Identification d'une plante médicinale africaine par le DNA barcoding et étude de composés à activité anti-HIV de cette plante

Yue Zheng

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

THÈSE

présentée par

Yue ZHENG

pour obtenir le grade de **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Virologie

Identification d'une plante médicinale africaine par le DNA barcoding et étude de composés à activité anti-HIV de cette plante

soutenue le 3 décembre 2015

Membres du jury :

Dr. André STEINMETZ	Directeur de thèse
Dr. Carole DEVAUX	Co-Directeur de thèse
Prof. Charles BOUCHER	Rapporteur externe
Dr. Guy COLLING	Rapporteur externe
Dr. Jean-Christophe PAILLART	Examinateur
Prof. Dominique SCHOLS	Membre invité
Dr. Jean-Claude SCHMIT	Membre invité

Luxembourg Institute of Health, Luxembourg

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. André STEINMETZ and Dr. Carole DEVAUX for their continuous support. Without their kind, professional, patient, motivated guidance and help I would never have been able to finish this study. André and Carole are the kindest people I ever met. I appreciate a lot their immense knowledge and humble attitude in scientific research as well as in daily life. During the past four years, they continuously guided me through my research and helped me in other fields as well to let me grow up in science. They presented themselves as great examples I would like to follow.

I would also like to say my great thanks to Dr. Jean-Claude SCHMIT who initiated this project, and to Prof. Dominique SCHOLS who has collaborated with us and gave a lot of suggestions and help. I also thank Prof. Jean-Christophe PAILLART, Prof. Charles BOUCHER, Dr. Guy COLLING, Prof. Dominique SCHOLS and Dr. Jean-Claude SCHMIT for accepting to be member of my Thesis Committee and for spending their precious time to read and evaluate my thesis and give their insightful comments.

I thank Dr. Serge SCHNEIDER from Laboratoire National de Santé (Luxembourg) for advice on the preparation of the crude extract and Xianwen YANG (CRP-Santé, Luxembourg) for providing the purified compounds.

My sincere thanks must go to all my colleagues of the Laboratory of Retrovirology and of the Laboratory of Cellular and Molecular Oncology at Luxembourg Institute of Health. Especially, thanks to Martin MULINGE who kindly brought the raw plant material from Kenya, to Andy CHEVIGNE who has given precious suggestions to the

project, to Gilles ISERENTANT, Manuel COUNSON, Morgane LEMAIRE and Katrin NEUMANN who have provided admirable helps on experiments, and to Xavier DERVILLEZ who kindly allowed me to also participate in his project where I learned a lot. All their precious support and the fun we had together have given me great encouragements.

Last but not least, I would like to thank my parents for their love and spiritual support throughout all my life, and my friend Lu ZHANG for her great help and company during my time abroad.

Table of Contents

Summary

This project focuses on the identification and the characterization of the anti-HIV activity of an African plant. Several pieces of wood of an unknown plant were originally brought to Brussels by an HIV positive African man from Congo at the early 2000s, who claimed to his doctor that he was treating himself with a plant decoction from his local place and felt better. Although this patient passed away from AIDS, an extraction of these pieces of wood was then tested in the Laboratory of Retrovirology at CRP-Santé of Luxembourg. The anti-HIV activity was confirmed but was accompanied with cytotoxicity. To further investigate its anti-HIV mode-of-action, more materials were needed. At that time, the African plant was unknown and further research was stopped due to the lack of material.

My PhD project is a following-up to the previous study aiming in a first part to identify the African plant and, in a second part, to assess the anti-HIV activity of the African plant, to decipher its mode-of-action, and to purify its active components.

To identify the African plant, DNA barcoding approach was used. The sequence of the nuclear markerITS2 of the African wood was first compared to sequences from GeneBank using blast. The sequence of the ITS2 barcode of the African wood was most similar to that of several Cassia and Senna species. The highest similarity was obtained with *Cassia grandis* with about 86% identity. Therefore, seeds of fifty two Cassia and Senna species were purchased and geminated. Sequences of five barcoding markers - ITS2, matK, trnH-psbA spacer, rbcL and trnL, were compared. It reveals that the African plant was indeed *Cassia abbreviata*. This conclusion was further confirmed by the sequences obtained with the leaf and bark of *Cassia abbreviata* brought from Kenya. *Cassia abbreviata* is well known for its medicinal use in Africa against a number of diseases including malaria and AIDS, but its mechanisms of protection are still unclear.

Besides the identification of the African plant, the phylogenetic relationship of Cassia and Senna species were studied. All the sequences of the chloroplast marker psbA-trnH and the nuclear marker ITS2 of fifty two Cassia and Senna species were further analyzed using the MUSCLE (multiple sequence comparison by log-expectation) alignment method and phylogenetic trees were then generated using neighbor joining clustering method. In the two phylogenetic trees, the species divide into five major clades one of which comprising exclusively species with accepted Cassia names. A typical "insertion" (35-67bp) in the psbA-trnH spacer and a "deletion" (9-13bp) in the ITS2 locus were found in the Cassia clade. Sequence identities between Cassias and Sennas range from 41 to 100% in the psbA-trnH locus and from 67 to 100% in the ITS2 locus. This high sequence divergence between the various species not only allows to unambiguously discriminate between Cassias and Sennas; it also defines the two loci as excellent complementary barcode markers for plant identification.

To study the anti-HIV mode-of-action of *Cassia abbreviata*, a crude extract (CE) obtained from bark after ethanol and ethyl acetate extraction was first tested in several anti-HIV assays. CE showed its protection on different cell lines expressing either the co-receptor CCR5 or CXCR4 such as on MT-4 cells against HIV-1 IIIB (X4 tropic strain) (IC_{50}=21.75±1.20 µg/ml; CC_{50}=272.45±110.38 µg/ml; SI=12.69±5.78), and on U373-CD4-CCR5/CXCR4 cells infected by pseudotype particles pNL4.3Δ*Env*Luc-Bal (R5 tropic strain) (IC_{50}=67.40±4.21 µg/ml) and pNL4.3Δ*Env*Luc-HXB2 (X4 tropic strain) (IC_{50}=13.37±8.46 µg/ml). CE also inhibited viral infection of primary human peripheral blood mononuclear cells (PBMCs) by HIV-1 ADA-M (R5 tropic strain, IC_{50}=13.53±12.99 µg/ml) and HIV-1 IIIB (X4 tropic strain, IC50=40.77±4.04 µg/ml) as well as by two clinical isolates harboring multi-drug resistance to nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs)

(IC_{50}=23.06±5.92 µg/ml and10.47±2.09 µg/ml, respectively) without displaying any cytotoxicity (CC_{50}>1000 µg/ml). To decipher at which step of HIV replication cycle CE functions, we performed a multi-dosing time assay using infection of U373-CD4-CCR5 and U373-CD4-CXCR4 cells by pseudotype particles. The inhibitory effect of CE was apparent when CE was added before infection, pre-incubated with the virus but not with the cells, and when incubated during infection but not after infection. These results indicate that CE inhibited HIV infection at an early stage of the HIV entry process independently of co-receptor usage.

To further elucidate the active components of *Cassia abbreviata*, fifty-seven compounds were then isolated using different solvents according to their polarity using Sephadex LH-20, CC cover silica gel and preparative thin layer chromatography (TLC) by Dr Xianwen Yang at the South China Sea Institute of Oceanology, Guangzhou, China. All the purified compounds were screened in an anti-HIV assay against pseudotype particles pNL4.3Δ*Env*Luc-Bal and pNL4.3Δ*Env*Luc-HXB2 when the compounds were added only at the time of the infection to identify the compounds targeting HIV entry. Four compounds (oleanolic acid, palmitic acid, taxifolin and piceatannol), showed anti-HIV activity as previously reported and three compounds (YCA 6-34, YCA 6-37 and YCA 7-9) presented an unindentified structure while demonstrating anti-HIV activity (against Bal: IC_{50}=29.61±6.14 µg/ml; IC_{50}=27.00±20.12 µg/ml; IC_{50}=12.86±3.13 µg/ml, respectively;. CC_{50}>100 µg/ml for the three compounds). Among the four known anti-HIV compounds, piceatannol showed the most anti-HIV potentials (against Bal: IC_{50}=47.46±6.52 µM; against HXB2: IC_{50}=10.28±5.74 µM). This activity was confirmed on PBMCs infected by HIV-1 ADA-M (IC_{50}=19.91±0.22 µM) and HIV-1 IIIB (IC_{50}=24.22±7.13 µM) as well as by the two multi-drug resistant clinical isolates (IC_{50}=37.72±12.54 µM; IC_{50}=8.04±3.07 µM, respectively).

Interestingly, piceatannol was previously described as an HIV-1 protease inhibitor and

an integrase inhibitor using enzyme based assays, but our results demonstrated that piceatannol acts at an earlier stage of the infection. Therefore, we investigated the mechanism of action of piceatannol using a multi-dosing time assay. Similarly to CE, piceatannol inhibited viral infection when added at the time of infection other than when added after infection. Piceatannol slightly inhibited HIV infection at high concentration when it was pre-incubated with cells before infection but not with the virus. This indicates that piceatannol may target the cells and may have a different anti-HIV mechanism than the other compounds contained in the CE of *Cassia abbreviata* targeting mainly the virus. Additionally, both CE and piceatannol showed a synergistic effect with the HIV entry inhibitors Enfuvirtide (T-20, fusion inhibitor) and AMD3100/Maraviroc (CXCR4 and CCR5 co-receptor inhibitors), supporting that their anti-HIV mechanisms differ from T20 and AMD3100/Maraviroc. To confirm these data, we performed fusion assays between Hela-P4-CXCR4 cells and Hela-Env-Lai cells. We did not found any effect of piceatannol on the fusion process whereas *CE of Cassia abbreviata*, T20 and AMD3100 showed a significant inhibition.

During the HIV entry process, the virus binds successively to the CD4 cellular receptor and to the co-receptors CXCR4/CCR5. We first tested whether CE and piceatannol could affect the binding of gp120 with the CD4 receptor using a home-made ELISA.CE inhibited the interaction between CD4 and gp120 in a dose-dependent manner, while piceatannol did not. No significant differences in CD4, CXCR4 and CCR5 binding between cells incubated or not with the drugs was further observed using different antibodies against CD4, CXCR4 and CCR5 suggesting no direct interaction of the compounds with the target cells receptors. In line with this observation, we observed that piceatannol and not the CE inhibit the infection of pseudotype particles of vesicular stomatitis virus (VSV) G proteins pNL4.3ΔEnvLuc-VSVG (IC_{50}=79.23±17.20µM) suggesting that piceatannol inhibits viruses entry by inhibiting the virus attachment to the cells in a non-specific manner whereas the main active components of CE target more specifically gp120.

To finally evaluate the potential activity of piceatanol as microbicide, we showed using an *in vitro* dual-chamber assay mimicking female genital tract that piceatannol could cross the layer of epithelial cells without affecting its confluence and inhibit HIV infection in the lower chamber. In addition, piceatannol did not affect the expression of the activation markers CD25 and CD69 on PBMCs indicating that piceatannol does not stimulate target cells to potentiate further HIV infection. Piceatannol inhibits also human Herpes Simplex Virus-1 and -2 (HSV-1 and HSV-2). Taken together, these data showed the potential use of piceatannol as a microbicide.

In conclusion, during this PhD project, we identified the African anti-HIV plant as *Cassia abbreviata* and the phylogenetic relationship among cassia and senna species. We purified fifty-seven pure compounds from *Cassia abbreviata*, and identified 4 known compounds and three compounds with an uncharacterized structure active against HIV infection. We showed that *Cassia abbreviata* contained several compounds inhibiting HIV entry by targeting the binding between gp120 and CD4. Since the four known compounds inhibit HIV infection through other mechanisms, we will verify if one of the three novel compounds might interact with the gp120/CD4 binding when their complete structure will be elucidated. We deciphered a novel mechanism of action of piceatannol, the most active known compounds, inhibiting HIV entry by preventing the virus attachment probably by adsorption on the cell surface. The structure of piceatannol could be modified in the future to further improve its anti-HIV activity by performing structure-activity relationship and structure-based optimization studies.

Résumé

Mon projet de thèse porte sur une plante africaine possédant des propriétés antivirales contre le virus de l'immunodéficience humaine (VIH). Un jeune homme congolais atteint du SIDA arriva à Bruxelles en l'an 2000 avec des morceaux de bois d'un arbre local dont l'écorce était utilisée au Congo pour faire des décoctions utilisées dans le traitement de nombreuses maladies et notamment le Syndrome d'Immunodéficience acquise (SIDA). Ce bois fut analysé au laboratoire de Retrovirologie au CRP-Santé et une activité antivirale (et cytotoxique) pouvait être mise en évidence dans l'extrait. Le jeune homme mourut du SIDA en 2003 et en raison de manque de matériel et du fait que la plante restait non identifiée les recherches sur le sujet furent suspendues.

J'ai repris ce sujet en 2010 dans le cadre de mes travaux de thèse de doctorat après des premiers tests prometteurs d'identification du bois grâce à l'approche du "DNA barcoding". Par comparaison avec des séquences disponibles dans Genbank, ces tests ont en effet suggéré que le bois, avec une identité de séquence de 86% avec *Cassia grandis* pour le marqueur étudié (ITS2), venait d'un arbre/arbuste du genre Cassia ou du genre apparenté Senna, les deux appartenant à la grande famille des Légumineuses. Les espèces de ces deux genres sont surtout répandues dans les régions tropicales du globe et sont souvent utilisées en médecine traditionnelle par les populations locales.

Nous avons pu nous procurer des graines d'une cinquantaine d'espèces de ces deux genres et nous avons fait une étude plus détaillée pour les deux marqueurs ITS2 (internal transcribed spacer 2 d'origine nucléaire) et le "trnH-psbA spacer" (d'origine chloroplastique). Cette étude a montré qu'une seule espèce, à savoir *Cassia*

abbreviata, montrait une identité >97% avec les deux marqueurs de notre plante. Cette identité a pu être confirmée par une analyse d' échantillon de bois fournis par M. Mulinge (South Eastern Kenya University). Les légères différences observées pouvaient être attribuées à la variabilité géographique.

Disposant des séquences nucléotidiques de ces deux marqueurs pour ces espèces nous avons effectué une étude phylogénétique en utilisant le logiciel d'alignement MUSCLE (**MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation), et la "Nearest Neighbour Analysis" pour l'assemblage de l'arbre phylogénétique. L'alignement des sequences nucléotidiques des deux marqueurs a permis de mettre en évidence des particularités structurales permettant de clairement différencier les Cassias et les Sennas. Il s'agit notamment, pour les Cassia, d'une insertion spécifique de 35 à 67 bp dans le marqueur chloroplastique (trnH-psbA) et une délétion de 9 à 13 pb dans le marqueur nucléaire (ITS2). Par ailleurs, des motifs nucléotidiques conservés dans des petits groupes d'espèces nous ont permis d'établir une phylogénie basée sur ces données moléculaires. Aussi la grande variabilité des séquences pour chacun des deux marqueurs a permis de clairement distinguer les différentes espèces et rendent ces deux marqueurs des cibles très utiles pour l'identification des espèces végétales de ces deux genres - et sans doute la majorité des espèces végétales - par le "DNA barcoding". Si cette approche n'est pas encore acceptée par les taxonomistes classiques, elle est pourtant très puissante et a l'avantage de ne pas être basée sur des traits morphologiques de la fleur; elle permet d'identifier une plante à partir de matériel foliaire, racinaire, cortical ou autre, pourvu qu'il ne soit pas contaminé par du matériel d'une autre espèce.

Afin d'étudier le mode d'action par lesquel la plante exerce son activité anti-VIH, un extrait brut (CE) a été obtenu par extraction à l'alcool et à l'acétate d'éthyle à partir de l'écorce de bois fourni par M. Mulinge. Cet extrait brut montrait une protection sur

différentes lignées cellulaires exprimant soit le co-récepteur CCR5 soit CXCR4 comme sur les cellules MT-4 infectées par le HIV IIIB (souche X4 tropique) (IC_{50} = 21.75±1.20 µg/ml; CC_{50} = 272.45±110.38 µg/ml; S = 12.69±5.78), ainsi que sur les cellules U373-CD4-CCR5/CXCR4 infectées par des particules pseudotypées pNL4.3Δ*Env*Luc-Bal (souche R5 tropique) (IC_{50} = 67.40±4.21 µg/ml) et pNL4.3Δ*Env*Luc-HXB2 (souche X4 tropique) (IC_{50} = 13.37±8.46 µg/ml). L'extrait brut inhibait aussi l'infection de cellules primaires mononucléaires du sang périphérique (PBMC) par la souche HIV-1 ADA-M (R5 tropique, IC_{50} = 13.53±12.99 µg/ml) et la souche HIV-1 IIIB (X4 tropique, IC_{50} = 40.77±4.04 µg/ml), ainsi que par deux isolats cliniques présentant une résistance à plusieurs médicaments et notamment aux analogues nucléosidiques ou nucléotidiques, inhibiteurs de la transcriptase inverse (NRTIs), aux inhibiteurs non-nucléosidiques de la transcriptase inverse (NNRTIs) et aux inhibiteurs de protéase (PIs) (IC_{50} = 23.06±5.92 µg/ml et IC_{50} = 10.47±2.09 µg/ml). L'extrait ne montrait pas d'effet cytotoxique (CC_{50}>1000 µg/ml). Afin d'identifier l'étape du cycle de replication virale au niveau de laquelle l'extrait brut pouvait agir nous avons effectué un test "multi-dosing time" en infectant des cellules U373-CD4-CCR5 and U373-CD4-CXCR4 par des particules pesudotypées. L'effet inhibiteur de l'extrait brut (CE) a pu être détecté quand l'extrait était ajouté avant infection, pré-incubé avec le virus mais pas avec les cellules, et quand il était ajouté pendant mais pas après infection. Ces résultats indiquent que le CE inhibe l'infection par le VIH à un stade précoce du processus d'entrée du virus, et ceci indépendamment de l'utilisation du co-récepteur.

Afin d'identifier les composés actifs de la plante, 57 composés ont pu être purifiés et identifiés par Yang Xianwen (South China Sea Institute of Oceanology, Guangzhou, Chine) par des approches chromatographiques (Sephadex LH-20, chromatographie sur colonne (CC) sur gel de silice, chromatographie en couche mince) permettant de séparer les composés selon leur polarité. Les composés purifiés ont été testés ensuite dans un test anti-VIH contre des particules virales pseudotypées

pNL4.3Δ*Env*Luc-Bal et pNL4.3Δ*Env*Luc-HXB2, les composés purifiés étant ajoutés seulement au moment de l'infection afin d'identifier ceux ciblant l'entrée du virus. Quatre composés (acide oléanolique, acide palmitique, taxifoline et piceatannol) ont montré une activité antivirale et cette activité a déjà décrite antérieurement par d'autres laboratoires. Trois autres composés (YCA 6-34, YCA 6-37, et YCA 7-9) présentaient une nouvelle structure et montraient une activité anti-VIH contre la souche de virus Bal (avec des IC_{50} de 29.61±6.14 µg/ml, de 27.00±20.12 µg/ml, respectivement de 12.86±3.13 µg/ml). Pour les trois composés, l'index de cytotoxicité (CC_{50}) était supérieur à 100 µg/ml.

Parmi les 4 composés connus, le piceatannol montrait le plus grand potentiel anti-VIH (avec un IC_{50}=47.46±6.52 µM contre la souche Bal, et un IC_{50}=10.28±5.74 µM contre la souche HXB2). Cette activité était confirmée sur des PBMCs infectées par la souche HIV-1 ADA-M (IC_{50} = 19.91±0.22 µM), et la souche HIV-1 IIIB (IC_{50} = 24.22±7.13 µM), ainsi que par les deux isolats cliniques résistant à plusieurs médicaments (IC_{50} = 37.72±12.54 µM et IC_{50} = 8.04±3.07 µM).

Dans des études antérieures basées sur des tests enzymatiques, le piceatannol était décrit comme un inhibiteur de protéase et un inhibiteur de l'intégrase du VIH. En revanche, nos résultats ont montré que le piceatannol agit à un stade plus précoce de l'infection. Pour cette raison nous avons décidé d'étudier plus en détail le mécanisme d'action du piceatannol en utilisant le test "multi-dosing time". Comme l'extrait brut, le piceatannol inhibait l'infection virale quand il était ajouté au moment de l'infection plutôt qu'après infection. Le piceatannol inhibitait l'infection virale légèrement à forte concentration quand il était préincubé avec les cellules avant l'infection mais pas quand il était préincubé avec le virus. Ceci suggère que le piceatannol cible préférentiellement les cellules et qu'il agit par un mécanisme différent de celui des autres composés présents dans l'extrait brut de la plante qui cibleraient surtout le

virus. De plus, l'extrait brut ainsi que le piceatannol montraient un effet synergique avec les inhibiteurs d'entrée virale Enfuvirtide (T-20, inhibiteur de fusion) et AMD3100/Maraviroc (inhibiteurs des co-récepteurs CXCR4 et CCR5), ce qui suggère que leurs mécanismes diffèrent de ceux de T20 et AMD3100/Maraviroc. Afin de confirmer ces résultats nous avons effectué des tests de fusion entre les cellules Hela-P4-CXCR4 et les cellules Hela-Env-Lai. Nous n'avons pas observé d'effet du piceatannol sur le processus de fusion, alors que l'extrait brut, T20, ainsi que l'AMD3100 montraient une inhibition significative.

Lors du processus d'entrée, le virus se lie successivement au récepteur cellulaire CD4 et aux co-récepteurs CXCR4/CCR5. Nous avons testé si l'extrait brut et le piceatannol pouvaient affecter l'attachement de gp120 avec le récepteur CD4 en utilisant un ELISA "fait maison". Alors que l'extrait brut inhibait l'interaction entre CD4 et gp120 d'une manière dose-dépendante, le piceatannol ne montrait pas d'effet inhibiteur. Nous n'avons pas observé de différence significative entre cellules incubées et cellules non-incubées avec les composés pour l'attachement au récepteur CD4 en utilisant le clone CD4 RPA-4 ciblant le domaine 1 de CD4 responsable de l'attachement à gp120 ou à CXCR4/CCR5. En accord avec cette observation, nous avons constaté que le piceatannol, et pas l'extrait brut, inhibait l'infection par des particules pseudotypées des protéines G pNL4.3ΔEnvLuc-VSVG (IC_{50} = 79.23±17.20 µM) du virus de la stomatite vésiculaire (VSV), ce qui suggère que le piceatannol inhibe l'entrée virale par inhibition non spécifique de l'attachement du virus aux cellules, alors que les composés actifs de l'extrait brut ciblent plus spécifiquement gp120.

Enfin, afin d'évaluer l'activité potentielle du piceatannol comme microbicide nous avons montré, en utilisant un essai *in vitro* de double-chambre simulant le tract génital femelle, que le piceatannol pouvait traverser une couche de cellules épithéliales sans

affecter sa confluence et inhiber l'infection virale dans la chambre inférieure. Aussi, le piceatannol n'affectait pas l'expression des marqueurs d'activation CD25 et CD69 sur les PBMC, indiquant que le piceatannol ne stimule pas les cellules cibles, ceci pouvant augmenter l'infection par le VIH. D'autre part, le piceatannol inhibe également l'infection par le virus de l'herpes humain (HSV-1 and HSV-2). Considérés dans leur ensemble, ces résultats montrent l'utilisation potentielle du piceatannol comme microbicide.

En résumé, lors de mon travail de thèse, nous avons identifié la plante à activité anti-VIH comme étant *Cassia abbreviata* et nous avons établi des relations phylogénétiques parmi des espèces Cassia et Senna. Nous avons purifié 57 composés de cette plante, parmi lesquels nous avons identifié 7 avec une activité antivirale, dont 4 connus et 3 dont la structure n'est pas encore caractérisée. Nous avons montré que l'extrait de *Cassia abbreviata* contenait plusieurs composés inhibant l'entrée du VIH en ciblant majoritairement l'attachement entre gp120 et CD4. Il sera intéressant dans le futur de vérifier si un des trois nouveaux composés, une fois leur structure connue, peut affecter l'interaction gp120/CD4. Nous avons identifié un nouveau mode d'action par lequel le piceatannol, le plus actif des composés connus, inhibe l'entrée virale en empêchant l'attachement du virus, probablement par adsorption à la surface cellulaire. Des études de relation entre structure et activité (SAR) ainsi que des études d'optimisation structurale devraient permettre d'améliorer d'avantage l'activité anti-VIH du piceatannol et des nouveaux composés identifiés à partir de *Cassia abbreviatta.*

Abbreviations

3TC	Lamivudine
ABC	Abacavir
AI	Attachment inhibitor
AIDS	Acquired immunodeficiency syndrome
APOBEC3G	Apolipoprotein B messenger RNA (mRNA)-editing enzyme catalytic polypeptide-like 3 G protein
APV	Amprenavir
ART	Antiretroviral therapy
ATV	Atazanavir
AZT	Azidothymidine
BST2	Bone marrow stromal cell antigen 2
BVM	Bevirimat
CA	Capsid
cART	Combination antiretroviral therapy
CBC	Cap binding complex
CC_{50}	Half-maximal cytotoxic concentration
$CCID_{50}$	Cell culture infectious dose 50%
CCL5	C-C motif ligand 5
CCR5	C-C chemokine receptor 5
CHMP	Charged multi-vesicular protein
CI	Combination index
COBI	Cobicistat
CPE	Cytopathic effect
CPRG	Chlorophenol red--β-D-galactopyranoside
CRF	Circulating recombinant form
CTD	C-terminal domain
CXCL12	C-X-C motif chemokine 12
CXCR4	C-X-C chemokine receptor 4
d4T	Stavudine
DC	Dendritic cell
DC-SIGN	Dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin
ddC	Zalcitabine
ddI	Didanosine

DLV	Delavirdine
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DRV	Darunavir
DTG	Dolutegravir
EC_{50}	Half-maximal effective concentration
EC_{95}	95% of maximal effective concentration
EFV	Efavirenz
ELISA	Enzyme-linked immunosorbent assay
Env	Envelope
ESCRT	Endosomal sorting complexes required for transport
ETR	Etravirine
EVG	Elvitegravir
FBS	Fetal bovine serum
FDA	Food and Drug Administration of the United States
FIPV	Feline corona virus
FITC	Fluorescein isothiocyanate
FPV	Fosamprenavir
FTC	Emtricitabine
gal-C	Galactocerebroside
GALT	Gut-associated lymphoid tissue
HAART	Highly active antiretroviral therapy
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HRS	Hepatocyte growth factor – regulated tyrosine kinase substrate
HSV	Herpers simplex virus
HTLV-III	Human T-lymphotropic virus type III
ICAM3	Intercellular adhesion molecule 3
IDU	Injecting drug users
IDV	Indinavir
IN	Integrase
INI	Integrase inhibitor
LAV	Lymphadenopathy associated virus
LEDGF/p75	Lens epithelium-derived growth factor/p75
LPS	Lipopolysaccharide
LPV/s	Lopinavir/ritonavir

MA	Matrix
mAb	Monoclonal antibody
MFI	Mean of fluorescence intensity
MHC-1	Major histocompatibility complex-1
MI	Maturation inhibitor
MMR	Macrophage mannose specific receptor
MOI	Multiplicity of infection
MSM	Men who have sex with men
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MUSCLE	Multiple sequence comparison by log-expectation
MVC	Maraviroc
NC	Nucleocapsid
Nef	Negative factor
NFV	Nelfinavir
NJ	Neighbour joining
NK	Natural killer cell
NLR	Nucleotide oligomerization domain (NOD)-like receptors
NMR	Nuclear magnetic resonance
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside/nucleotide reverse transcriptase inhibitor
NTD	N-terminal domain
NVP	Nevirapine
OD	Optical density
PAMP	Pathogen-associated molecular patterns
PBMC	Peripheral blood mononuclear cell
PE	Phycoerythrin
PHA-P	Phytohemagglutinin
PI	Protease inhibitor
$PI(4,5)P_2$	Phosphatidylinositol (4,5) diphosphate
PR	Protease
PRR	Pattern recognition receptor
RAFI	Rigid amphipathic fusion inhibitor
RAL	Raltegravir
Rev	Regulation of virion expression
RIP	Ribosome inactivating protein
RLR	Retinoid acid-inducible gene (RIG)-like receptors

RNA	Ribonucleic acid
RNase H	Ribonuclease H
RPV	Rilpivitrine
RRE	Rev response element
RSV	Respiratory syncytial virus
RT	Reverse transcriptase
RTV	Ritonavir
SAMHD1	Sterile alpha motif (SAM) and histidine-aspartic (HD)-containing protein 1
SAR	Structure-activity relationship
SIV	Simian immunodeficiency virus
SIVcpz	SIV from chimpanzees
SIVcpz*Pts*	SIV from *P. t. schweinfurthii* chimpanzee
SIVcpz*Ptt*	SIV from *P. t. troglodytes* chimpanzee
SIVgor	SIV from gorillas
SIVmac	SIV from macaque
SIVrcm	SIV from red-capped mangabeys
SIVsmm	SIV from sooty mangabey
SP1	Spacer peptide 1
SP2	Spacer peptide 2
SQV	Saquinavir
STAM	Signal transducing adaptor molecule
SU	surface
TAM	Thymidine analogue mutation
Tat	Transactivator of transcription
TDF	Tenofovir
TEER	Trans epithelial electric resistance
TLC	Thin layer chromatography
TLR	Toll-like receptor
TM	transmembrane
TPV	Tipranavir
TRIM5α	Tripartite-motif-containing 5α
UV	Ultraviolet
Vif	Viral infectivity factor
Vpr	Viral protein R
Vpu	Viral protein U

Vpx	Viral protein X
VSV	Vesicular stomatitis virus
ZDV	Zidovudine

List of tables

List of figures

List of Annexes

Chapter I

Introduction

AIDS refers to the acquired immunodeficiency syndrome, which was first clinically observed in 1981 in the United States (CDC,1981). The patients who develop AIDS suffer from a progressive failure of the immune system and subsequent life-threatening opportunistic infections and cancers to thrive. The cause of AIDS was then identified to be the human immunodeficiency virus (HIV). HIV was first isolated in early 1980s by two separate research groups, and first named HTLV-III (human T-lymphotropic virus type III) by Robert Gallo's group in the United States (Gallo, et al.,1984), and LAV (lymphadenopathy associated virus) by Luc Montagnier's group in France (Barre-Sinoussi, et al.,1983). In 1986, HIV-1 was confirmed to be the same virus, and the former HTLV-III and LAV were given the name of HIV-1.The same year, a new type of virus leading to AIDS was discovered in patients from Western Africa (Clavel, et al.,1986) and named HIV-2.

1. HIV & AIDS epidemic

As of 2014, 36.9 million (34.3 million - 41.4 million) people globally were estimated to be living with HIV (UNAIDS,2015). Most of them, 25.8 million (24.0 million – 28.7 million) were living in sub-Saharan Africa (Figure 1), where women account for more than half of that total number. Globally, there were about 1.4 million (1.2 million – 1.5 million) newly HIV-infected people. Compared to 2000, new HIV infections have fallen by 35%. Among these newly infected people, 220 000 (190 000 – 260 000) were children, which have declined by 58% since 2000.

Although antiretroviral therapies have been developed and widely used in clinics, 1.2 million (980 000 – 1.6 million) people died from AIDS-related illnesses in 2014, and 25.3 million people have died of AIDS-related illnesses cumulatively since 2000. HIV has been reported to be one of the leading causes of death for women of reproductive age worldwide. On the other hand, owing to the antiretroviral therapies, the number of AIDS-related deaths has decreased by 42% since the peak in 2004. Today about 15 million people have access to antiretroviral therapies. About 41% of all adults, 32% of all children and 73% of pregnant women living with HIV are on treatment.

In contrast to sub-Saharan Africa, the number of HIV infections has been increasing by 80% from 2004 to 2013 in Europe (ECDC,2014). In Luxembourg the number of new cases has doubled since the early 2000s, to 85 new cases in 2014, due to an increasing number of infections in injecting drug users (IDU) and men who have sex

with men (MSM) (CRP-Santé, 2014; Arendt, et al., 2015).

Importantly the UNAIDS report emphasizes that 19 million of the 37 million people living with HIV today are not aware of their HIV infection. UNAIDS announced recently their new 90-90-90 target for 2020 which should substantially reduce transmission of HIV: 90% of all people living with HIV will know their HIV status, 90% of all people diagnosed with HIV will receive sustained antiretroviral therapy and 90% of all people receiving antiretroviral therapy will have viral suppression (UNAIDS,2014). This is based on the principle that early start of therapy has not only clinical benefits for the individual patient, but prevents further transmission of HIV and has also a public health benefit.

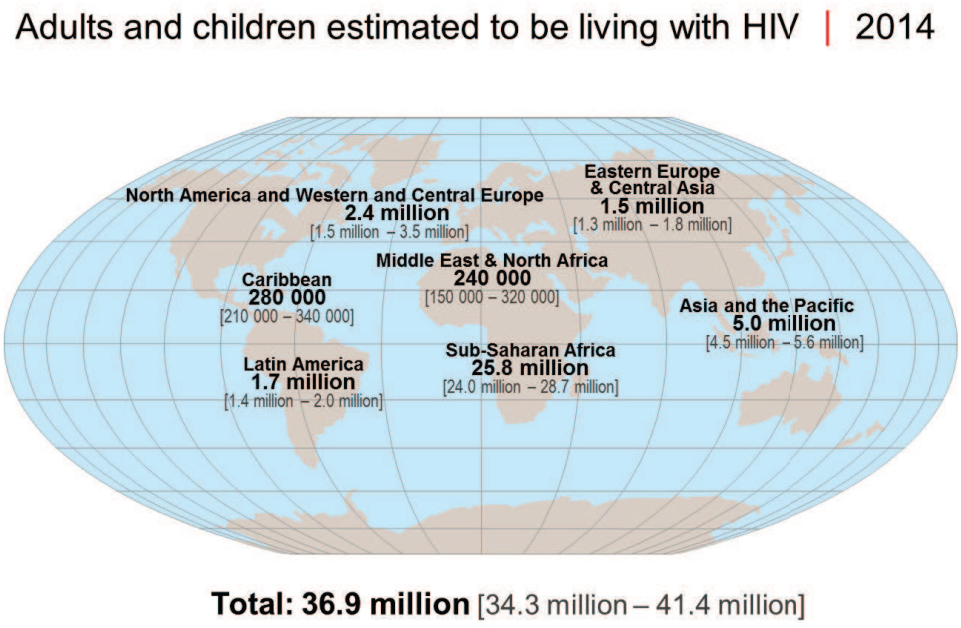


Figure 1. Adults and children estimated to be living with HIV in 2014 (UNAIDS,2015).

2. HIV pathology

2.1 HIV transmission

HIV transmission occurs through body fluids, including blood, semen, pre-seminal fluid, rectal fluids, vaginal fluids, and breast milk from an HIV-infected person (*https://aidsinfo.nih.gov/*). Up to now, the commonly known HIV transmission routes are having risky sexual contact, sharing injecting equipment, receiving blood transfusions from an HIV-infected person, and from an HIV-positive mother to her baby. Biologically, to be able to let an HIV transmission occur, the infected body fluids must contain sufficient viruses, and come in contact with a mucous membrane or damaged tissue or bloodstream. Compared to HIV-2, HIV-1 is more virulent and infective, and is responsible for worldwide epidemic. The risk of HIV transmission is mainly correlated to HIV viral load and transmission route.

HIV viral load is the amount of HIV in the body fluids of an HIV-infected person. It is clinically measured in blood, expressed in HIV-1 RNA copies present per millilitre of plasma. It has been proved that a higher viral load is associated with a higher risk of HIV transmission. Each log increment in the viral load has been reported to be associated with a rate ratio of 2.45 for seroconversion and heterosexual transmission of HIV-1 is rare among persons with levels of less than 1500 copies of HIV-1 RNA per millilitre (Quinn, et al.,2000).

HIV transmission routes have various risk levels for HIV infection. More than 80% of HIV-1 infections in adults are through risky sexual activities exposing a mucosal membrane to the virus. The rest 20% are through percutaneous and intravenous inoculations (Cohen, et al.,2011). The risk estimates for sexual transmission of HIV, per sex act, range widely, from 0.5% to 3.38% for receptive anal intercourse; 0.06% to 0.16% for insertive anal intercourse; 0.08% to 0.19% for receptive vaginal intercourse (i.e. male-to-female); and approximately 0.05% to 0.1% for insertive vaginal intercourse (i.e., female-to-male) (*http://www.phac-aspc.gc.ca/index-eng.php*). For people who inject drugs, the risk of transmission per injection from a contaminated needle has been estimated to be between 0.7% and 0.8%. The risks of transmission through oral intercourse and mother-to-child are relatively low.

2.2 Stages of HIV-1 infection

After primary HIV transmission and infection, there is a long period before HIV-1 infected individuals develop AIDS (Figure 2). It ranges from 3 to 20 years (Cohen, et al.,2011), on average about 10 years. Pathologically, it can be divided into three main stages: acute HIV-1 infection, clinical latency, and AIDS.

2.2.1 Acute HIV-1 infection

Acute HIV-1 infection is the most crucial stage of HIV infection and represents the period from HIV acquisition until seroconversion. The key characters of this period are the appearances and changes of the viral load and of the antibodies in the blood. Without any antiretroviral therapy initiation, a peak of viremia is able to reach millions copies per ml blood coinciding with a significant depletion of $CD4^+$ T lymphocytes (Brenchley, et al.,2004). Acute HIV-1 infection approximately lasts 2-3 months. During this period, about 70% of HIV-1 infected patients may develop influenza or mononucleosis-like symptoms (Pilcher, et al.,2004). The most common symptoms include fever, lymphadenophathy, rash, myalgia and malaise. According to the results of standard clinical laboratory tests, acute HIV-1 infection can be divided into six phases indicated in Figure 3 (Cohen, 2011).

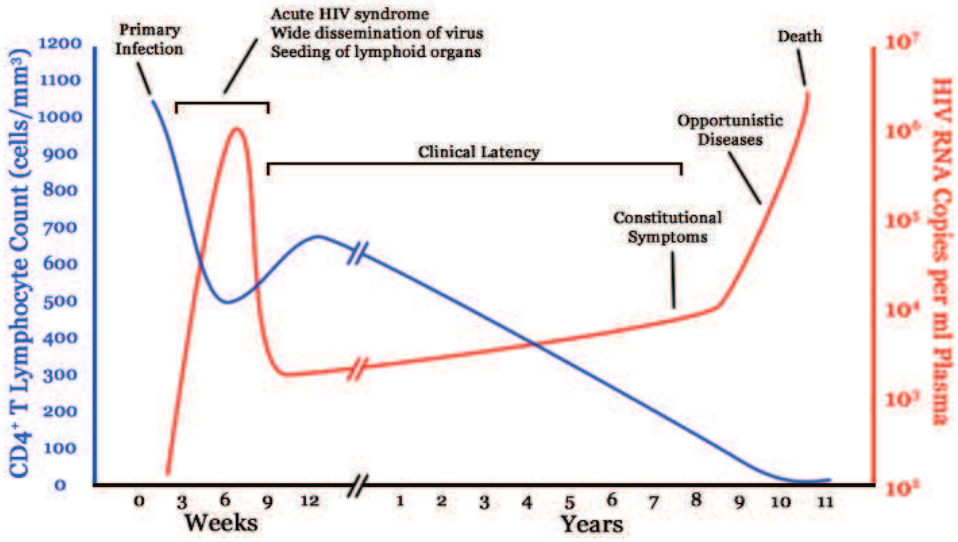


Figure 2. HIV time course (Wikimedia Commons)

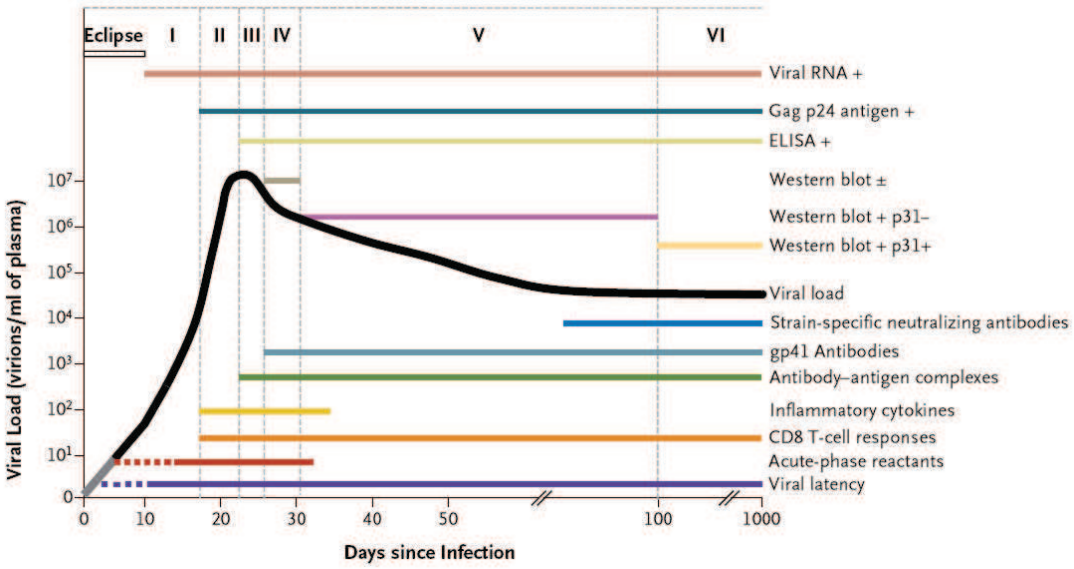


Figure 3. Natural history and immunopathogenesis of HIV-1 infection (Cohen, et al.,2011).

The eclipse phase is the period before viral RNAs become detectable in the plasma. Immediately after exposure and transmission, HIV-1 starts to replicate. The host begins to respond, producing cytokines such as interferon, leading to the classical symptoms of an acute infection (Perry, et al.,2005, Stacey, et al.,2009). During this period, the virus cannot be detected in the plasma. The eclipse phase of HIV infection is normally very short. Afterward, viral RNA increases rapidly during phase I and II and achieves a peak at the beginning of phase III. Along with the rise of viral RNA, immune cells start to respond and produce a large amount of inflammatory cytokines, resulting in a drop of the viral load later on (Stacey, et al.,2009). Antibodies specific for the enzyme-linked immunosorbent assay (ELISA) of HIV-1 proteins appear from phase III on. More specific antibodies can be detected at phase IV including gp41 antibodies (Tomaras, et al.,2009) and allow the confirmed diagnosis by the western blot method in clinics. The HIV-1 viral load drops rapidly during phase V and is sustained at a certain level at the end of HIV-1 infection. Neutralizing antibodies for specific strains and escaped mutants only occur around 3 months after primary infection (Wei, et al.,2003).

2.2.2 Clinical latency

The clinical latency stage is also known as chronic infection or asymptomatic stage. Without antiretroviral treatment, it can last for many years with an average of about eight years (Cohen, et al.,2011). Generally, HIV-1 infected individuals do not have any symptoms or very few symptoms in this period, although HIV-1 viral load is still

detectable. $CD4^+$ T cells progressively decline. Only a small group of HIV-1 infected individuals, called HIV controllers, are able to maintain a high level of $CD4^+$ T cells without any clinical interruption for several years (Pereyra, et al.,2008).

2.2.3 AIDS

AIDS is termed for the condition that an HIV-1 infected individual has less than 200 $CD4^+$ T cells per microliter and develops HIV-related symptoms. In this phase, the immune system of HIV-1 infected individuals is not able to function correctly, which makes them face the big challenge from opportunistic infections. Opportunistic infections can be caused by many organisms such as bacteria, fungi, parasites and other viruses (CDC,2014). Within the immune deficiency, HIV-1 infected individuals have more tendencies to develop cancers as well, such as Kaposi's sarcoma which is the most common cancer for HIV-1 infected individuals.

2.3 Immune response

2.3.1 Innate immune response

The innate immune system is the first line of the human body defense, which immediately responds to the primary HIV infection aiming to nonspecifically clear the invading pathogens. During this process, pattern recognition receptors (PRR), including the toll-like receptors (TLRs), the retinoid acid-inducible gene (RIG)-like receptors (RLRs) and the nucleotide oligomerization domain (NOD)-like receptors (NLR), are used to recognize the pathogen-associated molecular patterns (PAMP) of viruses and subsequently trigger innate immune cells and induce a series of inflammatory cytokines (Carrington, et al.,2012). Dendritic cells (DCs) play a central role, responding very early, in the innate immune response to HIV. After activation, DCs can release the interferons α and other cytokines, resulting in the inhibition on viral infection (Manches, et al.,2014). Natural killer cells (NKs) are also essential for the innate immune system. But at early HIV infection, viruses can evade the recognition by NKs through Nef proteins (Alter, et al.,2006). Lipopolysaccharide (LPS) is induced as well in the responses to HIV, and further triggers the inflammation cascade and activates circulating immune cells, which thereby further stimulates the viral replication (Kadoki, et al.,2010).

2.3.2 Adaptive immune response

HIV specific $CD8^+$ T cells play an essential role, associated with other cells like T helper cells and T regular cells, in the adaptive immune response to HIV. They can directly kill the infected $CD4^+$ T cells through the cytolytic effects as well as keep inducing cytokines (Davenport, et al.,2010). After acute HIV infection, the majority of $CD4^+$ T cells are significantly depleted. The earliest specific T cell response is mainly based on Env and Nef proteins (Turnbull, et al.,2009). Gag specific T cells and Pol specific T cells appear much later. In the meantime of $CD8^+$ T cells killing viruses, HIV rapidly mutates its epitope to evade, often through the effect on TCR recognition, HLA allele binding, and epitope processing. A viraemia set point can be achieved when the adaptive T cell response and HIV evasion are in a balance (Streeck, et al.,2009).

B cells are also very important in the adaptive response to HIV, due to the ability of antibody production. Both neutralizing and non-neutralizing antibodies have been found in the circulation of HIV infected patients (Moir, et al.,2009). Env specific antibodies are produced at the very early time of B cell response mainly targeting gp120 and gp41 and rarely targeting the conserved region. They are not effective against HIV due to the rapid viral mutations. After HIV acute infection, the numbers of both naïve B cells and memory B cells are decreased.

2.3.3 Restriction factors

Some host expressed proteins have been found to be able to suppress viral replication and so called restriction factors, such as the apolipoprotein B messenger RNA (mRNA)-editing enzyme catalytic polypeptide-like 3 G protein (APOBEC3G), tripartite-motif-containing 5α (TRIM5α), bone marrow stromal cell antigen 2 (BST2)/tetherin and the sterile alpha motif (SAM) and histidine-aspartic (HD)-containing protein 1 (SAMHD1). APOBEC3G is also known as CEM15, which is able to catalyze the deamination of cytidine to uridine of viral genes (Browne, et al.,2009). The Vif protein can hijack the cellular E3 ubiquitin ligase which is composed of Elongin B, Elongin C, Cullin5 and Rbx2 and cause APOBEC3G degradation (Stanley, et al.,2008, Wolfe, et al.,2010). TRIM5α can directly bind to HIV capsids and block reverse transcription. It has been reported that TRIM5α poorly inhibits retroviruses found naturally in the same host species, but is very efficient on others. BST2/Tetherin is also known as CD317, which can retain the nascent virions on the cell surface and thereby, after accumulation, virions are internalized within endosomes (Neil,2013). The gene encoding human SAMHD1 has been reported to restrict the HIV infection of DCs and other immune cells

through the depletion of the intracellular pool of deoxynucleoside triphosphates (dNTPs) (Lahouassa, et al.,2012).

2.3.4 Microbial translocation

Gut-associated lymphoid tissue (GALT) contains a large amount of activated $CD4^+CCR5^+$ T cells, which is the major place of viral replication. HIV infection can lead to a serious damage of mucosa directly through the disruption of the intestinal epithelial barrier by increasing cell apoptosis and destroying the tight junctions (Estes, et al.,2010), and can cause microbial overgrowth due to the loss of Th17 cells (Chege, et al.,2011) resulting in a less clearance of microbial products such as LPS (Balagopal, et al.,2009) (Figure 4). Subsequently microbial products translocate from the gastrointestinal lumen into the circulation. Microbial translocation has been proposed to be associated with systemic immune activation and inflammation which are key drivers of AIDS and AIDS related diseases progression (Sandler, et al.,2012). High levels of LPS, soluble CD14 as well as bacterial 16S rDNA are found in plasma of HIV infected individuals, and were correlated with high frequency of activated $CD4^+$ and $CD8^+$ T cells and high viral load (Wallet, et al.,2010). Recently, Kristoff et al. have provided a direct evidence that blocking LPS translocation using sevelamer in the gut can dramatically reduce the immune activation and inflammation and slightly reduce viral replication of SIV-infected pigtailed macaques (Kristoff, et al.,2014).

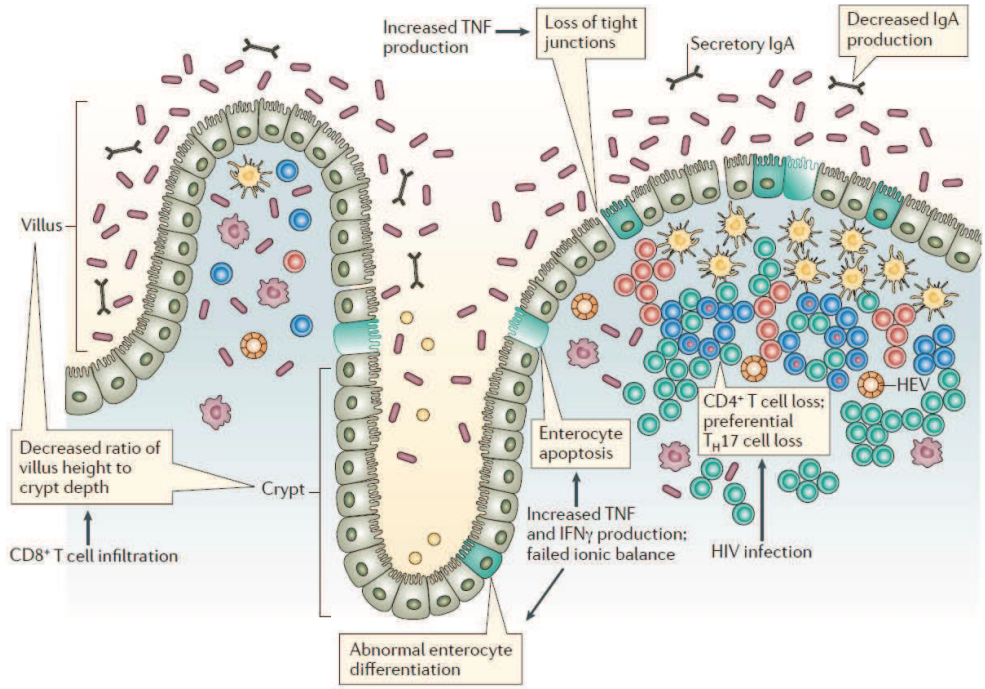


Figure 4. The intestinal epithelium in an HIV-infected individual (Sandler, et al.,2012)

3. HIV origin

Although HIV was identified as a retrovirus belonging to the lentivirus family only in 1980s, it existed already a long time before. In 1987, HIV-2 was identified to be closer to a simian virus than to HIV-1 (Chakrabarti, et al.,1987). It indicated that HIV-1 and HIV-2 might derive from simian viruses. In 1989 and 1990, close simian relatives of HIV-1 and HIV-2 were found in sooty mangabeys (Hirsch, et al.,1989) and chimpanzees (Huet, et al.,1990) respectively. HIV-1 and HIV-2 were then proved to be the result of zoonotic transfers of viruses infecting primates in Africa (Hahn, et al.,2000) (Figure 5). Simian immunodeficiency viruses (SIV) have been found in various different primates from sub-Saharan Africa, including African green monkeys, sooty mangabeys, mandrills, chimpanzees and other primates (Worobey, et al.,2010). Both HIV-1 and HIV-2 were found to be in the same phylogenetic cluster with SIVs (Figure 6).

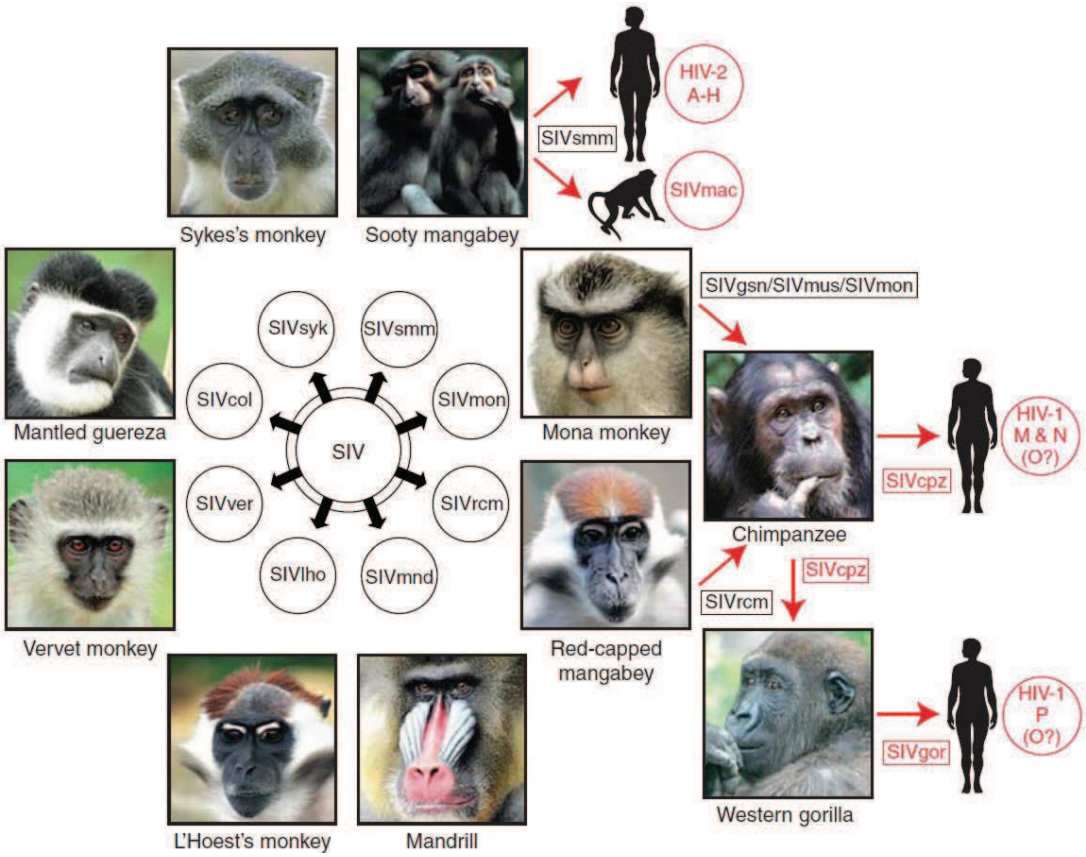


Figure 5. Origin of human AIDS viruses (Sharp, et al.,2011).

