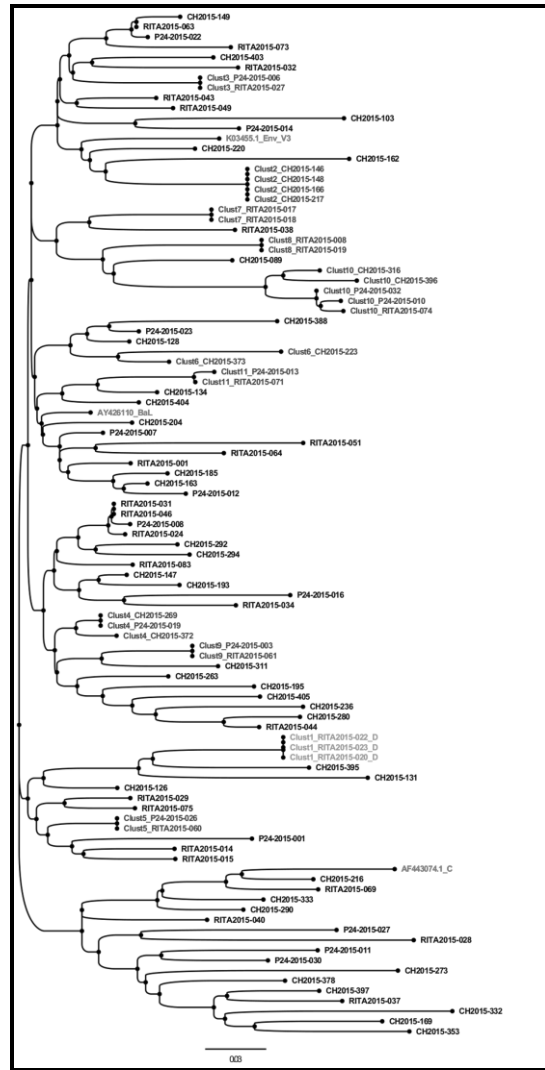


Figure XXXVIII. (Article 2, Figure 3). Molecular phylogenetic reconstruction of HIV-1 transmission clusters among newly and chronically HIV-infected individuals using *env* gp120 V3 loop-derived sequences (108 bp).

Note: The cluster tree contains 101 sequences from HIV-1-infected individual's *env*-V3 sequences, including 98 from the study cohort and three (3) reference sequences introduced in the analyses as controls (AF443074.1_C, AY426110_BaL and K03455.1-HXB2). Each tip represents an individual patient. A total of 31 HIV-1 sequences grouped into 11 distinct small HIV-1 transmission clusters with 2-5 members per cluster. These clusters are labelled Clust1 to Clust11 and are depicted using different colors. Clusters including sequence names are shown in blue for subtype B (Clust2 to Clust11) and orange for subtype D (Clust1). The control sequence names are shown in red. Sequences that did not cluster are shown in black. Each depicted cluster satisfied $\geq 99\%$ support on bootstrap resampling of 1000 replicates and 10% of pairwise distance ²². Abbreviations: RITA: Recent infection status determined by

recent infection testing algorithm³⁰; p24: recent infection status determined p24 antigen positivity and CH: chronically infected sequences. Agreement between the HIV-1 partial envelope fragment lengths of with the V3 loop derived sequences as independent tools to perform HIV-1 transmission clustering.

The *env*-V3 sequence-based clustering reproduced 10 of the 15 clusters previously identified with the *env* 1070 bp sequence length. An additional cluster (C6) was identified only with *env*-V3 loop sequence-based clustering presented in Figure 4.

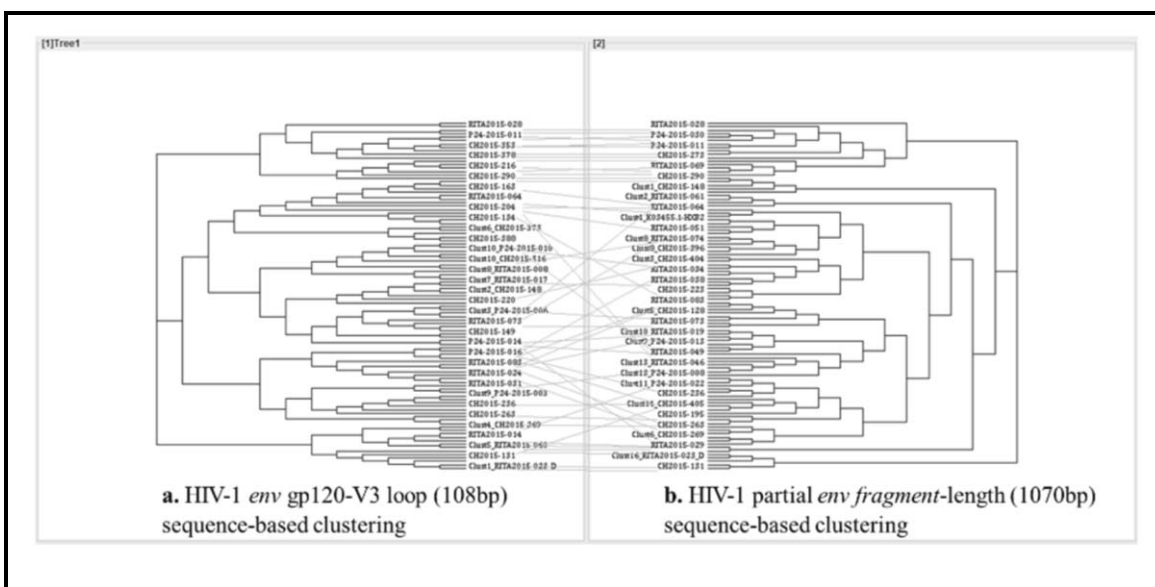


Figure XXXIX. (Article 2, Figure 4). Tanglegrams of rooted phylogenetic trees and networks comparing the HIV-1 envelope gp120 loop 3 (a) with the HIV-1 partial envelope length (b) sequence-based clustering.

Note: The linked clusters and sequences between the two trees are connected by a gray line. The sensitivity, specificity, positive predictive value and negative predictive value of *env* gp120-V3 in determining TCs were 70%, 98.6%, 95.5% and 88.8%, respectively, compared to the partial *env* fragment length (1070 bp) derived sequences as the gold standard (Table 2). The concordance between the gold standard and *env* gp120 V3 cluster determination was assessed by Cohen’s Kappa coefficient and demonstrated a significantly moderate agreement =0.59, $p < 0.00001$ (Table 2).

Tableau X. (Article2, Table 2). Performance of the *env* gp120-V3 sequences in determining HIV transmission clustering compared to the use of the full-*env* fragment.

Accuracy parameters	Value % (n/N)	95% CI
Sensitivity	61.9 (26/42)	45.6 – 76.4
Specificity	95.0 (57/60)	86.1 - 99
Positive predictive value	89.7 (26/29)	72.6 – 97.8
Negative predictive value	78.1 (57/73)	66.9 – 86.9

Note: Table 2 presents the performance of the *env* gp120-V3 sequenced-based clustering (108 bp) using partial HIV-1 envelope fragment length (1070 bp) as the gold standard. The sensitivity and specificity were 61.9% (45.6- 76.4) and 95% (86.1–99), respectively. In addition, the positive predictive value and negative predictive value were 89.7% (72.6 – 97.8) and 78.1% (66.9 – 86.9), respectively. Furthermore, significant and moderate agreement was documented using Cohen’s Kappa coefficient (kappa= 0.59, p <0.00001). Abbreviations: n= number of clusters identified by *env* gp120 V3 loop derived sequences, N= number of clusters identified by the gold standard.

Agreement between the partial HIV-1 envelope fragment sequences length with the gp120 C2V3C3 regions derived sequences as independent tools for identifying HIV-1 transmission clusters.

We also compared the partial *env* fragment length (1070 bp) with the gp120 C2V3C3 regions (HXB2 nt positions 6813 to 7376 ≈564 bp) to improve the moderate sensitivity associated with the V3 sequence-based clustering identification. Surprisingly, the gp120 C2V3C3 regions identified only 6 of the 15 clusters (40%) observed using the partial *env* fragment sequence at pairwise genetic distance (PWD) of ≤ 10%. This unexpected result represented less than the number of clusters identified by V3 loop-derived sequences alone estimated to 66.66% compared to the same gold standard.

Description of factors associated with HIV transmission clustering.

The recently HIV-1-infected subjects defined by RITA 30 represented 53.4% of the TCs, followed by acutely infected individuals (EIA-p24+) at 23.3%, and chronically infected individuals at 23.3% (Figure 1, Table 1). The HIV-1 subtypes B and D comprised 93% and 7% of clusters, respectively. The distribution of the HIV-1 transmission clusters by epidemiologic, clinical and risk factors of acquisition is presented in Table 3.

Tableau XI. (Article 2, Table 3). Summary statistics of the distribution of HIV-1 transmission clusters by factors

Parameters	HIV transmission clustering	
	No (%)	Yes (%)
Age		
<38	23 (39.66)	27 (65.85)
≥38	35 (60.34)	14 (34.15)
Gender		
Male	46 (79.31)	39 (95.12)
Female	12 (20.69)	2 (4.88)
HIV- subtype		
B	42 (70)	39 (92.86)
non-B	18 (30)	3 (7.14)
Ethnicity		
White	34 (85)	25 (83.33)
Black	16 (4)	5 (16.16)
Motivation for HIV testing		
Symptomatic person	25 (50)	16 (51.61)
Screening in an asymptomatic person	17 (24)	13 (41.94)
Confirmation of previous positive test	8 (16)	2 (6.45)
HIV-1 clinical status		
Asymptomatic	19 (38)	15 (48.39)
Acute infection	18 (36)	12 (38.71)
Symptomatic infection + AIDS	13(26)	4 (12.90)
HIV transmission risk factors		
MSM	32 (64)	23 (74.19)
MSM/IDU+IDU	3 (6)	2 (6.45)
Heterosexual	15(30)	6 (19.35)
Log¹⁰ HIV viral load		
<5	29 (59.18)	15 (48.39)
≥5	20 (40.82)	16 (51.61)
CD4 count value		
<350	24 (51.06)	14 (45.16)
≥350	23 (48.94)	17 (54.84)
HIV infection status by test		
Chronic infection by RITA testing ³⁰	33 (55)	15 (35.71)
Acute/early infection by EIA-p24 testing	9 (15)	10 (23.81)
Recent by RITA testing ³⁰	18 (30)	17 (40.48)
HIV-1 infection status		
Recent (acute/early recent by RITA)	27 (45)	27 (64.29)
Chronic	33 (55)	15 (35.71)

Table 3 presents the summary statistics of the demographic, epidemiologic, clinical and risk factors of HIV-1 infected individual's associate with inclusion in or not in clusters. The mean age of HIV-1 infected individuals included in clusters was 33.34 years [18-58], the mean CD4 count was 390.22 cell/mm³ [10-871], and the mean HIV-1 viral load was 5.12

log₁₀ copies/mL [1.60-8.20]. Mean age (<38.8 versus ≥38.8 years) and HIV-1 subtype (B versus non-B) were identified as two factors significantly associated with HIV-1 transmission clusters: mean age (OR= 0.25, 95% CI [0.04-0.66], P=0.002) and HIV-1 subtype (OR= 0.17, 95% CI [0.10-0.61], P=0.011) (Table 4).

Tableau XII. (Article 2, Table 4) Factors associated with HIV-1 transmission clustering using logistic regression analysis.

Factors	Bivariate analysis		Multivariate analysis		
	N	p	OR	95% CI	P
Ethnicity					
Black	21	0.74			NS
Nonblack	60	Ref.			
Gender					
Male	85	0.041			NS
Female	14	Ref			
Age (mean)					S
< 38.8 years	52	0.001	0.25	0.04-0.66	0.002
≥ 38.8 years	47	Ref.			
HIV transmission route					NS
MSM intercourse	55	0.18			
Heterosexual intercourse	21	0.83			
IDU & MSM / IDU intercourse	5	Ref.			
Motivation for HIV testing					NS
Screening in an asymptomatic person	30	0.20			
Symptomatic person	41	0.27			
Confirmation of previous positive test	10	Ref.			
Clinical status					NS
Acute infections	13	0.94			
Chronic infection symptoms	6	0.18			
Nonspecific diseases and symptoms	17	0.38			
AIDS	11	0.23			
Asymptomatic	34	Ref.			
HIV subtype					S
B	81	0.009	0.17	0.10-0.61	0.011
Non-B	21	Ref.			
HIV-1 infection status					NS
Recent by RITA (293) + recent by EIA-p24+	54	0.057			
Chronic by RITA (293)	48	Ref			
CD4 cell count (cells/mm³)					NS
≥500	22	0.895			
<500	56	Ref			
HIV-1 viral load (copies/mL)					NS
≥100000	36	0.345			
<100000	44	Ref.			
HIV-1 viral load (copies/mL)					NS
≥1500	77	0.337			
<1500	3	Ref			
VL_log10					NS
>5	36	0.345			
<5	44	ref			

Table 4 presents results of logistic regression analysis including: bivariate analysis shown that the HIV-1 subtype, the gender and the age means of infected individuals were statistically significant ($p < 0.05$). However, all factors with a p value ≤ 0.25 were subjected to multivariate analysis. For multivariate analysis, HIV-1 subtypes (B versus non-B), OR=0.25[0.04-0.66], $P=0.002$ and patient age mean (< 38.8 years compared to those aged ≥ 38.8 years), OR=0.17 [0.10-0.61], $P=0.011$, constitute two factors significantly associated with HIV-1 transmission clustering in this study population. Footnote Ref: variable of comparison. The statistically significant variables were indicated by their P values. Abbreviation: NS: statistically nonsignificant, S: statistically significant.

4. DISCUSSIONS

HIV-1 envelope sequence inference was used to reveal HIV-1 transmission clustering (network) among newly diagnosed individuals in 2015 in Quebec, Canada. The first analyses used a partial *env* fragment length (1070 bp) that identified 15 small transmission clusters (TCs), including 42 HIV-infected individuals. Using only the *env* gp120 V3 sequences (108 bp), we identified 11 small clusters, thus reproduced 66.7% (10/15) of clusters that were previously detected by the partial-*env* fragment sequence length which has been considered as the gold standard. We used the envelope loop 3 (V3) fragment as comparative because it is the most conserved of the HIV-1 *env* hypervariable regions, and its sequences are frequently used to predict coreceptor tropism 45-51 and may be available for intention-to-treat analysis using entry inhibitors in clinics.

The agreement between these two approaches was moderate (Cohen's Kappa coefficient (κ) =0.59). We tested different cutoff values of V3 sequences (1 to 20%) compared to partial-*env* fragment sequences, but it did not enhance the accuracy of the agreement of the approaches. This observation underlies and confirms that the length of HIV-1 genome sequences analyzed inflects on TC determination ²⁷. Therefore, using only *env* V3 loop sequences may underestimate HIV-1 transmission clusters.

Although, we included two constant regions of the gp120 (C2, C3) with the V3 loop and compared to HIV-1 partial *env* fragment in order to hope increasing degree of sensitivity of the number of clusters estimate. Unexpectedly, sequence derived from the gp120_C2-V3-C3 regions identified only 40% of the clusters, which is lesser than the number of clusters identified by the sequences derived from the V3 loop alone (66.66%) compared to the HIV-1 partial *env* fragment. The three regions decreased the sensitivity associated with the use of a larger fragment comprising 2 constant regions and the V3 loop. It may be that not all have a good degree of conservation of the nucleotide sequences. The nucleotides sequences diversity introduced by the two constant regions (C2 and C3) may have contributed to increasing the genetic distance threshold between individuals up to (> 10-15%) contributing to reduce the sensitivity of the estimated number of clusters. We believe that increasing the

length of the sequence up to 524 bp, including only constant regions (C2, C3 with V3), may not increase the degree of sensitivity of cluster estimates compared partial *env* fragment length.

However, the small number of *env* sequence datasets used in this study did not yield formal conclusions. Further studies including a large *env* sequence dataset and the combination of different *env* segments can help confirm the current result.

We have proceeded to new phylogenetic analyses concerning the gp120 C2V3C3 clustering and reached the same conclusions. We will consider in the future large sequence dataset and may evaluate different segments of the envelope sequence.

However, in screening for public health surveillance purposes, it may be useful as V3 loop sequencing techniques are routinely performed in clinical laboratories to inform virus tropism and the use of CCR5 inhibitors in treatment. A recent study has shown that the V3 loop is one of most predictable segments of the HIV-1 envelope and that using its sequences contributes to estimating the recency of an infection ³¹. Hence, it may also be useful for real-time HIV-1 transmission cluster detection in complement to existing methods.

Previous studies have already demonstrated that the near full-length genome (FG) sequences (9719 bp) of HIV-1 are the best tool for estimating the extent of HIV clusters ^{27,52}. However, the methodological and cost constraints associated with full genome sequencing preclude its use in routine clinical epidemiology studies. The estimate of HIV TCs (size and extent) not only depends on the selected gene and sequence length ²². It may negatively (underestimation) or positively (overestimation) affect the following: the cutoff value defined as genetic distance ²², the type of sequencing (Sanger or next generation sequencing), the number of sequences in datasets, the number of variable sites, or the period covered by the study ^{4,27}, the timely availability of sequence data ⁵³ and the molecular phylogenetic methods used for transmission chain defining ⁵⁴⁻⁵⁶. This study covered a one-year period (2015) that may underestimate transmission events occurring outside of this period. If sequencing is performed on a regular basis, it may help track the early founder or nascent clusters before its growth that will be identified in the following periods and help adapt prevention strategies for at-risk populations. Considering CDC Guide, June 2018, the identification of

most linked sequences over a short period of time could possibly indicate that transmission occurs rapidly within a group ⁵⁷.

The cutoff value of the pairwise genetic distance (PWD) used for this study was $\leq 10\%$, as proposed by Novitsky et al. ²². Because there is no gold standard concerning HIV envelope sequence-based clustering, we considered this cutoff value in reference to Novitsky et al. ²² studies. This is a most recent and was used PWD method to defining clustering of HIV-1 envelope sequence that was more adapted to the current study than others. Using a small cutoff value may underestimate cluster sizes and, in other hand, using a large cutoff value may overestimate them. Therefore, the cutoff value has to be well defined ^{22, 58-61} according to the Novitsky et al. ²² studies. Especially, in considering the specificity of each of the different

HIV-1 subtypes and CRFs, it may also be important to determine the best cutoff value when using the full-length HIV-1 envelope fragment (GP160) or its specific subregions or domains (GP120, GP41). The HIV-1 genome regions (*gag*, *pol*, or *env*) source of sequences for cluster estimates and establishment of correlation between them may also be considered. Further studies that may also evaluate the performance of using the near full-length HIV-1 genome and its three regions (*gag*, *pol*, or *env*) adjusted by subtypes as independent tools for determining HIV-1 transmission clusters is also encouraged using large sequence datasets.

Compared to earlier studies conducted by Brenner et al. ^{8,62,63} in Québec for longer periods of time and using a large sequence dataset of the HIV-1 *pol* region (n=1277), 30 large TCs (20+/cluster) occurring over 13 years (2002 to 2015) ⁶³ were identified. Lubelchek et al. also conducted a similar study in Chicago using a large dataset of *pol* gene sequences and identified a single large transmission cluster in a total of 26 clusters observed ¹¹. Many HIV-1 transmission cluster studies have used *pol* gene sequences as these sequences are often available from clinical laboratories performing drug resistance testing ⁵³. These findings contrast those of the current study, where no large cluster has been identified. The reason for the difference may be due to the limited *env* sequence data used in this study of analyzed (n= 102) and the restricted period (one year) ^{8, 62, 63 11}. Nevertheless, for this one-year period, we were able to identify and highlight the existence of HIV-1 clustering between newly HIV-

1-diagnosed patients. The current study aimed to track the nascent or forming clusters in real time to help adapt earlier prevention strategies that may limit the formation of large clusters. Bivariate analysis identified that newly HIV-1-infected individuals are significantly more associated with HIV-1 transmission clustering than chronically infected individuals, OR=2.20, P=0.04 using Fisher's exact test. This observation is in agreement with results presented by Brenner et al. ⁶³, Denis et al. ^{53,64} and Miller et al. ⁶⁵. The results may be explained by the extreme fitness early founder viruses that successfully establish infections ⁶⁶⁻⁶⁸ and the higher viral load during acute infection.

Multivariate analyzes identified two factors that were likely associated with transmission clusters: the mean age (<38.8 years) and the HIV-1 B subtype. Individuals aged less than 38.8 years were more likely to form TCs than those aged over 38.8 years, OR= 3.01, p=0.001 (bivariate) and OR=0.25, p=0.002 (multivariate) analysis. This result reflects the profile of the HIV epidemics in Quebec, Canada, where individuals aged between 30 and 39 years accounted for 28% of reported cases in 2016 ⁶⁹. Individuals infected by the HIV-1 B subtype were more likely to form clusters (OR= 0.17) than non-B subtypes. OR=5.57, p=0.009 (bivariate) and (OR=.0.17, p=0.01 (multivariate) analysis). This finding also reflects the most prevalent subtype circulating in Quebec. The HIV-infected MSM population is generally infected by HIV-1 B viruses and represents a large proportion of new cases in Quebec (67.90%) ⁶⁹.

The objective of the current study was to evaluate tools to detect newly HIV-infected individuals who have the potential to transmit and sustain HIV epidemics as early as possible. Newly HIV-infected individuals are generally unaware of their infection status ^{70,71} and therefore can contribute to the spread of infection. ^{65,72} High viral loads observed during acute infections ⁷³⁻⁷⁵ also constitute a factor that enhances HIV transmission of new HIV-infected individuals. HIV-1 viral load (> 10 000 copies/ml) and CD4 counts >350 generally constitute biological factors significantly associated with HIV-1 transmission clustering ⁶¹. Our study did not find a significant association of viral load or CD4 counts with clustering. The difference may be due to the fact that only 23% of clusters occurred in individuals with acute infection; therefore, all other clusters occurred in individuals with dampened viral load

(chronic or recent infection < 6 months). The timing of viral load assessment after diagnostic testing might also be a factor. Bivariate analysis identified gender (male) as significantly associated with transmission clusters, OR=5.08, p=0.02. This reflects the HIV epidemic profile in Québec, with MSM being the highest at-risk group ¹. In contrast to studies conducted in the US by Lubelchek et al. ¹¹ and Paz-Bailey et al.⁷⁶, black race was not significantly associated with clusters in the current study. This finding reflects the demographics of the HIV epidemic of Quebec, where black people are not overrepresented ^{69,1}. In general, the results of this current study suggest the importance of real-time follow-up of transmission networks regarding the early clusters of transmission established in a single year from new HIV-infected individuals. The short-term assessment of early cluster building may help improve quick responses to prevent HIV transmission by identifying the nascent or forming clusters and populations concerned with at high risk.

5. CONCLUSIONS

Our results confirm the presence of small HIV transmission clusters among HIV-1-infected individuals in Quebec. HIV subtype B-infected individuals and individuals less than 38.8 years of age are two factors significantly associated with HIV transmission clustering. The HIV-1 partial *env* fragment lengths derived sequences were able to detect clusters of transmission contributing to the persistence of the HIV epidemic in Quebec. Although, less sensitive than the sequencing of partial *env* fragments, the V3-derived sequences were able to identify HIV transmission clusters with moderate agreement. The latter tools (V3 sequence) may be useful for short-term assessment of nascent HIV-1 transmission clustering in support of the existing methods in screening purposes.

Competing interests

The authors have no competing interests to declare.

Acknowledgments

We are grateful to all the technicians of the molecular biology, serodiagnosis and virology departments at the provincial public health reference laboratory (LSPQ), Montreal, Canada, for their technical support.

Special thanks to Lyne Désautels, Sylvie-Nancy Beaulac and Lise Leblanc. We are grateful to Diane Sylvain and Maureen Hastie of the provincial HIV surveillance program of Quebec at the “Institut national de santé publique du Québec”, Quebec, Canada, for providing patients’ epidemiological and clinical data.

We thank Dr Art FY. Poon, PhD. Assistant professor and researcher in viral evolution at the Department of Pathology and Laboratory Medicine, University of Western Ontario, Canada, and Dr Paul Wankah, MD, at the Université de Sherbrooke, Québec, Canada, for revising this manuscript. We also thank Dr Céline Scornavacca, researcher-CNRS, Institut des Sciences de l’Evolution-CC64, Université Montpellier II, France, for advices and help for bioinformatics analysis.

We thank Dr. Réjean Dion from the LSPQ, Dr. Drissa Sia from the “Université du Québec en Outaouais, UQO”, Quebec, Canada and, Adama Kazienga, Hasselt University, Faculty of Science, Belgium, for their statistical analysis support.

We thank Drs. Andrés Finzi, Petronela Ancuta, and Nicolas Chomont of the Hospital Research Center of the University of Montreal (CRCHUM) for their advisement on study conception. Special thanks to the HIV-1-infected patients of Quebec from whom serum samples were used in the present study.

Sequences data

All HIV-1 envelope nucleotide sequences for partial fragment sequences were deposited and are available in the GenBank Sequence Database (NCBI) under accession numbers KY946519 to KY946713.

Funding

This study was supported by the Islamic Development Bank, the “Fonds de recherche du Québec – Santé” (FRQS), Réseau SIDA/Maladies infectieuses, Québec, Canada, and the Genome Canada Grant.

Alexis Kafando, PhD student, is the beneficiary of: 1) Islamic Development Bank Merit Scholarship Programme for High Technology, for 3 Year PhD (2013-2016), ID: 600014438, Jeddah, Saudi Arabia; 2) Bourse d'exemption des droits de scolarité supplémentaires pour étudiants étrangers of the Université de Montréal, Montréal, Québec, Canada; 3) Bourse de fin d'études doctorales of the Faculté des Études supérieures et postdoctorales (FESP) of the Université de Montréal, Québec, Canada; 4) Bourse d'étude of Dre Tremblay's laboratory at the Centre Hospitalier de l'Université de Montréal (CRCHUM). The laboratory is funded by the FRQS RESEAU SIDA, Québec, Canada. 5) Alexis Kafando is a beneficiary of the « Programme de prêts et bourses du ministère de l'Éducation et de l'enseignement supérieur du gouvernement du Québec », Canada and; 6) He is public officer of the Centre Muraz biomedical research center, ministry of health, Burkina Faso.

Cécile L. Tremblay, head of this study, is Director of the Pfizer/University of Montreal, Chair on Translational HIV Research and a scholar from the “Fonds de recherche du Québec - santé (FRQS), Québec, Canada.

Authors’ contributions.

- 1- CLT, HC, AK, BS, CM, FDB, EF, AC, and MEF: contribute to the conception and design of this study.
- 2- AK, EF, BS, CM, FDB, MS, AC, HC contribute to acquisition, analysis, and interpretation of data.
- 3- AK, EF, MNS, BS, CM, FDB, HC: contribute to the bioinformatics and statistical analysis of data
- 4- All authors contributed to drafting the work and revising it critically for important intellectual content and provided final approval of the version to be published.

6. REFERENCES

1. Raphaël Bitera MF, Michel Alary, Cécile Tremblay, Raymond Parent, Diane Sylvain, Maureen Hastie. Programme de surveillance de l'infection par le virus de l'immunodéficience humaine (VIH) au Québec - Rapport annuel 2015. In: Québec Indspd, <http://www.inspq.qc.ca>. eds. Quebec: gouvernement du Québec (2016); 2016.
2. TA. B. Molecular Phylogenetics. Vol Chapter 16. 2nd ed. ed. England: Oxford: New York: Wiley-Liss, 2002. 2002.
3. Chan PA, Hogan JW, Huang A, et al. Phylogenetic Investigation of a Statewide HIV-1 Epidemic Reveals Ongoing and Active Transmission Networks Among Men Who Have Sex with Men. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2015;70(4):428-435.
4. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjornsson J. Defining HIV-1 transmission clusters based on sequence data. *Aids*. 2017;31(9):1211-1222.
5. English S, Katzourakis A, Bonsall D, et al. Phylogenetic analysis consistent with a clinical history of sexual transmission of HIV-1 from a single donor reveals transmission of highly distinct variants. *Retrovirology*. 2011;8.
6. Brenner BG, Wainberg MA. Future of Phylogeny in HIV Prevention. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2013;63:S248-S254.
7. Brenner B, Wainberg MA, Roger M. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. *AIDS*. 2013;27(7):1045-1057.
8. Brenner BG, Roger M, Routy JP, et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis*. 2007;195(7):951-959.

9. Podsiadlo L, Polz-Dacewicz M. Molecular evolution and phylogenetic implications in clinical research. *Ann Agric Environ Med*. 2013;20(3):455-459.
10. Resik S, Lemey P, Ping LH, et al. Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. *AIDS Res Hum Retroviruses*. 2007;23(3):347-356.
11. Lubelchek RJ, Hoehnen SC, Hotton AL, Kincaid SL, Barker DE, French AL. Transmission Clustering Among Newly Diagnosed HIV Patients in Chicago, 2008 to 2011: Using Phylogenetics to Expand Knowledge of Regional HIV Transmission Patterns. *J Acquir Immune Defic Syndr*. 2015;68(1):46-54.
12. Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science*. 1998;280(5371):1884-1888.
13. Moulard M, Hallenberger S, Garten W, Klenk HD. Processing and routage of HIV glycoproteins by furin to the cell surface. *Virus Res*. 1999;60(1):55-65.
14. Haim H, Salas I, Sodroski J. Proteolytic Processing of the Human Immunodeficiency Virus Envelope Glycoprotein Precursor Decreases Conformational Flexibility. *J Virol*. 2013;87(3):1884-1889.
15. Hartley O, Klasse PJ, Sattentau QJ, Moore JP. V3: HIV's switch-hitter. *AIDS Res Hum Retroviruses*. 2005;21(2):171-189.
16. Checkley MA, Luttge BG, Freed EO. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J Mol Biol*. 2011;410(4):582-608.
17. Starcich BR, Hahn BH, Shaw GM, et al. identification and characterization of conserved and variable regions in the envelope gene of HTLV-III LAV, the retrovirus of AIDS. *Cell*. 1986;45(5):637-648.
18. Freed EO. HIV-1 replication. *Somatic cell and molecular genetics*. 2001;26(1-6):13-33.

19. Yu F, Lu L, Du LY, Zhu XJ, Debnath AK, Jiang SB. Approaches for Identification of HIV-1 Entry Inhibitors Targeting gp41 Pocket. *Viruses*. 2013;5(1):127-149.
20. Douglas NW, Munro GH, Daniels RS. HIV/SIV glycoproteins: Structure-function relationships. *Journal of Molecular Biology*. 1997;273(1):122-149.
21. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS*. 2017;31(9):1211-1222.
22. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. Impact of Sampling Density on the Extent of HIV Clustering. *Aids Research and Human Retroviruses*. 2014;30(12):1226-1235.
23. Novitsky V, Wang R, Rossenkhan R, Moyo S, Essex M. Intra-host evolutionary rates in HIV-1C *env* and *gag* during primary infection. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2013;19:361-368.
24. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(21):7552-7557.
25. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent Viral Evolutionary Changes Associated with the Progression of Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology*. 1999;73(12):10489-10502.
26. Herbeck JT, Nickle DC, Learn GH, et al. Human Immunodeficiency Virus Type 1 *env* Evolves toward Ancestral States upon Transmission to a New Host. *Journal of Virology*. 2006;80(4):1637-1644.

27. Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. Importance of Viral Sequence Length and Number of Variable and Informative Sites in Analysis of HIV Clustering. *Aids Research and Human Retroviruses*. 2015;31(5):531-542.
28. McClure P, Curran R, Boneham S, Ball JK. A polymerase chain reaction method for the amplification of full-length envelope genes of HIV-1 from DNA samples containing single molecules of HIV-1 provirus. *Journal of Virological Methods*. 2000;88(1):73-80.
29. Cornelissen M, Gall A, Vink M, et al. From clinical samples to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing. *Virus Research*. 2017;239:10-16.
30. Serhir B, Hamel D, Doualla-Bell F, et al. Performance of Bio-Rad and Limiting Antigen Avidity Assays in Detecting Recent HIV Infections Using the Quebec Primary HIV-1 Infection Cohort. *Plos One*. 2016;11(5).
31. Kafando A, Fournier E, Serhir B, et al. HIV-1 envelope sequence-based diversity measures for identifying recent infections. *Plos One*. 2017;12(12).
32. Asin-Milan O, Wei Y, Sylla M, Vaisheva F, Chamberland A, Tremblay CL. Performance of a clonal-based HIV-1 tropism phenotypic assay. *J Virol Methods*. 2014;204:53-61.
33. Sanders-Buell E, Saliminen, S.O., McCutchan, F.E., 1995. In: . Sequencing primers for HIV1. Los Alamos National Laboratory. 1995;pp. 1521–1527.
34. Cleghorn FR, Jack N, Carr JK, et al. A distinctive clade BHIV type 1 is heterosexually transmitted in Trinidad and Tobago. *Proc Natl Acad Sci U S A*. 2000;97(19):10532-10537.
35. Hunt M, Gall A, Ong SH, et al. IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics*. 2015;31(14):2374-2376.

36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016;33(7):1870-1874.
37. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013;30(4):772-780.
38. Ragonnet-Cronin M, Hodcroft E, Hue S, et al. Automated analysis of phylogenetic clusters. *BMC Bioinform*. 2013;14:9.
39. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology*. 2012;61(6):1061-1067.
40. Scornavacca C, Zickmann F, Huson DH. Tanglegrams for rooted phylogenetic trees and networks. *Bioinformatics*. 2011;27(13):i248-i256.
41. Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T. A web-based genotyping resource for viral sequences. *Nucleic Acids Res*. 2004;32:W654-W659.
42. Pineda-Pena AC, Faria NR, Imbrechts S, 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. *Infect Genet Evol*. 2013;19:337-348.
43. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ : British Medical Journal*. 1995;311(7003):485-485.
44. Berger JO, Delampady M. Testing Precise Hypotheses. *Statistical Science*. 1987;2(3):317-335.
45. Berger EA, Doms RW, Fenyö EM, et al. A new classification for HIV-1. *Nature*. 1998;391:240.

46. Fätkenheuer G, Pozniak AL, Johnson MA, et al. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. *Nature Medicine*. 2005;11:1170.
47. Gulick RM, Su Z, Flexner C, et al. Phase 2 study of the safety and efficacy of vicriviroc, a CCR5 inhibitor, in HIV-1-Infected, treatment-experienced patients: AIDS clinical trials group 5211. *The Journal of infectious diseases*. 2007;196(2):304-312.
48. Clotet B. CCR5 Inhibitors: Promising yet Challenging. *The Journal of infectious diseases*. 2007;196(2):178-180.
49. Dorr P, Westby M, Dobbs S, et al. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial agents and chemotherapy*. 2005;49(11):4721-4732.
50. Heger E, Kaiser R, Knops E, et al. Results of the first international HIV-1 coreceptor proficiency panel test. *Journal of Clinical Virology*. 2017;93:53-56.
51. Montagna C, De Crignis E, Bon I, et al. V3 Net Charge: Additional Tool in HIV-1 Tropism Prediction. *Aids Res Hum Retrovir*. 2014;30(12):1203-1212.
52. Yebra G, Hodcroft EB, Ragonnet-Cronin ML, et al. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. *Scientific Reports*. 2016;6.
53. Dennis AM, Herbeck JT, Brown AL, et al. Phylogenetic Studies of Transmission Dynamics in Generalized HIV Epidemics: An Essential Tool Where the Burden is Greatest? *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2014;67(2):181-195.

54. Villandré L, Labbe A, Brenner B, Ibanescu R-I, Roger M, Stephens DA. Assessing the role of transmission chains in the spread of HIV-1 among men who have sex with men in Quebec, Canada. *PLoS One*. 2019;14(3):e0213366.
55. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS (London, England)*. 2017;31(9):1211-1222.
56. Rose R, Lamers SL, Dollar JJ, et al. Identifying Transmission Clusters with Cluster Picker and HIV-TRACE. *AIDS Res Hum Retroviruses*. 2017;33(3):211-218.
57. (CDC) CfDCaP. HIV Cluster and Outbreak Detection and Response. Fact Sheet: HIV Molecular Cluster Detection and Response In: Division of HIV/AIDS Prevention NCFHA, Viral Hepatitis, STD, and TB Prevention,, ed. Vol DRAFT Version 2.04 JUNE 2018.
58. Brenner BG, Roger M, Routy JP, et al. High rates of forward transmission events after acute/early HIV-1 infection. *The Journal of infectious diseases*. 2007;195(7):951-959.
59. Brenner BG, Roger M, Moisi DD, et al. Transmission networks of drug resistance acquired in primary/early stage HIV infection. *AIDS (London, England)*. 2008;22(18):2509-2515.
60. English S, Katzourakis A, Bonsall D, et al. Phylogenetic analysis consistent with a clinical history of sexual transmission of HIV-1 from a single donor reveals transmission of highly distinct variants. *Retrovirology*. 2011;8:54-54.
61. Aldous JL, Pond SK, Poon A, et al. Characterizing HIV transmission networks across the United States. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2012;55(8):1135-1143.
62. Brenner BG, Roger M, Stephens D, et al. Transmission Clustering Drives the Onward Spread of the HIV Epidemic Among Men Who Have Sex With Men in Quebec. *Journal of Infectious Diseases*. 2011;204(7):1115-1119.