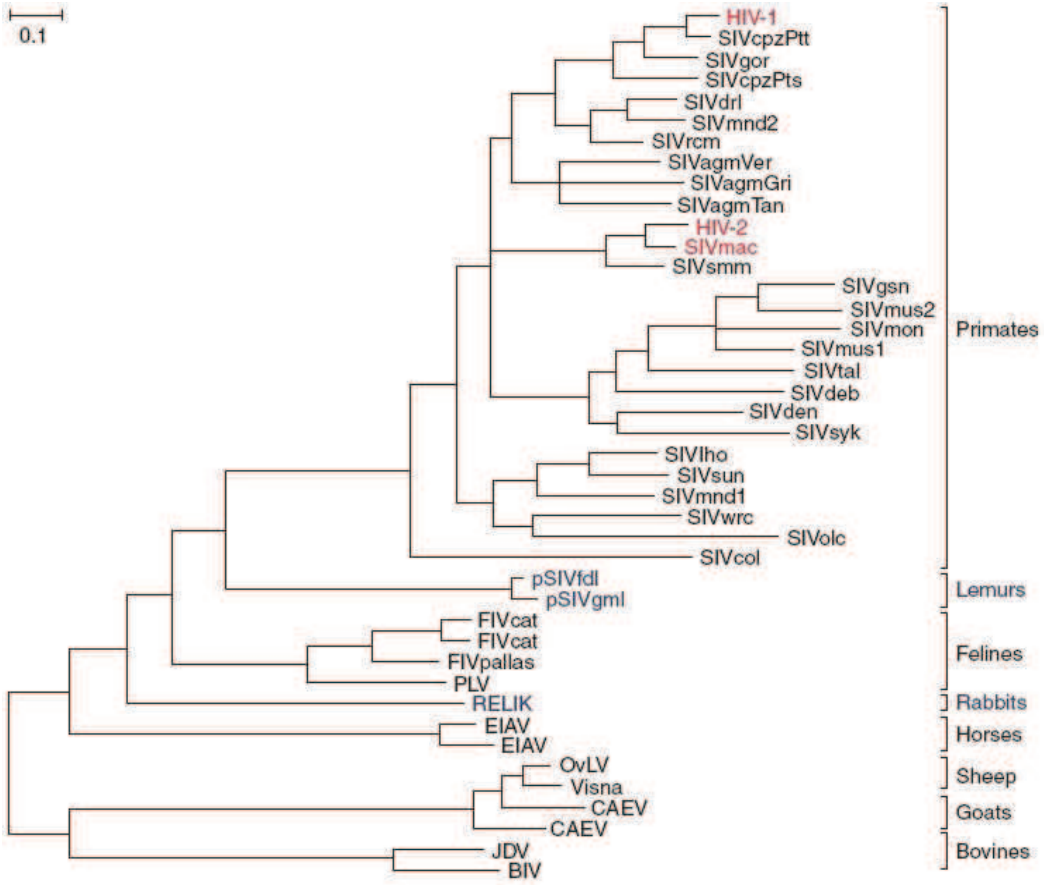


Figure 6. Phylogeny of lentiviruses (Sharp, et al.,2011).

3.1 Origins of HIV-1

Phylogenetic studies have shown that HIV-1 is most closely related to SIVcpz (SIV from chimpanzees) and SIVgor (SIV from gorillas).

There are two chimpanzee species: the common chimpanzee (*Pan troglodytes*) and the bonobo (*Pan paniscus*). The common chimpanzees have been divided into four subspecies according to their mitochondrial sequences: western chimpanzee (*P. t. verus*), Nigeria-Cameroonian chimpanzee (*P. t. ellioti* or *P. t. vellerosus*), central chimpanzee (*P. t. troglodytes*) and eastern chimpanzee (*P. t. schweinfurthii*) (Colin,2001). These four subspecies have non-overlapping ranges across Western and Central Africa (Figure 7) (Sharp, et al.,2011). The central and eastern chimpanzees were identified as the natural SIVcpz reservoir, and established as SIVcpz*Ptt* (SIV from *P. t. troglodytes* chimpanzee) and SIVcpz*Pts* (SIV from *P. t. schweinfurthii* chimpanzee). SIVcpz is believed to be a result of recombination of

different SIVs. It has been proved that the 5' half of the genome, nef gene and 3' LTR of SIVcpz are very close to SIVrcm (SIV from red-capped mangabeys), while other parts of the SIVcpz genome including the vpu, tat, rev env genes are close to a cluster of SIVs infecting several *Cercopithecus* species, such as greater spotnosed (*C. nictitans*), mustached (*C. cephus*), and mona (*C. mona*) monkeys (Bailes, et al.,2003).

Virology of SIVcpz and HIV-1 is quite similar as well. SIVcpz has been mainly revealed to be non-pathogenic in its natural host, but some cases showed that it was able to affect the chimpanzee's immune system and decrease the life time of infected chimpanzees. SIVcpz transmits mainly through sexual activities, as well as from mother to infants (Apetrei, et al.,2004). Migration of infected females constitutes a major route of virus transmission between communities.

Two types of gorilla are living in sub-Saharan Africa: western lowland gorilla (*G. gorilla*) and eastern gorillas (*G. beringer*) (Figure 8) (Sharp, et al.,2011, D'Arc, et al.,2015). SIVgor infection occurs more often on the western lowland gorillas than on the eastern gorillas (Van Heuverswyn, et al.,2006). Due to a lack of study on the natural infection in the wild-living gorillas, SIVgor's effect on its natural host is still not clear. But from the phylogenetic studies, SIVgor is believed to be derived from SIVcpz (Takehisa, et al.,2009, D'Arc, et al.,2015). Gorillas are found to act as a potential reservoir for human infection (LeBreton, et al.,2014).

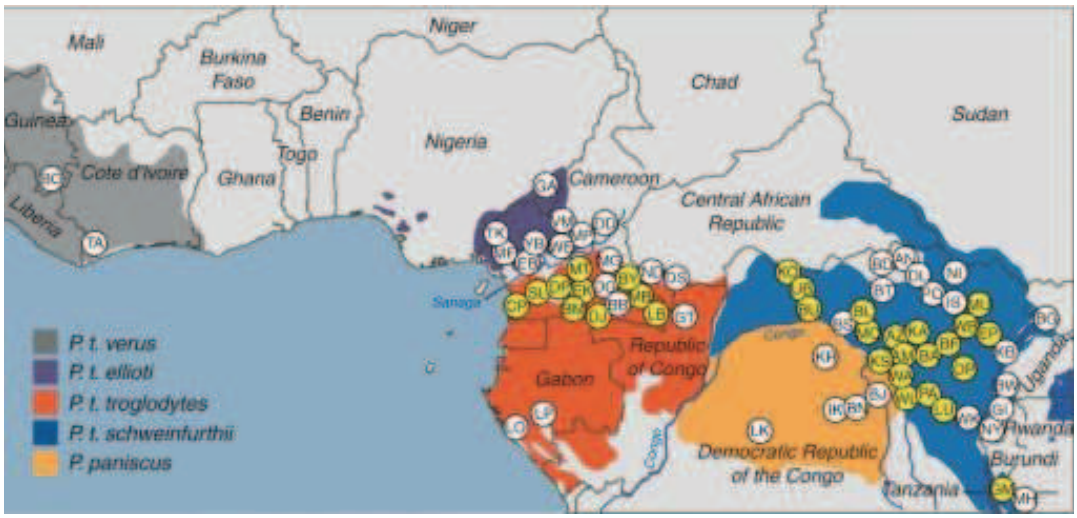


Figure 7. Geographic distribution of chimpanzees and SIVcpz in sub-Saharan Africa (Sharp, et al.,2011).

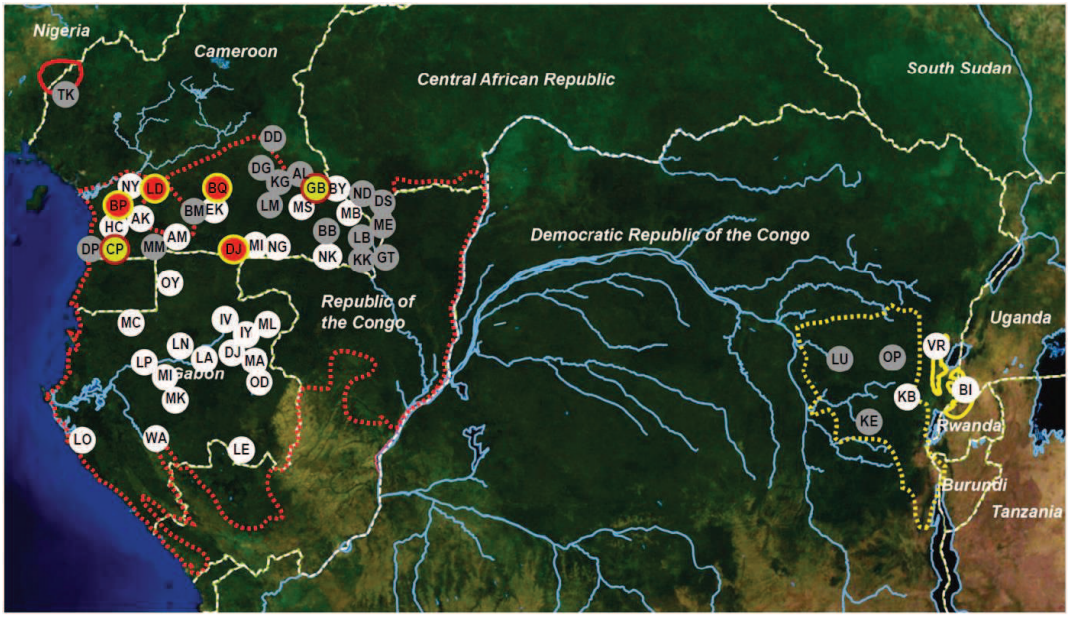


Figure 8. Geographic distribution of SIVgor in wild-living gorillas (D'Arc, et al.,2015).

3.2. HIV-1 subtype

HIV-1 strains can be classified into four groups: the 'major' group M, the 'outlier' group O, the non-M/non-O groups N, and a new group P. HIV-1 groups M and N are most closely related to SIVcpz*Ptt* strains, originally from southern and south-central Cameroon respectively (Figure 9) (Sharp, et al.,2011). HIV-1 group O and P are very close to SIVgor strains, originally from west-central Africa (D'Arc, et al.,2015).

More than 90% of HIV-1 infections belong to the HIV-1 group M (Tatt, et al.,2001). Group M includes subtypes A, B, C, D, F, G, H, J and K, as well as CRFs (circulating recombinant forms). HIV-1 subtypes A, B and C are the most widespread strains (Figure 10). Subtype A is predominant in Central and East Africa and in eastern European countries (former Soviet Union). Subtype B is the main genetic form in western and central Europe, in the Americas, in Australia, and in eastern Asia. Subtype C predominates in South and East Africa, India and Nepal, and is responsible for almost 50% of all HIV infections (Osmanov, et al.,2002). Group O is restricted to West and Central Africa and accounts for less than 5% of all HIV infections in those areas (D'Arc, et al.,2015). Group N and group P were discovered in Cameroon in 1998 (Simon, et al.,1998) and in 2009 (Plantier, et al.,2009) respectively . Both group N and group P are extremely rare. In Europe, the HIV epidemic was initiated by subtype B and increasing HIV genetic diversity has been now described (Chaix, et al.,2013,

Neogi, et al.,2014). In Luxembourg, the majority of new infections are currently caused by HIV subtypes other than B and an increased prevalence of recombinant strains was observed (Struck, et al.,2015). Analysis of 410 non-B HIV infections between 2002-2012 in Luxembourg, Ghent and Antwerp showed that 30% of non-B infections occurred now in the native Western European population, indicating a significant spread of originally African and Asian subtypes among the local population [(Dauwe K et al, in press] as it was recently reported in the case of a Dutch cohort (Hofstra Marije,2015).

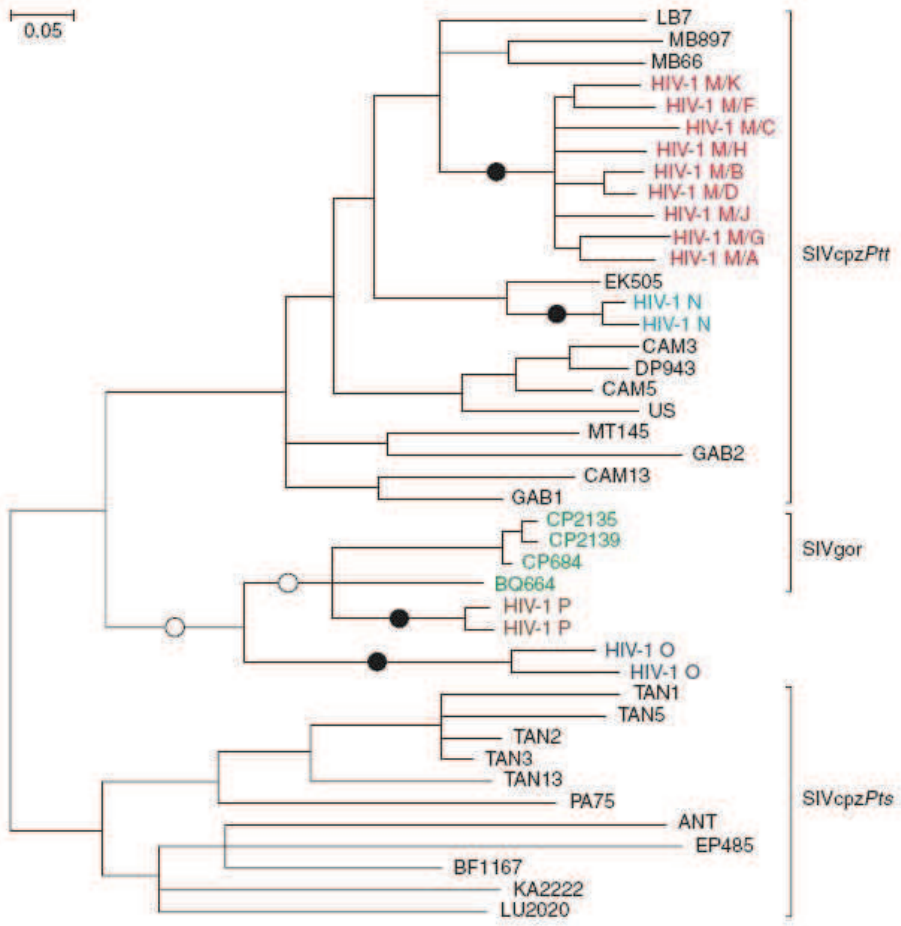


Figure 9. HIV-1 origins (Sharp, et al.,2011).

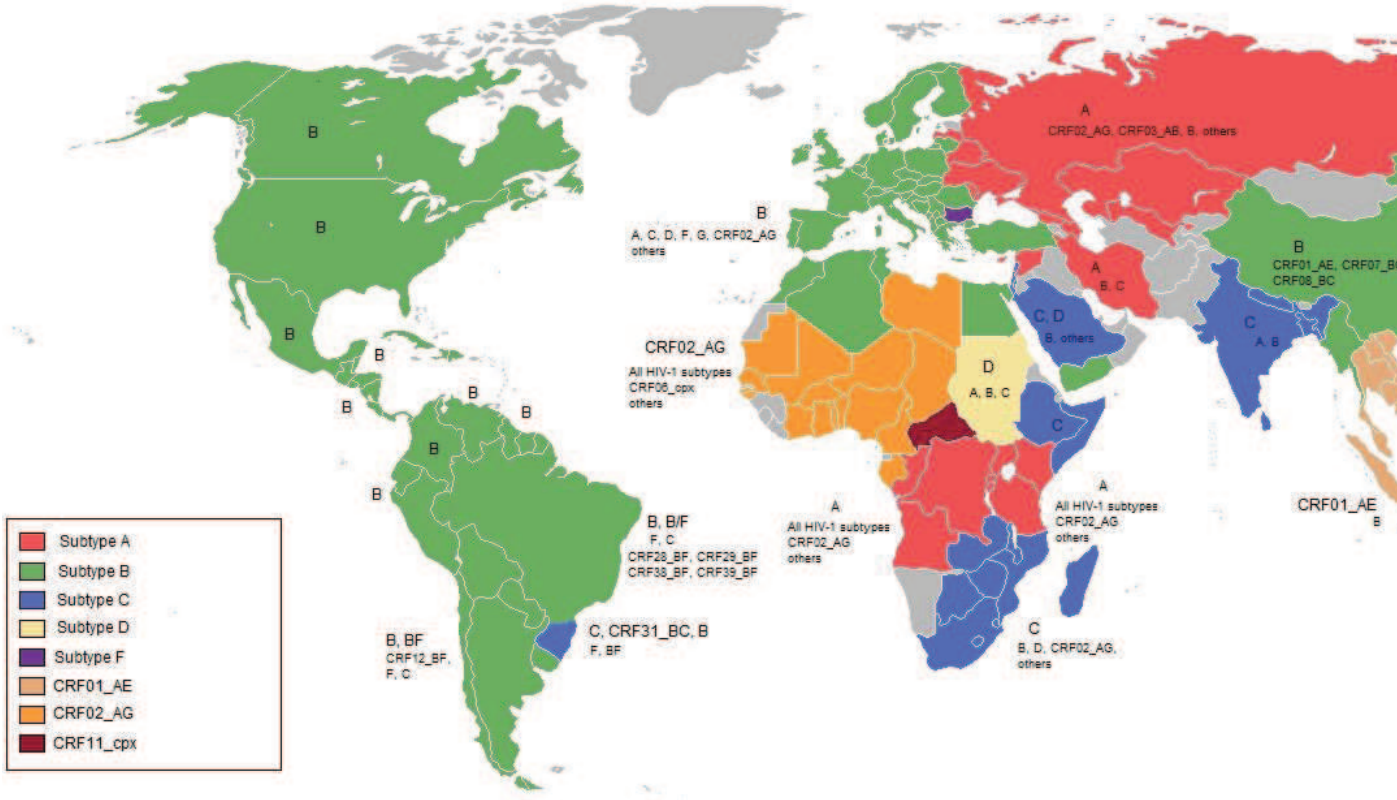


Figure 10. Global distribution of HIV-1 group M strains (Santos, et al.,2010).

3.3 Origins of HIV-2

The origin of HIV-2 was resolved before the origin of HIV-1 (Gao, et al.,1992). Phylogenetic studies have shown that HIV-2 is closer to SIVmac (SIV from macaque) and SIVsmm (SIV from sooty mangabey) (Figure 11) (Sharp, et al.,2011). The early isolated HIV-2 strains in West Africa were found to be derived from to local SIVsmm (Chen, et al.,1997). Transmission was believed to occur via the sooty mangabeys that have been caught by local people and used for agricultural tasks.

3.4 HIV-2 subtypes

HIV-2 is restricted to West Africa, where it was first discovered. It can be categorized into eight subtypes A to H, similar to HIV-1. HIV-2 subtypes A and B have been identified to be responsible for human epidemic (Damond, et al.,2001). HIV-2 subtypes C to H were only found in single individuals (Damond, et al.,2004).

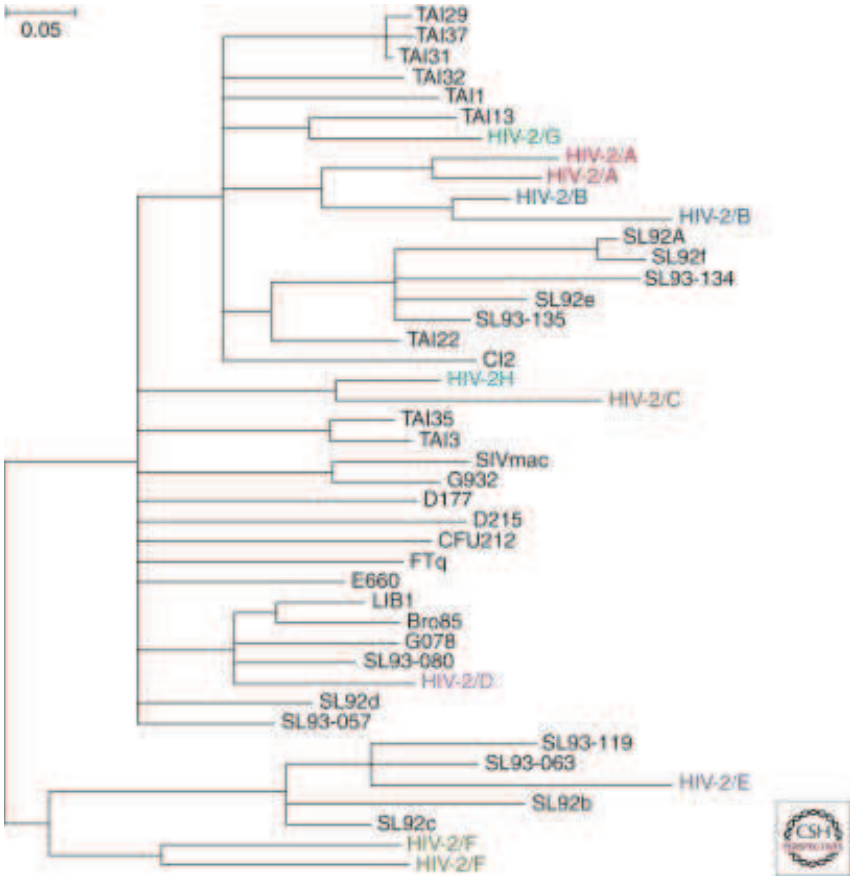


Figure 11. HIV-2 origin (Sharp, et al.,2011).

The phylogenetic relationships of representative SIVsmm and HIV-2 strains are shown for a region of the viral *gag* gene. SIVsmm and SIVmac are shown in black.

4. HIV structure and genome

4.1 HIV structure

The HIV mature virion is nearly spherical, with a diameter of around 100 nm (Figure 12) (Coffin JM, et al.,1997). It is enveloped by a lipid bilayer membrane, which is derived from the host cell membrane. Besides the items obtained from the host, the most crucial outer proteins of the HIV virion are the surface glycoprotein gp120 and the trans-membrane protein gp41. The inner structure of the HIV virion contains the matrix protein (MA, p17), which connects to the lipid membrane; the capsid protein (CA, p24), which constitutes the core shell of the virion, surrounds the RNA genome and the nucleocapsid (NC, p7); and the viral enzymes integrase (IN), reverse transcriptase (RT) and protease (PR).

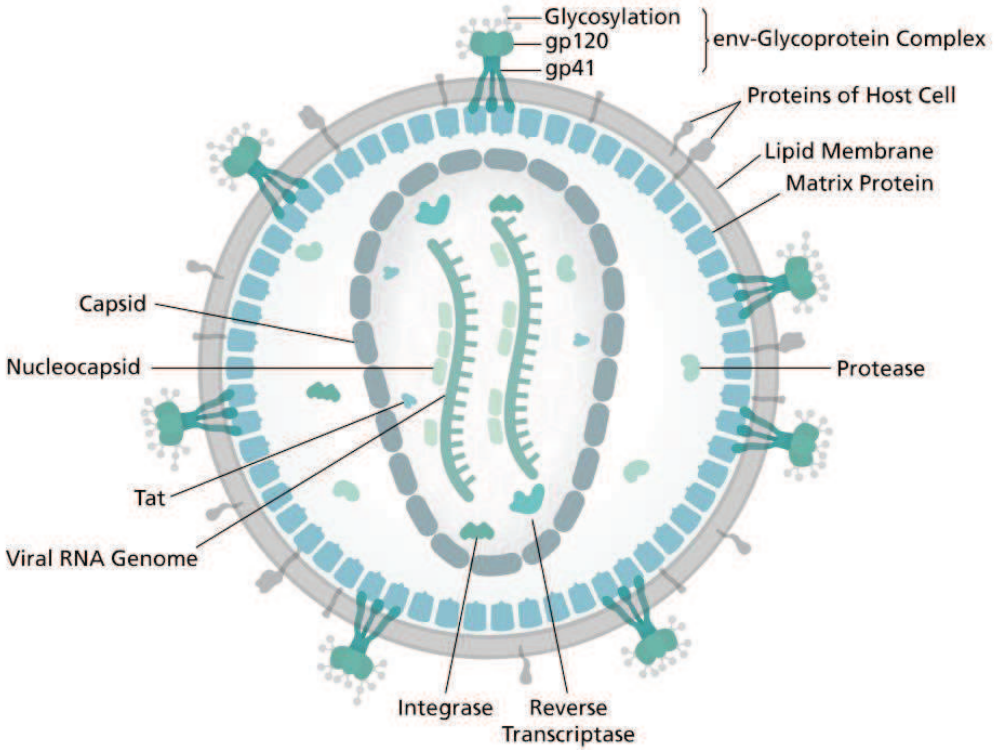


Figure 12. HIV structure (Wikimedia Commons).

4.2 HIV genome

The HIV genome is composed of two identical copies of single-stranded RNA molecules (Coffin JM, et al.,1997). It is approximately 9.2kb in length and contains nine open reading frames. The three largest reading frames encode the structural genes *gag*, *pol* and *env*, which are transcribed and translated into the Gag, Pol and Env polyproteins respectively. Gag is post-translationally processed into matrix (MA), capsid (CA), nucleocapsid (NC) and the P6 protein. Pol contains three viral enzymes: the integrase (IN), the reverse transcriptase (RT) and the protease (PR). Env encodes the protein gp160, which can be cleaved into gp120 (surface, SU) and gp41 (transmembrane, TM) proteins. Both HIV-1 and HIV-2 have the same basic structure, but present different combinations of regulatory and accessory genes (Figure 13) (Fanales-Belasio, et al.,2010). HIV-1 reading frames encode Vif, Vpr, Vpu, Tat, Rev and Nef, while HIV-2 reading frames encode Vif, Vpx, Vpr, Tat, Rev and Nef. Figure 14 is a more detailed genome structure of the reference strain HIV-1 HXB2 strain.

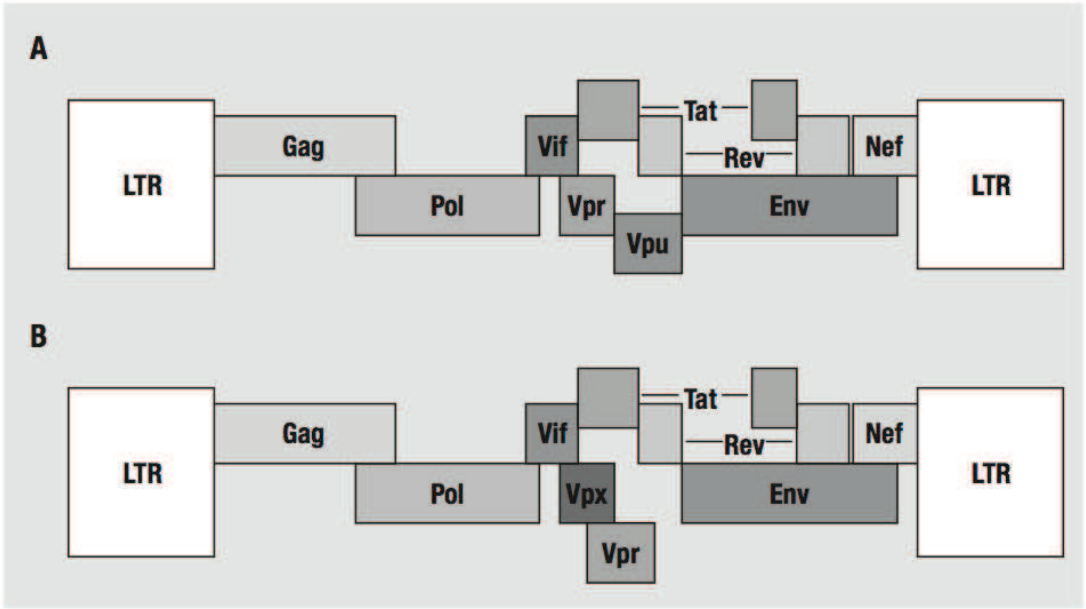


Figure 13. Organisation of the HIV genome (Fanales-Belasio, et al.,2010).
A. HIV-1 genome. B. HIV-2 genome.

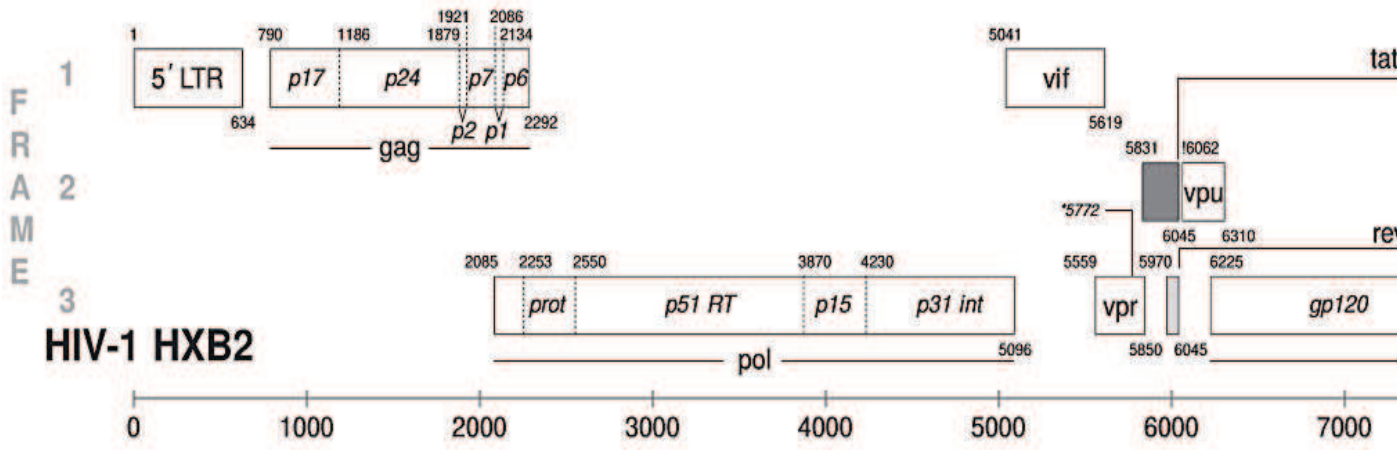


Figure 14. HIV-1 HXB2 strain genome.

From http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

4.2.1 Structural genes and their encoded proteins

4.2.1.1 Gag

Gag (group specific antigen) is encoded by the *gag* gene, and is also known as p55 (full length about 55 kDa). It serves multiple functions in viral assembly, budding and maturation processes (Göttlinger,2001). The Gag protein has four major domains including the matrix (MA) at the N terminus, the capsid (CA), the nucleocapsid (NC), and the p6 protein at the C terminus, as well as two spacer peptides SP1 and SP2 (Figure 15) (Henderson, et al.,1992, Freed,2015). These domains are cleaved during viral maturation.

MA, also known as p17, is about 17 kDa in size, with five alpha helices and a three-strand mixed beta sheet. The N terminus of MA is responsible for myristoylation, which is crucial for viral assembly and interaction with the cell membrane (Bryant, et al.,1990). Within the Gag protein, MA domain is capable to interact with the Env protein, with nucleic acids as well as with the lipid membrane. It is involved in Gag trafficking and recruitment of viral and host proteins (Ghanam, et al.,2012). After cleavage, the MA protein stays at the inner side of lipid membrane of the virions.

CA, also known as p24, is about 24 kDa in molecular weight, is divided into 3 parts - the C-terminal domain (CTD), the N-terminal domain (NTD) and a flexible linker in between (Momany, et al.,1996). CTD has four alpha helices and a three-strand helix, essential for dimerization (Gamble, et al.,1997), while NTD has seven alpha helices and a proline-rich loop and plays a key activity in early capsid formation (Gitti, et al.,1996). The flexible linker is capable to correct core assembly, viral replication and infectivity (Jiang, et al.,2011). In mature virions, CA forms the capsid which is the structural core of virions (Gelderblom,1991).

NC, also known as p7, is about 7 kDa, containing two specific zinc fingers which allow the binding of packaging signals (Bess, et al.,1992, Berkowitz, et al.,1996). NC has been reported to be essential not only for RNA binding, but also for Gag-Gag interaction and mature protein cleavage during viral replication and assembly processes (Burniston, et al.,1999, Freed,2015).

P6, is about 6 kDa, containing two helices (Henderson, et al.,1992). It has been shown that the p6 protein has multiple functions including binding to the accessory protein Vpr and recruiting host machineries for viral budding (Freed,2002).

SP1, also known as p2, is structurally connected with the C terminus of CA and is functionally important for viral CA assembly (Krausslich, et al.,1995).

SP2, also known as p1, has been reported to function in the incorporation between Gag and Pol (Hill, et al.,2002).

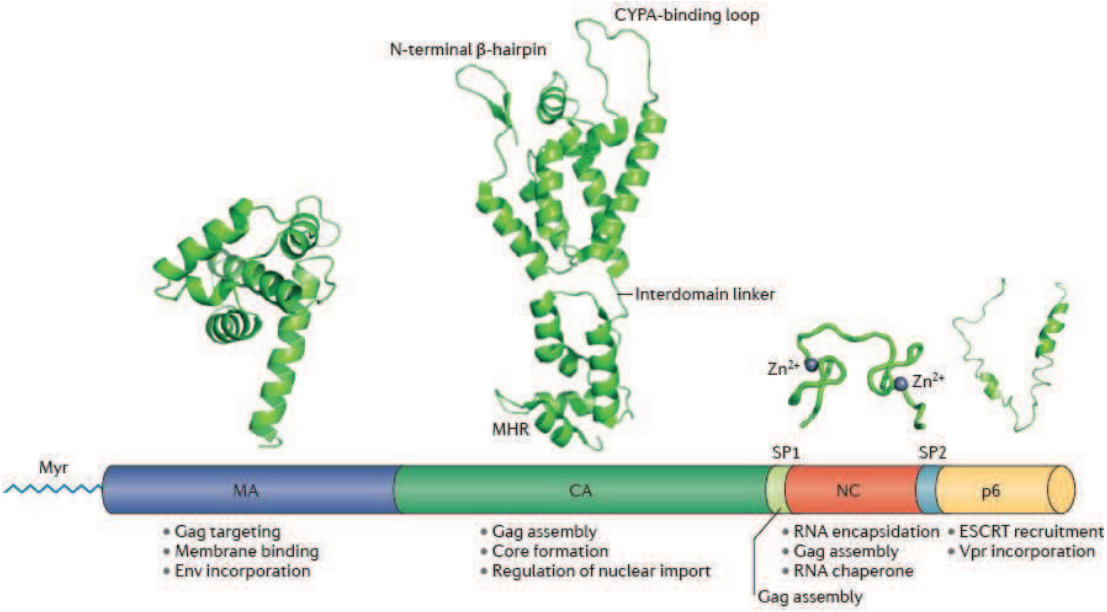


Figure 15. The structure and functions of Gag (Freed,2015).

4.2.1.2 Pol

Pol (DNA polymerase) is encoded by the *pol* gene and expressed as Gag-Pol precursor protein of about 160 kDa (Pr160-Gag-Pol, p160) (Hill, et al.,2001). Pol encodes the viral enzymes protease (PR), reverse transcriptase (RT) and integrase (IN) (Coffin JM, et al.,1997).

PR, also known as p10, has about 99 amino acids. It functions as a homodimer with only one active site (Restle, et al.,1990). PR is an enzyme required for proteolytic processing of viral polyproteins, which is critical for viral assembly, maturation and infectivity (Freed,2015). PR has to cleave viral polyproteins not only at the specific sites but also in a specific order (Figure 16) (Huang, et al.,2013).

RT has two subunits - one of about 66 kDa (p66), and the other one of about 51 kDa (p51) (di Marzo Veronese, et al.,1986). The p66 subunit contains a DNA polymerase domain and a ribonuclease H (RNase H) domain, while the p51 subunit only has a DNA polymerase domain (Kohlstaedt, et al.,1992). The absence of RNase H domain

in the p51 subunit is due to the proteolytic cleavage (Huang, et al.,2013). RT is the enzyme required during reverse transcription. In this process, the viral RNA is transcribed into its complementary DNA (Hu, et al.,2012).

IN is a 32 kDa protein encoded by the C terminal portion of the *pol* gene. It has three major domains – the N terminal domain, the central domain and the C terminal domain (Chiu, et al.,2004). IN is an enzyme required during integration of viral DNA into the host genome. The N terminal domain is able to bind Zn^{2+} (Engelman, et al.,1992, Zheng, et al.,1996). The central domain carries the catalytic activity (Jenkins, et al.,1995). The C terminal domain has a non-specific DNA binding activity and is involved in the stability of pre-integration complex (Chen, et al.,2000, Delelis, et al.,2008).

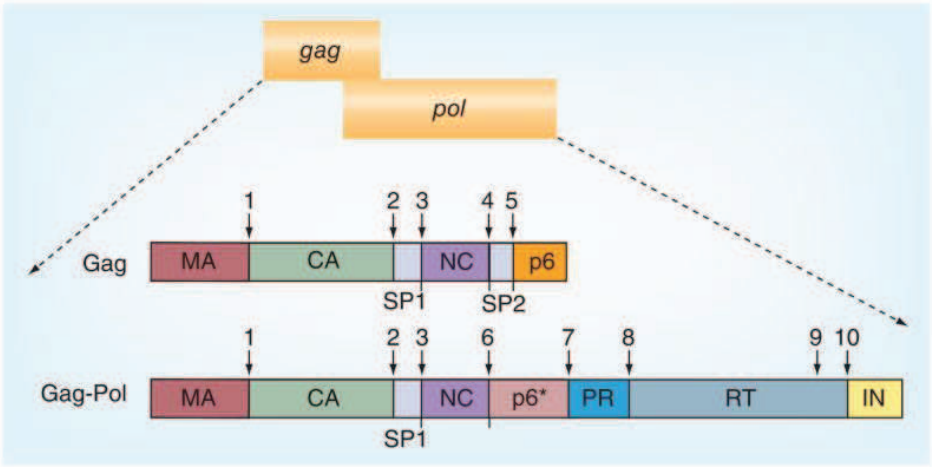


Figure 16. HIV-1 proviral genome and the protease cleavage sites in the Gag and GagPol (Huang, et al.,2013).

4.2.1.3 Env

The *Env* gene encodes a 160 kDa protein called gp160, which forms the viral envelope. It exists as a trimer on the viral surface (Lu, et al.,1995). The gp160 protein is cleaved by the host protease furin into two products – gp120 and the gp41. The gp120 protein is a glycoprotein located at the surface of virions, containing five hypervariable regions V1 to V5 (Merk, et al.,2013). It is known that the V3 loop is responsible for binding to the CD4 receptor and to the CCR5/CXCR4 co-receptor on the cell membrane and determines the viral tropism (Hung, et al.,1999, Dettin, et al.,2003). The gp41 protein is a transmembrane protein, containing heptad repeat domains which are able to form a six-helix bundle after exposure and lead to the fusion of viral and cell membranes (Chan, et al.,1997, Blumenthal, et al.,2012).

4.2.2 Regulatory genes and coded proteins

4.2.2.1 Tat

The Tat protein (Trans-activator of transcription) is encoded by the *tat* gene, composed of two exons of the HIV genome (Watson, et al.,1999). It is about 9-14 kDa, depending on the viral strains. The first exon encodes the first 72 amino acids, while the rest are encoded by the second exon (Kuppuswamy, et al.,1989). The Tat protein has a basic domain responsible for RNA binding (Karn, et al.,1991) and a protein transport domain which allows Tat to cross the cell membrane and translocate into the nucleus (Fawell, et al.,1994). The Tat protein is an early regulatory protein in viral replication, which significantly enhances viral DNA transcription and helps unspliced viral mRNA transport out of the nucleus (Vives, et al.,1997).

4.2.2.2 Rev

Rev is an approximately 13 kDa protein encoded by the *rev* gene. It exists as a homo-tetramer in the virions (Pollard, et al.,1998). The Rev protein contains an arginine-rich domain which allows its free transportation between cell cytoplasm and nucleus (Truant, et al.,1999). The Rev protein is crucial to the transition of viral transcription from early to late phase (Zhao, et al.,2012). It is produced by completely spliced viral transcripts at the early phase and is able to bind non-spliced viral mRNA at the rev response element (RRE) which helps viral structure protein producing mRNA exportation from nuclear to cytoplasm (Fernandes, et al.,2012).

4.2.3 Accessory genes and coded proteins

4.2.3.1 Nef

Nef protein stands for negative regulatory factor. It is a 27 – 32 kDa myristoylated protein encoded by the *nef* gene. It contains a flexible N terminal domain, a globular core domain, and a C terminal disordered loop (Lee, et al.,1996). The Nef protein has multiple functions including making contact with plasma and nuclear membranes (Fujii, et al.,1996), down-regulating host cellular proteins (Das, et al.,2005) and affecting virus infectivity (Basmaciogullari, et al.,2014). It decreases the expression of CD4 on the cell surface by increasing CD4 endocytosis (Laguette, et al.,2009) and reduces the major histocompatibility complex-1 (MHC-1) which facilitates virus replication and prevents the elimination of infected cells by the immune system (Vigerust, et al.,2005).

4.2.3.2 Vpu

Vpu (viral protein unique) is encoded by the *vpu* gene and translated from vpu-env bicstronic mRNA. It is a 16 kDa integral membrane protein unique to HIV-1 (Strebel, et al.,1989). The N terminal domain of Vpu is a transmembrane anchor, which is able to augment virus release from the plasma membrane (Bour, et al.,2003). The C terminal domain of Vpu can be phosphorylated and down-regulate CD4 expression in the endoplasmic reticulum (Tiganos, et al.,1997).

4.2.3.3 Vif

Vif (viral infectivity factor) is a 23 kDa protein encoded by the *vif* gene. It is required for viral replication in non-permissive cells, such as lymphocytes, macrophages and certain cell lines (Gabuzda, et al.,1992). The Vif protein is critical for viral replication via interaction with APOBEC3G (Henriet, et al.,2009).

4.2.3.4 Vpr

Vpr (viral protein R) is a 14 kDa protein encoded by the *vpr* gene (Cohen, et al.,1990, Bukrinsky, et al.,1999). It is incorporated into mature virions (Muller, et al.,2000). Each virion has about 100 copies of Vpr (Coffin JM, et al.,1997). It is a component of the pre-integration complex and facilitates viral genome entry into the nucleus (Depienne, et al.,2000). It is also able to arrest cell division by accumulating at the G2 phase of the cell cycle and preventing cell entry into mitosis (Romani, et al.,2015).

4.2.3.5 Vpx

Vpx is a protein encoded by the *vpx* gene, unique to HIV-2 (Coffin JM, et al.,1997). It has a similar size as the Vpr protein and is believed to share the function of Vpr (Khamsri, et al.,2006). In HIV-2 replication, Vpx functions in the transport of the pre-integration complex into the nucleus (Pancio, et al.,2000, Cheng, et al.,2008). The Vpx protein is able to neutralize the host factor SAMHD1 which is an enzyme responsible for blocking HIV replication in some cells such as dendritic cells, macrophages and monocytes (Baldauf, et al.,2012, Hofmann, et al.,2012).

5. The HIV-1 replication cycle

Figure 17 depicts a general HIV replication cycle in $CD4^+$ T cells (Barre-Sinoussi, et al.,2013). In a first step of the viral replication cycle, the HIV virion makes contact with the cell membrane. Then glycoproteins on the virion surface bind to the cell receptor CD4 and co-receptor CXCR4/CCR5 successively (Hung, et al.,1999, Dettin, et al.,2003, Blumenthal, et al.,2012). After fusion of the viral membrane with the cell membrane, the virion inserts its core into the cell. The viral capsid is uncoated and the HIV RNA and proteins are released into the cell cytoplasm. HIV RNA is reverse-transcribed into its complementary DNA, which is then integrated into the host genome (Delelis, et al.,2008, Hu, et al.,2012). New viral proteins are produced using the host machinery. Viral proteins are translocated to the cell membrane and assembled. After budding from the cell membrane, the virion hijacks envelope proteins and part of the lipid membrane from the host cell, and develops into a new mature virus (Freed,2015).

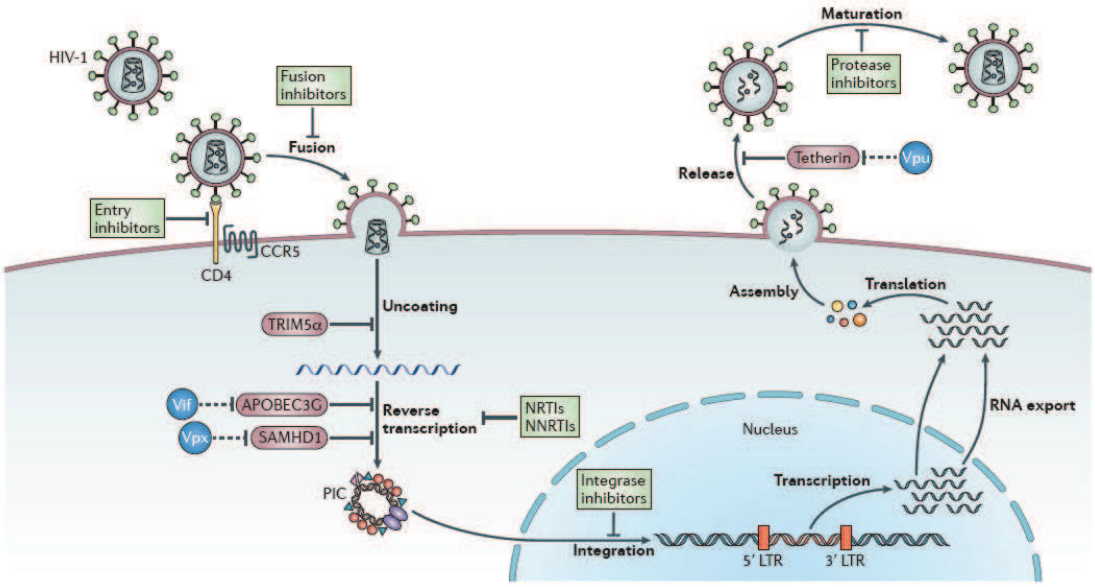


Figure 17. Schematic overview of the HIV-1 replication cycle (Barre-Sinoussi, et al.,2013).

5.1 HIV-1 attachment and entry

HIV attachment is a relatively non-specific process, driven by the virion envelope and the negatively charged heparin sulphate proteoglycans on the cell surface (Saphire, et al.,2001, Liu, et al.,2002). In some cases, the attachment of the virions to the host cell

can be driven by the interaction with specific molecules, such as the dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin (DC-SIGN) (Tsegaye, et al.,2010), the macrophage mannose specific receptor (MMR) (Nguyen, et al.,2003), α4β7 integrin (Perez, et al.,2014), and galactocerebroside (gal-C) (Harouse, et al.,1991). DC-SIGN, also known as CD209, is a C-type lectin present on macrophages and dendritic cells, which has a very high affinity to the intercellular adhesion molecule 3 (ICAM3, CD50) (Geijtenbeek, et al.,2000). It can interact with the HIV-1 envelope resulting in viral particle endocytosis (Kwon, et al.,2002) and can transmit HIV from the macrophages to the $CD4^+$ T cells by recognizing ICAM3 (Trumpfheller, et al.,2003). MMR is a protein of about 180 kDa that protrudes in an extended conformation from the macrophage cell surface (Pontow, et al.,1992). Many studies have shown that MMR is able to mediate the initial association of HIV and macrophages (Nguyen, et al.,2003), especially for those cells which lack DC-SIGN expression. The gut-homing receptor, integrin α4β7, has been proved to mediate HIV entry by interaction with the V2 loop of HIV-1 gp120 both *in vitro* and *in vivo* (Arthos, et al.,2008, Perez, et al.,2014, Tassaneetrithep, et al.,2014). Gal-C is the specific receptor for HIV-1 gp120 in neurons and glial cells (Harouse, et al.,1991).

5.2 HIV-1 fusion

The fusion between HIV-1 virion and cell membrane is mediated by viral envelope proteins and cell receptor and co-receptors (Blumenthal, et al.,2012). The HIV-1 envelope is composed by two subunits: glycoprotein gp120 and transmembrane protein gp41, present in trimers – three gp120 molecules non-covalently combined with three gp41 molecules (Decroly, et al.,1994, Lu, et al.,1995, Merk, et al.,2013). Gp120 is able to bind CD4 Domain 1, which causes the rearrangement of the V1/V2 loop of gp120 and subsequently the repositioning of V3 loop (Figure 18) (Wilen, et al.,2012). After V3 loop exposure, the CD4-gp120 complex is able to interact with a secondary receptor on the cell surface, which is called HIV co-receptor. CCR5 and CXCR4 are the two major co-receptors. After coming close to the cell membrane, the fusion peptide HIV-1 gp41 can insert into the cell membrane and form a fusion complex with a six-helix bundle (Markosyan, et al.,2009, Buzon, et al.,2010). It drives the insertion of viral RNA and enzymes into the cell cytoplasm via the fusion pore formed between HIV-1 virions and host cells.

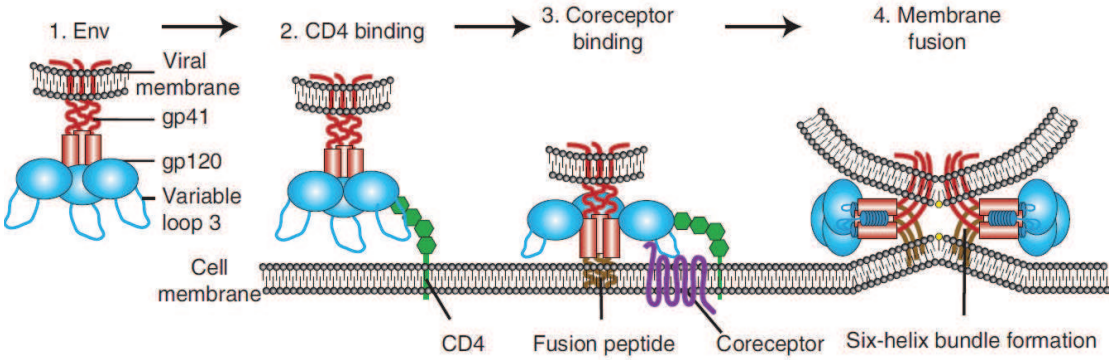


Figure 18. Overview of HIV entry (Wilen, et al.,2012).

5.3 Reverse transcription

After HIV virion entry into the host cell, the viral RNA is transcribed into its complementary DNA by the reverse transcriptase (RT). This process is initiated from a host tRNA primer, $tRNA_{lys3,}$ which hybridizes to the primer binding site located at the 5' end of the viral genome (Hu, et al.,2012). Using the viral the RNA genome as template, RT copies viral RNA and generates a RNA/DNA hybrid. The RNA/DNA hybrids can be degraded by RNase H leaving the DNA strand (Gao, et al.,2001, Nowotny, et al.,2007). Finally, a double-strand DNA is produced.

5.4 Integration

After completion of the reverse transcription of the viral RNA into its complementary DNA in the cell cytoplasm, integrase binds to the LTR regions of the newly formed viral DNA to generate the pre-integration complex with other host proteins and cleaves at the 3' end of each viral DNA, thereby generating a reactive intermediate with a 3' hydroxyl group (Delelis, et al.,2008). The pre-integration complex is translocated into the nucleus of the host cells via nuclear pore complexes (Piller, et al.,2003). The host protein lens epithelium-derived growth factor/p75 (LEDGF/p75) is believed to tether the pre-integration complex and the host DNA (Poeschla,2008). The 3' hydroxyl groups of the viral DNA attack a selected palindromic region of the viral DNA in a staggered fashion and join with the host DNA 5' phosphate ends. The unpaired region between viral DNA and host DNA is repaired by host enzymes, such as polymerase, nuclease and ligase (Yoder, et al.,2000).

5.5 Viral RNA and protein production

After integration into the host genome, the viral DNA is transcribed into mRNA. Some of these mRNAs are fully spliced into smaller pieces, exported to the cell cytoplasm, and finally translated into Tat, Rev and Nef proteins (Karn, et al.,2012). Tat and Rev are transported between from the cytoplasm and nucleus rapidly. With the combination of Tat and Rev, the incompletely spliced and unspliced mRNAs in the nucleus are exported into the cytoplasm, where the former have the potential to express Env, Vif, Vpu and Vpr and the latter are able to produce Gag and Gag-Pol precursor proteins or be packaged into new virions as genomic RNA (Stoltzfus,2009).

5.6 HIV assembly

The assembly of new HIV virions occurs at the host cell membrane (Figure 19), mainly driven by the Gag protein (Meng, et al.,2013). The Gag protein is composed of the matrix (MA) domain, the capsid (CA) domain, the nucleocapsid (NC) domain, the P6 protein domain and two spacer peptide domains (SP1 and SP2) (Henderson, et al.,1992). The CA, SP1 and NC domains facilitate the formation of Gag-Gag contacts, while the main interaction (I) domain is located at the NC domain (Sandefur, et al.,2000). The uncleaved Gag polyprotein is able to capture the viral genomic RNAs by its NC domain, while the viral genomic RNAs also play a structural scaffolding role in Gag dimer formation (Burniston, et al.,1999). The Gag protein associates with the inner viral membrane by insertion of the myristoyl group of its MA domain into the lipid bilayer of the cell membrane (Bryant, et al.,1990). The MA domain also binds to the acidic phospholipids on the cell membrane, especially phosphatidylinositol (4,5) diphosphate ($PI(4,5)P_2$) (Chukkapalli, et al.,2008), which is critical for directing the Gag proteins to the cell membrane. The Gag-Gag and Gag-Pol proteins finally form a layer at the inner side of the cell membrane and become a spherical particle. The P6 domain is able to recruit other viral constituents such as Vpr, Vif and Nef proteins (Gottlinger, et al.,1991), which are also contained in the new spherical particles. The viral envelope glycoprotein gp160 translocates to the cell membrane by anchoring its transmembrane domain gp41 to the cell membrane, which is independent of the Gag protein (Checkley, et al.,2011). In addition, other small cellular RNAs, including $tRNA^{Lys3}$, 7SL, and other rRNAs (Cen, et al.,2002, Keene, et al.,2010), are packaged in the new virions.