63. Brenner BG, Ibanescu RI, Hardy I, et al. Large cluster outbreaks sustain the HIV epidemic among MSM in Quebec. *AIDS*. 2017;31(5):707-717.
64. Dennis AM, Hue S, Hurt CB, et al. Phylogenetic insights into regional HIV transmission. *Aids*. 2012;26(14):1813-1822.
65. Miller WC, Rosenberg NE, Rutstein SE, Powers KA. Role of acute and early HIV infection in the sexual transmission of HIV. *Current Opinion in Hiv and Aids*. 2010;5(4):277-282.
66. Lythgoe KA, Fraser C. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proceedings of the Royal Society B-Biological Sciences*. 2012;279(1741):3367-3375.
67. Alizon S, Fraser C. Within-host and between-host evolutionary rates across the HIV-1 genome. *Retrovirology*. 2013;10.
68. Parrish NF, Gao F, Li H, et al. Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(17):6626-6633.
69. Bourgeois AC EM, Awan A, Jonah L, Varsaneux O, Siu W. HIV in Canada—Surveillance Report, 2016. *Can Commun Dis Rep*. 2017;43(12):55.
70. Daar ES, Pilcher CD, Hecht FM. Clinical presentation and diagnosis of primary HIV-1 infection. *Current Opinion in Hiv and Aids*. 2008;3(1):10-15.
71. Weinhardt LS, Kelly JA, Brondino MJ, et al. HIV transmission risk behavior among men and women living with HIV in 4 cities in the United States. *Jaids-Journal of Acquired Immune Deficiency Syndromes*. 2004;36(5):1057-1066.
72. Vergis EN, Mellors JW. Natural history of HIV-1 infection. *Infectious Disease Clinics of North America*. 2000;14(4):809-+.

73. Anderson RM, May RM. EPIDEMIOLOGICAL PARAMETERS OF HIV TRANSMISSION. *Nature*. 1988;333(6173):514-519.
74. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. *Journal of Infectious Diseases*. 2008;198(5):687-693.
75. Wawer MJ, Gray RH, Sewankambo NK, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *Journal of Infectious Diseases*. 2005;191(9):1403-1409.
76. Paz-Bailey G, Smith A, Masciotra S, et al. Early HIV Infections Among Men Who Have Sex with Men in Five Cities in the United States. *Aids and Behavior*. 2015;19(12):2304-2310.

Address correspondence to:

Dre Cécile L. Tremblay, MD, FRCPC

Département de Microbiologie, infectiologie et immunologie

Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM)

[REDACTED]

[REDACTED]

Tel.: [REDACTED]

FAX: [REDACTED]

E-mail: [REDACTED]

E-mail addresses of co-authors:

AK: [REDACTED]

BS: [REDACTED]

FDB: [REDACTED]

EF: [REDACTED]

MNS: [REDACTED]

CM: [REDACTED]

MS: [REDACTED]

AC: [REDACTED]

MEF: [REDACTED]

HC: [REDACTED]

CLT: [REDACTED]

ARTICLE 3

HIV-1 ENVELOPE GLYCOPROTEIN AMINO ACIDS SIGNATURES ASSOCIATED WITH CLADE B TRANSMITTED/FOUNDER AND RECENT VIRUSES (668).

Authors

Alexis Kafando,¹ Christine Martineau,² Mohamed El-Far,³ Eric Fournier,² Florence Doualla-Bell,² Bouchra Serhir,² Adama Kazienga,⁴ Mohamed Ndongo Sangaré,⁵ Mohamed Sylla,³ Annie Chamberland,³ Hugues Charest,^{1,2} and Cécile L. Tremblay^{1,2,3}

Publié dans le journal: Viruses. 2019 Nov 1;11(11):1012. doi: 10.3390/v11111012.

¹ Département de microbiologie, infectiologie et immunologie, Faculté de médecine, Université de Montréal, Montréal, Québec QC [REDACTED] [REDACTED] Canada; [REDACTED] (A.K.); [REDACTED] (H.C.)

² Laboratoire de santé publique du Québec, Institut national de santé publique du Québec, Sainte-Anne-de-Bellevue, Québec QC [REDACTED], Canada; [REDACTED] (C.M.); [REDACTED] (E.F.); [REDACTED] (F.D.-B.); [REDACTED] (B.S.). ³ Centre de recherche du centre hospitalier de

l'Université de Montréal, Montréal, Québec QC [REDACTED] [REDACTED] Canada; [REDACTED] (M.E.-F.); [REDACTED] (M.S.); [REDACTED] (A.C.), ⁴ Faculty of Science, Hasselt University, [REDACTED]

Hasselt, Belgium; [REDACTED]. ⁵ Département de médecine sociale et préventive, École de santé publique, Université de Montréal, Montréal QC [REDACTED] Québec, Canada; [REDACTED]. * Correspondence: Dre Cécile L. Tremblay, MD, FRCPC.

Département de Microbiologie, infectiologie et immunologie, Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM), [REDACTED]

[REDACTED]

[REDACTED] E-mail: [REDACTED]

CONTRIBUTION DE M KAFANDO ALEXIS À L'ARTICLE 3

M. Kafando Alexis a contribué à l'article 3 pour la conception de l'étude selon les directives du Dre Cécile Tremblay. Il a passé en revue les procédures expérimentales de l'étude sous la supervision du Dr Hugues Charest, Dre Christine Martineau et du Dre Annie Chamberland. M Kafando a réalisé toutes les expériences de laboratoire passant par l'identification des échantillons disponibles dans la serothèque du Laboratoire de santé publique du Québec (sous le contrôle du Dre Bouchra Serhir), à l'extraction des ARN du VIH, aux analyses de biologie moléculaire (PCR), aux séquençages (NGS) et à l'assemblage des séquences. M Kafando a réalisé les analyses bio-informatiques et une partie des analyses statistiques (supervision Dr Mohamed Ndong Sangaré et de M Adama Kazienga). Il a également assuré la synthèse de tous les résultats de l'étude et rédigé le premier manuscrit qu'il a soumis à l'ensemble de l'équipe de recherche pour les révisions et commentaires. Il a par la suite rédigé le manuscrit final, identifié le journal en concertation avec l'équipe et procédé à sa soumission. Il a également préparé les séquences de l'enveloppe du VIH-1 de tous les sujets de l'étude qu'il a soumis dans la base de données de Genbank en collaboration avec Éric Fournier.

M Kafando a interagi avec les éditeurs du journal et les reviseurs de l'article et a apporté toutes les corrections nécessaires au manuscrit en collaboration avec les coauteurs. M Kafando a rédigé la version finale du manuscrit soumise au journal qui a été publiée.

En somme, M. Kafando a participé à la conception de l'étude, à la réalisation des expériences de laboratoire et a été l'interface entre les différents coauteurs, les éditeurs et les reviseurs pour la rédaction du manuscrit.

Les contributions de chacun et de chacune des coauteurs sont décrites plus en détail dans l'article.

RÉSUMÉ

Contexte : Les variant transmis/fondateurs (TF) du VIH-1 sont des virus sélectionnés pendant la phase aiguë parmi une multitude de virions exposés ou présents au cours de la transmission chez l'individu exposé. Ces variants possèdent des propriétés génotypiques et fonctionnelles particulières au niveau des glycoprotéines de leur enveloppe leur permettant d'établir efficacement l'infection chez le nouvel hôte et de conduire par la suite à une dissémination systémique de la maladie. Décrypter l'ensemble des déterminants génétiques ou changements discrets (polymorphisme) au niveau de la glycoprotéine de l'enveloppe de ces variants permettrait de fournir des indices pour la conception de vaccins préventifs.

Méthodes : Cent vingt-six séquences consensus de l'enveloppe du VIH-1 de clade B provenant de sujets infectés en phase aiguë et récents non traités au Québec ont été comparés à cent cinq séquences d'individus infectés en phase chronique aussi de clade B sélectionné dans la base de données de séquences virales de Los Alamos (LANL). Un séquençage à haut débit (nouvelle génération de séquençage) et des analyses moléculaires ont permis d'obtenir ces séquences. Des outils moléculaires (Weblogo) des séquences virales ont permis d'identifier les signatures ou déterminants génétiques associés à chaque type viral (TF, récent ou chronique).

Résultats : La présente étude révèle la présence d'une signature génétique (mutation d'acide aminé) au niveau de la séquence de l'enveloppe virale des virus TF par rapport aux virus d'infections chroniques. Cette sélection positive consiste en une substitution d'arginine (R) par une isoleucine (I) en position 841 (R841I) située dans la queue cytoplasmique de la GP41, particulièrement au niveau du segment 1 des peptides lytiques de lentivirus (LLP-1), OR = 0.2, 95% [0.09, 0.44], $p = 0.00001$. Une sélection assez importante de l'isoleucine (I) (K6I), localisée en position 6 du peptide signal de l'enveloppe, par les virus d'infection chroniques par rapport aux virus TF, OR = 3.26, IC à 95% [1.76-6.02], $P = 0.0001$ est significativement observée.

Conclusions : La région LLP-1 du domaine cytoplasmique de la GP41 possède une séquence beaucoup plus conservée. Il joue un rôle majeur dans la réplication virale et intervient dans le

trafic intracellulaire et l'incorporation dans les virions. La signature génétique identifiée consistante en une mutation de l'arginine par une isoleucine associée aux virus TF pourrait constituer des facteurs clés permettant le maintien de la capacité de réplication du virus et au succès de la transmission au cours de phase aiguë de l'infection.

Mots clés: VIH-1, infection précoce et récente, virus transmis/fondateur, acide amines, signatures génétiques, peptide signal, domaine cytoplasmique, segment 1 de peptide lytique de lentivirus.

ABSTRACT

Background: HIV-1 transmitted/founder viruses (TF) are selected during the acute phase of infection from a multitude of virions present during transmission. They possess the capacity to establish infection and viral dissemination in a new host. Deciphering the discrete genetic determinant of infectivity in their envelope may provide clues for vaccine design.

Methods: One hundred twenty-six clade B HIV-1 consensus envelope sequences from untreated acute and early infected individuals were compared to 105 sequences obtained from chronically infected individuals using next generation sequencing and molecular analyses.

Results: We identified an envelope amino acid signature associated with TF viruses. They are more likely to have an isoleucine (I) in position 841 instead of an arginine (R). This mutation of R to I (R841I) in the gp41 cytoplasmic tail (gp41CT), specifically in lentivirus lytic peptides segment 1 (LLP-1), is significantly enriched compared to chronic viruses (OR = 0.2, 95% CI (0.09, 0.44), $p = 0.00001$). Conversely, a mutation of lysine (K) to isoleucine (I) located in position six (K6I) of the envelope signal peptide was selected by chronic viruses and compared to TF (OR = 3.26, 95% CI (1.76–6.02), $p = 0.0001$).

Conclusions: The highly conserved gp41 CT_ LLP-1 domain plays a major role in virus replication in mediating intracellular traffic and envelope incorporation into virions in interacting with encoded matrix protein. The presence of an isoleucine in gp41 in the TF viruses' envelope may sustain its role in the successful establishment of infection during the acute stage.

Keywords: HIV-1, acute/early infection, transmitted/founder viruses, recent viruses, envelopes, amino acids, genetic signatures, signal peptides, cytoplasmic domain, lentivirus lytic peptides segment 1.

1. INTRODUCTION

HIV-1 genetic diversity due to frequent mutation rates, polymorphisms, recombination events and altered pattern of glycosylation within the envelope (Env) [1] drives HIV-1 escape from broadly neutralizing antibodies (bnAbs) and other mechanisms developed by the immune system [2,3,4]. HIV-1 envelope polymorphism determines the functional properties of the virus during disease progression [4] and contributes to the rapid evolution of HIV-1 toward the establishment of a viral reservoir early after infection [5,6], which impedes HIV-1 cure and vaccine efforts [7,8]. For 80% to 90% of heterosexual transmission cases, it is estimated that a single HIV virus strain is capable of establishing successful transmission and is referred to as a transmitted/founder (TF) virus [9,10,11,12]. However, if the recipient presents genital ulceration or if transmission occurs via other routes, more than one strain can contribute to the initiation of infection [10,13,14].

The envelope glycoprotein of HIV-1 is a complex trimeric glycoprotein found on the surface of the virus and composed of gp120 subunit spikes linked in a noncovalent interaction with gp41. It is encoded by a viral genome and interacts with host cells and plays an essential role in the virus replication cycle by mediating the fusion between viral and cellular membranes during the entry process [15]. The HIV-1 envelope glycoprotein harbors the transmembrane domain [15]. The HIV-1 envelope structure starts with an N-terminal signal peptide (SP) that guides the nascent *env* to the endoplasmic reticulum (ER), where *env* synthesis and post translational modifications take place [16]. The gp120 glycoprotein is subdivided into five conserved subdomains (C1–C5) and five hypervariable glycosylated loops (V1–V5) [15,17,18,19] involved in HIV-1 pathogenesis and viral escape. HIV-1 gp41 consists of 345 amino acids [20], thus facilitating viral protein and host cell membrane fusion [21,22]. Gp41 has three subdomains [23] and seven functional domains, including the ectodomain that mediates membrane fusion [21,24] and the conserved transmembrane domain (TMD) [25], which acts as a membrane anchor that prevents the release of gp160 into the lumen of the endoplasmic reticulum (ER) [26,27], mediates fusion of the viral envelope with the host cell membrane [15] and enhances cell to cell and virus to cell fusion [28], and the cytoplasmic

domain, also called the cytoplasmic tail (CD) [15,29,30]. The latter includes three lentivirus lytic peptides segment 1, 2 and (LLPs) [30,31]. LLPs contribute to the surface expression of *env* [32,33], *env* incorporation into viral particles [34,35], fusogenicity [36,37] and its localization in lipid rafts [31].

The envelope glycoprotein of HIV-1 is encoded by a viral genome and this protein mediates the first contact with host cells. Even discrete or predominant, an amino acid change (mutations, insertions and deletions) in this viral specific region merit further analysis. Such changes may constitute an important genetic signature developed at the acute stage of infection and selected by TF viruses. They may constitute as keys factors that influence viral fitness and enhancing HIV transmission to new recipients.

The current study characterized the HIV-1 envelope variable region with regard to loop length, N-linked glycosylation sites and the V3-positive net charge. We also screened the *env* full-length amino acid sequences to identify mutation patterns occurring in acute/early stages of HIV infection.

2. MATERIALS AND METHODS

2.1. Description of specimens

We used specimens from HIV-1-infected individuals obtained from the Laboratoire de santé publique du Québec (LSPQ) of the Institut national de santé publique (INSPQ) serobank collection (original samples) from first time individual HIV diagnosis before enrollment in any antiretroviral therapy. The majority HIV-1 infected population is men who have sex with men (MSM) and someone injecting drug users (IDU). The LSPQ is the Public Health Laboratory of the province of Quebec, where confirmation testing of HIV infection is performed on all HIV-positive tests in Quebec. All the LSPQ specimens used in this study were residual sera first collected for routine diagnostic purposes between year 1991 and 2015. Specimens deriving from acute HIV-1 infection contained TF viruses (Fiebig stages 1 and 2) and were selected based on p24 antigen positivity in the absence of HIV-1 antibodies [38]. The recent (RC; Fiebig 3 to 5) and chronic (CH) HIV-1-infected individuals' serum samples were determined by recent testing algorithms (RITA) performed at the LSPQ [39]. The mean duration of recent infection (MRDI) is defined as less than 136 days of infection [39]. Confirmed untreated chronic clade B HIV-1 envelope sequences were also obtained from the Los Alamos sequence databases (LANL) as part of the data set for chronic infection in addition to chronic viruses' envelope sequences obtained among the public health laboratory of Quebec (LSPQ) serobank sample collection. The genetic diversity in the HIV-1 envelope sequence of TF viruses [40] may govern the functional properties and provide clues on viral transmission success and immune evasion [9]. Recent HIV viruses usually correspond to those identified at an early stage of infection within the 6 month period [40,41]. For this study, we considered early HIV infection a period within mean duration of recent infection (MDRI) of 136 days [39]. HIV-1 viruses identified at this period were defined as recent viruses. The distribution of the specimens and sequences per category of infection are described in Figure 1 and referred nomenclature in supplemental Table S1.

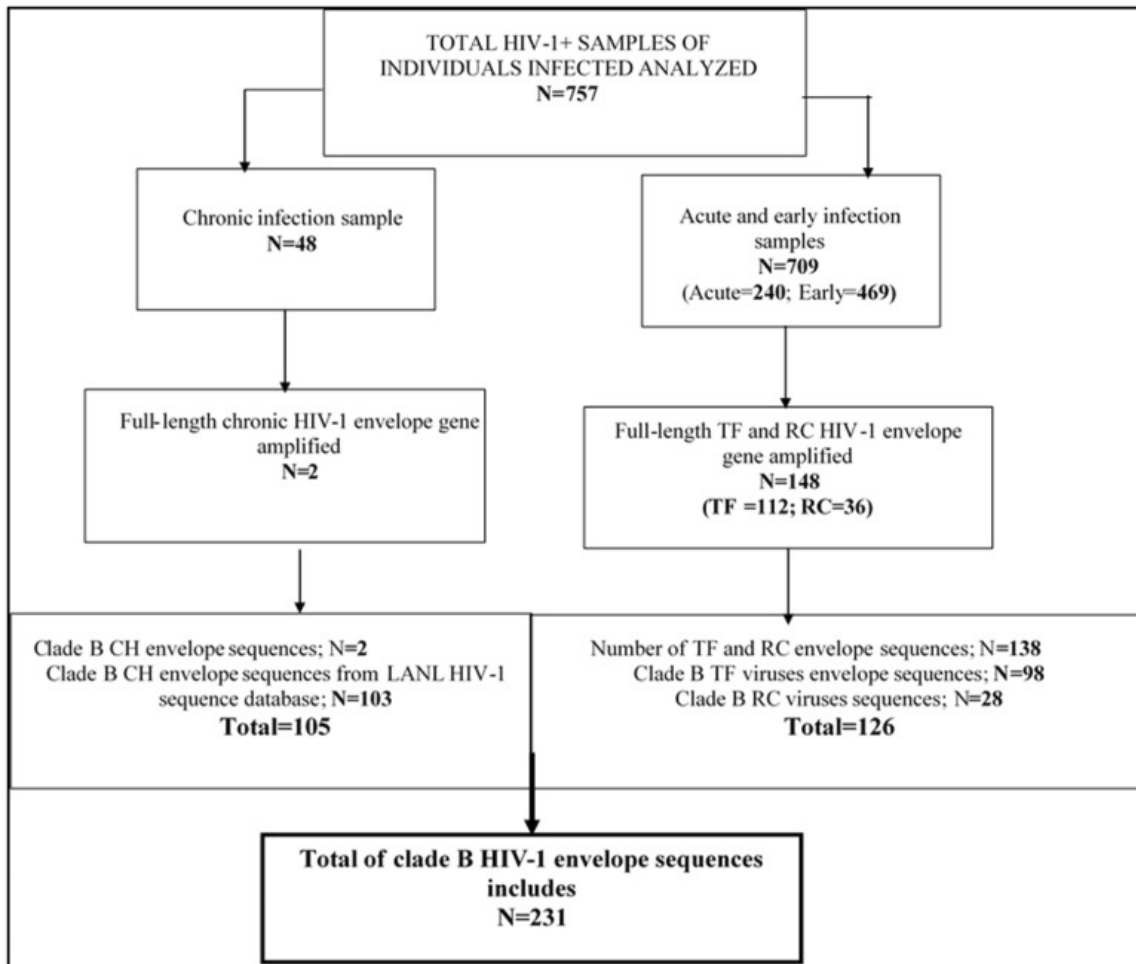


Figure XXXX. (Article 3, Figure 1): Flow chart showing the numbers of HIV-1 positive samples analyzed, and consensus envelope sequences of HIV-1-infected individuals include in the study.

From a total of 243 sequences obtained, 95% (231) of subtype B HIV-1 sequences were included in the current study. The non-B HIV-1 subtype envelope sequences representing 8.7% (four for transmitted/founder (TF) and eight for recent (RC) viruses) of the LSPQ serobank HIV-1 positive samples were excluded for subsequent genetic analysis because they may have influenced the results. The HIV-1 envelope sequences of transmitted/founder (TF) viruses derived from acute infection, recent viruses (RC) for early infection and chronic viruses for chronic infection.

2.2. HIV-1 RNA extraction

One hundred (100) microlitres of unique serum from each HIV-1-infected subject was used for HIV-1 RNA extraction using BioRobot MDx automated viral RNA extraction. The

QIAamp®Virus BioRobot® MDx Kit (Qiagen, Valencia, CA, USA) was used. To respect the minimum 350-µL sample volume required for BioRobot automate extraction, we diluted each 100-µL serum sample with 250 µL Dulbecco's Modified Eagle's Medium (DMEM; Sigma-Life Science, Oakville, Ontario, CA, USA). Extraction was conducted automatically according to the manufacturer's protocol. One positive and one negative control sera were always included in each panel of extraction for quality control. Suspension of extracted RNA (approximately 60–80 µL) was immediately used for reverse transcription or stored at –80 °C for reference use.

2.3. Reverse transcription (RT-PCR).

Reverse transcription of the full-length HIV-1 RNA envelope was performed using the following primers: SG3-lo forward and SG3-up reverse [42,43], which cover 4.7 kb fragments of *env* length. The extracted RNA (5 µL) was reverse-transcribed in a total volume of 50 µL using the Superscript III One-Step RT-PCR system with Platinum® Taq DNA polymerase ((Invitrogen, Carlsbad, CA, USA), which included 2.0 µL SuperScript™ III RT/Platinum® Taq mix (10 U/µL), 25 µL of 2X reaction mix at a concentration of (05 µM), 2.0 µL of each primer (SG3-lo and SG3-up) at a concentration of (10 µM), 0.5 µL of RNase OUT at concentration of (2 U/µL) completed with 13.5 µL of diethylpyrocarbonate (DEPC) water.

The RNA product (10 µL) was first denatured at 65 °C for 5 min in a thermocycler (Applied Biosystems (ABI) GeneAmp PCR System 9700). The denatured RNA (5 µL) was added to the 45 µL reaction mix containing primers, RNase OUT and the Platinum® Taq DNA polymerase. The reaction mix was then placed in an Applied Biosystems (ABI) thermocycler for cDNA synthesis. The thermal profile was as follow: 53 °C for 30 min and 94 °C for 2 min, followed by 40 cycles at 94 °C for 2 min for denaturing, 55 °C for 30 s for annealing, 68 °C for 4 min for extension and 68 °C for 5 min with a final hold at 4 °C. The PCR product was immediately amplified (nested PCR) or stored at 4–8 °C for future use.

2.4. Second amplification

After RT-PCR, a nested polymerase chain reaction (nested PCR) of the full-length HIV-1 envelope gene (GP160) was amplified using appropriate primers that covered a fragment of 3.10 kb. The following primers were used: *env* up forward (5'-

GTTTCTTTTAGGCATCTCCTATGGCAGGAAGAAG-3'), HXB2 nucleotides (nts) position (5957–5983) and *env-lo* reverse (5'-GTTTCTTCCAGTCCCCCTTTTCTTTTAAA AAG-3'), HXB2 nts position (9063–9088). For PCR amplification, cDNA (2 µL) was amplified in a total volume of 50 µL using the Expand™ High Fidelity PCR System (3.5 U/µL) enzyme (Roche Life Sciences, Mannheim, Germany). For the preparation of the reaction mix, (2 µL) of cDNA was added to 0.4 µL of the expand high fidelity enzyme mix, 5.0 µL of the expand high fidelity buffer at (10×) concentration, 1.5 µL of dNTPs at concentration of (10 mM) and 2.0 µL for each primer (*env up* and *env-lo*) at a concentration of 10 µM) completed with 37.10 µL of diethylpyrocarbonate (DEPC) water.

For amplification in the ABI 9700 thermocycler, the following temperatures were used: 94 °C for 2 min, followed by 45 cycles of denaturing at 94 °C for 15 s, annealing or hybridization at 55 °C for 30 s, and extension or elongation at 68 °C for 2 min, followed by 68 °C for 7 min and a hold at 4 °C. The PCR products were immediately visualized on a 1% agarose gel by electrophoresis and purified using DNA purification kits from QIAGEN and stored at –20 °C before sequencing.

2.5 DNA Sequencing and Sequence Assembly

Two µL of input purified DNA was quantified by a Nanodrop and the appropriate concentration was established. In addition, 5 µL (0.2 ng/µL) of input purified DNA was also quantified by iQ™5 Optical System Software, (Bio-Rad Laboratories Ltd. Ontario, Canada) using PicoGreen dsDNA Quantification Reagent). Full-length gp160 of the HIV-1 viral envelope gene was sequenced using MiSeq (Illumina, San Diego, CA, USA) a next generation sequencing (NGS) method with a MiSeq® Reagent Kit (San Diego, CA, USA). The Nextera XT DNA library prep kit (Illumina, San Diego, CA) was used for library preparation and the manufacturer's protocols were respected. The Illumina MiSeq system was edited using MiSeq Reporter, a bioinformatics data analysis software built into the MiSeq. The workflow using the Nextera XT DNA library kit contains the following steps: (1) tagmentation of genomic DNA, (2) PCR amplification, (3) PCR Clean-up, (4) library normalization and (5) library pooling for MiSeq sequencing were strictly respected. After cycle sequencing, gigabase data

provided by Illumina MiSeq were transferred and stored into a securely cloud-based genomics computing environment named BaseSpace Sequence Hub.

2.6. Data Management and Analysis

Sequence analysis was performed by cycle-sequencing using Illumina MiSeq. The data produced were viewed by a sequencing analysis viewer (SAV) as recommended by the manufacturer. Individual sequence fragments were assembled using an IVA (iterative virus assembler) and consensus sequences were identified by each specimen. Only the sequences that represented most of 1% of the viral population were retained for subsequent analyses to reduce potential recombination artifacts that may influence viral sequence diversity. Consensus sequences obtained from NGS and IVA assembly obtain from original data and those obtain from Los Alamos HIV-1 sequence database were aligned using Clustal Wallis and conducted in Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets (MEGA7.0) software (www.megasoftware.net) [44]. A reference “HXB2” (GenBank accession number: K03455.1) envelope sequence, gp160 (amino acids residues 512–856 of full genome numbering) was included in the alignment. Each sequence was also aligned in the blastx homepage with online software (<http://blast.ncbi.nlm.nih.gov>) and screened to identify potential protein products encoded by a nucleotide query of each sequence. This blastx search ensures that all sequences correspond to the amplified full-length gp160 HIV-1 envelope. All ambiguous or gaps sequences were excluded from subsequent analyses.

We used the online Los Alamos sequence database tools to determine the characteristics of the HIV-1 variable regions (GP120- V1 to V5 loops). This online tool provides results of the HIV N-linked glycosylation site, the loop length and the V3 loop net charge (NC) where we used default setting that have computed with KRH = (+) and DE = (-). The following link was used: https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/index.html.

Determination of the HIV-1 envelope amino acid sequences logo and frequencies for three category of infection stages (TF, RC and CH) were performed using WebLogo Versions 2.8 and 3 (<http://weblogo.berkeley.edu/logo.cgi>) [45,46]. The WebLogo tools generate sequence logo, graphical representations of the patterns within a multiple sequence alignment. HIV-1 envelope sequences subjected to WebLogo analyses (N = 231)) were first submitted to

multiple sequence alignments using MEGA 7.0 software. The HXB2 *env* GP160 sequences were introduced in alignment for numbering purposes [15,29,31,47,48,49]. Aligned envelope sequences for each category of viruses (chronic, transmitted/founder and recent) were then downloaded separately in fasta files. Sequences files for each category of infection were subsequently uploaded individually and analyzed in WebLogo online software. Weblogo analysis identify and counts number of individuals amino acids selected at each position of the *env* sequence length (1-856) and report frequency from the population. Sequence data of each category of infection were further downloaded in plain text output format that reported the total count of selected amino acids at each position of *env* aligned sequence. We finally generated three files of amino acid counts for each of the three categories of viruses (TF, RC and CH) and proceeded to statistical comparison; position by position and amino by amino acid, between them and reported the significant difference. We referred to HXB2 *env* sequences introduced as reference in MEGA 7.0 multiple sequence alignments to identify the exact position of the amino acids in *env* by checking in black the boxes: without (w/o) gaps. Localization (sub-regions or domains of *env*) for any amino acid changes referred to HXB2 numbering as summarized in Supplemental Table S2.

Envelope nucleotides and amino acid sequences for all full-length HIV-1 transmitted founder, recent and chronic viruses were deposited and are available in the GenBank sequence database under the accession MK076153-MK076292. HIV-1 envelope sequence data qualifiers are also available in Supplemental Table S3. The background information of selected LANL chronic clade B HIV-1 envelope sequences were reported in the Supplemental Table S4.

2.7. Statistical Analyses

We used descriptive statistics, the mean and nucleotide composition across HIV-1 envelope gp160 length to estimate the amino acid differences between transmitted founder with recent and chronic sequences using the HXB2 envelope sequence as a reference. Descriptive statistics were performed using proportion and means or median for qualitative and quantitative variables, respectively, as well as ridge plots. The Kruskal Wallis, Wilcoxon and

Chi square tests were used to compare the different parameters per type of infection. The Wald test statistics with logistic regression model were also used. Stata version 14, R version 3.5.1 and SPSS version 24 were used as statistical software. A p-value less than 0.05 was considered statistically significant. Moreover, the p-values were adjusted using the Benjamin Hochberg procedure for multiple comparisons.

2.8. Ethics Approval and Consent to Participate

Ethical approval was given by the “Comité d’éthique et de la recherche (CÉR) des Centres hospitaliers affiliés à l’Université de Montréal (CHUM) ; *Number :2015-5569, CE14-344CA*. It was yearly renewed since 2015 by our Institutional review board. All samples were anonymized before application in this study. No nominal information was used for analysis or data management. This manuscript did not contain any individual data in any form whatsoever to publish.

3. RESULTS

A total (N) of 757 specimens from acute and early HIV-1 infections based respectively on EIA-p24 antigen positive, Western blot (WB) negative and WB positive with the presence of HIV-1 antibodies and qualified by a recent infection testing algorithms (RITA) [37,39] (Figure 1). Chronic clade B HIV-1 virus envelope sequences were selected from Los Alamos HIV sequence databases and constitute a part of chronic sequences that were obtained from the LSPQ serobank collection (Figure 1, Supplemental Table S1).

From acute HIV-1-infection serum samples (N = 469) classified as TF viruses, we obtained 98 consensus individual clade B HIV-1 envelope sequences after a molecular evolutionary genetic analysis. Five of the one hundred tree sequences were excluded because they had short envelope amino acids sequences lengths (<856 bp) after MEGA 7 multiple sequence alignment. The nested-RT-PCR success rate of acute infection samples was 23% (102/469), as presented in Figure 1.

From the total of early HIV-1 infection samples (N = 240) where viruses identified as classified as recent HIV viruses, twenty-eight (28) HIV-1 consensus envelope sequences were obtained

(Figure 1). This result corresponded to an RT-PCR amplification success rate of 15% (36/240). Eight (8) non-B HIV-1 envelope sequences were excluded for molecular analysis (Figure 1).

Of forty-eight (48) chronic HIV-1 infection samples collected from LSPQ serobank samples collections, only two HIV-1 envelope sequences were finally obtained after analysis, which demonstrated a fair result after HIV RNA amplification (4%; Figure 1).

The repartition of clade B HIV-1 envelope sequences (one sequence per individual) in total is as follow: chronic (CH): 45.46% (N = 105 includes); TF viruses: 42.42% (N = 98 includes, N = 4 non-B HIV-1 was excluded) and recent (RC): 12.12% (N = 28 were included, N=8 non-B HIV-1 were excluded) as shown in Figure 1 and Supplemental Table S3. A total of 105 HIV-1 B chronic envelope sequences included two HIV-1 envelope sequences of LSPQ serobank sample collections and LANL chronic HIV-1 clade B envelope sequences (Figure 1). The background information of LANL selected chronic HIV-1 envelope sequences is available in Supplemental Table S4.

Thus, a total of 231 clade B HIV-1 full-length consensus envelope sequences were included in this analysis.

3.1. Characteristics of HIV-1 Envelope Variable Regions

The first objective of the current study was to determine the characteristics of the HIV-1 variable regions, including the number of N-glycosylation sites, the loop length and the V3 loop positive net charge between TF, RC and CH viruses. As presented in Figure 2d and Supplemental Table S5, the differences in the numbers of N-glycosylation sites of the HIV-1 *env* GP120 loop 3 (V3) were statistically significant using the Kruskal–Wallis test and Wald test with logistic regression model. These concerns CH (median/range: 2 (2,2)) and TF (median/range: 2 (2,2)), (OR = 0.58, 95% CI (0.36–0.93), p = 0.026), between RC (median/range: 2 (2,3) and TF: 2 (2,2) (OR = 0.37, 95% CI (0.19–0.73), p = 0.004 and CH (median/range: 2 (2,2) and RC (median/range: 2 (2,3), (OR = 2.03, 95% CI (0.98–4.21), p = 0.05).

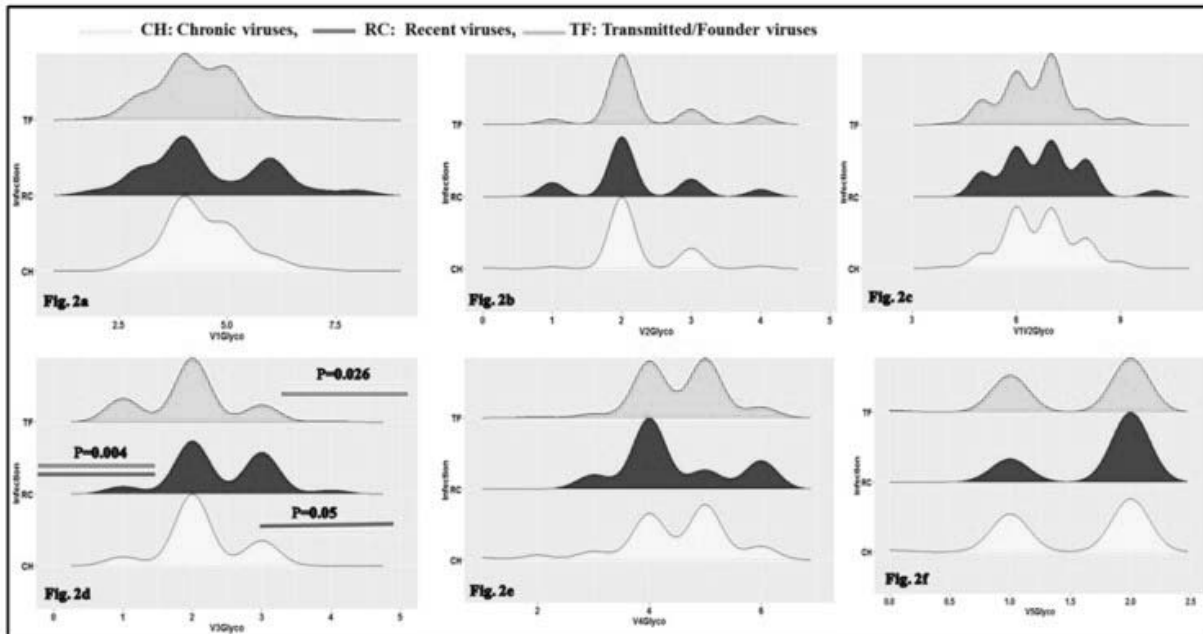


Figure XXXI. (Article 3, Figure 2): Ridge plot comparing the HIV-1 envelope variable loop number of N-glycosylation sites between TF, RC and chronic (CH) viruses.

The boxes represent a density plot of a number of N-glycosylation sites for: *env* V1 loop (a), *env* V2 loop (b), *env* V1V2 loop (c), *env* V3 loop (d), *env* V4 loop (e) and *env* V5 loop (f) for CH, RC and TF viruses respectively. In box, the top (green), middle (blue) and bottom (yellow) represent respectively number of N-glycosylation sites for CH, RC and TF viruses envelope sequences respectively. The X-axis represents sequence loops number of N-glycosylation sites and the Y-axis the density of sequences number of N-glycosylation sites for each timeline category of viruses (CH, RC and TF). As shown in Figure 2d, the differences in *env* V3 loop numbers of N-glycosylation sites between CH and TF, $p = 0.026$; RC and TF, $p = 0.004$ and CH and RC, $p = 0.05$ were statistically significant using Wald test with logistic regression model.

The difference in the HIV-1 *env* GP120 loop 5 (V5) lengths was statistically significant using the Kruskal–Wallis test and Wald test with logistic regression model (Figure 3f and Supplemental Table S5). These differences concerned RC (median/range: 15 (13, 15) and TF (median/range: 13 (12, 14)), OR = 0.65, 95% CI (0.49–0.86), $p = 0.003$ and between CH (median/range: 15 (12, 14)) and RC (median/range: 15 (13, 15), OR = 1.44, 95% CI (1.12–1.86), $p = 0.004$.