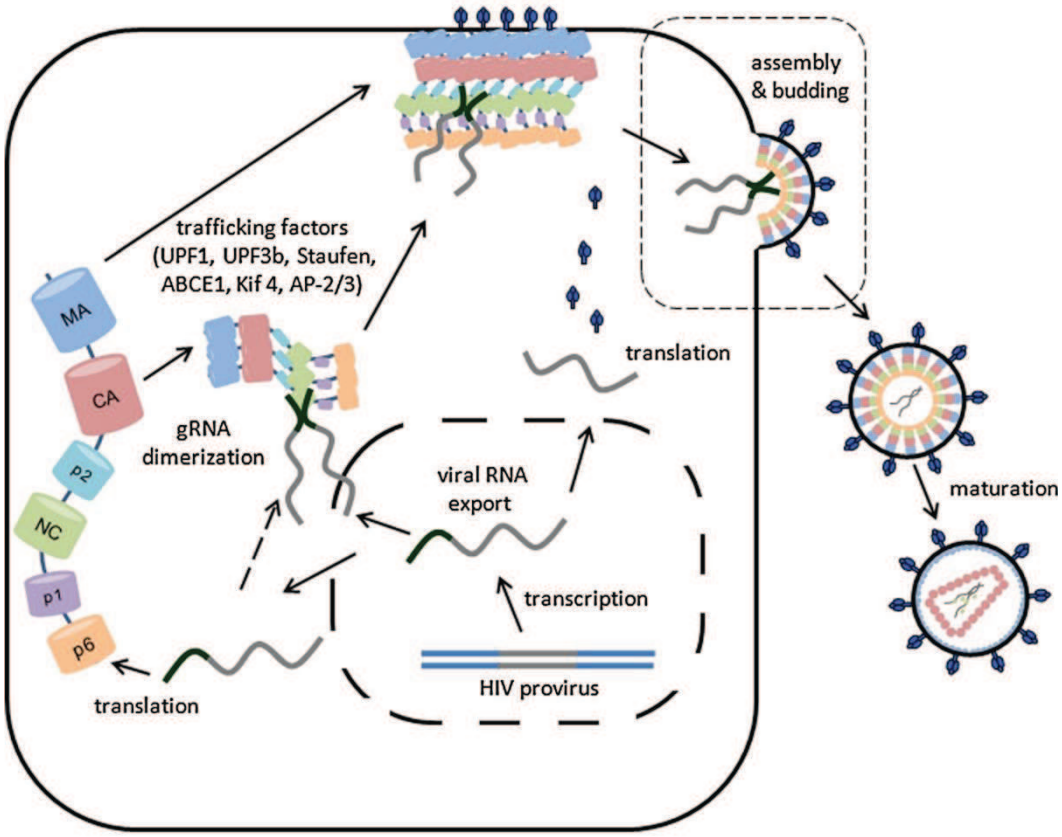


Figure 19. HIV assembly (Meng, et al.,2013).

5.7 HIV budding

After the new virion is assembled and forms a spherical particle on the cell membrane, it will bud from the membrane carrying along a portion of lipid bilayer. This step does not occur spontaneously. Several studies have shown that when certain domains are mutated, the new virion is still able to be assembled but remains tethered at the cell membrane (Gottlinger, et al.,1991, Karetnikov, et al.,2010). These domains are called viral late (L) domains, including the PTAP and YPXnL motifs (Demirov, et al.,2002, Dilley, et al.,2010).

HIV budding is mainly mediated by the endosomal sorting complexes required for transport (ESCRT) pathways (Figure 20) (Meng, et al.,2013). The ESCRT factors are able to facilitate membrane deformation and fission from within the necks of thin, cytoplasmic membrane vesicles and tubules (McCullough, et al.,2013). The ESCRT pathway contains five heterooligomeric proteins – ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-III and VPS4, and some accessory proteins such as ALIX (Henne, et al.,2011). ESCRT-0 is composed of the signal transducing adaptor molecule (STAM) and the hepatocyte growth factor – regulated tyrosine kinase substrate (HRS) (Kojima, et

al.,2014). HRS is able to facilitate Vpu-induced tetherin downregulation and degradation, thus enhancing HIV release (Janvier, et al.,2011). ESCRT-I contains TSG101, VPS28, VPS37 and MVB12, while TSG101 and VPS28 have been reported to be crucial for virion budding (Martin-Serrano, et al.,2003). ESCRT-II contains Vps25/EAP20, Vps22/EAP30 and Vps36/EAP45, which are recruited and co-localize with ESCRT-I on the Gag assembly site (Langelier, et al.,2006). The P6 domain of the Gag protein is able to interact with TSG101 by its PTAP motif (Demirov, et al.,2002). ESCRT-III includes 12 members of the charged multi-vesicular protein (CHMP) family, which is able to bind to the ESCRT-II/ESCRT-I complex via CHMP6 (Morita, et al.,2011). ESCRT-III is believed to mediate membrane scission getting energy either from VPS4 depolymerization or ESCRT-III filament procession. ALIX has been also reported to be utilized for virion budding by recruitment of NEDD4-1 ubiquitin ligase and by interaction between its V domain and the YPXnL motif of the P6 domain of the Gag protein (Sette, et al.,2010).

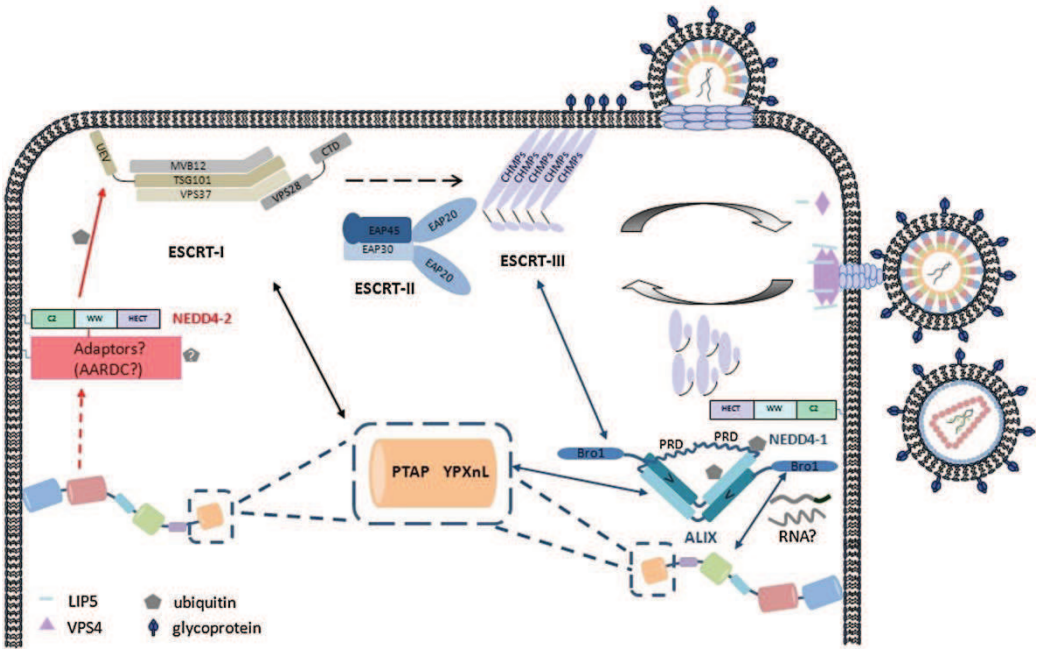


Figure 20. HIV budding via ESCRT pathway (Meng, et al.,2013).

5.8 HIV maturation

HIV maturation starts with the budding process. It is initiated by a proteolytic cleavage of the Gag-Gag and Gag-Pol polyproteins, which generate the matrix and the capsid proteins, the nucleocapsid, the P6 protein, the protease, the reverse transcriptase and the integrase (Konnyu, et al.,2013). These proteins are rearranged dramatically to create a mature virion, which is able to infect other cells (Freed,2015).

6. HIV-1 tropism

HIV-1 tropism is mainly determined by the target cell and co-receptor usage. HIV-1 can infect a variety of immune cells such as $CD4^+$ T cells and macrophages. Early HIV-1 isolates were only able to be replicated in activated peripheral blood T-lymphocytes and monocyte-derived macrophages, but not in T cell lines (Roos, et al.,1992). According to this observation, viral isolates were classified into macrophage-tropic (M tropic) or T cell line-tropic (T tropic), or dual tropic (D tropic). HIV-1 enters cells mainly using co-receptor CCR5, CXCR4, or both (Clapham, et al.,2001). Based on co-receptor usage, HIV-1 isolates can be classified as R5 tropic, X4 tropic and dual tropic (R5X4 tropic). Generally two HIV-1 tropism classification systems can be co-related. It has been shown that *in vitro*, R5 viruses could infect primary lymphocytes and macrophages, while X4 viruses could infect T cell lines. But there is a controversy: some X4 isolates were shown to be able to infect macrophages as well (Jayakumar, et al.,2005). *In vivo*, HIV-1 R5 viruses cause most HIV transmission, but HIV-1 X4 viruses drive to AIDS more rapidly (Schuitemaker, et al.,1992, Locher, et al.,2005). In addition, HIV-1 tropism is not constant. During the disease procession, HIV-1 may switch to other tropisms (Connor, et al.,1997, Ribeiro, et al.,2006, Sarmati, et al.,2010) and broaden their targets.

HIV-1 tropism can be assessed by either phenotypic or genotypic methods. In phenotype assays, a recombinant virus containing viral isolate-derived envelope sequences is used to infect CCR5 or CXCR4 reporter cell line (Mulinge, et al.,2013), and thus distinguishes viral tropism by its co-receptor usage. In the genotypic method, the gp120 V3 loop is the major determinant of HIV tropism (Obermeier, et al.,2012). According to the sequence of the gp120 V3 loop, co-receptor usage can be predicted. Generally, the V3 region of R5 viruses has relatively lower numbers of positively charged amino acid residues, while the V3 region which uses CXCR4 has a surplus of positively charged amino acid residues (Coetzer, et al.,2006). The V3 loop of dual tropic viruses is similar to that of X4 tropic, and other regions of the viral envelope are required as well to predict dual tropic viruses.

7. Anti-HIV drugs

7.1 Clinical antiretroviral therapy

Since HIV was identified, more than 30 drugs have been developed for antiretroviral therapy (ART) (Arts, et al.,2012). Antiretroviral drugs are designed to inhibit different steps of the HIV replication cycle (Figure 21) (Sierra-Aragon, et al.,2012). According to different drug targets, antiretroviral drugs approved by the Food and Drug Administration of the United States (FDA) are mainly classified into nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors, integrase inhibitors (INIs) and co-receptor inhibitors. Since the 1990s, antiretroviral drugs were first used as monotherapy, and then in dual therapy, and in multi-therapy today. Antiretroviral therapies have significantly reduced the morbidity and mortality of HIV infection (Quinn,2008).

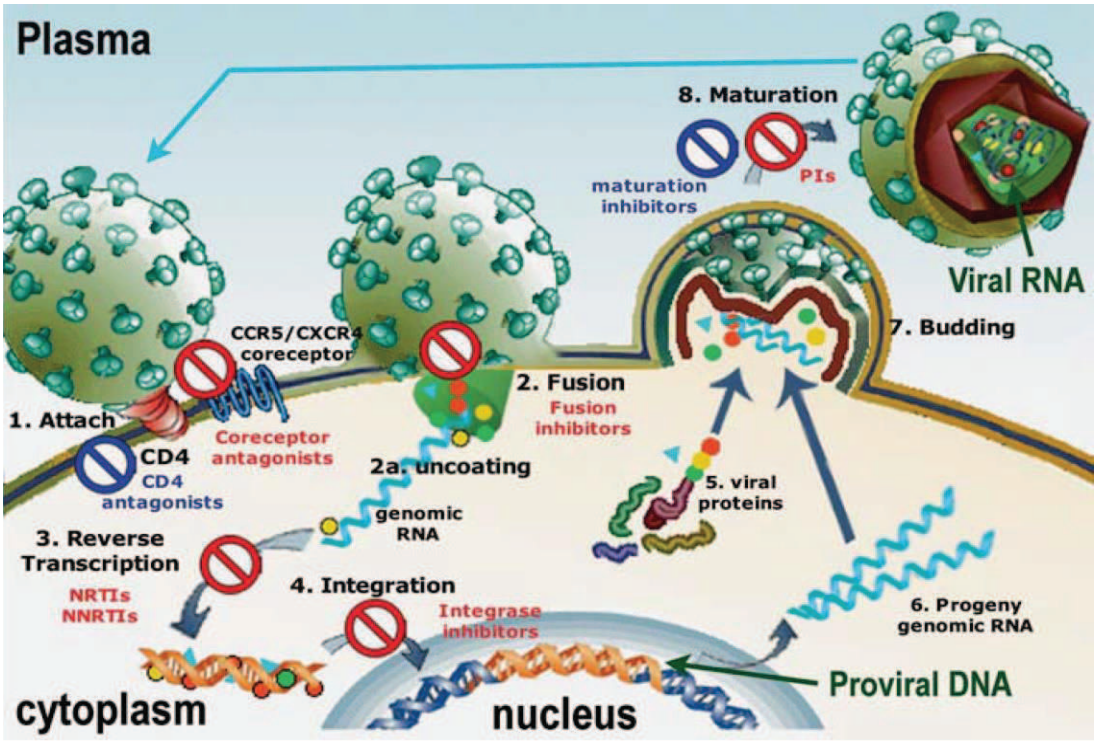


Figure 21. HIV-1 life cycle and the antiretroviral drug intervention points (Sierra-Aragon, et al.,2012).

7.1.1 Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs)

NRTIs are analogues of natural nucleosides and nucleotides. They are administrated as prodrugs and act as competitive substrate of reverse transcriptase (RT) during HIV replication after phosphorylation by cellular enzymes (Mitsuya, et al.,1985, Hart, et al.,1992). NRTIs lack a 3'-hydroxyl group at the sugar ring, resulting in chain termination during DNA synthesis by halting 5'-3' phosphodiester bond formation. Noteworthy, ZDV/AZT was the first antiretroviral drug approved by FDA in 1987 (Young,1988). Up to now, eight NRTIs have (ever) been approved by FDA. Seven of them are still being used in clinic, including abacavir (ABC, Ziagen), didanosine (ddI, Videx), emtricitabline (FTC, Emtriva), lamivudine (3TC, Epivir), stavudine (d4T, Zerit), tenofovir (TDF, Viread) and Zidovudine (ZDV, Retrovir)/azidothymidine (AZT, Retrovir) (Table 1). Zalcitabine (ddC, Hivid) was approved in 1992 and its production was discontinued by the manufacturer in 2006 (*www.fda.gov*).

Drug resistance develops quickly under NRTIs therapy. Accumulated mutations in the coding gene of RT (central region of *pol* gene) lead to the change of amino acids, and result in the selection of HIV strains resistant to NRTIs. It has been reported that mutations such as M184V/I, K65R/E/N, L74V and Q151M are able to decrease the affinity of RT for NRTIs in favour of promoting interaction between RT and natural nucleosides and nucleotides (Quan, et al.,1996, White, et al.,2006, Fourati, et al.,2013, Iyidogan, et al.,2014). Thymidine analogue mutations (TAMs) including 41L, T215Y, K70R, D67N and K219Q/E are known to be mainly selected by ZDV and d4T and involved in the resistance to most NRTIs, which are able to remove incorporated NRTIs from the DNA chain by phosphorolysis (Cozzi-Lepri, et al.,2005, Yahi, et al.,2005, Praparattanapan, et al.,2012).

7.1.2 Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

NNRTIs target reverse transcriptase (RT) by binding to a non-active site of the enzyme and do not act as competitive substrate inhibitors (Ivetac, et al.,2009). NNRTIs are able to inhibit the polymerase activity by leading a conformational change of p66 subunit of RT where the polymerase active site is located (Sluis-Cremer, et al.,2004). Up to now, five NNRTIs have been approved by FDA (Table 2), including delavirdine (DLV, Rescriptor), efavirenz (EFV, Sustiva) and nevirapine (NVP, Viramune), as well as the second generation NNRTIs etravirine (ETR, Intelence) and rilpivitrine (RPV, Edurant). All of them are still in use clinically.

Table 1. Nucleoside/nucleotide reverse transcriptase inhibitors approved by FDA

Generic Name (Other names and acronyms)	Brand Name	FDA Approval Date
zidovudine	Retrovir	March 19, 1987
(azidothymidine, AZT, ZDV)		
didanosine	Videx	October 9, 1991
(delayed-release didanosine, dideoxyinosine, enteric-coated didanosine, ddI, ddI EC)	Videx EC	October 31, 2000
Zalcitabine*	Hivid	June 19, 1992
(dideoxycytidine, ddC)		
stavudine	Zerit	June 24, 1994
(d4T)		
lamivudine	Epivir	November 17, 1995
(3TC)		
abacavir	Ziagen	December 17, 1998
(abacavir sulfate, ABC)		
tenofovir disoproxil fumarate	Viread	October 26, 2001
(tenofovir DF, TDF)		
emtricitabine	Emtriva	July 2, 2003
(FTC)		

* discontinued by manufacturer in 2006

Modified from https://aidsinfo.nih.gov/, accessed on 15/07/2015

Table 2. Non-nucleoside reverse transcriptase inhibitors approved by FDA

Generic Name (Other names and acronyms)	Brand Name	FDA Approval Date
nevirapine	Viramune	June 21, 1996
(extended-release nevirapine, NVP)	Viramune XR (extended release)	March 25, 2011
delavirdine	Rescriptor	April 4, 1997
(delavirdine mesylate, DLV)		
efavirenz	Sustiva	September 17, 1998
(EFV)		
etravirine	Intelence	January 18, 2008
(ETR)		
rilpivirine	Edurant	May 20, 2011
(rilpivirine hydrochloride, RPV)		

Modified from https://aidsinfo.nih.gov/, accessed on 15/07/2015

NNRTIs are able to develop drug resistance by a single mutation in the aa98-108 and aa178-190 regions, such as frequent mutations at A98, L100, K101, K103, V106, V108, E138, V179, Y181, Y188 and G190 (Tantillo, et al.,1994, Menendez-Arias,2010, Brumme, et al.,2013). Mutations in the aa225-238 region, such as point mutations at P225, P236, F227, M230, Y232, K238, Y318, are normally less frequent and lead to drug resistance in association with other mutations (Harrigan, et al.,2002, De Clercq,2004, Ren, et al.,2008). Cross resistance happened among NNRTIs, resulting in treatment failure (Antinori, et al.,2002).

7.1.3 Protease inhibitors (PIs)

PIs inhibit HIV protease activity by mimicking the substrates. PIs competitively bind to the catalytic site of HIV protease, especially two Asp at position 25 of each subunit (Lv, et al.,2015). Up to now, ten PIs have been approved by FDA, including atazanavir (ATV, Reyataz), amprenavir (APV, Agenerase), fosamprenavir (FPV, Telzir), Darunavir (DRV, Prezista), saquinavir (SQV, Invirase), ritonavir (RTV, Norvir), lopinavir/ritonavir (LPV/r, Kaletra), indinavir (IDV, Crixivan), nelfinavir (NFV, Viracept) and tipranavir (TPV, Aptivus) (Table 3). Among these PIs, APV has been further developed into FPV and RTV. RTV is further used as a pharmacokinetic enhancer (ritonavir) because it can inhibit the activity of the metabolizing enzyme cytochrome P450 CYP3A4 (Oldfield, et al.,2006), and thus boosts the circulating concentration of the antiretroviral drugs.

PIs have been reported to represent high generic barriers to develop drug resistance (Luber,2005). Accumulated mutations at substrate-binding sites are able to lead to the reduction of PI incorporation by changing the conformation of HIV protease (Tie, et al.,2005). Mutations such as D30N, V32I, L33F, M46I/L, I47A/V, G48V/M, I50V/L, I54V/T/A/L/M, L76V, V82A/T/F/S, I84V, N88D/S and L90M have been revealed to be crucial in clinic (Johnson, et al.,2006, Fun, et al.,2011). Other mutations at the HIV protease cleavage sites of the *gag* and *pol* genes are associated with PI resistance as well (Cote, et al.,2001, Dam, et al.,2009).

Table 3. Protease inhibitors approved by FDA.

Generic Name (Other names and acronyms)	Brand Name	FDA Approval Date
saquinavir	Invirase	December 6, 1995
(saquinavir mesylate, SQV)		
ritonavir	Norvir	March 1, 1996
(RTV)		
indinavir	Crixivan	March 13, 1996
(indinavir sulfate, IDV)		
nelfinavir	Viracept	March 14, 1997
(nelfinavir mesylate, NFV)		
amprenavir	Agenerase	April 15, 1999
(APV)		
Lopinavir/ritonavir	Kaletra	September 15, 2000
(LPV/r)		
atazanavir	Reyataz	June 20, 2003
(atazanavir sulfate, ATV)		
fosamprenavir	Lexiva	October 20, 2003
(fosamprenavir calcium, FOS-APV, FPV)		
tipranavir	Aptivus	June 22, 2005
(TPV)		
darunavir	Prezista	June 23, 2006
(darunavir ethanolate, DRV)		

Modified from https://aidsinfo.nih.gov/, accessed on 15/07/2015

7.1.4 Integrase inhibitors (INIs)

INIs, also known as strand transfer inhibitors, bind to the catalytic site of HIV integrase and reduce proviral DNA incorporation (McColl, et al.,2010). So far, only three INIs have been approved by FDA: raltegravir (RAL, Isentress), dolutegravir (DTG, Tivicay) and elvitegravir (EVG, Vitekta) (Table 4). All three INIs contain two main active components, a metal-binding pharmacophore and a hydrophobic group, which allow INIs interact with magnesium ions and the preintegration complex (Craigie,2001, Grobler, et al.,2002).

Drug resistance of RAL and EVG can be due to either a single mutation or accumulated mutations (Mesplede, et al.,2012). Mutations such as T66A/I/K, E92Q, E138K/A, G140S/A, Y143R/C/H, S147G, Q148H/R/K and N155H have been observed in clinic (Hu, et al.,2010, Blanco, et al.,2011, Garrido, et al.,2011, Abram, et al.,2013). DTG has

a much higher genetic barrier, which has not shown any resistance given as first-line therapy (Mesplede, et al.,2014).

Table 4. Integrase inhibitors approved by FDA

Generic Name (Other names and acronyms)	Brand Name	FDA Approval Date
raltegravir (raltegravir potassium, RAL)	Isentress	October 12, 2007
dolutegravir (DTG)	Tivicay	August 13, 2013
elvitegravir (EVG)	Vitekta	September 24, 2014

Modified from https://aidsinfo.nih.gov/, accessed on 15/07/2015

7.1.5 Co-receptor inhibitors

Co-receptor inhibitors are designed to target co-receptors CCR5/CXCR4 which are crucial for HIV entry. Maraviroc (MVC, Selzentry/Celsentri) is the only co-receptor inhibitor in use clinically, approved by FDA on August 6, 2007. MVC is a CCR5 antagonist, binding to the hydrophobic transmembrane pocket (Dorr, et al.,2005). Although MVC does not bind to the sites required for CCR5 agonists and viral envelope, it is able to prevent viral envelope interaction with CCR5 and stop HIV entry by altering the conformation of CCR5 (Van Der Ryst,2015).

Resistance to MVC has been observed in patients taking long-term treatment but rarely in naïve patients (Seclen, et al.,2010, Roche, et al.,2013). Mutations in the V3 loop of gp120 are capable to alter co-receptor usage (Hu, et al.,2000, Platt, et al.,2001, Garcia-Perez, et al.,2015) but cannot be used to predict the primary resistance (Delobel, et al.,2013). To adapt, some resistant strains are also able to use MVC-bound receptors for HIV entry (Roche, et al.,2011). So far, no signature mutation has been approved that can predict MVC resistance.

AMD3100 is a CXCR4 antagonist, but has failed in clinical trials due to its severe side effects (Donzella, et al.,1998, Hendrix, et al.,2004).

7.1.6 Fusion inhibitors

So far, enfuvirtide (T-20, Fuzeon) approved by FDA on March 13, 2003 is the only one fusion inhibitor used in clinics (Fung, et al.,2004). It is a 36-amino acid peptide, directly derived from the HIV-1 HXB2 strain. It is an analogue peptide of the heptad repeat region HR2 of envelope gp41 protein, which is able to stop fusion by competitively binding to HR1 and halting the formation of the six helix bundle (Gallo, et al.,2001, Moore, et al.,2003). Mutations in the HR1 of gp41 have revealed to be resistant to T-20, including G36D/S/V, I37T, V38A/E/M, Q40H, N42T/D, N43K/D and L45M (Greenberg, et al.,2004, Lu, et al.,2006).

7.1.7 Antiretroviral combination drugs

Since antiretroviral monotherapy started in the 1990s, a selection of HIV resistant strains was observed subsequently. Antiretroviral drug combinations were then developed to increase the effect and overcome the resistance. The highly active antiretroviral therapy (HAART) was one of the successes of HIV infection treatment (Quinn,2008) and is now termed cART (combination antiretroviral therapy). Recently, in order to overcome the high pill burden and improve the drug inherence, several new one-pill fix-dosed combinations have been licensed into clinics, such as Stribild (EVG/COBI/FTC/TDF) and Triumeq (DTG/ABC/3TC) (Greig, et al.,2015, Murrell, et al.,2015) including INSTI.

The recent guidelines of the Department of Health and Human Services of US 2015, recommends now that an antiretroviral regimen generally consists of two NRTIs (one of which is FTC or 3TC) plus an INSTI, NNRTI, or a pharmacokinetic enhancer (cobicistat or ritonavir) boosted PI (DHHS,2015). The new version of European antiretroviral guideline is going to be announced soon by the European AIDS Clinical Society at the 15th European AIDS Conference on October 21 of 2015.

7.2 Other classes of potential antiretroviral agents

7.2.1 Attachment inhibitors (AIs)

HIV attachment inhibitors are designed to inhibit the binding of the HIV gp120 proteins to the CD4 on the cell surface. BMS-626529 and its phosphonooxymethyl ester

prodrug BMS-663068 are the first generation AIs (Zhu, et al.,2015). BMS-626529 is proposed to bind to the conserved outer domain of unliganded HIV-1 gp120 proteins which is under the antiparallel β20-β21 sheet and close to the CD4 binding site, and thereby block the CD4-gp120 interaction (Langley, et al.,2015). *In vitro* and *in vivo* studies have shown the antiretroviral activity of both BMS-663068 and BMS-626529 on a wide range of HIV-1 strains, as well as its efficiency on the combination of ATV/r or of TDF and RAL (Brinson, et al.,2014). Mutations at the amino acid sites such as S375M, M426L, M434I and M475I of gp120 are mostly linked to the reductions in susceptibility to BMS-626529 (Zhou, et al.,2014).

Ibalizumab (TNX-355) is a humanized IgG4 monoclonal antibody which can bind to the second domain of CD4 (Burkly, et al.,1992). Unlike the common AIs, it is believed to reduce the flexibility of CD4 and thus inhibit CD4-bound gp120 to interact with co-receptors CCR5 or CXCR4, instead of blocking CD4-gp120 binding. It has been shown that the viral susceptibility to ibalizumab is associated with the variable region 5 of the gp120 proteins (Pace, et al.,2013a).

7.2.2 Viral gene inhibitors

ABX464 is the first member of a new class of antiretroviral agents, targeting viral gene expression. ABX464 inhibits HIV replication both *in vitro* and in humanized mouse models (Campos, et al.,2015). It binds to the cellular cap binding complex (CBC) regulating viral mRNA splicing and transport from the nucleus to the cytoplasm which are generally regulated by the Rev protein in cells (Berkhout, et al.,2015). A phase I study has shown that ABX464 was well tolerated up to 150mg and did not induce any serious adverse event and clinically abnormal result (Campos, et al.,2015a). A phase II study has been just initiated in 2015. Compared to other antiretroviral drugs which target viral components, ABX464 is believed to be less potent to develop drug resistance.

7.2.3 Maturation inhibitors

HIV-1 maturation inhibitors (MIs) are a new class of inhibitors that target the final step of the viral replication cycle. Bevirimat (BVM), also known as PA-457 or MPC-4326, is the first generation of MIs, initially isolated from *Syzygium claviflorum* (Martin, et al.,2008). This compound inhibits the proteolytic process of Gag by blocking the cleavage between the capsid (CA, p24) and the spacer peptide 1 (SP1). It has been

shown that polymorphisms of SP1 residues 6-8 are crucial for BVM susceptibility (Adamson, et al.,2010). About half of the patients receiving BVM did not respond in a phase 2 clinical trial (McCallister S, et al.,2008). BMS-955176 and GSK2838232 are second generation MIs under clinical development (Olender, et al.,2015). Both target the CA-SP1 cleavage site and have potentials to inhibit BVM resistant viruses.

7.2.4 Broadly neutralizing antibodies

A number of broadly neutralizing antibodies of HIV-1 able to inhibit HIV infection by blocking the HIV entry process have been identified. The monoclonal antibodies (mAbs) PG9 and PG16 bind to the V1/V2 loop of the gp120 proteins, while PGT128 can target the V3 loop (Pejchal, et al.,2011). PG9 and PG16 have been tested as bispecific antibodies PG9-ibalizumab and PG16-ibalizumab *in vitro* and showed great potency to restrict viral infection (Pace, et al.,2013b). The mAbs b12 (Saphire, et al.,2007), VRC01 and 3BNC117 (Scheid, et al.,2011) target the CD4-binding site. VRC01 and its related clone VRC07 combined with PGT121 have shown to inhibit viral infection on rhesus macaques (Diane L. Bolton, et al.,2015). 3BNC117 has shown to inhibit HIV infection in clinics and be well tolerated (Caskey, et al.,2015). The mAbs 2F5, 4E10 and 10E8 are believed to interact with the membrane-proximal external region of the gp41 proteins (Nelson, et al.,2007, Huang, et al.,2012).

7.2.5 Broad-spectrum antiretroviral agents

Compared to other antiretroviral agents targeting viral proteins, antiretroviral agents targeting less variable host cell factors such as the lipid membrane, are believed to have a broad antiretroviral activity and less potential to develop drug resistance. Recently, rigid amphipathic fusion inhibitors (RAFIs) and LJ001 have been investigated *in vitro* representing a new class of broad spectrum antivirals against enveloped viruses. RAFIs contain relatively larger hydrophilic heads and small hydrophobic tails, which are able to inhibit the infection of enveloped viruses by regulating lipid membrane curvature (St Vincent, et al.,2010). LJ001 has been reported to have an activity similar to RAFIs, inhibiting infection of a great range of enveloped viruses including HIV, hepatitis C virus (HCV), vesicular stomatitis virus (VSV), influenza virus, Ebola virus, Marburg virus and Yellow Fever viruses by conferring a positive curvature to the membrane (Wolf, et al.,2010).

7.3 Anti-HIV natural products

Natural products are very good candidates for new drug development, due to their rich resources, low toxicity to organisms, low drug resistance and low cost. Natural products have been used for thousands of years in the world, especially in Asian and African countries which have a long history of traditional medicines. Since the 19th century, when organic chemistry was developed, single chemical compounds started to play a crucial role in modern medicine. Morphine was first isolated from *Papaver somniferum* by Friedrich Sertürnerwhich in 1804 as the first active component obtained from an herb, and more and more compounds have been then isolated from natural resources and their bioactivities deciphered against various diseases. Recently, many natural products including alkaloids, saccharides, coumarins, flavonoids, lignans, tannins, terpenes and proteins have been discovered to inhibit HIV infection.

7.3.1 Alkaloids

Alkaloids are secondary metabolites of amino acids mainly existing in plants, bacteria and fungi. Some of them can inhibit HIV infection by acting on various targets. For instance, didehydro-Cortistatin A, an analogue of the natural alkaloid Cortistatin A isolated from the marine sponge *Corticium simplex*, has been reported to inhibit HIV infection by targeting the Tat proteins (Mousseau, et al.,2012). It can specifically bind to the trans-activation-responsive element binding domain of the Tat proteins, thereby blocking HIV viral RNA production. The half-maximal effective concentration (EC_{50}) was extremely low, less than 0.1 nM, while the half-maximal cytotoxic concentration (CC_{50}) was around 100 μM. Other alkaloids, such as polycitone A isolated from *Polycitor sp.* (Loya, et al.,1999), and Michellamines A and B isolated from Ancistrocladus species (Boyd, et al.,1994), inhibit HIV by targeting reverse transcriptase; Lindechunine A isolated from roots of *Lindera chunii Merr.* targets the integrase (Zhang, et al.,2002).

7.3.2 Saccharides

Saccharides, especially sulphated polysaccharides such as rhamnan sulfate isolated from *Monostromo latissimum*, have been known for their antiviral activity for a long time (Lee, et al.,1999). It was suggested that they inhibit HIV infection by targeting not only viral adsorption but also a late stage of the HIV replication cycle. In a collaboration

with other groups in Viet Nam and Russia, we also observed anti-HIV activity of sulfated polysaccharides: indeed, we showed that fucoidans, derived from marine brown algae, inhibited HIV infection with IC_{50} ranging from 0.33 to 0.7 µg/ml, by blocking HIV entry into cells (Thuy, et al.,2015).

7.3.3 Coumarins

Coumarins have a basic benzopyrone structure, with a variety of substitutions. Coumarins and their derivates have been reported to inhibit HIV infection at different HIV replication steps, including integration, reverse transcription and proteolysis. A modified coumarin, 3,3'-(2-naphthalenomethylene)-bis-(4,7-dihydroxycoumarin) has been described to have a better effect than its parent compounds inhibiting HIV integrase with IC_{50} at 4.2 µM (Zhao, et al.,1997). Calanolide A isolated from *Calophyllum lanigerum*, is known to inhibit HIV infection as a non-nucleoside reverse transcriptase inhibitor (Kashman, et al.,1992). Noteworthy, Tipranavir (TPV, Aptivus), is a synthetic coumarin--based protease inhibitor which was approved by US FDA in 2005 (King, et al.,2006).

7.3.4 Flavonoids

Flavonoids are a wide class of secondary metabolites of plants. Epigallocatechin gallate, which is the main polyphenol in green tea, is able to bind to CD4 and thus inhibit the interaction between gp120 and CD4 (Williamson, et al.,2006). Baicalin isolated from *Scutellaria baicalensis* has been reported to decrease the replication of HIV by blocking HIV entry but not affecting gp120-CD4 interaction (Li, et al.,2000). Taxifolin, isolated from *Juglans mandshurica*, protected MT-4 cells against HIV infection with a 100% inhibitory concentration of 25 µg/ml (Min, et al.,2002).

7.3.5 Lignans

Several lignans showing anti-HIV activity have been isolated from *Schisandra propinqua var. sinensis* (Li, et al.,2009) and *Schisandra lancifolia* (Yang, et al.,2010). Among them, Tiegusanin G was the most active, inhibiting HIV infection with an IC_{50} of 7.9 µM. Dibenzocyclooctadine lignan and its derivatives inhibited various HIV strains including wild-type HIV-1, HIV-2 and drug resistant HIV-1 (Han, et al.,2015). They were shown to target HIV reverse transcriptase.

7.3.6 Tannins

Tannins are known to inhibit HIV infection by targeting reverse transcriptase, the protease and the fusion peptide gp41. Tetragalloylquinic acids, punicalin and punicacortcin were described as the first identified anti-HIV tannins targeting reverse transcriptase (Nonaka, et al.,1990). Three ellagitannins containing a hexahydroxydiphenoyl unit linked to the O-3 and O-6 positions of the sugar strongly inhibited HIV protease (Xu, et al.,2000). Tannins from different sources inhibited HIV infection by interfering with the protein gp41 and blocking six-helix bundle formation (Lu, et al.,2004).

7.3.7 Terpenes

Terpenes, especially the triterpenes and their derivatives, are known to have various bioactivities. Betulinic acid, oleanolic acid and ursolic acid are the typical compounds representing three different types of triterpenes: lupane, oleanane and ursane. Betulinic acid and its derivatives were reported to inhibit HIV infection either by targeting the protein gp41 or by disrupting viral assembly and budding (Cichewicz, et al.,2004). Bevirimat (BVM, PA-457 or MPC-4326), first isolated from *Syzygium claviflorum*, is a betulinic acid-like compound and a first generation HIV maturation inhibitor (Martin, et al.,2008). It inhibits the proteolytic process of Gag by blocking the cleavage between the capsid (CA, p24) and the spacer peptide 1 (SP1). Oleanolic acid was described to inhibit HIV infection on PBMCs due to its effect on HIV protease (Mengoni, et al.,2002). The 3-O-diglycoryl-ursolic acid inhibited HIV infection with an IC_{50} of about 0.31 μM (Kashiwada, et al.,2000).

7.3.8 Proteins

Lectins are carbohydrate-binding proteins, which can be isolated from various sources including plants, bacteria and worms. Griffithsin is a marine plant protein isolated from *Griffithsia*, known to inhibit HIV by interfering with the glycans on the surface of virions (Mori, et al.,2005). Other lectins such as cyanovirin-N isolated from *Nostoc ellipsosporum* (Esser, et al.,1999) and concanavalin A isolated from *Canavalia ensiformis* (Matsui, et al.,1990) have also been described to have anti-HIV activities.

Ribosome inactivating proteins (RIP) are a family of natural occurring proteins existing in plants, bacteria and fungi which can inhibit protein synthesis at the ribosome level

(Stirpe,2013). One member of this family, MAP30, was initially isolated from *Momordica charantia*, and has been reported to inhibit HIV-1 due to its ribosome inactivating activity and its inhibitory effect on integrase (Lee-Huang, et al.,1995).

In conclusion, a broad variety of natural products, especially plant derived natural products, have been described for their anti-HIV potentials and gave many ideas for new drug development. In the early 2000s, the Retrovirology Laboratory at CRP-Santé detected an anti-HIV activity in plant materials originally brought by an HIV positive African man from Congo to his doctor in Brussels. The identity of this plant remained unknown for about 10 years, until 2012 when I identified it using the DNA barcoding technology.

8. Species identification and DNA barcoding

The DNA barcoding technology is a molecular and bioinformatic research tool originally developed by Hebert *et al*. on insects in 2003 and progressively adapted to an increasing number of other living species including plants (Hebert, et al.,2003).

The basic concept of DNA barcoding is to use one or few short genetic markers in the DNA of an organism to identify it as belonging to a particular species. The general steps of DNA barcoding are first to build a DNA barcode library of known species with standard genetic markers, ideally the more genetic markers the better, and second, to compare the barcode sequence of the unknown sample to the DNA barcode library and identify it with the matching species.

Compared to the classical method using morphological knowledge, the DNA barcoding method has expanded its capacity of identification using all of the life history stages of an organism. For instance, DNA barcoding allows species identification not only via flowers but also via samples like seeds, seedlings, pieces of branches, stems and roots; animal species can be also identified via eggs or parts of their bodies. With the DNA barcoding method, microorganisms lacking morphological information can be identified as well. The DNA barcoding method allows even non-taxonomists to correctly and rapidly identify the samples.

8.1 DNA barcode pipeline

Figure 22 describes a general pipeline for DNA barcoding. It fits for all kinds of biological resources in some ways. The pipeline has four main components, which constitute the basis of DNA barcoding. They are the specimens, laboratory analysis, databases and data analysis.

8.1.1 Specimens

The basic components of the DNA barcoding pipeline are specimens. Since the capacity of identification using DNA barcoding method has been expanded, theoretically, all kinds of samples from an organism at any history

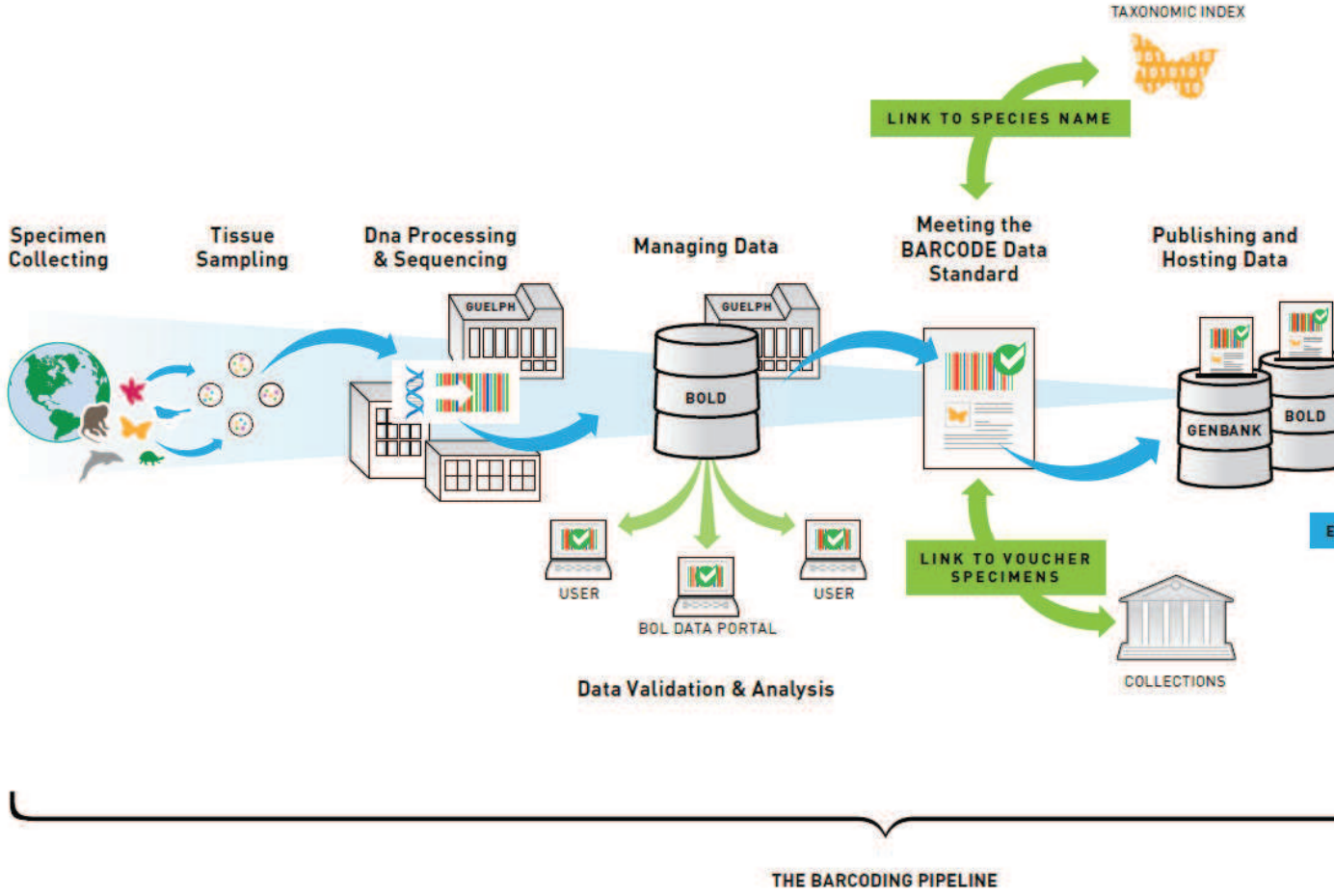


Figure 22. The barcoding pipeline

Accessed from *http://www.barcodeoflife.org/content/about/what-dna-barcoding*

step of life can be collected. The original specimens to fill the DNA barcode databases are mainly from natural history museums, herbaria, zoos, botanical gardens, seed banks and frozen tissues, all well identified by taxonomists. It is strongly recommended to record detail information of collections, including who collected the sample where and when, and which part of the organism the sample is from. To be well prepared for further studies, samples have to be stored in a proper way, kept clean and avoiding contaminations from other species.

8.1.2 Laboratory analysis

To acquire the barcode sequences of specimens, high quality DNA has to be obtained. Proper primers have to be designed and used. So far, the 5' end region of a mitochondrial gene cytochrome oxidase subunit 1 (*co1*) is suited for most metazoans, including mammals, fishes and insects (Hebert, et al.,2003), while the internal transcribed spacer (ITS) gene works fine with fungi (Schoch, et al.,2012) (Table 5). Due to the low rate of evolution on the *coxI* gene in the land plants, other barcoding regions are used. The Consortium for the Barcode of Life (CBOL) suggested matK and rbcL to be the standard loci for plant DNA barcoding (CBOL,2009). ITS2 (Chen, et al.,2010) and psbA-trnH (Kress, et al.,2009) have been shown to work in the study of land plants as well.

Table 5. List of the main taxa-specific DNA barcoding recommended by The Barcode of Life Data Systems (BOLD) (Casiraghi, et al.,2010)

Group	Campaign	Main marker	Reference
Metazoans	ACG Parasitoids	*coxI*	http://www.barcodinglife.org ('Campaigns' section)
	ABBI-all birds barcoding initiative	*coxI*	http://www.barcodingbirds.org
	All Caddis DNA barcoding	*coxI*	http://www.barcodinglife.org ('Campaigns' section)
	Ants of the world	*coxI*	http://www.barcodinglife.org ('Campaigns' section)
	Barcoding fish (FishBOL)	*coxI*	http://www.fishbol.org
	Barcoding mammals of the world	*coxI*	http://www.barcodinglife.org ('Campaigns' section)
	Bee barcoding initiative	*coxI*	http://www.barcodinglife.org ('Campaigns' section)
	Mosquitos of the world	*coxI*	http://www.lepbarcoding.org
Fungi	All fungi barcoding	ITS	http://www.allfungi.com
Plants	Plant DNA barcoding	*MatK* + *rbcL*	http://www.barcoding.si.edu/plant.working.group.html

8.1.3 Databases

A DNA barcode library is essential for species identification via DNA barcoding. To build up such a library, first, the efforts from taxonomic experts are required. The original sequences used for DNA barcodes are generated by taxonomic experts, who have to identify the species using morphological knowledge and import the sequence data into the databases.

So far, three biological sequence databases are most widely used: GenBank in the US (*http://www.ncbi.nlm.nih.gov/genbank/*); the European Nucleotide Achive of the European Molecular Biology Lab (ENA-EMBL, *http://www.ebi.ac.uk/ena*); and the DNA Data Bank of Japan (DDBJ, *http://www.ddbj.nig.ac.jp/*). All three databases collaborate with The Barcode of Life Data Systems (BOLD) (*http://www.boldsystems.org/*) and recommend a relative standard submission of new barcode sequences.

8.1.4 Data Analysis

The basic DNA barcoding data analysis consists in aligning each query sequence to a set of referenced sequences, which can be easily done online using the webpage of several databases. For instance, the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information (NCBI, US) (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*) and of the DNA Databank of Japan (DDBJ) (*http://blast.ddbj.nig.ac.jp/*) are the two webpages commonly used to identify a query sequence. Among different algorithms, the Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar,2004) is commonly used for alignment of multiple sequences, while the neighbour-joining method (Saitou, et al.,1987) is widely used to generate phylogenetic trees.

8.2 Commonly used barcode markers

Generally, a good barcode marker region should be easily amplified and sequenced in many different specimens. Thus, it must have a conserved part which allows it to be used as a universal PCR primer to amplify the target sequences. It also should contain a significant genetic variability and divergence that allows it to be clearly discriminated from that of other species. The sequences amplified from a good barcode marker should be of appropriate length, not too long (because not easy to sequence) and not too short (because lack of useful information).

8.2.1 *co1*

The *co1* gene codes for cytochrome oxidase subunit 1, which is an essential part of enzyme for the electron transport in mitochondria. The 5' end portion of *co1* has been recommended to be the universal barcode marker for animals. It fits all the criteria of a good barcode marker. It is highly abundant in the cells and relatively easy to amplify. The sequence length of this region is relatively stable among species, around 600 bp. Moreover, it has shown its discriminating power in different populations of animals, such as birds, butterflies, fish, mosquitos (Hebert, et al.,2003).

8.2.2 *matK*

The *matK* gene has been recommended as the common barcode marker for land plants by CBOL. It is a chloroplast-encoded gene, located in a large single copy region, between the two exons of the split tRNA-lysine gene (*trnK*) (Sugita, et al.,1985). It has an 83% amplification rate and discriminates well between gymnosperms but has a low amplification rate (less than 10%) in cryptogams (ferns, mosses, lichen, algae and fungi) (Li, et al.,2015).

8.2.3 *rbcL*

The *rbcL* gene has also been recommended to be one of the common barcode markers for land plants by CBOL, in combination with *matK*. It codes for ribulose-1,5,-bisphosphate carboxylase/oxygenase (RuBisCo), which is an important enzyme for carbon fixation in plants (Feller, et al.,2008). Compared to *matK*, *rbcL* has a slow evolutionary rate. It does not have a strong discriminatory power, but in combination with other barcode markers, such as *matK* and *psbA-trnH,* it can provide a relatively high accuracy of species identification (N, et al.,2015).

8.2.4 *psbA-trnH* spacer

The *psbA-trnH* spacer is a non-coding intergenic region on the chloroplast DNA (Degtjareva, et al.,2012). It shows a high divergence among species, with a relatively high frequency of insertions and deletions. The *psbA-trnH* gene has been widely used in land plant identifications and has shown its power of discrimination among various populations. Its length varies much among species, ranging from 100 bp to more than 1000 bp (Hollingsworth, et al.,2011). The short sequences cannot be used for identification due to the lack genetic information, but they can provide useful information for phylogenetic studies.

8.2.5 *ITS*

ITS refers to the internal transcribed spacer, which separates the coding regions of the ribosomal RNA genes. In eukaryotes, the *ITS* gene has two loci – ITS1 and ITS2. ITS1 is situated between 18S and 5.8S rRNA genes, while ITS2 sits between 5.8S and 25S in plants (28S in animals) rRNA genes (Lafontaine, et al.,2001). Currently, *ITS* is a common barcode marker for fungi. Also, due to its high amplification power in fungi (Schoch, et al.,2012), it is recommended by CBOL to be a supplementary locus for

land plants, in order to detect contaminations by fungal sequences. Recently, the China Plant BOL group has shown that the *ITS* region can discriminate seed plants better than other plastid barcodes (China Plant, et al.,2011). So far the *ITS* loci are not as widely used as *matK*, *rbcL*, but there is no doubt that they can be good barcode markers.

8.3 Expanded applications

DNA barcoding is relatively new as it was developed only a little more than 10 years ago. Since its initial application for biological species identification it has been applied in other fields, such as food traceability, herbal medicine control and species evolution.

8.3.1 Food traceability

Food quality and safety is always an important issue. To evaluate the quality and safety of food, the agencies of different goverments use different guidelines. The main techniques in use are immunological assays, electrophoretical, as well as chromatography-based approaches such as high performance liquid chromatograpy (HPLC) and thin-layer chromatograpy (TLC). Since its development the DNA barcoding technique has shown its power in biological species identification. Moreover, in 2008, it was proposed by the Food and Drug Administration of the United States for food traceability of fish-based commercial products. The DNA barcoding method has also been used to trace dairy products, edible plants and mixtures of food products (Rolla, et al.,2013).

8.3.2 Herbal medicine control

Traditional medicines have been used for a long time and their demand is still increasing. One of the main characters of traditional medicines is that they often

consist of a mixture of several species. Therefore, DNA barcoding is often used for the identification of raw materials in the mixtures (Techen, et al.,2014).

8.3.3 Species evolution

DNA barcoding has also been used in species evolutionary studies. A phylogenetic tree can be generated from the barcode sequences using different algorithms, estimating the evolutionary relationships among species. In plants, *matK* and the *psbA-trnH* spacer have a high evolving rate, and are commonly used in phylogenetic studies. This evolutionary information is also useful for ecological and societal studies (Kress, et al.,2015).

Chapter II

Results

Part 1. Identification of an African medicinal plant with anti-HIV activities by DNA barcoding and phylogenetic study of Cassia and Senna species

Yue ZHENG, Katrin NEUMANN, Andre STEINMETZ
Laboratory of Cellular and Molecular Oncology,
Luxembourg Institute of Heath,
Luxembourg

Abstract

Using the DNA barcoding technology on the nuclear ITS2 marker I have identified an African medicinal plant with anti-HIV activities as *Cassia abbreviata*. Three closely related sequences could be generated from three samples from three different sources and may correspond to the three geographically distinct subspecies identified so far (*Cassia abbreviata* subsp. *abbreviata*, subsp. *beareana* and subsp. *kassneri*). I have then studied the phylogenetic relationship of about 70 Senna and Cassia species using the nuclear marker ITS2 and the chloroplast marker trnH-psbA spacer. While the trnH-psbA spacer differs significantly in size (180-400 bp for the trimmed sequences analyzed) among the species considered, ITS2 shows relatively little size variation (277-319 bp). In the phylogenetic trees derived from the sequence alignments of the two markers the species divide into a clade comprising exclusively Cassia species and one Senna clade with five subclades. Typical for the Cassia species is a 35-67 bp "insertion" in the trnH-psbA spacer and a 9-13 bp "deletion" in the ITS2 locus. Sequence similarities between the species analyzed range from 41 to 100% in the trnH-psbA spacer and from 67 to 100% in the ITS2 locus. This wide range of sequence divergence not only allows to unambiguously discriminate between the vast majority of Cassia and Senna species; it also defines the two loci as excellent complementary barcode markers for plant identification.

Introduction

In the year 2000 an HIV-positive young man from the Democratic Republic of the Congo came to Brussels with small chunks of wood (Figure 23) from a local medicinal plant and claimed to his doctor that treating himself with a decoction from this plant had improved his condition. An extract from the wood was subsequently tested for anti-HIV activity in the Laboratory of Retrovirology at the Public Research Center for Health (now Luxembourg Institute of Health), in collaboration with the Laboratory of Toxicology of the National Health Laboratory (Luxembourg), and an antiviral activity was detected, together with a cytotoxic activity. In 2003 the young man died of AIDS without leaving further information regarding the identity of the plant. Since more material was needed for further studies and the species remained unknown, the research was discontinued (until 2011).

Figure 23. An example of a piece of African wood which had been analyzed for anti-HIV activity and which I used for identification.

My PhD project is a follow-up to the previous study. In this first part I will describe how I identified the plant using the DNA barcoding technology. This step involved the analysis of barcode markers of over 50 Cassia and Senna species of which we were able to purchase seeds online (*http://www.sunshine-seeds.de/* and *http://www.rareplants.de*). I will then compare the sequences of the two barcode markers (ITS2 and trnH-psbA spacer) which we determined for these species, together with publicly available sequences, in order to establish the phylogenic relationships between the Cassia and Senna species that I analyzed in the frame of the identification

study. In the second part I will describe my study on the anti-HIV activities present in an ethanol extract from the wood of *Cassia abbreviata*, as well as of 57 purified components.