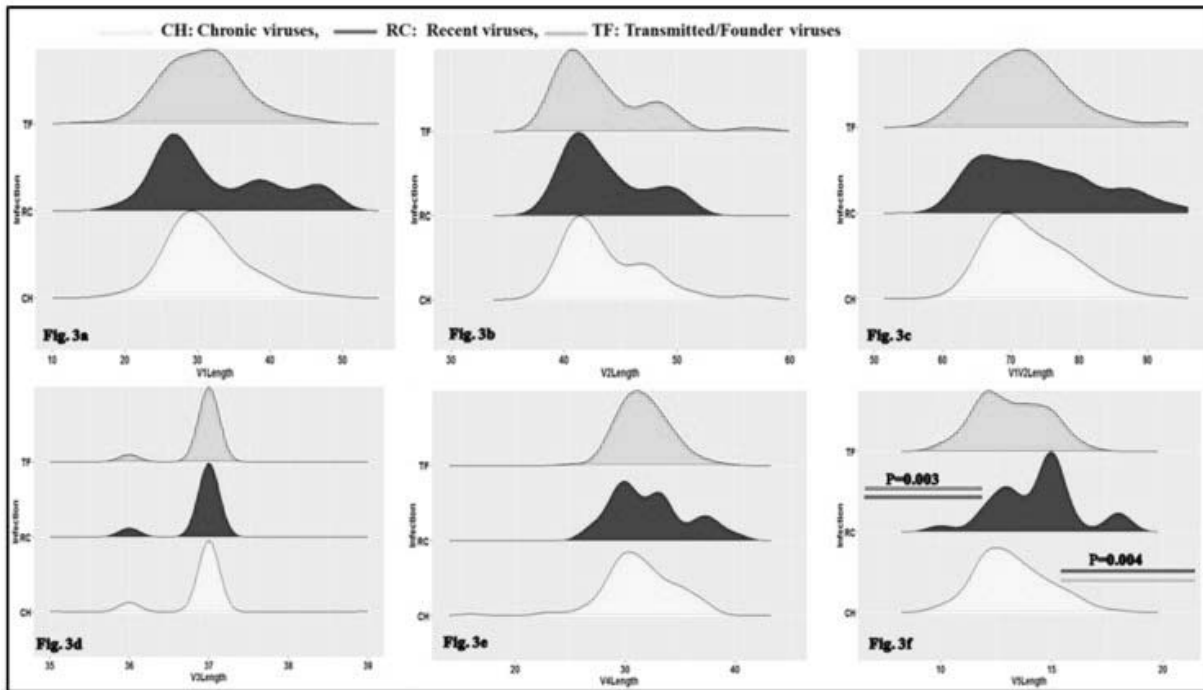


Figure XXXII. (Article 3, Figure 3): Ridge plot comparing the clade B HIV-1 envelope variable loop length between CH, RC and TF viruses.

The boxes represent the density plot of: *env* V1 loop lengths (a), *env* V2 loop lengths (b), *env* V1V2 loop length (c), *env* V3 loop length (d), *env* V4 loop length (e) and *env* V5 loop length (f) for CH, RC and TF viruses respectively. For each box, the top (green), middle (blue) and bottom (yellow) represent, respectively, TF, RC and CH viruses sequences. X-axis presents sequence loop lengths and the Y-axis the loop length density for each timeline category of viruses (CH, RC and TF). As presented in Figure 2f, the differences in the HIV-1 *env* GP120 V5 loop lengths between RC and TF, $p = 0.003$ and CH and RC, $p = 0.004$ are statistically significant using Wald Test with regression logistic model.

The positive net charge of the HIV-1 *env* GP120 loop 3 (V3) was also statistically significant between CH (median/range: 5 (3, 6), RC (median range: 4 (3, 5.5) and TF (median/range: 4 (3, 5), $p = 0.040$ and specifically, between CH (median/range: 5 (3, 6), and TF (median/range: 4 (3, 5), viruses (OR = 0.82, 95% CI (0.69–0.98), $p = 0.038$) using the Kruskal–Wallis test and Wald test with logistic regression model (Figure 4; Supplemental Table S5).

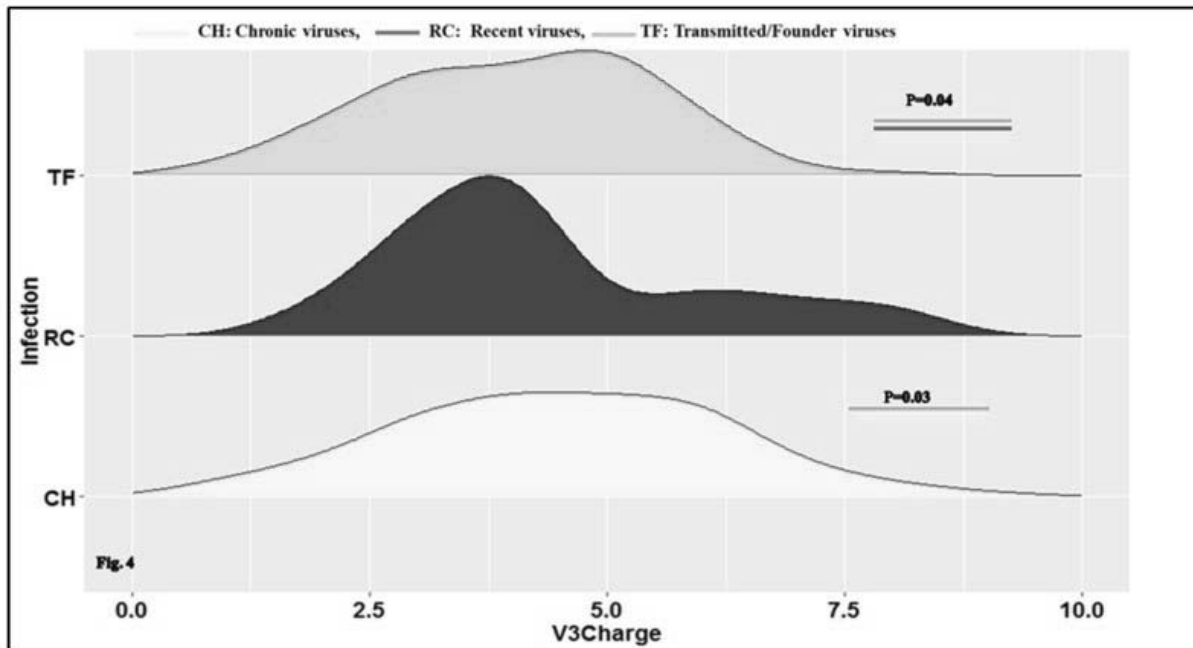


Figure XXXXIII. (Article 3, Figure 4). Ridge plot comparing the clade B HIV-1 envelope variable region gp120 loop 3 (V3) net positive between CH, RC and TF viruses.

The box represents a density plot of the V3 positive net charge. The top (green), middle (blue) and bottom (yellow) represent the TF, RC and CH viruses' V3 sequences net charge, respectively. The X-axis represents the number of charges for the HIV-1 *env* gp120V3 loop and the Y-axis represents the density of sequence charges of HIV-1-infected individuals for the CH, RC and TF viruses respectively. As shown in Figure 4, the difference of the V3 positive net charge was significant between CH, RC and TF, $p = 0.04$. Importantly, the difference in HIV-1 *env* V3 loop net charge was statistically significant between CH and TF viruses, $p = 0.03$ using Wald with regression logistic model. No significant difference was observed between RC and TF, $p > 0.05$.

3.2. Clade B HIV-1 Envelope Amino Acids Signatures Associated to Transmitted/Founders and Recent Viruses Compared to Chronic

The second objective of this study was to screen full-length HIV-1 envelope sequences to identify genetic characteristics (mutation patterns) associated with transmitted/founder and recent viruses compared to chronic ones. As presented in figures 6, 7 and Table 1, two genetic signatures were identified.

The first significant amino acids enrichment difference between CH and TF was observed in the HIV-1 envelope gp41 cytoplasmic tail, specifically in the Lentivirus Lytic peptide 1 (LLP-1). It concerns a substitution of an arginine (R) by an isoleucine (I) at position 841 (R841I) in reference to HXB2 *env* sequence numbering (Figure 5, Table 1).

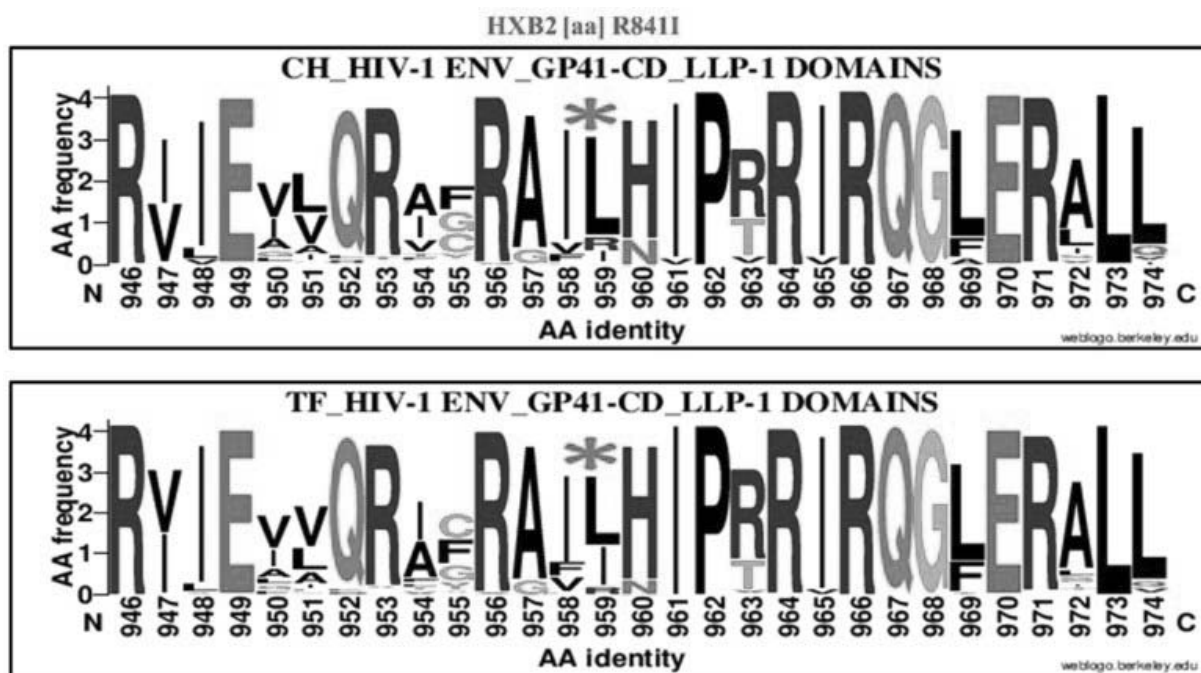


Figure XXXIV. (Article 3, Figure 5). Genetic signature identified under the HIV-1 envelope GP41 Lentivirus Lytic peptide 1 (LLP-1) associate to clade B HIV-1 TF viruses compared to chronic (CH) using WebLogo.

The X axis represents amino acids (AA) composing the LLP-1 sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the LLP-1 sequence for each category of infection. The top line box represents the chronic (CH) viruses envelope GP41 LLP-1 sequence (N = 105) and the bottom line for TF viruses (N = 98). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The isoleucine (I) amino acid signature was located at position 959 of alignment (HXB2 position R841I) and identified by a red asterisk. The isoleucine was highly enriched in TF viruses LLP-1 domains, 33.68% (32/95), compared to CH virus LLP-1 domains, 9.52% (10/105), OR = 0.2, 95% CI (0.09–0.44), $p < 0.00001$ using the hi-squared (χ^2) test.

The second genetic signature was identified in the HIV-1 envelope signal peptide (SP). It concerns a substitution of lysine (K) by an isoleucine at position six of HXB2 numbering (Figure 6, Table 1). The substitution of lysine (K) for isoleucine (I; K6I) in the *env* SP at position six was highly enriched in chronic viruses, 79.04% (83/105) compared to TF viruses, 53.60% (52/97), OR = 3.26, 95% CI (1.76–6.02), $p = 0.0001$ using the chi-squared (Chi2) test.

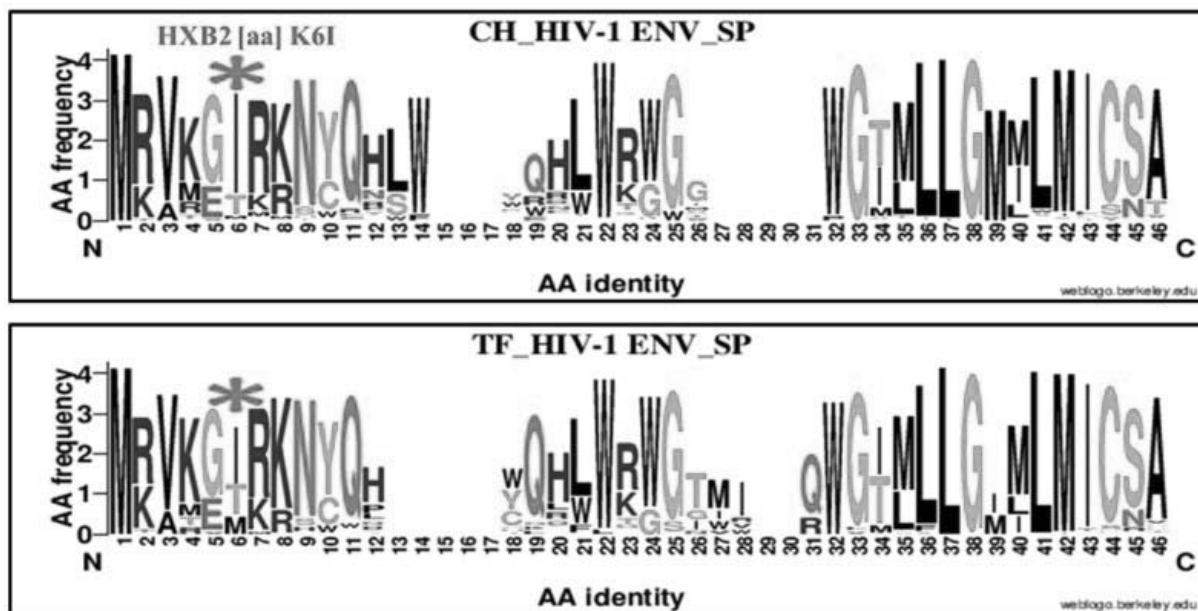


Figure XXXXV. (Article 3, Figure 6). Genetic signature identified under the HIV-1 envelope signal peptide (SP) associate to clade B HIV-1 chronic compared to TF viruses using WebLogo.

The X axis represents sequences and amino acids (AA) identities composing the SP (direction N to C). The X axis represents amino acids (AA) composing the LLP-1 sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the SP sequence for each category of infection. The top line box represents chronic (CH) viruses envelope SP sequence (N = 105) and the bottom line for TF viruses (N = 98). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The isoleucine (I) amino acid signature was located at position six of alignment (HXB2 position, K6I) and identified by a red asterisk. *NB. The lysine (K) was found at position 6 of HXB2 env SP. However, as reported in figure 6, the TF like chronic viruses selected other amino acids than K at the same position, importantly for isoleucine (I).*

Other significant amino acids mutation patterns that distinguish TF from CH HIV-1 envelope sequences were also found less significant amino signatures in the GP120 C1 VI, V5 loops and GP 41 fusion peptide (FP), Kennedy Epitope (KE), loop and Fusion peptide proximal region (FPPR; Table 1).

Tableau XIII. Table 1. Summary statistics of the important amino acids change among the HIV-1 envelope sequences of chronic and transmitted/founder viruses.

HXB2		Env	Amino Acids			CH			TF			Chi ² Test		
Position	Amino Acids	Subregion/ Domain	Alignment Position	Change	Genetic Signature	YES	NO	Total	YES	NO	Total	OR	95% CI	B-H adjusted p-values
841	R	LLP-1	959	I	R841I	10	95	105	32	63	95	0.2	0.09–0.44	0.00001
6	K	SP	6	I	K6I	83	22	105	52	45	97	3.26	1.76–6.02	0.0001
62	D	C1	81	E	D62E	20	85	105	4	94	98	5.52	1.89–16.04	0.006
514	G	ECD	628	T	G514T	9	22	31	26	12	38	0.18	0.06–0.52	0.006
24	M	SP	40	I	M24I	38	66	104	16	81	97	2.91	1.50–5.64	0.006
743	R	HIR/KE	865	R	743R	45	60	105	62	35	97	0.42	0.24–0.74	0.008
153	E	V1	206	E	153E	93	11	104	71	26	97	3.09	1.44–6.59	0.008
744	R	HIR/KE	862	R	744R	78	27	105	53	44	97	2.39	1.32–4.32	0.008
717	F	HIR/KE	834	F	717F	86	19	105	62	35	97	2.55	1.34–4.85	0.008
717	F	HIR/KE	834	L	I717L	19	86	105	35	62	97	0.39	0.20–0.74	0.008
154	I	V1	207	M	I154M	35	70	105	52	46	98	0.44	0.25–0.77	0.008
744	R	HIR/KE	862	T	R744T	15	90	105	30	67	97	0.37	0.18–0.74	0.008
841	R	LLP-1	959	L	R841L	81	24	105	56	39	95	2.35	1.27–4.31	0.009
621	Q	GP 41 Loop	738	D	Q621D	16	88	104	31	67	98	0.39	0.20–0.77	0.009
464	L	V5	566	N	L464N	24	45	69	42	31	73	0.39	0.20–0.77	0.009
543	Q	FPPR	657	Q	543Q	63	42	105	76	22	98	0.43	0.23–0.79	0.009

Table 1 presents the important HIV-1 envelope sequence polymorphisms of chronically infected individuals compared to TF viruses of individuals infected that are considered as genetic signatures. The bold typeface defines the important *env* genetic signatures identified and discussed in manuscript. The gray character referred to sequence alignment position of amino acid change. Abbreviations: V1—variable loop 1, CP—cytoplasmic domain/tails, EC—endocytosis domain, SP—Signal peptides, C2—constant domain 2, C3—constant domain 3, ECD-Loop—Ectodomain-loop region, CP-KE—cytoplasmic domain-Kennedy epitope, ECD-CHR—Ectodomain-C-hepta-repeat, CP-LLP-1—Cytoplasmic Tail-Lentivirus Lytic peptide 1. MSD—Membrane-spanning domain.

3.3. HIV-1 Envelope Genetic Signatures Among Transmitted/Founder and Recent Viruses Compared to Chronic

Four important genetics signatures were also identified when combining TF and RC compared to chronic viruses (Figure 7, Figure 8, Figure 9 and Figure 10). The first one was located in the GP120 V1 loop at position 153 (Figure 7) and did not constitute a change. However, it identified a high enrichment of glutamic acid (E; 153E) in CH viruses, 89.42% (93/104) and 65.60% (82/105) in RC (+TF) ones, OR = 4.43, 95% CI (2.16–9.05), p = 0.000001 using the Chi2 test.

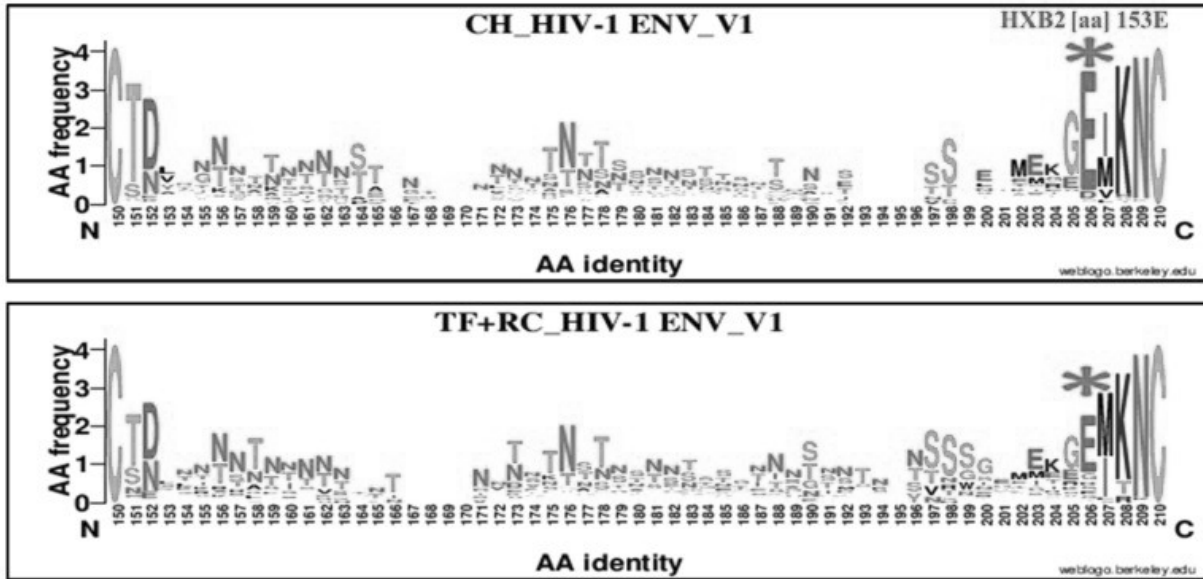


Figure XXXVI. (Article 3, Figure 7). Genetic signature identified under the HIV-1 envelope GP120 V1 loop associate to clade B HIV-1 chronic compared to recent viruses.

The X axis represents amino acids (AA) composing the V1 loop sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the V1 sequence for each category of infection. The top line box represents chronic virus envelope V1 loop sequences (N = 105) and the bottom for RC viruses within 136 days MDRI and TF (N = 126). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The glutamic acid (E) amino signature was located at position 206 of alignment (HXB2 position 153E) and was identified by a red asterisk.

The second was found in the *env* signal peptide at position 24 and identifies a substitution of a methionine (M) by an Isoleucine (I; M24I; Figure 8). The isoleucine mutation was enriched at 36.53% (38/104) for CH and 12.08% (16/125) for RC (+TF), OR = 3.92%, 95% CI (2.04–7.53), p = 0.00001 using the Chi2 test.

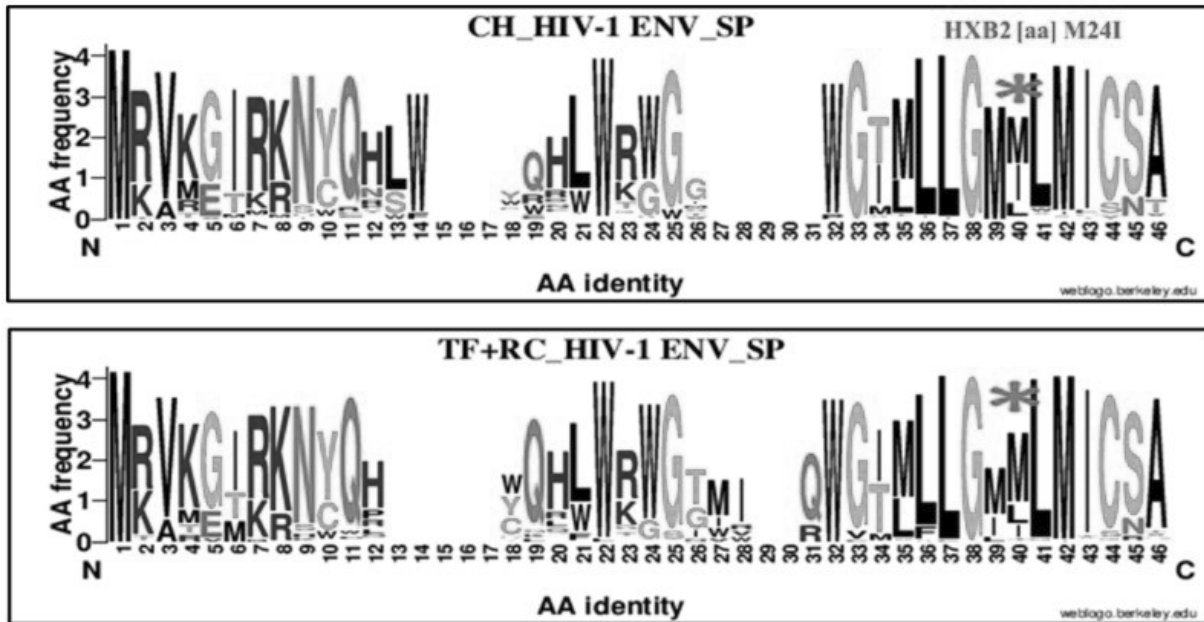


Figure XXXVII. (Article 3, Figure 8). Genetic signature identified under the HIV-1 envelope signal peptide (SP) associate to clade B HIV-1 chronic compared to recent viruses.

The X axis represents amino acids (AA) composing the *env* SP sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the SP sequence for each category of infection. The top line box represents chronic virus envelope SP sequences (N = 105) and the bottom for RC viruses within 136 MDRI and TF (N = 126). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The isoleucine (I) amino signature was located at position 40 of alignment (HXB2 position M24I) and is identified by a red asterisk.

The third signature was located at position 621 of HIV-1 *env* GP41 loop domain. It consisted of glutamine (Q) substitution by aspartic acid (D; Figure 9). The aspartic acid was enriched at 15.38% (16/104) for CH and 41.60%, (52/125) for RC (+TF), OR = 0.25%, 95% CI (0.13–0.48), p = 0.00001 using the Chi2 test.

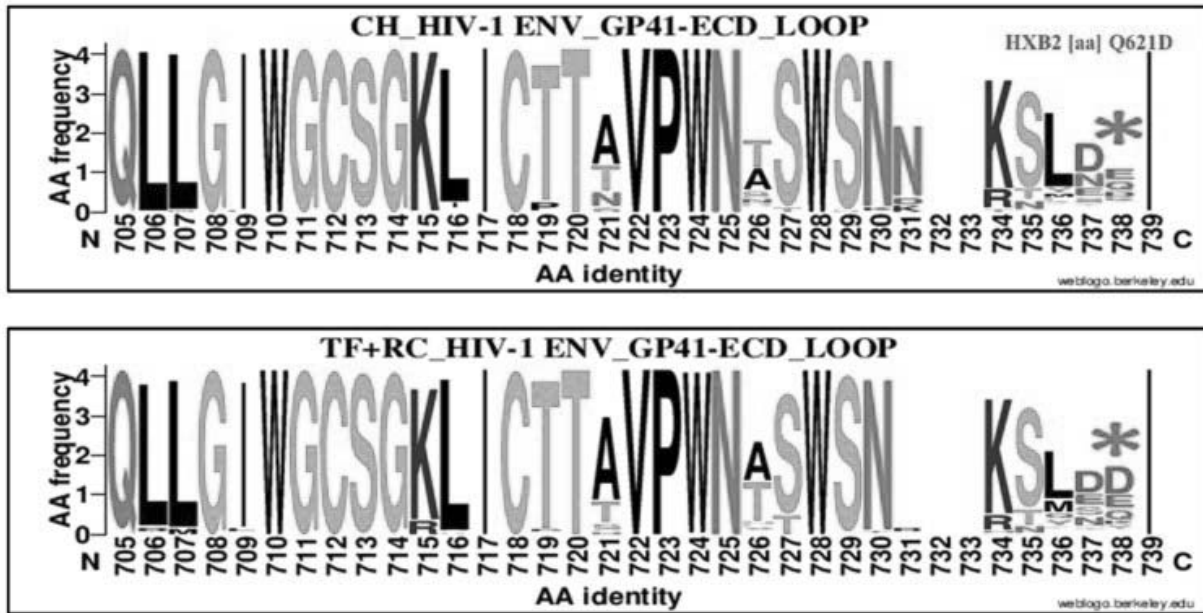


Figure XXXVIII. (Article 3, Figure 9). Genetic signature identified under the HIV-1 envelope GP41 cytoplasmic tail (CT or CD) loop domain associate to clade B HIV-1 chronic compared to recent viruses.

The X axis represents amino acids (AA) composing the GP41 CT loop sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the GP41 CT loop sequence for each category of infection. The top line box represents chronic virus envelope GP41 CT loop sequences represented individuals (N = 105) and the bottom for RC viruses within 136 MDRI and TF (N = 126). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The aspartic acid (D) amino signature was located at position 738 of alignment (HXB2 position Q621D) and identified by a red asterisk.

The last amino acid mutation patterns that distinguish chronic from recent viruses were located in the HIV-1 *env* GP 41 cytoplasmic tail specifically at position 751 (Figure 10). Specifically, they were located between the NF- κ B activation (NA) and the highly immunogenic region, also called Kennedy Epitope (KE; Figure 10). It consisted of an aspartic acid (D) substitution by a valine (V; D751V). The valine was enriched at 60.95% (64/105) for CH and 31.20% (39/125) for RC (+TF), OR = 0.25%, 95% CI (1.99–5.92), p = 0.00001 using the Chi2 test.

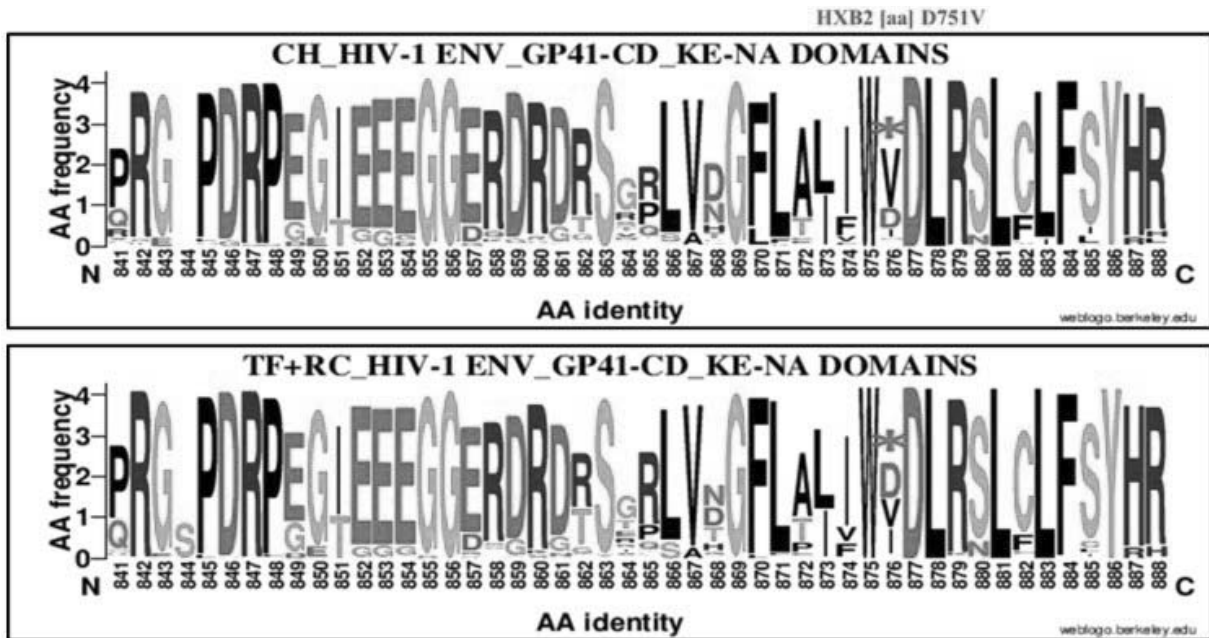


Figure XXXIX. (Article 3, Figure 10). Genetic signature identified under the HIV-1 envelope GP41 cytoplasmic tail (CT or CD) between NF- κ B activation (NA) and Kennedy Epitope (KE) domains associate to clade B HIV-1 chronic compared to recent viruses.

The X axis represents amino acids (AA) composing the GP41 CT NA and KE sequence (direction N to C). The Y axis represents the normalized AA frequency identified at each position of the GP41 CT, NA and KE sequence for each category of infection. The top line box represents chronic virus envelope GP41 CT NA and KE sequences represented individuals (N = 105) and the bottom for RC viruses within 136 MDRI and TF (N = 126). As indicated for Weblogo analysis, the overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding amino or nucleic acid at that position. The valine (V) signature was located at position 876 of alignment and position 751 referring to HXB2 numbering (D751V) and identified by a red asterisk.

The complete profile of statistically significant clade B HIV-1 envelope amino acid genetic signatures that distinguish recently from chronic viruses is summarized in Table 2.

Tableau XIV. Table 2. Major amino acid signatures among HIV-1 envelope sequences between chronic and recent viruses.

HXB2		Env		Amino Acid		CH			TF + RC			Chi ² test		
Position	Amino acid	Subregion/Domain	Alignment position	Change	Genetic signature	YES	NO	TOTAL	YES	NO	TOTAL	OR	95% CI	B-H Adjusted p value
153	E	V1	206	E	I53E	93	11	104	82	43	125	4.43	2.16, 9.05	0.00001
24	M	SP	40	I	M24I	38	66	104	16	109	125	3.92	2.04, 7.53	0.00001
621	Q	GP 41 CT Loop	738	D	Q621D	16	88	104	52	73	125	0.25	0.13, 0.48	0.00001
751	D	CT (HIR/KE)	876	V	D751V	64	41	105	39	86	125	3.44	1.99, 5.92	0.00001
6	K	SP	6	I	K6I	83	22	105	69	56	125	3.06	1.70, 5.48	0.0003
33	K	C1	49	Q	K33Q	39	55	94	22	102	124	3.28	1.78, 6.06	0.0003
717	F	EC (YSPL) and HIR/KE	834	F	F717F	86	19	105	73	52	125	3.22	1.75, 5	0.0003
717	F	EC (YSPL) and HIR/KE	834	L	F717L	19	86	105	52	73	125	0.31	0.16, 0.56	0.0003
747	R	HIR/KE NA	865	R	R747R	45	60	105	85	40	125	0.35	0.20, 0.60	0.0003
132	T	V2	151	T	T132T	86	19	105	75	51	126	3.07	1.67, 5.64	0.0004
154	I	V1	207	M	I154M	35	70	105	73	53	126	0.36	0.21, 0.62	0.0004
360	I	C3	441	V	I360V	52	52	104	33	93	126	2.81	1.62, 4.88	0.0004
737	R	HIR/KE	862	T	R737T	15	90	105	44	81	125	0.3	0.16, 0.58	0.0006
841	R	LLP-1	959	I	R841I	10	95	105	35	87	122	0.26	0.12, 0.55	0.0006
744	R	HIR/KE	862	R	R744R	78	27	105	65	60	125	2.66	1.52, 4.65	0.0009
236	T	Loop D	312	S	T236T	9	95	104	33	93	126	0.26	0.12, 0.58	0.001

Table 2 presents the statistically significant clade B HIV-1 envelope sequence polymorphisms among chronic compared to recent viruses of individuals HIV-1 infected that are considered as genetic signatures. The bold typeface defines the important *env* genetic signatures identified and discussed in manuscript. The gray character referred to sequence alignment position of amino acid changes. Abbreviations: V1—variable loop 1, CP—cytoplasmic domain/tails, EC—endocytosis domain, SP—Signal peptides, C—constant domain 2, C3—constant domain 3, ECD-Loop—Ectodomain-loop region, CP-KE—cytoplasmic domain-Kennedy epitope HIR-KE— Highly Immunogenic region (HIR) also called Kennedy epitope. ECD-CHR—Ectodomain-C-hepta-repeat, CP-LLP-1—Cytoplasmic tail-lentivirus lytic peptide 1.

4. DISCUSSIONS

The main objective of the current study was to determine the characteristics of the clade B HIV-1 envelope variable loop in terms of sequence length, number of N-glycosylation sites and net charge. It also aimed at identifying the principal amino acid signatures associated with TF and RC founder virus strains compared to chronic viruses. The TF and RC HIV-1 viruses envelope glycoprotein mutation patterns determine the success of viral transmission and its evolution during HIV-1 infections. Identifying such genetic signatures may help improve HIV-1 prevention and inform vaccine design. The current study included 103 untreated HIV-1 clade B HIV-1 consensus envelope sequences from different cohorts available in the Los Alamos sequence databases (Figure 1), in addition to two sequences derived from LSPQ serobank chronically infections. To limit the selection bias in LANL sequences, we carefully identified consensus sequences (one/patient) from clearly untreated chronically HIV-1-infected individuals from the North America region (the United States of America and Canada) that had been previously included in published articles [1,9]. We failed to obtain more HIV-1 chronic envelope sequences from all study participants derived from LSPQ serobank collections in order to make comparisons between TF and CH derived in the same context. This was due to the lower amplification success rate obtained in this study for those samples. Multiple factors may have affected HIV-1 envelope amplification success rate including sample quality such as the long-term storage, viral RNA extraction procedures, primers and enzymes as well as the viral loads of infected individuals (VL < 20,000 copies/ml). Depending on the length the HIV-1 genome to be amplified and specifically for *env* gene, the procedure is known to be challenging [9,50,51,52].

4.1. Clade B HIV-1 Envelope Variable Loop Characteristics

The first objective of the current study was to characterize HIV-1 TF viruses envelope variable regions, which include the V1/V2, V3, V4 and V5 loop lengths, their number of N-linked glycosylation sites and the V3 loop positive net charge.

Our results show that the V3 loop numbers of N-glycosylation sites of TF viruses were significantly less glycosylated than the chronic ones (Figure 2 and Table S5). The *env* V3 loop of TF viruses were less positively charged than chronic viruses. (Figure 4 and Table S5). This observation confirms earlier findings of a decreased positive net charge of TF viruses V3 loop sequences compared to chronic [53,54,55,56,57,58]. The positive net charge of HIV-1 envelope hypervariable loop three modulated the viral phenotype and tropism [59] at different stages of infection. The lower decreased charge of TF viruses V3 loop may constitute a regulating factor of viral phenotype during transmission.

In this study, The V1/V2 loop length and number of N-glycosylation sites did not differ between TF and CH HIV-1 viruses envelopes identified by earlier studies [60]. A shorter V1/V2 length and a fewer number of N-glycosylation sites have been associated to TF viruses in previous studies [60]. Most of these characteristics have been observed for clades A, C and D of HIV-1 [1,14,58,61]. This could reflect a difference among clades, as our study compared clade B HIV-1 *env* V1 and V2 loops [14,57,58].

We also observed that HIV-1 *env* GP120 loop 5 (V5) length of TF viruses was significantly shorter than RC and CH viruses (Figure 3 and Table S1). The V5 loop has been found to be necessary for viral structure integrity maintenance, negatively affected virus assembly and virus entry [62] and constituted neutralizing determinants recognized by broadly neutralizing monoclonal antibodies [63,64]. It also participates in CD4 binding sites (CD4bs) formation [65]. The shorter loop length of TF compared to RC and CH viruses suggest that V5 sequence loop length modelling at the acute stage of infection plays an important role in the virus transmission process and subsequently to the disease progression.