With the recent development of the DNA barcoding technology (Hebert, et al.,2003) it has become possible to identify living species on the basis of selected adequate DNA markers using minute amounts of any tissue from an organism. While the mitochondrial cytochrome oxidase I (COI) gene is a commonly used marker for animal identification, several markers have been proposed in the case of plants. The latter include several chloroplast loci such as matK (maturase K), rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) (CBOLPlantWorkingGroup,2009), the trnH-psbA spacer, and the nuclear intergenic transcribed spacer (ITS2) located in the rDNA region and separating the 5.8S and 28S ribosomal RNA genes (Chen *et al*., 2010).

Prerequisites for species identification by the barcoding technology are that: (1) at the outset the species have been correctly identified by classical taxonomists using morphological features; (2) reliable barcodes for the identified species are available in public databases; (3) primer pairs are used that can amplify a maximum number of species; (4) the material is not contaminated by another organism.

DNA barcoding utilizes primers derived from highly conserved genes which, by PCR amplification, produce DNA fragments of a size that can be fully sequenced in a single step (ideally 500 to 700 bp) and which are most informative. Due to functional constraints coding sequences are less prone to mutations (which are to a large extent nucleotide substitutions in the wobble position of the codons) or multiples-of-3 nucleotide deletions or insertions. Non-coding sequences generally evolve more freely and faster and show therefore higher sequence differences than coding regions. The changes not only include a higher degree of nucleotide substitutions but also insertions/deletions of various lengths which can occur at any position in the targeted

sequences. Non-coding regions are therefore much more informative regarding the identification of species and can even reveal differences between closely related species and therefore be useful markers for phylogenetic studies.

The initial choice for the ITS2 marker for this study was guided by a then recent publication by Chen et al. (Chen, et al.,2010) who described this region as a novel DNA barcode for identifying medicinal plant species. This marker was described as easy to amplify and to generate reliable sequences. As a noncoding DNA region its mutation rate was expected to be high; on the other hand the multiple copies of the rRNA genes increase the risk of unreadable sequences (in case of nucleotide deletions) but also offered possibilities of even higher discrimination (in case of single nucleotide changes between the various copies of the rRNA genes).

Materials and Methods

Plant materials

Seeds of *Senna* and *Cassia* species were purchased from *Sunshine Seeds* (SS) (*http://www.sunshine-seeds.de*) and from *Rare Plants* (RP) (*http://www.rareplants.de*) (see Table 6). Seeds from most of these species could be brought to germination for DNA extraction, PCR amplification and sequence analysis. *Senna marilandica* material was obtained from the Botanical Institute of the University of Strasbourg (France).
DNA isolation and amplification

Total DNA from hypocotyls or young leaves from *Senna* and *Cassia* species was isolated using the CTAB method described by Saghai-Maroof *et al*. (Saghai-Maroof, et al.,1984) with minor modifications. In short, the plant material (about 100 mg) was placed in a 1.5 ml tube, frozen in liquid nitrogen, ground in the presence of glass beads (125-150 mm ø) using a Silama S5 mixer (Ivoclar Vivadent, Schaan, Li), and

subsequently incubated with 300 µl CTAB buffer (2% CTAB, 1% PVP, 100 mM Tris HCl pH8.0, 20 mM EDTA and 1.4 M NaCl) at 65°C for 10 minutes. Following addition of an equal volume of chloroform and vortexing, the tubes were centrifuged at 15000 g at room temperature for 5 min. The upper phase was transferred to a new tube and an equal volume of isopropanol was added. After 10 minutes incubation at room temperature, the tubes were centrifuged at 15000 g for 10 min. The pellet was subsequently washed with 1 ml 70% ethanol. After 10 min centrifugation the liquid phase was discarded and the pellet was dissolved in 100 µl 10 mM Tris-HCl pH 8.0 containing RNAse (1 mg/ml). PCR amplification was performed in a RoboCycler (Stratagene) in 25-µl reaction mixtures containing 2 µl of the genomic DNA solution (approximately 30 ng DNA), the corresponding primer pair and Advantage2 DNA Polymerase mix (Clontech). The primers used were the following: for trnH-psbA 5'-GTTATGCATGAACGTAATGCTC-3' (forward) and 5'-CGCGCATGGTGGATTCACAATCC-3' (reverse) (Ma, et al.,2010); for ITS2 5'-ATGCGATACTTGGTGTGAAT-3' (forward) and 5'-GACGCTTCTCCAGACTACAAT-3' (reverse) (Chen, et al.,2010). PCR conditions were for trnH-psbA: 94°C 5 min; 94°C 1 min, 55°C 1 min, 72°C 1.5 min, 30 cycles; 72°C 7 min; and for ITS2: 94°C 5 min; 94°C 30 s, 56°C 30 s, 75°C 45 s, 40 cycles; 75°C 10 min. PCR products were purified on NucleoSpin columns (Macherey-Nagel) and sequenced by Seqlab (Göttingen, Germany).

Data analysis

Sequences were aligned using MUSCLE software [Geneious V8.0.5 (Biomatters Ltd.)] with default parameters. All alignments were checked visually for optimization. Sequences were subsequently trimmed at the ends and only clearly readable sequences were kept for analysis. For each of the two markers, the phylogenetic tree was generated using the Neighbor Joining Method and the Tamura-Nei Genetic Distance Model (Geneious V8.0.5).

Databases consulted

DNA sequence databases: A number of sequences were also downloaded for analysis following a BLAST search of public databases (NCBI). They are marked by their accession numbers. Plant databases consulted for accepted names, synonyms as well as species descriptions and habitats include: The Plant List (*http://www.theplantlist.org/*), the Natural Resources Conservation Service (NRCS) USDA (*http://plants.usda.gov/*), the Encyclopedia of Life (*http://eol.org*); PlantNET, the New South Wales Flora Online (*http://plantnet.rbgsyd.nsw.gov.au/*), as well as the Atlas of Living Australia (*http://www.ala.org.au/*). Species descriptions were also taken from seed provider websites (*http://www.sunshine-seeds.de/*; *http://www.rareplants.de/*; *http://www.desert-tropicals.com*).

Results

I. Identification of the African plant

The African wood ITS2 sequence is highly related to that of Cassia grandis.

DNA from the wood was isolated and successfully amplified using the ITS2 primer pair (see Materials and Methods). Sequencing was performed by SeqLab (Göttingen, Germany). While a subsequent BLAST Search did not show a perfect or near-perfect match, it neverthess revealed a number of Cassia and Senna species with significant scores of which that of Cassia grandis (FJ009820) was highest (85.5% sequence similarity) (see Figure 24).

Since *Cassia grandis* is native to South America, and our wood came from a tree native to Africa, we surmised that the tree was very likely a related African species. Cassias are, together with the related Sennas, native to tropical and subtropical regions of the

globe, particularly to Central and South America, South-East Asia, Australia, and also to Africa (see Table 6). Many of the Cassia and Senna species have been used by the local populations to treat diseases.

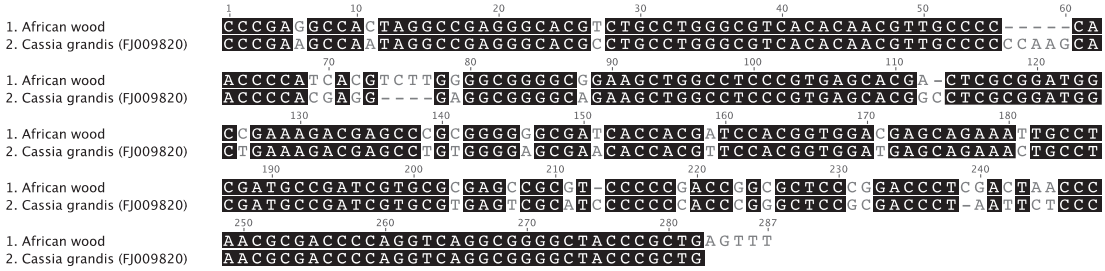


Figure 24. Alignment of the nucleotide sequences of the ITS2 from the African wood and *Cassia grandis* (FJ009820).

The African wood comes from a Cassia abbreviata tree

In order to further identify the plant we ordered seeds of Cassia and Senna species available from online seed companies (Rare Plants RP, and Sunshine Seeds SS, see Table 6). In total, seeds from 45 species were obtained, DNA was isolated from hypocotyls and amplified using the ITS2 marker primer pair. An analysis of the subsequently obtained sequences as well as of sequences from databases revealed that, with over 97% sequence similarity, *Cassia abbreviata* had a nearly perfect score (Table 7).

Accepted scientific name	Common name	Native from	Purchased	
Senna petersiana	Monkey Pod, Eared Senna	Africa		www.
Senna didymobotrya	African Senna, Popcorn Cassia	Africa (Central Africa, Madagascar)		www.
Cassia abbreviata	Long Pod Cassia, Sjambok pod	Africa (E, NE, S and W-C)		www.
Senna italica ssp. aradoides	Eland's Senna	Africa (South)		www.
Senna singueana	Winter Cassia	Africa (Tropical)		www.
Chamaecrista nigricans	Black Grain	Africa (Tropical), Australia		
Senna alexandrina (Mil.)	Alexandrian Senna, Egyptian Senna	Africa (Upper Egypt, Sudan)		
Senna italica	Senegal Senna	Africa, Asia		
Senna uniflora	Oneleaf Senna	America (Brazil, Mexico)		
Senna dariensis		America (Central & South)		
Cassia grandis	Pink Shower Tree	America (Central and South)	www.rareplants.de	www.
Senna pallida	Twin Flowered Cassia	America (Central), Caribbean		
Senna roemeriana	Twoleaf Senna, Roemer Senna	America (Central), Mexico		
Senna obtusifolia	Chinese Senna, American sicklepod, Java bean	America (Southern US, Caribbean)		
Senna alata	Candle Bush	America Central (Mexico)	www.rareplants.de	www.
Chamaecrista absus	Tropical Sensitive Pea	America N.& S., Africa, S. Asia, Australia		
Senna lindheimeriana	Lindheimer's Senna, Velvet Leaf Senna	America North		www.
Senna purpusii	Baja California Senna	America North		www.
Senna hebecarpa	American Senna, Wild Senna	America North (East)		www.
Cassia fasciculata	Partridge Pea	America North (Eastern USA)		
Senna covesii	Desert Senna	America North (SW USA)		www.
Senna covesii	Coues' Cassia	America North (SW USA)		
Senna roemeriana	Cheery Orange Senna, Two Leaf Senna	America North (SW USA)		www.
Senna spectabilis	Cassia Amarilla	America North (USA (NM, OK, TX))	www.rareplants.de	www.
Senna marilandica	American Senna	America North (USA Central and East)		
Cassia floribunda	Golden Showy Cassia, Devils Finger	America South		www.
Cassia glauca	Glaucous Cassia, Scrambled Egg Bush	America South		www.
Cassia leptophylla	Gold Medaillon tree	America South		www.
Senna bicapsularis	Butterfly Cassia, Winter Cassia	America South	www.rareplants.de	www.
Senna floribunda	Golden Showy Cassia, Devils Finger	America South		
Senna multijuga	False Sicklepod, November Shower	America South		www.
Senna pendula var. glabrata	Climbing Cassia	America South		www.
Senna reticulata	Golden Lantern tree, Carrion Crow Bush	America South	www.rareplants.de	www.
Cassia sp.Oliveira 610		America South ???		
Senna australis		America South (Brazil)		
Senna gardneri		America South (Brazil)		
Senna appendiculata		America South (Brazil)	www.rareplants.de	
Senna coquimbensis	Alcaparra	America South (Chile)		
Senna corymbosa	Texas Senna, Bush Senna, Argentina Senna	America South (Northern)		
Senna bicapsularis	Christmas Senna, Winter Cassia	America South (Northern), West Indies		
Senna occidentalis	Coffee Senna	America Tropical (?), now pantropical		www.
Senna hirsuta	Stinking Cassia, Woolly Senna	America(Tropical), South West USA		
Chamaecrista nictitans	Sensitive Partridge Pea, Small Partridge Pea,	Americas (temperate and tropical)		
Cassia roxburghii	Red Cassia, Ceylon Senna	Asia (India, Sri Lanka)		
Senna auriculata	Matura Tea Tree, Avaram Senna	Asia (India, Sri Lanka)		
Senna sophera	Sophera Senna	Asia (India)		
Senna siamea	Cassod Tree	Asia (S & SE)		
Senna surattensis	Golden Senna, Bush Senna, Scrambled Egg Tree	Asia (SE, India), Australia (Tropical)		www.
Cassia bakeriana	Pink Cassia, Dwarf Apple Blossom Tree	Asia (SE)		www.
Cassia javanica	Java Cassia, Pink Shower	Asia (SE)	www.rareplants.de	www.
Cassia nodosa	Pink Lady	Asia (SE)		www.

Senna glauca	Limestone Cassia	Asia (SE), Australia		
Senna tora	Sickle Wild-Sensitive Plant	Asia (South)		
Senna surattensis	Golden Senna, Bush Senna, Scrambled Egg Tree	Asia (Tropical), Australia		
Senna sulfurea	Sulphur-flowered Senna, Sulphur Cassia	Asia, Africa		
Cassia renigera	Burmese Pink Cassia	Asia, Burma	www.rareplants.de	www.s
Cassia fistula	Golden Shower tree	Asia, India, SE Asia	www.rareplants.de	www.s
Cassia marginata	Ceylon Senna, Red Cassia	Asia, India, Sri Lanka		www.s
Senna montana		Asia, Peninsular India		
Cassia artemisioides ssp. artemisioides	Silver Cassia, Feathery Cassia	Australia	www.rareplants.de	
Cassia brewsteri	Brewster's Cassia	Australia		www.s
Cassia eremophila	Punty Bush, Scented Cassia	Australia	www.rareplants.de	
Cassia glutinosa ssp. chatelainiana	Green Cassia	Australia		www.s
Senna artemisioides	Feathery Cassia, Silver Cassia	Australia		www.s
Senna barclayana	Pepperleaf Senna	Australia		www.s
Senna eremophila	Desert Cassia	Australia		www.s
Senna filifolia	Desert Cassia	Australia		www.s
Senna hamersleyensis	Clay Cassia	Australia		www.s
Senna odorata	Fragrant Senna	Australia		
Senna petiolaris	Woody Cassia	Australia		www.s
Senna phyllodinea	Silvery Cassia	Australia		
Senna pleurocarpa var. angustifolia	Firebush	Australia		www.s
Senna pleurocarpa var. pleurocarpa	Firebush	Australia		www.s
Senna venusta	Candlestick Cassia	Australia		www.s
Chamaecrista mimosoides	Tea Senna, Japanese Tea	Australia (West), America (Tropical)		
Senna glutinosa ssp. chatelainiana	Green Acacia, Glutinous Acacia	Australia (Western)		
Senna helmsii	Blunt-leaved Cassia	Australia (Western)		www.s
Senna notabilis	Cockroach Bush	Australia, Western Australia		www.s
Senna planitiicola	Arsenic Bush	Australia, Western Australia		www.s

Table 6. *Cassia* and *Senna* species discussed in the present study. In grey background are the species from w
in our lab in the frame of the project. They were either puchased from Rara Plants (RP) or from Sunshine See
was obtained from the Botanical Garden of the University of Strasbourg. Sequences from the other spe
databases (Genbank) and are shown on white background. Some *Chamaecrista* species have been included
section.

	Cassia sp....	Cassia gra...	Cassia lep...	Cassia ab...	African wo...	Cassia ab...	Cassia bre...	Cassia gra...	Cassia fist...	Cassia rox...	Cassia jav...	Cassia jav...	Cassia jav...	Cassia no...	Cassia jav...
Cassia sp. Oliveira 61...		100%	81.73%	82.87%	83.22%	82.52%	77.17%	76.10%	75.81%	77.85%	74.81%	77.99%	78.83%	77.14%	77.14%
Cassia grandis (FJ009...	100%		81.73%	82.87%	83.22%	82.52%	77.17%	76.10%	75.81%	77.85%	74.81%	77.99%	78.83%	77.14%	77.14%
Cassia leptophylla (SS)	81.73%	81.73%		83.44%	83.77%	83.77%	84.94%	83.75%	82.32%	85.39%	83.59%	84.52%	87.01%	85.40%	85.40%
Cassia abbreviata (KY)	82.87%	82.87%	83.44%		98.57%	97.86%	80.71%	76.56%	76.13%	78.57%	76.63%	77.74%	79.55%	77.53%	77.53%
African wood	83.22%	83.22%	83.77%	98.57%		98.57%	81.35%	76.88%	76.77%	79.22%	77.01%	78.39%	80.19%	78.16%	78.16%
Cassia abbreviata (SS)	82.52%	82.52%	83.77%	97.86%	98.57%		81.03%	76.56%	76.45%	78.90%	76.63%	78.06%	79.87%	78.16%	78.16%
Cassia brewsteri (SS)	77.17%	77.17%	84.94%	80.71%	81.35%	81.03%		86.52%	86.31%	87.50%	85.71%	87.82%	89.10%	86.56%	86.56%
Cassia grandis (SS)	76.10%	76.10%	83.75%	76.56%	76.88%	76.56%	86.52%		84.83%	86.25%	84.67%	86.88%	86.88%	84.45%	84.45%
Cassia fistula (SS)	75.81%	75.81%	82.32%	76.13%	76.77%	76.45%	86.31%	84.83%		85.90%	85.11%	87.22%	86.86%	84.06%	84.06%
Cassia roxburghii (JQ3...	77.85%	77.85%	85.39%	78.57%	79.22%	78.90%	87.50%	86.25%	85.90%		99.62%	95.79%	96.42%	91.11%	91.11%
Cassia javanica (FJ009...	74.81%	74.81%	83.59%	76.63%	77.01%	76.63%	85.71%	84.67%	85.11%	99.62%		95.06%	96.17%	89.59%	89.59%
Cassia javanica (JQ30...	77.99%	77.99%	84.52%	77.74%	78.39%	78.06%	87.82%	86.88%	87.22%	95.79%	95.06%		95.79%	90.85%	90.85%
Cassia javanica ssp no...	78.83%	78.83%	87.01%	79.55%	80.19%	79.87%	89.10%	86.88%	86.86%	96.42%	96.17%	95.79%		91.75%	91.75%
Cassia nodosa (SS)	77.14%	77.14%	85.40%	77.53%	78.16%	78.16%	86.56%	84.45%	84.06%	91.11%	89.59%	90.85%	91.75%		100%
Cassia javanica (SS)	77.14%	77.14%	85.40%	77.53%	78.16%	78.16%	86.56%	84.45%	84.06%	91.11%	89.59%	90.85%	91.75%	100%	

Table 7. Heat map showing the sequence similarities between the ITS2 markers (in %) of the Cassia species analyzed. Black boxes indicate >95% sequence identity, grey boxes indicate 75 to 95% sequence identity, and white boxes indicate <75% sequence identity.

That the sample of African wood stems from a *Cassia abbreviata* tree was further supported by the fact that *Cassia abbreviata* is native to Africa, and more precisely to West, Central, and East Africa (Figure 25). A sample of wood from a *Cassia abbreviata* tree in Kenya provided by M. Mulinge (South Eastern Kenya University) showed indeed 98.6% sequence similarity with our African wood ITS2 sequence, the minor sequence difference most likely reflecting the geographical variability in the marker sequence analyzed (as a reminder: the original wood was from the Democratic Republic of the Congo). Three geographical subspecies of *Cassia abbreviata* have indeed been described: *Cassia abbreviata* subsp. *abbreviata* (Mosambique, Zambia, Zimbabwe, Tanzania and Democratic Republic of the Congo), subsp. *beareana* (Botswana, Kenya, Mosambique, Namibia, Zimbabwe, Somalia, Tanzania), and subsp. *kassneri* (Kenya and Tanzania) (http://memim.com/cassia-abbreviata.html).

Figure 25. Distribution of *C. abbreviata* in Africa. Three subspecies have been identified: subsp. *abbreviata*, subsp. *beareana*, and subsp. *kassneri* (see text above)

Figure *26. Cassia abbreviata. A: Tree. B: Flowers are zygomorphic (have bilateral symmetry). Note the sigmoidal-shaped basis of the filaments of the three long curved stamens (characteristic for Cassia species; see also Figure 32). C: open pod (a few seeds are still visible at the top and at the bottom)* www.africanplants.senckenberg.de

Uses of Cassia abbreviate

Cassia abbreviata (Figure 26) is used for many diseases including haematuria (smoked leaves), toothache (root infusion) gastrointestinal disorders, venereal diseases, pneumonia, uterus complaints, heavy menstruation, stomach-ache, bilharzia, snakebites (root decoction), malaria and blackwater fever (root extracts). The wood is termite resistant and therefore the poles are used for house construction (http://www.prota4u.org/protav8.asp?en=1&p=Cassia+abbreviata+Oliv.).

II Phylogenetic analysis of Cassia and Senna species

Cassia and *Senna* are two closely related genera of the *Cassiinae* subtribe which is the largest of five forming the *Cassieae* tribe of the *Leguminosae* family of flowering plants. They are mostly trees or shrubs growing predominantly in tropical and subtropical regions including Australia, South-East Asia, Central and East Africa, as well as Central and South America. Many of the species have been used in local traditional medicines for the treatment of various diseases.

The taxonomic position of many species of the two genera was not clearly defined for a long time and many species formerly classified as Cassias have been transferred over the years to the *Senna* or *Chamaecrista* genus. It was only in 1982 when Irwin and Barneby published a comprehensive taxonomical study on Cassia and Senna species based on floral morphological traits (see Figure 32). And yet, *The Plant List* (*http://www.theplantlist.org*), an online encyclopedia in charge of determining which are accepted species names in the *International Plant Names Index (IPNI)* database, created by authoritative institutions like the Royal Botanical Gardens Kew (UK) and The Missouri Botanical Garden (US) and launched in December 2010, still lists many accepted Cassia names with Senna synonyms and vice versa. The situation is therefore rather confusing even now, especially since very little effort has been made

in recent years to use DNA data in support of their identification and classification. Only one detailed phylogenetic study on nearly 90 Senna species combining morphological features like floral symmetry and extrafloral nectaries as well as DNA sequence data, has been published in 2006 by Marazzi *et al*. (Marazzi, et al.,2006).

We have used the DNA data generated during the identification study of *Cassia abbreviata* to establish the phylogenetic relationship between the various species analyzed. In addition to the ITS2 marker we studied the chloroplast marker trnH-psbA spacer which is a non-coding region, and which is therefore fast evolving. In total, seeds from 45 species of the two genera could be purchased from online seed companies. The ITS2 region and the trnH-psbA spacer could be successfully amplified and sequenced for the vast majority of the species considered in the study (43/45 for each marker). For 33 species clear sequences were obtained for the two markers. While the amplified trnH-psbA spacer fragments exhibited a considerable size difference (from about 210 bp to 470 bp before trimming, 196 to 400 bp after trimming), the ITS2 region was less variable in length (277 to 312 bp after trimming). *Cassia* and *Senna* sequences for the two markers available in public databases are also included in the present analysis.

Internal Transcribed Spacer 2 (ITS2)

The internal transcribed spacer 2 (ITS2) has been proposed by Chen *et al.* (Chen, et al.,2010) as one of the most valuable barcoding marker for medicinal plants. It is indeed one of the markers that can be most successfully amplified and sequenced and shows high sequence diversity among plant species. Although in general the PCR-amplified ITS2 sequences are easily readable, the chromatograms occasionally contain double peaks, sometimes 1 bp shifts (due to single nucleotide deletions) starting at a specific point, indicating a heterogeneity in the template DNA. This is not surprising considering the existence, in eukaryotic genomes, of multiple copies of

rDNA genes which are are not absolutely identical. As an example, the diploid *Arabidopsis* contains approximately 1200-1500 rDNA repeats with some, but limited sequence variation (Mentewab, et al.,2011). Also, extensive pyrosequencing revealed frequent intra-genomic variations in internal transcribed spacer regions of nuclear ribosomal DNA of many plant species (Song, et al.,2012).

We have included in our study 67 ITS2 sequences from 56 species. Of these, 43 sequences were determined in our laboratory (from plant material grown from purchased seeds) and 24 sequences are from public databases. The optimized alignment of the sequences (Figure 27) reveals only a few small gaps, the longest of which being 29 bp. The ungapped sequences range from 277 to 319 nucleotides in length. The ITS2 sequences are very GC-rich (61 to 71%). The sequence identities between the various species listed range from 57 to 100%. The alignment shown in Figure 27 reveals 6 regions (A to F) which significantly contribute to the differentiation of the *Cassia* and *Senna* species, the most striking feature being a 15 bp deletion (region F) present in 14 sequences and corresponding to 9 different species, all with accepted *Cassia* names (*The Plant List*). These same 14 *Cassia* sequences also share other unique features which include an AC-PuAC sequence in positions 45-50 (region A), a highly variable region B, and a conserved A at position 195 (region C). All the species from the *Senna* subclades share a 5 bp deletion at position 220 (region D), a 2 bp deletion at position 253 (region E) and a 3 bp deletion at position 261 (region E). Clear outlier sequences (53 to 65% sequence similarity with the other sequences listed) are those from *Chamaecrista absus* (KC817015 and FJ009832), *Chamaecrista nictitans* (FJ009855) and *Cassia fasciculata* (reclassified as *Chamaecrista fasciculata*). The phylogenetic tree constructed using the sequence alignments from Figure 27 shows the subdivision of the species into one major *Cassia* clade as well as 5 *Senna* subclades (Figure 29).

The ***Cassia* clade** includes 14 sequences from 9 species: *C. abbreviata* (SS, CD, KY),

C. leptophylla (SS), *C. brewsteri* (SS), *C. fistula* (SS), *C. javanica* (SS, JQ301831), *C. javanica subsp. nodosa* (SS, FJ980413), *C. roxburghii* (JQ301841), *Cassia* sp Oliveira 610, and *C. grandis* (SS, FJ009820). The three sequences corresponding to *C. abbreviata* (Sunshine Seeds, Democratic Republic of Congo, and Kenya) differ among each other by 4 or 6 nucleotides (see positions 7, 80, 165, 222, 230, 266, and 301 in Figure 27) and may correspond to the three different subspecies (subsp. *abbreviata*, *beareana*, and *kassneri*; see Figure 25). Two different sequences were found for *C. javanica* (JQ301831 and SS), for *C. javanica* subsp. *nodosa* (FJ980413 and SS), and for *C. grandis* (FJ009820 and SS). The existence of different sequences for *C. javanica* may not be surprising as they may come from different subspecies (at least 6 *C. javanica* subspecies have been described: subsp. *javanica*, subsp. *agnes (*also *indochinensis)*, subsp. *nodosa*, subsp. *renigera,* subsp. *pubifolia,* subsp. *microcalyx*). The fact that two sequences (differing by as many as 27 nucleotides) were found for subsp. *nodosa* however suggests that misidentification or mislabeling may be at the origin of two sequences for this marker. This not only the case for *C. javanica* subsp. *nodosa* but also for other species, for instance *C. grandis*. No subspecies has been identified for the latter species; therefore the existence of two significantly different sequences (FJ009820 and Sunshine Seeds, differing by 75 nucleotides) is unexpected. It is worth noting that a sequence [*Cassia* sp. Oliveira 610 (FJ009819)] identical to that of *C. grandis* (FJ009820) was identified by the same group; the two latter sequences differ by at least 50 nucleotides (including a 40 bp gap at positions 73-112) present in all the other Cassia ITS2 sequences. Our *C. grandis* (Sunshine Seeds) has been identified on the basis of the trnH-psbA spacer sequence (see above) as well as of the matK and rbcL sequences (data not shown). The *C. grandis* (FJ009820) ITS2 sequence may therefore not correspond to that of a *C. grandis*; its identity as well as that of *Cassia* sp. Oliveira 610 (FJ009819) remains unclear.

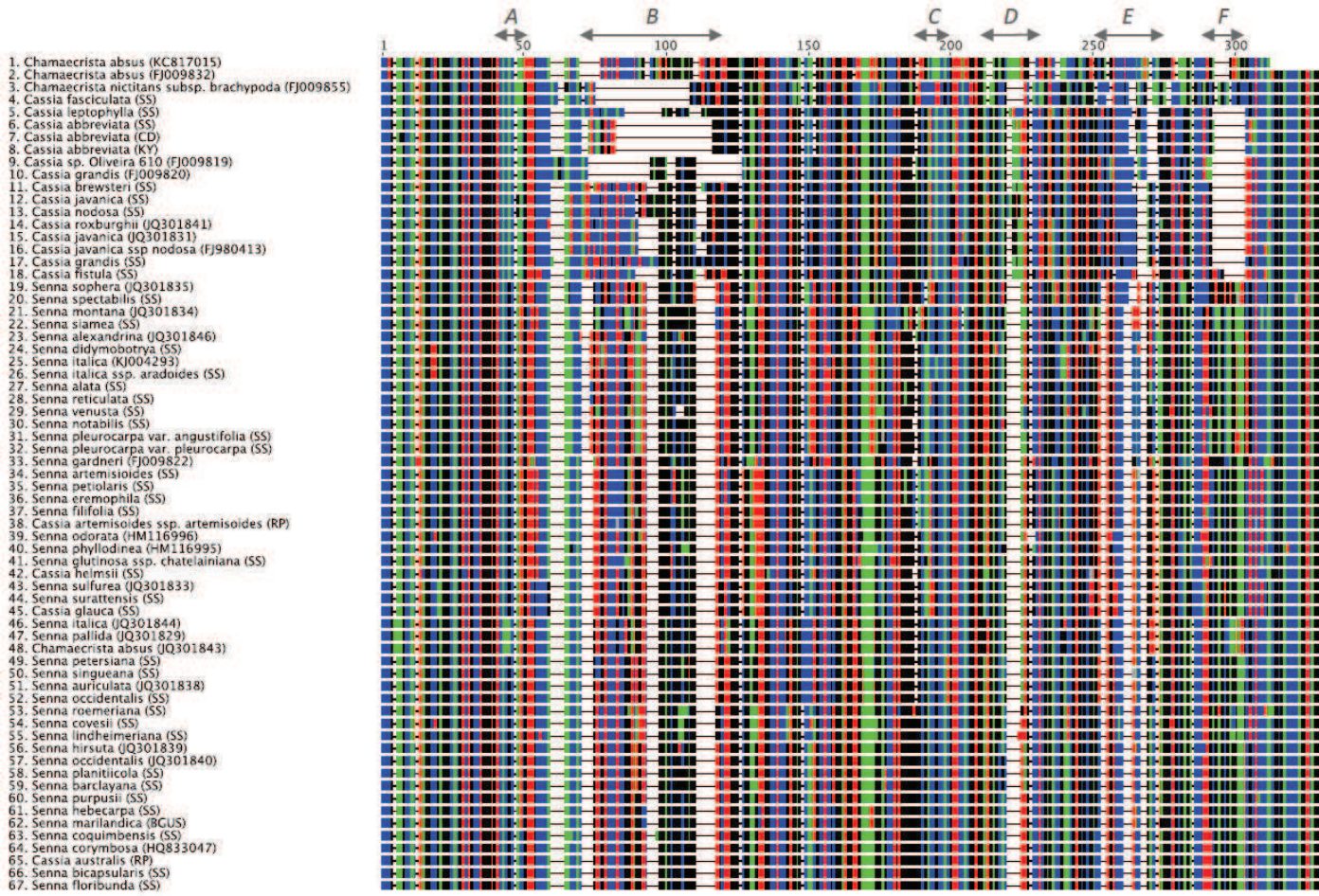


Figure 27. Alignment of the ITS2 sequences. MUSCLE Align software was used for a first alignment. The align by visual inspection. Color codes are the following: A = green; G = black; C = blue; T = red. The ITS2 region

***Senna* subclade 1** comprises the sequences *S. alexandrina* (JQ301846), *S. didymobotrya* (SS), *S. italica* (KJ004293), *S. italica* subsp. *aradoides* (SS), *S. alata* (SS), *S. reticulata* (SS), S. *venusta* (SS), *S. notabilis* (SS), *S. pleurocarpa* var. *angustifolia* and var. *pleurocarpa* (SS). All the species of this subclade share the following unique features: a stretch of 4 A at position 59, 2 A at position 90, the CATT sequence at position 149, 2 T at positions 172 and 212, 2 G at position 235, and a GTG at position 103. The *S. pleurocarpa* ssp *angustifolia* and subsp. *pleucocarpa* sequences show an example of how an intragenomic heterogeneity in the rDNA spacer can be used to discriminate between two very closely related subspecies: indeed, while subsp. *pleurocarpa* has a G:A ratio of 10:1 at position 176, subsp. *angustifolia* has a G:A ratio of 1:1 (Figure 28A). *S. pleurocarpa*, *S. venusta* and *S. notabilis* (differing among themselves by 6 to 13 bp) form a small group of native Australian species in this subclade.

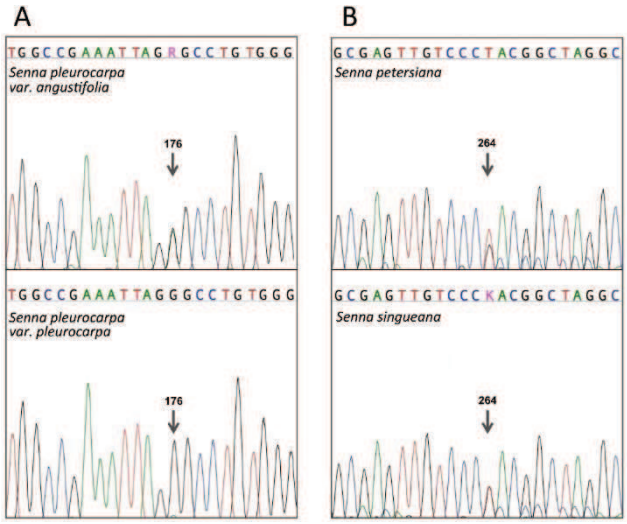


Figure 28. Intragenomic heterogeneities in the ITS2 region of closely related *Senna* species. A. Chromatographs of *Senna pleurocarpa* var. *pleurocarpa* and *Senna pleurocarpa* var. *angustifolia* ITS2 sequence showing a double peak (G:A=1:1 in var. *angustifolia* and G:A = 10:1 in var. *pleurocarpa*) at position 176. B. Chromatographs of *Senna petersiana* and *Senna singueana* ITS2 sequence showing a double peak (T:G= 6.6:4.1 for *S. petersiana* and T:G=1:1 for *S. singueana*) at position 264. Numerous additional double peaks confirm the intragenomic sequence heterogeneity in the multiple copies of the ITS2 region of these two species.

Identical ITS2 sequences have been found for *S. alata* and *S. reticulata*. While *S. alata* is native to Mexico, *S. reticulata* is native to the Amazonian floodplains (Brazil) and they have not been described to be closely related genetically, which suggests one of the two has been misnamed or misidentified.

The two different sequences for *S. italica* (KJ004293 and subsp. *aradoides*) differ by 4 nucleotides, suggesting that KJ004293 corresponds to that of a different *S. italica* subspecies. Two other subspecies have indeed been described for *S. italica*: subsp. *italica* and subsp. *micrantha*. The two *S. italica* sequences are also very similar to that of *S. didymobotrya* (differing from the latter by 4 and 8 nucleotides respectively). *S. italica* and *S. didymobotrya* are both native to Africa, hence the similarity of their ITS2 sequences (97%) is not surprising. A database search revealed a third *S. italica* sequence (JQ301844) not belonging to Senna subclade 1 but to subclade 3, and differing only by one nucleotide from that of *S. pallida* (JQ301829) (T instead of G at position 240). *S. pallida* is native to South America, and it is very likely that the *S. italica* sequence (JQ301844) corresponds to that from a misidentified/misnamed subspecies/variety of the former.

S. alexandrina (JQ301846), another native African species, shows at least 24 nucleotides difference (about 92% identity) with the most closely related *Senna* species of this subclade and therefore may have evolved over a longer period than *S. didymobot*rya and *S. italica* from a common African ancestral species.

***Senna* subclade 2** comprises *S. artemisioides* (SS), *S. petiolaris* (SS), *S. eremophila* (RP), *S. filifolia* (SS), *C. artemisioides* subsp. *artemisioides* (RP), *S. odorata* (HM116996), *S. phyllodinea* (HM116995), *S. glutinosa* subsp. *chatelainiana* (SS), *C. helmsii* (SS), *S. sulfurea* (JQ301833), *S. surattensis* (SS), and *C. glauca* (SS). Unique features conserved in all these species include: TT at position 49-50, ATC at position 87-89, T at position 127, CCA at position 149-151, and GGT at position 269-271.

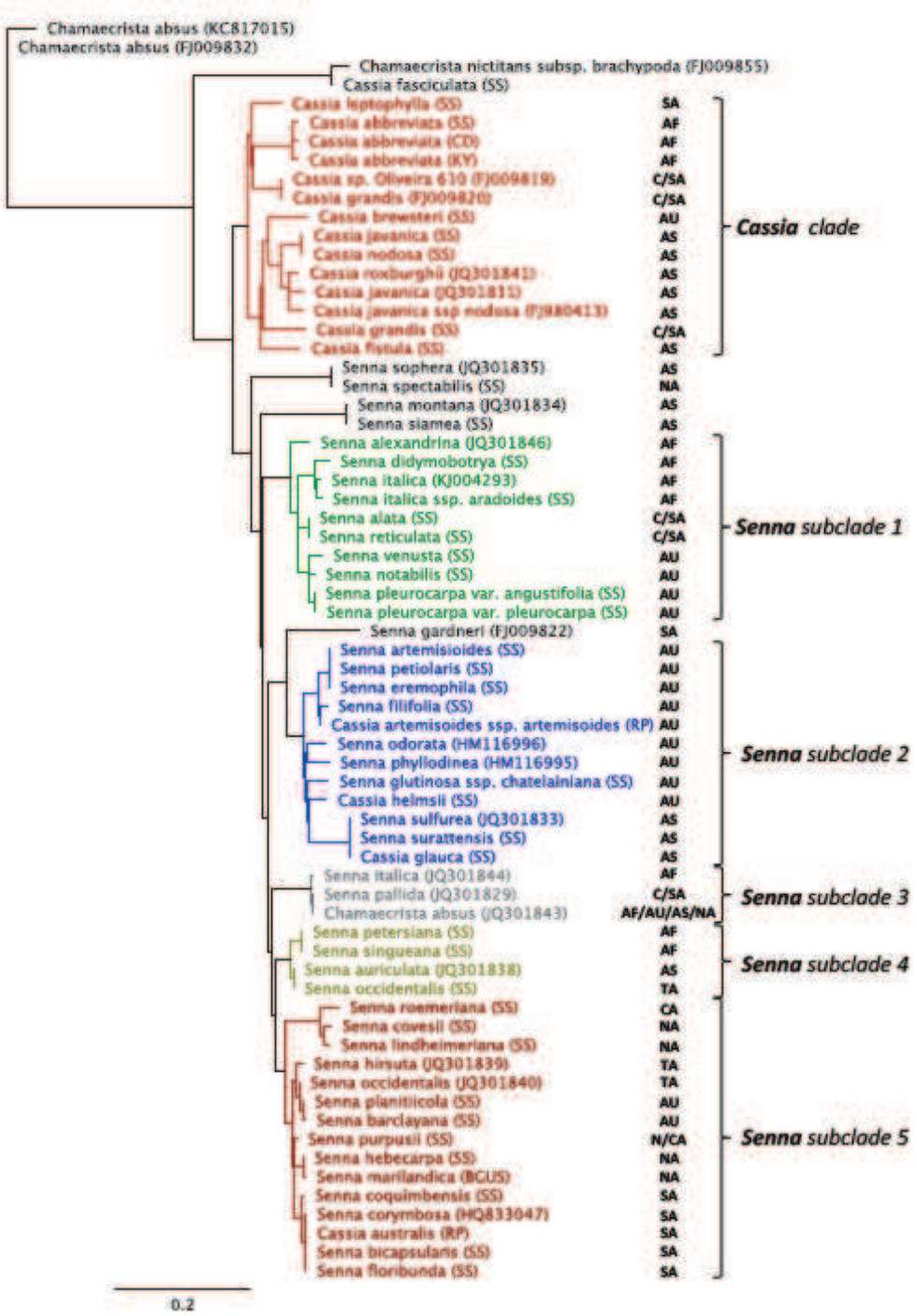


Figure 29. Phylogenetic tree built from the aligned ITS2 sequences. Continents to which the various species are native are indicated: AU = Australia; AS = Asia; AF = Africa; NA = North America; SA = South America; CA = Central America; TA = Tropical America. Note the sequences that are most likely derived from misidentified species: *Senna italica* JQ301844 and *Chamaecrista absus* JQ301843 in *Senna* subclade 3. Other inconsistencies are discussed in the text.

S. surattensis, *C. glauca* and *S. sulfurea* and have identical ITS2 sequences. While the former is an accepted species name which occasionally (and possibly also here) has been erroneously used to refer to *S. sulfurea*, *S. sulfurea* is also an accepted name while *C. glauca* Lam. is a synonym of the latter *(http://www.theplantlist.org/tpl/record/ild-29621*).

Five species of this subclade differ by 10 nucleotides or less: *S. filifolia*, *C. artemisioides* subsp. *artemisioides*, *S. eremophila*, *S. artemisoides*, and *S. petiolaris*. All five species have been considered to be subspecies of *S. artemisoides* (*S. filifolia* is occasionally also considered to be a synonym of *S. eremophila*). The fact that the *S. artemisioides* (Sunshine Seeds) sequence is not identical to that of *C. artemisioides* subsp. *artemisioides* suggests that the two might correspond to different subspecies. As to *C. helmsii*, it is also described as a subspecies of *S. artemisoides* (*S. artemisoides* subsp. *helmsii*, see *http://www.ala.org.au/*); however its ITS2 sequence diverges significantly from the *S. artemisioides* sequence (20 to 25 nucleotides difference), as do the sequences of *S. odorata*, *S. phyllodinea* and *S. glutinosa* subsp. *chatelainiana*. Unique for *S. odorata* is a 3-bp deletion at position 227.

S. gardneri, which is genetically quite distant from the other Cassias/Sennas (with *C. helmsii* being closest with 44 bp difference), has a unique stretch of 5 "G"s at position 291.

***Senna* subclade 3** comprises *S. italica* (JQ301844), *S. pallida* (JQ301829) and *Chamaecrista absus* (JQ301843). The *S. italica* (JQ301844) and *S. pallida* (JQ301829) sequences differ among themselves by 1 nucleotide, suggesting that they are very closely related. However their different geographical distribution (*S. pallida* native to Central America, *S. italica* native to Africa) suggests that one of the two species was misidentified. This is most likely *S. italica* (JQ301844), since two very similar *S. italica* ITS2 sequences (differing among themselves by only 4 nucleotides) clustering with

another African species (*S. didymobotrya*) have been described in the *Senna* subclade 1. As to the *Chamaecrista absus* sequence (determined by the same group), it is identical to the *S. pallida* (JQ301829) sequence, suggesting an error in data submission, since two other *Chamaecrista absus* ITS2 sequences (KC817015 and FJ009832) sharing 99.6% identity are available (see outgroup in Figure 26); they differ by >40% from the JQ301843 sequence. Also, on the basis of the trnH-psbA spacer sequence *S. pallida* was shown to be very different from *Chamaecrista absus* [only 42% identity between the *Chamaecrista absus* (HQ161768) and *S. pallida* (HQ161754) sequences; see above].

***Senna* subclade 4** comprises 4 sequences: *S. petersiana* (SS), *S. singueana* (SS), *S. auriculata* (JQ301838), and *S. occidentalis* (SS). The ITS2 sequences of *S. petersiana* and *S. singueana* are virtually identical, differing only by a different T:G ratio at position 264: in *S. petersiana* this ratio is 6.6:4.1, while in *S. singueana* it is 1:1 (see Figure 28B). These different ratios reflect a sequence heterogeneity in the multiple copies of rDNA units. The two species are native to Africa. *S. auriculata* and *S. occidentalis* have identical ITS2 sequences. Since *S. auriculata* is native to Asia (India, Sri Lanka), while *S. occidentalis* is native to South America, it is likely that one of the two species has been misnamed or mislabeled. This is most likely *S. occidentalis* since a second sequence [*S. occidentalis* (JQ301840)] differing by 20 nucleotides has been identified (see subclade 5).

***Senna* subclade 5** comprises the species *S. roemeriana* (SS)*, S. covesii* (SS), *S. lindheimeriana* (SS), *S. hirsuta* (JQ301839), *S. occidentalis* (JQ301840), *S. planitiicola* (SS), *S. barclayana* (SS), *S. purpusii* (SS), *S. hebecarpa* (SS), *S. marilandica* (USTR), *S. coquimbensis* (SS), *S. corymbosa* (HQ833047), *S. australis* (RP), *S. bicapsularis* (SS), and *S. floribunda* (SS).

Unique features of this subclade include a GGGGGGG stretch at position 183 (with a

T at position 187 for *S. roemeriana*), and a GAC at position 269-271. The *S. covesii*, *S. lindheimeriana*, *S. roemeriana* ITS2 sequences differ among themselves by 7 to 15 bp. All three species are native to North America. The *S. hirsuta* (JQ301839) and *S. occidentalis* (JQ301840) sequences are very similar (98.7% sequence identity); the two sequences (which are from tropical American species) also share high sequence identities (97-99%) with the two native Australian species *S. planitiicola* and *S. barclayana*, and a little less (about 96%) with the North American species *S. hebecarpa* (Eastern US), *S. marilandica* (Eastern US), and with the native South American species *S. coquimbensis* (Chile), *S. corymbosa* (Argentina), *S. australis* (Brazil), *S. bicapsularis* (Central and South America), and *S. floribunda* (Caribbean). The ITS2 sequences of the latter four species are identical. This is not surprising at least for *S. corymbosa* and *S. floribunda* since *Cassia corymbosa* (Ortega) has been considered a synonym for *S. floribunda* which is an accepted name (theplantlist). Regarding our *S. bicapsularis* ITS2 sequence, it is identical to that of *S. bicapsularis* (HQ833043), suggesting that this species has been correctly identified. It is interesting to note that the chloroplast trnH-psbA spacer sequence of our *S. bicapsularis* shows a significant sequence difference (12 nucleotides difference, or 92.9% identity) with that of *S. floribunda* and 16 nucleotides difference (88.1% identity) with that of *S. pendula* var. *glabrata* (GU135439) which is also native to South America (see above under trnH-psbA spacer), suggesting that these species are best identified via their trnH-psbA marker. As to *S. australis*, no other barcode sequence is available in the databases, hence it is difficult to assess whether it has been correctly identified/named and whether it can be identified using a specific barcode marker. *S. hebecarpa* and *S. marilandica* differ by a single nucleotide (A/G at position 271). Since both are native to the Eastern US, their close genetic relationship is not surprising. *S. purpusii*, native to Central and North America (Mexico, Baja California), is closely related (only 6 to 8 bp difference) to the other North American (*S. hirsuta*, *S. purpusii*, *S. hebecarpa*, *S. marilandica*) as well as South American (*S. coquimbensis*, *S. corymbosa*, *S. australis*, *S. bicapsularis*, *S. floribunda*) and, surprisingly, also to Australian (*S. barclayana*, *S.*

planitiicola) species of this clade.

Two pairs of sequences not fitting into any of the listed Senna subclades include *S. sophera* and *S. spectabilis* as well as *S. montana* and *S. siamea*. The *S. sophera* (JQ301835) and *S. spectabilis* (RP) ITS2 sequences are virtually identical, showing only double peaks at positions 297 (A/C) and 317 (C/T) in the S*. spectabilis* sequence, which suggests a very close genetic relationship; however since the two species are native to different continents (Asia and Tropical America respectively) it is most likely that one of the two has been misidentified/mislabeled. The very similar trnH-psbA spacer sequence (11 bp difference, including an 8 bp indel) of *S. spectabilis* (RP) and *S. spectabilis* (HQ161761) suggests that the former was correctly identified and that the two sequences could correspond to two *S. spectabilis* subspecies/varieties; two varieties have indeed been described: var. *excelsa,* native to South America (Brazil), and var. *spectabilis,* native to the Caribbean (see link *http://www.eol.org*). In support of the correct identification of *S. spectabilis* (RP) is also the fact that its rbcL barcode marker is identical to that of *S. spectabilis* (Sequence ID CAATB229-11.rbcL; *http://www.boldsystems.org*; and unpublished data). That *S. spectabilis* and *S. sophera* are distinct and not very closely related species is supported by the trnH-psbA sequences (68-nucleotide difference, or 80% identity) between *S. spectabilis* (RP) and *S. sophera* (HQ161760) (see trnH-psbA spacer sequence alignment and tree, Figs 8 and 9), a difference which could be expected indeed for species native to different continents. The *S. siamea* (SS) and *S. montana* (JQ301834) ITS2 sequences are identical. The very similar trnH-psbA spacer sequences of the two species (1-6 bp difference, see above) indicate that they are genetically very closely related. This is also supported by the fact that they are both native to the same continent (Asia).

trnH-psbA spacer

The trnH-psbA spacer is a marker located on the chloroplast DNA present in multiple

identical copies in the plant cells and is therefore expected to yield clear sequences without double peaks on the chromatographs. The successful amplification of this marker from DNA extracted from the African wood allowed to confirm the identity of the latter as *Cassia abbreviata*. As it encompasses the non-coding region separating the chloroplast genes trnH (gene encoding the $tRNA^{His}$) and psbA (gene encoding the D1 polypeptide of the photosystem II reaction center complex) its high variability in sequence as well as in length indicates that this region is not subjected to structural contraints unlike the ITS2 marker, which, being transcribed to yield a precursor transcript, has to adopt a specific secondary structure required for correct cleavage into the two ribosomal subunits 5.8S and 28S rRNAs.

In total, 71 sequences (from 59 species) ranging from 196 bp (*S. pallida*) to 400 bp (*C. roxburghii*) in length, were considered for the analysis of the chloroplast trnH-psbA spacer region: 43 unambiguous reads from DNA isolated in our lab from plantlets grown from commercially available seeds, as well as 28 sequences available in databases. The 71 sequences were aligned using the *MUSCLE Align* software. The optimized alignment generated following visual inspection includes 8 gaps longer than 3 bp (G1 to G8 in Figure 30). All the sequences of this marker are AT-rich (about 75% A+T) contrasting with the GC-rich nuclear ITS2 sequence (see Figure 27). The sequence identity, based on the sequence alignments, ranges from 41 to 100%. In the phylogenetic tree generated from the aligned sequences, the various species cluster into 6 major (sub) clades (Figure 31). Outlier sequences include *Cassia fasciculata* (which has been re-classified as *Chamaecrista fasciculata* and whose trnH-psbA marker sequence is identical to that of *Chamaecrista nictitans*), and *Senna italica* (HQ161769). The latter shows high sequence similarity with *Chamaecrista mimosoides* (HQ161757): the only differences are found at positions 379-382 (ATTT in *S. italica* vs --AG in *Ch. mimosoides*). The most relevant features of the various clades and of some of their members will be described in the following lines.

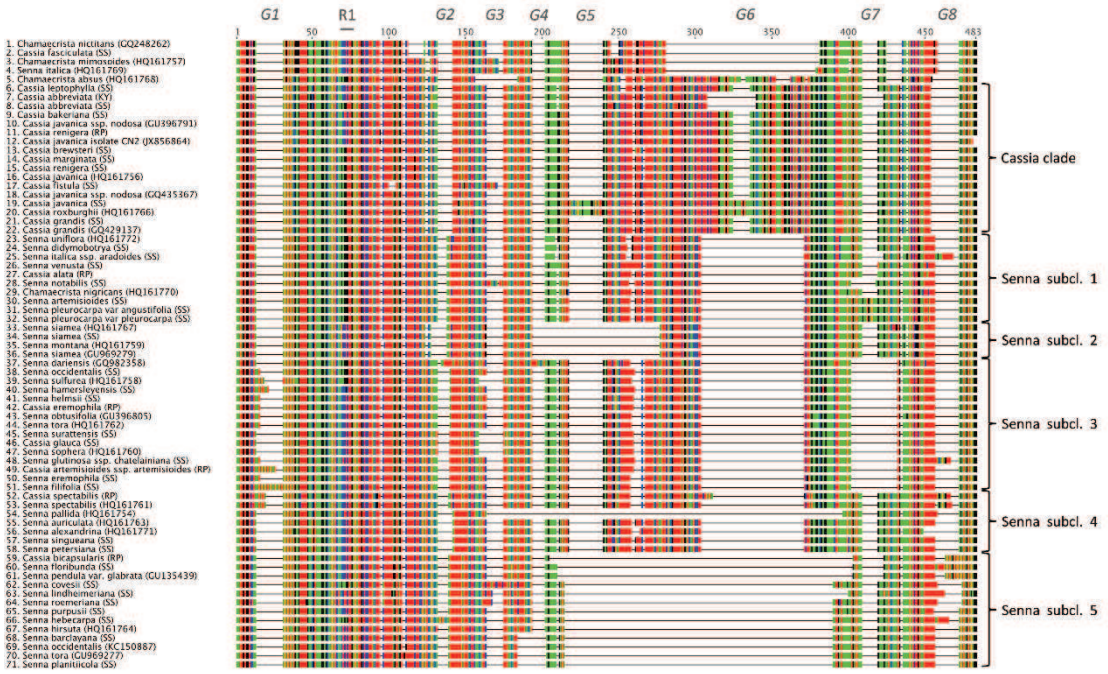


Figure 30. Alignment of the trnH-psbA spacer sequences. MUSCLE Align software was used for a first alignment. Alignments were subsequently optimized by visual inspection. (Sub) clades were determined, primarily based on the presence of conserved insertions/deletions (gaps G1-G8). Note the "insertion" in G6 and the conserved G8 for the *Cassia* species; the single large gap G6 and purine-rich pentanucleotide in R1 (position 69-73) for *Senna* subclade 1; the two large gaps G5 and G6 for *Senna* sublade 2 species; the two large gaps G6 and G7 for the species of *Senna* subclade 3; the unique large gap G6 for species and the pyrimidine-rich pentanucleotide in R1 (position 69-73) for *Senna* subclade 4 species (this subclade also contains *S. pallida* with the longest gap spanning G3 to G6). *Senna* subclade 5 species are characterized by a long gap spanning G5 and G6. Color codes are the following: A = green; G = black; C = blue; T = red. The trnH-psbA spacer region is very AT-rich (75%).