4.2. Clade B HIV-1 Envelope Amino Acids Signatures Associate to Transmitted/Founder and Recent Viruses

The second objective of this study was to identify specific mutation patterns across the HIV-1 envelope that may be considered as a genetic signature of TF viruses. We first compared the TF viruses envelope sequences derived of acutely HIV-1-infected individuals (Fiebig stages 1 to 2) [66,67] with those from chronically infected individuals. The first important point mutation identified consisted of the substitution of an arginine (R) by an isoleucine (I) at

position 959 of alignment and referred to HXB2 numbering to position 841 (R841I; Figure 5). This mutation was located in the C-terminal of the cytoplasmic (CT/D), specifically in the lentivirus lytic peptide segment 1 (LLP-1) [68,69]. The cytoplasmic tail or domain of GP41 is important for HIV-1 replication and pathogenesis by regulating rapid clathrin-mediated endocytosis that induces low levels of *env* expression on the cell surface [30,70,71]. This phenomenon contributes to limiting humoral immune pressure to HIV-1 [30]. It is also known that GP41 CT contributes to *env* incorporation into virions by interacting with viral matrix protein [37] and also for cellular-transcription factor NF- κ B activation [30]. The CTs of HIV-1 of GP41 have also been shown to have an impact on gp120 and ECD conformation and mutations in this domain also impacted recognition and neutralization of antibodies [15,32,72].

The R841I signature associated to clade B HIV-1 TF viruses was located at position 841 of the GP41 cytoplasmic tail in this LLP-1. It was reported that the LLP-1 mutations affect *env* association with lipid rafts [31,70] and reduce *env* incorporation, infectivity and the replication process for certain viral phenotypes [73]. The results of our current study reveal the selection of isoleucine by TF viruses (Table 1), which may contribute to HIV-1 gp41 CT functions. Earlier studies have highlighted the importance of gp41 cytoplasmic tail domain (CD) in HIV-1 in transmission and pathogenesis [30]. Lee, S. F. et al. (2002) previously demonstrated that a single deletion of one of the two adjacent valine residues located at positions 832 and 833 and Ile-830, Ala-836 and Ile-840 significantly contributed to the reduction of *env* steady-state expression [74]. The R841I substitution identified in the acute stage of infection of our study may constitute a key factor to enhance LLP-1 functions. It would be necessary to evaluate its functional implications in the HIV-1 transmission process and viral replication.

The second important amino acid signature identified was K6I (within the signal peptide), highly enriched in chronic vs. TF viruses (Figure 6, Table 1). The lower selection of this mutation for the TF viruses may represent a strategy for virus resistance to early immune responses. Gnanakaran S et al. 2011 [1] showed that a histidine signature at position 12 (H12) in the signal peptide was highly enriched in TF viruses compared to chronic HIV-1

envelope sequences. The histidine amino acids that were normally located at position 12 were substituted by arginine (R) or proline (P) during acute infection [1,75]. This H12 signature was found to increase envelope incorporation in pseudo-viruses in vitro [75,76]. The current study identified and highly selection of the isoleucine (K6I) associated to chronic compared to TF viruses (Figure 6, Table 1). This mutation constituted an amino acid signature enriched during disease progression over chronic infection.

The HIV-1 envelope signal peptide plays an important role in virus interaction with host cells during transmission and its evolution toward the chronic stage. It contributes to increased *env* gp120 transport and the secretion and expression of *env* on the cell membrane surface [77]. As reported by previous studies, a natural variation in the N-terminal signal peptide (SP) of the HIV envelope significantly impacts the antigenicity and molecular mass of mature gp120 and its glycosylation and interaction with DC-SIGN [78]. The SP is also likely subjected to antibody-mediated immune pressure [77]. Compared to the Gnanakaran et al. (2011) study, phylogenetic analysis methods and the numbers of sequence datasets used for amino acid signature estimate may explain the different results. The Gnanakaran et al. study used consensus and corrected phylogenetic tree analyses [1,76] for amino acid signature estimates, whereas our study used the WebLogo online-based application to map and determine the amino acid estimates. In addition, the definition of the TF viruses was also different. The current study considered TF viruses as the sequences of HIV-1-infected individuals sampled during acute infection Fiebig 1 to 2 stages [40,41], whereas Gnanakaran et al. (2011) considered as recent viruses those identified at an early stage of HIV-1 infection covering the Fiebig stages 2 to 5 [1,76].

We believed that identifying the HIV-1 envelope genetic signature very early after infection might lead to a better identification of important genetic polymorphism observed during disease transmission. Therefore, we compared chronic virus envelope sequences with those of TF and recent ones to determine a specific mutation pattern. Four significant HIV-1 envelopes amino acid signatures were identified (Figure 7, Figure 8, Figure 9 and Figure 10 and Table 2). Three were highly associated to chronic viruses and one to recent ones. The mutation patterns associated with chronic viruses consisted of glutamic acid (E) at position

153 in the V1 loop (153E), a methionine (M) substitution by isoleucine at position 24 in the signal peptide (M24I) and aspartic acid (D) substitution by a valine (V) in the cytoplasmic tail (D751V; Figure 7, Figure 9 and Figure 10 and Table 2). The amino acid signature associated with recent and TF viruses together was located in the GP41 ectodomain, specifically in the loop domain (Figure 9 and Table 2). It consisted of glutamine (Q) substitution by aspartic acid (D) at position 621 (Q621D). The HIV-1 envelope genetic signatures identified for chronic and recent viruses constituted results of accumulated mutations during disease progression.

In summary, we identified an important HIV-1 envelope amino acid genetic signature associate to the GP41 cytoplasmic tail, specifically in the lentivirus lytic peptides associated TF compared to chronic viruses. It would be interesting to conduct phenotypic studies to further evaluate the role of isoleucine substitutions in viral *env* function (R841I) including other frequent mutations. Overall, this study provided new evidence related to genetic characteristics of HIV-1 envelope sequences associate with clade B TF, RC and CH viruses. As other genetic analysis, the Weblogo would have been impacted results of sequence genetic profiles if different groupings were considered. Careful verification to ensure that all sequences have the same lengths and do not contain gaps could be considered.

5. CONCLUSIONS

The current study identified the presence of different point mutations patterns in the HIV-1 envelope, specifically in the GP41 cytoplasmic tail lentivirus lytic peptide segment 1 (gp41 CT_LLP-1) significantly associated with TF viruses. The LLP-1 domains of GP41 CT play an important role in the virus replication and pathogenesis. The R841I mutation identified in this segment may be considered as a specific genetic signature, as well as its phenotypic properties during HIV-1 transmission merits further study. The HIV-1 transmission is complex, multifactorial and this mutation profile identified could not be the only contributing factors to the disease transmission. But understanding and identifying such early envelope molecular determinants may provide clues for the design of an HIV vaccine.

Acknowledgments

We are grateful to all lab technicians of the molecular biology, serodiagnosis and virology departments at “the laboratoire de santé publique Québec” (LSPQ) of the “Institut national de santé publique du Québec” (INSPQ) Montreal, Canada, for technical support. We would also like to thank Lyne Désautels, Sylvie-Nancy Beaulac and Lise Leblanc and colleagues. We are grateful to Diane Sylvain and Maureen Hastie of the provincial HIV surveillance program of Quebec at the INSPQ, Quebec, Canada, for providing patients’ epidemiological and clinical data. We thank Réjean Dion (médecin-conseil) at the LSPQ/INSPQ, Sia Drissa at the “Université du Québec en Outaouais”, Quebec, Canada, for statistical analysis advising. We also thank Art FY. Poon, PhD. assistant professor and researchers in viral evolution at the Department of Pathology and Laboratory Medicine, University of Western Ontario, Canada, Nicolas Chomont, PhD. Associate research professor in the Department of Microbiology, Infectiology and Immunology at the University of Montreal. and at the University of Montreal Hospital Research Centre (CRCHUM), Quebec, Canada and Paul Wankah, MD, at the University of Sherbrooke, Québec, Canada; for their precious advising and revising this manuscript. Finally, we thank Andrés Finzi and Petronela Ancuta at the CRCHUM and University of Montreal and Guy Lemay at the department of microbiology, infectiology and

immunology of the University of Montreal for advisement on study conception. We also acknowledge the HIV-1-infected patients of Quebec from whom serum samples were obtained in the present study.

Supplementary Materials

The following are available online at www.mdpi.com/xxx/s1, Table S1: Defining of the different timeline categories of HIV-1 infection status and referred nomenclatures reported in manuscript, Table S2: Limit of HIV-1 envelope sub-regions and domains referred to HXB2 *env* gp160 sequence numbering, Table S3: HIV-1 TF viruses envelope sequences data qualifiers, Table S4: Background information of LANL clade B HIV-1 chronic viruses envelope sequences, Table S5: Descriptive statistics of HIV-1 envelope variable region characteristics including the number of N-glycosylation sites, the loop lengths and the V3 loop net charge.

Author Contributions

C.L.T. H.C. A.K. (Alexis Kafando), M.E.-F. B.S. C.M. F.D.-B and. A.C.: contributed to the conception and study design; A.K. (Alexis Kafando), E.F. B.S. C.M. F.D.-B. M.S. A.C. H.C. contributed to data acquisition, analysis and interpretation; A.K. (Alexis Kafando), E.F. A.K. (Adama Kazienga) M.N.S. B.S. C.M. F.D.-B. H.C.: contributed to the bioinformatics and statistical analysis of the data. All authors contributed to drafting the work or critically revising the manuscript for important intellectual content and provided approval of the final version to be published.

Funding

This study was funded by the Islamic Development Bank, Jeddah, Saudi Arabia (Grant number: 600014438), the AIDS and Infectious Disease Network (FRQS), Quebec, Canada, and Genome Canada. Alexis Kafando received PhD scholarships of (1) Islamic Development Bank Merit Scholarship Programme for High Technology, for 3 Year PhD. (2013–2016), ID: 600014438, Jeddah, Saudi Arabia. (2) A bourse d'exemption des droits de scolarité supplémentaires pour étudiants étrangers of the Université de Montréal, Montréal, Québec, Canada. (3) A bourse de fin d'études doctorales of the Faculté des Études supérieures et postdoctorales (FESP) of the Université de Montréal, Montréal, Québec, Canada. (4) A bourse

d'étude of Dre Tremblay's laboratory at the University of Montreal Hospital Research Centre (CRCHUM) funded by the (FRQS RESEAU SIDA-MI), Québec, Canada. (5) Kafando is also a beneficiary of the "Programme de prêts et bourses du ministère de l'Éducation et de l'enseignement supérieur du gouvernement du Québec, Canada" and (6) he is an employee of the Centre Muraz biomedical research center, Ministry of Health, Burkina Faso. Cécile L. Tremblay, corresponding author, is a Director of the Pfizer/University of Montreal Chair on Translational HIV Research, Director of the AIDS and Infectious Disease Network of the Fonds de recherche du Québec–Santé (FRQS), Canada. Cecile Tremblay's laboratory at University of Montreal Hospital Research Centre (CRCHUM) was funded by FRQS, Quebec, Canada.

Conflicts of Interest

All authors have no competing interests to declare.

6. ANNEXES (ARTICLE 3)

Supplementary Materials

Tableau XV. (Annexe 1, Table S1, Article 3). Samples description: timeline categories of HIV-1 infection.

Table S1: Samples description: defining timeline categories of HIV-1 infection and referred nomenclatures.					
Infection category	Virus type	Fiebig stages	Infection duration	Biomarkers	Sample deriv from
Acute infection	Transmitted/founder viruses (TF)	Fieb. 1-2	14-21 days	HIV-RNA+ and p24 Ag+, Western blot -, HIV antibody-	Acutely infect individuals
Early infection	Recent viruses (RC)	Fieb. 3-5	≤ 136 days according RITA testing, <i>Sherir B. et al. 2016.</i>	HIV p24 Ag-, Western blot +, HIV antibody +, qualified as recent by RITA testing	Recently infect individuals
Chronic infection	Chronic viruses (CH)	Established	> 6 months according RITA testing, <i>Sherir B. al. 2016.</i>	HIV antibody +	Chronically infected individuals

Table S1 presents the defining timeline categories of HIV-1 status and referred nomenclatures frequently used in manuscript. Abbreviations RITA Recent infection testing algorithm; Ag Antigen; (-) negative, (+); positive; WB Western Blot; Fieb Fiebig stage

Tableau XVI. (Annexe 2, Table S2, Article 3). Summary of the HIV-1 subregions and domains locations according HXB2 envelope (gp160) numbering.

Table S2: Summary of the HIV-1 envelope subregions and domains locations according HXB2 envelope (gp160) numbering.

Order	Envelope subdomain	HXB2 Position	Alignment position	Sequence length
	GP120	1-811	1-627	512
1	SP	M1-A30	1-42	42
2	C1	T31-K130	43-145	103
3	V1	C131-C157	146-204	59
4	V2	S126-C196	141-270	130
5	C2	N197-N295	271-370	100
6	Loop D	V275-T283	350-358	9
7	V3	C294-C331	371-408	38
8	C3	N332-Y384	409-466	58
9	V4	C385-C418	467-520	54
10	CD4 binding loop	S364-H374	445-456	12
11	C4	R419-G459	521-561	41
12	V5	N460-R469	566-577	12
13	C5	P470-R511	578-627	50
	GP41	512-856	628-977	345
1	ECD	A512-N677	628-797	170
	FP	A512-G527	628-644	17
	FPPR	T529-L545	646-662	17
	NHR	S546-L581	663-698	36
	PFD	A582-Q590	699-707	9
	Loop	Q591-I622	708-742	35
	CHR	W623-E659	743-779	37
	MPER	W673-F683	795-805	11
2	TMD or MSD	N677-S 716	N794-S833	40
3	CD	N706-L856	N832-977	151
	EC (GYSPL)	G711-L715	832-835	5
	HIR/KE (P-S)	P724-S745	844-866	25
	NA (C-L)	C764-R770	885-891	7
	LLP-2	Y768-R788	889-909	21
	LLP-3	G789-L815	910-936	27
	LLP-1 (R-E)	R828-S55L	949-976	28
	EC (LL)	L855-L856	976-977	2

Abbreviations. GP: Glycoproteins; ECD: Ectodomain; TMD: Transmembrane domain, CD: Cytoplasmic domain; MSD: membrane-spanning domain; EC: endocytosis; HIR: highly immunogenic region; KE: Kennedy epitope, NA: NF- κ B activation; LLP: lentivirus lytic peptide, FPPR: Fusion peptide proximal region Refs [15, 37, 41, 42, 44, 61-63, 81]

Tableau XVII. (Annexe 3, Table S5, Article 3). Descriptive statistics of HIV-1 variable regions characteristics.

Table S5. Descriptive statistics of HIV-1 envelope variable regions characteristics.

Parameters	Infection status			Logistic Regression analysis							
	CH (N=105)	RC (N=28)	TF (N=98)	CH vs RC		CH vs TF		RC vs TF		CH vs RC vs TF	
				OR	P	OR	P	OR	P	OR	P
HIV-1 Env variables regions characteristics											
V1 Glyco Num, median (Range)	4 (4,5)	4 (4,6)	4 (4,5)	0.6543105	0.165	0.9486806	0.796	1.507161	0.195		
V2 Glyco Num, median (Range)	2 (2,2)	2 (2,2.5)	2 (2,2)	0.8293438	0.655	0.5753516	0.064	0.7065884	0.437		
V1+V2 Glyco Num, median (Range)	7 (6,7)	7 (6,7.5)	7 (6,7)	0.7234489	0.248	0.758415	0.151	1.080748	0.793		
V3 Glyco Num, median (Range)	2 (2,2)	2 (2,3)	2 (2,2)	2.032284	0.057	0.5856403	0.026	0.3794798	0.004		
V4 Glyco Num, median (Range)	5 (4,5)	4 (4,5)	5 (4,5)	1.128342	0.681	1.148025	0.457	1.026063	0.926		
V5 Glyco Num, median (Range)	2 (1,2)	2 (1.5, 2)	2 (1, 2)	1.09692	0.824	0.9732193	0.923	0.8946218	0.783		
V1 Length, median (Range)	30 (28, 34)	29.5 (26.5, 38.5)	31 (27, 34)	1.044247	0.212	0.9907996	0.729	0.9538506	0.167		
V2 Length, median (range)	43(41, 46)	42.5 (41, 46)	42 (40, 45)	0.9889286	0.846	0.9732534	0.453	0.9834442	0.756		
V1+V2 length, median (Range)	72 (68, 77)	73 (66.5, 80)	72 (67, 76)	1.032432	0.309	0.9837571	0.458	0.9638805	0.188		
V3 Length, median (Range)	37 (37, 37)	37 (37, 37)	37 (37, 37)	1.129842	0.695	1.215851	0.543	1.186667	0.808		
V4 Length, median (Range)	31 (29, 33)	32 (30,34)	31.5 (30, 33)	1.099581	0.158	1.050612	0.317	0.9167963	0.280		
V5 Length, median (Range)	15 (12, 14)	15 (13, 15)	13 (12, 14)	1.44851	0.004	0.9790804	0.813	0.6537583	0.003		
V3 net Charge, median (Range)	5 (3, 6)	4 (3, 5.5)	4 (3, 5)	0.9419158	0.640	0.8264031	0.038	0.8718934	0.347	-1.655425	0.040

Table S5 presents the summary statistics of HIV-1 envelope variables regions characteristics (numbers of N-glycosylation sites, loop lengths and V3 net charge for transmitted/founder (TF) viruses compared to recent (RC) and chronic (CH) ones. Statistically significant results were shown in red in table.

7. REFERENCES

1. Gnanakaran S., Bhattacharya T., Daniels M., Keele B.F., Hraber P.T., Lapedes A.S., Shen T., Gaschen B., Krishnamoorthy M., Li H., et al. Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *PLoS Pathog.* 2011;7:e1002209.
2. Shaw T.I., Zhang M. HIV N-linked glycosylation site analyzer and its further usage in anchored alignment. *Nucleic Acids Res.* 2013;41:W454–W458.
3. Benjelloun F., Lawrence P., Verrier B., Genin C., Paul S. Role of Human Immunodeficiency Virus Type 1 Envelope Structure in the Induction of Broadly Neutralizing Antibodies. *J. Virol.* 2012;86:13152–13163.
4. Arrildt K.T., Labranche C.C., Joseph S.B., Dukhovlina E.N., Graham W.D., Ping L.H., Schnell G., Sturdevant C.B., Kincer L.P., Mallewa M., et al. Phenotypic Correlates of HIV-1 Macrophage Tropism. *J. Virol.* 2015;89:11294–11311.
5. Pierson T., McArthur J., Siliciano R.F. Reservoirs for HIV-1: Mechanisms for Viral Persistence in the Presence of Antiviral Immune Responses and Antiretroviral Therapy. *Annu. Rev. Immunol.* 2000;18:665–708.
6. Finzi D., Blankson J., Siliciano J.D., Margolick J.B., Chadwick K., Pierson T., Smith K., Lisziewicz J., Lori F., Flexner C., et al. Latent infection of CD4(+) T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* 1999;5:512–517.
7. Van Regenmortel M.H.V. Development of a Preventive HIV Vaccine Requires Solving Inverse Problems Which Is Unattainable by Rational Vaccine Design. *Front. Immunol.* 2018;8:2009.

8. Ensoli B., Cafaro A., Monini P., Marcotullio S., Ensoli F. Cna Menges in HIV vaccine research for treatment and prevention. *Front. Immunol.* 2014;5:11.
9. Keele B.F., Giorgi E.E., Salazar-Gonzalez J.F., Decker J.M., Pham K.T., Salazar M.G., Sun C., Grayson T., Wang S., Li H., et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA.* 2008;105:7552–7557.
10. Joseph S.B., Swanstrom R., Kashuba A.D.M., Cohen M.S. Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. *Nat. Rev. Genet.* 2015;13:414–425.
11. Kariuki S.M., Selhorst P., Ariën K.K., Dorfman J.R. The HIV-1 transmission bottleneck. *Retrovirology.* 2017;14:22.
12. Shaw G.M., Hunter E. HIV transmission. *Cold Spring Harb. Perspect. Med.* 2012;2:a006965.
13. Bar K.J., Li H., Chamberland A., Tremblay C., Routy J.P., Grayson T., Sun C., Wang S., Learn G.H., Morgan C.J., et al. Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users. *J. Virol.* 2010;84:6241–6247.
14. Frost S.D.W., Liu Y., Pond S.L.K., Chappey C., Wrin T., Petropoulos C.J., Little S.J., Richman D.D. Characterization of Human Immunodeficiency Virus Type 1 (HIV-1) Envelope Variation and Neutralizing Antibody Responses during Transmission of HIV-1 Subtype B. *J. Virol.* 2005;79:6523–6527.
15. Checkley M.A., Luttge B.G., Freed E.O. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J. Mol. Biol.* 2011;410:582–608.

16. Upadhyay C., Feyznehzad R., Yang W., Zhang H., Zolla-Pazner S., Hioe C.E. Alterations of HIV-1 envelope phenotype and antibody-mediated neutralization by signal peptide mutations. *PLoS Pathog.* 2018;14:e1006812.
17. Freed E.O. HIV-1 replication. *Somat. Cell Mol. Genet.* 2001;26:13–33.
18. Moore J.P., Willey R.L., Lewis G.K., Robinson J., Sodroski J. Immunological evidence for interactions between the first, 2nd, and 5th conserved domains of the gp120 surface glycoprotein of human-immunodeficiency-virus type-1. *J. Virol.* 1994;68:6836–6847.
19. Starcich B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S., Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C., et al. Identification and characterization of conserved and variable regions in the envelope gene of htlv-iii lav, the retrovirus of aids. *Cell.* 1986;45:637–648.
20. Douglas N., Munro G., Daniels R. HIV/SIV glycoproteins: Structure-function relationships. *J. Mol. Biol.* 1997;273:122–149.
21. Shang L., Yue L., Hunter E. Role of the Membrane-Spanning Domain of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein in Cell-Cell Fusion and Virus Infection. *J. Virol.* 2008;82:5417–5428.
22. Shang L., Hunter E. Residues in the membrane-spanning domain core modulates conformation and fusogenicity of the HIV-1 envelope glycoprotein. *Virology.* 2010;404:158–167.
23. Weiss C.D. HIV-1 gp41: Mediator of fusion and targets for inhibition. *AIDS Rev.* 2003;5:214–221.
24. Markovic I., Clouse K. Recent Advances in Understanding the Molecular Mechanisms of HIV-1 Entry and Fusion: Revisiting Current Targets and Considering New Options for Therapeutic Intervention. *Curr. HIV Res.* 2004;2:223–234.

25. Chen B., Chou J.J. Structure of the transmembrane domain of HIV-1 envelope glycoprotein. *FEBS J.* 2017;284:1171–1177.
26. Haffar O.K., Dowbenko D.J., Berman W. Topogenic analysis of the human immunodeficiency virus type-1 envelope glycoprotein, gp160, in microsomal-membranes. *J. Cell Biol.* 1988;107:1677–1687.
27. Berman W., Nunes W.M., Haffar O.K. Expression of membrane-associated and secreted variants of gp160 of the human immunodeficiency virus type-1 in vitro and in continuous cell-lines. *J. Virol.* 1988;62:3135–3142.
28. Salzwedel K., Johnston P.B., Roberts S.J., Dubay J.W., Hunter E. Expression and characterization of glycopospholipid-anchored human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* 1993;67:5279–5288.
29. Env Feature database: HXB2 Genome Annotation. [(accessed on 30 October 2019)];2017 Available online: www.hiv.lanl.gov.
30. Postler T.S., Desrosiers R.C. The tale of the long tail: The cytoplasmic domain of HIV-1 gp41. *J. Virol.* 2013;87:2–15.
31. Yang P., Ai L.S., Huang S.C., Li H.F., Chan W.E., Chang C.W., Ko C.Y., Chen S.S. The Cytoplasmic Domain of Human Immunodeficiency Virus Type 1 Transmembrane Protein gp41 Harbors Lipid Raft Association Determinants. *J. Virol.* 2010;84:59–75.
32. Edwards T.G., Wyss S., Reeves J.D., Zolla-Pazner S., Hoxie J.A., Doms R.W., Baribaud F. Truncation of the Cytoplasmic Domain Induces Exposure of Conserved Regions in the Ectodomain of Human Immunodeficiency Virus Type 1 Envelope Protein. *J. Virol.* 2002;76:2683–2691.

33. Bültmann A., Muranyi W., Seed B., Haas J. Identification of Two Sequences in the Cytoplasmic Tail of the Human Immunodeficiency Virus Type 1 Envelope Glycoprotein That Inhibit Cell Surface Expression. *J. Virol.* 2001;75:5263–5276.
34. Murakami T., Freed E.O. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc. Natl. Acad. Sci. USA.* 2000;97:343–348.
35. Piller S.C., Dubay J.W., Derdeyn C.A., Hunter E. Mutational Analysis of Conserved Domains within the Cytoplasmic Tail of gp41 from Human Immunodeficiency Virus Type 1: Effects on Glycoprotein Incorporation and Infectivity. *J. Virol.* 2000;74:11717–11723.
36. Jiang J., Aiken C. Maturation-Dependent Human Immunodeficiency Virus Type 1 Particle Fusion Requires a Carboxyl-Terminal Region of the gp41 Cytoplasmic Tail. *J. Virol.* 2007;81:9999–10008.
37. Kalia V., Sarkar S., Gupta P., Montelaro R.C. Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. *J. Virol.* 2003;77:3634–3646.
38. Cohen M.S., Shaw G.M., McMichael A.J., Haynes B.F. Acute HIV-1 Infection. *N. Engl. J. Med.* 2011;364:1943–1954.
39. Serhir B., Hamel D., Doualla-Bell F., Routy J.P., Beaulac S.-N., Legault M., Fauvel M., Tremblay C. Quebec Primary HIV infection study group Performance of Bio-Rad and Limiting Antigen Avidity Assays in Detecting Recent HIV Infections Using the Quebec Primary HIV-1 Infection Cohort. *PLoS ONE.* 2016;11:e0156023.
40. Brian R., Wood M., David H., Spach M.D. Acute and Recent HIV Infection. Section 1: Screening and Diagnosis, Topic 4: Acute and Recent HIV Infection 2019 April 24th.

[(Accessed on 8 July 2019)]; 2019 Available online: <https://www.hiv.uw.edu/go/screening-diagnosis/acute-recent-early-hiv/core-concept/all#tables>.

41. Panel on Antiretroviral Guidelines for Adults and Adolescents. National Institute of Health (NIH), Department of Health and Human Services. AIDSinfo; 2018. [(accessed on 3 October 2019)]. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. Considerations for Antiretroviral Use in Special Patient Populations: Acute and Recent (early) HIV Infection. Available online: <https://aidsinfo.nih.gov/guidelines>.
42. Revilla A., Delgado E., Christian E.C., Dalrymple J., Vega Y., Carrera C., González-Galeano M., Ocampo A., de Castro R.O., Lezaún M.J., et al. Construction and phenotypic characterization of HIV type 1 functional envelope clones of subtypes G and F. *AIDS Res. Hum. Retrovir.* 2011;27:889–901.
43. Shcherbakova N.S., Shalamova L.A., Delgado E., Fernández-García A., Vega Y., Karpenko L.I., Ilyichev A.A., Sokolov Y.V., Shcherbakov D.N., Pérez-Álvarez L., et al. Short communication: Molecular epidemiology, phylogeny, and phylodynamics of CRF63_02A1, a recently originated HIV-1 circulating recombinant form spreading in Siberia. *AIDS Res. Hum. Retrovir.* 2014;30:912–919.
44. Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 2016;33:1870–1874.
45. Schneider T.D., Stephens R. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* 1990;18:6097–6100.
46. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. WebLogo: A sequence logo generator. *Genome Res.* 2004;14:1188–1190.

47. Yi H.A., Diaz-Rohrer B., Saminathan P., Jacobs A. The Membrane Proximal External Regions of gp41 from HIV-1 Strains HXB2 and JRFL Have Different Sensitivities to Alanine Mutation. *Biochemistry*. 2015;54:1681–1693.
48. Yu F., Lu L., Du L., Zhu X., Debnath A.K., Jiang S. Approaches for Identification of HIV-1 Entry Inhibitors Targeting gp41 Pocket. *Viruses*. 2013;5:127–149.
49. Hunt R. Virology—Chapter Seven Part Nine Human Immunodeficiency Virus and Aids Structure: The Genome and Proteins of HIV. In: Hunt R.C., editor. *Microbiology and Immunology On-line*. University of South Carolina School of Medicine; Columbia, CA, USA: 2016.
50. Archary D., Gordon M.L., Green T.N., Coovadia H.M., Goulder P.J., Ndung’U T. HIV-1 subtype C envelope characteristics associated with divergent rates of chronic disease progression. *Retrovirology*. 2010;7:92.
51. McClure P., Curran R., Boneham S., Ball J.K. A polymerase chain reaction method for the amplification of full-length envelope genes of HIV-1 from DNA samples containing single molecules of HIV-1 provirus. *J. Virol. Methods*. 2000;88:73–80.
52. Cornelissen M., Gall A., Vink M., Zorgdrager F., Binter Š., Edwards S., Jurriaans S., Bakker M., Ong S.H., Gras L., et al. From clinical sample to complete genome: Comparing methods for the extraction of HIV-1 RNA for high-throughput deep sequencing. *Virus Res*. 2017;239:10–16.
53. Balasubramanian C., Chillemi G., Abbate I., Capobianchi M.R., Rozera G., Desideri A. Importance of V3 Loop Flexibility and Net Charge in the Context of Co-Receptor Recognition. A Molecular Dynamics Study on HIV gp120. *J. Biomol. Struct. Dyn*. 2012;29:879–891.
54. De Wolf F., Hogervorst E., Goudsmit J., Fenyö E.-M., Rübsamen-Waigmann H., Holmes H., Galvão-Castro B., Karita E., Wasi C., Sempala S., et al. Syncytium-Inducing and Non-

- Syncytium-Inducing Capacity of Human Immunodeficiency Virus Type 1 Subtypes Other Than B: Phenotypic and Genotypic Characteristics. *AIDS Res. Hum. Retrovir.* 1994;10:1387–1400.
55. Kaleebu P., Nankya I.L., Yirell D.L., Shafer L.A., Kyosiimire-Lugemwa J., Lule D.B., Morgan D., Beddows S., Weber J., Whitworth J.A. Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D—Results from a rural Ugandan cohort. *J. Acquir. Immune Defic. Syndr.* 2007;45:28–33.
56. Baalwa J., Wang S., Parrish N.F., Decker J.M., Keele B.F., Learn G.H., Yue L., Ruzagira E., Ssemwanga D., Kamali A., et al. Molecular identification, cloning and characterization of transmitted/founder HIV-1 subtype A, D and A/D infectious molecular clones. *Virology.* 2013;436:33–48.
57. Wilen C.B., Parrish N.F., Pfaff J.M., Decker J.M., Henning E.A., Haim H., Petersen J.E., Wojcechowskyj J.A., Sodroski J., Haynes B.F., et al. Phenotypic and Immunologic Comparison of Clade B Transmitted/Founder and Chronic HIV-1 Envelope Glycoproteins. *J. Virol.* 2011;85:8514–8527.
58. Finzi A., Pacheco B., Xiang S.-H., Pancera M., Herschhorn A., Wang L., Zeng X., Desormeaux A., Kwong P.D., Sodroski J. Lineage-Specific Differences between Human and Simian Immunodeficiency Virus Regulation of gp120 Trimer Association and CD4 Binding. *J. Virol.* 2012;86:8974–8986.
59. Wyatt R. The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science.* 1998;280:1884–1888.
60. Chohan B., Lang D., Sagar M., Korber B., Lavreys L., Richardson B., Overbaugh J. Selection for Human Immunodeficiency Virus Type 1 Envelope Glycosylation Variants with Shorter V1-V2 Loop Sequences Occurs during Transmission of Certain Genetic Subtypes and May Impact Viral RNA Levels. *J. Virol.* 2005;79:6528–6531.

61. Helseth E., Olshevsky U., Furman C., Sodroski J. Human-Immunodeficiency-Virus Type-1 Gp120 Envelope Glycoprotein Regions Important For Association With The Gp41 Transmembrane Glycoprotein. *J. Virol.* 1991;65:2119–2123.
62. Yuan T., Li J., Zhang M.-Y. HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry. *PLoS ONE.* 2013;8:e69789.
63. Pejchal R., Doores K.J., Walker L.M., Khayat R., Huang P.-S., Wang S.-K., Stanfield R.L., Julien J.-P., Ramos A., Crispin M., et al. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science.* 2011;334:1097–1103.
64. Zhou T., Georgiev I., Wu X., Yang Z.Y., Dai K., Finzi A., Kwon Y.D., Scheid J.F., Shi W., Xu L., et al. Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Sci.* 2010;329:811–817.
65. Zolla-Pazner S., Cardozo T. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nat. Rev. Immunol.* 2010;10:527–535.
66. Cohen M.S., Gay C.L., Busch M.P., Hecht F.M. The Detection of Acute HIV Infection. *J. Infect. Dis.* 2010;202:S270–S277.
67. Fiebig E.W., Wright D.J., Rawal B.D., E Garrett P., Schumacher R.T., Peddada L., Heldebrant C., Smith R., Conrad A., Kleinman S.H., et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: Implications for diagnosis and staging of primary HIV infection. *AIDS.* 2003;17:1871–1879.
68. Venable R.M., Pastor R.W., Brooks B.R., Carson F.W. Theoretically Determined Three-Dimensional Structures for Amphipathic Segments of the HIV-1 gp41 Envelope Protein. *AIDS Res. Hum. Retrovir.* 1989;5:7–22.

69. Eisenberg D., Wesson M. The most highly amphiphilic alpha-helices include two amino acid segments in human immunodeficiency virus glycoprotein 41. *Biopolymers*. 1990;29:171–177.
70. Fernandez M.V., Freed E.O. Meeting Review: 2018 International Workshop on Structure and Function of the Lentiviral gp41 Cytoplasmic Tail. *Viruses*. 2018;10:613.
71. Kao S.M., Miller E.D., Su L. A leucine zipper motif in the cytoplasmic domain of gp41 is required for HIV-1 replication and pathogenesis in vivo. *Virology*. 2001;289:208–217.
72. Kalia V., Sarkar S., Gupta P., Montelaro R.C. Antibody Neutralization Escape Mediated by Point Mutations in the Intracytoplasmic Tail of Human Immunodeficiency Virus Type 1 gp41. *J. Virol*. 2005;79:2097–2107.
73. Newman J.T., Sturgeon T.J., Gupta P., Montelaro R.C. Differential functional phenotypes of two primary HIV-1 strains resulting from homologous point mutations in the LLP domains of the envelope gp41 intracytoplasmic domain. *Virology*. 2007;367:102–116.
74. Lee S.F., Ko C.Y., Wang C.T., Chen S.S.L. Effect of Point Mutations in the N Terminus of the Lentivirus Lytic Peptide-1 Sequence of Human Immunodeficiency Virus Type 1 Transmembrane Protein gp41 on Env Stability. *J. Biol. Chem*. 2002;277:15363–15375.
75. Asmal M., Hellmann I., Liu W., Keele B.F., Perelson A.S., Bhattacharya T., Gnanakaran S., Daniels M., Haynes B.F., Korber B.T., et al. A Signature in HIV-1 Envelope Leader Peptide Associated with Transition from Acute to Chronic Infection Impacts Envelope Processing and Infectivity. *PLoS ONE*. 2011;6:e23673.
76. Gonzalez M.W., DeVico A.L., Lewis G.K., Spouge J.L. Conserved Molecular Signatures in gp120 Are Associated with the Genetic Bottleneck during Simian Immunodeficiency Virus (SIV), SIV-Human Immunodeficiency Virus (SHIV), and HIV Type 1 (HIV-1) Transmission. *J. Virol*. 2015;89:3619–3629.

77. Pfeiffer T., Pisch T., Devitt G., Holtkotte D., Bösch V. Effects of signal peptide exchange on HIV-1 glycoprotein expression and viral infectivity in mammalian cells. *FEBS Lett.* 2006;580:3775–3778.
78. Yolitz J., Schwing C., Chang J., Van Ryk D., Nawaz F., Wei D., Cicala C., Arthos J., Fauci A.S. Signal peptide of HIV envelope protein impacts glycosylation and antigenicity of gp120. *Proc. Natl. Acad. Sci. USA.* 2018;115:2443–2448.

CHAPITRE IV
DISCUSSIONS GÉNÉRALES ET PERSPECTIVES
DE L'ÉTUDE

1. CONTRIBUTIONS SCIENTIFIQUES DE CETTE THÈSE

L'ensemble des travaux inscrits dans cette thèse présente de nouvelles données concernant l'enveloppe des virus transmis/fondateurs du VIH-1. Ces évidences pourront avoir un impact sur l'épidémiologie moléculaire de cette infection, sur la caractérisation des infections récentes ainsi que sur la conception de vaccins thérapeutiques grâce à l'identification de signatures moléculaires associées à l'enveloppe des virus transmetteurs de l'infection par le VIH-1 susceptible d'informer le design d'immunogène capable d'induire des réponses anticorps neutralisants dans le cadre de recherches vaccinales contre le VIH-1. Ces données s'inscrivent dans trois thématiques conformément aux objectifs de notre étude et discutés ci-dessous.