The ***Cassia* clade** includes the following sequences: *C. leptophylla* (SS), *C. abbreviata* (SS, KY), *C. bakeriana* (SS), *C. javanica* (SS, HQ161756), *C. javanica* subsp. *nodosa* (GU396791, GQ435367), *C. renigera* (SS, RP), *C. brewsteri* (SS), *C. marginata* (SS), *C. grandis* (SS, GQ429137), *C. fistula* (SS), and *C. roxburghii* (HQ161766). The ungapped trnH-psbA spacer sequences of this clade range from 341 bp (*C. abbreviata*)

to 400 bp (*C. roxburghii*); the size differences are mainly due to a shorter G6 gap (32 bp in *C. abbreviata*, and 11 bp in the other remaining species of this clade except for *C. roxburghii* and *C. javanica* (SS) which have inserts in G5 and G6. All the members of this clade have a C at position 38, a stretch of 3-5 Ts (while all other species listed have only 1 or 2) at position 392, a C at position 435, and they all have a stretch of only 5 Ts in the region 449). The *C. roxburghii* (HQ161766) sequence is the longest of all the trnH-psbA spacer sequences listed. This is due to the fact that it contains two head-to-tail copies of two duplicated segments, one 22 bp long (TAACATAAGAAAAAGGATATTG) at positions 188 and 218, and one 11 bp long (TGAAATGTAAA) at positions 314 and 325. We determined an identical sequence for a *C. javanica* (SS) which may in fact correspond to *C. javanica* (sensu Bojer), which is a synonym of *C. roxburghii* DC (*http://www.theplantlist.org/tpl/record/ild-1108*).

Two *C. abbreviata* trnH-psbA spacer sequences have been determined in our laboratory: one from hypocotyls grown from purchased seeds (SS), and one from leaves collected from a tree in Kenya (KY). While the two sequences share a unique 32 bp gap (vs 11 bp) at position 308, as well as a 14 bp gap (vs 10 bp) at position 409 as compared to the other members of this clade, they differ from each other by a 5 nucleotide stretch (ACCTC vs GAGGT at position 69), a trinucleotide (CAT vs TGC at position 246), and single nucleotide changes at position 285 (G vs A) and at position 289 (T vs G). This demonstrates the existence of an intraspecies variability for the trnH-psbA marker in *C. abbreviata*, possibly reflecting different geographical origins (see also Figure 25).

C. fistula is the only species in this clade that has a 5-bp deletion at position 100, and a 7-bp insertion at position 164. All the sequences available for *C. fistula* in the databases (accession numbers JX856861, JX856862, GQ435368, HQ161755) are identical to the sequence we determined, suggesting that, unlike for *C. abbreviata* and other *Cassia* species (see below), there is no sequence variation for this marker in *C.*

fistula. In support of this, no subspecies of *C. fistula* has been described.

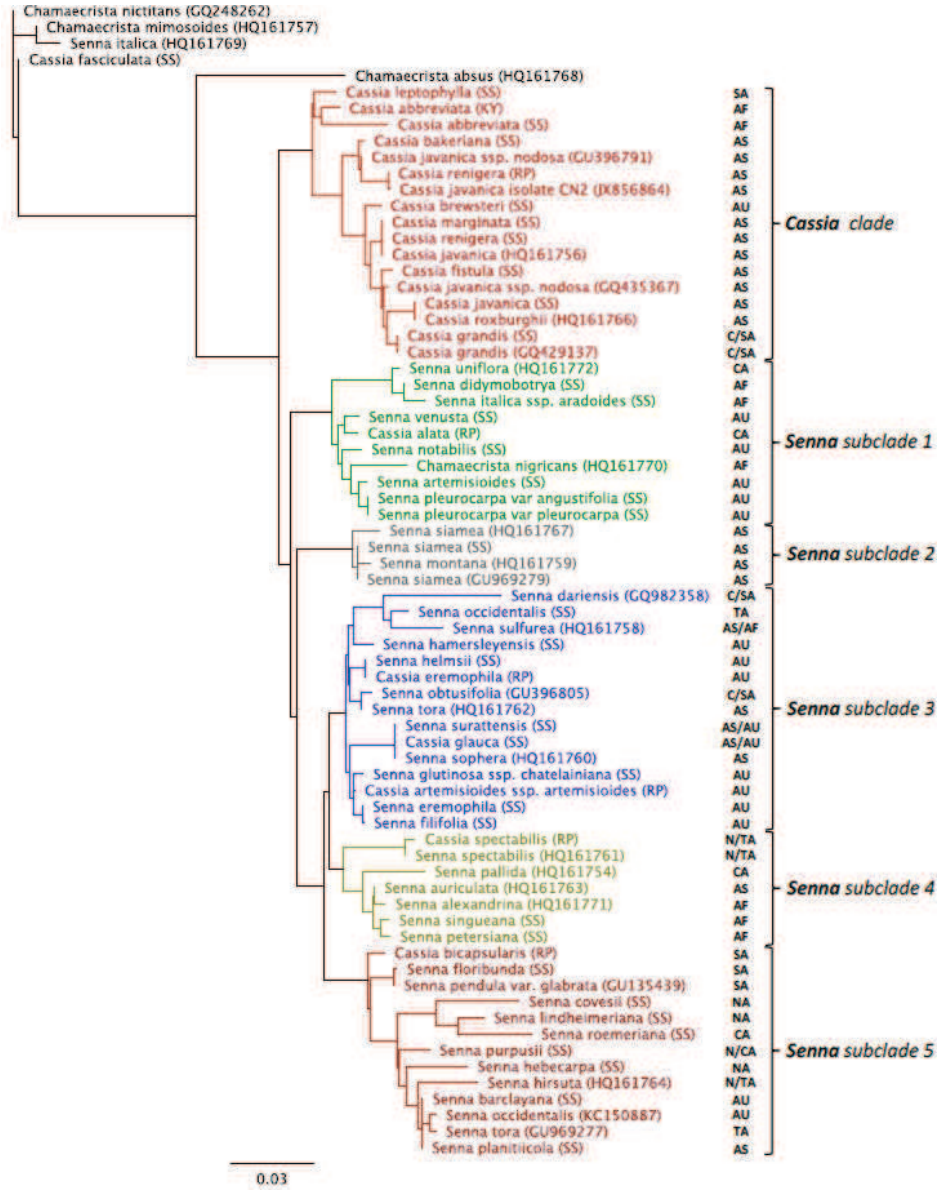


Figure 31. Phylogenetic tree built from the aligned trnH-psbA spacer sequences. Continents from which the various species are native are indicated: AU = Australia; AS = Asia; AF = Africa; NA = North America; SA = South America; CA = Central America; TA = Tropical Amercia. Note the sequences that are most likely derived from misidentified species: *Chamaecrista nigricans* (HQ161770) in *Senna* subclade 1; the two *Senna tora* sequences (*S. tora* HQ161762 in S. subclade 2 and *S. tora* GU969277 in S. subclade 5); the *S. italica* HQ161769 sequence in the *Chamaecrista* outlier group. *Cassia fasciculata* (as named by Sunshine Seeds) is in fact *Chamaecrista fasciculata*. Other inconsistencies see text.

The *C. grandis* sequence we determined is identical to the sequence found in the database (accession number GQ429137). *C. grandis* is the only species within the *Cassia* and *Senna* clades that has a G (instead of a T) at position 155.

Two sequences (GU396791 and GQ435367) are available in the databases for *C. javanica* subsp. *nodosa.* They differ by a pentanucleotide at positions 69-73 (ACCTC vs. GAGGT), and by positions 146 and 396 (A vs. C). The GU396791 sequence differs by 8 nucleotides (positions 69-73, 117, 264, 266) from the *C. javanica* sequence HQ161756 and by one nucleotide (position 71) from the *C. bakeriana* sequence. The GQ435367 sequence differs from the *C. javanica* sequence HQ161756 by 5 nucleotides (positions 117, 146, 264, 266, 396) and from the *C. bakeriana* sequence by 7 nucleotides (positions 69-73, 146, 396). Three *C. javanica* sequences are available: JX856864, HQ161756, as well as the sequence we determined from seedlings (Sunshine Seeds). While the former two sequences differ by 12 bp, the sequence from Sunshine Seeds plants shows 39 nucleotide difference with *C. javanica* (HQ161756) and 46 bp difference with *C. javanica* (JX856863). This difference is predominantly due to the presence, in the Sunshine Seeds sequence, of two "inserts" which are also present in the *C. roxburghii* (JQ301841) sequence (see discussion under *C. roxburghii* above). Confusion is also raised by *C. marginata* (SS), which, as a synonym *of C. roxburghii* as well, should have the same sequence. *C. marginata* however lacks the two inserts, and is identical to *C. javanica* (HQ161756) and *C. renigera* (SS). *C. renigera* is an accepted name and is a synonym of *C. javanica* subsp. *renigera* (*http://www.theplantlist.org/tpl/record/ild-1103*). This may explain the sequence identity of the latter with that of *C. javanica* isolate CN2 (JX856864), as well as the identity of the second *C. renigera* sequence (RP; see below) with *C. javanica*. The two *C. renigera* trnH-psbA spacer sequences (SS and RP) differ by 11 nucleotides. While the *C. renigera* (SS) sequence (identical to *C. javanica* HQ161756) shares the same pentanucleotide (GAGGT) with *C. javanica* subsp. *nodosa* (GQ435367) at positions 69-73, the *C. renigera* (RP) sequence (identical to *C. javanica* isolate CN2

(JX856864) shares at these same positions the ACCTC sequence with *C. javanica* subsp. *nodosa* (GU396791). The *C. javanica* subsp. *nodosa* GQ435367 and *C. renigera* (SS) sequences are not identical though: they differ at 5 positions (C vs G at position 117; C vs A at position 146; G vs T at position 264; T vs - at position 266; and C vs A at position 396). Similar for GU396791 and the *C. renigera* (RP) sequence: they share the ACCTC sequence at position 69-73, but they differ at positions 39 (G vs C), 94 (T vs G), and 281 (C vs T). All this illustrates the complexity and confusion generated by the high resolution power of DNA-based plant identification methods, a confusion often due to incorrect or incomplete naming within a same or between highly related and morphologically very similar species.

A distinctive feature of *C. leptophylla* in this clade is a 4-bp gap at positions 251-254, a gap also found in *Chamaecrista absus* (HQ161768). *C. leptophylla* is also the only species of the *Cassia* and *Senna* species listed here that has a C (instead of an A) at position 447.

The two species most closely related to *C. brewsteri* are *C. javanica* subsp. *nodosa* GQ435367 (differences at positions 261, 264, 266, 368 and 396) and *C. grandis* (differences at positions positions 155, 261, 264, 368 and 396). *C. brewsteri* is the only one of all *Cassia* and *Senna* species listed that has an A at positions 261 and 264. This low sequence divergence may be surprising since the three species are native to three different continents: Australia for *C. brewsteri*, Asia for *C. javanica*, and South America for *C. grandis*.

The ***Senna* subclade 1** includes the following species: *S. uniflora* (HQ161772), *S. didymobotrya* (SS), *S. italica* subsp. *aradoides* (SS), *S. venusta* (SS), *S. alata* (RP), *S. notabilis* (SS), *S. artemisioides* (SS), *S. pleurocarpa* var. *pseudocarpa* (SS) and *S. pleurocarpa* var. *angustifolia* (SS) (as well as the *Chamaecrista nigricans* HQ161770 sequence, see discussion below). Their ungapped spacer lengths range from 293 to

328 bp. All 9 species share a 67/68-bp gap G6 (which is also present in the species of subclades 2, 3 and 4 but which is much longer in subclade 5).

S. artemisioides, *S. pleurocarpa* var. *pseudocarpa* and var. *angustifolia* have an identical trnH-psbA spacer sequence, and this sequence is also the only one of all species listed to have a 10 bp insertion (GAATGATAAA) at position 409. Although the three species are native to Australia, the high similarity (only 3 nucleotides difference) of the *S. artemisioides* sequence with that of the two *S. pleurocarpa* subspecies is surprising (especially since the ITS2 sequence differs quite considerably (by 53 nucleotides) between *S. artemisioides* and *S. pleurocarpa*; see below in ITS2).

S. italica subsp. *aradoides* shares the same 14 bp gap at position 409 with *C. abbreviata*. It has a unique 12 bp insert (AGAATTTTTTTT) at position 457.

Characteristic for the species *S. notabilis* are a 10-bp insertion (TTTAACA-TTA) at position 164, and a 16-bp gap G7 (position 402). Other differences between species of this clade include single nucleotide deletions/insertions and single nucleotide changes. The BLAST search for related sequences also revealed a *Chamaecrista nigricans* (HQ161770) sequence which has all the features of the Senna subclade 1 species. Unique features include the GC dinucleotide at positions 216-217, and a C at position 448 and 452. This species was most likely misidentified.

***Senna* subclade 2,** with the sequences *S. siamea* (HQ161767, GU969279 and SS) and *S. montana* (HQ161759), is characterized by a unique long gap spanning positions 194-276. The three *Senna siamea* sequences differ among themselves by 1, 3 and 4 nucleotides respectively; they differ from the *S. montana* sequence by 3, 4 and 6 nucleotides respectively.

***Senna* subclade 3** includes the following sequences: *S. dariensis* (GQ982358), *S.*

occidentalis (SS), *S. sulfurea* (HQ161758), *S. hamersleyensis* (SS)*, C. helmsii* (SS), *C. eremophila* (*RP*), *S. obtusifolia* (GU396805), *S. tora* (HQ161762), *S. surattensis* (SS), *C. glauca* (SS), *S. sophera* (HQ161760), *S. glutinosa* subsp. *chatelainiana* (SS), *C. artemisoides* subsp. *artemisioides* (RP), *S. eremophila* (SS), and *S. filifolia* (SS). Their ungapped spacer length ranges from 289-309 bp. Distinctive for these species is the 30 bp gap (G7) at position 402 (in most *Cassia* and *Senna* species listed here this gap is shorter (10, 14 or 15 bp long). The most relevant differences between species of this subclade are found in the region extending from position 14 to 30 where most of them have an AT-rich insert of variable length: these include *S. hamersleyensis (*ATATTAAT), *S. occidentalis, S. obtusifolia* (GU396805)*, S. tora* (HQ161762)*, C. helmsii, S. eremophila, C. eremophila*, *S. glutinosa* subsp. *chatelainiana* (AT), *S. sulfurea* (ATATA), *C. artemisioides* (ATATATATATAAT), and *S. filifolia* (ATATATATATAATATAT). *S. glutinosa* subsp. *chatelainiana* is the only species of this subclade with an insertion (10 bp) at position 457.

S. dariensis, S. occidentalis and S. sulfurea have a purine-rich pentanucleotide at position 69 (similar to the subclade 1 species) while in the other species of this subclade this stretch consists almost exclusively of pyrimidines. Of all the species listed *S. dariensis* is the only one to have an AAAAAAATTTTTTATTTA sequence in the G2 region (positions 128-147) and a duplicated TTTAACAT sequence at positions 186-193 and 194-201.

S. sophera (HQ161760), *C. glauca* (Sunshine Seeds) and *S. surattensis* have identical trnH-psbA spacer sequences; while *C. glauca* [var. *suffruticosa* (Koenig ex Roth) Baker] is an accepted synonym of *S. surattensis* [(Burm.f.) H.S.Irwin & Barneby] and the identity of the two sequences is therefore expected, *S. sophera* is an accepted name (*http://www.theplantlist.org/tpl/search?q=Senna+sophera*) and is considered to be a distinct species. The sequence identity of the *S. sophera* and the *C. glauca* / *S. surattensis* trnH-psbA marker sequence suggests that the two species are very closely

related though, which may not be surprising since all three species are native to Asia.

The *S. obtusifolia* (GU396805) *and S. tora* (HQ161762) trnH-psbA spacer sequences differ by a single nucleotide (G instead of A at position 438). The latter sequence differs considerably (by 126 nucleotides) from the *S. tora* sequence in subclade 5 (GU969277). Since the *S. tora* name has been erroneously used in the past to refer to *Senna obtusifolia* (L) H.S.Irvin & Barneby (*http://www.theplantlist.org/tpl/record/ild-1085*), the former sequence (*S. tora* (HQ161762)) may in fact correspond to that of *S. obtusifolia* (which is native to Central and South America, while *S. tora* is native to Asia).

The *C. eremophila* and *S. eremophila* trnH-psbA sequences differ by 6 nucleotides (positions 140, 163, 247, 255, 256, and 373). Since the seeds of the two species were obtained from different providers (*Rare Plants* and *Sunshine Seeds* respectively) the differences might be attributed to different ecotypes (different geographical origins). Alternatively one of the species may have been incorrectly identified; in support of the latter hypothesis is the fact that the sequence of the *C. eremophila* (*Rare Plants*) marker is identical to that of *C. helmsii* (*Sunshine Seeds*). No barcode marker sequence neither for *C. helmsii* nor for *C.* or *S. eremophila* is available in databases to verify the accurate identity of these species.

The trnH-psbA spacer sequence of *S. occidentalis* (Sunshine Seeds) in this subclade 3 differs significantly (only 58.8% sequence identity) from that described in subclade 5 (*S. occidentalis* KC150887). *S. occidentalis* is supposed to be native to Tropical America but is now a pantropical weed. No sequence identical to this one is presently described in the database and it is very likely that one of the two has been incorrectly named.

***Senna* subclade 4** comprises the sequences *S. spectabilis* (RP, HQ161761), *S. pallida* (HQ161754), *S. auriculata* (HQ161763), *S. alexandrina* (HQ161771), *S. singueana*

(SS), and *S. petersiana* (SS). The species of this subclade, with the exception of *S. pallida*, share a number of features with the Senna subclade 1 members: these include the number and the organization of the gaps, as well as the size of gap G6 (67 bp). *S. pallida*, due to a long gap extending from position 164 to 395, has the shortest spacer (180 bp) of all the Cassia and Senna species listed. They are native to Asia (*S. auriculata*), Africa (*S. alexandrina*, *S. singueana*, and *S. petersiana*), and Tropical America (*S. pallida*, *S. spectabilis*). The major difference between the subclade 1 and subclade 4 members is the presence of a GAGGG sequence at positions 69-73 in clade 1 members while clade 4 members have a CCCTC in the same position. *S. spectabilis* is the only species of clade 4 that has an (AT-rich) insert (ATATTA) at position 14. All other species that have an insert in this position are from subclade 3. *S. spectabilis* is also the only species listed that has a C at position 42, a G at position 92, and a 10 bp insert at position 458. Of the two different *S. spectabilis* sequences available one (*Rare Plants*) has an 8 bp insert at position 305; they also differ in the 10 bp insert at position 457 (TAGATAGTTT- vs. -AGATAGTTTT). The very similar trnH-psbA spacer sequence (11 bp difference, including an 8 bp indel) of *S. spectabilis* (Rare Plants) and *S. spectabilis* (HQ161761) suggests that *S. spectabilis* (Rare Plants) was correctly identified and that the two sequences could correspond to two *S. spectabilis* subspecies/varieties; two varieties have indeed been described: var. *excelsa* (native to South America (Brazil) and var. *spectabilis* (native to the Caribbean) (see link *eol.org*).

The *S. alexandrina* sequence is unique among the species listed in that it lacks the stretch of "T"s at position 449.

S. auriculata differs from *S. singueana* at 2 positions (247 and 400) and from *S. petersiana* at 3 positions (250, 400 and 434). *S. petersiana* differs from *S. singueana* at positions 241, 250 and 434.

***Senna* subclade 5** comprises 13 sequences: *S. bicapsularis* (RP), *S. floribunda* (SS), *S. pendula var. glabrata* (GU135439), *S. covesii* (SS), S. *lindheimeriana* (SS), *S. roemeriana* (SS), *S. purpusii* (SS), *S. hebecarpa* (RP), *S. hirsuta* (HQ161764), *S. barclayana* (SS), *S. planitiicola* (SS), *S. occidentalis* (KC150887), and *S. tora* (GU969277. The ungapped spacer lengths of these species are between 218 and 251 bp. The shorter spacers are mainly due to the presence of a long gap G6 (>150 bp) centering around position 330. In 9 of the species (*S. roemeriana*, *S. covesii*, *S. hebecarpa*, *S. purpusii*, *S. occidentalis*, *S. tora*, *S. barclayana*, *S. planitiicola*, *S. hirsuta*) this gap is 178 bp long, in *S. lindheimeriana* it is 188 bp, in *S. floribunda* and *S. pendula* it is 197 bp, and in *S. bicapsularis* it is 217 bp long. Four species (*S. occidentalis*, *S. tora*, *S. barclayana* and *S. planitiicola)* also have an 11-bp longer gap G4. While the *S. barclayana* and *S. planitiicola* sequences are identical, *S. tora* (GU969277) is the only of all species listed that has a 2 bp insertion (TG) at position 109 and differs from the three other species by 3 nucleotides (including the 2 bp insert at position 109 and nucleotide change at position 111). *S. barclayana* and *S. planitiicola* are native to Australia, which may explain their close genetic relationship. *S. occidentalis* is native to Tropical America and would therefore be expected to be genetically more distant. A more distant genetic relationship would also be expected for *S. tora* which is native to Asia. Another *S. tora* (HQ161762) and *S. occidentalis* (Sunshine Seeds) sequence have been described in subclade 3. The *S. occidentalis* (Sunshine Seeds) and *S. occidentalis* (KC150887) sequences differ by 131 nucleotides and it is most likely that one of them has been misnamed. The same is the case for the two *S. tora* sequences (HQ161762 and GU969277) which differ by 126 nucleotides.

C. bicapsularis and *S. pendula var. glabrata* have an extra ATATAGTAT sequence at position 463 which is repeated immediately downstream (position 472). The same nonanucleotide sequence is present in the *S. floribunda* sequence but in this case it is preceded by an ATTTTT hexanucleotide (position 457). *C. bicapsularis* differs from *S. pendula* var. *glabrata* and *S. floribunda* by the A at position 103, the 5 bp longer large

gap G6, and the AT (instead of GA) at position 403 (immediately following G6). All three species are native to South America.

Unique features of a few other species of this clade include: for *S. hebecarpa* two insertions: ATATACA at position 132, and ATTTTTTTT at position 457 (a repeat of the immediate upstream sequence); for *S. covesii* one insertion (TTTATTCTTTC) at position 164. *S. hebecarpa* and *S. covesii* are native to North America.

S. lindheimeriana and *S. roemeriana* share the same TTTC insert at position 164 but differ by 8 nucleotide changes (positions 11, 46, 102, 140, 163, 184, 423, 480); also the *S. lindheimeriana* spacer has a 10 bp-longer gap G6 (390-399), and a 5 bp-longer T-stretch at position 449.

Discussion

A detailed taxomomical study of Cassia, Senna, as well as of Chamaecrista species was published in 1982 by Irwin and Barneby (Irwin, et al.,1982). This study illustrates the massive amount of knowledge, expertise and patience that is required for correct identification using the classical approach (see excerpt and illustration in Figure 32). It comes therefore not as a surprise that quite a few species have been misnamed or incorrectly identified. Species identification by DNA barcoding is therefore a promising approach allowing much faster and more accurate identifications without requiring expert taxonomical knowledge. One of the prerequisites is that the barcodes for the individual species and subspecies have been first established by the expert taxonomists on the basis of earlier morphological descriptions. Only recently, a molecular phylogeny using both molecular and morphological data has been proposed for Senna species (Marazzi, et al.,2006).

Key to the Genera of subtribe Cassiinae

1. Filaments of 3 abaxial antesepalous stamens sigmoidally curved (below middle outward from the fl's vertical axis, thence inward) and many times longer than their anther, this dorsifixed subversatile and introrsely dehiscent by slits, the filaments of 2 abaxial antepetalous and of the remaining 5 adaxial stamens straight and shorter, their anthers dehiscent mostly by basal pores; pedicels 2-bracteolate at or shortly above base; pod elongate, cylindric or variously compressed, indehiscent, pulpy or pithy within; seeds 1- or 2-seriate, their funicle filiform; seed-coat smooth, exareolate; no extrafloral nectaries; trees or early arborescent shrubs; all continents, obligately tropical. **Cassia**

1. Filaments of all stamens straight or simply incurved and either shorter than or not over twice as long as their anther, if 3 abaxial ones longer than the rest these 2 antepetalous with 1 antesepalous between them, and the anthers in any case terminally dehiscent by slit or pore; pedicels either bracteolate or ebracteolate, but if bracteoles present these (with extremely rare exceptions) inserted above or near middle of pedicel and the pod then elastically dehiscent; pod various, but if resembling that of Cassia the pedicels ebracteolate; funicle and seed variable; extrafloral nectaries common but not universal.
   2. Bracteoles 0; pod either indehiscent or inertly dehiscent through 1 or both sutures, if through one only then follicular, if through both then the valves tardily separating but not coiling, or (infrequently) the valves breaking up into 1-seeded joints; androecium commonly zygomorphic, the stamens tending to dwindle from abaxial to adaxial side of the fl, the 3 adaxial members commonly staminodal, but sometimes all 10 subequal fertile; anther-thecae naked along the sutures; extrafloral nectaries (when present) mounded, claviform or phalloid, secreting nectar from a convex surface; funicle filiform; seed-coat either smooth or minutely rugulose but not pitted, often charged on each face (or on margin) with a closed areole; trees, woody vines, shrubs, herbs; all continents and Oceania, a few extratropical. **Senna**

   2. Bracteoles 2; pod elastically dehiscent, the valves coiling; androecium (suberratically) actinomorphic, the 2 cycles of stamens bearing anthers of different lengths but never zygomorphically dwindhng from abaxial to adaxial side of fl; anther-thecae ciliolate along the sutures; extrafloral nectaries, when present, dish- or cup-shaped, rarely flat, secreting nectar from a concave (flat) surface; funicle deltately dilated; seed-coat either smooth or pitted, but exareolate; all continents and Oceania, but primarily American, one section {Chamaecrista) highly differentiated also in Africa, less so in Asia, Australia and the Pacific, a few extratropical in both Old and New World. **Chamaecrista**

from : The American Cassinae. A Synoptical Revision of Leguminosae Tribe Cassieae subtribe Cassinae in the New World. Howard S. Irwin and Rupert C. Barneby. Memoirs of the New York Botanical Garden, Bronx, New York (1982) p. 1-2

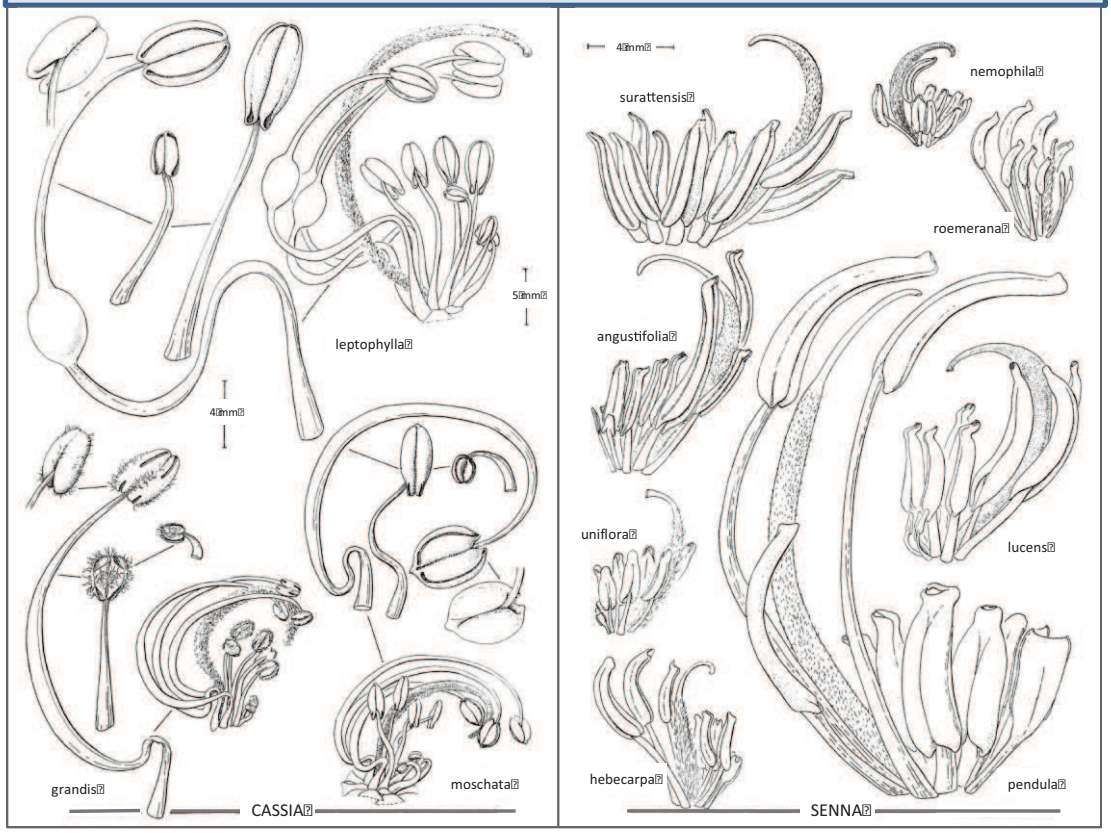


Figure 32. Stamen structure of Cassia and Senna flowers (Irwin, et al.,1982). Note the sigmoidal basis of the three long (antesepalous) stamens which is typical for Cassia species.

Presently the *Plant List* (*http://www.theplantlist.org*) includes 1396 scientific plant names of species for the genus *Cassia*. Of these, only 87 are accepted names while the others are either synonyms or unassessed. The list also includes 332 scientific plant names of species rank for the genus *Senna*. Of these, 263 are accepted species names.

Identification of Cassia abbreviata by DNA barcoding

Cassia abbreviata is one of the few Cassia species native to Africa. Three other native African species with accepted Cassia names are *Cassia afrofistula*, *Cassia sieberiana*, and *Cassia burtii*. *Cassia afrofistula* (Kenyan shower) is native to East Africa. *Cassia sieberiana* (drumstick tree) is distributed from Senegal and Gambia east to Democratic Republic of the Congo and Uganda. *Cassia burtii* is native to Mosambique and Tanzania. For none of the three latter species were there seeds available for analysis, hence their ITS2 sequences could not be compared and their phylogenetic relationships to Cassia abbreviata could not be established.

It was only in 2013 that the *Cassia afrofistula* rbcL (JX572386) and matK (JX518010) barcode markers became publicly available. The two markers show 2-nucleotide differences with those from *C. abbreviata* which we also sequenced and which were identical to those of the sequences in the databases. For *Cassia sieberiana* and *Cassia burtii* no sequence is available to date for any marker for comparison.

Cassia abbreviata, *Cassia sieberiana*, *Cassia burtii* and *Cassia afrofistula* are medicinal plants and are used in local medicine.

trnH-psbA spacer and ITS2 as barcode markers for Cassia, Senna and Chamaecrista species

We have studied the phylogenetic relationship of 65 *Cassia* and *Senna* as well as of 4 *Chamaecrista* species using the nuclear ITS2 and/or the chloroplast trnH-psbA spacer sequences. Our analysis of aligned sequences of the two markers revealed that several conserved structural features allow to sort the species into six major clusters/clades. One of these clades contains exclusively species with accepted *Cassia* names, predominantly characterized by a unique "insertion" in G6 of the trnH-psbA spacer sequence and a unique deletion in the ITS2 sequence. In addition to the five other subclades which comprise *Senna* species, the study includes an outgroup comprising some members of the related *Chamaecrista* genus. The unique features mentioned above, found in a plastid as well as a nuclear marker, can therefore be considered as molecular fingerprints for the taxonomic assignment of *Cassia* species. Accordingly, all the *Cassia* species names listed here in the *Senna* subclades should take the *Senna* genus name, as should be the case for *Cassia artemisioides*, C. *barclayana*, C. *chatelainiana*, C. *eremophila*, C. *hamersleyensis*, C. *helmsii*, C. *italica*, C. *notabilis*, C. *pleurocarpa* and C. *venusta* in *The Plant List*. As to *Chamaecrista nigricans* (HQ161770) (trnH-psbA spacer) and *Chamaecrista absus* (JQ301843) (ITS2), their position in a *Senna* subclade is most likely due to a misidentification, mislabeling, or data submission error.

Comparison with other published Cassia and Senna barcode/phylogeny data

Two publications describing data closely related to this work will be discussed in the following lines. One deals with the study of several barcode markers (including the trnH-psbA spacer and the ITS2 region) from medicinal *Cassia* and *Senna* species in India (Purushothaman, et al.,2014). These sequences are available under the accession numbers HQ161753 to HQ161772 (psbA-trnH) and JQ301828 to JQ301847 (ITS2) and have been included in our alignments and in the discussion. Some of the data in the publication are conflicting with other published data; for instance, regarding the trnH-psbA spacer, HQ161770 is identified as *Chamaecrista nigricans*, but it

clusters with *Senna* subclade 1 species (see discussion under *Senna* subclade above). Also the *S. tora* (HQ161762) sequence is very similar to the *S. obtusifolia* (GU396805) sequence (1 nucleotide difference), and the *S. italica* (HQ161769) sequence is very similar to that of *Chamaecrista mimosoides* HQ161757) (4 nucleotides difference). As to the ITS2 marker, the *Chamaecrista absus* (JQ301843) sequence is identical to that of *S. pallida* (JQ301829) and shows only 1 nucleotide difference with that of *S. italica* (JQ301844).

A detailed phylogenetic study by Marazzi *et al*. (Marazzi, et al.,2006), based on the chloroplast markers rpS16 intron, rpL16 intron and matK and involving 98 taxa of which 87 were ascribed to *Senna*, resolved the latter species into seven major clades. Twenty of these species have also been considered in our study for their trnH-psbA and/or ITS2 markers. Of these, 17 species clustering together in 3 different Senna subclades, were found to cluster together with the same species in the Marazzi study (see Table 8). These species are: *S. didymobotrya*, *S. italica*, *S. alata*, *S. reticulata*, *S. notabilis*, *S. pleurocarpa* and *S. venusta* (*Senna* subclade 1 and Marazzi clade II); *S. artemisioides*, *S. odorata*, *S. glutinosa*, *S. dariensis* (*Senna* subclades 3 (trnH-psbA) and 2 (ITS2) and Marazzi clade IV); *S. purpusii*, *S. bicapsularis*, *S. corymbosa*, *S. occidentalis*, *S. barclayana*, and *S. hirsuta* (*Senna* subclade 5 and Marazzi clade VII). The co-clustering of species in a same subclade for five (chloroplast and nuclear) markers is in support of their correct identification and their evolution from a common ancestor for the species of a same clade.

ITS2		trnH-psbA spacer		Combined matrix of rpL16 intron, rpS16 intron and matK (Marazzi *et al.* 2005)	
subclade 1	*S. didymobotrya* (SS) *S. italica* (SS) *S. alata* (SS) *S. reticulata* (SS) *S. notabilis* (SS) *S. pleurocarpa* (SS) *S. venusta* (SS)	subclade 1	*S. didymobotrya* (SS) *S. italica* (SS) *S. alata* (SS) *S. notabilis* (SS) *S. pleurocarpa* (SS) *S. venusta* (SS)	clade II	*S. didymobotrya* *S. italica* *S. alata* *S. reticulata* *S. notabilis* *S. pleurocarpa* *S. venusta*
subclade 2	*S. artemisioides* (SS) *S. odorata* (HM116996) *S. glutinosa* (SS)	subclade 3	*S. artemisioides* (SS) *S. glutinosa* (SS) *S. dariensis* (GQ982358)	clade IV	*S. artemisioides* *S. odorata* *S. glutinosa* *S. dariensis*
subclade 5	*S. purpusii* (SS) *S. bicapsularis* (SS) *S. corymbosa* (HQ833047) *S. occidentalis* (JQ301840) *S. barclayana* (SS) *S. hirsuta* (JQ301840)	subclade 5	*S. purpusii* (SS) *S. bicapsularis* (RP) *S. occidentalis* (KC150887) *S. barclayana* (SS) *S. hirsuta* (HQ161764)	clade VII	*S. purpusii* *S. bicapsularis* *S. corymbosa* *S. occidentalis* *S. barclayana* *S. hirsuta*

Table 8. Co-clustering of Senna species based on 5 markers (trnH-psbA spacer, ITS2, rpl16 intron, rps16 intron, and matK).

Inconsistencies with/between publicly available data

A number of other inconsistencies between data from databases as well as our data emerged during this study: starting with the *S. italica* (HQ161769) trnH-psbA sequence which is clearly a *Chamaecrista* sequence (it shows 95 to 99% sequence identity with trnH-psbA sequences from *Chamaecrista* species and at most 80% sequence identity with sequences from *Senna* or *Cassia* species); two *S. tora* trnH-psbA sequences (HQ161762 and GU969177) differing by 126 bp; two *S. occidentalis* trnH-psbA sequences (Sunshine Seeds and KC150887) differing by 131 bp; very similar sequences for trnH-psbA sequences from species native to different continents, such as *S. occidentalis* (KC150887) and *S. tora* (GU969277) which differ by only 3 nucleotides.

Inconsistencies are also found between ITS2 data from databases as well as our own data: the *C. grandis* sp. Oliveira 610 (FJ009819) sequence, identical to *C. grandis* (FJ009820), is very different (only 75.4% sequence identity) from the *C. grandis* sequence we determined from seedlings obtained from two different providers. In

addition, the matK and trnH-psbA spacer sequences of our *C. grandis* DNA are identical to those published in databases (JQ587550 series, and GQ429137 respectively) suggesting that our *C. grandis* was correctly identified. This raises the question as to the identity of the species from which the FJ009819/FJ009820 sequence was obtained. The trnH-psbA sequence has all the Cassia features, and most likely corresponds to that of a *Cassia* species but needs further identification. Another conflicting case is the sequence identity between the *S. sophera* (JQ301835) and the *S. spectabilis* (SS) ITS2 marker: *S. sophera* is native from Asia while *S. spectabilis* is native to North America.

The (near) sequence identity between *S. artemisioides*, *S. petiolaris* and *S. eremophila* is not surprising since *S. petiolaris* is considered a subspecies of *S. artemisioides* (*http://bie.ala.org.au/search?q=Senna+petiolaris*), while *S. eremophila* (sensu auct) is a misapplied name that has been erroneously used in the past to refer to *C. artemisioides* DC. and which may have been used for identification. Nor is the identity between the ITS2 marker sequence of the three species *S. sulfurea*, *S. surattensis*, and *C. glauca* surprising. *S. sulfurea* is an accepted name while *C. glauca* is a synonym, as is *C. surattensis* subsp. *glauca* (Lam.) K.Larsen & S.S.Larsen. *C. surattensis* auct. has also been used in the past to refer to *S. sulfurea* (*http://www.theplantlist.org/tpl/record/ild-29621*). It is worth noting though that *Cassia glauca* and *Senna glauca* are considered to be different species, the latter being a synonym of *S. timoriensis* (DC) H.S.Irvin & Barneby.

Inconsistencies are also noted in the ITS2 sequences of the three species in the *Senna* subclade 3: *Senna italica* (JQ301844), *Senna pallida* (JQ301829), and *Chamaecrista absus* (JQ301843), with one nucleotide difference between *S. italica* and the other two species. The three species are native to different continents (Africa, Central and South America for *Senna italica*; North and Central America for *Senna pallida*; and Africa, Australia, Asia for *Chamaecrista absus*), and the position of *Chamaecrista absus* in

this clade strongly suggests that this species has been misidentified. It may be worth noting that the three sequences have been submitted by the same laboratory and a possibility of an error during data submission is not excluded.

Conclusions

I have been able to unambiguously identify the African plant using the DNA barcoding technology. It is *Cassia abbreviata*, a species native to tropical Africa, and of which 3 subspecies have been described: subsp. *abbreviata*, subsp. *beareana*, and subsp. *kassneri*. The analysis of the barcode markers ITS2 and trnH-psbA spacer of over 50 *Senna* and *Cassia* species revealed genus-specific features (insertions/deletions) for each of the two markers, thereby allowing to differentiate between the *Cassia* and the *Senna* genus. Highly variable sequences in the two markers allow to discriminate between the vast majority of the species in the *Senna* as well as the *Cassia* genus, thereby establishing the two markers as useful barcode markers for the species of the two genera. A not insignificant number of species whose ITS2 and trnH-psbA spacer sequences have been deposited in public databases appear to have been misidentified or misnamed by the authors.

We cannot guarantee that all our seeds came from correctly identified species. Most of the species we studied however share identical marker sequences with same-name species in databases and are therefore most probably correctly identified and named. A reliable database established by expert traditional and molecular taxonomists is an absolute prerequisite for correct species identification by DNA barcoding. In this sense the Consortium for the Barcode of Life (CBOL) is a welcome initiative. Unfortunately however the CBOL markers for plant identification only include the chloroplast coding sequences rbcL and matK which, in our hands, were more difficult to amplify than the non-coding ITS2 and trnH-psbA markers.

References

CBOLPlantWorkingGroup. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America,* 2009, 106, 12794-12797

Chen, S., Yao, H., Han, J., et al. Validation of the its2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one,* 2010, 5, e8613

Hebert, P. D., Cywinska, A., Ball, S. L., et al. Biological identifications through DNA barcodes. *Proceedings. Biological sciences / The Royal Society,* 2003, 270, 313-321

Ma, X. Y., Xie, C. X., Liu, C., et al. Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast psba-trnh intergenic region. *Biological & pharmaceutical bulletin,* 2010, 33, 1919-1924

Marazzi, B., Endress, P. K., Queiroz, L. P., et al. Phylogenetic relationships within senna (leguminosae, cassiinae) based on three chloroplast DNA regions: Patterns in the evolution of floral symmetry and extrafloral nectaries. *American journal of botany,* 2006, 93, 288-303

Mentewab, A. B., Jacobsen, M. J. and Flowers, R. A. Incomplete homogenization of 18 s ribosomal DNA coding regions in arabidopsis thaliana. *BMC research notes,* 2011, 4, 93

Purushothaman, N., Newmaster, S. G., Ragupathy, S., et al. A tiered barcode authentication tool to differentiate medicinal cassia species in india. *Genetics and molecular research : GMR,* 2014, 13, 2959-2968

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., et al. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America,* 1984, 81, 8014-8018

Song, J., Shi, L., Li, D., et al. Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. *PloS one,* 2012, 7, e43971

Part 2. Natural compounds derived from *Cassia abbreviata* prevent HIV-1 entry and infection

Yue Zheng[1]*, Xianwen Yang[1], Dominique Schols[2], Manuel Counson[3], Andy Chevigné[3], Gilles Iserentant[3], Martin Mulinge[1], Jean-Claude Schmit[3,4], André Steinmetz[1] and Carole Seguin-Devaux[3]

[1] Laboratory of Cellular and Molecular Oncology, Luxembourg Institute of Health, Luxembourg

[2] Laboratory of Virology and Chemotherapy, Department of Microbiology and Immunology, Rega Institute for Medical Research, KU Leuven, 3000 Leuven, Belgium

[3]Department of Infection and Immunity, Unit of Infectious Diseases, Luxembourg Institute of Health, Esch.sur-Alzette, Luxembourg

[4]Service National of Infectious Diseases, Centre Hospitalier de Luxembourg, Luxembourg.

* corresponding author

yue.zheng@lih.lu

tel: + 352 26 97 330

Abstract

Natural products represent rich sources of bioactive compounds with a strong value in therapeutics. *Cassia abbreviata* is widely used in Africa for treating many diseases including HIV infection but its active components remain unknown. In this study, we showed that the active components of a crude extract of *Cassia abbreviata* inhibit HIV entry by targeting gp120 and preventing its binding with the cellular receptor CD4. Fifty-seven compounds were isolated and seven compounds retained anti-HIV activity. Oleanolic acid, palmitic acid, taxifolin, piceatannol, and three other compounds presented an unknown structure. Among the characterized compounds, piceatannol showed the best anti-HIV activity profile when evaluated against various HIV-1 strains (IC_{50} values ranging from 8.04 to 47.46 µM, CC_{50}>300 µM). Unlike CE, piceatannol functions as a non-specific inhibitor against HIV, via potential interaction with the cell membrane. Finally, piceatannol inhibited HIV infection in an *in vitro* dual-chamber assay mimicking the female genital tract as well as HSV infection emphasizing its potential as a microbicide. The structure of the natural products isolated from *Cassia abbreviata* may be used to provide lead compounds of therapeutic relevance in human viral infection.

Key words: Natural products, Cassia abbreviata, HIV entry, HSV, microbicide.

Introduction

The Human Immunodeficiency Virus (HIV) is the cause of the Acquired Immune Deficiency Syndrome (AIDS), leading to a progressive failure of the immune system, and to life-threatening opportunistic infections and cancers to thrive. In 2013, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that there were nearly 35 million people living with HIV (http://www.unaids.org/). The vast majority of these people are located in low- and middle-income countries, particularly in Sub-Saharan Africa. Combined antiretroviral therapy (cART) has significantly decreased both the mortality and morbidity of AIDS, has largely improved the life of HIV-infected patients [1] and was the leading factor in diminishing the number of new HIV-infected cases worldwide. Despite the considerable success of cART, these drugs are still facing a lot of challenges for life-long adherence such as drug resistance, side effects, and high financial cost [2]. For economic and cultural reasons, women are more prone to HIV infection than men in these countries [3]. Therefore, development of new anti-HIV drugs as well as potent and safe microbicides are still required to facilitate treatment as prevention in the absence of an effective HIV vaccine.

Since the beginning of the epidemic, around 30 drugs were released for clinical practice, inhibiting HIV-1 at different stages of the HIV life cycle. Nevertheless, some treatment-experienced patients have viral mutants resistant to multiple drugs, drug toxicities resulting from long-term antiretroviral therapy and contraindications limiting seriously their treatment options. Novel therapeutic strategies could take advantage of the multiple events involved in the entry process [4]. HIV-1 attaches to cell membrane, engaging its surface envelope glycoprotein gp120 to bind successively to the CD4 receptor and either to the CXCR4 or CCR5 co-receptor. These initial events trigger conformational changes in HIV-1 envelope glycoproteins (gp120 and gp41) leading to the membrane fusion process [5]. New compounds targeting HIV entry through interactions with cellular proteins/membranes or viral membranes might be less prone to selecting for drug resistance than compounds targeting enzymes involved in the process of HIV replication [6] and could display additionally broad antiviral effects.

In the past decades, a high number of drugs have been developed directly and indirectly from natural products and their derivatives, which have shown their critical power in medical therapies, such as the pain killer morphine, the anti-malarial drug artemisinin, and the anticancer drug paclitaxel [7]. Compared to other therapies, natural products are less toxic to organisms, less sensitive for drug resistance, and much less

costly. *Cassia abbreviata* is a tropical tree, indigenous to South-East of Arica, widespread in African countries and commonly used in the local medicines [8]. Roots, barks and leaves of *Cassia abbreviata* are taken orally as decoction or chewed, for healing abnormal pain, fever, cough, snake bite, malaria, blood vomit, rash, syphilis, diarrhea, gonorrhoea and in particular for HIV infection [9]. These properties may well be attributed to a variety of compounds including alkaloids, tannins, anthraquinones, and flavonoids [10]. Nevertheless *Cassia abbreviata*'s antiviral activity has not been well studied.

In the present study, we have investigated the anti-HIV activity of *Cassia abbreviata's* crude extract and isolated fifty-seven compounds. Several compounds including oleanolic acid, palmitic acid, piceatannol and taxifolin were identified as inhibitors of HIV entry. The active components of the crude extract (CE) targeted the viral protein gp120 whereas piceatannol prevented viral entry through a potential interaction with cell membrane and was efficient in a dual chamber system mimicking the female genital tract.

Results

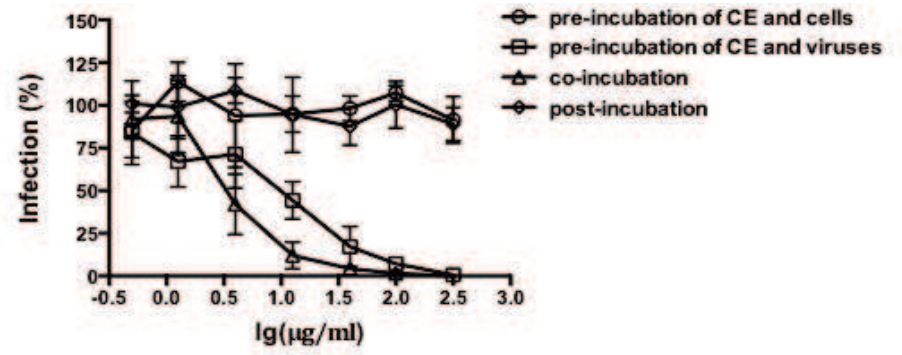
Crude extracts and purified compounds of *Cassia abbreviata* inhibit HIV entry

The anti-HIV activity of *Cassia abbreviata*'s crude extract (CE) was first confirmed on MT4 cells and on human peripheral blood mononuclear cells (PBMCs). CE inhibited HIV-1 infection in MT4 cells infected with the reference strains HIV-1 IIIB (CXCR4 tropic, X4, IC_{50} = 21.75 ±1.20 µg/ml) and HIV-2 ROD (IC_{50} = 34.2 µg/ml) at non-toxic concentrations (CC_{50} was higher than 1000 µg/ml, data not shown) as well as in PBMCs infected with HIV-1 IIIB (X4, IC_{50} = 40.77 ± 4.04 µg/ml), HIV-1 ADA-M (CCR5 tropic, R5, IC_{50} = 13.53±12.99) and 2 HIV-1 clinical isolates carrying several drug resistance mutations to nucleoside/nucleotide reverse transcriptase inhibitor (NRTI), non-nucleoside reverse transcriptase inhibitor (NNRTI) and protease inhibitors (PI) (IC_{50} of 23.06±5.92 µg/ml and 10.47±2.09 µg/ml) without displaying any cytotoxicity on PBMCs by FACS analysis (data not shown).

We first sought to characterize at which step of infection the CE of *Cassia abbreviata* inhibited HIV and whether CE contained compounds acting on the virus or on the host cells. Multi-dosing time assay experiments were performed with pseudotype viruses

pNL4.3Δ*Env*Luc-HXB2 and pNL4.3Δ*Env*Luc-BAL allowing a single virus life cycle. The inhibitory effect of CE was apparent when CE was added before infection when incubated with the virus and not the cells, and when incubated during infection but not after infection (Fig 33). These results indicate that CE inhibited HIV infection at an early stage of the HIV entry process independently of co-receptor usage (IC_{50} HXB2 = 13.37±8.46 µg/ml and IC_{50} BAL = 67.40±4.21 µg/ml when added during infection) and interact with the virus (IC_{50} HXB2 = 2.57±0.05 µg/ml, IC_{50} BAL= 11.41±1.63 µg/ml when pre-incubated with the virus).

A.



B.

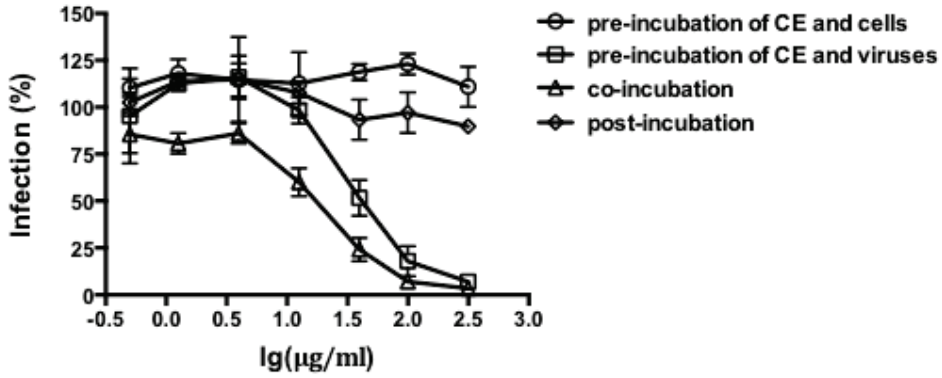


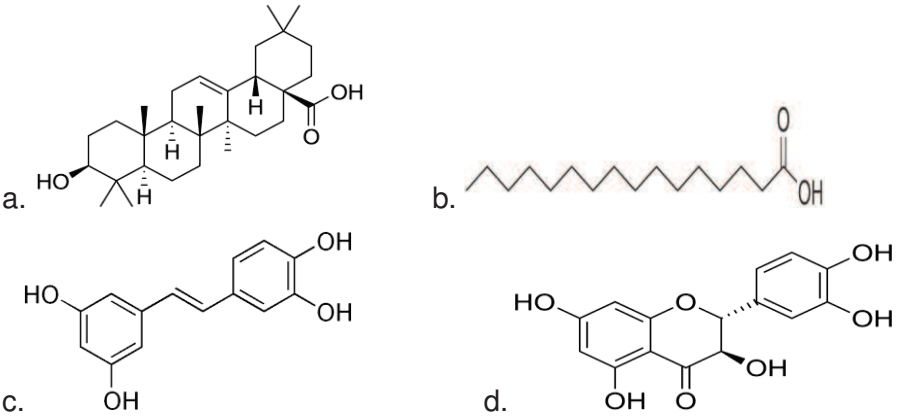
Figure 33. Crude extract of *Cassia abbreviatta* inhibits HIV entry into cells

CE was tested on multi-dosing time assay **(a)**: using U373-CD4-CXCR4 cells against pseudotype virus pNL4.3Δ*Env*Luc-HXB2 and **(b)** using U373-CD4-CCR5 cells against pseudotype virus pNL4.3Δ*Env*Luc-BAL. CE was added at different points of HIV infection, either 2h before infection (pre-incubation with cells or viruses), or during infection (co-incubation), or 2h after infection (post-incubation). CE inhibited HIV infection when co-incubated with cells and viruses and pre-incubated with viruses. Three independent experiments were performed in triplicates.

We next purified and identified fifty-seven compounds from *Cassia abbreviata* and assessed their anti-HIV activities on U373-CD4-CXCR4 and U373-CD4-CCR5 cells

infected with pseudotpye viruses when the compounds were added only at the time of infection to identify the compounds targeting HIV entry. Oleanolic acid, palmitic acid, piceatannol and taxifolin (Fig 34A) were identified as pure compounds. Oleanolic acid, palmitic acid and piceatannol inhibited both X4 and R5 infection as previously described (IC_{50} of 34.87±9.09 µM, 87.48±16.12 µM and 10.28±5.74 µM for X4 infection, respectively and IC_{50} of 288.5±8.20 µM, 639.6 ±5.20 µM, and 47.46±6.52 µM for R5 infection) while taxifolin inhibits only X4 infection (Fig 34B, IC_{50} = 239.92±3.21 µM). Among other purified compounds, three uncharacterized compounds inhibit HIV infection (IC_{50} around 20 µg/ml for R5 viruses and around 50 µg/ml for X4 viruses) whereas the rest of the compounds did not show any anti-HIV activity.

A.



B.

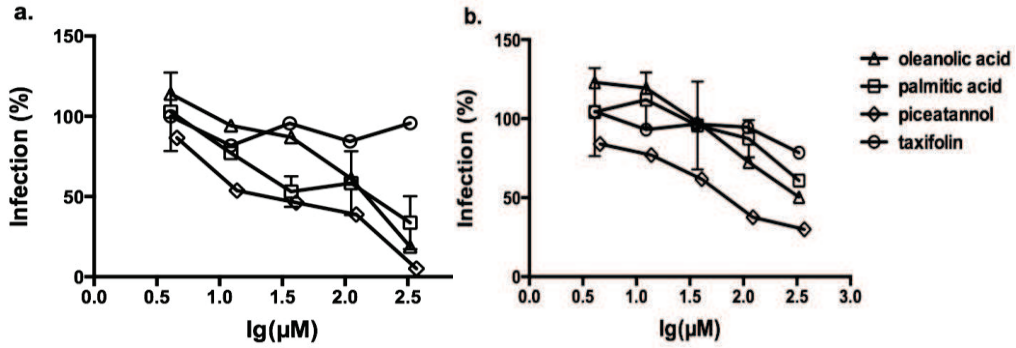


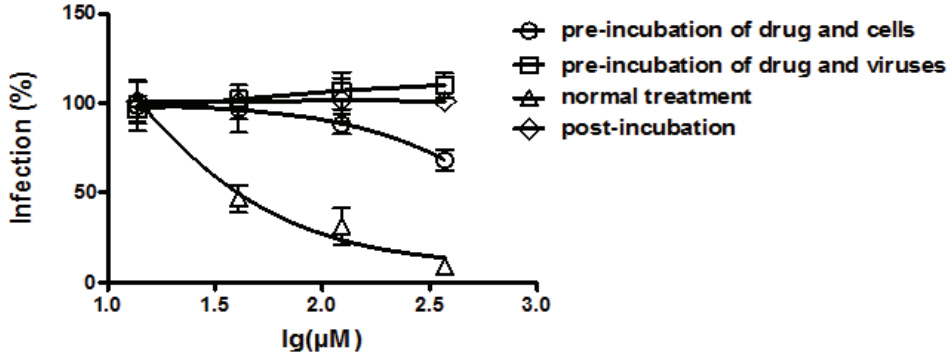
Figure 34. Anti-HIV activity of oleanolic acid, palmitic acid, piceatannol and taxifolin. A. Chemical structures. **(a)** oleanolic acid; **(b)** palmitic acid; **(c)** piceatannol; **(d)** and taxifolin. **B. Evaluation of HIV activity**. Compounds were tested on **(a)** U373-CD4-CXCR4 cells infected with pseudotype virus pNL4.3Δ*Env*Luc-HXB2 and **(b)** U373-CD4-CCR5 cells infected with pseudotype virus pNL4.3Δ*Env*Luc-BAL. Piceatannol showed the strongest anti-HIV activity (a. IC_{50}=10.28±5.74 µM; b. IC_{50}=47.46±6.52 µM; CC_{50}>300µM). Three independent experiments were performed in triplicates.

Crude extract of *Cassia abbreviata* and piceatannol have a synergistic effect with HIV entry inhibitors

Among the characterized compounds, piceatannol showed the strongest anti-HIV activity on U373-CD4-CXCR4/CCR5 cells (IC_{50} HXB2 was 10.28±5.74 µM and IC_{50} BAL was 47.46±6.52 µM), on PBMCs against HIV-1 IIIB and HIV-1 ADA-M (IC_{50} = 24.22 ±7.13 µM and IC_{50} = 19.91±0.22 µM, respectively, and two HIV-1 primary clinical isolates harboring multi-drug resistance to NRTI, NNRTI and PI (IC_{50} of 37.72± 12.54 and 8.04 ± 3.07 µM) without showing any cytotoxic effects (CC_{50} higher than 300µM, data not shown). Since the role of piceatannol on HIV entry was not as previously reported, we tempted to decipher at which step of the HIV replication cycle piceatannol functions using the multi-dosing time assay (Fig 35A). Similarly to CE, piceatannol inhibited viral infection when added at the time of infection other than when added after infection. In contrast to CE, piceatannol weakly inhibited HIV infection at high concentration when it was pre-incubated with cells before infection but not with the virus. This indicates that piceatannol may target the cells and may have a different anti-HIV mechanism than the other compounds contained in the CE of *Cassia abbreviata* targeting mainly the virus. Additionally, both CE and piceatannol showed a synergistic effect with the HIV entry inhibitors Enfuvirtide (T-20, fusion inhibitor) and AMD3100/Maraviroc (CXCR4 and CCR5 co-receptor inhibitors) (Fig 35B) suggesting that both CE and piceatannol act on synergy with the current entry inhibitors and that their antiviral mechanisms might differ. To confirm these data, we performed fusion assays between Hela-P4-CXCR4 cells and Hela-Env-Lai cells. We did not find any effect of piceatannol on the fusion process whereas CE of *Cassia abbreviata*, T20 and AMD3100 showed a significant inhibition (data not shown).

A.

a.



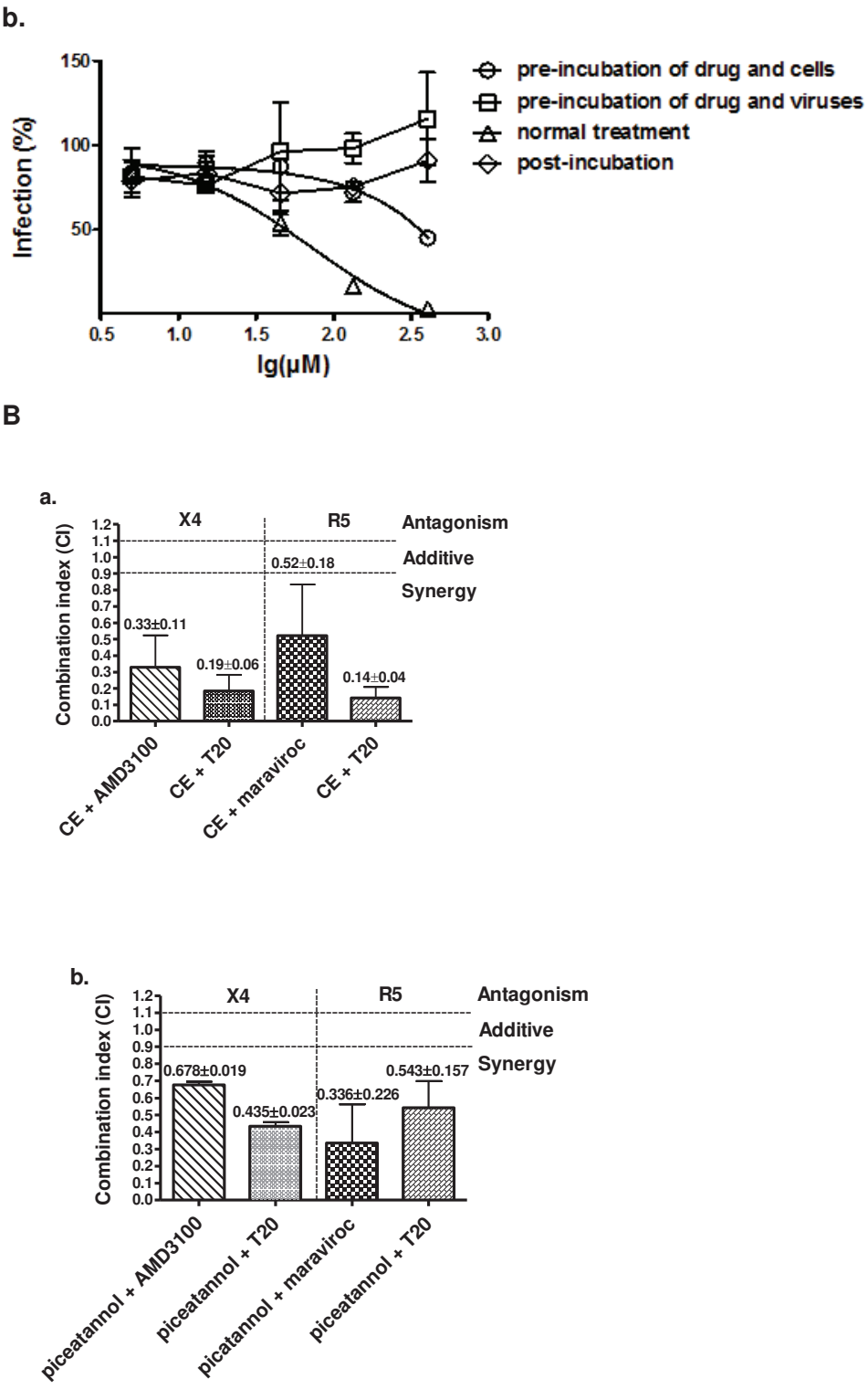


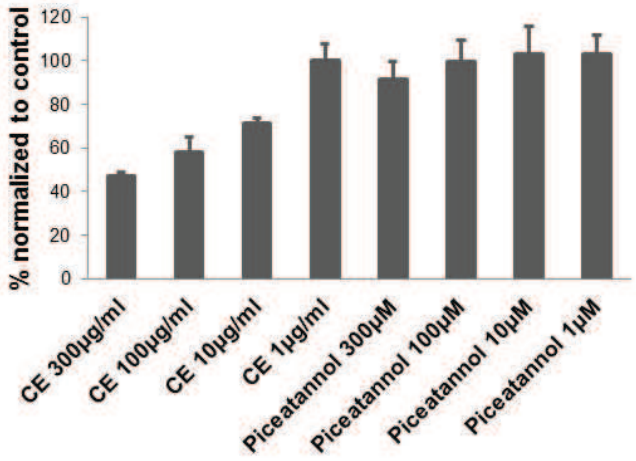
Figure 35. Piceatannol inhibits HIV entry into cells. A. Effects of Piceatannol in a multi-dosing time assay. Multi-dosing time assay was performed **(a)**: using U373-CD4-CXCR4 cells against pseudotype virus pNL4.3Δ*Env*Luc-HXB2 and **(b)** using U373-CD4-CCR5 cells against pseudotype virus pNL4.3Δ*Env*Luc-BAL. Piceatannol was added either 2h before infection (pre-incubation with cells or viruses), or during infection (co-incubation), or 2h after infection (post-incubation). Piceattanol inhibited

HIV infection when co-incubated with cells and viruses and when pre-incubated with the cells only at the highest concentrations. Three independent experiments were performed in triplicates. **B. Synergistic activity of the crude extracts of *Cassia abbreviata* and piceatannol with anti-HIV entry inhibitors**. CE **(a)** and piceatannol **(b)** were combined with co-receptor inhibitors (AMD3100 for CXCR4, maraviroc for CCR5) and the fusion inhibitor enfuviritide (T20) respectively, and tested on U373-CD4-CXCR4/U373-CD4-CCR5 cells against pseudotype particles pNL4.3Δ*Env*Luc-HXB2/BAL. Combination Index (CI) at 95% maximal effective concentration (EC_{95}) level was calculated using CompuSyn (ComboSyn, USA). According to Chou and Talalay's method, CI >1.1 means antagonism, CI<0.9 means synergy and 0.9<CI<1.1 means additive effect of the drugs. Data are shown as mean±SD of three independent experiments performed in triplicates

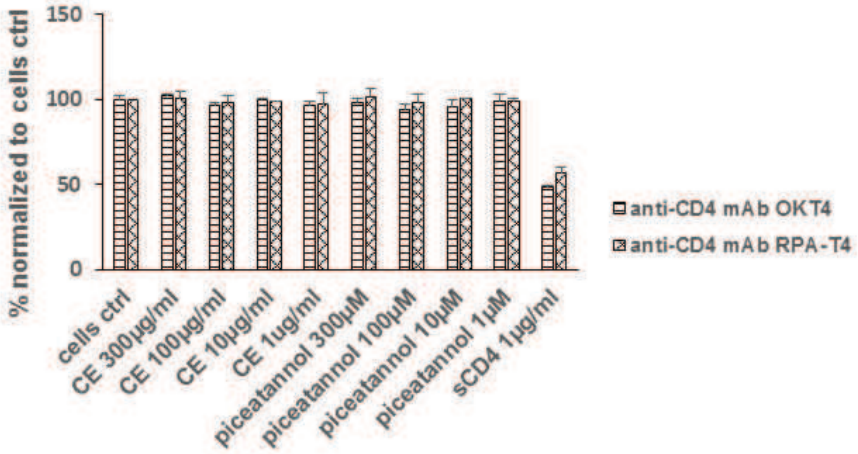
Piceatannol does not bind to the viral protein gp120 nor to the CD4, CXCR4, CCR5 receptors while CE of *Cassia abbreviata* affects gp120/CD4 interaction

During the HIV entry process, the virus binds successively to the CD4 cellular receptor and to the co-receptors CXCR4/CCR5. Therefore we first tested whether CE and piceatannol could affect the binding of gp120 with the CD4 receptor. As shown in Figure 36A, CE inhibited the interaction between CD4 and gp120 in a dose-dependent manner, while piceatannol did not. No significant differences were further observed in CD4 binding (Fig 36B) and in CXCR4/CCR5 binding (Fig 36C) between the cells incubated with or without drugs while sCD4 and the chemokines CXCL12 and CCL5 inhibited CD4 and CXCR4/CCR5 binding in this assay, respectively. Internalisation assays confirmed that the compounds did not induce receptor internalisation or inhibit chemokine-induced receptor internalisation (data not shown). In line with this observation, we observed that piceatannol but not CE inhibit the infection of pseudotype particles of vesicular stomatitis virus (VSV) G proteins pNL4.3ΔEnvLuc-VSVG (IC50=79.23±17.20µM, Table 9) supporting that piceatannol inhibits viruses entry by preventing the virus attachment to the cells in a non-specific manner whereas the main active components of CE more specifically target gp120. In addition CE of *Cassia abbreviata* and piceatannol were further tested on other viruses: CE and piceatannol showed anti-viral activity on HSV-1 (herpes simplex virus) and HSV-2 while having no effect on influenza (H1N1, H3N2, B), para-influenza, hepatitis C virus, coxsackie virus, respiratory syncytial virus, reovirus, sindbisvirus, punta toro virus, yellow river virus, feline corona virus, and feline herpes virus (Table 9).

A.



B.



C.

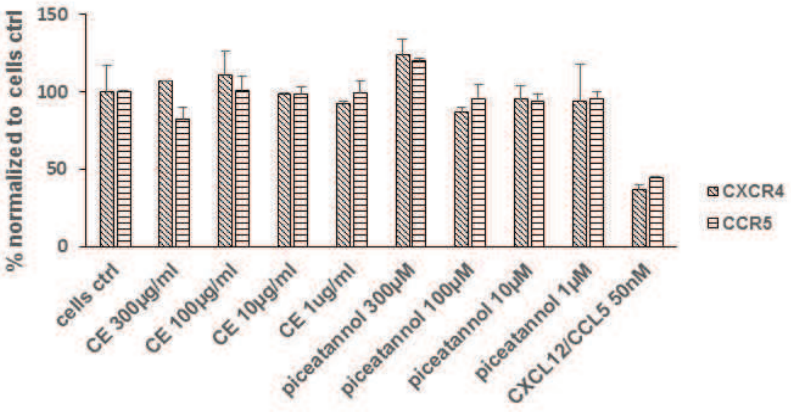


Figure 36. Binding activity of Cassia abbreviata's crude extract and piceatannol A. CE but not piceatannol affects gp120 and CD4 interaction. Human soluble CD4 coated plate was first incubated with CE or piceatannol for 1h, and then with HIV gp120 protein for another 1h. Anti-HIV-1 gp120 antibody was used to detect CD4-gp120 interaction. Value of OD_{492}-OD_{630} was calculated. Data were normalized to control

(without CE/piceatannol treatment).**B. CE and piceatannol do not bind to the CD4 receptor.** U373-CD4-CXCR4/CCR5 cells were incubated with either anti-CD4 antibody OKT4 or RPA-T4 in the presence of CE and piceatannol. sCD4 were used as a positive control for binding with U373-CD4-CXCR4 and U373-CD4-CCR5 cells. **C. CE and Piceatannol do not bind to CXCR4 or CCR5 co-receptors.** U373-CD4-CXCR4/CCR5 cells were incubated with either the anti-CXCR4 antibody 12G5 or the anti-CCR5 antibody 2D7 in the presence of CE and piceatannol. The chemokines CXCL12 and CCL5 were used as positive controls for binding with U373-CD4-CXCR4 and U373-CD4-CCR5 cells, respectively. Signals were measured by flow cytometry. Data was normalized to the MFI of the control cells. Three independent assays were performed.

Table 9. Evaluation of CE and piceatannol against various classes of viruses

Cells	Viruses	CE (µg/ml)	Piceatannol (µM)
HEL	Herpes simplex virus-1 (KOS)	46.7 ± 2.9	47.5 ± 3.5
HEL	Herpes simplex virus-2 (G)	39.5 ± 5.5	45.0 ± 0.0
HEL	Herpes simplex virus-1 TK KOS ACV	45.0 ± 0.0	45.4 ± 0.0
U373	Vesicular stomatitis virus	>100	79.2± 17
MDCK	Influenza A/H1N1 A/Ned/378/05	>100	>100
MDCK	Influenza A/H3N2 A/HK/7/87	>100	>100
MDCK	Influenza B B/Ned/537/05	>100	>100
Huh 7-D	Hepatitis C virus	>100	>100
hela	Coxsackie virus B4	>100	>100
vero	Coxsackie virus B4	>100	>100
Hela	Respiratory syncytial virus	>100	>100
vero	Para-influenza-3 virus	>100	>100
vero	Reovirus-1	>100	>100
vero	sindbisvirus	>100	>100
vero	Punta toro virus	>100	>100
vero	Yellow fever virus	>100	>100
CRFK	Feline corona virus (FIPV)	>100	>100
CRFK	Feline herpes virus	>100	>100
HEL	Human corona virus	>100	>100

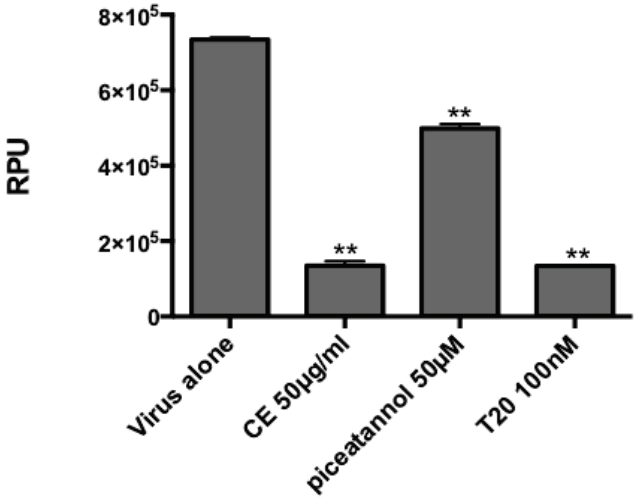
IC_{50} (inhibitory concentration of viral replication by 50%).

CE, crude extract of *Cassia abbreviata*,

Microbicide activity of piceatannol in a dual-chamber system

To see the potential of piceatannol as a microbicide, CE and piceatanol were tested in a dual-chamber system mimicking the epithelium female genital tract. In this assay, HeLa cells (human cervical epithelial carcinoma cells) were grown in the upper chamber, until a confluent layer of a minimum of 150 Ohm/cm^2 was achieved while TZM-Bl target cells were infected in the lower chamber. As shown in Figure 37A, both CE and piceatannol could cross the layer of epithelial cells and inhibit HIV infection in the lower chamber without affecting the confluence layer's TEER. At a concentration around the IC_{50} of the compounds, CE decreased by more than 4 times HIV infection (from 100 to 18.41%) while piceatannol diminished by 33% (from 100 to 66.9 %) without any toxicity on the epithelial cells (data not shown). For potential microbicide application, anti-HIV compounds should not stimulate target cells. We next evaluated the expression of early and late activation markers when PBMCs were incubated with a series of concentrations of CE and piceatannol during 24 hours. PHA increased the expression of CD25 (early activation marker) and CD69 (late activation marker) from 6% to 34% and from 5 % to 62%, respectively, whereas CE and piceatannol did not have any effect alone. (Fig 37B). Finally, we tested whether CE and piceatannol could prevent transmission of DC-SIGN (dendritic cell-specific ICAM-3 grabbing non integrin) captured virus to CD4$^+$ T cells in a DC-SIGN-mediated HIV transmission assay: at the highest concentration evaluated 100 µg/ml CE and 50 µM piceatannol did not protect CD4$^+$ T cells from subsequent HIV replication suggesting that these compounds could not block this HIV DC-SIGN-T cell mediated infection pathway.

A.



B.

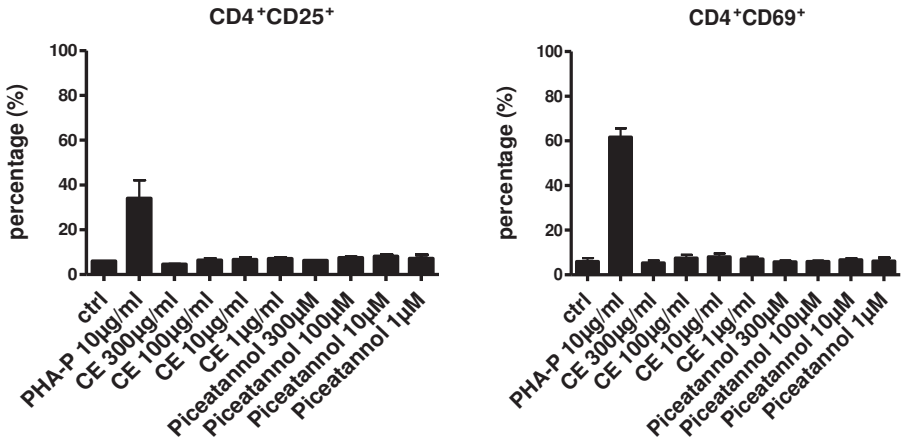


Figure 37. A. CE and piceatannol inhibit HIV infection in a dual chamber system. CE (50 µg/ml), piceatannol (50µM) and T20 (100nM) were added in the upper chamber to HeLa cells with HIV-1 ADA-M (200pg P24) when TEER (Trans Epithelial Electric Resistance) values reached a minimum of 150 Ohm/cm^2. After 24h, luciferase value of TZM-Bl cell lysate was measured. **B. CE and piceatannol do not activate PBMCs.** PBMCs were treated with CE, piceatannol or PHA-P for 24h. The expression of CD4, CD25 and CD69 were measured by flow cytometry.

Discussion

Traditional medicine has been used for thousands of years in many countries where it is still commonly used because of its significant therapeutic effects. Many antiviral agents have been identified from plant sources but their mechanisms have been only partly characterised. *Cassia abbreviata* is indigenous to Africa, where it is commonly used to heal abnormal pain, fever, cough, snake bite, blood vomit, rash, diarrhea, as well as infectious diseases such as malaria, syphilis, gonorrhoea and HIV/AIDS [9,10]. Nevertheless, *Cassia abbreviata*'s bioactive mode-of-actions were not investigated. Our study provided for the first time in-depth investigation of its active components as well as insight into their anti-HIV mode of action.

Our study first confirmed the anti-HIV activity of *Cassia abbreviata*, which has been already described by Leteane's group showing a weak anti-HIV activity of an ethanol extract but failing to elucidate its anti-HIV mechanism [11]. By performing a secondary ethyl acetate partition of the ethanol extraction, we obtained a crude extract of *Cassia abbreviata* (CE) from the aqueous phase displaying a better anti-HIV activity without any cell toxicity. HIV entry involves multiple steps that require both viral and cellular

membrane proteins. After attachment, the HIV-1 envelope glycoprotein gp120 binds to the cellular receptor CD4 and subsequently to the co-receptor CCR5/CXCR4, resulting in a further change of conformation to expose the fusion protein gp41 and to mediate the fusion process [12]. According to our findings, the active components of CE inhibited both X4 and R5 tropic viral strains, target HIV-1 and not the cells, had no effect on the availability and binding activity of CD4, CXCR4 and CC5 expressed on the cell surface, and inhibit the interaction between gp120 and CD4. CD4 and co-receptor binding sites as well as variable loops and glycans of gp120 were proposed as therapeutic targets (Acharya et al, 2015). For instance, several small-molecule attachment inhibitors targeting the conserved CD4 binding region within gp120 have been described [13-17]. BMS-626529 and its phosphonooxymethyl ester prodrug BMS-663068 are currently the most promising HIV attachment inhibitors blocking CD4-gp120 interaction [19]. Using homology models, BMS-626529 was proposed to bind to the unliganded conformation of gp120 within the structurally conserved outer domain, under the antiparallel b20–b21 sheet, and adjacent to the CD4 binding loop. This drug exhibits a unique mechanism of action by inhibiting both CD4-induced and CD4-independent formation of the “open state” four stranded gp120 bridging sheet. We assume that the active compounds of CE display some specific mechanisms on gp120 conformation or binding with CD4.

To further investigate the anti-HIV components of *Cassia abbreviata*, we succeeded to isolate fifty-seven compounds. These compounds were all assessed in an assay targeting HIV entry, as well as in an assay targeting the late stage of HIV infection (data not shown). Among them, oleanolic acid, palmitic acid, piceatannol, taxifolin, and three so-far-unidentified compounds were found to inhibit HIV entry, while none of the compounds were found to inhibit a later stage of HIV infection. Oleanolic acid has been reported to inhibit HIV-1 infection on H9 cells [20], PBMCs and macrophages (IC_{50} ranging between 20 to 50 µM) [21]. The authors have proposed oleanolic acid as a protease inhibitor using a non-cell-based assay. In our study, we observed a similar IC_{50} for oleanolic acid against HIV-1 but we proposed a different mechanism since we did not find any effect of oleanic acid when incubated after infection (data not shown). Canki et al. has shown that palmitic acid isolated from *Sargassum fusiforme* inhibited viral infection on T cell line, primary peripheral blood lymphocytes and macrophages via binding to CD4 [22]. They have determined the palmitic acid binding epitope for CD4 containing a hydrophobic methyl and methylene groups located away from its carboxyl terminal which was responsible for blocking CD4-gp120 interaction. Taxifolin isolated from *Juglans mandshurica* has shown to protect MT-4 cells against HIV infection with

a 100% inhibitory concentration of 25 µg/ml [23] but its mechanism of action was not studied. Since oleanolic acid, palmitic acid and taxifolin exhibited a relative weak anti-HIV activity in our hands, we further focused our work on piceatannol's effect on HIV entry.

Clouser et al. have reported that piceatannol inhibited HIV-1 replication with an IC_{50} at 21.4 µM in agreement with our results [24]. Although Tewtrakul et al. have isolated piceatannol from *Cassia garrettiana* and claimed it as an HIV-1 integrase inhibitor and a protease inhibitor using non-cell-based assays [25-26], we currently demonstrated, using several viral assays that piceatannol inhibited HIV entry. Taking into consideration that piceatannol did not affect neither CD4, CCR5/CXCR4 binding, nor the interaction between CD4 and gp120 as well as the fusion process, it is therefore tempting to speculate that piceatannol may interact with virus attachment by adsorbing at either the cell surface or the virus surface. Resveratrol and its analogue piceatannol were described to interact preferentially with headgroup region of lipid bilayer in model membranes composed of phosphatidylcholine [27]. Piceatannol showed a preference to bind to the cells as compared to the viruses but only at concentrations above the IC_{50} value. This could result of a low affinity of piceatannol with the cell membrane explaining that the effect is maximal only when the compound is added together with cells and viruses. Piceatannol also showed a synergistic effect with both co-receptor and fusion inhibitors, which indicates that piceatannol employs a different mechanism to prevent HIV entry. This hypothesis is confirmed with the observation that piceatannol inhibited VSV infection since VSV enters into target cells by endocytosis and not by interfering with a specific cellular receptor.

Structurally, piceatannol (3,3',4,5'-trans-trihydroxystilbene) is a natural analogue of resveratrol (3,4',5'-trans-trihydroxystilbene). Resveratrol has shown its great therapeutic effect on cancers in clinics and in cardio- and neuroprotection but relative low activity on HIV by acting as a ribonucleotide reductase inhibitor [24]. Many analogues of resveratrol have been reported as well to inhibit HIV infection [28]. However, only resveratrol did potentiate the effect of decitabine, an inducer of lethal mutagenesis, to inhibit HIV-1 infectivity [24] and 3,3',4,4',5,5'-hexahydroxy-trans-stilbene was shown recently to inhibit HIV entry before the fusion step with an IC_{50} around 1 µM [29]. Both piceatannol and 3,3',4,4',5,5'-hexahydroxy-trans-stilbene display better anti-HIV activity than resveratrol suggesting that the additional hydroxyl groups to the basic stilbene rings may increase the anti-HIV activity by strengthening compound affinity or by facilitating its action on the membrane surface.

Curcumin is another antiviral compound displaying structural similarity to piceatannol, having two phenols connected by a carbon chain. Curcumin can affect viral membrane fluidity and can confer a positive curvature to block viral entry [30] such as rigid amphipathic fusion inhibitors (RAFIs) and LJ001 affecting the membrane of HIV and other enveloped viruses [6,31]. We have tested piceatannol pre-incubating either with viruses or with cells for 5 min at 37C before infection, similar to the assay described for curcumin [30], but we failed to observe a similar inhibitory effect as curcumin. We cannot exclude nevertheless that a lower affinity of piceatannol for the viral membrane as compared to curcumin may also explain this difference. Further experiments directly targeting membrane fluidity and curvature need to be performed.

Importantly, we have shown that both CE and piceatannol inhibited HIV infection in an *in vitro* dual-chamber model mimicking the epithelium of the female genital tract [32], suggesting that the active components can cross the epithelial barrier without any toxicity on epithelial cells and no further activation of PBMCs. This data indicates the potential of piceatannol to be used as a lead structure for microbicides although it did not prevent cell-to-cell and DC-SIGN-mediated viral transmission *in vitro*. The structure of more potent synthetic derivatives should be resolved to reach nevertheless a therapeutic efficacy in humans. Interestingly, at the IC_{50} concentration, CE of *Cassia abbreviata* achieved a similar potency as T20 suggesting that all the different anti HIV compounds from CE may harbor a synergistic effect, and among them piceatannol would represent only one piece. We will therefore elucidate the structure of our new anti-HIV compounds and assess whether these compounds target gp120. Lastly, we showed the broad spectrum antiviral activity of *Cassia abbreviata* and piceatannol for HIV and HSV but not for other viruses. The dual antiviral activity of piceatannol and CE is a crucial added value for microbicidal applications since genital HSV-2 infection has been shown to potentiate HIV transmission and infection [33].

In conclusion, we have shown that the main active components of *Cassia abbreviata* prevent HIV by targeting gp120. We have also isolated seven compounds blocking HIV entry and deciphered the mode of action of piceatannol. Ultimately, there is an ongoing need for new potent classes of antiretroviral drugs with improved safety and tolerability profiles to sustain long-life antiretroviral therapy. Optimized synthetic derivatives from piceatannol and from our new uncharacterized components could provide potent antiviral lead compounds in the near future.

Materials and methods

Plant extraction and compounds purification

Barks and roots of *Cassia abbreviata* were collected in Kenya. All materials were pulverized before extraction. The crude extract (CE) was obtained through a first extraction with 95% ethanol and a second extraction with ethyl acetate and dried. To purify the compounds, the concentrate of theethanol phase was suspended in deionized water, successively partitioned with $CHCl_3$, EtOAc, and *n*-BuOH and further subjected to column chromatography over silica gel. The $CHCl_3$ and EtOAc extracts were combined and subjected to column chromatography over silica gel eluting with a gradient $CHCl_3$-MeOH (0->100%), followed by column chromatography over octadecyl silane and sephadex LH-20. Preparative thin layer chromatography were further used to purify the compounds. UV and NMR data were collected from UV-2550 spectrometer and Bruker Avance 500 or 600 NMR spectrometers to characterise the compounds. Piceatannol was purchased from Sigma (Belgium). AMD3100 and maraviroc were obtained through NIH AIDS Reagent Program (Division of AIDS, NIAID, NIH). T-20 (Enfuvirtide) was purchased from Eurogentec (Belgium).

Cells and cell cultures

HEK293, MT-4, U373-CD4-CXCR4, U373-CD4-CCR5, HeLa, TZM-Bl cell lines were obtained through NIH AIDS Reagent Program. HeLa-P4-CXCR4-LTRLacZ and Hela-Env-Lai cells were kindly given by Pr Marc Alizon, Institute Pasteur, Paris. HEK 293 cells, HeLa cells and TZM-Bl cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Lonza, The Netherlands) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Lonza, The Netherlands). MT-4 cells were maintained in RPMI-1640 medium (Lonza, The Netherlands) containing 10% FBS, 2mM L-glutamine (Lonza, The Netherlands). U373-CD4-CXCR4 and U373-CD4-CCR5 cells were cultured in DMEM supplemented 10%FBS, 0.2mg/ml geneticin (Gibco, Germany), 0.1mg/ml hygromycin B (Invitrogen, Belgium) and 1µg/ml puromycin (Invivogen, France). HeLa-P4-CXCR4-LTRLacZ cells were cultured in DMEM containing 10% FBS, 2mM L-glutamine and 0.5mg/ml geneticin. Hela-Env-Lai cells were cultured in DMEM containing 10% FBS, 2mM L-glutamine and 2µM MTX-Amethopterin (Sigma, Belgium). Peripheral blood mononuclear cells (PBMCs) were isolated from healthy donors' buffycoats (Red Cross of Luxembourg) using Ficoll-Hypaque gradient. PBMCs were cultured in RPMI-1640 medium containing 10% FBS, 2mM L-glutamine. All other

cell lines including HEL, Vero, MDCK and CRFK were cultured in RPMI-1640 medium containing 10% FBS, 2 mM L-glutamine and 0.075 M $NaHCO_3$.

Cellular assays for HIV replication inhibition

MT-4 cells were incubated with CE and HIV-1 IIIB (100 TCID50) for 5 days. Viral inhibition and drug toxicity were evaluated in parallel using MTT ((3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Sigma, Belgium) by measuring A_{540} and A_{690} using a POLARstar Omega Plate Reader **(**BMG Lab Technologies, Belgium)**.** Value of OD_{540} - OD_{690} of each well was calculated. IC50 was calculated using GraphPrism version 6 (GraphPad software, Inc, USA) PBMCs were stimulated using 10µg/ml phytohemagglutinin (PHA-P, Sigma, Belgium) for 48h and then recombinant IL-2 (10U/ml, Roche, Germany) for another 24h before HIV infection. Stimulated PBMCs were infected by the HIV-1 reference strains IIIB/ADA-M or clinical isolates in the presence of CE replaced every other day after infection. Viral infection was monitored 7 days after infection by measuring P24 production in supernatants by-ELISA (Perkin Elmer, Belgium). U373-CD4-CXCR4/CCR5 cells were infected by pseudotype virus pNL4.3Δ*env*Luc-HXB2/BAL [34] by spinoculation at 1200g during 2h in the presence of CE or purified compounds removed just after infection. After 48h, the luciferase value of the cells lysate was measured (Luciferase System kit, Promega, The Netherlands). Cytotoxicity of drugs on PBMCs was measured by flow cytometry using near-IR fluorescent reactive dye (Life technologies, Belgium) and FACSCanto analysis (BD biosciences, Belgium). For multi-dosing time assay**,** several conditions were performed: 2h pre-incubation of drugs with U373-CD4-CXCR4/CCR5 cells, 2h pre-incubation of drugs with the pseudotype virus pNL4.3Δ*env*Luc-HXB2/BAL, 2h co-incubation of drugs, cells and viruses during the spinoculation step, and 2h post-incubation of drugs with infected cells. For drug combination assay, U373-CD4-CXCR4/CCR5 cells were co-incubated with CE or piceatannol combined or not with the CXCR4 inhibitor AMD3100 (Sigma, Belgium) or with the CCR5 inhibitor Maraviroc (Sigma, Belgium) or with the fusion inhibitor Enfuvirtide (T-20) during infection with the pseudotype viruses by spinoculation as described above. The combination index (CI) was calculated at the EC_{95}-level using CompuSyn software (ComboSyn, USA). Data were analysed using GraphPrism v6-0 (GraphPad software, Inc, USA). Fusion assay was performed between two HeLa cell lines expressing either CXCR4 (HeLa-P4-CXCR4-LTRLacZ) or X4 envelopes (Hela-Env-Lai cells). Drugs were incubated with Hela-P4-CXCR4-LTRLacZ cells during 30min and the mixtures were added to Hela-Env-Lai cells for 24h. Chlorophenol red-β -D-galactopyranoside (CPRG, Roche,

Germany) was added. OD_{562} was measured using a POLARstar Omega Plate Reader **(**BMG Lab Technologies, Belgium**).** IC50 were determined using GraphPrism v6-0.

Binding between gp120 and CD4

The binding of CE and piceatannol to gp120 was determined by ELISA. Soluble human CD4 was immobilized on 96-well plates and incubated for 1h with 100 ng HIV-1 gp120 protein (Fitzgerald, USA) incubated or not with increasing concentrations of the compounds. Bound gp120 was detected using a sheep anti-HIV-1 gp120 (Aalto Bio Reagents, Ireland) and an anti-sheep HRP. Optical densities were measured using a POLARstar Omega Plate Reader (BMG Lab Technologies, Belgium). Data were analysed using GraphPrism v6-0.

Flow cytometry assays

To investigate if CE and piceatannol could interact with the CD4 receptor and the HIV CXCR4 and CCR5 co-receptors, binding assays were performed by a FACSCanto flow cytometry. Binding competition assay between increasing concentrations of compounds and FITC-conjugated mouse anti-human CD4 clone RPA-4 (Biolegend, Belgium), PE-conjugated mouse anti-Human CXCR4 (CD184) clone 12G5 (BD Pharma, Belgium), and PE-conjugated mouse anti-human CCR5 (CD195) clone 2D7 (BD Pharma, Belgium) to U373-CD4-CXCR4 and U373-CD4-CCR5 cells was first tested. The chemokines CXCL12 and CCL5 (50 nM) (Peprotech, United Kingdom) were used as positive controls, near-IR fluorescent reactive dye (Life technologies, Belgium) was added simultaneously to evaluate cell viability. After 1 h of incubation at 4°C, cells were washed with FACS buffer (1% bovine serum albumin, 0.1% N3Na in PBS, Sigma, Belgium) and Mean Fluorescence Intensity was measured. Receptor internalisation assays were further performed: compounds were incubated alone (agonist mode) or in the presence of 50 nM chemokines (CXCL12/CCL5) (antagonist mode) with MT-4 and U373-CD4-CCR5 cells at 37°C for 1 h. Internalisation was stopped after 30 min by addition of NaN_3 (0.1%) and incubating the cells on ice. Cells were stained with anti-human CXCR4 (CD184) clone 4G10 (BD, Belgium) and anti-human CCR5 (CD195) clone: eBio T21/8, (eBioscience, Austria). Near-IR fluorescent reactive dye was added simultaneously. After 1h of incubation at 4°C, cells were washed with FACS buffer and a secondary antibody R-Phycoerythrin Affinipure F(ab')2 fragment goat anti-mouse IgG (H+L) (Jackson ImmunoResearch, United States) was added. Mean Fluorescence Intensity was measured by FACSCanto. Data were analysed using GraphPrism v6-0. The expression of the activation markers CD25 and CD69 was measured on PBMCs after incubation with CE, piceatannol or

phytohemagglutinin (PHA-P, Sigma) 10µg/ml for 24h at 37°C. PBMCs were stained with conjugated antibodies (anti-CD4 mAb conjugated with FITC, anti-CD25 mAb conjugated with PE/Cy7 and anti-CD69 mAb conjugated with PE, Biolegend, Belgium) and the near-IR fluorescent reactive dye (Invitrogen, Belgium). Mean Fluorescence Intensity was analysed using a FACSCanto flow cytometer. Cell activation was assessed by the expression of CD25 and CD69 on PBMCs after incubation with CE, piceatannol or PHA-P 10 µg/ml for 24 h at 37°C. PBMCs were stained with conjugated antibodies (anti-CD4 mAb conjugated with FITC, anti-CD25 mAb conjugated with PE/Cy7 and anti-CD69 mAb conjugated with PE, Biolegend, Belgium) and the near-IR fluorescent reactive dye (Invitrogen, Belgium) together for 30 min at 4°C. After washing, samples were analysed using FACSCanto.

Dual-chamber system and DC-SIGN transmission assay

Dual –chamber transmission assay was performed as previously described [32]. Briefly, HeLa cells (10^4 cells) were seeded into an upper chamber of a 24-well transwell plate (Corning, Belgium), while TZM-bl cells (10^5 cells) were seeded in the lower chamber. TEER (Trans Epithelial Electric Resistance) was measured using Millicell-ERS Volt-Ohm Meter every day as previously described. After 4 days, when TEER reached 150 Ohm/cm^2, HIV-1 ADA-M (200pg) and drugs were added to the upper chamber. 24h after infection luciferase value of TZM-bl cells lysate was measured using the Luciferase System Kit (Promega, The Netherlands) and the POLARstar Omega Plate Reader (BMG Lab Technologies, Belgium). Data were analysed using GraphPrism. For the HIV-1/DC-SIGN transmission assay Raji.DC-SIGN$^+$ cells ($5x10^5$ cells/200 µL) were exposed to high amounts of HIV-1 HE (100 µL; ~$2.5x10^6$ pg/mL of p24) for 1 h at 37°C. Test compounds were diluted in a 96-well plate (100 µL) and pre-incubated for 30 min. with CD4$^+$ T cells ($1x10^5$ cells/50 µL). The virus-exposed Raji.DC-SIGN cells were extensively washed, resuspended and cocultivated with the CD4$^+$ T cells and giant cell formation was evaluated light microscopically.

Broad spectrum antiviral assays

The protocols of the antiviral assays were previously described [35]. Briefly, viruses were tested on different cell lines; HEL cells for HSV-1 KOS, HSV-2 G, HSV-1 TK KOS ACV; hela cells for VSV, coxsackie virus B4 and RSV; vero cells: coxsackie virus B4, para-influenza-3 virus, reovirus-1, sindbisvirus, punta toro virus and yellow fever virus; MDCK cells for influenza A/H1N1 1/Ned/378/05, influenza A/H3N2 A/HK/7/87 and influenza B B/Ned/537/05; CRFK cells for human corano virus, feline corona virus and feline herpes virus; Huh-7D cells for HCV. Confluent cells were infected by 100 $CCID_{50}$,

and incubated with drugs. The viral CPE was recorded as soon as it reached completion in the virus control group. The minimal cytotoxicity was visualized using microscope on the cell morphology. For HCV, 6*103 cells were incubated with 0.5 MOI JC1 and drugs for 4 h. After washing, drugs were added to cells and incubated for 48 h. Luciferase was then measured using the luciferase system kit (Promega, the Netherlands).

References

1. Este, J. A. & Cihlar, T. Current status and challenges of antiretroviral research and therapy. *Antiviral Res.* 85, 25-33, (2010).
2. Palmisano, L. & Vella, S. A brief history of antiretroviral therapy of HIV infection: success and challenges. *Ann Ist Super Sanita.* 47, 44-48, (2011).
3. Glynn, J.R., et al. Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia. *AIDS*. Suppl 4 :S51-60, (2001).
4. Vigant, F., Santos, N.C. & Lee, B. Broad-spectrum antivirals against viral fusion. *Nature Reviews Microbiology*. 13, 426-437, (2015).
5. Wilen, C.B., Tilton, J. C. & Doms, R.W. HIV: cell binding and entry. *Cold Spring Harb Perspect Med*. 8, 1-13, (2012).
6. St.Vincent, M.R., et al. Rigid amphipathic fusion inhibitors, small molecule antiviral compounds against enveloped viruses. *PNAS*. 107, 17339-17344, (2010).
7. Dias, D.A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites*. 2, 303-336, (2012).
8. Maurice, M. I. Handbook of African medicinal plants (second edition). *CRC Press*. 1993.
9. Ribeiro, A., et al. Ethnobotanical survey in Canhane Village, district of Massingir, Monzambique: medicinal plants and traditional knowledge. *J. Ethnobio. Ethnomed*. 6, 33-43, (2010).
10. Mongalo, N.I. & Mafoko, B.J. Cassia abbreviata Oliv. a review of its ethnomedicinal uses, toxicology, phytochemistry, possible propagation techniques and pharmacology. *African Journal of Pharmacy and Pharmacology.* 7, 2901-2906, (2013).
11. Leteane, M.M, et al. Old plant newly discovered: Cassia sieberiana D.C and Cassia abbreviata Oliv. root extracts inhibit in vitro HIV-1c replication in peripheral blood mononuclear cells (PBMCs) by different modes of action. *J. Ethnopharmacol.* 141, 48-56, (2012).

12. Blumenthal, R., Durell, S. & Viard, M. HIV entry and envelope glycoprotein-mediated fusion. *J Biol Chem*. 287, 49-60, (2012).
13. Guo, Q. et al. Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120-CD4 interactions. *J Virol*. 77, 10528-10536, (2003).
14. Ho, H.T., et al. Envelope conformational changes induced by human immunodeficiency virus type 1 attachment inhibitors prevent CD4 binding and downstream entry events. *J Virol*. 80, 4017-4025, (2006).
15. Lin, P.F., et al. A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding. *Proc Natl Acad Sci USA*. 100, 11013-11018, (2003).
16. Wang, T., et al. Discovery of 4-benzoyl-1-[(4-methoxy-1H- pyrrolo [2,3-b] pyridin-3-yl) oxoacetyl] -2- (R)-methylpiperazine (BMS-378806): a novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions. J Med Chem. 46, 4236–4239, (2003).
17. Wang, T., et al. Inhibitors of human immunodeficiency virus type 1 (HIV-1) attachment. 5. An evolution from indole to azaindoles leading to the discovery of 1- (4-benzoylpiperazin-1-yl) -2-(4,7-dimethoxy-1H-pyrrolo[2,3-c]pyridin-3-yl)ethane -1,2-dione (BMS-488043), a dug candidate that demonstrates antiviral activity in HIV-1-infected subjects. *J Med Chem.* 52, 7778-7787, (2009).
18. Langley, D.R., et al. Homology models of the HIV-1 attachment inhibitor BMS-626529 bound to gp120 suggest a unique mechanism of action. *Proteins*. 83, 331-350, (2015).
19. Zhu, L. et al. Analysis of the pharmacokinetic interactions between BMS-626529, the active moiety of the HIV-1 attachment inhibitor prodrug BMS-663068, and ritonavir or ritonavir-boosted atazanavir in healthy subjects. *Antimicrob. Agents Chemother.* 7, 35-70, (2015).
20. Kashiwada, Y., et al. Anti-AIDS agents. 30. Anti-HIV activity of oleanolic acid, pomolic acid, and structurally related triterpenoids. *J. Nat. Prod*. 61, 1090-1095, 1998
21. Mengoni, F. et al. In vitro anti-HIV activity of oleanolic acid on infected human mononuclear cells. Planta Med. 68, 111-114, (2002).
22. Lee, D. Y. W., et al. Palmitic acid is a novel CD4 fusion inhibitor that blocks HIV entry and infection. *AIDS Research and Human Retroviruses.* 25, 1231-1243, (2009).

23. Min, B.S., et al. Anti-Human Immunodeficiency Virus-Type 1 activity of constituents from Juglans mandshurica. *Arch Pharm Res Vol.* 25, 441-445, (2002).
24. Clouser, C.L., et al. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. *Bioorg Med Chem Lett.* 22, 6642-6646, (2012).
25. Tewtrakul, S., et al. HIV-1 protease inhibitory substances from Cassia garrettiana. *Songklanakarin J. Sci. Technol*. 29, 145-149, (2007).
26. Bunluepuech, K., Wattanapiromsakul, C. & Tewtrakul, S. Anti-HIV-1 integrase activity of compounds from Cassia garrettiana heartwood. *Songklanakarin J. Sci. Technol*. 35, 665-669, (2013).
27. Wesolowska, O., et al. Interaction of the chemopreventive agent resveratrol and its metabolite, piceatannol with model membranes. *Biochimica et Biophysica Acta,* 1788, 1851-1860, (2009).
28. Piao, Z.S., et al. Synthesis and HIV-1 inhibitory activity of natural products isolated from Gnetum parvifolium and their analogues. *Yao Xue Xue Bao.* 45, 1509-1545, (2010).
29. Han, Y.S., et al. A resveratrol analog termed 3,3'4,4',5,5'-hexahydroxy-trans-stilbene is a potent HIV-1 inhibitor. *J Med Virol*. 87, 2054-2060, (2015).
30. Anqqakusuma, C., et al. Turmeric curcumin inhibits entry of all hepatitis C virus genotypes into human liver cells. *Gut*. 63, 1137-1149, (2014).
31. Wolf, M.C., et al. A broad-spectrum antiviral targeting entry of enveloped viruse. *PNAS.* 107, 3157-3062, (2010).
32. Pasetto, S., Pardi, V. & Murata, R.M. Anti-HIV-1 activity of flavonoid myricetin on HIV-1 infection in a dual-chamber in vitro model. *PLOS ONE*. 9, 1101-1118, (2014).
33. Baeten, J.M., et al. Hormonal contraceptive use, herpes simplex virus infection, and risk of HIV-1 acquisition among Kenyan women. *AIDS*. 21, 1771-1777, (2007).
34. Baatz, F., et al. Impact of the HIV-1 env genetic context outside HR1-HR2 on resitance to the fusion inhibitor enfuvirtide and viral infectivity in clinical isolates. *PLOS ONE*. 6, 21535-21546, (2011).
35. Gordts, S., et al. The low-cost compound lignosulfonic acid (LA) exhibits broad-spectrum anti-HIV and anti-HSV activity and has potential for microbicidal applications. *PLOS ONE*. 10, 1371-1383, (2015).

Acknowledgements

This work was supported by the “Fonds National de la Recherche” (FNR) of Luxembourg for the PhD grant of Yue Zheng [PHD AFR 1189522] and “la Fondation Recherche sur le SIDA”, the KU Leuven (GOA 10/014 and PF/10/018) and the Foundation of Scientific Research (FWO no. G-0528-12),

The authors would like to thank Sandra Claes, Evelyne Van Kerckhove, Eric Fonteyn, Leentje Persoons and Lies Van den Heurck for excellent technical assistance

.

Chapter III

General discussions

Half of the Nobel Prize in Physiology or Medicine 2015 has just been awarded to Youyou Tu for her discoveries of artemisinin concerning a novel therapy against Malaria. In 1960s, Youyou Tu and her colleagues were in a large Chinese national project embarking upon effort to develop new malaria therapies. They had collected over 2000 potential traditional remedies and tested 380 of them on mice. In an ancient book named *Zhouhoubeijifang* written by Hong Ge at 400 AD, Youyou Tu found the preparation method and usage of *Artemisia annua* which inspiriting her. She finally purified the active compound artemisinin which represents a new class of anti-malaria therapies and has been used for healing malaria all over the world. Youyou Tu and artemisinin's story has highlighted the enormous value of traditional medicines and ethnopharmacology.

Traditional medicine, also termed as complementary medicine in some countries, is a very important part of modern health care, which is often underestimated. Traditional medicines, especially traditional herbal medicines, have been used for thousands of years in many countries, where it is still commonly used. WHO has announced that eighty percent of African population use some form of traditional herbal medicine. Because of its significant therapeutic effects, the request of traditional medicine usage spreads to more countries and keeps a rapid increase. A US government survey revealed that US adults spend $33.9 billion out-of-pocket on visits to complementary and alternative medicine (CAM) practitioners and purchases of CAM products, classes, and materials (Nahin, et al., 2009). The global herbal supplement and remedy industry is forecast to reach $105 billion by 2017 (Global Industry In, 2015).

Ethnopharmacology has contributed a lot on drug development from traditional herbal medicines. It is not only related to the bio-use of pharmaceuticals but also related to the identification of medicinal plants, preparation of pharmaceuticals and social-medicinal aspects. Drug discovery from traditional medicines is an ethical attitude towards affordable and potent lead compounds, but it requires a multidisciplinary approach: ethnobiology, chemistry, biomedicine, physics/mathematics, computational modeling to explain and improve the mode of action of the compounds. Based on thousand-year experiences, the preparation of most traditional medicines and their bioactivity and toxicity have been well described and evaluated in human, which make it much easier to find new bioactivity related compounds in herbs compared to discover a new drug from zero.

Natural molecules from medicinal plants have become the rich resource of new drugs discovery, especially along with the development of organic chemistry and synthetic chemistry. Despite the fact that there are numerous plant derived compounds exhibiting inhibitory effects against a broad spectrum of pathogens having been reported, only a few of them reach the clinics, especially in the western countries. It costs about 15 years averagely for a new drug entering into the clinic and a large amount of money is required. In the case of anti-HIV drugs, there was also a lack of representative *ex vivo* models and cost-effective *in vivo* models. Although primates represent a good model for HIV research, ethical, financial and logistical considerations limit the use of primates in research. Many efforts have thus been conducted to develop humanized mice models susceptible to HIV as a preclinical tool for drug development that should facilitate the workflow of HIV drugs into clinics.

This thesis was focused on a traditional African plant, on which we have performed an in-depth investigation including the plant identification, the extraction and the isolation of compounds, as well as the elucidation of the anti-HIV mode of action. Several pieces of wood of this African plant were originally brought by an HIV infected patient from Congo at Saint-Pierre Hospital in Brussels in the early 2000s. This patient claimed that he had taken a decoction from this plant and felt an improvement of his condition. Although this patient died from AIDS because he did not receive antiretroviral therapy on time in Congo, we confirmed in 2001 the anti-HIV activity of the plant using a crude extract. However, the name of the plant was not disclosed by the patient and classical taxonomy could not be used to identify the plant. A first set of compounds purification was initiated at the Toxicology Laboratory of the National Health Laboratory of Luxembourg by Dr Serge Schneider but the single components purified at that time did not show any anti-HIV activity.

This thesis project was initiated 10 years after the first HIV tests performed with the plant in Luxembourg thanks to the emergence of the DNA barcoding approach. Using the DNA barcoding approach we were able to identify the plant as *Cassia abbreviata*. This plant is known in Africa to have therapeutic activities on many diseases including AIDS. The fact that we have been able to correctly identify the plant just from a small amount of wood material underlines the power of this technology which can be used by scientists having little or no taxonomical knowledge in the field to identify species. Classically, plants are identified by expert classical taxonomists using morphological features primarily from the flower, but also from other parts/organs of the plant (leaf shape and organization, fruit, etc). In contrast, the DNA-based barcoding technology relies on selected DNA markers which are present in all

parts of the plant and which can be amplified easily with a very limited number of primer pairs and whose sequences can be analyzed using appropriate software. It requires though that the plants have been correctly identified first by classical taxonomists and that the appropriate and reliable marker sequences are available for comparison. In principle every plant whose barcode marker sequences have been determined should be able to be identified correctly at any time in just a few days from even tiny amounts of material. The barcoding technology is presently used in many species identification programs for plants as well as animals (insects, birds, fish, etc).