Les résultats de la première partie de nos travaux publiés dans le journal « **PLOS ONE** » portent le titre : « **Identification des nouvelles infections par le VIH-1 en utilisant des mesures de la diversité génétique des séquences de l'enveloppe** » pour « *HIV-1 envelope sequence-based diversity measures for identifying recent infections* » (292). L'étude propose d'identifier la meilleure méthode permettant de prédire efficacement quelles sont les personnes nouvellement infectées en opposition à celles chroniquement infectées parmi quatre mesures de diversité génétique de la séquence de l'enveloppe du VIH-1. Ces quatre mesures incluent : le pourcentage de diversité, le pourcentage de complexité, le nombre d'haplotypes et l'entropie de Shannon (356, 669, 670). Ces mesures de diversité génétique pourraient ainsi constituer des biomarqueurs moléculaires pour l'identification des nouvelles infections par le VIH-1.

Nos travaux ont démontré que l'entropie de Shannon (356) de trois courts segments de l'enveloppe du VIH-1 (gp120 C2_1, gp120 C2_3 et la gp120_V3) prédit fortement la période récente de l'infection par le VIH-1. Ces trois segments sont localisés dans deux sous-régions (gp120 C2 et V3) de l'enveloppe du VIH-1 qui sont fortement associées à l'évolution de la maladie au stade précoce de l'infection (671). La valeur seuil de l'entropie de Shannon (S) des séquences virales des trois segments de l'enveloppe du VIH-1 identifiés qui ont donné les meilleurs résultats de prévision de la période récente de l'infection par le VIH est

respectivement de : 0.054 pour la gp120-C2_1, de 0.059 pour la gp120-C2_3 et de 0.083 pour la gp120-V3.

Dans un contexte de disponibilité des séquences, des réactifs, des amorces et de la maîtrise de la technologie de séquençage associée aux trois segments identifiés, la mesure de diversité génétique (entropie de Shannon) de la séquence de la boucle V3 serait plus facile à intégrer dans un programme de surveillance. Sa technique de séquençage est déjà bien maîtrisée dans de nombreux laboratoires en faveur des tests de tropisme viral par rapport au séquençage d'un fragment plus long de l'enveloppe du VIH-1. Un court fragment constitue un gain en termes de temps, de contraintes techniques et de réduction des coûts. L'Organisation mondiale de la santé n'a pas encore homologué une méthode de référence (gold standard) pour le diagnostic des infections récentes par le VIH-1. Néanmoins, le test de l'antigénémie p24 est utilisé jusqu'à présent, mais ce marqueur présente des limites à cause de sa présence d'une durée assez courte dans le sang (moins de 22 jours après l'infection) (279). Les méthodes sérologiques sont jusqu'à présent les plus utilisées, mais présentent aussi des limites liées à l'identification de faux récents. Les mauvaises classifications liées aux épreuves sérologiques obligent l'utilisation d'algorithmes pour accroître leurs performances et leurs sensibilités (293). Pourtant, les personnes nouvellement infectées sont celles qui présentent une capacité accrue à transmettre et à disséminer la maladie dans la population. L'entropie de Shannon est l'une des meilleures mesures de diversité génétique de l'enveloppe du VIH-1. Elle constitue un jalon de plus dans la méthodologie appliquée à la détection des nouvelles infections. Cette méthode couplée à un algorithme de détection des infections récentes contribuerait à améliorer la performance de ces différentes épreuves pour le diagnostic des nouvelles infections par le VIH-1. Comparativement à d'autres régions du génome du VIH-1 (*gag et pol*), nous pensons que des résultats similaires pourraient être obtenus. Il reste cependant à pouvoir identifier le ou les segments les plus prédictibles pour chacune des régions afin d'établir une comparaison objective. Pour ce faire, d'autres études sont encouragées afin de déterminer sur l'ensemble de la séquence du génome viral, le ou les segments dont la mesure de la diversité génétique permettraient de distinguer une infection récente par le VIH-1 par rapport à une infection chronique. Des études similaires

conduites sur les séquences du gène *gag* démontrent que les segments les plus informatifs pour l'identification des nouvelles infections se situent dans les régions variables de ce gène comme la *p17* et *p2/p7/p1/p6* (356). Puisque le gène *env* constitue le plus variable du génome du VIH-1, nous fondons l'hypothèse qu'il constitue le meilleur prédicteur pour l'identification des nouvelles infections. Cependant, des études comparatives méritent d'être conduites.

Les résultats de la deuxième partie de nos travaux publiés dans le journal « **AIDS Research and Human Retroviruses** » a porté sur « **Analyses phylogénétiques des séquences du gène de l'enveloppe du VIH-1 pour l'évaluation et la construction à court terme des réseaux naissants de la transmission de l'infection par le VIH-1** » pour « *A Short-Term Assessment of Nascent HIV-1 Transmission Clusters Among Newly Diagnosed Individuals Using Envelope Sequence-Based Phylogenetic Analyses* » (667). L'objectif de cette étude a été de reconstituer les chaînes de transmission du VIH-1 (réseau ou *cluster*) parmi les nouveaux cas diagnostiqués au Québec au cours de l'année 2015. Il s'est agi plus spécifiquement : (1) d'analyser et de comparer la capacité d'un fragment plus long (1070pb) à celui d'un fragment plus court (108 pb, loop 3) de la séquence de l'enveloppe du VIH-1 à prédire et à identifier de façon concordante les chaînes de transmission de la maladie ; (2) et d'identifier les facteurs épidémiologiques susceptibles de contribuer à la formation des réseaux de transmission du VIH-1 au Québec. Brièvement, les résultats de cette étude ont mis en évidence la formation de petits réseaux de transmission du virus (2-5 individus par chaîne) parmi les personnes nouvellement infectées au cours de l'année 2015. Cette étude a permis de démontrer l'existence d'une relation statistiquement significative entre le fragment long et celui plus court couvrant uniquement la boucle V3 de l'enveloppe virale pour la prédiction des réseaux de transmission du VIH-1 dans la population étudiée. Bien que cette relation (*degré d'accord, de concordance*), coefficient Kappa ($k=0.59$), l'utilisation des séquences de la boucle V3 de l'enveloppe du VIH-1 à la place de celles d'un fragment de 1070 pb de la séquence *env* pourraient être utilisées en première intention dans le cadre d'un dépistage intégré à un programme de surveillance des nouvelles infections. Nos travaux ont également permis d'associer deux facteurs à la formation des réseaux de transmission du VIH-1 dans la

population : l'âge moyen (<38.8 ans) et le sous-type viral (majoritairement le sous type B). Des études antérieures conduites par Bluma Brenner BG, et al. (2011, 2013 ,2017) (672-674) sur plusieurs années au Québec avait permis d'identifier de larges réseaux reflétant la dynamique de la transmission après plusieurs années (674). Cette étude diffère cependant de la nôtre, car elle a utilisé principalement des séquences du gène *pol* obtenues dans le cadre du programme de génotypage pour la détermination de la résistance aux traitements antirétroviraux institué dans toute la province. Notre étude a mis l'accent sur l'évaluation de la capacité des séquences du gène de l'enveloppe à identifier aussi efficacement les chaînes de transmission naissantes de la maladie pour une année. Notre étude s'est concentrée sur les réseaux développés au cours d'une année, car elle entend vérifier l'état de leur croissance le plus précocement possible. Le choix d'une année répond aussi à l'objectif de répondre au besoin d'un suivi en temps réel de la dynamique de la formation des réseaux de transmission de la maladie et de pouvoir les identifier le plus précocement possible afin de mieux les circonscrire.

Le troisième article avec pour titre : « **Identification de signatures génétiques au niveau de la glycoprotéine de l'enveloppe du VIH-1 associées aux virus transmis/fondateurs et récents de sous-types B** » pour « *HIV-1 Envelope Glycoprotein Amino Acids Signatures Associated with Clade B Transmitted/Founder and Recent Viruses* » publié dans le journal « **VIRUSES** » constitue la dernière partie de cette thèse. Cette étude a décrit les caractéristiques moléculaires des virus transmis/fondateurs qui sont des variants du VIH-1 sélectionnés parmi une multitude de virions exposés au cours de la transmission. Puisque ces virus arrivent à résister à tous les mécanismes de défense induits par le système immunitaire tant chez l'hôte qu'au niveau du receveur, ils possèdent en effet les déterminants moléculaires et/ou fonctionnels leur caractérisant. Ces propriétés moléculaires ou polymorphismes développés très tôt par les virus TF au niveau de leurs enveloppes au cours de la transmission constituent des signatures génétiques. La caractérisation moléculaire de ces différents changements opérant au niveau de la séquence de la glycoprotéine de l'enveloppe des virus TF permet de mieux comprendre les mécanismes sous-jacents de l'établissement de l'infection (goulot d'étranglement) susceptible d'informer les champs de recherche sur un vaccin anti-VIH-1. Les

résultats de nos travaux ont décrit plusieurs traits génétiques qui constituent des signatures caractéristiques du virus TF. Mais une substitution/mutation localisée dans le domaine cytoplasmique précisément dans le LLP-1 de la gp41 semble être la plus importante. Elle consiste en une substitution de l'arginine (R) par une isoleucine (I) et localisée à la position 841 (R841I). Le domaine cytoplasmique (CT) de la gp41 est hautement impliqué dans la fusion membranaire entre les protéines virales et cellulaires au cours du cycle de réplication du VIH-1. Il régule le trafic intracellulaire, mais il facilite l'assemblage des protéines virales et module l'expression des glycoprotéines de surface de l'enveloppe du VIH-1 (244, 245). Il contient trois peptides nommés « *Lentivirus lytic peptide (LLP) domains 1-3* » (675) qui contribuent à l'expression de l'enveloppe à la surface cellulaire (676, 677), à l'incorporation de la particule virale (678, 679), à la fusion (247, 680), et à la localisation de la particule virale dans les parois lipidiques de la cellule hôte (675). Le rôle du domaine cytoplasmique est amplement décrit dans cet ouvrage au chapitre I au point 3.6. Au regard des fonctions spécifiques au domaine cytoplasmique, il est possible que la mutation R841I identifiée dans le LLP-1 puisse constituer un des mécanismes développés par les virus TF contribuant à la capacité de résistance du virus face aux mécanismes de défense développés par le système immunitaire pour faciliter le processus de fusion des protéines virales à la membrane cellulaire. En effet, il a été démontré que des mutations au niveau de la séquence du LLP-1 affectent l'association de l'enveloppe avec les radeaux lipidiques (681, 682), elles réduisent son incorporation et l'infectiosité virale de même que le processus de réplication pour certains phénotypes viraux (683). La mutation R841I pourrait jouer un rôle inverse en contribuant au succès de l'incorporation et à la fusion virale.

Une autre mutation (substitution de la lysine (K), *normalement présente à la position 6 du peptide signal (SP) de la séquence de référence HXB2*, par l'isoleucine (I)) est abondamment sélectionnée par les virus chroniques par rapport aux virus TF. Bien que l'isoleucine soit faiblement sélectionnée au niveau de la SP par les virus TF, elle pourrait aussi constituer un mécanisme d'échappement au système immunitaire au cours de la transmission VIH-1 après la phase aiguë de l'infection par le VIH-1 en contribuant à la N-glycosylation de la gp120 de l'enveloppe du VIH-1 (684). L'identification des différents traits génétiques ou

polymorphisme au niveau de la séquence de l'enveloppe du VIH-1 au cours de la transmission et particulièrement à un domaine particulier constitue une des clés nécessaires à la flexibilité conformationnelle du virus et de résistance contre les mécanismes de défense de l'hôte. Gnanakaran *et al.* (503) ont précédemment identifié d'autres signatures génétiques au niveau de la séquence du peptide signal consistant principalement à la sélection de l'histidine (H) en position 12 (H12) et d'une substitution de la leucine (L) par une phénylalanine (F) en position 721 au niveau du domaine cytoplasmique de la gp41 (L721F) de l'enveloppe du VIH-1. Cette étude diffère cependant de la nôtre non seulement par le nombre de séquences analysées, aussi, par la durée de l'infection (fenêtre) définie pour la sélection des virus dits TF. Pour notre étude, toutes les séquences virales de virus TF provenaient de spécimens de patients infectés par le VIH-1 au stade Fiebig 1 à 2 définis uniquement par la présence de l'antigène p24 en absence d'anticorps (phase aigüe) contrairement pour celle de Gnanakaran *et al.* (2011) (475). Pour cette dernière, les spécimens provenaient de patients infectés par le VIH-1 au stade Fiebig 1 à 5, mais aussi a inclus un plus grand nombre de séquences que la nôtre. Nous avons sélectionné les spécimens de patients infectés par le VIH-1 au stade précoce de la maladie. Parce que les mutations génétiques qui s'opèrent le plus tôt refléteraient mieux les mécanismes moléculaires susceptibles d'influencer et de contribuer à la transmission avant toute accumulation de mutations dans le temps.

En somme, cette partie de nos travaux a permis de caractériser et d'identifier des mutations pouvant être considérées comme des signatures génétiques pour les virus TF mais aussi pour les virus chroniques (K6I). La plus importante étant celle localisée dans la région cytoplasmique de la GP41 et spécifiquement au niveau du LLP-1 (R841I), développée par les virus TF au cours de la primo-infection. Ces signatures génétiques spécifiques pourraient constituer de facteurs importants à considérer pour la compréhension des événements moléculaires précoces développés par le VIH-1 au cours de la transmission du VIH-1. Lesquelles pourront servir de repère pour la conception de vaccins et d'inhibiteurs d'entrée du VIH-1 ? Cependant, des études complémentaires portant sur l'évaluation des propriétés fonctionnelles de ces mutations apporteront plus d'informations sur leurs impacts dans le

processus de réplication et de la transmission du VIH-1. Les évidences qui en découleront permettront de mieux informer les champs de recherche sur un vaccin anti-VIH-1. Notamment en optimisant la construction de meilleurs vaccins candidats ou la sélection de meilleurs immunogènes d'enveloppe du VIH-1 susceptibles d'induire des anticorps neutralisants de large spectre contre le VIH-1 qui prendraient en compte les caractéristiques moléculaires de la séquence de l'enveloppe des virus transmis/fondateurs.

En matière de santé publique, nous proposons d'utiliser les méthodes moléculaires pour le diagnostic des nouvelles infections par le VIH-1 basées sur l'entropie de Shannon, une mesure de diversité génétique de des trois segments (gp120_C2_1 et C2_3 et V3) de l'enveloppe du VIH-1. Cette méthode pourrait être couplée à l'ensemble des méthodes existantes pour améliorer les performances des tests et des méthodes de diagnostic des nouvelles infections par le VIH-1. Nous avons démontré que les séquences du gène *env* d'une taille plus grande (1070 paires de bases) et celles d'une taille réduite de 118 paires de bases incluant seulement les séquences de la boucle V3, pourraient être utilisées pour identifier les réseaux ou cluster naissants de transmission de l'infection par le VIH-1. Nos travaux ont pu identifier en l'espace d'une année la formation de petits réseaux de transmission du VIH-1 principalement chez les sujets de moins de 40 ans infectés par le clade B au Québec. Ces évidences bien que ne constituant pas un portrait global de la transmission du VIH-1 au Québec corroborent avec les données de l'agence de santé publique du Canada de 2018 qui présentent les jeunes de 30-39 comme les plus affectés par le VIH-1(16). Nos résultats devraient permettre aux autorités de santé publique d'améliorer les stratégies de lutte et de prévention des nouvelles transmissions du VIH-1 en ciblant les populations concernées. Nos travaux ont aussi identifié des signatures génétiques au niveau de la séquence de l'enveloppe du VIH-1. Ces signatures permettront d'optimiser la sélection d'immunogènes ou de candidats vaccins prenant en compte les spécificités polymorphiques de certains sous-régions ou domaines de l'enveloppe du VIH-1 comme le domaine cytoplasmique. Pour ce dernier cas, nous estimons au regard de nos résultats que l'efficacité du vaccin RV144 aurait pu avoir plus d'efficacité s'il avait inclus les caractéristiques du domaine cytoplasmique de la gp41 dans la construction du candidat vaccin : (*ALVAC-HIV (vCP1521) est un vecteur viral*

(canarypox) qui code pour les gènes *gag*, *pol*, et le **domaine transmembranaire de la gp41 d'une souche du VIH-1 de clade B**, mais aussi du fragment *gp120* de l'enveloppe d'un isolat de CRF01-AE) (510).

De façon générale, nos travaux contribueront à l'atteinte des objectifs 90-90-90 à l'horizon 2020 de l'ONUSIDA. À savoir que : d'ici 2020, 90% des personnes vivant avec le VIH connaîtront leur statut sérologique, 90% de toutes les personnes infectées par le VIH dépisté recevront un traitement antirétroviral durable et qu'enfin, 90% des personnes recevant un traitement antirétroviral auront une charge virale durablement supprimée. Pris ensemble, cet ouvrage apporte de nouvelles connaissances susceptibles de contribuer à la lutte et à la prévention de l'infection par le VIH-1.

Sur le plan scientifique, nos travaux ont permis de publier trois articles dans des journaux à comité de lecture sur des thématiques portant sur la caractérisation moléculaire de l'enveloppe des virus transmis/fondateurs du VIH-1. Ces différents résultats sont susceptibles de contribuer à l'avancement de la science et d'ouvrir les perspectives d'autres études pour le futur. Nos différentes études ont permis également de mettre à disposition plus de 400 séquences du gène *env* du VIH-1 de sujets infectés au Canada aux stades précoces et chroniques dans les bases de données internationales (*Los Alamos HIV sequence database*) qui serviront pour des études futures.

2. PERSPECTIVES

Comme toute œuvre humaine, nous sommes conscients que ce travail présente des limites qui méritent d'être améliorées. Dans le cadre des travaux présentés ici, nous suggérons que des études futures puissent évaluer les propriétés fonctionnelles des mutations (R841I, K6I, Q621D, taille de la boucle V5) pour mieux comprendre leurs importances et implications dans le processus de transmission du VIH-1. Ces études complémentaires contribueront également à mieux cerner leurs impacts dans la résistance du VIH-1 contre les mécanismes de défense développés par le système immunitaire en début d'infections.

Ces données complémentaires qui en résulteront pourront servir de bases pour le développement de moyen de prévention (identification d'immunogènes ou de conception de candidats vaccins dans le cadre de vaccin et inhibiteurs d'entrée contre le VIH-1).

Nous suggérons d'inclure dans les programmes de surveillance en santé publique l'analyse de la diversité génétique des séquences du gène *env* des segments de la C2 et V3 de la gp120 comme autres biomarqueurs pour la détection des nouvelles infections par le VIH-1. Nous suggérons de prendre en compte dans le cadre de programmes de surveillance en santé publique, à défaut d'un fragment de plus de 1000 paires de bases de la séquence de l'enveloppe du VIH-1, celui de la boucle V3 (118 paires de bases) parmi les méthodes ou biomarqueurs pour l'identification des réseaux ou clusters de transmission du VIH-1.

CONCLUSIONS

Les différents travaux présentés dans cette thèse ont montré l'importance de l'utilisation des méthodes moléculaires (*sequence-based diversity measures*) pour l'identification des nouvelles infections à VIH-1 dans la population. Les résultats démontrent que l'entropie de Shannon de trois segments de l'enveloppe du VIH-1 constitue une des mesures de diversité génétique capable de prédire et identifier correctement les nouvelles infections à VIH-1. Les personnes nouvellement infectées par le VIH-1 jouent un rôle important dans la dissémination et la transmission de l'infection dans la population. Il serait nécessaire d'associer l'entropie de Shannon aux nombreux tests de diagnostic précoce de l'infection dans un algorithme afin d'améliorer les performances de ces épreuves.

Nos travaux ont aussi permis de démontrer qu'utiliser des séquences virales d'un fragment d'une taille assez réduite comme celles de la boucle V3 de l'enveloppe du VIH-1 pouvait prédire les réseaux de transmission de l'infection avec une efficacité presque équivalente qu'un fragment d'une taille plus longue. À défaut de pouvoir séquencer l'enveloppe complète du VIH-1, il est suggéré d'utiliser celle de la boucle V3 dans certaines mesures pour les analyses phylogénétiques pour déterminer les réseaux de transmission du VIH-1. Il serait nécessaire que ces différentes méthodes soient intégrées à celles déjà existantes dans le cadre d'un algorithme de détection des infections récentes par le VIH-1. Nos résultats ont permis de démontrer que le facteur âge et le sous-type viral infectant l'individu étaient associés à la formation des réseaux de transmission du VIH-1 dans la population étudiée (Québec). L'identification de ces facteurs contributifs ou communs devrait permettre de mieux orienter les stratégies de riposte dans la province. Il serait nécessaire d'intégrer un suivi en temps réel des réseaux de transmission du VIH dans les programmes de santé publique pour mieux dénicher les réseaux de transmission naissants afin de les circonscire à temps.

Nos travaux présentés dans cette thèse ont également permis d'identifier des signatures génétiques principalement dans le peptide signal, la boucle V1 et dans le domaine cytoplasmique de la gp41 qui pourront être considérée comme de caractéristiques

particulières des virus transmis/fondateurs. Les évènements précoces qui ont lieu précisément au niveau de ces domaines de l'enveloppe virale et qui sont fortement impliqués dans la réplication virale méritent une attention particulière et de plus amples évaluations. Ces caractéristiques pourraient constituer des clés du succès de la transmission et les bases de développement d'un vaccin. Il serait important de mener des études complémentaires afin de mieux comprendre leurs implications fonctionnelles dans la pathogenèse virale et au cours de la transmission du VIH-1 et de l'évolution de la maladie.

ANNEXES GÉNÉRALES

Protocole d'amplification de l'enveloppe (gp160) du VIH-1



Annexe 1 : Protocole de Reverse Transcriptase PCR (RT-PCR).

Date : _____ Fait par : _____

Extraction : (QiAmp Viral RNA mini kit (Qiagen)=100µl plasma) : PR-BM-010 PR-BM-085

Dénaturation de l'ARN (V : 10µl d'échantillon)	Début	Fin	Initiales
Incubation 65°C ± 2 °C, 5 minutes			

Bain-marie :

thermomètre :

thermocycleur :

Nombre total d'échantillons à analyser +1 (n + _) : _____

N° de lot de la trousse « SuperScript™ III One-Step RT-PCR System with Platinum® Taq DNA Polymerase » : _____ ; date de péremption : _____

Préparation du mélange réactionnel (Reaction mixture for RT-PCR)				
Réactifs	N° lot ou date de préparation	Date de péremption	Volume (µl) par réaction X nb_Ech+1	Volume total (µl) dans le mélange réactionnel
DEPC WATER			13,5	
2X Reaction Mix (5 µM)	S/O	S/O	25,0	
Primers SG3-up (10µM): 5' TACAGTGCAGGGGAAAGAATAATAGACATAATA-3' HXB2 (4809 – 4841)			2,0	
Primers SG3-lo(10µM): 5'- AGACCCAGTACAGGCRARAAGC-3' HXB2(9523 – 9544)			2,0	

RNase out (2U/μl)			0,5	
SuperScript III RT / Platinum® Taq Mix (10U/μl)	S/O	S/O	2,0	
Volume du mélange réactionnel			45,0	
Volume de la préparation d'acide nucléique (Extrait ARN)			5,0	
Volume final			50,0	

Thermocycleur L_____

Utiliser le ABI 9700 en
mode PE 9600

Programmes

_____	53 °C	60 mn	35 cycles
_____	94 °C	2 min	
_____	94 °C	15 s	
_____	55 °C	30 s	
_____	68 °C	5min	
_____	68 °C	10 min	
_____	04 °C	infini	

Micropipettes utilisées		
Réactifs	préPCR	

Annexe 2 : Protocole d'amplification par RT-PCR nichée de l'enveloppe du VIH-1.



Protocole d'Amplification de l'enveloppe (gp160) du VIH-1 par Nested PCR

Date : _____ Fait par : _____

Nombre total d'échantillons à analyser +1 (n+1) : _____

N° de lot de la trousse Expand High Fidelity PCR System: _____

Préparation du mélange réactionnel (Reaction mixture for Nested-PCR)				
Réactifs	N° lot ou date de fabrication	Date de péréemption	Volume (µl) par réaction X Nb Ech	Volume total (µl) dans le mélange réactionnel
DEPC water			37,10	
Expand® Buffer 10X			5,0	
dNTPs (10 mM)			1,5	
Primers <i>env-up*</i> (10 µM) : 5'-GTTTCTTTTAGGCATCTCCTATGGCAGGAAGAAG-3' HXB2 (5957 – 5983)			2,0	
Primers <i>env-lo*</i> (10 µM): 5'-GTTTCTCCAGTCCCCCTTTTCTTTAAAAAG-3' HXB2 (9063 – 9088):			2,0	
Expand High Fidelity PCR System(3.5U/µl)			0,4	
Volume du mélange réactionnel			48,0	
Volume de la préparation d'acide nucléique :			2,0	
Volume final			50	

Thermocycleur L_____

Programmes

Utiliser le ABI 9700 en

Micropipettes utilisées		
Réactifs		post-PCR

_____	94 °C	2 min	45 cycles
_____	94 °C	15 s	
_____	55 °C	30 s	
_____	68 °C	5 min	
_____	68 °C	10 min	

BIBLIOGRAPHIE

1. Centers for Disease C. Pneumocystis pneumonia--Los Angeles. MMWR Morbidity and mortality weekly report. 1981;30(21):250-2.
2. Freed EO. HIV-1 replication. Somatic Cell and Molecular Genetics. 2001;26(1-6):13-33.
3. Lapointe N, Michaud J, Pekovic D, Chausseau JP, Dupuy JM. Trans-placental transmission of HTLV-III virus. New England Journal of Medicine. 1985;312(20): 1325-6
4. Centers for Disease C. Update on acquired immune deficiency syndrome (AIDS)--United States. MMWR Morbidity and mortality weekly report. 1982;31(37):507-8, 13-4.
5. Sepkowitz KA. AIDS - The first 20 years. New England Journal of Medicine. 2001;344(23):1764-72.
6. Gallo RC, Sarin PS, Gelmann EP, Robertguroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human t-cell leukemia-virus in acquired immune-deficiency syndrome (AIDS). Science. 1983;220(4599):865-7.
7. Gelmann EP, Popovic M, Blayney D, Masur H, Sidhu G, Stahl RE, et al. Proviral dna of a retrovirus, human t-cell leukemia-virus, in 2 patients with AIDS. Science. 1983;220(4599):862-5.
8. Barresinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. ISolation of a t-lymphotropic retrovirus from a patient at risk for acquired immune-deficiency syndrome (AIDS). Science. 1983;220(4599):868-71.
9. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 1984;224(4648):497-500.

10. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*. 1984;224(4648):500-3.
11. Singer HL. Institut-pasteur v united-states - the aids patent dispute, the contract disputes act and the international exchange of scientific-data. *American Journal of Law & Medicine*. 1989;15(4):439-59.
12. Lever AML, Berkhout B. 2008 Nobel prize in medicine for discoverers of HIV. *Retrovirology*. 2008;5:91.
13. Case K. Nomenclature - human-immunodeficiency-virus. *Annals of Internal Medicine*. 1986;105(1):133.
14. Clavel F, Guetard D, Brunvezinet F, Chamaret S, Rey MA, Santosferreira MO, et al. Isolation of a new human retrovirus from west-african patients with aids. *Science*. 1986;233(4761):343-6.
15. UNAIDS. UNAIDS data 2020. Geneva, Switzerland: Access 13 January 2020. <https://www.unaids.org/fr/resources/fact-sheet>.
16. Haddad N RA, Weeks A, Popovic N, Siu W, Archibald C. Le VIH au Canada — Rapport de surveillance, 2018. *Relevé des maladies transmissibles au Canada* 2019;45(12):334–43.
17. Bourgeois AC EM, Awan A, Jonah L, Varsaneux O, Siu W. HIV in Canada—Surveillance Report 2016. *Can Commun Dis Rep*; 2017.
18. PA L. Human immunodeficiency viruses and their replication. In: Publishers PL-R, editor. *Fields virology*. 3rd ed ed1996. p. 1881–952.
19. Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS - AIDS as a zoonosis: Scientific and public health implications. *Science*. 2000;287(5453):607-14.

20. Gurtler L, Aepfelbacher M, Bauerfeind U, Blumel J, Burger R, Gartner B, et al. Human Immunodeficiency Virus (HIV). *Transfusion Medicine and Hemotherapy*. 2016;43(3):203-22.
21. Baltimore D. Viral rna-dependent dna polymerase - RNA-dependent dna polymerase in virions of rna tumour viruses. *Nature*. 1970;226(5252): 1209–1211.
22. Sundquist WI, Krausslich HG. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine*. 2012;2 (7).
23. Henderson. Diagram of the HIV virus: US National Institute of Health (redrawn by :Carl Henderson); access, 3 June 2005.https://commons.wikimedia.org/wiki/File:HIV_Virion-en.png.
24. PK V. Historical Introduction to the General Properties of Retroviruses. In: Coffin JM HS, Varmus HE, editor. *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.
25. Klaver B, Berkhout B. Comparison of 5' and 3' long terminal repeat promoter function in human-immunodeficiency-virus. *Journal of Virology*. 1994;68(6):3830-40.
26. Brown PH, Tiley LS, Cullen BR. Efficient polyadenylation within the human-immunodeficiency-virus type-1 long terminal repeat requires flanking u3-specific sequences. *Journal of Virology*. 1991;65(6):3340-3.
27. Levy JA. Pathogenesis of human immunodeficiency virus infection. *Microbiological reviews*. 1993;57(1):183-289.
28. Sauter D, Unterweger D, Vogl M, Usmani SM, Heigele A, Kluge SF, et al. Human Tetherin Exerts Strong Selection Pressure on the HIV-1 Group N Vpu Protein. *Plos Pathogens*. 2012;8(12).

29. Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng LW, et al. Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontiers in Bioscience*. 2004;9:2388-413.
30. Anderson JL, Hope TJ. Recent Insights into HIV Accessory Proteins. *Current infectious disease reports*. 2003;5(5):439-50.
31. Vicenzi E, Poli G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens*. 2013;81(2):61-71.
32. Kuiken C LT, Hahn B, Mullins J, Wolinsky S, Foley B, Apetrei C, Mizrachi I, Rambaut A, Korber B. HIV Sequence Compendium 2012. NM,: Los Alamos National Lab; 2012.
33. Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, et al. Adaptation of HIV-1 to its human host. *Molecular Biology and Evolution*. 2007;24(8):1853-60.
34. Takehisa J, Kraus MH, Ayoub A, Bailes E, Van Heuverswyn F, Decker JM, et al. Origin and Biology of Simian Immunodeficiency Virus in Wild-Living Western Gorillas. *Journal of Virology*. 2009;83(4):1635-48.
35. Vidricaire G. Étude des étapes précoces du cycle de réplication du virus d'immunodéficience humaine de type 1 dans les cellules trophoblastiques : Vers une compréhension de la transmission materno-fœtale. Thèse de doctorat, Université Laval, Canada. 2006.
36. Freed EO. HIV-1 Gag proteins: Diverse functions in the virus life cycle. *Virology*. 1998;251(1):1-15.
37. Pearl LH, Taylor WR. A structural model for the retroviral proteases. *Nature*. 1987;329(6137):351-4.

38. Imamichi T. Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Curr Pharm Des.* 2004;10(32):4039-53.
39. Piana S, Carloni P, Rothlisberger U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* 2002;11(10):2393-402.
40. Todd MJ, Freire E. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins.* 1999;36(2):147-56.
41. Todd MJ, Semo N, Freire E. The structural stability of the HIV-1 protease. *J Mol Biol.* 1998;283(2):475-88.
42. Toth G, Borics A. Flap opening mechanism of HIV-1 protease. *J Mol Graph Model.* 2006;24(6):465-74.
43. Velazquez-Campoy A, Muzammil S, Ohtaka H, Schon A, Vega S, Freire E. Structural and thermodynamic basis of resistance to HIV-1 protease inhibition: implications for inhibitor design. *Curr Drug Targets Infect Disord.* 2003;3(4):311-28.
44. Brik A, Wong CH. HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem.* 2003;1(1):5-14.
45. Park J, Morrow CD. Mutations in the Protease Gene of Human Immunodeficiency Virus Type 1 Affect Release and Stability of Virus Particles. *Virology.* 1993;194(2):843-50.
46. Perez MA, Fernandes PA, Ramos MJ. Substrate recognition in HIV-1 protease: a computational study. *J Phys Chem B.* 2010;114(7):2525-32.
47. Pettit SC, Clemente JC, Jeung JA, Dunn BM, Kaplan AH. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. *J Virol.* 2005;79(16):10601-7.

48. Krausslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci U S A*. 1989;86(3):807-11.
49. de Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, zur Megede J, et al. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol*. 2003;77(17):9422-30.
50. Barrie KA, Perez EE, Lamers SL, Farmerie WG, Dunn BM, Sleasman JW, et al. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology*. 1996;219(2):407-16.
51. Wilson W, Braddock M, Adams SE, Rathjen PD, Kingsman SM, Kingsman AJ. HIV expression strategies-ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell*. 1988;55(6):1159-69.
52. Rice AP. The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies. *Current pharmaceutical design*. 2017;23(28):4098-102.
53. Zhou C, Rana TM. A bimolecular mechanism of HIV-1 Tat protein interaction with RNA polymerase II transcription elongation complexes. *J Mol Biol*. 2002;320(5):925-42.
54. Sheridan PL, Sheline CT, Milocco LH, Jones KA. Tat and the HIV-1 promoter: A model for RNA-mediated regulation of transcription. *Seminars in Virology*. 1993;4(1):69-80.
55. Fackler OT. Spotlight on HIV-1 Nef: SERINC3 and SERINC5 Identified as Restriction Factors Antagonized by the Pathogenesis Factor. *Viruses*. 2015;7(12):6730-8.

56. Rosa A, Chande A, Ziglio S, De Sanctis V, Bertorelli R, Goh SL, et al. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*. 2015;526(7572):212-7.
57. Usami Y, Wu Y, Gottlinger HG. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*. 2015;526(7572):218-23.
58. Inuzuka M, Hayakawa M, Ingi T. Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. *J Biol Chem*. 2005;280(42):35776-83.
59. Li M, Waheed AA, Yu J, Zeng C, Chen HY, Zheng YM, et al. TIM-mediated inhibition of HIV-1 release is antagonized by Nef but potentiated by SERINC proteins. *Proc Natl Acad Sci U S A*. 2019;116(12):5705-14.
60. Desimmie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *Journal of molecular biology*. 2014;426(6):1220-45.
61. Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell*. 2003;12(3):591-601.
62. Goila-Gaur R, Strebel K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology*. 2008;5:51.
63. Malim MH. Natural resistance to HIV infection: The Vif–APOBEC interaction. *Comptes Rendus Biologies*. 2006;329(11):871-5.
64. Cadima-Couto I, Goncalves J. Towards Inhibition of Vif-APOBEC3G Interaction: Which Protein to Target? *Advances in Virology*. 2010;2010:10.

65. Goncalves J, Santa-Marta M. HIV-1 Vif and APOBEC3G: Multiple roads to one goal. *Retrovirology*. 2004;1(1):28.
66. Neil SJ. The antiviral activities of tetherin. *Curr Top Microbiol Immunol*. 2013;371:67-104.
67. Sauter D. Counteraction of the multifunctional restriction factor tetherin. *Front Microbiol*. 2014;5:163.
68. Varthakavi V, Smith RM, Bour SP, Strebel K, Spearman P. Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(25):15154-9.
69. Tokarev A, Skasko M, Fitzpatrick K, Guatelli J. Antiviral activity of the interferon-induced cellular protein BST-2/tetherin. *AIDS Res Hum Retroviruses*. 2009;25(12):1197-210.
70. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*. 2008;451(7177):425-30.
71. Pujol FM, Laketa V, Schmidt F, Mukenhirn M, Müller B, Boulant S, et al. HIV-1 Vpu Antagonizes CD317/Tetherin by Adaptor Protein-1-Mediated Exclusion from Virus Assembly Sites. *Journal of Virology*. 2016;90(15):6709.
72. Douglas JL, Viswanathan K, McCarroll MN, Gustin JK, Früh K, Moses AV. Vpu Directs the Degradation of the Human Immunodeficiency Virus Restriction Factor BST-2/Tetherin via a β TrCP-Dependent Mechanism. *Journal of Virology*. 2009;83(16):7931.
73. Hauser H, Lopez LA, Yang SJ, Oldenburg JE, Exline CM, Guatelli JC, et al. HIV-1 Vpu and HIV-2 Env counteract BST-2/tetherin by sequestration in a perinuclear compartment. *Retrovirology*. 2010;7(1):51.

74. Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, et al. Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell*. 2009;139(3):499-511.
75. Van Damme N, Goff D, Katsura C, Jorgenson RL, Mitchell R, Johnson MC, et al. The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe*. 2008;3(4):245-52.
76. Le Tortorec A, Neil SJ. Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *J Virol*. 2009;83(22):11966-78.
77. Li SX, Barrett BS, Guo K, Santiago ML. Tetherin/BST-2: Restriction Factor or Immunomodulator? *Curr HIV Res*. 2016;14(3):235-46.
78. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Segal E, et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*. 2011;474(7353):654-7.
79. Leavy O. Antiviral immunity: SAMHD1-stopping HIV in its tracks. *Nat Rev Immunol*. 2011;11(7):440.
80. Lim ES, Emerman M. HIV: Going for the watchman. *Nature*. 2011;474(7353):587-8.
81. Baldauf H-M, Stegmann L, Schwarz S-M, Ambiel I, Trotard M, Martin M, et al. Vpx overcomes a SAMHD1-independent block to HIV reverse transcription that is specific to resting CD4 T cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114(10):2729-34.
82. Bakir TM. The role of SAMHD1 expression and its relation to HIV-2 (Vpx) gene production. *Saudi Pharm J*. 2018;26(6):903-8.