During our identification study we used material obtained from germinating seeds from about 45 Cassia and Senna species that we ordered from online companies. This was the only way for us to obtain material from a wide range of species within a short period of time. Given the importance of the plant in Africa, there was a good chance that the seeds could be commonly obtained from one of these companies. This, however, also raised an important issue as to the correct identification of the species from which the seeds - sold by the company - were obtained. Also, as we experienced during our studies, labeling errors or mishandling of the seeds were real in that mislabeled seeds can occasionally be shipped. This was the case for instance for what we expected to be *Senna atomaria* seeds but which in fact revealed to be from *Leucaena leucocephala*, a species from the Mimosae tribe, also of the Legume family. Moreover, neither the exact geographical location nor the person who identified the plants were known to us, hence the information we had on all these species was very limited. It turned out however that we correctly identified *Cassia abbreviata*, which was our main concern, but it also showed that the vast majority of the seeds were from correctly identified species. Only in a few cases, e.g. *Cassia javanica*, for which several subspecies are known (*C. javanica* subsp. *javanica*, subsp. *nodosa*, subsp. *indochinensis*, subsp. *bartonii*) different sequences were obtained. This is most likely due to an incomplete naming (i.e. they were named *C. javanica* without regard to the subspecies). Alternative names may also have been used in some cases by the persons who identified the species for the providing company. One example could be *Cassia javanica* sensu Boer whose name is considered in *The Plant List* as a synonym of *Cassia roxburghii*. *Cassia javanica* and *Cassia roxburghii* however have clearly distinct trnH-psbA sequences with two duplications present in *Cassia roxburghii*. Our sequence analysis and comparison with published sequences however showed that the majority of the species have been correctly named, although the name of the subspecies was not always given.

We also used the sequence alignment of the two markers, ITS2 and trnH-psbA spacer, to establish a phylogenetic classification of the species studied, in order to assess their genetic relationships. It can be assumed indeed that closely related species basically produce identical or very similar compounds, while more distant species could be expected to lack some compounds and produce more different compounds. Interestingly the Cassia and Senna species are nearly all used as local medicines in the tropical countries where they grow, suggesting that their therapeutic activities are basically conserved.

The comparison of the sequences obtained in our lab as well as of sequences from publicly available databases documented by the sequence alignments as well as by the phylogenetic trees for each of the two markers show that the Cassia species can clearly be discriminated from the Senna species which form 5 subclades. Therefore the two markers have a high discriminating power for these and probably also for species from other genera. Regarding the Senna species it is interesting to see that the same species in appear to be linked in a same cluster, and this for quite a few markers, for instance ITS2, trnH-psbA, rpl16 intron, rps16 intron and matK.

This analysis also reveals some inconsistencies like *Senna italica* trnH-psbA sequence (HQ161769) which is found with the Chamaecrista species; on the other hand a *Chamaecrista absus* ITS2 sequence (JQ301843) clusters with Senna species in the Senna subclade 3, and a *Chamaecrista nigricans* trnH-psbA spacer sequence (HQ161770) clusters with Sennas in subclade 1. Other inconsistencies include very different trnH-psbA spacer sequences for two *Senna tora* species (HQ161762 in Senna subclade 3, and GU969277 in Senna subclade 5), these sequences were determined by different groups and it is clear that one of the two plants must have been misnamed of incorrectly identified. Also identical or very similar sequences for some species native to different continents suggest a misnaming/misidentification problem. In some cases, identical sequences have been found for different species from a same continent (examples). This is acceptable since these species may have diverged only recently and the genetic changes may have affected genes responsible for the minor morphological changes, but not the two marker regions ITS2 and trnH-psbA spacer.

We therefore conclude from our barcoding studies that the technology is extremely powerful, and useful for identification of raw materials of medicinal plant. It allows fast identification of species to scientists without knowledge in taxonomy, provided there are reliable sequences in

the databases used for the comparisons. This underlines the importance of the Consortium for the Barcode of Life (CBOL) initiative which aims to establish barcode libraries together with expert taxonomists. It also reveals that many sequences in public databases have been determined on misidentified species.

Despite the significant number of available antiretroviral agents, there is an ongoing need for new classes of antiretroviral agents that can provide potent and durable activity against HIV-1, primarily due to the development of drug resistance to existing compounds and the need for improved safety and tolerability profiles of treatment regimen. In particular treatment-experienced patients have limited treatment options owing to the presence of multi-drug resistant viral mutants causing reduced drug susceptibility, emergence of drug toxicities from long-term antiretroviral therapy and contraindications from the need to manage concurrent infections (Wilson et al., 2009). To manage the complex and highly individualized requirements of treatment-experienced patients, new classes of antiretrovirals offering less future drug resistance potential are still required.

Cassia abbreviata is indigenous to Africa, where it is commonly used to heal abnormal pain, fever, cough, snake bite, blood vomit, rash, diarrhea, as well as infectious diseases such as malaria, syphilis, gonorrhoea and HIV/AIDS (Kisangau et al, 2011). Nevertheless, *Cassia abbreviata*'s bioactive mode-of-actions were not investigated. Presently, we have completed the study of its anti-HIV activity and further focused on its mechanism and active components. We have also expanded the study on the antiviral activity of *Cassia abbreviata* on other viruses.

Our study first confirmed the anti-HIV activity of *Cassia abbreviata* by testing a crude aqueous extract (CE) and showed that CE inhibited HIV infection by targeting HIV entry via interacting with gp120. To further investigate the anti-HIV components, we have succeeded to isolate fifty-seven compounds and assessed their anti-HIV activities. Oleanolic acid, palmitic acid, piceatannol, taxifolin and another three so far uncharacterised compounds were found to inhibit HIV entry. Among them, piceatannol showed the highest activity and interesting mode-of-action. Similar to CE, piceatannol inhibited both R5 and X4 tropic HIV reference strains as well as clinical isolates harboring NRTIs, NNRTIs and PIs resistance.

HIV entry process contains multiple steps that require both viral and cellular membrane proteins. To enter target cells, virions have first to attach to cell membrane mainly driven by the virion envelope and the negatively charged heparin sulphate proteoglycans on the cell surface. The HIV-1 envelope is composed by two subunits: glycoprotein gp120 and transmembrane protein gp41, present in trimers – three gp120 molecules non-covalently combined with three gp41 molecules (Decroly, et al.,1994, Lu, et al.,1995, Merk, et al.,2013). After viral attachment to the cell membrane, which is a non-specific process via interaction of the virion envelope to negatively charged heparin sulfate proteoglycans, gp120 binds to CD4 causing the rearrangement of the V1/V2 loop of gp120 and the repositioning of V3 loop (Wilen, et al.,2012). CD4-gp120 complex then subsequently binds to a secondary receptor CCR5/CXCR4 on the cell surface, leading to the exposure of fusion peptide gp41 and the formation of the fusion complex.

Since our results showed that CE inhibited the interaction between CD4 and gp120 and had no effect on the availability and binding activity of CD4 expressing on the cell surface, it is tempting to speculate that the active compounds of CE may interfere with the binding of gp120 and CD4 or display some specific mechanism on gp120 conformation related to CD4 binding site. Recently, a series of small-molecule attachment inhibitors targeting the conserved CD4 binding region within gp120 have been described, such as BMS-378806, BMS-488043, BMS-088, BMS-049, BMS-626529 and BMS-663068 (Guo et al., 2003; Ho et al., 2006; Wang et al., 2003; Wang et al., 2009). Among these, BMS-626529 and BMS-663068 are currently the most promising HIV attachment inhibitors blocking CD4-gp120 interaction (Langley et al., 2015; Zhu et al., 2015). C7 position 2-amino-benzimidazole carboxamide moiety of all these BMS small molecules have been shown to be correlated to the effect on the V3 loop exposure of gp120. By optimizing structure, BMS-626529 has better anti-HIV activity than others, and its phosphonooxymethyl ester prodrug BMS-663068 has a better pharmacokinetic profile. Using homology models, BMS-626529 was proposed to bind to the unliganded conformation of gp120 (UNLIG) within the structurally conserved outer domain, under the anti-parallel β20-β21 sheet, which is adjacent to the CD4 binding site. BMS-626529/UNLIG complex blocks the central hydrophobic cavity of gp120 folding into CD4 binding conformation and the formation of a four stranded (β3-β2-β21-β20) bridging sheet, and thereby sheds the exposure of CD4 binding site of gp120. Since piceatannol does not target gp120, we will keep searching for the active compounds of CE targeting gp120 and confirm the activity on gp120 binding by Surface Plasmon Resonance to provide more binding kinetics information. We will also perform mutagenesis and homology studies to describe the targeting pocket of these compounds.

In our study, piceatannol shows its anti-HIV activity on various HIV strains displaying as a non-specific HIV entry inhibitor, as well as inhibits HSV. Not surprisingly, a natural analogue of piceatannol, reserveratrol has been described on its antiviral activity against HIV and HSV, and proposed to function at the early stage of infection (Faith, 2006). Resveratrol was also shown to inhibit HSV in the vagina of mice and limit extravaginal diseases (Docherty et al., 2005). Additionally, 3,3',4,4',5,5'-hexahydroxy-trans-stilbene, structurally close to reserveratrol and piceatannol, has also reported to inhibit HIV infection before fusion process. It is suggested that the additional hydroxyl groups to the basic stribene rings may increase the anti-HIV activity by strengthening compound affinity.

Recently, a new class of broad antiviral compounds has been described to target lipid membrane. Curcumin is another antiviral compound displaying structural similarity to piceatannol, having two phenols connected by a carbon chain. Curcumin can affect viral membrane fluidity and can confer a positive curvature to block viral entry (Anggakusuma et al, 2012). Similarly, rigid amphipathic fusion inhibitors (RAFIs), and Broad-SAVE (a class of thiazolidine-based lipophilic compounds, eg LJ001), broadly inhibit enveloped viruses by targeting the biophysical process of fusion between virus membrane and cell membrane (St. Vincent et al, 2010; Wolf et al, 2010). The fusion of lipid membranes is believed to contain intermediates and obtain energy from the elastic energy of spontaneous curvature of lipid membranes. Compounds like RAFIs containing a large polar head group and a relative small hydrocarbon tails can confer a positive curvature by bending the lipid layers away from polar heads, and thereby block membrane fusion. By compared study on dUY11 (a synthetic exampar RAFI) and LJ001, Vigant et al. described that the perylene group in RAFIs have a similar photosensitizing activity like LJ001, whose antiviral activity is light dependent. Nevertheless, both piceatannol and its analogue resveratrol have been shown to interact with headgroup region of phosphatidylcholine bilayers (Wesolowska et al., 2009). Although in our assays, piceatannol showed no inhibition on viral fusion, we cannot exclude that piceatannol inhibits viral entry by affecting membrane curvature like curcumin. To prove this hypothesis, more experiments focusing on the membrane fluidity and curvatures are planned to be performed.

Women have shown to be more likely to be infected by HIV during unprotected heterosexual intercourse than men (Glynn et al., 2001). Especially women in sub-Saharan Africa are in a high risk of HIV infection due to gender inequities. Microbicides are designed to prevent women from sexually transmitted infections such as HIV and HSV infections. Gel-based on tenofovir

have been shown to significantly inhibit viral infection in vagina (Rosenberg et al., 2012). In this study, we have shown the potential use as microbicide of CE and piceatannol *in vitro* that both inhibited HIV infection in a mimicked female genital tract model and did not stimulate PBMCs activation as well as inhibited infection of various HSV strains *in vitro*. Both CE and piceatannol have no effect on DC-SIGN dependent HIV transmission. Nevertheless, piceatannol has been described to inhibit LPS induced inflammation in cell lines and in rats suppressing pro-inflammatory cytokines production (Piotrowska et al., 2012). This may help to reduce the inflammation in mucosa and attenuate immune activation. To be a good microbicide, antivirals are required to highly prevent HIV acquisition, and demonstrate safety, stability and acceptability, and no adverse interaction with other anti-HIV drugs. In the study, we have shown that both CE and piceatannol have synergistic effect with co-receptor inhibitors AMD3100 and maraviroc, and fusion inhibitor enfuvirtide. In future studies, we will test piceatannol in combination with other classes of anti-HIV drugs and also perform structure optimization study to improve the anti-HIV activity.

More than thirty years after HIV identification, many efforts have been launched all over of the world and more than thirty anti-HIV drugs have been licensed. Antiretroviral therapies are a tremendous success, having dramatically reduced HIV-related morbidity, mortality and transmission. UNAIDS has announced the 90-90-90 goal: by 2020, 90% of all people living with HIV will know their HIV status; 90% of all people with diagnosed HIV infection with receive sustained antiretroviral therapy: 90% of all people receiving antiretroviral therapy will have viral suppression (UNAIDS, 2014). To achieve this ambitious target, more antitretroviral drugs are required to improve efficiency, and reduce the long term toxicity, and decrease the cost. Our study has shown piceatannol as a good structure candidate to be further optimized for HIV treatment and we expect that our uncharacterized compounds will follow the same course.

Chapter IV
Conclusion and Perspectives

In this thesis project, we have succeeded to identify an unknown wood from Africa as *Cassia abbreviata* using the DNA barcoding technology. The barcode markers ITS2 and trnH-psbA have shown their great power in the discrimination among Cassia and Senna species. Especially, a unique insertion and deletion were observed in the ITS2 and trnH-psbA sequences of real Cassia species, which could be used as special molecular fingerprint for Cassia species. During analysis, we have also noticed that some ITS2 and trnH-psbA spacer sequences in the public databases appear to have been misidentified or misnamed by the authors. A reliable database established by expert traditional and molecular taxonomists is an absolute prerequisite for correct species identification by DNA barcoding.

We have performed an in-depth antiviral investigation of *Cassia abbreviata*. Fifty-seven compounds have been succeeded to be isolated from *Cassia abbreviata*. Seven anti-HIV compounds targeting HIV entry have been found: oleanolic acid, palmitic acid, piceatannol, taxifolin and another three so far uncharacterized compounds. A crude aqueous extract (CE) and piceatannol have shown *Cassia abbreviata*'s broad spectrum antiviral activity against HIV and HSV, while piceatannol also inhibits VSV. CE was proven to inhibit HIV infection by targeting gp120, while we unraveled a new mechanism of action for piceatannol as a non-specific HIV entry inhibitor. Both CE and piceatannol have shown their inhibitory effect on HIV infection in a female genital tract model and no stimulation on PBMCs' activation. Based on its low toxicity, and long experience used in traditional medicine, Cassia Abbreviatta has a great potential to be used as food supplementary.

In the future, we will continue to study CE and other anti-HIV compounds derived from *Cassia abbreviata*. We will 1) decipher the structure and antiviral activity of the three so far uncharacterized compounds; 2) investigate which anti-HIV component of CE target gp120 and perform Surface Plasmon Resonance to obtain binding kinetics information; 3) work on mutagenesis and homology models to find the exact targeting

pocket on the viral gp120 protein; 4) optimize piceatannol's structure and initiate structure- activity relationship studies; 5) elucidate piceatannol's and derivatives antiviral mechanisms, by testing them in assays related to membrane fluidity and membrane curvature; 5) further investigate the synergy between piceatannol and the other anti-HIV uncharacterized compounds derived from *Cassia abbreviata*.

Bibliography

Abram, M. E., Hluhanich, R. M., Goodman, D. D., et al. Impact of primary elvitegravir resistance-associated mutations in hiv-1 integrase on drug susceptibility and viral replication fitness. *Antimicrob Agents Chemother,* 2013, 57, 2654-2663
Adamson, C. S., Sakalian, M., Salzwedel, K., et al. Polymorphisms in gag spacer peptide 1 confer varying levels of resistance to the hiv- 1 maturation inhibitor bevirimat. *Retrovirology,* 2010, 7, 36
Alter, G. and Altfeld, M. Nk cell function in hiv-1 infection. *Current molecular medicine,* 2006, 6, 621-629
Antinori, A., Zaccarelli, M., Cingolani, A., et al. Cross-resistance among nonnucleoside reverse transcriptase inhibitors limits recycling efavirenz after nevirapine failure. *AIDS Res Hum Retroviruses,* 2002, 18, 835-838
Apetrei, C., Robertson, D. L. and Marx, P. A. The history of sivs and aids: Epidemiology, phylogeny and biology of isolates from naturally siv infected non-human primates (nhp) in africa. *Front Biosci,* 2004, 9, 225-254
Arendt, V., Biwersi, G., Devaux, C., et al. *Comite de surveillance du sida: Rapport d'activité 2014*. Luxembourg, June 2015.
Arthos, J., Cicala, C., Martinelli, E., et al. Hiv-1 envelope protein binds to and signals through integrin alpha4beta7, the gut mucosal homing receptor for peripheral t cells. *Nat Immunol,* 2008, 9, 301-309
Arts, E. J. and Hazuda, D. J. Hiv-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med,* 2012, 2, a007161
Bailes, E., Gao, F., Bibollet-Ruche, F., et al. Hybrid origin of siv in chimpanzees. *Science,* 2003, 300, 1713
Balagopal, A., Ray, S. C., De Oca, R. M., et al. Kupffer cells are depleted with hiv immunodeficiency and partially recovered with antiretroviral immune reconstitution. *AIDS,* 2009, 23, 2397-2404
Baldauf, H. M., Pan, X., Erikson, E., et al. Samhd1 restricts hiv-1 infection in resting cd4(+) t cells. *Nat Med,* 2012, 18, 1682-1687
Barre-Sinoussi, F., Chermann, J. C., Rey, F., et al. Isolation of a t-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (aids). *Science,* 1983, 220, 868-871
Barre-Sinoussi, F., Ross, A. L. and Delfraissy, J. F. Past, present and future: 30 years of hiv research. *Nat Rev Microbiol,* 2013, 11, 877-883
Basmaciogullari, S. and Pizzato, M. The activity of nef on hiv-1 infectivity. *Front Microbiol,* 2014, 5, 232
Berkhout, B. and van der Velden, Y. U. Abx464: A good drug candidate instead of a magic bullet. *Retrovirology,* 2015, 12, 64
Berkowitz, R., Fisher, J. and Goff, S. P. Rna packaging. *Curr Top Microbiol Immunol,* 1996, 214, 177-218
Bess, J. W., Jr., Powell, P. J., Issaq, H. J., et al. Tightly bound zinc in human immunodeficiency virus type 1, human t-cell leukemia virus type i, and other retroviruses. *J Virol,* 1992, 66, 840-847
Blanco, J. L., Varghese, V., Rhee, S. Y., et al. Hiv-1 integrase inhibitor resistance and its clinical implications. *J Infect Dis,* 2011, 203, 1204-1214
Blumenthal, R., Durell, S. and Viard, M. Hiv entry and envelope glycoprotein-mediated fusion. *J Biol Chem,* 2012, 287, 40841-40849
Bour, S. and Strebel, K. The hiv-1 vpu protein: A multifunctional enhancer of viral particle release. *Microbes and infection / Institut Pasteur,* 2003, 5, 1029-1039
Boyd, M. R., Hallock, Y. F., Cardellina, J. H., 2nd, et al. Anti-hiv michellamines from ancistrocladus korupensis. *J Med Chem,* 1994, 37, 1740-1745
Brenchley, J. M., Schacker, T. W., Ruff, L. E., et al. Cd4+ t cell depletion during all stages of hiv disease occurs predominantly in the gastrointestinal tract. *J Exp Med,* 2004, 200, 749-759

Brinson, C., Lalezari, J., Gulam, L. H., et al. Hiv-1 attachment inhibitor prodrug bms-663068 in antiretroviral-experienced subjects: Week 24 sub-group analysis. *Journal of the International AIDS Society,* 2014, 17, 19529
Browne, E. P., Allers, C. and Landau, N. R. Restriction of hiv-1 by apobec3g is cytidine deaminase-dependent. *Virology,* 2009, 387, 313-321
Brumme, C. J., Huber, K. D., Dong, W., et al. Replication fitness of multiple nonnucleoside reverse transcriptase-resistant hiv-1 variants in the presence of etravirine measured by 454 deep sequencing. *J Virol,* 2013, 87, 8805-8807
Bryant, M. and Ratner, L. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci U S A,* 1990, 87, 523-527
Bukrinsky, M. and Adzhubei, A. Viral protein r of hiv-1. *Reviews in medical virology,* 1999, 9, 39-49
Burkly, L. C., Olson, D., Shapiro, R., et al. Inhibition of hiv infection by a novel cd4 domain 2-specific monoclonal antibody. Dissecting the basis for its inhibitory effect on hiv-induced cell fusion. *J Immunol,* 1992, 149, 1779-1787
Burniston, M. T., Cimarelli, A., Colgan, J., et al. Human immunodeficiency virus type 1 gag polyprotein multimerization requires the nucleocapsid domain and rna and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J Virol,* 1999, 73, 8527-8540
Buzon, V., Natrajan, G., Schibli, D., et al. Crystal structure of hiv-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog,* 2010, 6, e1000880
Campos, N., Myburgh, R., Garcel, A., et al. Long lasting control of viral rebound with a new drug abx464 targeting rev - mediated viral rna biogenesis. *Retrovirology,* 2015, 12, 30
Campos, N, Myburgh, R, Garcel, A, et al. Durable control of viral rebound in humanized mice by abx464 targeting rev functions. In *Conference on Retroviruses and Opportunistic Infections, Seatle, USA, 23th-26th Feb 2015.*
Carrington, M. and Alter, G. Innate immune control of hiv. *Cold Spring Harb Perspect Med,* 2012, 2, a007070
Casiraghi, M., Labra, M., Ferri, E., et al. DNA barcoding: A six-question tour to improve users' awareness about the method. *Briefings in bioinformatics,* 2010, 11, 440-453
Caskey, M., Klein, F., Lorenzi, J. C., et al. Viraemia suppressed in hiv-1-infected humans by broadly neutralizing antibody 3bnc117. *Nature,* 2015, 522, 487-491
CBOL, C. P. W. A DNA barcode for land plants. *Proc Natl Acad Sci U S A,* 2009, 106, 12794-12797
CDC. First report of aids. *MMWR,* 1981, 30, 2
Cen, S., Javanbakht, H., Kim, S., et al. Retrovirus-specific packaging of aminoacyl-trna synthetases with cognate primer trnas. *J Virol,* 2002, 76, 13111-13115
Chaix, M. L., Seng, R., Frange, P., et al. Increasing hiv-1 non-b subtype primary infections in patients in france and effect of hiv subtypes on virological and immunological responses to combined antiretroviral therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America,* 2013, 56, 880-887
Chakrabarti, L., Guyader, M., Alizon, M., et al. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature,* 1987, 328, 543-547
Chan, D. C., Fass, D., Berger, J. M., et al. Core structure of gp41 from the hiv envelope glycoprotein. *Cell,* 1997, 89, 263-273
Checkley, M. A., Luttge, B. G. and Freed, E. O. Hiv-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J Mol Biol,* 2011, 410, 582-608
Chege, D., Sheth, P. M., Kain, T., et al. Sigmoid th17 populations, the hiv latent reservoir, and microbial translocation in men on long-term antiretroviral therapy. *AIDS,* 2011, 25, 741-749
Chen, J. C., Krucinski, J., Miercke, L. J., et al. Crystal structure of the hiv-1 integrase catalytic core and c-terminal domains: A model for viral DNA binding. *Proc Natl Acad Sci U S A,* 2000, 97, 8233-8238

Chen, S., Yao, H., Han, J., et al. Validation of the its2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One,* 2010, 5, e8613
Chen, Z., Luckay, A., Sodora, D. L., et al. Human immunodeficiency virus type 2 (hiv-2) seroprevalence and characterization of a distinct hiv-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J Virol,* 1997, 71, 3953-3960
Cheng, X., Belshan, M. and Ratner, L. Hsp40 facilitates nuclear import of the human immunodeficiency virus type 2 vpx-mediated preintegration complex. *J Virol,* 2008, 82, 1229-1237
China Plant, B. O. L. G., Li, D. Z., Gao, L. M., et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (its) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci U S A,* 2011, 108, 19641-19646
Chiu, T. K. and Davies, D. R. Structure and function of hiv-1 integrase. *Curr Top Med Chem,* 2004, 4, 965-977
Chukkapalli, V., Hogue, I. B., Boyko, V., et al. Interaction between the human immunodeficiency virus type 1 gag matrix domain and phosphatidylinositol-(4,5)-bisphosphate is essential for efficient gag membrane binding. *J Virol,* 2008, 82, 2405-2417
Cichewicz, R. H. and Kouzi, S. A. Chemistry, biological activity, and chemotherapeutic potential of betulinic acid for the prevention and treatment of cancer and hiv infection. *Med Res Rev,* 2004, 24, 90-114
Clapham, P. R. and McKnight, A. Hiv-1 receptors and cell tropism. *Br Med Bull,* 2001, 58, 43-59
Clavel, F., Brun-Vezinet, F., Guetard, D., et al. [lav type ii: A second retrovirus associated with aids in west africa]. *C R Acad Sci III,* 1986, 302, 485-488
Coetzer, M., Cilliers, T., Ping, L. H., et al. Genetic characteristics of the v3 region associated with cxcr4 usage in hiv-1 subtype c isolates. *Virology,* 2006, 356, 95-105
Coffin JM, Hughes SH and HE, V. *Retroviruses*. New York: Cold Spring Harbor Laboratory Press,
Cohen, E. A., Terwilliger, E. F., Jalinoos, Y., et al. Identification of hiv-1 vpr product and function. *J Acquir Immune Defic Syndr,* 1990, 3, 11-18
Cohen, M. S., Shaw, G. M., McMichael, A. J., et al. Acute hiv-1 infection. *N Engl J Med,* 2011, 364, 1943-1954
Colin, G. *Primate taxonomy*. Washington. DC: Smithsonian Institution Press,
Connor, R. I., Sheridan, K. E., Ceradini, D., et al. Change in coreceptor use correlates with disease progression in hiv-1--infected individuals. *J Exp Med,* 1997, 185, 621-628
Cote, H. C., Brumme, Z. L. and Harrigan, P. R. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. *J Virol,* 2001, 75, 589-594
Cozzi-Lepri, A., Ruiz, L., Loveday, C., et al. Thymidine analogue mutation profiles: Factors associated with acquiring specific profiles and their impact on the virological response to therapy. *Antiviral therapy,* 2005, 10, 791-802
Craigie, R. Hiv integrase, a brief overview from chemistry to therapeutics. *J Biol Chem,* 2001, 276, 23213-23216
CRP-Santé. *Communique a l'occasion de la journee mondiale contre le sida*. Luxembourg, 1.12.2014.
D'Arc, M., Ayouba, A., Esteban, A., et al. Origin of the hiv-1 group o epidemic in western lowland gorillas. *Proc Natl Acad Sci U S A,* 2015, 112, E1343-1352
Dam, E., Quercia, R., Glass, B., et al. Gag mutations strongly contribute to hiv-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog,* 2009, 5, e1000345
Damond, F., Descamps, D., Farfara, I., et al. Quantification of proviral load of human immunodeficiency virus type 2 subtypes a and b using real-time pcr. *J Clin Microbiol,* 2001, 39, 4264-4268

Damond, F., Worobey, M., Campa, P., et al. Identification of a highly divergent hiv type 2 and proposal for a change in hiv type 2 classification. *AIDS Res Hum Retroviruses,* 2004, 20, 666-672
Das, S. R. and Jameel, S. Biology of the hiv nef protein. *The Indian journal of medical research,* 2005, 121, 315-332
Davenport, M. P. and Petravic, J. Cd8+ t cell control of hiv--a known unknown. *PLoS Pathog,* 2010, 6, e1000728
De Clercq, E. Non-nucleoside reverse transcriptase inhibitors (nnrtis): Past, present, and future. *Chemistry & biodiversity,* 2004, 1, 44-64
Decroly, E., Vandenbranden, M., Ruysschaert, J. M., et al. The convertases furin and pc1 can both cleave the human immunodeficiency virus (hiv)-1 envelope glycoprotein gp160 into gp120 (hiv-1 su) and gp41 (hiv-i tm). *J Biol Chem,* 1994, 269, 12240-12247
Degtjareva, G. V., Logacheva, M. D., Samigullin, T. H., et al. Organization of chloroplast psba-trnh intergenic spacer in dicotyledonous angiosperms of the family umbelliferae. *Biochemistry. Biokhimiia,* 2012, 77, 1056-1064
Delelis, O., Carayon, K., Saib, A., et al. Integrase and integration: Biochemical activities of hiv-1 integrase. *Retrovirology,* 2008, 5, 114
Delobel, P., Cazabat, M., Saliou, A., et al. Primary resistance of ccr5-tropic hiv-1 to maraviroc cannot be predicted by the v3 sequence. *J Antimicrob Chemother,* 2013, 68, 2506-2514
Demirov, D. G., Orenstein, J. M. and Freed, E. O. The late domain of human immunodeficiency virus type 1 p6 promotes virus release in a cell type-dependent manner. *J Virol,* 2002, 76, 105-117
Depienne, C., Roques, P., Creminon, C., et al. Cellular distribution and karyophilic properties of matrix, integrase, and vpr proteins from the human and simian immunodeficiency viruses. *Experimental cell research,* 2000, 260, 387-395
Dettin, M., Ferranti, P., Scarinci, C., et al. Is the v3 loop involved in hiv binding to cd4? *Biochemistry,* 2003, 42, 9007-9012
DHHS, D. o. H. a. H. S. Guidelines for the use of antiretroviral agents in hiv-1-infected adults and adolescents. 2015,
di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., et al. Characterization of highly immunogenic p66/p51 as the reverse transcriptase of htlv-iii/lav. *Science,* 1986, 231, 1289-1291
Diane L. Bolton, Amarendra Pegu, Keyun Wang, et al. Efficacy of hiv-1 monoclonal antibody immunotherapy in acute shiv-infected macaques. In *Conference on Retroviruses and Opportunistic Infections, Seatle, USA, 23th-26th Feb 2015.*
Dilley, K. A., Gregory, D., Johnson, M. C., et al. An lypsl late domain in the gag protein contributes to the efficient release and replication of rous sarcoma virus. *J Virol,* 2010, 84, 6276-6287
Docherty, J. J., Fu, M. M., Hah, J. M., et al. Effect of resveratrol on herpes simplex virus vaginal infection in the mouse. *Antiviral research,* 2005, 67, 155-162
Donzella, G. A., Schols, D., Lin, S. W., et al. Amd3100, a small molecule inhibitor of hiv-1 entry via the cxcr4 co-receptor. *Nat Med,* 1998, 4, 72-77
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. *Antimicrob Agents Chemother,* 2005, 49, 4721-4732
ECDC, E. c. f. d. p. a. c. Hiv/aids surveillance in europe 2013. 2014,
Edgar, R. C. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 2004, 32, 1792-1797
Engelman, A. and Craigie, R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J Virol,* 1992, 66, 6361-6369
Esser, M. T., Mori, T., Mondor, I., et al. Cyanovirin-n binds to gp120 to interfere with cd4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but

does not affect the cd4 binding site on gp120 or soluble cd4-induced conformational changes in gp120. *J Virol,* 1999, 73, 4360-4371
Estes, J. D., Harris, L. D., Klatt, N. R., et al. Damaged intestinal epithelial integrity linked to microbial translocation in pathogenic simian immunodeficiency virus infections. *PLoS Pathog,* 2010, 6, e1001052
Faith, S. A., Sweet, T. J., Bailey, E., et al. Resveratrol suppresses nuclear factor-kappab in herpes simplex virus infected cells. *Antiviral research,* 2006, 72, 242-251
Fanales-Belasio, E., Raimondo, M., Suligoi, B., et al. Hiv virology and pathogenetic mechanisms of infection: A brief overview. *Ann Ist Super Sanita,* 2010, 46, 5-14
Fawell, S., Seery, J., Daikh, Y., et al. Tat-mediated delivery of heterologous proteins into cells. *Proc Natl Acad Sci U S A,* 1994, 91, 664-668
Feller, U., Anders, I. and Mae, T. Rubiscolytics: Fate of rubisco after its enzymatic function in a cell is terminated. *Journal of experimental botany,* 2008, 59, 1615-1624
Fernandes, J., Jayaraman, B. and Frankel, A. The hiv-1 rev response element: An rna scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. *RNA biology,* 2012, 9, 6-11
Fourati, S., Visseaux, B., Armenia, D., et al. Identification of a rare mutation at reverse transcriptase lys65 (k65e) in hiv-1-infected patients failing on nucleos(t)ide reverse transcriptase inhibitors. *J Antimicrob Chemother,* 2013, 68, 2199-2204
Freed, E. O. Viral late domains. *J Virol,* 2002, 76, 4679-4687
Freed, E. O. Hiv-1 assembly, release and maturation. *Nat Rev Microbiol,* 2015, 13, 484-496
Fujii, Y., Otake, K., Fujita, Y., et al. Clustered localization of oligomeric nef protein of human immunodeficiency virus type 1 on the cell surface. *FEBS letters,* 1996, 395, 257-261
Fun, A., van Maarseveen, N. M., Pokorna, J., et al. Hiv-1 protease inhibitor mutations affect the development of hiv-1 resistance to the maturation inhibitor bevirimat. *Retrovirology,* 2011, 8, 70
Fung, H. B. and Guo, Y. Enfuvirtide: A fusion inhibitor for the treatment of hiv infection. *Clinical therapeutics,* 2004, 26, 352-378
Gabuzda, D. H., Lawrence, K., Langhoff, E., et al. Role of vif in replication of human immunodeficiency virus type 1 in cd4+ t lymphocytes. *J Virol,* 1992, 66, 6489-6495
Gallo, R. C., Salahuddin, S. Z., Popovic, M., et al. Frequent detection and isolation of cytopathic retroviruses (htlv-iii) from patients with aids and at risk for aids. *Science,* 1984, 224, 500-503
Gallo, S. A., Puri, A. and Blumenthal, R. Hiv-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by cxcr4 in the hiv-1 env-mediated fusion process. *Biochemistry,* 2001, 40, 12231-12236
Gamble, T. R., Yoo, S., Vajdos, F. F., et al. Structure of the carboxyl-terminal dimerization domain of the hiv-1 capsid protein. *Science,* 1997, 278, 849-853
Gao, F., Yue, L., White, A. T., et al. Human infection by genetically diverse sivsm-related hiv-2 in west africa. *Nature,* 1992, 358, 495-499
Gao, H. Q., Sarafianos, S. G., Arnold, E., et al. Rnase h cleavage of the 5' end of the human immunodeficiency virus type 1 genome. *J Virol,* 2001, 75, 11874-11880
Garcia-Perez, J., Staropoli, I., Azoulay, S., et al. A single-residue change in the hiv-1 v3 loop associated with maraviroc resistance impairs ccr5 binding affinity while increasing replicative capacity. *Retrovirology,* 2015, 12, 50
Garrido, C., Soriano, V., Geretti, A. M., et al. Resistance associated mutations to dolutegravir (s/gsk1349572) in hiv-infected patients--impact of hiv subtypes and prior raltegravir experience. *Antiviral research,* 2011, 90, 164-167
Geijtenbeek, T. B., Torensma, R., van Vliet, S. J., et al. Identification of dc-sign, a novel dendritic cell-specific icam-3 receptor that supports primary immune responses. *Cell,* 2000, 100, 575-585
Gelderblom, H. R. Assembly and morphology of hiv: Potential effect of structure on viral function. *AIDS,* 1991, 5, 617-637

Ghanam, R. H., Samal, A. B., Fernandez, T. F., et al. Role of the hiv-1 matrix protein in gag intracellular trafficking and targeting to the plasma membrane for virus assembly. *Front Microbiol,* 2012, 3, 55
Gitti, R. K., Lee, B. M., Walker, J., et al. Structure of the amino-terminal core domain of the hiv-1 capsid protein. *Science,* 1996, 273, 231-235
Global Industry In. Global herbal medicine industry report 2015. *QYreserach*.January 2015.
Göttlinger, H. G. *Hiv-1 gag: A molecular machine driving viral particle assembly and release*.
Gottlinger, H. G., Dorfman, T., Sodroski, J. G., et al. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A,* 1991, 88, 3195-3199
Greenberg, M. L. and Cammack, N. Resistance to enfuvirtide, the first hiv fusion inhibitor. *J Antimicrob Chemother,* 2004, 54, 333-340
Greig, S. L. and Deeks, E. D. Abacavir/dolutegravir/lamivudine single-tablet regimen: A review of its use in hiv-1 infection. *Drugs,* 2015, 75, 503-514
Grobler, J. A., Stillmock, K., Hu, B., et al. Diketo acid inhibitor mechanism and hiv-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proc Natl Acad Sci U S A,* 2002, 99, 6661-6666
Guo, Q., Ho, H. T., Dicker, I., et al. Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120-cd4 interactions. *Journal of virology,* 2003, 77, 10528-10536
Hahn, B. H., Shaw, G. M., De Cock, K. M., et al. Aids as a zoonosis: Scientific and public health implications. *Science,* 2000, 287, 607-614
Han, Y. S., Xiao, W. L., Xu, H., et al. Identification of a dibenzocyclooctadiene lignan as a hiv-1 non-nucleoside reverse transcriptase inhibitor. *Antivir Chem Chemother,* 2015, 24, 28-38
Harouse, J. M., Bhat, S., Spitalnik, S. L., et al. Inhibition of entry of hiv-1 in neural cell lines by antibodies against galactosyl ceramide. *Science,* 1991, 253, 320-323
Harrigan, P. R., Salim, M., Stammers, D. K., et al. A mutation in the 3' region of the human immunodeficiency virus type 1 reverse transcriptase (y318f) associated with nonnucleoside reverse transcriptase inhibitor resistance. *J Virol,* 2002, 76, 6836-6840
Hart, G. J., Orr, D. C., Penn, C. R., et al. Effects of (-)-2'-deoxy-3'-thiacytidine (3tc) 5'-triphosphate on human immunodeficiency virus reverse transcriptase and mammalian DNA polymerases alpha, beta, and gamma. *Antimicrob Agents Chemother,* 1992, 36, 1688-1694
Hebert, P. D., Ratnasingham, S. and deWaard, J. R. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biological sciences / The Royal Society,* 2003, 270 Suppl 1, S96-99
Henderson, L. E., Bowers, M. A., Sowder, R. C., 2nd, et al. Gag proteins of the highly replicative mn strain of human immunodeficiency virus type 1: Posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J Virol,* 1992, 66, 1856-1865
Hendrix, C. W., Collier, A. C., Lederman, M. M., et al. Safety, pharmacokinetics, and antiviral activity of amd3100, a selective cxcr4 receptor inhibitor, in hiv-1 infection. *J Acquir Immune Defic Syndr,* 2004, 37, 1253-1262
Henne, W. M., Buchkovich, N. J. and Emr, S. D. The escrt pathway. *Dev Cell,* 2011, 21, 77-91
Henriet, S., Mercenne, G., Bernacchi, S., et al. Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (vif) and the human apobec-3g and apobec-3f restriction factors. *Microbiology and molecular biology reviews : MMBR,* 2009, 73, 211-232
Hill, M. K., Hooker, C. W., Harrich, D., et al. Gag-pol supplied in trans is efficiently packaged and supports viral function in human immunodeficiency virus type 1. *J Virol,* 2001, 75, 6835-6840
Hill, M. K., Shehu-Xhilaga, M., Crowe, S. M., et al. Proline residues within spacer peptide p1 are important for human immunodeficiency virus type 1 infectivity, protein processing, and genomic rna dimer stability. *J Virol,* 2002, 76, 11245-11253

Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., et al. An african primate lentivirus (sivsm) closely related to hiv-2. *Nature,* 1989, 339, 389-392
Ho, H. T., Fan, L., Nowicka-Sans, B., et al. Envelope conformational changes induced by human immunodeficiency virus type 1 attachment inhibitors prevent cd4 binding and downstream entry events. *Journal of virology,* 2006, 80, 4017-4025
Hofmann, H., Logue, E. C., Bloch, N., et al. The vpx lentiviral accessory protein targets samhd1 for degradation in the nucleus. *J Virol,* 2012, 86, 12552-12560
Hofstra, M. T., Pingen, M., Koekkoek, K., et al. Sexual networks across risk groups persistently contribute to local spread of hiv. In *Conference on Retroviruses and Opportunistic Infections, Seatle, USA, 23th-26th Feb 2015.*
Hollingsworth, P. M., Graham, S. W. and Little, D. P. Choosing and using a plant DNA barcode. *PLoS One,* 2011, 6, e19254
Hu, Q., Trent, J. O., Tomaras, G. D., et al. Identification of env determinants in v3 that influence the molecular anatomy of ccr5 utilization. *J Mol Biol,* 2000, 302, 359-375
Hu, W. S. and Hughes, S. H. Hiv-1 reverse transcription. *Cold Spring Harb Perspect Med,* 2012, 2,
Hu, Z. and Kuritzkes, D. R. Effect of raltegravir resistance mutations in hiv-1 integrase on viral fitness. *J Acquir Immune Defic Syndr,* 2010, 55, 148-155
Huang, J., Ofek, G., Laub, L., et al. Broad and potent neutralization of hiv-1 by a gp41-specific human antibody. *Nature,* 2012, 491, 406-412
Huang, L. and Chen, C. Understanding hiv-1 protease autoprocessing for novel therapeutic development. *Future Med Chem,* 2013, 5, 1215-1229
Huet, T., Cheynier, R., Meyerhans, A., et al. Genetic organization of a chimpanzee lentivirus related to hiv-1. *Nature,* 1990, 345, 356-359
Hung, C. S., Vander Heyden, N. and Ratner, L. Analysis of the critical domain in the v3 loop of human immunodeficiency virus type 1 gp120 involved in ccr5 utilization. *J Virol,* 1999, 73, 8216-8226
Ivetac, A. and McCammon, J. A. Elucidating the inhibition mechanism of hiv-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations. *J Mol Biol,* 2009, 388, 644-658
Iyidogan, P. and Anderson, K. S. Current perspectives on hiv-1 antiretroviral drug resistance. *Viruses,* 2014, 6, 4095-4139
Janvier, K., Pelchen-Matthews, A., Renaud, J. B., et al. The escrt-0 component hrs is required for hiv-1 vpu-mediated bst-2/tetherin down-regulation. *PLoS Pathog,* 2011, 7, e1001265
Jayakumar, P., Berger, I., Autschbach, F., et al. Tissue-resident macrophages are productively infected ex vivo by primary x4 isolates of human immunodeficiency virus type 1. *J Virol,* 2005, 79, 5220-5226
Jenkins, T. M., Hickman, A. B., Dyda, F., et al. Catalytic domain of human immunodeficiency virus type 1 integrase: Identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci U S A,* 1995, 92, 6057-6061
Jiang, J., Ablan, S. D., Derebail, S., et al. The interdomain linker region of hiv-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology,* 2011, 421, 253-265
Johnson, V. A., Brun-Vezinet, F., Clotet, B., et al. Update of the drug resistance mutations in hiv-1: Fall 2006. *Topics in HIV medicine : a publication of the International AIDS Society, USA,* 2006, 14, 125-130
Kadoki, M., Choi, B. I. and Iwakura, Y. The mechanism of lps-induced hiv type i activation in transgenic mouse macrophages. *Int Immunol,* 2010, 22, 469-478
Karetnikov, A. and Suomalainen, M. Tethered virions are intermediates in the assembly and release of hiv-1 particles. *Virology,* 2010, 407, 289-295
Karn, J., Dingwall, C., Finch, J. T., et al. Rna binding by the tat and rev proteins of hiv-1. *Biochimie,* 1991, 73, 9-16
Karn, J. and Stoltzfus, C. M. Transcriptional and posttranscriptional regulation of hiv-1 gene expression. *Cold Spring Harb Perspect Med,* 2012, 2, a006916

Kashiwada, Y., Nagao, T., Hashimoto, A., et al. Anti-aids agents 38. Anti-hiv activity of 3-o-acyl ursolic acid derivatives. *J Nat Prod,* 2000, 63, 1619-1622
Kashman, Y., Gustafson, K. R., Fuller, R. W., et al. The calanolides, a novel hiv-inhibitory class of coumarin derivatives from the tropical rainforest tree, calophyllum lanigerum. *J Med Chem,* 1992, 35, 2735-2743
Keene, S. E., King, S. R. and Telesnitsky, A. 7sl rna is retained in hiv-1 minimal virus-like particles as an s-domain fragment. *J Virol,* 2010, 84, 9070-9077
Khamsri, B., Murao, F., Yoshida, A., et al. Comparative study on the structure and cytopathogenic activity of hiv vpr/vpx proteins. *Microbes and infection / Institut Pasteur,* 2006, 8, 10-15
King, J. R. and Acosta, E. P. Tipranavir: A novel nonpeptidic protease inhibitor of hiv. *Clin Pharmacokinet,* 2006, 45, 665-682
Kisangau, D. P., Herrmann TM, Lyaruu HVM, et al. Traditional knowledge, use practices and conservation of medicinal plants for hiv/aids care in rural tanzania. *Ethnobot. Res. Appl.,* 2011, 9, 14
Lu, M., Blacklow, S. C. and Kim, P. S. A trimeric structural domain of the hiv-1 transmembrane glycoprotein. *Nat Struct Biol,* 1995, 2, 1075-1082
Kohlstaedt, L. A., Wang, J., Friedman, J. M., et al. Crystal structure at 3.5 a resolution of hiv-1 reverse transcriptase complexed with an inhibitor. *Science,* 1992, 256, 1783-1790
Kojima, K., Amano, Y., Yoshino, K., et al. Escrt-0 protein hepatocyte growth factor-regulated tyrosine kinase substrate (hrs) is targeted to endosomes independently of signal-transducing adaptor molecule (stam) and the complex formation with stam promotes its endosomal dissociation. *J Biol Chem,* 2014, 289, 33296-33310
Konnyu, B., Sadiq, S. K., Turanyi, T., et al. Gag-pol processing during hiv-1 virion maturation: A systems biology approach. *PLoS Comput Biol,* 2013, 9, e1003103
Krausslich, H. G., Facke, M., Heuser, A. M., et al. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol,* 1995, 69, 3407-3419
Kress, W. J., Erickson, D. L., Jones, F. A., et al. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in panama. *Proc Natl Acad Sci U S A,* 2009, 106, 18621-18626
Kress, W. J., Garcia-Robledo, C., Uriarte, M., et al. DNA barcodes for ecology, evolution, and conservation. *Trends in ecology & evolution,* 2015, 30, 25-35
Kristoff, J., Haret-Richter, G., Ma, D., et al. Early microbial translocation blockade reduces siv-mediated inflammation and viral replication. *J Clin Invest,* 2014, 124, 2802-2806
Kuppuswamy, M., Subramanian, T., Srinivasan, A., et al. Multiple functional domains of tat, the trans-activator of hiv-1, defined by mutational analysis. *Nucleic acids research,* 1989, 17, 3551-3561
Kwon, D. S., Gregorio, G., Bitton, N., et al. Dc-sign-mediated internalization of hiv is required for trans-enhancement of t cell infection. *Immunity,* 2002, 16, 135-144
Lafontaine, D. L. and Tollervey, D. The function and synthesis of ribosomes. *Nature reviews. Molecular cell biology,* 2001, 2, 514-520
Laguette, N., Bregnard, C., Bouchet, J., et al. Nef-induced cd4 endocytosis in human immunodeficiency virus type 1 host cells: Role of p56lck kinase. *J Virol,* 2009, 83, 7117-7128
Lahouassa, H., Daddacha, W., Hofmann, H., et al. Samhd1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol,* 2012, 13, 223-228
Langelier, C., von Schwedler, U. K., Fisher, R. D., et al. Human escrt-ii complex and its role in human immunodeficiency virus type 1 release. *J Virol,* 2006, 80, 9465-9480
Langley, D. R., Kimura, S. R., Sivaprakasam, P., et al. Homology models of the hiv-1 attachment inhibitor bms-626529 bound to gp120 suggest a unique mechanism of action. *Proteins,* 2015, 83, 331-350
LeBreton, M., Switzer, W. M., Djoko, C. F., et al. A gorilla reservoir for human t-lymphotropic virus type 4. *Emerg Microbes Infect,* 2014, 3, e7

Lee-Huang, S., Huang, P. L., Chen, H. C., et al. Anti-hiv and anti-tumor activities of recombinant map30 from bitter melon. *Gene,* 1995, 161, 151-156
Lee, C. H., Saksela, K., Mirza, U. A., et al. Crystal structure of the conserved core of hiv-1 nef complexed with a src family sh3 domain. *Cell,* 1996, 85, 931-942
Lee, J. B., Hayashi, K., Hayashi, T., et al. Antiviral activities against hsv-1, hcmv, and hiv-1 of rhamnan sulfate from monostroma latissimum. *Planta Med,* 1999, 65, 439-441
Li, B. Q., Fu, T., Dongyan, Y., et al. Flavonoid baicalin inhibits hiv-1 infection at the level of viral entry. *Biochem Biophys Res Commun,* 2000, 276, 534-538
Li, X., Yang, Y., Henry, R. J., et al. Plant DNA barcoding: From gene to genome. *Biological reviews of the Cambridge Philosophical Society,* 2015, 90, 157-166
Li, X. N., Pu, J. X., Du, X., et al. Lignans with anti-hiv activity from schisandra propinqua var. Sinensis. *J Nat Prod,* 2009, 72, 1133-1141
Liu, J. and Thorp, S. C. Cell surface heparan sulfate and its roles in assisting viral infections. *Med Res Rev,* 2002, 22, 1-25
Locher, C. P., Witt, S. A., Kassel, R., et al. Differential effects of r5 and x4 human immunodeficiency virus type 1 infection on cd4+ cell proliferation and activation. *J Gen Virol,* 2005, 86, 1171-1179
Loya, S., Rudi, A., Kashman, Y., et al. Polycitone a, a novel and potent general inhibitor of retroviral reverse transcriptases and cellular DNA polymerases. *Biochem J,* 1999, 344 Pt 1, 85-92
Lu, J., Deeks, S. G., Hoh, R., et al. Rapid emergence of enfuvirtide resistance in hiv-1-infected patients: Results of a clonal analysis. *J Acquir Immune Defic Syndr,* 2006, 43, 60-64
Lu, L., Liu, S. W., Jiang, S. B., et al. Tannin inhibits hiv-1 entry by targeting gp41. *Acta Pharmacol Sin,* 2004, 25, 213-218
Lu, M., Blacklow, S. C. and Kim, P. S. A trimeric structural domain of the hiv-1 transmembrane glycoprotein. *Nat Struct Biol,* 1995, 2, 1075-1082
Luber, A. D. Genetic barriers to resistance and impact on clinical response. *MedGenMed : Medscape general medicine,* 2005, 7, 69
Lv, Z., Chu, Y. and Wang, Y. Hiv protease inhibitors: A review of molecular selectivity and toxicity. *Hiv/Aids,* 2015, 7, 95-104
Manches, O., Frleta, D. and Bhardwaj, N. Dendritic cells in progression and pathology of hiv infection. *Trends in immunology,* 2014, 35, 114-122
Markosyan, R. M., Leung, M. Y. and Cohen, F. S. The six-helix bundle of human immunodeficiency virus env controls pore formation and enlargement and is initiated at residues proximal to the hairpin turn. *J Virol,* 2009, 83, 10048-10057
Martin-Serrano, J., Zang, T. and Bieniasz, P. D. Role of escrt-i in retroviral budding. *J Virol,* 2003, 77, 4794-4804
Martin, D. E., Salzwedel, K. and Allaway, G. P. Bevirimat: A novel maturation inhibitor for the treatment of hiv-1 infection. *Antivir Chem Chemother,* 2008, 19, 107-113
Matsui, T., Kobayashi, S., Yoshida, O., et al. Effects of succinylated concanavalin a on infectivity and syncytial formation of human immunodeficiency virus. *Med Microbiol Immunol,* 1990, 179, 225-235
McCallister S, Lalezari J, Richmond G, et al. Hiv-1 gag polymorphisms determine treatment respose to bevirimat (pa-457). Antivitral Therapy. Sitges, Spain.
McColl, D. J. and Chen, X. Strand transfer inhibitors of hiv-1 integrase: Bringing in a new era of antiretroviral therapy. *Antiviral research,* 2010, 85, 101-118
McCullough, J., Colf, L. A. and Sundquist, W. I. Membrane fission reactions of the mammalian escrt pathway. *Annu Rev Biochem,* 2013, 82, 663-692
Menendez-Arias, L. Molecular basis of human immunodeficiency virus drug resistance: An update. *Antiviral research,* 2010, 85, 210-231
Meng, B. and Lever, A. M. Wrapping up the bad news: Hiv assembly and release. *Retrovirology,* 2013, 10, 5
Mengoni, F., Lichtner, M., Battinelli, L., et al. In vitro anti-hiv activity of oleanolic acid on infected human mononuclear cells. *Planta Med,* 2002, 68, 111-114