83. St Gelais C, Wu L. SAMHD1: a new insight into HIV-1 restriction in myeloid cells. *Retrovirology*. 2011;8(1):55.
84. Wu L. SAMHD1: a new contributor to HIV-1 restriction in resting CD4+T-cells. *Retrovirology*. 2012;9(1):88.
85. Peterlin BM, Trono D. Hide, shield and strike back: How HIV-infected cells avoid immune eradication. *Nature Reviews Immunology*. 2003;3(2):97-107.
86. Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science*. 1998;280(5371):1884-8.
87. Freed EO, Martin MA. The role of human-immunodeficiency-virus type-1 envelope glycoproteins in virus-infection. *Journal of Biological Chemistry*. 1995;270(41):23883-6.
88. Dalgleish AG, Beverley PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) ANTIGEN is an essential component of the receptor for the aids retrovirus. *Nature*. 1984;312(5996):763-7.
89. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The T4 gene encodes the aids virus receptor and is expressed in the immune-system and the brain. *Cell*. 1986;47(3):333-48.
90. Chesebro B, Buller R, Portis J, Wehrly K. Failure of human immunodeficiency virus entry and infection in cd4-positive human-brain and skin cells. *Journal of Virology*. 1990;64(1):215-21.
91. Clapham PR, Blanc D, Weiss RA. Specific cell-surface requirements for the infection of cd4-positive cells by human-immunodeficiency-virus type-1 and type-2 and by simian immunodeficiency virus. *Virology*. 1991;181(2):703-15.

92. Sattentau QJ, Moore JP. Conformational-changes induced in the human-immunodeficiency-virus envelope glycoprotein by soluble cd4 binding. *Journal of Experimental Medicine*. 1991;174(2):407-15.
93. Thali M, Moore JP, Furman C, Charles M, Ho DD, Robinson J, et al. Characterization of conserved human-immunodeficiency-virus type-1 gp120 neutralization epitopes exposed upon gp120-cd4 binding. *J Virol*. 1993;67(7):3978-88.
94. Labrosse B, Treboute C, BreLOT A, Alizon M. Cooperation of the V1/V2 and V3 domains of human immunodeficiency virus type 1 gp120 for interaction with the CXCR4 receptor. *Journal of Virology*. 2001;75(12):5457-64.
95. Saunders CJ, McCaffrey RA, Zharkikh I, Kraft Z, Malenbaum SE, Burke B, et al. The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *Journal of Virology*. 2005;79(14):9069-80.
96. Zolla-Pazner S, Cardozo T. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nature Reviews Immunology*. 2010;10(7):527-35.
97. Deng HK, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. 1996;381(6584):661-6.
98. Wu LJ, Gerard NP, Wyatt R, Choe H, Parolin C, Ruffing N, et al. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*. 1996;384(6605):179-83.
99. Trkola A, Dragic T, Arthos J, Binley JM, Olson WC, Allaway GP, et al. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature*. 1996;384(6605):184-7.

100. Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. Identification of the envelope v3 loop as the primary determinant of cell tropism in HIV-1. *Science*. 1991;253(5015):71-4.
101. Shioda T, Levy JA, Chengmayer C. Macrophage and t-cell line tropisms of hiv-1 are determined by specific regions of the envelope gp120 gene. *Nature*. 1991;349(6305):167-9.
102. Berger EA, Doms RW, Fenyo EM, Korber BTM, Littman DR, Moore JP, et al. A new classification for HIV-1. *Nature*. 1998;391(6664):240.
103. Furuta RA, Wild CT, Weng YK, Weiss CD. Capture of an early fusion-active conformation of HIV-1 gp41. *Nature Structural Biology*. 1998;5(4):276-9.
104. Abrahamyan LG, Markosyan RM, Moore JP, Cohen FS, Melikyan GB. Human immunodeficiency virus type 1 Env with an intersubunit disulfide bond engages coreceptors but requires bond reduction after engagement to induce fusion. *Journal of Virology*. 2003;77(10):5829-36.
105. Koshiba T, Chan DC. The prefusogenic intermediate of HIV-1 gp41 contains exposed C-peptide regions. *Journal of Biological Chemistry*. 2003;278(9):7573-9.
106. Si ZH, Madani N, Cox JM, Chruma JJ, Klein JC, Schon A, et al. Small-molecule inhibitors of HIV-1 entry block receptor-induced conformational changes in the viral envelope glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(14):5036-41.
107. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell*. 1997;89(2):263-73.
108. Lu M, Blacklow SC, Kim PS. A Trimeric structural domain of the hiv-1 transmembrane glycoprotein. *Nature Structural Biology*. 1995;2(12):1075-82.

109. Yang J, Xia XY, He X, Yang SL, Ruan YH, Zhao QB, et al. A new pattern-based method for identifying recent HIV-1 infections from the viral sequence. *Science China-Life Sciences*. 2012;55(4):328-35.
110. Karageorgos L, Li P, Burrell C. Characterization of hiv replication complexes early after cell-to-cell infection. *Aids Research and Human Retroviruses*. 1993;9(9):817-23.
111. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *Journal of Virology*. 2002;76(11):5667-77.
112. Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, et al. HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. *Plos Pathogens*. 2011;7(12).
113. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature*. 2011;472(7343):361-5.
114. Yamashita M, Perez O, Hope TJ, Emerman M. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *Plos Pathogens*. 2007;3(10):1502-10.
115. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5 alpha restricts HIV-1 infection in Old World monkeys. *Nature*. 2004;427(6977):848-53.
116. Sebastian S, Luban J. TRIM5a selectively binds a restriction-sensitive retroviral capsid. *Retrovirology*. 2005;2.
117. Stremlau M, Perron M, Lee M, Li Y, Song B, Javanbakht H, et al. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5 alpha restriction factor.

Proceedings of the National Academy of Sciences of the United States of America. 2006;103(14):5514-9.

118. Campbell EM, Weingart J, Sette P, Opp S, Sastri J, O'Connor SK, et al. TRIM5 alpha-Mediated Ubiquitin Chain Conjugation Is Required for Inhibition of HIV-1 Reverse Transcription and Capsid Destabilization. *Journal of Virology*. 2016;90(4):1849-57.
119. Nakayama EE, Miyoshi H, Nagai Y, Shioda T. A Specific Region of 37 Amino Acid Residues in the SPRY (B30.2) Domain of African Green Monkey TRIM5 α Determines Species-Specific Restriction of Simian Immunodeficiency Virus SIVmac Infection. *Journal of Virology*. 2005;79(14):8870-7.
120. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5 α restricts HIV-1 infection in Old World monkeys. *Nature*. 2004;427:848.
121. Yap MW, Nisole S, Stoye JP. A single amino acid change in the SPRY domain of human Trim5 alpha leads to HIV-1 restriction. *Current Biology*. 2005;15(1):73-8.
122. Song BW, Gold B, O'HUigin C, Javanbakht H, Li X, Stremlau M, et al. The B30.2(SPRY) domain of the retroviral restriction factor TRIM5 alpha exhibits line age-specific length and sequence variation in primates. *Journal of Virology*. 2005;79(10):6111-21.
123. Goff SP. HIV Replication trimmed back. *Nature*. 2004;427(6977): 791-3.
124. Ikeda Y, Ylinen LMJ, Kahar-Bador M, Towers GJ. Influence of gag on human immunodeficiency virus type 1 species-specific tropism. *Journal of Virology*. 2004;78(21):11816-22.
125. Owens CM, Song BW, Perron MJ, Yang PC, Stremlau M, Sodroski J. Binding and susceptibility to postentry restriction factors in monkey cells are specified by distinct

- regions of the human immunodeficiency virus type I capsid. *Journal of Virology*. 2004;78(10):5423-37.
126. Kootstra NA, Munk C, Tonnu N, Landau NR, Verma IM. Abrogation of postentry restriction of HIV-1-based lentiviral vector transduction in simian cells. *Proc Natl Acad Sci U S A*. 2003;100(3):1298-303.
127. Hatzioannou T, Cowan S, von Schwedler UK, Sundquist WI, Bieniasz PD. Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid. *Journal of Virology*. 2004;78(11):6005-12.
128. Ylinen LMJ, Keckesova Z, Wilson SJ, Ranasinghe S, Towers GJ. Differential restriction of human immunodeficiency virus type 2 and simian immunodeficiency virus SIVmac by TRIM5 alpha alleles. *Journal of Virology*. 2005;79(18):11580-7.
129. Chatterji U, Bobardt MD, Stanfield R, Ptak RG, Pallansch LA, Ward PA, et al. Naturally occurring capsid substitutions render HIV-1 cyclophilin A independent in human cells and TRIM-cyclophilin-resistant in owl monkey cells. *J Biol Chem*. 2005;280(48):40293-300.
130. Rose KM, Marin M, Kozak SL, Kabat D. Transcriptional regulation of APOBEC3G, a cytidine deaminase that hypermutates human immunodeficiency virus. *Journal of Biological Chemistry*. 2004;279(40):41744-9.
131. Henriot S, Richer D, Bernacchi S, Decroly E, Vigne R, Ehresmann B, et al. Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. *Journal of Molecular Biology*. 2005;354(1):55-72.
132. Dupuy LC, Kelly NJ, Elgavish TE, Harvey SC, Morrow CD. Probing the importance of tRNA anticodon: Human immunodeficiency virus type 1 (HIV-1) RNA genome complementarity with an HIV-1 that selects tRNA(Glu) for replication. *Journal of Virology*. 2003;77(16):8756-64.

133. Jiang M, Mak J, Ladha A, Cohen E, Klein M, Rovinski B, et al. Identification of transfer-rnas incorporated into wild-type and mutant human-immunodeficiency-virus type-1. *Journal of Virology*. 1993;67(6):3246-53.
134. Mak J, Jiang M, Wainberg MA, Hammarskjold ML, Rekosh D, Kleiman L. Role of pr160(gag-pol) in mediating the selective incorporation of trna(lys) into human-immunodeficiency-virus type-1 PARTICLES. *Journal of Virology*. 1994;68(4):2065-72.
135. He JL, Choe S, Walker R, Dimarzio P, Morgan DO, Landau NR. Human-immunodeficiency-virus type-1 viral-protein-r (vpr) arrests cells in the g(2) phase of the cell-cycle by inhibiting p34(cdc2) activity. *Journal of Virology*. 1995;69(11):6705-11.
136. Jacque JM, Mann A, Enslin H, Sharova N, Brichacek R, Davis RJ, et al. Modulation of HIV-1 infectivity by MAPK, a virion-associated kinase. *Embo Journal*. 1998;17(9):2607-18.
137. Yang XY, Gabuzda D. Regulation of human immunodeficiency virus type 1 infectivity by the ERK mitogen-activated protein kinase signaling pathway. *Journal of Virology*. 1999;73(4):3460-6.
138. Perez-Caballero D, Hatzioannou T, Zhang FW, Cowan S, Bieniasz PD. Restriction of human immunodeficiency virus type 1 by TRIM5-CypA occurs with rapid kinetics and independently of cytoplasmic bodies, ubiquitin, and proteasome activity. *Journal of Virology*. 2005;79(24):15567-72.
139. Stremlau M, Song BW, Javanbakht H, Perron M, Sodroski J. Cyclophilin A: An auxiliary but not necessary cofactor for TRIM5 alpha restriction of HIV-1. *Virology*. 2006;351(1):112-20.
140. Hulme AE, Perez O, Hope TJ. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(24):9975-80.

141. Fassati A, Goff SP. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *Journal of Virology*. 2001;75(8):3626-35.
142. Arhel N. Revisiting HIV-1 uncoating. *Retrovirology*. 2010;7:96.
143. Isel C, Ehresmann C, Marquet R. Initiation of HIV Reverse Transcription. *Viruses-Basel*. 2010;2(1):213-43.
144. Kogel D, Aboud M, Flugel RM. Molecular biological characterization of the human foamy virus reverse-transcriptase and ribonuclease-h domains. *Virology*. 1995;213(1):97-108.
145. Molling K, Bolognesi DP, Bauer H, Busen W, Plassmann HW, Hausen P. Association of viral reverse transcriptase with an enzyme degrading rna moiety of rna-dna hybrids. *Nature-New Biology*. 1971;234(51): 240-3
146. Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. Strand transfer events during HIV-1 reverse transcription. *Virus Research*. 2008;134(1-2):19-38.
147. Charneau P, Mirambeau G, Roux P, Paulous S, Buc H, Clavel F. HIV-1 reverse transcription - a termination step at the center of the genome. *Journal of Molecular Biology*. 1994;241(5):651-62.
148. Zennou V, Petit C, Guetard D, Nerhbass U, Montagnier L, Charneau P. HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell*. 2000;101(2):173-85.
149. Ao Z, Yao XJ, Cohen TA. Assessment of the role of the central DNA flap in human immunodeficiency virus type 1 replication by using a single-cycle replication system. *Journal of Virology*. 2004;78(6):3170-7.
150. Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, et al. HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *Embo Journal*. 2007;26(12):3025-37.

151. De Rijck J, Debyser Z. The central DNA flap of the human immunodeficiency virus type 1 is important for viral replication. *Biochemical and Biophysical Research Communications*. 2006;349(3):1100-10.
152. Dharan A, Talley S, Tripathi A, Mamede JI, Majetschak M, Hope TJ, et al. KIF5B and Nup358 Cooperatively Mediate the Nuclear Import of HIV-1 during Infection. *Plos Pathogens*. 2016;12(6).
153. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science*. 2008;319(5865):921-6.
154. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, et al. A role for LEDGF/p75 in targeting HIV DNA integration. *Nature medicine*. 2005;11(12):1287-9.
155. Cai M, Huang Y, Zheng R, Wei SQ, Ghirlando R, Lee MS, et al. Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nature Structural Biology*. 1998;5(10):903-9.
156. Bowerman B, Brown PO, Bishop JM, Varmus HE. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes & Development*. 1989;3(4):469-78.
157. Brown PO, Bowerman B, Varmus HE, Bishop JM. Retroviral integration - structure of the initial covalent product and its precursor, and a role for the viral in protein. *Proceedings of the National Academy of Sciences of the United States of America*. 1989;86(8):2525-9.
158. Lee K, Ambrose Z, Martin TD, Oztop I, Mulky A, Julias JG, et al. Flexible Use of Nuclear Import Pathways by HIV-1. *Cell Host & Microbe*. 2010;7(3):221-33.
159. Veillette M, Bichel K, Pawlica P, Freund SMV, Plourde MB, Pham QT, et al. The V86M mutation in HIV-1 capsid confers resistance to TRIM5 alpha by abrogation of cyclophilin

- A-dependent restriction and enhancement of viral nuclear import. *Retrovirology*. 2013;10.
160. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, et al. The Requirement for Cellular Transportin 3 (TNPO3 or TRN-SR2) during Infection Maps to Human Immunodeficiency Virus Type 1 Capsid and Not Integrase. *J Virol*. 2010;84(1):397-406.
161. Engelman A, Bushman FD, Craigie R. Identification of discrete functional domains of hiv-1 integrase and their organization within an active multimeric complex. *Embo Journal*. 1993;12(8):3269-75.
162. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *Journal of Biological Chemistry*. 2003;278(1):372-81.
163. Turlure F, Devroe E, Silver PA, Engelman A. Human cell proteins and human immunodeficiency virus DNA integration. *Frontiers in Bioscience*. 2004;9:3187-208.
164. Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D, et al. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *Journal of Biological Chemistry*. 2005;280(27):25517-23.
165. Kalpana GV, Marmon S, Wang WD, Crabtree GR, Goff SP. Binding and stimulation of hiv-1 integrase by a human homolog of yeast transcription factor SNF5. *Science*. 1994;266(5193):2002-6.
166. Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, et al. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Research*. 2013;41(6):3924-36.

167. Craigie R, Bushman FD. HIV DNA Integration. Cold Spring Harbor Perspectives in Medicine. 2012;2(7).
168. Yoder KE, Bushman FD. Repair of gaps in retroviral DNA integration intermediates. J Virol. 2000;74(23):11191-200.
169. Yoder K, Sarasin A, Kraemer K, McIlhatton M, Bushman F, Fishel R. The DNA repair genes XPB and XPD defend cells from retroviral infection. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(12):4622-7.
170. Li Y, Flanagan PM, Tschochner H, Kornberg RD. RNA polymerase-ii initiation-factor interactions and transcription start site selection. Science. 1994;263(5148):805-7.
171. Zhang DY, Carson DJ, Ma J. The role of TFIIB-RNA polymerase II interaction in start site selection in yeast cells. Nucleic Acids Res. 2002;30(14):3078-85.
172. Dayton AI, Sodroski JG, Rosen CA, Goh WC, Haseltine WA. The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. Cell. 1986;44(6):941-7.
173. Fisher AG, Feinberg MB, Josephs SF, Harper ME, Marselle LM, Reyes G, et al. The trans-activator gene of HTLV-III is essential for virus replication. Nature. 1986;320(6060):367-71.
174. Feng S, Holland EC. HIV-1 tat trans-activation requires the loop sequence within tar. Nature. 1988;334(6178):165-7.
175. Roy S, Delling U, Chen CH, Rosen CA, Sonenberg N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev. 1990;4(8):1365-73.

176. Dingwall C, Ernberg I, Gait MJ, Green SM, Heaphy S, Karn J, et al. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc Natl Acad Sci U S A.* 1989;86(18):6925-9.
177. Garcia-Martinez LF, Mavankal G, Neveu JM, Lane WS, Ivanov D, Gaynor RB. Purification of a Tat-associated kinase reveals a TFIID complex that modulates HIV-1 transcription. *Embo j.* 1997;16(10):2836-50.
178. Feinberg MB, Baltimore D, Frankel AD. The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. *Proc Natl Acad Sci U S A.* 1991;88(9):4045-9.
179. Asamitsu K, Fujinaga K, Okamoto T. HIV Tat/P-TEFb Interaction: A Potential Target for Novel Anti-HIV Therapies. *Molecules.* 2018;23(4):933.
180. Tiley LS, Madore SJ, Malim MH, Cullen BR. The VP16 transcription activation domain is functional when targeted to a promoter-proximal RNA sequence. *Genes Dev.* 1992;6(11):2077-87.
181. Southgate CD, Green MR. The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function. *Genes Dev.* 1991;5(12b):2496-507.
182. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell.* 1998;92(4):451-62.
183. Zhu Y, Pe'ery T, Peng J, Ramanathan Y, Marshall N, Marshall T, et al. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* 1997;11(20):2622-32.

184. Mancebo HS, Lee G, Flygare J, Tomassini J, Luu P, Zhu Y, et al. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev.* 1997;11(20):2633-44.
185. Zhou M, Deng L, Kashanchi F, Brady JN, Shatkin AJ, Kumar A. The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proc Natl Acad Sci U S A.* 2003;100(22):12666-71.
186. Hong SW, Hong SM, Yoo JW, Lee YC, Kim S, Lis JT, Lee DK. Phosphorylation of the RNA polymerase II C-terminal domain by TFIIF kinase is not essential for transcription of *Saccharomyces cerevisiae* genome. *Proc Natl Acad Sci U S A.* 2009 Aug 25;106(34):14276-80
187. Malim MH, Cullen BR. HIV-1 structural gene-expression requires the binding of multiple rev monomers to the viral rre - implications for hiv-1 latency. *Cell.* 1991;65(2):241-8.
188. MATYSIAK J. Intégration du VIH-1: Contrôle et régulation de l'interaction fonctionnelle entre l'intégrase et la chromatine. Thèse de doctorat, Université de Bordeaux; France, 2016.
189. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature.* 1989;338(6212):254-7.
190. Sundquist WI, Krausslich HG. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine.* 2012;2(7):24.
191. Gheysen D, Jacobs E, Deforesta F, Thiriart C, Francotte M, Thines D, et al. ASsembly and release of hiv-1 precursor pr55gag virus-like particles from recombinant baculovirus infected insect cells. *Cell.* 1989;59(1):103-12.

192. Veillette M. Rôle de la conformation des glycoprotéines de l'enveloppe du VIH-1 dans la réponse cytotoxique cellulaire dépendante des anticorps et impact des protéines virales Nef et Vpu. Thèse de doctorat, Université de Montréal, Canada, 2016.
193. Finzi A, Orthwein A, Mercier J, Cohen EA. Productive human immunodeficiency virus type 1 assembly takes place at the plasma membrane. *J Virol.* 2007;81(14):7476-90.
194. Hermida-Matsumoto L, Resh MD. Localization of human immunodeficiency virus type 1 Gag and Env at the plasma membrane by confocal imaging. *J Virol.* 2000;74(18):8670-9.
195. Ivanchenko S, Godinez WJ, Lampe M, Krausslich HG, Eils R, Rohr K, et al. Dynamics of HIV-1 Assembly and Release. *Plos Pathogens.* 2009;5(11):12.
196. Corbin A, Grigorov B, Roingard P, Darlix JL, Muriaux D. Revisiting HIV-1 assembly. *M S-Medecine Sciences.* 2008;24(1):49-55.
197. Ghazi-Tabatabai S, Saksena S, Short JM, Pobbati AV, Veprintsev DB, Crowther RA, et al. Structure and disassembly of filaments formed by the ESCRT-III subunit Vps24. *Structure.* 2008;16(9):1345-56.
198. Hanson PI, Roth R, Lin Y, Heuser JE. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. *Journal of Cell Biology.* 2008;180(2):389-402.
199. Teis D, Saksena S, Emr SD. Ordered Assembly of the ESCRT-III Complex on Endosomes Is Required to Sequester Cargo during MVB Formation. *Developmental Cell.* 2008;15(4):578-89.
200. Weiss ER, Gottlinger H. The Role of Cellular Factors in Promoting HIV Budding. *Journal of Molecular Biology.* 2011;410(4):525-33.
201. Elia N, Sougrat R, Spurlin TA, Hurley JH, Lippincott-Schwartz J. Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its

- role in abscission. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(12):4846-51.
202. Jouvenet N, Simon SM, Bieniasz PD. Visualizing HIV-1 Assembly. *Journal of Molecular Biology*. 2011;410(4):501-11.
203. Van Engelenburg SB, Shtengel G, Sengupta P, Waki K, Jarnik M, Ablan SD, et al. Distribution of ESCRT Machinery at HIV Assembly Sites Reveals Virus Scaffolding of ESCRT Subunits. *Science*. 2014;343(6171):653-6.
204. Bieniasz PD. Late budding domains and host proteins in enveloped virus release. *Virology*. 2006;344(1):55-63.
205. Morita E, Sundquist WI. Retrovirus budding. *Annual Review of Cell and Developmental Biology*. 2004;20:395-425.
206. Morita E, Sandrin V, Chung HY, Morham SG, Gygi SP, Rodesch CK, et al. Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *Embo Journal*. 2007;26(19):4215-27.
207. Scarlata S, Carter C. Role of HIV-1 Gag domains in viral assembly. *Biochimica Et Biophysica Acta-Biomembranes*. 2003;1614(1):62-72.
208. Ono A, Freed EO. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(24):13925-30.
209. Hill M, Tachedjian G, Mak J. The packaging and maturation of the HIV-1 pol proteins. *Current HIV research*. 2005;3(1):73-85.
210. Checkley MA, Luttmann BG, Freed EO. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *Journal of Molecular Biology*. 2011;410(4):582-608.

211. Shcherbakov DN, Bakulina AY, Karpenko LI, Ilyichev AA. Broadly Neutralizing Antibodies against HIV-1 As a Novel Aspect of the Immune Response. *Acta Naturae*. 2015;7(4):11-21.
212. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu Rev Immunol*. 1999;17:657-700.
213. Daecke J, Fackler OT, Dittmar MT, Krausslich HG. Involvement of clathrin-mediated endocytosis in human immunodeficiency virus type 1 entry. *J Virol*. 2005;79(3):1581-94.
214. Miyauchi K, Kim Y, Latinovic O, Morozov V, Melikyan GB. HIV Enters Cells via Endocytosis and Dynamin-Dependent Fusion with Endosomes. *Cell*. 2009;137(3):433-44.
215. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. *Nature*. 1997;387(6631):426-30.
216. Center RJ, Leapman RD, Lebowitz J, Arthur LO, Earl PL, Moss B. Oligomeric structure of the human immunodeficiency virus type I envelope protein on the virion surface. *J Virol*. 2002;76(15):7863-7.
217. Haim H, Salas I, Sodroski J. Proteolytic Processing of the Human Immunodeficiency Virus Envelope Glycoprotein Precursor Decreases Conformational Flexibility. *Journal of Virology*. 2013;87(3):1884-9.
218. Moulard M, Hallenberger S, Garten W, Klenk HD. Processing and routage of HIV glycoproteins by furin to the cell surface. *Virus Research*. 1999;60(1):55-65.
219. Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W. Inhibition of furin-mediated cleavage activation of hiv-1 glycoprotein-GP160. *Nature*. 1992;360(6402):358-61.

220. Land A, Braakman I. Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum. *Biochimie*. 2001;83(8):783-90.
221. Moulard M, Decroly E. Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim Biophys Acta-Rev Biomembr*. 2000;1469(3):121-32.
222. Hunt R. Human immunodeficiency virus and aids structure: the genome and proteins of HIV. *Virology-chapter seven, part nine*. In: *Microbiology and Immunology On-line*. University of South Carolina School of Medicine August 28, 2016.
223. Yu F, Lu L, Du LY, Zhu XJ, Debnath AK, Jiang SB. Approaches for Identification of HIV-1 Entry Inhibitors Targeting gp41 Pocket. *Viruses-Basel*. 2013;5(1):127-49.
224. www.hiv.lanl.gov. Env Feature database: HXB2 Genome Annotation. 2017.
225. Yu F, Lu L, Du L, Zhu X, Debnath AK, Jiang S. Approaches for Identification of HIV-1 Entry Inhibitors Targeting gp41 Pocket. *Viruses-Basel*. 2013;5(1):127-49.
226. Douglas NW, Munro GH, Daniels RS. HIV/SIV glycoproteins: Structure-function relationships. *Journal of Molecular Biology*. 1997;273(1):122-49.
227. Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, Hendrickson WA, et al. Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci U S A*. 2000;97(16):9026-31.
228. Chen B, Vogan EM, Gong HY, Skehel JJ, Wiley DC, Harrison SC. Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature*. 2005;433(7028):834-41.
229. Pancera M. Structure, function and antigenicity of the HIV1 envelope glycoproteins. *Virologie (Montrouge)*. 2005;9(6):457-72.

230. Wyatt R KP, Hendrickson WA, Sodroski. JG. Structure of the Core of the HIV-1 gp120 Exterior Envelope Glycoprotein. In: Theoretical Biol. & Biophys. Group LANL, LA-UR 01-3860 [Read it online: SODROSKI], editor.: Los Alamos National Laboratory (LANL); 1998.
231. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. 1998;393(6686):648-59.
232. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type-1 recombinant human-immunodeficiency-virus envelope glycoprotein (gp120) expressed in chinese hamster ovary cells. *Journal of Biological Chemistry*. 1990;265(18):10373-82.
233. Willey RL, Rutledge RA, Dias S, Folks T, Theodore T, Buckler CE, et al. Identification of conserved and divergent domains within the envelope gene of the acquired-immunodeficiency-syndrome retrovirus. *Proc Natl Acad Sci U S A*. 1986;83(14):5038-42.
234. Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H, et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III LAV, the retrovirus of aids. *Cell*. 1986;45(5):637-48.
235. Moore JP, Willey RL, Lewis GK, Robinson J, Sodroski J. Immunological evidence for interactions between the first, 2nd, and 5th conserved domains of the gp120 surface glycoprotein of human-immunodeficiency-virus type-1. *Journal of Virology*. 1994;68(11):6836-47.
236. Yuan TT, Li JJ, Zhang MY. HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry. *PLoS One*. 2013;8(8):9.

237. Allan JS, Coligan JE, Barin F, McLane MF, Sodroski JG, Rosen CA, et al. Major glycoprotein antigens that induce antibodies in aids patients are encoded by HTLV-III. *Science*. 1985;228(4703):1091-4.
238. Montefiori DC, Robinson WE, Mitchell WM. ROle of protein n-glycosylation in pathogenesis of human immunodeficiency virus type-1. *Proc Natl Acad Sci U S A*. 1988;85(23):9248-52.
239. Bernstein HB, Tucker SP, Hunter E, Schutzbach JS, Compans RW. Human-immunodeficiency-virus type-1 envelope glycoprotein is modified by o-linked oligosaccharides. *J Virol*. 1994;68(1):463-8.
240. Termini JM, Church ES, Silver ZA, Haslam SM, Dell A, Desrosiers RC. Human Immunodeficiency Virus and Simian Immunodeficiency Virus Maintain High Levels of Infectivity in the Complete Absence of Mucin-Type O-Glycosylation. *J Virol*. 2017;91(19):13.
241. Gallaher WR, Ball JM, Garry RF, Griffin MC, Montelaro RC. A general-model for the transmembrane proteins of hiv and other retroviruses. *Aids Research and Human Retroviruses*. 1989;5(4):431-40.
242. Zolla-Pazner S. Identifying epitopes of HIV-1 that induce protective antibodies. *Nature Reviews Immunology*. 2004;4(3):199-210.
243. Gallo SA, Finnegan CM, Viard M, Raviv Y, Dimitrov A, Rawat SS, et al. The HIV Env-mediated fusion reaction. *Biochimica Et Biophysica Acta-Biomembranes*. 2003;1614(1):36-50.
244. Lodge R, Gottlinger H, Gabuzda D, Cohen EA, Lemay G. The intracytoplasmic domain of gp41 mediates polarized budding of human-immunodeficiency-virus type-1 in mdck cells. *J Virol*. 1994;68(8):4857-61.

245. Blot G, Janvier K, Le Panse S, Benarous R, Berlioz-Torrent C. Targeting of the human immunodeficiency virus type 1 envelope to the trans-Golgi network through binding to TIP47 is required for Env incorporation into virions and infectivity. *J Virol.* 2003;77(12):6931-45.
246. Freed EO, Martin MA. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *J Virol.* 1996;70(1):341-51.
247. Kalia V, Sarkar S, Gupta P, Montelaro RC. Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. *J Virol.* 2003;77(6):3634-46.
248. Miller MA, Mietzner TA, Cloyd MW, Robey WG, Montelaro RC. Identification of a calmodulin-binding and inhibitory peptide domain in the hiv-1 transmembrane glycoprotein. *Aids Research and Human Retroviruses.* 1993;9(11):1057-66.
249. Rice AP, Carlotti F. Mutational analysis of the conserved cysteine-rich region of the human-immunodeficiency-virus type-1 tat protein. *J Virol.* 1990;64(4):1864-8.
250. Lopez-Verges S, Camus G, Blot G, Beauvoir R, Benarous R, Berlioz-Torrent C. Tail-interacting protein TIP47 is a connector between Gag and Env and is required for Env incorporation into HIV-1 virions. *Proc Natl Acad Sci U S A.* 2006;103(40):14947-52.
251. Brasseur R, Vandenbranden M, Cornet B, Burny A, Ruyschaert JM. Orientation into the lipid bilayer of an asymmetric amphipathic helical peptide located at the n-terminus of viral fusion proteins. *Biochimica Et Biophysica Acta.* 1990;1029(2):267-73.
252. Pierson TC, Doms RW. HIV-1 entry and its inhibition. *Cellular Factors Involved in Early Steps of Retroviral Replicaton.* 2003;281:1-27.

253. Eckert DM, Kim PS. Mechanisms of viral membrane fusion and its inhibition. *Annual Review of Biochemistry*. 2001;70:777-810.
254. Melikyan GB, Markosyan RM, Hemmati H, Delmedico MK, Lambert DM, Cohen FS. Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. *Journal of Cell Biology*. 2000;151(2):413-23.
255. Colman PM, Lawrence MC. The structural biology of type I viral membrane fusion. *Nature Reviews Molecular Cell Biology*. 2003;4(4):309-19.
256. Celum CL, Coombs RW, Lafferty W, Inui TS, Louie PH, Gates CA, et al. Indeterminate human-immunodeficiency-virus type-1 western blots - seroconversion risk, specificity of supplemental tests, and an algorithm for evaluation. *Journal of Infectious Diseases*. 1991;164(4):656-64.
257. L. Morand-Joubert. La « non-progression » de l'infection à VIH : une origine multifactorielle . *Virologie*. 2001;5(2):113-9.
258. Buchbinder SP, Katz MH, Hessel NA, Omalley PM, Holmberg SD. Long-term hiv-1 infection without immunological progression. *Aids*. 1994;8(8):1123-8.
259. Easterbrook PJ. Non-progression in HIV-infection. *Aids*. 1994;8(8):1179-82.
260. Sheppard HW, Lang W, Ascher MS, Vittinghoff E, Winkelstein W. The characterization of nonprogressors - long-term hiv-1 infection with stable CD4+ T-cell levels. *Aids*. 1993;7(9):1159-66.
261. Baltimore D. Lessons from people with nonprogressive hiv-infection. *New England Journal of Medicine*. 1995;332(4):259-60.
262. Pantaleo G, Menzo S, Vaccarezza M, Graziosi C, Cohen OJ, Demarest JF, et al. Studies in subjects with long-term nonprogressive human-immunodeficiency-virus infection. *New England Journal of Medicine*. 1995;332(4):209-16.

263. Laeyendecker O, Redd AD, Lutalo T, Gray RH, Wawer M, Ssempijja V, et al. Frequency of Long-Term Nonprogressors in HIV-1 Seroconverters From Rakai Uganda. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*. 2009;52(3):316-9.
264. F. Okulicz J. Elite Controllers and Long-term Nonprogressors: Models for HIV Vaccine Development? 2012.
265. Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity*. 2007;27(3):406-16.
266. Okulicz JF, Marconi VC, Landrum ML, Wegner S, Weintrob A, Ganesan A, et al. Clinical Outcomes of Elite Controllers, Viremic Controllers, and Long-Term Nonprogressors in the US Department of Defense HIV Natural History Study. *Journal of Infectious Diseases*. 2009;200(11):1714-23.
267. Grabar S, Selinger-Leneman H, Abgrall S, Pialoux G, Weiss L, Costagliola D. Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV. *Aids*. 2009;23(9):1163-9.
268. O'Connell KA, Bailey JR, Blankson JN. Elucidating the elite: mechanisms of control in HIV-1 infection. *Trends in Pharmacological Sciences*. 2009;30(12):631-7.
269. Gillespie SL, Chinen J, Paul ME, Shearer WT. 39 - Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome. In: Rich RR, Fleisher TA, Shearer WT, Schroeder HW, Frew AJ, Weyand CM, editors. *Clinical Immunology (Fifth Edition)*. London: Content Repository Only!; 2019. p. 545-60.e1.
270. Ioannidis JPA, Rosenberg PS, Goedert JJ, Ashton LJ, Benfield TL, Buchbinder SP, et al. Effects of CCR5-Delta 32, CCR2-641, and SDF-1 3' A alleles on HIV-1 disease progression: An international meta-analysis of individual-patient data. *Annals of Internal Medicine*. 2001;135(9):782-95.

271. Mulherin SA, O'Brien TR, Ioannidis JPA, Goedert JJ, Buchbinder SP, Coutinho RA, et al. Effects of CCR5-Delta 32 and CCR2-64I alleles on HIV-1 disease progression: the protection varies with duration of infection. *Aids*. 2003;17(3):377-87.
272. Gillespie SL, Paul ME, Chinen J, Shearer WT. 37 - HIV infection and acquired immunodeficiency syndrome. In: Rich RR, Fleisher TA, Shearer WT, Schroeder HW, Frew AJ, Weyand CM, editors. *Clinical Immunology (Fourth Edition)*. London: Content Repository Only!; 2013. p. 465-79.
273. Hutter G, Nowak D, Mossner M, Ganepola S, Mussig A, Allers K, et al. Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. *New England Journal of Medicine*. 2009;360(7):692-8.
274. Allers K, Hutter G, Hofmann J, Loddenkemper C, Rieger K, Thiel E, et al. Evidence for the cure of HIV infection by CCR5 Delta 32/Delta 32 stem cell transplantation. *Blood*. 2011;117(10):2791-9.
275. Gupta RK, Abdul-Jawad S, McCoy LE, Mok HP, Peppas D, Salgado M, et al. HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. *Nature*. 2019;568(7751):244-8.
276. De Pasquale M, Kourteva Y, Allos T, D'Aquila RT. Lower HIV provirus levels are associated with more APOBEC3G protein in blood resting memory CD4+ T lymphocytes of controllers in vivo. *PLoS One*. 2013;8(10):e76002-e.
277. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*. 2002;418(6898):646-50.
278. Learmont JC, Geczy AF, Mills J, Ashton LJ, Raynes-Greenow CH, Garsia RJ, et al. Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort. *N Engl J Med*. 1999;340(22):1715-22.

279. Cohen MS, Gay CL, Busch MP, Hecht FM. The Detection of Acute HIV Infection. *Journal of Infectious Diseases*. 2010;202:S270-S7.
280. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *Aids*. 2003;17(13):1871-9.
281. Robb ML, Ananworanich J. Lessons from acute HIV infection. *Current Opinion in HIV and AIDS*. 2016;11(6):555-60.
282. Nowak MA, Lloyd AL, Vasquez GM, Wiltrout TA, Wahl LM, Bischofberger N, et al. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *Journal of Virology*. 1997;71(10):7518-25.
283. Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV. Viral dynamics of acute HIV-1 infection. *Journal of Experimental Medicine*. 1999;190(6):841-50.
284. Quinn TC. Acute primary HIV infection. *Jama-Journal of the American Medical Association*. 1997;278(1):58-62.
285. Pierson T, McArthur J, Siliciano RF. Reservoirs for HIV-1: Mechanisms for Viral Persistence in the Presence of Antiviral Immune Responses and Antiretroviral Therapy. *Annual Review of Immunology*. 2000;18(1):665-708.
286. Daar ES, Moudgil T, Meyer RD, Ho DD. Transient high-levels of viremia in patients with primary human-immunodeficiency-virus type-1 infection. *New England Journal of Medicine*. 1991;324(14):961-4.
287. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MBA. Virus-specific cd8+ cytotoxic t-lymphocyte activity associated with control of viremia in primary human-immunodeficiency-virus type-1 infection. *Journal of Virology*. 1994;68(9):6103-10.

288. Koup RA, Safrit JT, Cao YZ, Andrews CA, McLeod G, Borkowsky W, et al. Temporal association of cellular immune-responses with the initial control of viremia in primary human-immunodeficiency-virus type-1 syndrome. *Journal of Virology*. 1994;68(7):4650-5.
289. Pantaleo G, Demarest JF, Soudeyns H, Graziosi C, Denis F, Adelsberger JW, et al. Major expansion of cd8+ t-cells with a predominant v-beta usage during the primary immune-response to HIV. *Nature*. 1994;370(6489):463-7.
290. Schmitz JE, Kuroda MJ, Santra S, Sasseville VG, Simon MA, Lifton MA, et al. Control of viremia in simian immunodeficiency virus infection by CD8(+) lymphocytes. *Science*. 1999;283(5403):857-60.
291. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and cd4 lymphocytes in HIV-1 infection. *Nature*. 1995;373(6510):123-6.
292. Kafando A, Fournier E, Serhir B, Martineau C, Doualla-Bell F, Sangare MN, et al. HIV-1 envelope sequence-based diversity measures for identifying recent infections. *Plos One*. 2017;12(12).
293. Serhir B, Hamel D, Doualla-Bell F, Routy JP, Beaulac S-N, Legault M, et al. Performance of Bio-Rad and Limiting Antigen Avidity Assays in Detecting Recent HIV Infections Using the Quebec Primary HIV-1 Infection Cohort. *Plos One*. 2016;11(5).
294. Moyo S, Wilkinson E, Novitsky V, Vandormael A, Gaseitsiwe S, Essex M, et al. Identifying Recent HIV Infections: From Serological Assays to Genomics. *Viruses-Basel*. 2015;7(10):5508-24.
295. Fearon M. The laboratory diagnosis of HIV infections. *The Canadian journal of infectious diseases & medical microbiology*. 2005;16(1):26-30.

296. Suligoi B, Galli C, Massi M, Di Sora F, Sciandra M, Pezzotti P, et al. Precision and accuracy of a procedure for detecting recent human immunodeficiency virus infections by calculating the antibody avidity index by an automated immunoassay-based method. *J Clin Microbiol.* 2002;40(11):4015-20.
297. Zaaijer HL, Vonexeloeblers P, Kraaijeveld T, Altena E, Lelie PN. Early detection of antibodies to hiv-1 by 3rd-generation assays. *Lancet.* 1992;340(8822):770-2.
298. Brust S, Duttmann H, Feldner J, Gurtler L, Thorstensson R, Simon F. Shortening of the diagnostic window with a new combined HIV p24 antigen and anti-HIV-1/2/O screening test. *Journal of Virological Methods.* 2000;90(2):153-65.
299. Henrard DR, Wu S, Phillips J, Wiesner D, Phair J. Detection of p24 antigen with and without immune-complex dissociation for longitudinal monitoring of human-immunodeficiency-virus type-1 infection. *Journal of Clinical Microbiology.* 1995;33(1):72-5.
300. Vallefucoco L, Mazzarella C, Portella G. Fourth generation assays for HIV testing. *Expert Rev Mol Diagn.* 2016;16(7):723-32.
301. Kassarjee R, Pilcher CD, Keating SM, Facente SN, McKinney E, Price MA, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. *Aids.* 2014;28(16):2439-49.
302. Xia X-Y, Ge M, Hsi JH, He X, Ruan Y-H, Wang Z-X, et al. High-Accuracy Identification of Incident HIV-1 Infections Using a Sequence Clustering Based Diversity Measure. *Plos One.* 2014;9(6).
303. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America.* 2008;105(21):7552-7.

304. Xia XY, Ge M, Hsi JH, He X, Ruan YH, Wang ZX, et al. High-Accuracy Identification of Incident HIV-1 Infections Using a Sequence Clustering Based Diversity Measure. *Plos One*. 2014;9(6):9.
305. Ragonnet-Cronin M, Aris-Brosou S, Joanisse I, Merks H, Vallee D, Caminiti K, et al. Genetic Diversity as a Marker for Timing Infection in HIV-Infected Patients: Evaluation of a 6-Month Window and Comparison With BED. *Journal of Infectious Diseases*. 2012;206(5):756-64.
306. Piatak M, Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, et al. High-levels of hiv-1 in plasma during all stages of infection determined by competitive PCR. *Science*. 1993;259(5102):1749-54.
307. Connor RI, Mohri H, Cao YZ, Ho DD. Increased viral burden and cytopathicity correlate temporally with cd4+ t-lymphocyte decline and clinical progression in human-immunodeficiency-virus type 1-infected individuals. *Journal of Virology*. 1993;67(4):1772-7.
308. Reinhart TA, Rogan MJ, Huddleston D, Rausch DM, Eiden LE, Haase AT. Simian immunodeficiency virus burden in tissues and cellular compartments during clinical latency and AIDS. *Journal of Infectious Diseases*. 1997;176(5):1198-208.
309. Chengmayer C, Seto D, Tateno M, Levy JA. Biologic features of hiv-1 that correlate with virulence in the host. *Science*. 1988;240(4848):80-2.
310. Fenyo EM, Morfeldtmanson L, Chiodi F, Lind B, Vongegerfelt A, Albert J, et al. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *Journal of Virology*. 1988;62(11):4414-9.
311. Kimata JT, Kuller L, Anderson DB, Dailey P, Overbaugh J. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence AIDS progression. *Nature Medicine*. 1999;5(5):535-41.

312. Serhir B. VIH confirmation (dépistage effectué avec un test de détection des anticorps) Guide des services : Confirmation sérologique de l'infection à VIH révisé le 2017-07-25. Québec: Laboratoire de santé publique du Québec; 2017.
313. National Center for HIV/AIDS VH, and TB Prevention (U.S.). Division of HIV/AIDS Prevention.; Association of Public Health Laboratories. Recommended laboratory HIV testing algorithm for serum or plasma specimens. Centers for Disease Control and Prevention; Updated January 2018.
314. Sharp PM, Shaw GM, Hahn BH. Simian immunodeficiency virus infection of chimpanzees. *Journal of Virology*. 2005;79(7):3891-902.
315. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science*. 2000;288(5472):1789-96.
316. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. The origins of acquired immune deficiency syndrome viruses: where and when? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*. 2001;356(1410):867-76.
317. Lemey P, Pybus OG, Rambaut A, Drummond AJ, Robertson DL, Roques P, et al. The molecular population genetics of HIV-1 group O. *Genetics*. 2004;167(3):1059-68.
318. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine*. 2011;1(1).
319. Hahn B. The Origin of HIV. In: Robert C. Gallo JMC, Mila Pollock & Bruce D. Walker, editor. *HIV/AIDS Research: Its History and Future* Genentech Center for the History of Molecular Biology and Biotechnology Cold Spring Harbor Laborator: Cold Spring Harbor Laborator Archives Meeting; 13-16 October 2016.

320. Keele BF, Van Heuverswyn F, Li YY, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. 2006;313(5786):523-6.
321. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Philosophical Transactions of the Royal Society B-Biological Sciences*. 2010;365(1552):2487-94.
322. Peeters M, Chaix M-L. Origin and genetic diversity of the Human Immunodeficiency Virus: where does it come from, where is it going? *Virologie*. 2013;17(3):119-31.
323. Gao F, Bailes E, Robertson DL, Chen YL, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*. 1999;397(6718):436-41.
324. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. 2008;455(7213):661-U57.
325. Wertheim JO, Worobey M. Dating the Age of the SIV Lineages That Gave Rise to HIV-1 and HIV-2. *Plos Computational Biology*. 2009;5(5).
326. Sauter D, Hue S, Petit SJ, Plantier JC, Towers GJ, Kirchhoff F, et al. HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef. *Retrovirology*. 2011;8.
327. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemee V, et al. A new human immunodeficiency virus derived from gorillas. *Nature Medicine*. 2009;15(8):871-2.
328. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of Putative HIV-1 Group P in Cameroon. *Journal of Virology*. 2011;85(3):1403-7.
329. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 2014;346(6205):56-61.

330. Santoro MM, Perno CF. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiology*. 2013;2013:481314.
331. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res*. 2012;169(2):415-29.
332. Kresge KJ. IAVI's AIDS vaccine blueprint promotes innovative approaches to evaluating lead candidates. *IAVI Rep*. 2006 Jul-Aug;10(4):20.
333. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science (New York, NY)*. 1988;242(4882):1168-71.
334. Sosin DV, Tchurikov NA. Molecular Mechanisms of HIV-1 Genetic Diversity. *Mol Biol*. 2017;51(4):483-95.
335. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature*. 2003;424(6944):94-8.
336. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse-transcriptase from HIV-1. *Science*. 1988;242(4882):1171-3.
337. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science*. 1988;242(4882):1171-3.
338. Ward D. Reverse transcriptase: Edited by A M Skalka and S P Goff. pp 432. Cold Spring Harbor Laboratory Press, NY. 1993. \$85 ISBN 0-87969-382-7. *Biochemical Education*. 1995;23(1):47-.
339. Berkhout B, Vastenhouw NL, Klasens BI, Huthoff H. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA*. 2001;7(8):1097-114.

340. Mansky LM. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS research and human retroviruses*. 1996;12(4):307-14.
341. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *Journal of virology*. 1995;69(8):5087-94.
342. Gao F, Chen Y, Levy DN, Conway JA, Kepler TB, Hui H. Unselected Mutations in the Human Immunodeficiency Virus Type 1 Genome Are Mostly Nonsynonymous and Often Deleterious. *Journal of virology*. 2004;78(5):2426-33.
343. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *Journal of virology*. 2010;84(19):9864-78.
344. Onafuwa-Nuga A, Telesnitsky A. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiology and molecular biology reviews : MMBR*. 2009;73(3):451-80.
345. Ramirez BC, Simon-Loriere E, Galetto R, Negroni M. Implications of recombination for HIV diversity. *Virus research*. 2008;134(1-2):64-73.
346. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus research*. 2012;169(2):415-29.
347. Holsinger KE. Hardy–Weinberg Law. In: Brenner S, Miller JH, editors. *Encyclopedia of Genetics*. New York: Academic Press; 2001. p. 912-4.
348. Peeters M, Chaix ML, Delaporte E. Genetic diversity and phylogeographic distribution of SIV: how to understand the origin of HIV. *M S-Medecine Sciences*. 2008;24(6-7):621-8.
349. Peeters M, D'Arc M, Delaporte E. Origin and Diversity of Human Retroviruses. *Aids Reviews*. 2014;16(1):23-34.

350. Kijak GH, McCutchan FE. HIV diversity, molecular epidemiology, and the role of recombination. *Current Infectious Disease Reports*. 2005;7(6):480-8.
351. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *Journal of virology*. 2002;76(22):11273-82.
352. Williamson S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. *Molecular biology and evolution*. 2003;20(8):1318-25.
353. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995;373:123.
354. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science (New York, NY)*. 1996;271(5255):1582-6.
355. Joseph SB, Swanstrom R, Kashuba ADM, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nature Reviews Microbiology*. 2015;13(7):414-25.
356. Wu JLW, Patterson-Lomba O, Novitsky V, Pagano M. A Generalized Entropy Measure of Within-Host Viral Diversity for Identifying Recent HIV-1 Infections. *Medicine*. 2015;94(42).
357. Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *Journal of Antimicrobial Chemotherapy*. 2003;51(2):229-40.
358. McBurney SP, Ross TM. Viral sequence diversity: challenges for AIDS vaccine designs. *Expert Review of Vaccines*. 2008;7(9):1405-17.

359. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nature Reviews Genetics*. 2004;5:52.
360. Butler IF, Pandrea I, Marx PA, Apetrei C. HIV genetic diversity: biological and public health consequences. *Current HIV research*. 2007;5(1):23-45.
361. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harb Perspect Med*. 2012;2(8).
362. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, et al. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol*. 2003;77(7):4449-56.
363. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. *The Journal of experimental medicine*. 1997;185(4):621-8.
364. Lee B, Sharron M, Montaner LJ, Weissman D, Doms RW. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proc Natl Acad Sci U S A*. 1999;96(9):5215-20.
365. Joseph SB, Swanstrom R. The evolution of HIV-1 entry phenotypes as a guide to changing target cells. *Journal of Leukocyte Biology*. 2018;103(3):421-31.
366. Schnell G, Joseph S, Spudich S, Price RW, Swanstrom R. HIV-1 replication in the central nervous system occurs in two distinct cell types. *PLoS Pathog*. 2011;7(10):e1002286.
367. Joseph SB, Arrildt KT, Swanstrom AE, Schnell G, Lee B, Hoxie JA, et al. Quantification of entry phenotypes of macrophage-tropic HIV-1 across a wide range of CD4 densities. *J Virol*. 2014;88(4):1858-69.
368. Lin NH, Kuritzkes DR. Tropism Testing in the Clinical Management of HIV-1 Infection. *Current opinion in HIV and AIDS*. 2009;4(6):481-7.

369. Jensen MA, van 't Wout AB. Predicting HIV-1 coreceptor usage with sequence analysis. *AIDS reviews*. 2003;5(2):104-12.
370. Weber J, Piontkivska H, Quinones-Mateu ME. HIV type 1 tropism and inhibitors of viral entry: clinical implications. *AIDS reviews*. 2006;8(2):60-77.
371. Whitcomb JM, Huang W, Fransen S, Limoli K, Toma J, Wrin T, et al. Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism. *Antimicrobial agents and chemotherapy*. 2007;51(2):566-75.
372. Schuitemaker H, Koot M, Kootstra NA, Dercksen MW, de Goede RE, van Steenwijk RP, et al. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocyctotropic to T-cell-tropic virus population. *J Virol*. 1992;66(3):1354-60.
373. Recordon-Pinson P, Soulie C, Flandre P, Descamps D, Lazrek M, Charpentier C, et al. Evaluation of the genotypic prediction of HIV-1 coreceptor use versus a phenotypic assay and correlation with the virological response to maraviroc: the ANRS GenoTropism study. *Antimicrobial agents and chemotherapy*. 2010;54(8):3335-40.
374. Raymond S, Recordon-Pinson P, Saliou A, Delobel P, Nicot F, Descamps D, et al. Improved V3 genotyping with duplicate PCR amplification for determining HIV-1 tropism. *The Journal of antimicrobial chemotherapy*. 2011;66(9):1972-5.
375. Johnston SH, Lobritz MA, Nguyen S, Lassen K, Delair S, Posta F, et al. A quantitative affinity-profiling system that reveals distinct CD4/CCR5 usage patterns among human immunodeficiency virus type 1 and simian immunodeficiency virus strains. *J Virol*. 2009;83(21):11016-26.

376. Duenas-Decamp MJ, Peters PJ, Burton D, Clapham PR. Determinants flanking the CD4 binding loop modulate macrophage tropism of human immunodeficiency virus type 1 R5 envelopes. *J Virol.* 2009;83(6):2575-83.
377. Dunfee RL, Thomas ER, Gorry PR, Wang J, Taylor J, Kunstman K, et al. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. *Proc Natl Acad Sci U S A.* 2006;103(41):15160-5.
378. Gorry PR, Taylor J, Holm GH, Mehle A, Morgan T, Cayabyab M, et al. Increased CCR5 affinity and reduced CCR5/CD4 dependence of a neurovirulent primary human immunodeficiency virus type 1 isolate. *J Virol.* 2002;76(12):6277-92.
379. Martin-Garcia J, Cao W, Varela-Rohena A, Plassmeyer ML, Gonzalez-Scarano F. HIV-1 tropism for the central nervous system: Brain-derived envelope glycoproteins with lower CD4 dependence and reduced sensitivity to a fusion inhibitor. *Virology.* 2006;346(1):169-79.
380. Peters PJ, Bhattacharya J, Hibbitts S, Dittmar MT, Simmons G, Bell J, et al. Biological analysis of human immunodeficiency virus type 1 R5 envelopes amplified from brain and lymph node tissues of AIDS patients with neuropathology reveals two distinct tropism phenotypes and identifies envelopes in the brain that confer an enhanced tropism and fusogenicity for macrophages. *J Virol.* 2004;78(13):6915-26.
381. Peters PJ, Sullivan WM, Duenas-Decamp MJ, Bhattacharya J, Ankghuambom C, Brown R, et al. Non-macrophage-tropic human immunodeficiency virus type 1 R5 envelopes predominate in blood, lymph nodes, and semen: implications for transmission and pathogenesis. *J Virol.* 2006;80(13):6324-32.
382. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A.* 2008;105(21):7552-7.

383. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol.* 2005;43(1):406-13.
384. Shriner D, Rodrigo AG, Nickle DC, Mullins JI. Pervasive genomic recombination of HIV-1 in vivo. *Genetics.* 2004;167(4):1573-83.
385. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol.* 1990;64(12):5840-50.
386. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol.* 2008;82(8):3952-70.
387. Keele BF. Identifying and characterizing recently transmitted viruses. *Curr Opin HIV AIDS.* 2010;5(4):327-34.
388. Duncan CJ, Sattentau QJ. Viral determinants of HIV-1 macrophage tropism. *Viruses.* 2011;3(11):2255-79.
389. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol.* 2007;25(12):1407-10.
390. Montagna C, De Crignis E, Bon I, Re MC, Mezzaroma I, Turriziani O, et al. V3 Net Charge: Additional Tool in HIV-1 Tropism Prediction. *Aids Res Hum Retrovir.* 2014;30(12):1203-12.

391. Zhou S, Bednar MM, Sturdevant CB, Hauser BM, Swanstrom R. Deep Sequencing of the HIV-1 env Gene Reveals Discrete X4 Lineages and Linkage Disequilibrium between X4 and R5 Viruses in the V1/V2 and V3 Variable Regions. *J Virol.* 2016;90(16):7142-58.
392. Fouchier RA, Groenink M, Kootstra NA, Tersmette M, Huisman HG, Miedema F, et al. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol.* 1992;66(5):3183-7.
393. Scarlatti G, Tresoldi E, Bjorndal A, Fredriksson R, Colognesi C, Deng HK, et al. In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nature medicine.* 1997;3(11):1259-65.
394. Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, et al. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science (New York, NY).* 1998;280(5371):1949-53.
395. Jensen MA, Coetzer M, van 't Wout AB, Morris L, Mullins JI. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol.* 2006;80(10):4698-704.
396. Cho MW, Lee MK, Carney MC, Berson JF, Doms RW, Martin MA. Identification of Determinants on a Dualtropic Human Immunodeficiency Virus Type 1 Envelope Glycoprotein That Confer Usage of CXCR4. *Journal of Virology.* 1998;72(3):2509-15.
397. Dwyer DE, Ge YC, Bolton WV, Wang B, Cunningham AL, Saksena NK. Subtype B isolates of human immunodeficiency virus type 1 detected in Australia. *Annals Academy of Medicine Singapore.* 1996;25(2):188-91.
398. Jeeninga RE, Hoogenkamp M, Armand-Ugon M, de Baar M, Verhoef K, Berkhout B. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J Virol.* 2000;74(8):3740-51.

399. Kim K, Dauphin A, Komurlu S, McCauley SM, Yurkovetskiy L, Carbone C, et al. Cyclophilin A protects HIV-1 from restriction by human TRIM5alpha. *Nat Microbiol.* 2019;4(12):2044-51.
400. Matsuoka S, Dam E, Lecossier D, Clavel F, Hance AJ. Modulation of HIV-1 infectivity and cyclophilin A-dependence by Gag sequence and target cell type. *Retrovirology.* 2009;6:21.
401. Sokolskaja E, Sayah DM, Luban J. Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *J Virol.* 2004;78(23):12800-8.
402. Braaten D, Luban J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. *EMBO J.* 2001;20(6):1300-9.
403. Cousins MM, Konikoff J, Laeyendecker O, Celum C, Buchbinder SP, Seage GR, III, et al. HIV Diversity as a Biomarker for HIV Incidence Estimation: Including a High-Resolution Melting Diversity Assay in a Multiassay Algorithm. *Journal of Clinical Microbiology.* 2014;52(1):115-21.
404. Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, Kalams SA, et al. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science.* 1997;278(5342):1447-50.
405. Sacha JB, Chung C, Rakasz EG, Spencer SP, Jonas AK, Bean AT, et al. Gag-specific CD8+ T lymphocytes recognize infected cells before AIDS-virus integration and viral protein expression. *Journal of immunology (Baltimore, Md : 1950).* 2007;178(5):2746-54.
406. Streeck H, Jolin JS, Qi Y, Yassine-Diab B, Johnson RC, Kwon DS, et al. Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *J Virol.* 2009;83(15):7641-8.

407. Altfeld M, Rosenberg ES, Shankarappa R, Mukherjee JS, Hecht FM, Eldridge RL, et al. Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. *The Journal of experimental medicine*. 2001;193(2):169-80.
408. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature medicine*. 1997;3(2):205-11.
409. Evans DT, O'Connor DH, Jing P, Dzuris JL, Sidney J, da Silva J, et al. Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nature medicine*. 1999;5(11):1270-6.
410. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Gansarov VV, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *The Journal of experimental medicine*. 2009;206(6):1253-72.
411. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(5):1890-5.
412. Nowak MA, Anderson RM, McLean AR, Wolfs TF, Goudsmit J, May RM. Antigenic diversity thresholds and the development of AIDS. *Science*. 1991;254(5034):963-9.
413. Sagar M, Kirkegaard E, Long EM, Celum C, Buchbinder S, Daar ES, et al. Human Immunodeficiency Virus Type 1 (HIV-1) Diversity at Time of Infection Is Not Restricted to Certain Risk Groups or Specific HIV-1 Subtypes. *J Virol*. 2004;78(13):7279-83.
414. Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, et al. Dual HIV-1 infection associated with rapid disease progression. *Lancet (London, England)*. 2004;363(9409):619-22.

415. Hazenberg MD, Otto SA, van Benthem BH, Roos MT, Coutinho RA, Lange JM, et al. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. *AIDS (London, England)*. 2003;17(13):1881-8.
416. Allen TM, Altfeld M. HIV-1 superinfection. *The Journal of allergy and clinical immunology*. 2003;112(5):829-35; quiz 36.
417. Fang G, Weiser B, Kuiken C, Philpott SM, Rowland-Jones S, Plummer F, et al. Recombination following superinfection by HIV-1. *AIDS (London, England)*. 2004;18(2):153-9.
418. Streeck H, Li B, Poon AFY, Schneidewind A, Gladden AD, Power KA, et al. Immune-driven recombination and loss of control after HIV superinfection. *The Journal of experimental medicine*. 2008;205(8):1789-96.
419. Jost S, Bernard MC, Kaiser L, Yerly S, Hirschel B, Samri A, et al. A patient with HIV-1 superinfection. *The New England journal of medicine*. 2002;347(10):731-6.
420. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous Substitution Rates Predict HIV Disease Progression as a Result of Underlying Replication Dynamics. *PLoS Comput Biol*. 2007;3(2):e29.
421. Levesque MC, Moody MA, Hwang KK, Marshall DJ, Whitesides JF, Amos JD, et al. Polyclonal B cell differentiation and loss of gastrointestinal tract germinal centers in the earliest stages of HIV-1 infection. *PLoS medicine*. 2009;6(7):e1000107.
422. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*. 2002;420(6916):678-82.
423. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. *Nature*. 2003;422(6929):307-12.

424. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science (New York, NY)*. 1997;278(5341):1295-300.
425. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci U S A*. 1997;94(24):13193-7.
426. Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, Cane P, et al. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS medicine*. 2005;2(4):e112.
427. Clavel F, Hance AJ. HIV drug resistance. *The New England journal of medicine*. 2004;350(10):1023-35.
428. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. *Viruses-Basel*. 2010;2(2):503-31.
429. Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS (Auckland, NZ)*. 2015;7:95-104.
430. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. *Viruses*. 2010;2(2):503-31.
431. van de Vijver DA, Wensing AM, Angarano G, Asjo B, Balotta C, Boeri E, et al. The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes. *Journal of acquired immune deficiency syndromes (1999)*. 2006;41(3):352-60.
432. Lech WJ, Wang G, Yang YL, Chee Y, Dorman K, McCrae D, et al. In vivo sequence diversity of the protease of human immunodeficiency virus type 1: presence of protease inhibitor-resistant variants in untreated subjects. *J Virol*. 1996;70(3):2038-43.

433. Loussert-Ajaka I, Ly TD, Chaix ML, Ingrand D, Saragosti S, Courouce AM, et al. HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. *Lancet (London, England)*. 1994;343(8910):1393-4.
434. Schable C, Zekeng L, Pau CP, Hu D, Kaptue L, Gurtler L, et al. Sensitivity of United States HIV antibody tests for detection of HIV-1 group O infections. *Lancet (London, England)*. 1994;344(8933):1333-4.
435. Apetrei C, Loussert-Ajaka I, Descamps D, Damond F, Saragosti S, Brun-Vezinet F, et al. Lack of screening test sensitivity during HIV-1 non-subtype B seroconversions. *AIDS (London, England)*. 1996;10(14):F57-60.
436. Dorn J, Masciotra S, Yang C, Downing R, Biryahwaho B, Mastro TD, et al. Analysis of genetic variability within the immunodominant epitopes of envelope gp41 from human immunodeficiency virus type 1 (HIV-1) group M and its impact on HIV-1 antibody detection. *Journal of clinical microbiology*. 2000;38(2):773-80.
437. Galli RA, Castriciano S, Fearon M, Major C, Choi KW, Mahony J, et al. Performance characteristics of recombinant enzyme immunoassay to detect antibodies to human immunodeficiency virus type 1 (HIV-1) and HIV-2 and to measure early antibody responses in seroconverting patients. *Journal of clinical microbiology*. 1996;34(4):999-1002.
438. van Binsbergen J, de Rijk D, Peels H, Dries C, Scherders J, Koolen M, et al. Evaluation of a new third generation anti-HIV-1/anti-HIV-2 assay with increased sensitivity for HIV-1 group O. *Journal of virological methods*. 1996;60(2):131-7.
439. van Binsbergen J, Keur W, vd Graaf M, Siebelink A, Jacobs A, de Rijk D, et al. Reactivity of a new HIV-1 group O third generation A-HIV-1/-2 assay with an unusual HIV-1 seroconversion panel and HIV-1 group O/group M subtyped samples. *Journal of virological methods*. 1997;69(1-2):29-37.

440. Branson BM, Stekler JD. Detection of acute HIV infection: we can't close the window. *The Journal of infectious diseases*. 2012;205(4):521-4.
441. Alexander TS. Human Immunodeficiency Virus Diagnostic Testing: 30 Years of Evolution. *Clinical and Vaccine Immunology*. 2016;23(4):249-53.
442. Chavez P, Wesolowski L, Patel P, Delaney K, Owen SM. Evaluation of the performance of the Abbott ARCHITECT HIV Ag/Ab Combo Assay. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2011;52 Suppl 1:S51-5.
443. Ly TD, Plantier JC, Leballais L, Gonzalo S, Lemee V, Laperche S. The variable sensitivity of HIV Ag/Ab combination assays in the detection of p24Ag according to genotype could compromise the diagnosis of early HIV infection. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2012;55(2):121-7.
444. Gaudy C, Moreau A, Brunet S, Descamps JM, Deleplanque P, Brand D, et al. Subtype B human immunodeficiency virus (HIV) type 1 mutant that escapes detection in a fourth-generation immunoassay for HIV infection. *Journal of clinical microbiology*. 2004;42(6):2847-9.
445. Salmona M, Delarue S, Delaugerre C, Simon F, Maylin S. Clinical Evaluation of BioPlex 2200 HIV Ag-Ab, an Automated Screening Method Providing Discrete Detection of HIV-1 p24 Antigen, HIV-1 Antibody, and HIV-2 Antibody. *Journal of clinical microbiology*. 2014;52(1):103-7.
446. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, et al. Diversity Considerations in HIV-1 Vaccine Selection. *Science*. 2002;296(5577):2354-60.
447. Fuchs SP, Desrosiers RC. Promise and problems associated with the use of recombinant AAV for the delivery of anti-HIV antibodies. *Molecular Therapy - Methods & Clinical Development*. 2016;3.

448. Morison L, Buve A, Zekeng L, Heyndrickx L, Anagonou S, Musonda R, et al. HIV-1 subtypes and the HIV epidemics in four cities in sub-Saharan Africa. *AIDS (London, England)*. 2001;15 Suppl 4:S109-16.
449. D'Souza MP, Livnat D, Bradac JA, Bridges SH. Evaluation of monoclonal antibodies to human immunodeficiency virus type 1 primary isolates by neutralization assays: performance criteria for selecting candidate antibodies for clinical trials. AIDS Clinical Trials Group Antibody Selection Working Group. *The Journal of infectious diseases*. 1997;175(5):1056-62.
450. Goulder PJ, Sewell AK, Lalloo DG, Price DA, Whelan JA, Evans J, et al. Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation. *The Journal of experimental medicine*. 1997;185(8):1423-33.
451. Beekman NJ, van Veelen PA, van Hall T, Neisig A, Sijts A, Camps M, et al. Abrogation of CTL epitope processing by single amino acid substitution flanking the C-terminal proteasome cleavage site. *Journal of immunology (Baltimore, Md : 1950)*. 2000;164(4):1898-905.
452. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin*. 2001;58(1):19-42.
453. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science*. 2000;288(5463):55-7.
454. Simon F, Mauclore P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine*. 1998;4(9):1032-7.

455. Foley B LT, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S, and Korber B. HIV Sequence Database Compendium. NM, LA-UR 17-25240.: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2017.
456. Jia L, Li L, Li HP, Liu SY, Wang XL, Bao ZY, et al. Recombination Pattern Reanalysis of Some HIV-1 Circulating Recombination Forms Suggest the Necessity and Difficulty of Revision (vol 9, e107349, 2014). Plos One. 2016;11(5).
457. McCutchan FE, Salminen MO, Carr JK, Burke DS. HIV-1 genetic diversity. Aids. 1996;10:S13-S20.
458. Ayouba A, Akoua-Koffi C, Calvignac-Spencer S, Esteban A, Locatelli S, Li H, et al. Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Cote d'Ivoire. Aids. 2013;27(15):2488-91.
459. Faria NR, Hodges-Mameletzis I, Silva JC, Rodes B, Erasmus S, Paolucci S, et al. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. Journal of General Virology. 2012;93:889-99.
460. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An african primate lentivirus (sivsm) closely related to HIV-2. Nature. 1989;339(6223):389-92.
461. Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, et al. Human infection by genetically diverse sivsm-related HIV-2 in west Africa. Nature. 1992;358(6386):495-9.
462. Ibe S, Yokomaku Y, Shiino T, Tanaka R, Hattori J, Fujisaki S, et al. HIV-2 CRF01_AB: First Circulating Recombinant Form of HIV-2. J AIDS-Journal of Acquired Immune Deficiency Syndromes. 2010;54(3):241-7.

463. Dilley KA, Ni N, Nikolaitchik OA, Chen J, Galli A, Hu W-S. Determining the Frequency and Mechanisms of HIV-1 and HIV-2 RNA Copackaging by Single-Virion Analysis. *Journal of Virology*. 2011;85(20):10499-508.
464. Chakrabarti L, Guyader M, Alizon M, Daniel MD, Desrosiers RC, Tiollais P, et al. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature*. 1987;328:543.
465. Franchini G, Gurgo C, Guo HG, Gallo RC, Collalti E, Fagnoli KA, et al. Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses. *Nature*. 1987;328:539.
466. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. *Aids*. 2014;28(10):1509-19.
467. del Romero J, Marincovich B, Castilla J, Garcia S, Campo J, Hernando V, et al. Evaluating the risk of HIV transmission through unprotected orogenital sex. *Aids*. 2002;16(9):1296-7.
468. Baggaley RF, Boily MC, White RG, Alary M. Risk of HIV-1 transmission for parenteral exposure and blood transfusion: a systematic review and meta-analysis. *Aids*. 2006;20(6):805-12.
469. Sperling RS, Shapiro DE, Coombs RW, Todd JA, Herman SA, McSherry GD, et al. Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. *N Engl J Med*. 1996;335(22):1621-9.
470. Baggaley RF, White RG, Boily MC. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol*. 2010;39(4):1048-63.

471. Kaplan EH, Heimer R. A model-based estimate of hiv infectivity via needle sharing. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1992;5(11):1116-8.
472. Boily MC, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, et al. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis.* 2009;9(2):118-29.
473. Jin FY, Jansson J, Law M, Prestage GP, Zablotska I, Imrie JCG, et al. Per-contact probability of HIV transmission in homosexual men in Sydney in the era of HAART. *Aids.* 2010;24(6):907-13.
474. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *Am J Epidemiol.* 1999;150(3):306-11.
475. Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, et al. Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. *PLoS pathogens.* 2011;7(9):e1002209-e.
476. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users. *Journal of Virology.* 2010;84(12):6241-7.
477. Frost SDW, Liu Y, Pond SLK, Chappey C, Wrin T, Petropoulos CJ, et al. Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. *J Virol.* 2005;79(10):6523-7.
478. Kariuki SM, Selhorst P, Arien KK, Dorfman JR. The HIV-1 transmission bottleneck. *Retrovirology.* 2017;14:19.

479. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *Journal of virology*. 2009;83(8):3556-67.
480. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLOS Pathogens*. 2010;6(5):e1000890.
481. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLOS Pathogens*. 2016;12(5):e1005619.
482. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users. *Journal of Virology*. 2010;84(12):6241-7.
483. Masharsky AE, Dukhovlina EN, Verevokhin SV, Toussova OV, Skochilov RV, Anderson JA, et al. A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia. *The Journal of infectious diseases*. 2010;201(11):1697-702.
484. Russell ES, Kwiek JJ, Keys J, Barton K, Mwapasa V, Montefiori DC, et al. The genetic bottleneck in vertical transmission of subtype C HIV-1 is not driven by selection of especially neutralization-resistant virus from the maternal viral population. *Journal of virology*. 2011;85(16):8253-62.
485. Boeras DI, Hraber PT, Hurlston M, Evans-Strickfaden T, Bhattacharya T, Giorgi EE, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *Proc Natl Acad Sci U S A*. 2011;108(46):E1156-E63.

486. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee FH, et al. Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5. *Journal of Virology*. 2013;87(5):2401-11.
487. Parrish NF, Wilen CB, Banks LB, Iyer SS, Pfaff JM, Salazar-Gonzalez JF, et al. Transmitted/Founder and Chronic Subtype C HIV-1 Use CD4 and CCR5 Receptors with Equal Efficiency and Are Not Inhibited by Blocking the Integrin alpha 4 beta 7. *Plos Pathogens*. 2012;8(5).
488. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *Journal of Virology*. 2008;82(8):3952-70.
489. Jiang CL, Parrish NF, Wilen CB, Li H, Chen Y, Pavlicek JW, et al. Primary Infection by a Human Immunodeficiency Virus with Atypical Coreceptor Tropism. *Journal of Virology*. 2011;85(20):10669-81.
490. Chalmet K, Dauwe K, Foquet L, Baatz F, Seguin-Devaux C, Van der Gucht B, et al. Presence of CXCR4-Using HIV-1 in Patients With Recently Diagnosed Infection: Correlates and Evidence for Transmission. *J Infect Dis*. 2012;205(2):174-84.
491. Huang W, Toma J, Stawiski E, Fransen S, Wrin T, Parkin N, et al. Characterization of Human Immunodeficiency Virus Type 1 Populations Containing CXCR4-Using Variants from Recently Infected Individuals. *Aids Research and Human Retroviruses*. 2009;25(8):795-802.
492. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *Journal of Experimental Medicine*. 2009;206(6):1273-89.

493. Ochsenbauer C, Edmonds TG, Ding HT, Keele BF, Decker J, Salazar MG, et al. Generation of Transmitted/Founder HIV-1 Infectious Molecular Clones and Characterization of Their Replication Capacity in CD4 T Lymphocytes and Monocyte-Derived Macrophages. *Journal of Virology*. 2012;86(5):2715-28.
494. Wilen CB, Parrish NF, Pfaff JM, Decker JM, Henning EA, Haim H, et al. Phenotypic and Immunologic Comparison of Clade B Transmitted/Founder and Chronic HIV-1 Envelope Glycoproteins. *Journal of Virology*. 2011;85(17):8514-27.
495. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, et al. Phenotypic properties of transmitted founder HIV-1. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(17):6626-33.
496. Arthos J, Cicala C, Martinelli E, Macleod K, Van Ryk D, Wei D, et al. HIV-1 envelope protein binds to and signals through integrin alpha(4)beta(7), the gut mucosal homing receptor for peripheral T cells. *Nature Immunology*. 2008;9(3):301-9.
497. McKinnon LR, Nyanga B, Chege D, Izulla P, Kimani M, Huibner S, et al. Characterization of a Human Cervical CD4(+) T Cell Subset Coexpressing Multiple Markers of HIV Susceptibility. *Journal of Immunology*. 2011;187(11):6032-42.
498. Choi JY, Pond SLK, Anderson CM, Richman DD, Smith DM. Molecular Features of the V1-V4 Coding Region of Sexually Transmitted Human Immunodeficiency Virus Type 1. *J Infect Dis*. 2017;215(10):1506-13.
499. Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, et al. Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *PLoS Pathog*. 2011;7(9):19.
500. Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, et al. Selection of HIV Variants with Signature Genotypic Characteristics during Heterosexual Transmission. *J Infect Dis*. 2009;199(4):580-9.

501. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, et al. Selection for human immunodeficiency virus type I envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol*. 2005;79(10):6528-31.
502. Nakamura KJ, Heath L, Sobrera ER, Wilkinson TA, Semrau K, Kankasa C, et al. Breast milk and in utero transmission of HIV-1 select for envelope variants with unique molecular signatures. *Retrovirology*. 2017;14:19.
503. Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, et al. Recurrent Signature Patterns in HIV-1 B Clade Envelope Glycoproteins Associated with either Early or Chronic Infections. *Plos Pathogens*. 2011;7(9).
504. Glidden DV, Anderson PL, Grant RM. Pharmacology supports on-demand PrEP. *Lancet HIV*. 2016;3(9):e405-e6.
505. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, et al. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. *N Engl J Med*. 2015;373(23):2237-46.
506. McCarthy M. PrEP on demand cuts HIV transmission in high risk men, study finds. *Bmj*. 2015;351:h6468.
507. Hsu DC, O'Connell RJ. Progress in HIV vaccine development. *Human Vaccines & Immunotherapeutics*. 2017;13(5):1018-30.
508. Shin SY. Recent update in HIV vaccine development. *Clinical and Experimental Vaccine Research*. 2016;5(1):6-11.
509. Stephenson KE, D'Couto HT, Barouch DH. New concepts in HIV-1 vaccine development. *Current Opinion in Immunology*. 2016;41:39-46.

510. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *New England Journal of Medicine*. 2009;361(23):2209-20.
511. Robb ML, Rerks-Ngarm S, Nitayaphon S, Pitisuttithum P, Kaewkungwal J, Kunasol P, et al. Risk behaviour and time as covariates for efficacy of the HIV vaccine regimen ALVAC-HIV (vCP1521) and AIDSVAX B/E: a post-hoc analysis of the Thai phase 3 efficacy trial RV 144. *Lancet Infectious Diseases*. 2012;12(7):531-7.
512. Lewis GK, DeVico AL, Gallo RC. Antibody persistence and T-cell balance: Two key factors confronting HIV vaccine development. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(44):15614-21.
513. Bonsignori M, Pollara J, Moody MA, Alpert MD, Chen X, Hwang KK, et al. Antibody-Dependent Cellular Cytotoxicity-Mediating Antibodies from an HIV-1 Vaccine Efficacy Trial Target Multiple Epitopes and Preferentially Use the VH1 Gene Family. *Journal of Virology*. 2012;86(21):11521-32.
514. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-Correlates Analysis of an HIV-1 Vaccine Efficacy Trial. *New England Journal of Medicine*. 2012;366(14):1275-86.
515. O'Connell RJ, Kim JH, Excler JL. The HIV-1 gp120 V1V2 loop: structure, function and importance for vaccine development. *Expert Review of Vaccines*. 2014;13(12):1489-500.
516. Montefiori DC, Karnasuta C, Huang Y, Ahmed H, Gilbert P, de Souza MS, et al. Magnitude and Breadth of the Neutralizing Antibody Response in the RV144 and Vax003 HIV-1 Vaccine Efficacy Trials. *Journal of Infectious Diseases*. 2012;206(3):431-41.

517. Gottardo R, Bailer RT, Korber BT, Gnanakaran S, Phillips J, Shen XY, et al. Plasma IgG to Linear Epitopes in the V2 and V3 Regions of HIV-1 gp120 Correlate with a Reduced Risk of Infection in the RV144 Vaccine Efficacy Trial. *Plos One*. 2013;8(9):16.
518. Tomaras GD, Ferrari G, Shen XY, Alam SM, Liao HX, Pollara J, et al. Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(22):9019-24.
519. Fourati S, Ribeiro SP, Blasco Tavares Pereira Lopes F, Talla A, Lefebvre F, Cameron M, et al. Integrated systems approach defines the antiviral pathways conferring protection by the RV144 HIV vaccine. *Nat Commun*. 2019;10(1):863.
520. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee F-H, et al. Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5. *Journal of Virology*. 2013;87(5):2401-11.
521. Baalwa J, Wang S, Parrish NF, Decker JM, Keele BF, Learn GH, et al. Molecular identification, cloning and characterization of transmitted/founder HIV-1 subtype A, D and A/D infectious molecular clones. *Virology*. 2013;436(1):33-48.
522. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, et al. Phenotypic properties of transmitted founder HIV-1. *Proc Natl Acad Sci U S A*. 2013;110(17):6626-33.
523. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, et al. Relative resistance of HIV-1 founder viruses to control by interferon-alpha. *Retrovirology*. 2013;10:146.
524. Hernan KG, Eric S-B, Agnes-Laurance C, Michael E, Nilu G, Rasmi T, et al. Cryptic Multiple HIV-1 Infection Revealed by Early, Frequent, and Deep Sampling during Acute Infection. *AIDS Research and Human Retroviruses*. 2014;30(S1):A58-A.

525. Chowdhury A, Hayes TL, Bosinger SE, Lawson BO, Vanderford T, Schmitz JE, et al. Differential Impact of In Vivo CD8+ T Lymphocyte Depletion in Controller versus Progressor Simian Immunodeficiency Virus-Infected Macaques. *J Virol.* 2015;89(17):8677-86.
526. Pontesilli O, Klein MR, Kerkhof-Garde SR, Pakker NG, de Wolf F, Schuitemaker H, et al. Longitudinal analysis of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte responses: a predominant gag-specific response is associated with nonprogressive infection. *J Infect Dis.* 1998;178(4):1008-18.
527. Saez-Cirion A, Lacabaratz C, Lambotte O, Versmisse P, Urrutia A, Boufassa F, et al. HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype. *Proc Natl Acad Sci U S A.* 2007;104(16):6776-81.
528. Gray GE, Laher F, Lazarus E, Ensoli B, Corey L. Approaches to preventative and therapeutic HIV vaccines. *Current Opinion in Virology.* 2016;17:104-9.
529. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, et al. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N Engl J Med.* 2012;366(14):1275-86.
530. Horwitz JA, Bar-On Y, Lu C-L, Fera D, Lockhart AAK, Lorenzi JCC, et al. Non-neutralizing Antibodies Alter the Course of HIV-1 Infection In Vivo. *Cell.* 2017;170(4):637-48.e10.
531. Florese RH, Robert-Guroff M. Non-neutralizing antibodies and vaccine-induced protection against HIV. *Future HIV Therapy.* 2007;1(3):259-71.
532. Zolla-Pazner S. Non-neutralizing Antibody Functions for Protection and Control of HIV in Humans and of SIV and SHIV in Non-human Primates. *AIDS (London, England).* 2016;30(16):2551-3.

533. Holl V, Peressin M, Decoville T, Schmidt S, Zolla-Pazner S, Aubertin A-M, et al. Nonneutralizing Antibodies Are Able To Inhibit Human Immunodeficiency Virus Type 1 Replication in Macrophages and Immature Dendritic Cells. *Journal of Virology*. 2006;80(12):6177-81.
534. Balazs AB, Chen J, Hong CM, Rao DS, Yang L, Baltimore D. Antibody-based protection against HIV infection by vectored immunoprophylaxis. *Nature*. 2011;481(7379):81-4.
535. Scheid JF, Horwitz JA, Bar-On Y, Kreider EF, Lu CL, Lorenzi JC, et al. HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. *Nature*. 2016;535(7613):556-60.
536. Caskey M, Schoofs T, Gruell H, Settler A, Karagounis T, Kreider EF, et al. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nat Med*. 2017;23(2):185-91.
537. Malbec M, Porrot F, Rua R, Horwitz J, Klein F, Halper-Stromberg A, et al. Broadly neutralizing antibodies that inhibit HIV-1 cell to cell transmission. *J Exp Med*. 2013;210(13):2813-21.
538. Bruel T, Guivel-Benhassine F, Amraoui S, Malbec M, Richard L, Bourdic K, et al. Elimination of HIV-1-infected cells by broadly neutralizing antibodies. *Nat Commun*. 2016;7:10844.
539. Leroux-Roels G, Maes C, Clement F, van Engelenburg F, van den Dobbelen M, Adler M, et al. Randomized Phase I: Safety, Immunogenicity and Mucosal Antiviral Activity in Young Healthy Women Vaccinated with HIV-1 Gp41 P1 Peptide on Virosomes. *PLoS One*. 2013;8(2):e55438.
540. Nyombayire J, Anzala O, Gazzard B, Karita E, Bergin P, Hayes P, et al. First-in-Human Evaluation of the Safety and Immunogenicity of an Intranasally Administered Replication-Competent Sendai Virus–Vectored HIV Type 1 Gag Vaccine: Induction of

Potent T-Cell or Antibody Responses in Prime-Boost Regimens. *The Journal of Infectious Diseases*. 2017;215(1):95-104.

541. Barouch DH, Tomaka FL, Wegmann F, et al. Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). *Lancet*. 2018;392(10143):232–243.

542. U.S. Food and Drug Administration NHA, Silver Spring, MD 20993. Antiretroviral drugs used in the treatment of HIV infection US: U.S. Food and Drug Administration; 2018 [updated April 12, 2018. Available from:

<https://www.fda.gov/ForPatients/Illness/HIVAIDS/Treatment/ucm118915.htm>.

543. Arts EJ, Hazuda DJ. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspectives in Medicine*. 2012;2(4):23.

544. Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nature medicine*. 2003;9(6):727-8.

545. Vanhamel J, Bruggemans A, Debyser Z. Establishment of latent HIV-1 reservoirs: what do we really know? *J Virus Erad*. 2019;5(1):3-9.

546. Davey RT, Jr., Bhat N, Yoder C, Chun TW, Metcalf JA, Dewar R, et al. HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc Natl Acad Sci U S A*. 1999;96(26):15109-14.

547. Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, Pomerantz RJ. The challenge of finding a cure for HIV infection. *Science (New York, NY)*. 2009;323(5919):1304-7.

548. Alexaki A, Liu YJ, Wigdahl B. Cellular Reservoirs of HIV-1 and their Role in Viral Persistence. *Current Hiv Research*. 2008;6(5):388-400.
549. Medzhitov R, Janeway CJ. Advances in immunology: Innate immunity. *New England Journal of Medicine*. 2000;343(5):338-44.
550. Mogensen TH. Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses. *Clinical Microbiology Reviews*. 2009;22(2): 240-73,
551. Mogensen TH, Melchjorsen J, Larsen CS, Paludan SR. Innate immune recognition and activation during HIV infection. *Retrovirology*. 2010;7:19.
552. Janeway CA. Approaching the Asymptote? Evolution and Revolution in Immunology. *Journal of Immunology*. 2013;191(9):4475-87.
553. Janeway CA. Approaching the asymptote - evolution and revolution in immunology. *Cold Spring Harbor Symposia on Quantitative Biology*. 1989;54:1-13.
554. Lund J, Sato A, Akira S, Medzhitov R, Iwasaki A. Toll-like receptor 9-mediated recognition of herpes simplex virus-2 by plasmacytoid dendritic cells. *Journal of Experimental Medicine*. 2003;198(3):513-20.
555. Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, et al. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science*. 2004;303(5663):1526-9.
556. Diebold SS, Kaisho T, Hemmi H, Akira S, Sousa CRE. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science*. 2004;303(5663):1529-31.
557. Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA and activation of NF-kappa B by Toll-like receptor 3. *Nature*. 2001;413(6857):732-8.

558. Wang T, Town T, Alexopoulou L, Anderson JF, Fikrig E, Flavell RA. Toll-like receptor 3 mediates West Nile virus entry into the brain causing lethal encephalitis. *Nature Medicine*. 2004;10(12):1366-73.
559. Geijtenbeek TBH, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GCF, Middel J, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*. 2000;100(5):587-97.
560. Gringhuis SI, van der Vlist M, van den Berg LM, den Dunnen J, Litjens M, Geijtenbeek TB. HIV-1 exploits innate signaling by TLR8 and DC-SIGN for productive infection of dendritic cells. *Nat Immunol*. 2010;11(5):419-26.
561. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunology*. 2004;5(7):730-7.
562. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, et al. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*. 2006;441(7089):101-5.
563. Campbell GR, Spector SA. CCL2 increases X4-tropic HIV-1 entry into resting CD4+ T cells. *J Biol Chem*. 2008;283(45):30745-53.
564. Kinter A, Catanzaro A, Monaco J, Ruiz M, Justement J, Moir S, et al. CC-chemokines enhance the replication of T-tropic strains of HIV-1 in CD4(+) T cells: role of signal transduction. *Proc Natl Acad Sci U S A*. 1998;95(20):11880-5.
565. Arrighi JF, Pion M, Garcia E, Escola JM, van Kooyk Y, Geijtenbeek TB, et al. DC-SIGN-mediated infectious synapse formation enhances X4 HIV-1 transmission from dendritic cells to T cells. *Journal of Experimental Medicine*. 2004;200(10):1279-88.

566. Haase AT. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature*. 2010;464(7286):217-23.
567. Sips M, Sciaranghella G, Diefenbach T, Dugast AS, Berger CT, Liu Q, et al. Altered distribution of mucosal NK cells during HIV infection. *Mucosal Immunology*. 2012;5(1):30-40.
568. Martin MP, Gao XJ, Lee JH, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genetics*. 2002;31(4):429-34.
569. Martin MP, Qi Y, Gao XJ, Yamada E, Martin JN, Pereyra F, et al. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature Genetics*. 2007;39(6):733-40.
570. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, Schneidewind A, et al. Differential natural killer cell-mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *Journal of Experimental Medicine*. 2007;204(12):3027-36.
571. Altfeld M, Goulder P. 'Unleashed' natural killers hinder HIV. *Nature Genetics*. 2007;39(6):708-10.
572. Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, Carlson JM, et al. HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature*. 2011;476(7358): 96-100.
573. Borrow P. Innate immunity in acute HIV-1 infection. *Current Opinion in Hiv and Aids*. 2011;6(5):353-63.
574. Hattab S. Impact of different antiretroviral therapy (ART) regimens on the evolution of soluble markers of inflammation and immune activation in HIV-infected patients. Thèse de doctorat, Université Pierre et Marie Curie, Paris-France, 2014.
575. Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, Heitman J, et al. Induction of a striking systemic cytokine cascade prior to peak viremia in acute human

immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J Virol.* 2009;83(8):3719-33.

576. Soper A, Kimura I, Nagaoka S, Konno Y, Yamamoto K, Koyanagi Y, et al. Type I Interferon Responses by HIV-1 Infection: Association with Disease Progression and Control. *Front Immunol.* 2018;8:1823-.
577. Bosinger SE, Utay NS. Type I interferon: understanding its role in HIV pathogenesis and therapy. *Curr HIV/AIDS Rep.* 2015;12(1):41-53.
578. Schoggins JW, MacDuff DA, Imanaka N, Gainey MD, Shrestha B, Eitson JL, et al. Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature.* 2014;505(7485):691-5.
579. Utay NS, Douek DC. Interferons and HIV Infection: The Good, the Bad, and the Ugly. *Pathog Immun.* 2016;1(1):107-16.
580. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature.* 2011;472(7344):481-5.
581. Pillai SK, Abdel-Mohsen M, Guatelli J, Skasko M, Monto A, Fujimoto K, et al. Role of retroviral restriction factors in the interferon-alpha-mediated suppression of HIV-1 in vivo. *Proc Natl Acad Sci U S A.* 2012;109(8):3035-40.
582. Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature.* 2013;502(7472):559-62.
583. Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, et al. MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature.* 2013;502(7472):563-6.

584. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5 α restricts HIV-1 infection in Old World monkeys. *Nature*. 2004;427(6977):848-53.
585. Baldauf HM, Pan X, Erikson E, Schmidt S, Daddacha W, Burggraf M, et al. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med*. 2012;18(11):1682-7.
586. Utay NS, Hunt PW. Role of immune activation in progression to AIDS. *Curr Opin HIV AIDS*. 2016;11(2):131-7.
587. Iyer SS, Bibollet-Ruche F, Sherrill-Mix S, Learn GH, Plenderleith L, Smith AG, et al. Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. *Proc Natl Acad Sci U S A*. 2017;114(4):E590-e9.
588. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferoth K, Aasa-Chapman MM, et al. Relative resistance of HIV-1 founder viruses to control by interferon- α . *Retrovirology*. 2013;10:146.
589. Betts MR, Ambrozak DR, Douek DC, Bonhoeffer S, Brenchley JM, Casazza JP, et al. Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: Relationship to viral load in untreated HIV infection. *Journal of Virology*. 2001;75(24):11983-91.
590. Migueles SA, Connors M. Frequency and function of HIV-specific CD8(+) T cells. *Immunology Letters*. 2001;79(1-2):141-50.
591. Gea-Banacloche JC, Migueles SA, Martino L, Shupert WL, McNeil AC, Sabbaghian MS, et al. Maintenance of large numbers of virus-specific CD8(+) T cells in HIV-infected progressors and long-term nonprogressors. *Journal of Immunology*. 2000;165(2):1082-92.

592. Shanker P, Russo M, Harnisch B, Patterson M, Skolnik P, Lieberman J. Impaired function of circulating HIV-specific CD8(+) T cells in chronic human immunodeficiency virus infection. *Blood*. 2000;96(9):3094-101.
593. Palmer LD, Weng NP, Levine BL, June CH, Lane HC, Hodes RJ. Telomere length, telomerase activity, and replicative potential in HIV infection: Analysis of CD4(+) and CD8(+) T cells from HIV-discordant monozygotic twins. *Journal of Experimental Medicine*. 1997;185(7):1381-6.
594. Sachsenberg N, Perelson AS, Yerly S, Schockmel GA, Leduc D, Hirschel B, et al. Turnover of CD4(+) and CD8(+) T lymphocytes in HIV-1 infection as measured by Ki-67 antigen. *Journal of Experimental Medicine*. 1998;187(8):1295-303.
595. Sloand EM, Young NS, Kumar P, Weichold FF, Sato T, Maciejewski JP. Role of Fas ligand and receptor in the mechanism of T-cell depletion in acquired immunodeficiency syndrome: Effect on CD4(+) lymphocyte depletion and human immunodeficiency virus replication. *Blood*. 1997;89(4):1357-63.
596. Estaquier J, Tanaka M, Suda T, Nagata S, Golstein P, Ameisen JC. Fas-mediated apoptosis of CD4(+) and CD8(+) T cells from human immunodeficiency virus-infected persons: Differential in vitro preventive effect of cytokines and protease antagonists. *Blood*. 1996;87(12):4959-66.
597. Cocchi F, Devico AL, Garzinodemo A, Arya SK, Gallo RC, Lusso P. Identification of rantes, mip-1-alpha, and mip-1-beta as the major hiv-suppressive factors produced by CD8(+) T-cells. *Science*. 1995;270(5243):1811-5.
598. Walker CM, Moody DJ, Stites DP, Levy JA. Lymphocytes-CD8+ can control hiv-infection invitro by suppressing virus-replication. *Science*. 1986;234(4783):1563-6.

599. Mackewicz CE, Blackbourn DJ, Levy JA. CD8(+) T-Cells suppress human-immunodeficiency-virus replication by inhibiting viral transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;92(6):2308-12.
600. Phillips RE, Rowlandjones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. Human-immunodeficiency-virus genetic-variation that can escape cytotoxic T-cell recognition. *Nature*. 1991;354(6353):453-9.
601. Price DA, Goulder PJR, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, et al. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(5):1890-5.
602. Zhang D, Shankar P, Xu Z, Harnisch B, Chen G, Lange C, et al. Most antiviral CD8 T cells during chronic viral infection do not express high levels of perforin and are not directly cytotoxic. *Blood*. 2003;101(1):226-35.
603. Appay V, Nixon DF, Donahoe SM, Gillespie GMA, Dong T, King A, et al. HIV-specific CD8(+) T cells produce antiviral cytokines but are impaired in cytolytic function. *Journal of Experimental Medicine*. 2000;192(1):63-75.
604. Migueles SA, Laborico AC, Shupert WL, Sabbaghian MS, Rabin R, Hallahan CW, et al. HIV-specific CD8(+) T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nature Immunology*. 2002;3(11):1061-8.
605. Trimble LA, Shankar P, Patterson M, Daily JP, Lieberman J. Human immunodeficiency virus-specific circulating CD8 T lymphocytes have down-modulated CD3 zeta and CD28, key signaling molecules for T-cell activation. *Journal of Virology*. 2000;74(16):7320-30.
606. Jones RB, Walker BD. HIV-specific CD8+ T cells and HIV eradication. *The Journal of Clinical Investigation*. 2016;126(2):455-63.

607. Youngblood B, Noto A, Porichis F, Akondy RS, Ndhlovu ZM, Austin JW, et al. Cutting edge: Prolonged exposure to HIV reinforces a poised epigenetic program for PD-1 expression in virus-specific CD8 T cells. *Journal of immunology (Baltimore, Md : 1950)*. 2013;191(2):540-4.
608. Warren JA, Clutton G, Goonetilleke N. Harnessing CD8(+) T Cells Under HIV Antiretroviral Therapy. *Front Immunol*. 2019;10:291.
609. El-Far M, Halwani R, Said E, Trautmann L, Doroudchi M, Janbazian L, et al. T-cell exhaustion in HIV infection. *Current HIV/AIDS Reports*. 2008;5(1):13-9.
610. Day CL, Kaufmann DE, Kiepiela P, Brown JA, Moodley ES, Reddy S, et al. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature*. 2006;443(7109):350-4.
611. Jones RB, Ndhlovu LC, Barbour JD, Sheth PM, Jha AR, Long BR, et al. Tim-3 expression defines a novel population of dysfunctional T cells with highly elevated frequencies in progressive HIV-1 infection. *The Journal of experimental medicine*. 2008;205(12):2763-79.
612. Kaufmann DE, Kavanagh DG, Pereyra F, Zaunders JJ, Mackey EW, Miura T, et al. Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. *Nature immunology*. 2007;8(11):1246-54.
613. Nikolova MH, Muhtarova MN, Taskov HB, Kostov K, Vezenkov L, Mihova A, et al. The CD160+ CD8high cytotoxic T cell subset correlates with response to HAART in HIV-1+ patients. *Cellular immunology*. 2005;237(2):96-105.
614. Tian X, Zhang A, Qiu C, Wang W, Yang Y, Qiu C, et al. The upregulation of LAG-3 on T cells defines a subpopulation with functional exhaustion and correlates with disease

- progression in HIV-infected subjects. *Journal of immunology* (Baltimore, Md : 1950). 2015;194(8):3873-82.
615. Trautmann L, Janbazian L, Chomont N, Said EA, Gimmig S, Bessette B, et al. Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nature medicine*. 2006;12(10):1198-202.
616. Petrovas C, Casazza JP, Brenchley JM, Price DA, Gostick E, Adams WC, et al. PD-1 is a regulator of virus-specific CD8+ T cell survival in HIV infection. *The Journal of experimental medicine*. 2006;203(10):2281-92.
617. Grabmeier-Pfistershammer K, Steinberger P, Rieger A, Leitner J, Kohrgruber N. Identification of PD-1 as a unique marker for failing immune reconstitution in HIV-1-infected patients on treatment. *Journal of acquired immune deficiency syndromes (1999)*. 2011;56(2):118-24.
618. Velu V, Shetty RD, Larsson M, Shankar EM. Role of PD-1 co-inhibitory pathway in HIV infection and potential therapeutic options. *Retrovirology*. 2015;12:14.
619. El-Far M, Pellerin C, Pilote L, Fortin JF, Lessard IA, Peretz Y, et al. CD160 isoforms and regulation of CD4 and CD8 T-cell responses. *Journal of translational medicine*. 2014;12:217.
620. Blackburn SD, Shin H, Haining WN, Zou T, Workman CJ, Polley A, et al. Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nature immunology*. 2009;10(1):29-37.
621. Lichtenfeld M, Kaufmann DE, Yu XG, Mui SK, Addo MM, Johnston MN, et al. Loss of HIV-1-specific CD8+ T cell proliferation after acute HIV-1 infection and restoration by vaccine-induced HIV-1-specific CD4+ T cells. *J Exp Med*. 2004;200(6):701-12.

622. Winni De Haes CPo, Guido Vanham and Joanna Rejman. "Wrapped Up" Vaccines in the Context of HIV-1 Immunotherapy. In: Metodiev K, editor. Immunodeficiency. Bulgaria2012. p. 50.
623. Wei XP, Decker JM, Wang SY, Hui HX, Kappes JC, Wu XY, et al. Antibody neutralization and escape by HIV-1. *Nature*. 2003;422(6929):307-12.
624. Watkins BA, Buge S, Aldrich K, Davis AE, Robinson J, Reitz MS, et al. Resistance of human immunodeficiency virus type 1 to neutralization by natural antisera occurs through single amino acid substitutions that cause changes in antibody binding at multiple sites. *Journal of Virology*. 1996;70(12):8431-7.
625. Mohan T, Bhatnagar S, Gupta DL, Rao DN. Current understanding of HIV-1 and T-cell adaptive immunity: Progress to date. *Microbial Pathogenesis*. 2014;73:60-9.
626. Schacker TW, Nguyen PL, Beilman GJ, Wolinsky S, Larson M, Reilly C, et al. Collagen deposition in HIV-1 infected lymphatic tissues and T cell homeostasis. *Journal of Clinical Investigation*. 2002;110(8):1133-9.
627. Sanchez-Merino V, Fabra-Garcia A, Gonzalez N, Nicolas D, Merino-Mansilla A, Manzardo C, et al. Detection of Broadly Neutralizing Activity within the First Months of HIV-1 Infection. *Journal of Virology*. 2016;90(11):5231.
628. Burton DR, Mascola JR. Antibody responses to envelope glycoproteins in HIV-1 infection. *Nature Immunology*. 2015;16(6):571-6.
629. Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. Antibodies in HIV-1 Vaccine Development and Therapy. *Science*. 2013;341(6151):1199-204.
630. West AP, Scharf L, Scheid JF, Klein F, Bjorkman PJ, Nussenzweig MC. Structural Insights on the Role of Antibodies in HIV-1 Vaccine and Therapy. *Cell*. 2014;156(4):633-48.

631. Landais E, Huang XY, Havenar-Daughton C, Murrell B, Price MA, Wickramasinghe L, et al. Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary Infection Cohort. *Plos Pathogens*. 2016;12(1).
632. Klasse PJ, Sattentau QJ. Occupancy and mechanism in anti body-mediated neutralization of animal viruses. *Journal of General Virology*. 2002;83:2091-108.
633. Eroshkin AM, LeBlanc A, Weekes D, Post K, Li ZW, Rajput A, et al. bNAber: database of broadly neutralizing HIV antibodies. *Nucleic Acids Research*. 2014;42(D1):D1133-D9.
634. Malbec M, Porrot F, Rua R, Horwitz J, Klein F, Halper-Stromberg A, et al. Broadly neutralizing antibodies that inhibit HIV-1 cell to cell transmission. *J Exp Med*. 2013;210(13):2813-21.
635. Bruel T, Guivel-Benhassine F, Amraoui S, Malbec M, Richard L, Bourdic K, et al. Elimination of HIV-1-infected cells by broadly neutralizing antibodies. *Nature Communications*. 2016;7.
636. Montefiori DC. Measuring HIV neutralization in a luciferase reporter gene assay. *Methods in molecular biology (Clifton, NJ)*. 2009;485:395-405.
637. Reh L, Magnus C, Schanz M, Weber J, Uhr T, Rusert P, et al. Capacity of Broadly Neutralizing Antibodies to Inhibit HIV-1 Cell-Cell Transmission Is Strain- and Epitope-Dependent. *PLoS Pathog*. 2015;11(7):34.
638. Karina Yusim BTMK, Christian Brander, Dan Barouch, Rob de Boer, Barton F. Haynes, Richard Koup, John P. Moore, Bruce D. Walker, and David I. Watkins. Best Neutralizing Antibodies. In: Karina Yusim BTMK, Christian Brander, Dan Barouch, Rob de Boer, Barton F. Haynes, Richard Koup, John P. Moore, Bruce D. Walker, and David I. Watkins, editor. *HIV Molecular Immunology 2016*. New Mexico (US): Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR 17-24847.; 2017.

639. Liao HX, Lynch R, Zhou TQ, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*. 2013;496(7446): 469-76
640. Haynes BF, Kelsoe G, Harrison SC, Kepler TB. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol*. 2012;30(5):423-33.
641. Bonsignori M, Hwang KK, Chen X, Tsao CY, Morris L, Gray E, et al. Analysis of a Clonal Lineage of HIV-1 Envelope V2/V3 Conformational Epitope-Specific Broadly Neutralizing Antibodies and Their Inferred Unmutated Common Ancestors. *J Virol*. 2011;85(19):9998-10009.
642. Walker LM. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target (vol 326, pg 289, 2009). *Science*. 2010;327(5968):959-959.
643. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011;477(7365):466-U117.
644. Fuchs SP, Desrosiers RC. Promise and problems associated with the use of recombinant AAV for the delivery of anti-HIV antibodies. *Molecular Therapy Methods & Clinical Development*. 2016;3:16068.
645. Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans. *Proc Natl Acad Sci U S A*. 2005;102(42):14943-8.
646. Krachmarov C, Pinter A, Honnen WJ, Gorny MK, Nyambi PN, Zolla-Pazner S, et al. Antibodies that are cross-reactive for human immunodeficiency virus type 1 clade A and clade B V3 domains are common in patient sera from cameroon, but their neutralization activity is usually restricted by epitope masking. *J Virol*. 2005;79(2):780-90.

647. Krachmarov CP, Honnen WJ, Kayman SC, Gorny MK, Zolla-Pazner S, Pinter A. Factors determining the breadth and potency of neutralization by V3-specific human monoclonal antibodies derived from subjects infected with clade A or clade B strains of human immunodeficiency virus type 1. *J Virol.* 2006;80(14):7127-35.
648. Kwong PD. Human immunodeficiency virus - Refolding the envelope. *Nature.* 2005;433(7028):815-6.
649. Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature.* 2002;420(6916):678-82.
650. Pantophlet R, Burton DR. GP120: Target for neutralizing HIV-1 antibodies. *Annual Review of Immunology.* Annual Review of Immunology. 24. Palo Alto: Annual Reviews; 2006. p. 739-69.
651. Pinter A, Honnen WJ, He YX, Gorny MK, Zolla-Pazner S, Kayman SC. The V1/V2 domain of gp120 is a global regulator of the sensitivity of primary human immunodeficiency virus type 1 isolates to neutralization by antibodies commonly induced upon infection. *J Virol.* 2004;78(10):5205-15.
652. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature.* 1998;393(6686):705-11.
653. Reynard FB, Verrier HIV1 Env N-glycans and their interactions with the immune system. *Virologie.*10(2):129-41.
654. Mascola JR, Montefiori DC. HIV-1: nature's master of disguise. *Nature Medicine.* 2003;9(4):393-4.

655. Manoury B, Hewitt EW, Morrice N, Dando PM, Barrett AJ, Watts C. An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature*. 1998;396(6712):695-9.
656. Yokoyama M, Naganawa S, Yoshimura K, Matsushita S, Sato H. Structural Dynamics of HIV-1 Envelope Gp120 Outer Domain with V3 Loop. *PLoS One*. 2012;7(5):9.
657. Quiñones-Kochs MI, Buonocore L, Rose JK. Role of N-Linked Glycans in a Human Immunodeficiency Virus Envelope Glycoprotein: Effects on Protein Function and the Neutralizing Antibody Response. *Journal of Virology*. 2002;76(9):4199-211.
658. Montefiori DC, Robinson WE, Jr., Mitchell WM. Role of protein N-glycosylation in pathogenesis of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences of the United States of America*. 1988;85(23):9248-52.
659. Huang X, Jin W, Hu K, Luo S, Du T, Griffin GE, et al. Highly conserved HIV-1 gp120 glycans proximal to CD4-binding region affect viral infectivity and neutralizing antibody induction. *Virology*. 2012;423(1):97-106.
660. Wang W, Nie J, Prochnow C, Truong C, Jia Z, Wang S, et al. A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization. *Retrovirology*. 2013;10:14.
661. Yen PJ, Herschhorn A, Haim H, Salas I, Gu C, Sodroski J, et al. Loss of a conserved N-linked glycosylation site in the simian immunodeficiency virus envelope glycoprotein V2 region enhances macrophage tropism by increasing CD4-independent cell-to-cell transmission. *J Virol*. 2014;88(9):5014-28.
662. Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, et al. Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol*. 1996;70(2):1100-8.

663. West AP, Jr., Scharf L, Horwitz J, Klein F, Nussenzweig MC, Bjorkman PJ. Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc Natl Acad Sci U S A*. 2013;110(26):10598-603.
664. Scharf L, Scheid JF, Lee JH, West AP, Jr., Chen C, Gao H, et al. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep*. 2014;7(3):785-95.
665. Chuang GY, Liou D, Kwong PD, Georgiev IS. NEP: web server for epitope prediction based on antibody neutralization of viral strains with diverse sequences. *Nucleic Acids Res*. 2014;42(Web Server issue):W64-71.
666. Chuang GY, Acharya P, Schmidt SD, Yang Y, Louder MK, Zhou T, et al. Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. *J Virol*. 2013;87(18):10047-58.
667. Kafando A, Serhir B, Doualla-Bell F, Fournier E, Sangare MN, Martineau C, et al. A short-term assessment of nascent HIV-1 transmission clusters among newly diagnosed individuals using envelope sequence-based phylogenetic analyses. *AIDS Res Hum Retroviruses*. 2019.
668. Kafando A, Martineau C, El-Far M, Fournier E, Doualla-Bell F, Serhir B, et al. HIV-1 Envelope Glycoprotein Amino Acids Signatures Associated with Clade B Transmitted/Founder and Recent Viruses. *Viruses*. 2019;11(11):1012.
669. Cousins MM, Swan D, Magaret CA, Hoover DR, Eshleman SH. Analysis of HIV Using a High Resolution Melting (HRM) Diversity Assay: Automation of HRM Data Analysis Enhances the Utility of the Assay for Analysis of HIV Incidence. *Plos One*. 2012;7(12):10.
670. Wu JW, Patterson-Lomba O, Novitsky V, Pagano M. A Generalized Entropy Measure of Within-Host Viral Diversity for Identifying Recent HIV-1 Infections. *Medicine*. 2015;94(42).

671. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(21):7552-7.
672. Brenner B, Wainberg MA, Roger M. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. *Aids*. 2013;27(7):1045-57.
673. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, Weinberg J, et al. Transmission Clustering Drives the Onward Spread of the HIV Epidemic Among Men Who Have Sex With Men in Quebec. *Journal of Infectious Diseases*. 2011;204(7):1115-9.
674. Brenner BG, Ibanescu RI, Hardy I, Stephens D, Otis J, Moodie E, et al. Large cluster outbreaks sustain the HIV epidemic among MSM in Quebec. *Aids*. 2017;31(5):707-17.
675. Yang PL, Ai LS, Huang SC, Li HF, Chan WE, Chang CW, et al. The Cytoplasmic Domain of Human Immunodeficiency Virus Type 1 Transmembrane Protein gp41 Harbors Lipid Raft Association Determinants. *Journal of Virology*. 2010;84(1):59-75.
676. Edwards TG, Wyss S, Reeves JD, Zolla-Pazner S, Hoxie JA, Doms RW, et al. Truncation of the cytoplasmic domain induces exposure of conserved regions in the ectodomain of human immunodeficiency virus type 1 envelope protein. *J Virol*. 2002;76(6):2683-91.
677. Bultmann A, Muranyi W, Seed B, Haas J. Identification of two sequences in the cytoplasmic tail of the human immunodeficiency virus type 1 envelope glycoprotein that inhibit cell surface expression. *J Virol*. 2001;75(11):5263-76.
678. Murakami T, Freed EO. The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(1):343-8.

679. Piller SC, Dubay JW, Derdeyn CA, Hunter E. Mutational analysis of conserved domains within the cytoplasmic tail of gp41 from human immunodeficiency virus type 1: Effects on glycoprotein incorporation and infectivity. *J Virol.* 2000;74(24):11717-23.
680. Jiang JY, Aiken C. Maturation-dependent human immunodeficiency virus type 1 particle fusion requires a carboxyl-terminal region of the gp41 cytoplasmic tail. *J Virol.* 2007;81(18):9999-10008.
681. Yang P, Ai LS, Huang SC, Li HF, Chan WE, Chang CW, et al. The cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein gp41 harbors lipid raft association determinants. *J Virol.* 2010;84(1):59-75.
682. Fernandez MV, Freed EO. Meeting Review: 2018 International Workshop on Structure and Function of the Lentiviral gp41 Cytoplasmic Tail. *Viruses.* 2018;10(11):613.
683. Newman JT, Sturgeon TJ, Gupta P, Montelaro RC. Differential functional phenotypes of two primary HIV-1 strains resulting from homologous point mutations in the LLP domains of the envelope gp41 intracytoplasmic domain. *Virology.* 2007;367(1):102-16.
684. Yolitz J, Schwing C, Chang J, Van Ryk D, Nawaz F, Wei D, et al. Signal peptide of HIV envelope protein impacts glycosylation and antigenicity of gp120. *Proc Natl Acad Sci U S A.* 2018;115(10):2443-8.