Merk, A. and Subramaniam, S. Hiv-1 envelope glycoprotein structure. *Curr Opin Struct Biol,* 2013, 23, 268-276
Mesplede, T., Quashie, P. K. and Wainberg, M. A. Resistance to hiv integrase inhibitors. *Curr Opin HIV AIDS,* 2012, 7, 401-408
Mesplede, T. and Wainberg, M. A. Is resistance to dolutegravir possible when this drug is used in first-line therapy? *Viruses,* 2014, 6, 3377-3385
Min, B. S., Lee, H. K., Lee, S. M., et al. Anti-human immunodeficiency virus-type 1 activity of constituents from juglans mandshurica. *Arch Pharm Res,* 2002, 25, 441-445
Mitsuya, H., Weinhold, K. J., Furman, P. A., et al. 3'-azido-3'-deoxythymidine (bw a509u): An antiviral agent that inhibits the infectivity and cytopathic effect of human t-lymphotropic virus type iii/lymphadenopathy-associated virus in vitro. *Proc Natl Acad Sci U S A,* 1985, 82, 7096-7100
Moir, S. and Fauci, A. S. B cells in hiv infection and disease. *Nature reviews. Immunology,* 2009, 9, 235-245
Momany, C., Kovari, L. C., Prongay, A. J., et al. Crystal structure of dimeric hiv-1 capsid protein. *Nat Struct Biol,* 1996, 3, 763-770
Moore, J. P. and Doms, R. W. The entry of entry inhibitors: A fusion of science and medicine. *Proc Natl Acad Sci U S A,* 2003, 100, 10598-10602
Mori, T., O'Keefe, B. R., Sowder, R. C., 2nd, et al. Isolation and characterization of griffithsin, a novel hiv-inactivating protein, from the red alga griffithsia sp. *J Biol Chem,* 2005, 280, 9345-9353
Morita, E., Sandrin, V., McCullough, J., et al. Escrt-iii protein requirements for hiv-1 budding. *Cell Host Microbe,* 2011, 9, 235-242
Mousseau, G., Clementz, M. A., Bakeman, W. N., et al. An analog of the natural steroidal alkaloid cortistatin a potently suppresses tat-dependent hiv transcription. *Cell Host Microbe,* 2012, 12, 97-108
Mulinge, M., Lemaire, M., Servais, J. Y., et al. Hiv-1 tropism determination using a phenotypic env recombinant viral assay highlights overestimation of cxcr4-usage by genotypic prediction algorithms for crf01_ae and crf02_ag [corrected]. *PLoS One,* 2013, 8, e60566
Muller, B., Tessmer, U., Schubert, U., et al. Human immunodeficiency virus type 1 vpr protein is incorporated into the virion in significantly smaller amounts than gag and is phosphorylated in infected cells. *J Virol,* 2000, 74, 9727-9731
Murrell, D. E., Moorman, J. P. and Harirforoosh, S. Stribild: A review of component characteristics and combination drug efficacy. *European review for medical and pharmacological sciences,* 2015, 19, 904-914
N, T. P. T., Duc, N. M., Sinh, N. V., et al. Application of DNA barcoding markers to the identification of hopea species. *Genetics and molecular research : GMR,* 2015, 14, 9181-9190
Nahin, R. L., Barnes, P. M., Stussman, B. J., et al. Costs of complementary and alternative medicine (cam) and frequency of visits to cam practitioners: United states, 2007. *National health statistics reports,* 2009, 1-14
Neil, S. J. The antiviral activities of tetherin. *Curr Top Microbiol Immunol,* 2013, 371, 67-104
Nelson, J. D., Brunel, F. M., Jensen, R., et al. An affinity-enhanced neutralizing antibody against the membrane-proximal external region of human immunodeficiency virus type 1 gp41 recognizes an epitope between those of 2f5 and 4e10. *J Virol,* 2007, 81, 4033-4043
Neogi, U., Haggblom, A., Santacatterina, M., et al. Temporal trends in the swedish hiv-1 epidemic: Increase in non-b subtypes and recombinant forms over three decades. *PLoS One,* 2014, 9, e99390
Nguyen, D. G. and Hildreth, J. E. Involvement of macrophage mannose receptor in the binding and transmission of hiv by macrophages. *Eur J Immunol,* 2003, 33, 483-493
Nonaka, G., Nishioka, I., Nishizawa, M., et al. Anti-aids agents, 2: Inhibitory effects of tannins on hiv reverse transcriptase and hiv replication in h9 lymphocyte cells. *J Nat Prod,* 1990, 53, 587-595

Nowotny, M., Gaidamakov, S. A., Ghirlando, R., et al. Structure of human rnase h1 complexed with an rna/DNA hybrid: Insight into hiv reverse transcription. *Mol Cell,* 2007, 28, 264-276
Obermeier, M., Symons, J. and Wensing, A. M. Hiv population genotypic tropism testing and its clinical significance. *Curr Opin HIV AIDS,* 2012, 7, 470-477
Oldfield, V. and Plosker, G. L. Lopinavir/ritonavir: A review of its use in the management of hiv infection. *Drugs,* 2006, 66, 1275-1299
Olender, S. A., Taylor, B. S., Wong, M., et al. Croi 2015: Advances in antiretroviral therapy. *Top Antivir Med,* 2015, 23, 28-45
Osmanov, S., Pattou, C., Walker, N., et al. Estimated global distribution and regional spread of hiv-1 genetic subtypes in the year 2000. *J Acquir Immune Defic Syndr,* 2002, 29, 184-190
Pace, C. S., Fordyce, M. W., Franco, D., et al. Anti-cd4 monoclonal antibody ibalizumab exhibits breadth and potency against hiv-1, with natural resistance mediated by the loss of a v5 glycan in envelope. *J Acquir Immune Defic Syndr,* 2013a, 62, 1-9
Pace, C. S., Song, R., Ochsenbauer, C., et al. Bispecific antibodies directed to cd4 domain 2 and hiv envelope exhibit exceptional breadth and picomolar potency against hiv-1. *Proc Natl Acad Sci U S A,* 2013b, 110, 13540-13545
Pancio, H. A., Vander Heyden, N. and Ratner, L. The c-terminal proline-rich tail of human immunodeficiency virus type 2 vpx is necessary for nuclear localization of the viral preintegration complex in nondividing cells. *J Virol,* 2000, 74, 6162-6167
Pejchal, R., Doores, K. J., Walker, L. M., et al. A potent and broad neutralizing antibody recognizes and penetrates the hiv glycan shield. *Science,* 2011, 334, 1097-1103
Pereyra, F., Addo, M. M., Kaufmann, D. E., et al. Genetic and immunologic heterogeneity among persons who control hiv infection in the absence of therapy. *J Infect Dis,* 2008, 197, 563-571
Perez, L. G., Chen, H., Liao, H. X., et al. Envelope glycoprotein binding to the integrin alpha4beta7 is not a general property of most hiv-1 strains. *J Virol,* 2014, 88, 10767-10777
Perry, A. K., Chen, G., Zheng, D., et al. The host type i interferon response to viral and bacterial infections. *Cell Res,* 2005, 15, 407-422
Pilcher, C. D., Eron, J. J., Jr., Galvin, S., et al. Acute hiv revisited: New opportunities for treatment and prevention. *J Clin Invest,* 2004, 113, 937-945
Piller, S. C., Caly, L. and Jans, D. A. Nuclear import of the pre-integration complex (pic): The achilles heel of hiv? *Curr Drug Targets,* 2003, 4, 409-429
Piotrowska, H., Kucinska, M. and Murias, M. Biological activity of piceatannol: Leaving the shadow of resveratrol. *Mutation research,* 2012, 750, 60-82
Plantier, J. C., Leoz, M., Dickerson, J. E., et al. A new human immunodeficiency virus derived from gorillas. *Nat Med,* 2009, 15, 871-872
Platt, E. J., Kuhmann, S. E., Rose, P. P., et al. Adaptive mutations in the v3 loop of gp120 enhance fusogenicity of human immunodeficiency virus type 1 and enable use of a ccr5 coreceptor that lacks the amino-terminal sulfated region. *J Virol,* 2001, 75, 12266-12278
Poeschla, E. M. Integrase, ledgf/p75 and hiv replication. *Cell Mol Life Sci,* 2008, 65, 1403-1424
Pollard, V. W. and Malim, M. H. The hiv-1 rev protein. *Annual review of microbiology,* 1998, 52, 491-532
Pontow, S. E., Kery, V. and Stahl, P. D. Mannose receptor. *Int Rev Cytol,* 1992, 137B, 221-244
Praparattanapan, J., Kotarathititthum, W., Chaiwarith, R., et al. Resistance-associated mutations after initial antiretroviral treatment failure in a large cohort of patients infected with hiv-1 subtype crf01_ae. *Current HIV research,* 2012, 10, 647-652
Quan, Y., Gu, Z., Li, X., et al. Endogenous reverse transcription assays reveal high-level resistance to the triphosphate of (-)2'-dideoxy-3'-thiacytidine by mutated m184v human immunodeficiency virus type 1. *J Virol,* 1996, 70, 5642-5645
Quinn, T. C. Hiv epidemiology and the effects of antiviral therapy on long-term consequences. *AIDS,* 2008, 22 Suppl 3, S7-12

Quinn, T. C., Brookmeyer, R., Kline, R., et al. Feasibility of pooling sera for hiv-1 viral rna to diagnose acute primary hiv-1 infection and estimate hiv incidence. *AIDS,* 2000, 14, 2751-2757
Ren, J. and Stammers, D. K. Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of hiv reverse transcriptase. *Virus research,* 2008, 134, 157-170
Restle, T., Muller, B. and Goody, R. S. Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. *J Biol Chem,* 1990, 265, 8986-8988
Ribeiro, R. M., Hazenberg, M. D., Perelson, A. S., et al. Naive and memory cell turnover as drivers of ccr5-to-cxcr4 tropism switch in human immunodeficiency virus type 1: Implications for therapy. *J Virol,* 2006, 80, 802-809
Roche, M., Jakobsen, M. R., Ellett, A., et al. Hiv-1 predisposed to acquiring resistance to maraviroc (mvc) and other ccr5 antagonists in vitro has an inherent, low-level ability to utilize mvc-bound ccr5 for entry. *Retrovirology,* 2011, 8, 89
Roche, M., Salimi, H., Duncan, R., et al. A common mechanism of clinical hiv-1 resistance to the ccr5 antagonist maraviroc despite divergent resistance levels and lack of common gp120 resistance mutations. *Retrovirology,* 2013, 10, 43
Rolla, G., Mietta, S., Raie, A., et al. Incidence of food anaphylaxis in piemonte region (italy): Data from registry of center for severe allergic reactions. *Internal and emergency medicine,* 2013, 8, 615-620
Romani, B., Shaykh Baygloo, N., Aghasadeghi, M. R., et al. Hiv-1 vpr protein enhances proteasomal degradation of mcm10 DNA replication factor through the cul4-ddb1[vprbp] e3 ubiquitin ligase to induce g2/m cell cycle arrest. *J Biol Chem,* 2015, 290, 17380-17389
Rosenberg, Z. F. and Devlin, B. Future strategies in microbicide development. *Best practice & research. Clinical obstetrics & gynaecology,* 2012, 26, 503-513
Roos, M. T., Lange, J. M., de Goede, R. E., et al. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J Infect Dis,* 1992, 165, 427-432
Saitou, N. and Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular biology and evolution,* 1987, 4, 406-425
Sandefur, S., Smith, R. M., Varthakavi, V., et al. Mapping and characterization of the n-terminal i domain of human immunodeficiency virus type 1 pr55(gag). *J Virol,* 2000, 74, 7238-7249
Sandler, N. G. and Douek, D. C. Microbial translocation in hiv infection: Causes, consequences and treatment opportunities. *Nat Rev Microbiol,* 2012, 10, 655-666
Santos, A. F. and Soares, M. A. Hiv genetic diversity and drug resistance. *Viruses,* 2010, 2, 503-531
Saphire, A. C., Bobardt, M. D., Zhang, Z., et al. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J Virol,* 2001, 75, 9187-9200
Saphire, E. O., Montero, M., Menendez, A., et al. Structure of a high-affinity "mimotope" peptide bound to hiv-1-neutralizing antibody b12 explains its inability to elicit gp120 cross-reactive antibodies. *J Mol Biol,* 2007, 369, 696-709
Sarmati, L., Parisi, S. G., Andreoni, C., et al. Switching of inferred tropism caused by hiv during interruption of antiretroviral therapy. *J Clin Microbiol,* 2010, 48, 2586-2588
Scheid, J. F., Mouquet, H., Ueberheide, B., et al. Sequence and structural convergence of broad and potent hiv antibodies that mimic cd4 binding. *Science,* 2011, 333, 1633-1637
Schoch, C. L., Seifert, K. A., Huhndorf, S., et al. Nuclear ribosomal internal transcribed spacer (its) region as a universal DNA barcode marker for fungi. *Proc Natl Acad Sci U S A,* 2012, 109, 6241-6246
Schuitemaker, H., Koot, M., Kootstra, N. A., 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 monocytotropic to t-cell-tropic virus population. *J Virol,* 1992, 66, 1354-1360
Seclen, E., Gonzalez Mdel, M., Lapaz, M., et al. Primary resistance to maraviroc in a large set of r5-v3 viral sequences from hiv-1-infected patients. *J Antimicrob Chemother,* 2010, 65, 2502-2504

Sette, P., Jadwin, J. A., Dussupt, V., et al. The escrt-associated protein alix recruits the ubiquitin ligase nedd4-1 to facilitate hiv-1 release through the lypxnl l domain motif. *J Virol,* 2010, 84, 8181-8192
Sharp, P. M. and Hahn, B. H. Origins of hiv and the aids pandemic. *Cold Spring Harb Perspect Med,* 2011, 1, a006841
Sierra-Aragon, S. and Walter, H. Targets for inhibition of hiv replication: Entry, enzyme action, release and maturation. *Intervirology,* 2012, 55, 84-97
Simon, F., Mauclere, P., Roques, P., et al. Identification of a new human immunodeficiency virus type 1 distinct from group m and group o. *Nat Med,* 1998, 4, 1032-1037
Sluis-Cremer, N., Temiz, N. A. and Bahar, I. Conformational changes in hiv-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Current HIV research,* 2004, 2, 323-332
St Vincent, M. R., Colpitts, C. C., Ustinov, A. V., et al. Rigid amphipathic fusion inhibitors, small molecule antiviral compounds against enveloped viruses. *Proc Natl Acad Sci U S A,* 2010, 107, 17339-17344
Stacey, A. R., Norris, P. J., Qin, L., 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, 3719-3733
Stanley, B. J., Ehrlich, E. S., Short, L., et al. Structural insight into the human immunodeficiency virus vif socs box and its role in human e3 ubiquitin ligase assembly. *J Virol,* 2008, 82, 8656-8663
Stirpe, F. Ribosome-inactivating proteins: From toxins to useful proteins. *Toxicon,* 2013, 67, 12-16
Stoltzfus, C. M. Chapter 1. Regulation of hiv-1 alternative rna splicing and its role in virus replication. *Adv Virus Res,* 2009, 74, 1-40
Strebel, K., Klimkait, T., Maldarelli, F., et al. Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J Virol,* 1989, 63, 3784-3791
Streeck, H., Jolin, J. S., Qi, Y., 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, 7641-7648
Struck, D., Roman, F., De Landtsheer, S., et al. Near full-length characterization and population dynamics of the human immunodeficiency virus type i circulating recombinant form 42 (crf42_bf) in luxembourg. *AIDS Res Hum Retroviruses,* 2015, 31, 554-558
Sugita, M., Shinozaki, K. and Sugiura, M. Tobacco chloroplast trna(uuu) gene contains a 2.5-kilobase-pair intron: An open reading frame and a conserved boundary sequence in the intron. *Proc Natl Acad Sci U S A,* 1985, 82, 3557-3561
Takehisa, J., Kraus, M. H., Ayouba, A., et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J Virol,* 2009, 83, 1635-1648
Tantillo, C., Ding, J., Jacobo-Molina, A., et al. Locations of anti-aids drug binding sites and resistance mutations in the three-dimensional structure of hiv-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. *J Mol Biol,* 1994, 243, 369-387
Tassaneetrithep, B., Tivon, D., Swetnam, J., et al. Cryptic determinant of alpha4beta7 binding in the v2 loop of hiv-1 gp120. *PLoS One,* 2014, 9, e108446
Tatt, I. D., Barlow, K. L., Nicoll, A., et al. The public health significance of hiv-1 subtypes. *AIDS,* 2001, 15 Suppl 5, S59-71
Techen, N., Parveen, I., Pan, Z., et al. DNA barcoding of medicinal plant material for identification. *Current opinion in biotechnology,* 2014, 25, 103-110
Thuy, T. T., Ly, B. M., Van, T. T., et al. Anti-hiv activity of fucoidans from three brown seaweed species. *Carbohydr Polym,* 2015, 115, 122-128
Tie, Y., Boross, P. I., Wang, Y. F., et al. Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of hiv-1 protease mutants with substrate analogs. *The FEBS journal,* 2005, 272, 5265-5277

Tiganos, E., Yao, X. J., Friborg, J., et al. Putative alpha-helical structures in the human immunodeficiency virus type 1 vpu protein and cd4 are involved in binding and degradation of the cd4 molecule. *J Virol,* 1997, 71, 4452-4460
Tomaras, G. D. and Haynes, B. F. Hiv-1-specific antibody responses during acute and chronic hiv-1 infection. *Curr Opin HIV AIDS,* 2009, 4, 373-379
Truant, R. and Cullen, B. R. The arginine-rich domains present in human immunodeficiency virus type 1 tat and rev function as direct importin beta-dependent nuclear localization signals. *Molecular and cellular biology,* 1999, 19, 1210-1217
Trumpfheller, C., Park, C. G., Finke, J., et al. Cell type-dependent retention and transmission of hiv-1 by dc-sign. *Int Immunol,* 2003, 15, 289-298
Tsegaye, T. S. and Pohlmann, S. The multiple facets of hiv attachment to dendritic cell lectins. *Cell Microbiol,* 2010, 12, 1553-1561
Turnbull, E. L., Wong, M., Wang, S., et al. Kinetics of expansion of epitope-specific t cell responses during primary hiv-1 infection. *J Immunol,* 2009, 182, 7131-7145
UNAIDS. 90-90-90 an ambitious treatment target to help end the aids epidemic. 2014,
UNAIDS. How aids changed everything - mdg6: 15 years, 15 lessons of hope from the aids response. 2015,
Van Der Ryst, E. Maraviroc - a ccr5 antagonist for the treatment of hiv-1 infection. *Frontiers in immunology,* 2015, 6, 277
Van Heuverswyn, F., Li, Y., Neel, C., et al. Human immunodeficiency viruses: Siv infection in wild gorillas. *Nature,* 2006, 444, 164
Vigerust, D. J., Egan, B. S. and Shepherd, V. L. Hiv-1 nef mediates post-translational down-regulation and redistribution of the mannose receptor. *Journal of leukocyte biology,* 2005, 77, 522-534
Vives, E., Brodin, P. and Lebleu, B. A truncated hiv-1 tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J Biol Chem,* 1997, 272, 16010-16017
Wallet, M. A., Rodriguez, C. A., Yin, L., et al. Microbial translocation induces persistent macrophage activation unrelated to hiv-1 levels or t-cell activation following therapy. *AIDS,* 2010, 24, 1281-1290
Wang, T., Yin, Z., Zhang, Z., et al. Inhibitors of human immunodeficiency virus type 1 (hiv-1) attachment. 5. An evolution from indole to azaindoles leading to the discovery of 1-(4-benzoylpiperazin-1-yl)-2-(4,7-dimethoxy-1h-pyrrolo[2,3-c]pyridin-3-yl)ethane -1,2-dione (bms-488043), a drug candidate that demonstrates antiviral activity in hiv-1-infected subjects. *Journal of medicinal chemistry,* 2009, 52, 7778-7787
Watson, K. and Edwards, R. J. Hiv-1-trans-activating (tat) protein: Both a target and a tool in therapeutic approaches. *Biochemical pharmacology,* 1999, 58, 1521-1528
Wei, X., Decker, J. M., Wang, S., et al. Antibody neutralization and escape by hiv-1. *Nature,* 2003, 422, 307-312
Wesolowska, O., Kuzdzal, M., Strancar, J., et al. Interaction of the chemopreventive agent resveratrol and its metabolite, piceatannol, with model membranes. *Biochimica et biophysica acta,* 2009, 1788, 1851-1860
White, K. L., Chen, J. M., Feng, J. Y., et al. The k65r reverse transcriptase mutation in hiv-1 reverses the excision phenotype of zidovudine resistance mutations. *Antiviral therapy,* 2006, 11, 155-163
Wilen, C. B., Tilton, J. C. and Doms, R. W. Hiv: Cell binding and entry. *Cold Spring Harb Perspect Med,* 2012, 2,
Williamson, M. P., McCormick, T. G., Nance, C. L., et al. Epigallocatechin gallate, the main polyphenol in green tea, binds to the t-cell receptor, cd4: Potential for hiv-1 therapy. *J Allergy Clin Immunol,* 2006, 118, 1369-1374
Wolf, M. C., Freiberg, A. N., Zhang, T., et al. A broad-spectrum antiviral targeting entry of enveloped viruses. *Proc Natl Acad Sci U S A,* 2010, 107, 3157-3162
Wolfe, L. S., Stanley, B. J., Liu, C., et al. Dissection of the hiv vif interaction with human e3 ubiquitin ligase. *J Virol,* 2010, 84, 7135-7139

Worobey, M., Telfer, P., Souquiere, S., et al. Island biogeography reveals the deep history of siv. *Science,* 2010, 329, 1487
Xu, H. X., Wan, M., Dong, H., et al. Inhibitory activity of flavonoids and tannins against hiv-1 protease. *Biol Pharm Bull,* 2000, 23, 1072-1076
Yahi, N., Fantini, J., Henry, M., et al. Structural analysis of reverse transcriptase mutations at codon 215 explains the predominance of t215y over t215f in hiv-1 variants selected under antiretroviral therapy. *Journal of biomedical science,* 2005, 12, 701-710
Yang, G. Y., Li, Y. K., Wang, R. R., et al. Dibenzocyclooctadiene lignans from schisandra wilsoniana and their anti-hiv-1 activities. *J Nat Prod,* 2010, 73, 915-919
Yoder, K. E. and Bushman, F. D. Repair of gaps in retroviral DNA integration intermediates. *J Virol,* 2000, 74, 11191-11200
Young, F. E. The role of the fda in the effort against aids. *Public health reports,* 1988, 103, 242-245
Zhang, C. F., Nakamura, N., Tewtrakul, S., et al. Sesquiterpenes and alkaloids from lindera chunii and their inhibitory activities against hiv-1 integrase. *Chem Pharm Bull (Tokyo),* 2002, 50, 1195-1200
Zhao, H., Neamati, N., Hong, H., et al. Coumarin-based inhibitors of hiv integrase. *J Med Chem,* 1997, 40, 242-249
Zhao, L. F., Iwasaki, Y., Nishiyama, M., et al. Liver x receptor alpha is involved in the transcriptional regulation of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase gene. *Diabetes,* 2012, 61, 1062-1071
Zheng, R., Jenkins, T. M. and Craigie, R. Zinc folds the n-terminal domain of hiv-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci U S A,* 1996, 93, 13659-13664
Zhou, N., Nowicka-Sans, B., McAuliffe, B., et al. Genotypic correlates of susceptibility to hiv-1 attachment inhibitor bms-626529, the active agent of the prodrug bms-663068. *J Antimicrob Chemother,* 2014, 69, 573-581
Zhu, L., Hruska, M., Hwang, C., et al. Pharmacokinetic interactions between bms-626529, the active moiety of the hiv-1 attachment inhibitor prodrug bms-663068, and ritonavir or ritonavir-boosted atazanavir in healthy subjects. *Antimicrob Agents Chemother,* 2015, 59, 3816-3822

Annexes I

Virological and immunological outcomes of elvitegravir-based regimen in a treatment-naïve HIV-2-infected patient

Correspondence

AIDS 2014, **28**:2323–2332

HLA-B*35-PX and HLA-B*35-PY subtype differentiation does not predict observed differences in level of HIV control in a Peruvian MSM cohort

It is well established that host genetics, as well as immunological and virological parameters, influence the rate of HIV disease progression [1,2]. The most consistent associations exist for human leukocyte antigen (HLA) alleles, mainly in the variable HLA-B locus. Of these, HLA-B*35 has been related to accelerated progression with the most deleterious effects linked to a group of subtypes, HLA-B*35 PX (including HLA-B*35 : 02, HLA-B*35 : 03 and HLA-B*35 : 04) [3–8]. These PX alleles differ by as few as one amino acid from other HLA-B*35 subtypes, referred to as HLA-B*35 PY alleles that show a somewhat reduced rate of disease progression [9]. PX and PY alleles have been differentiated on the basis of their F-pocket epitope anchor preference, with PY alleles only allowing for tyrosine (Y) in this position, while PX alleles can accommodate a range of residues [5,8,9]. Individuals expressing HLA-B*35 PX alleles have been reported to mount broader cytotoxic T cell (CTL) responses than PY individuals, which could lend PX alleles some advantage for viral control [8,10,11]. However, the increased affinity of PX alleles to immunoglobulin-like transcript 4 could have opposite effects [5]. In addition, the usefulness of discriminating PX and PY alleles needs to be questioned, as only few alleles have been assigned to either group and this separation is based on poorly characterized HLA-binding motifs.

We studied the association of HLA-B*35 alleles with viral control in an untreated cohort [12] of 235 HIV clade B infected individuals in Lima (Perú) and related these differences to known PX and PY classification, subtype-allele frequencies, T-cell responses restricted by these alleles and sequence similarities in regions forming the allele-specific F-pocket. The selected cohort is exceptionally well suited to study HLA-B*35 subtypes, as it includes 14 different B*35 subtypes, 10 of which are present at a frequency greater than 1% (Fig. 1a) [13]. When comparing viral loads between HLA-B*35-positive and negative individuals (Fig. 1a), all but three alleles showed median viral load above the population median (37 113 copies/ml). Differences were statistically significant for HLA-B*35 : 05 and HLA-B*35 : 09, alleles that are not classified as PX or PY [14]. When HLA-B*35 alleles were grouped into known PX or PY and unclassified alleles, all three groups showed higher viral load than HLA-B*35-negative individuals (Fig. 1b) in line with recently published data [15]. To test whether adding further putative PX and PY alleles would yield stronger differences between allele groups, we classified B*35 alleles on the basis of polymorphisms in the allele's F-pocket. Individuals expressing HLA-B*35 alleles with aspartic acid (D) at position 114 or serine (S) at position 116 showed overall reduced viral load (Fig. 1c). However, this classification did not match the PX/PY assignments, as both putative PX (HLA-B*35 : 09) and putative PY (HLA-B*35 : 05) alleles showed increased viral load and some alleles had changes others than tyrosine (Y) to S at position 116. We also attempted grouping unclassified HLA-B*35 alleles by their preferred F-pocket epitope anchor, but the poorly defined binding motifs for many alleles add great uncertainty to such an approach. To overcome limitations in binding motifs information and to group putative PX/PY alleles based on functional epitope response data, we defined the dominant HLA-B*35 restricted T-cell response in this cohort using overlapping peptide (OLP) reactivity data [16,17] and the HEPITOPE algorithm (http://www.hiv.lanl.gov/content/immunology/hepitopes/index.html) (Fig. 1d). Most OLP for which reactivity was associated with at least one HLA-B*35 allele contained a described HLA-B*35 epitope (nine out of 22) or a HLA-B*35 anchor motif (18/22). However, no associations between OLP reactivity and classification as PX/PY alleles emerged from these analyses.

Together, our data indicate that the classification of HLA-B*35 subtypes into PX and PY alleles does not satisfactorily explain differences in median VL. Given the severe limitations in allele-specific binding motifs information (which form the base of PX/PY classification), these results are not overly surprising. This does not imply that no differences exist in allele-specific epitope responses but, as it is, the classification of HLA-B*35 alleles into these two groups is incomplete and likely not readily explainable by allele sequence polymorphisms. Rather, future analyses will likely need to test for potential linkage disequilibrium with HLA-A or HLA-C alleles or components of the innate immune response (e.g. KIR polymorphisms) [18–21]. This will require larger cohorts to allow for statistically robust comparisons of the combined effect of HLA and KIR alleles on viral control. What the current data indicate though is that it is unlikely that, even from such expanded studies, a grouping of alleles with or without any evident shared pattern of sequence polymorphisms would explain in-vivo HIV control.

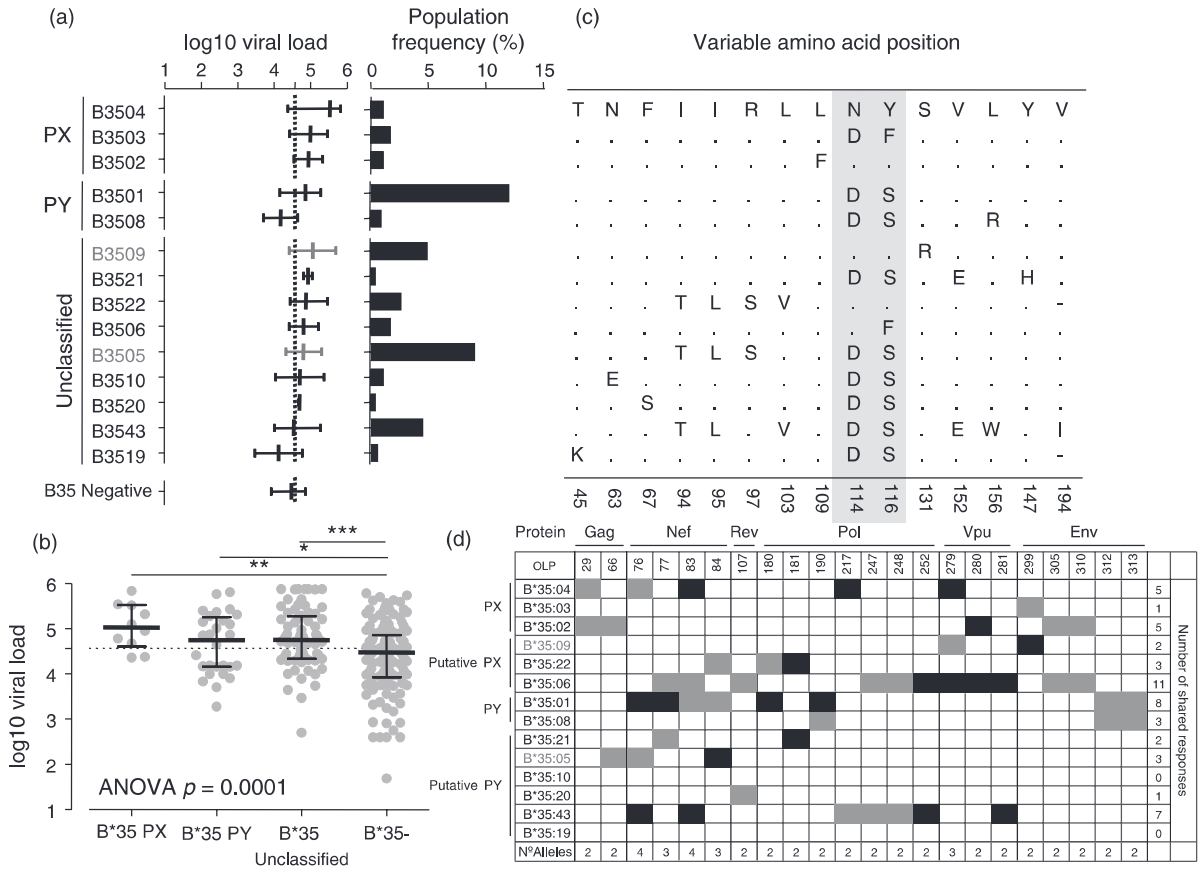


Fig. 1. Viral load distribution, dominant HLA-B35 subtype restricted T-cell responses and allele sequence polymorphisms in the Peruvian MSM cohort. (a) HLA*B35 subtype alleles classified as PX, PY or unclassified alleles and the median plasma viral load (VL) in individuals expressing these alleles are shown. Alleles with statistically significant differences in VL are indicated by grey bars ($P < 0.05$); population median VL is indicated by a dotted line. Population allele frequency of the various subtypes is indicated by bars. (b) VL of HLA-B*35 alleles classified as PX, PY and unclassified compared with individuals bearing no HLA-B*35 subtype alleles. VLs were compared using one-way ANOVA and two-by-two comparisons using Student's *t*-test (***$P < 0.001$, ** $P < 0.01$, *$P < 0.05$). (c) Variable positions in an amino acid alignment of the HLA-B*35 subtypes present. Positions 114 and 116 are shaded in gray as the only polymorphisms that showed significantly different VL ($P < 0.05$). HLA subtype sequence information was obtained from http://www.ebi.ac.uk/ipd/imgt/hla/allele.html. (d) Associations between responses to particular OLP and HLA-B*35 subtypes were determined using the HEPITOPE algorithm (http://www.hiv.lanl.gov/content/immunology/hepitopes) and the presence of defined epitopes in the OLP sequence was revised using ELF (http://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html). The OLP response profile is shown for each HLA-B*35 subtype allele with responses scoring with a one-sided Fisher's exact test of $P < 0.01$ marked in black and those with an association of $0.01 < P < 0.05$ in gray. The bottom line indicates the numbers of alleles that share reactivities to the same OLP. Only responses to OLP shared by at least two HLA alleles are indicated.

Acknowledgements

This work was supported by NIH-NIDCR R01 DE018925-04 (C.B.), the HIVACAT programme, the CUTHIVAC project of the EU FP7 programme and PI12/00529 grant of the Instituto de Salut Carlos III. C.B. is an ICREA (Institució Catalana de Recerca i Estudis Avançats) Senior Research Professor.

Conflicts of interest

There are no conflicts of interest.

Alex Olvera[a], Carmela Ganoza[b], Susana Pérez-Álvarez[a,c], William Hildebrand[d], Jorge Sanchez[b,e] and Christian Brander[a,c,f,g], *[a]IrsiCaixa AIDS Research Institute – HIVACAT, Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain, [b]Asociación Civil IMPACTA Salud y Educacion, Lima, Peru, [c]Universitat Autònoma de Barcelona, Barcelona, Spain, [d]University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, [e]Department of Global Health, University of Washington, Seattle, Washington, USA, [f]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, and [g]Universitat de Vic, Spain.*

Correspondence to Alex Olvera, IrsiCaixa AIDS Research Institute, Hospital Germans Trias i Pujol, Ctra. de Canyet s/n, 08916 Badalona, Barcelona, Spain. Tel: +34 934 656 374; e-mail: aolvera@irsicaixa.es

Received: 31 May 2014; revised: 1 July 2014; accepted: 2 July 2014.

References

1. McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, *et al*. **Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls.** *PLoS Pathog* 2013; **9**:e1003515.
2. Mothe B, Ibarrondo J, Llano A, Brander C. **Virological, immune and host genetics markers in the control of HIV infection.** *Dis Markers* 2009; **27**:105–120.
3. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, *et al*. **HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.** *Science* 1999; **283**:1748–1752.
4. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, Buchbinder S, *et al*. **AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis.** *Nat Med* 2005; **11**:1290–1292.
5. Huang J, Goedert JJ, Sundberg EJ, Cung TD, Burke PS, Martin MP, *et al*. **HLA-B*35-Px-mediated acceleration of HIV-1 infection by increased inhibitory immunoregulatory impulses.** *J Exp Med* 2009; **206**:2959–2966.
6. Matthews PC, Koyanagi M, Kloverpris HN, Harndahl M, Stryhn A, Akahoshi T, *et al*. **Differential clade-specific HLA-B*3501 association with HIV-1 disease outcome is linked to immunogenicity of a single Gag epitope.** *J Virol* 2012; **86**:12643–12654.
7. Scorza Smeraldi R, Fabio G, Lazzarin A, Eisera N, Uberti Foppa C, Moroni M, Zanussi C. **HLA-associated susceptibility to AIDS: HLA B35 is a major risk factor for Italian HIV-infected intravenous drug addicts.** *Hum Immunol* 1988; **22**:73–79.
8. Willberg CB, Garrison KE, Jones RB, Meiklejohn DJ, Spotts G, Liegler TJ, *et al*. **Rapid progressing allele HLA-B35 Px restricted anti-HIV-1 CD8R T cells recognize vestigial CTL epitopes.** *PLoS One* 2010; **5**:e10249.
9. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, *et al*. **Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS.** *N Engl J Med* 2001; **344**:1668–1675.
10. Jin X, Gao X, Ramanathan M Jr, Deschenes GR, Nelson GW, O'Brien SJ, *et al*. **Human immunodeficiency virus type 1 (HIV-1)-specific CD8R -T-cell responses for groups of HIV-1-infected individuals with different HLA-B*35 genotypes.** *J Virol* 2002; **76**:12603–12610.
11. Steers NJ, Currier JR, Kijak GH, di Targiani RC, Saxena A, Marovich MA, *et al*. **Cell type-specific proteasomal processing of HIV-1 Gag-p24 results in an altered epitope repertoire.** *J Virol* 2010; **85**:1541–1553.
12. Zuniga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, *et al*. **Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control.** *J Virol* 2006; **80**:3122–3125.
13. Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, *et al*. **Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities.** *J Virol* 2004; **78**:2187–2200.
14. Mori M, Wichukchinda N, Miyahara R, Rojanawiwat A, Pathipvanich P, Maekawa T, *et al*. **HLA-B*35: 05 is a protective allele with a unique structure among HIV-1 CRF01_AE-infected Thais, in whom the B*57 frequency is low.** *AIDS* ; **28**:959–967.
15. Juarez-Molina CI, Valenzuela-Ponce H, Avila-Rios S, Garrido-Rodriguez D, Garcia-Tellez T, Soto-Nava M, *et al*. **Impact of HLA-B*35 subtype differences on HIV disease outcome in Mexico.** *AIDS* 2014; May 21 [Epub ahead of print].
16. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, *et al*. **Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes.** *Nat Immunol* 2006; **7**:173–178.
17. Zuniga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, *et al*. **HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations.** *PLoS One* 2013; **8**:e74442.
18. Boulet S, Sharafi S, Simic N, Bruneau J, Routy JP, Tsoukas CM, Bernard NF. **Increased proportion of KIR3DS1 homozygotes in HIV-exposed uninfected individuals.** *AIDS* 2008; **22**:595–599.
19. Boulet S, Song R, Kamya P, Bruneau J, Shoukry NH, Tsoukas CM, Bernard NF. **HIV protective KIR3DL1 and HLA-B genotypes influence NK cell function following stimulation with HLA-devoid cells.** *J Immunol* 2010; **184**:2057–2064.
20. Hong HA, Paximadis M, Gray GE, Kuhn L, Tiemessen CT. **KIR2DS4 allelic variants: differential effects on in utero and intrapartum HIV-1 mother-to-child transmission.** *Clin Immunol* 2013; **149**:498–508.
21. Merino A, Malhotra R, Morton M, Mulenga J, Allen S, Hunter E, *et al*. **Impact of a functional KIR2DS4 allele on heterosexual HIV-1 transmission among discordant Zambian couples.** *J Infect Dis* 2011; **203**:487–495.

DOI:10.1097/QAD.0000000000000403

Ending the invisibility of sex workers in the US HIV/AIDS surveillance and prevention strategy

Globally, sex workers, that is, individuals who trade sex for money or other goods [1], suffer a disproportionate HIV burden [2]. Most nations provide sex worker-specific HIV infection and risk behavior estimates routinely in the Joint United Nations Programme on HIV/AIDS (UNAIDS) global epidemic reports. By contrast, sex workers remain conspicuously absent in US domestic HIV/AIDS policy. Our national surveillance system fails to disaggregate HIV risk, disease burden, and treatment outcomes for sex workers; rather sex exchange is grouped within 'high risk heterosexual sex'. The HIV/AIDS report form for case surveillance also omits sex exchange as a potential transmission mode. Our CDC website on sex work and HIV lacks both data and intervention recommendations specific to this population. This dearth of data stymies our ability to develop a national HIV response for those in sex work, and perpetuates their invisibility and marginalization in US policy and interventions.

Recent evidence confirms a robust sex industry in the United States [3], and sex exchange is documented in patient populations of sexually transmitted infections (STI) clinics [4], family planning clinics [5], and substance use treatment programs [6]. The 2010 high-risk heterosexual wave of the National HIV Behavior Surveillance System (NHBS) reported a past-year sex exchange prevalence of 16.6% among men and women combined [7]. Further, the HIV prevalence was significantly higher among exchangers (3.7 vs. 2.1%) [7] suggesting profound HIV implications.

Why the lack of HIV surveillance devoted to sex work in the United States? One likely explanation is the challenge in characterizing sex exchange, as it is widely recognized that many who exchange sex do not identify as sex workers, and that some are involved against their will and/or as minors. Within the NHBS and elsewhere, we lack clarity on the nature of sex trade with regard to context, consent or exploitation, degree of formality, and resources received. An evidence-informed response to this ambiguity requires surveillance with a breadth of individuals who trade sex, with enhanced measures to clarify their experiences, rather than data suppression for fear of misrepresentation.

Sex worker marginalization in US HIV policy arguably has deeper roots. The contentious US global policy context likely exacerbated the reluctance to acknowledge sex work within the United States. The 2003 US President's Emergency Plan for AIDS Relief (PEPFAR) included a policy known as the antiprostitution loyalty oath (APLO, the so-called 'Prostitution Pledge'), which required beneficiaries to explicitly oppose prostitution, its legalization, and sex trafficking. By conflating sex work and sex trafficking, the APLO furthered the marginalization of sex work. Needed research and intervention efforts, which might have helped evaluate risks related to HIV and violence across the spectrum of sex exchange did not occur. The 2013 US Supreme Court ruling the APLO unconstitutional should pave the way for a new era of policy, including sex industry-related research to monitor risk behavior and infection trends, and inform the extent of exploitation.

Why disaggregate sex work from other forms of high-risk sexual transmission? Globally, sex workers face significant, unique challenges that amplify HIV risk and exacerbate exclusion from essential HIV/STI services. Criminalization leaves sex workers vulnerable to pressure for unprotected sex, condom confiscation by police as evidence of sex work, and sexual violence with impunity [8–11]. Extensive physical and sexual violence perpetrated by clients, police, and others enable HIV risk behavior and infection [9,11–15]. The US HIV epidemic dynamics, and their social determinants, also demand exploration of exchange sex as a transmission route. As is the case globally, in the United States, HIV is situated in poverty. It is heavily concentrated among African-Americans; the rate among African American women is 20 times that of white women [16]. The confluence of sex and racial/ethnic discrimination, and economic deprivation strongly suggest sex exchange as a mechanism through which social determinants influence HIV.

The marginalization and criminalization of sex work through policy and practice are challenges to the success of the National HIV/AIDS Strategy and to control of the US HIV epidemic. Surveillance is a cornerstone of epidemiology. Without clarity on the extent and nature of sex work in the US population, we risk an incomplete and flawed understanding of HIV dynamics. And we risk the continued exclusion of sex workers from services they urgently need. We must not let discomfort with the sex industry undermine surveillance or evidence-based prevention and interventions. The improved policy climate post the Supreme Court decision is an opportunity for urgently needed reform. It is time for the US HIV surveillance system to include sex workers as a key population, and for sex workers and their communities to come out of the shadows and into urgently needed prevention, treatment, and care services.

Acknowledgements

Johns Hopkins Center for AIDS Research (JHU CFAR; NIAID 1P30AI094189; PI Chaisson).

Conflicts of interest

There are no conflicts of interest.

Michele R. Decker[a,b], Chris Beyrer[b,c] and Susan G. Sherman[b,c], *[a]Department of Population, Family & Reproductive Health, [b]Center for Public Health and Human Rights, and [c]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.*

Correspondence to Michele R. Decker, ScD, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, E4142, Baltimore, MD 21212, USA.
Tel: +1 410 502 2747; fax: +1 410 955 2303;
e-mail: mdecker@jhu.edu

Received: 13 June 2014; revised: 10 July 2014; accepted: 10 July 2014.

References

1. UNAIDS. *UNAIDS Guidance Note on HIV and Sex Work.* Geneva: UNAIDS; 2012.
2. Baral S, Beyrer C, Muessig K, *et al.* **Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis.** *Lancet Infect Dis* 2012; **12**:538–549.
3. Urban Institute. *The Hustle: Economics of the underground commercial sex industry.* Washington, DC: Urban Institute; 2014.
4. Gindi RM, Erbelding EJ, Page KR. **Sexually transmitted infection prevalence and behavioral risk factors among Latino and non-Latino patients attending the Baltimore City STD clinics.** *Sex Transm Dis* 2010; **37**:191–196.
5. Decker MR, Miller E, McCauley HL, *et al.* **Sex trade among young women attending family-planning clinics in Northern California.** *Int J Gynaecol Obstet* 2012; **117**:173–177.
6. Burnette ML, Lucas E, Ilgen M, Frayne SM, Mayo J, Weitlauf JC. **Prevalence and health correlates of prostitution among patients entering treatment for substance use disorders.** *Arc Gen psychiatry* 2008; **65**:337–344.
7. Miles IJ, Le BC, Wejnert C, Oster A, DiNenno E, Paz-Baily G. **HIV infection among heterosexuals at increased risk - United States 2010.** *MMWR Morb Mortal Wkly Rep* 2013; **62**:183–188.

8. Sex workers at risk: Condoms as evidence of prostitution in four US cities. Human Rights Watch; 2012.
9. Erausquin JT, Reed E, Blankenship KM. **Police-related experiences and HIV risk among female sex workers in Andhra Pradesh, India.** *J Infect Dis* 2011; **204 (Suppl 5)**:S1223–S1228.
10. Rhodes T, Simic M, Baros S, Platt L, Zikic B. **Police violence and sexual risk among female and transvestite sex workers in Serbia: qualitative study.** *BMJ* 2008; **337**:a811.
11. Decker MR, Crago AL, Chu SKH, *et al*. **Human rights violations against sex workers: burden and effect on HIV.** *Lancet* 2014.; [Epub ahead of print]
12. Decker MR, McCauley HL, Phuengsamran D, Janyam S, Seage GR 3rd, Silverman JG. **Violence victimisation, sexual risk and sexually transmitted infection symptoms among female sex workers in Thailand.** *Sex Transm Infect* 2010; **86**:236–240.
13. Decker MR, Wirtz AL, Baral SD, *et al*. **Injection drug use, sexual risk, violence and STI/HIV among Moscow female sex workers.** *Sex Transm Infect* 2012; **88**:278–283.
14. Deering KN, Shannon K, Blanchard JF, *et al*. **Experiencing occupational violence is associated with inconsistent condom use among female sex workers in Southern India: Emphasizing the need for structural HIV interventions.** *Can J Infect Dis Med Microbiol* 2011; **22**:90B.
15. Shannon K, Strathdee SA, Shoveller J, Rusch M, Kerr T, Tyndall MW. **Structural and environmental barriers to condom use negotiation with clients among female sex workers: implications for HIV-prevention strategies and policy.** *Am J Public Health* 2009; **99**:659–665.
16. CDC. *HIV Among African Americans Fact Sheet.* Atlanta, GA: Centers for Disease Control and Prevention; 2014.

DOI:10.1097/QAD.0000000000000411

Triple secondary neoplasms: penis, lip and oral cavity in an AIDS patient treated with pegylated liposomal doxorubicin for cutaneous Kaposi's sarcoma

Pegylated liposomal doxorubicin (PLD) is a first-line therapy for AIDS Kaposi's sarcoma [1]. Kaposi's sarcoma lesions contain dilated vascular spaces with extravasated erythrocytes; PLD uptake due to extravasation of liposomes was anticipated, offering advantage over conventional formulation. Skin Kaposi's sarcoma biopsies of patients under PLD reached 10–20 times doxorubicin concentrations to that of the normal skin [1].

Complete regression of Kaposi's sarcoma solely with highly active antiretroviral therapy (HAART) has been reported [2–5], questioning the need of chemotherapy for this angioproliferative disease produced by human herpes virus 8, as chemotherapy can increase immunosuppression. Worsening of Kaposi's sarcoma after HAART initiation is a manifestation of immune reconstitution syndrome [6,7], and without visceral involvement, it is a limited phenomenon; as immune function recovers, Kaposi's sarcoma regresses without further therapy.

Secondary malignancy is a concern from chemotherapy use, but its role is difficult to demonstrate as other factors can contribute to its development. A substantial high incidence of secondary neoplasms – 9% among Kaposi's sarcoma patients treated with PLD – was recently reported [8], and oral cavity squamous cell carcinoma has also been reported in PLD-exposed patients [9–12], raising concerns of a possible contributing effect of PLD in the development of these neoplasms.

Here, we present the case of a 37-year-old man with a 20-year history of cigarette smoking (average 5 cigarettes per day), who was diagnosed with AIDS on May 2011; he presented with wasting syndrome, oropharyngeal-candidiasis, disseminated-*Mycobacterium avium Intracelulare* and cutaneous Kaposi's sarcoma. His viral load was 631 471 copies/ml and $CD4^+$ 144 cells/ml.

The patient was started on Truvada + Kaletra and, 4 months later, the viral load was undetectable and the $CD4^+$ had risen to 196 cells/ml. Following this, he received five monthly cycles with bleomycin 17 units, vincristine 2 mg and dexamethasone 16 mg, after which monthly PLD was started. On October 2012, after eight cycles of PLD, the patient noted multiple erythematous plaques in his penis; his HIV viral load remained undetectable, but he had 73 cells/ml $CD4^+$. A penis biopsy revealed epidermoid carcinoma, which was successfully treated with topical 5% fluouracil. In November 2012, the patient complained of lip and oral mucosa lesions, for which the biopsies reported high-grade dysplasia. Several variants of HPV were identified: in the penis carcinoma and ureteral smear 16, in lips 33 and 35, in oral mucosa 35, in fresh oral mucosal smear 16 and 33, and in perianal condylomas 16 and 11.

To our knowledge, there are no reports on adverse effects of PLD in HIV patients with concomitant mucosal HPV infections. However, since PLD concentrates for prolonged periods in mucous membranes, it is likely that it might alter local immunity since. PLD immunosuppression could also increase HPV replication [13], and decrease $CD4^+$ cell counts [8] as for the case herein reported, in whom despite HIV viral suppression, $CD4^+$ cells decrease with PDL administration (Fig. 1).

Although there has been an increasing incidence of non-AIDS-defining malignancies in the HAART era, the incidence of cancers of the oral cavity and penis in HIV patients is rare (<1%) [14].

Kaposi's sarcoma remission can be achieved solely with HAART and thus the use of PLD in AIDS Kaposi's sarcoma should be questioned since it can chronically

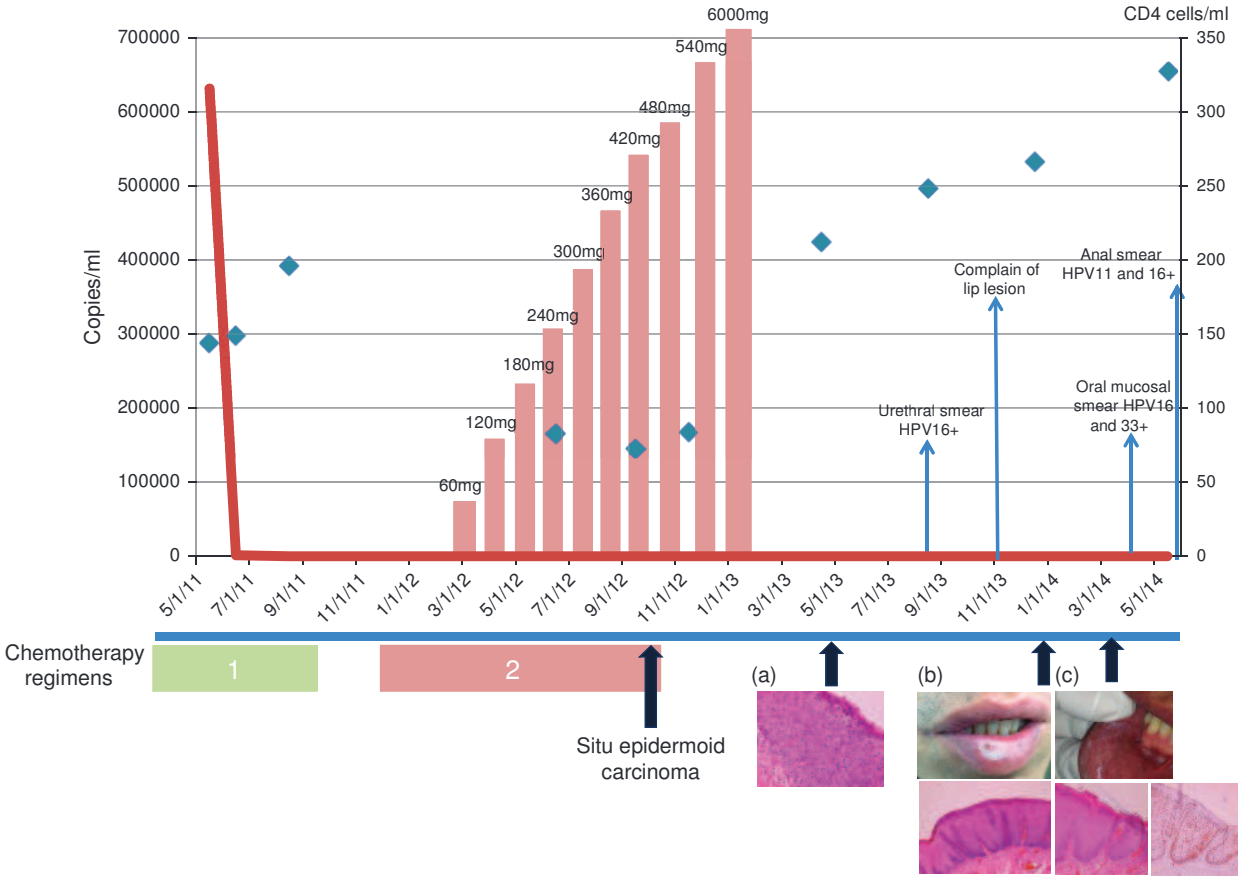


Fig. 1. Temporal events of patient's case. Administration of chemotherapy*: (1) bleomycin 17 mg + vincristine 2 g + dexamethasone 16 mg (5 cycles); (2) pegylated liposomal doxorubicin (PLD) 60 mg + dexamethasone 16 mg (10 cycles). Red line, HIV viral load; blue diamonds, $CD4^+$ cell count; pink bars, accumulated PLD. (a) Hematoxylin and eosin (HE) 40×. Penis mucosa with marked cytological atypia involving the entire epithelium. Note numerous atypical mitosis in the upper zones. *In-situ epidermoid penis carcinoma, human papillomavirus (HPV)16+. (b) HE 10×. Lip mucosa with acanthosis, cytologic atypia involving half epithelium. Note koilocytosis in the upper. High grade dysplasia (HGD) of the lip HPV 33 and 35+. (c) Oral mucosa. HE 4× and Ki67 immunostain. Oral mucosa with irregular acanthosis and cytologic atypia. Involving more than half of the epithelium. Note Ki67 positive in the entire epithelium, HGD of oral mucosa HPV35+.

affect the control of HPV infection and thus increase the risk of HPV-related neoplasms.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

Patricia Volkow[a], Marcela Lizano[b], Adela Carrillo-García[b], Delia Pérez-Montiel[c] and Pamela Garciadiego[a], *[a]Infectious Disease Department, Instituto Nacional de Cancerología, Mexico D.F., [b]Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología-Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico D.F., and [c]Pathology Department, Instituto Nacional de Cancerología, Mexico D.F., Mexico.*

Correspondence to Patricia Volkow, MD, Instituto Nacional de Cancerología, Av. San Fernando No. 22, Col. Sección XVI, Deleg. Tlalpan, 14080 Mexico, D.F., Mexico.
Tel: +52 55 5628 0400 ext. 355;
fax: +52 55 5628 0447; e-mail: pvolkowf@gmail.com

Received: 5 July 2014; revised: 16 July 2014; accepted: 16 July 2014.

References

1. Udhrain A, Skubitz KM, Northfelt DW. **Pegylated liposomal doxorubicin in the treatment of AIDS-related Kaposi's sarcoma.** *Int J Nanomed* 2007; **2**:345–352.
2. Murphy M, Armstrong D, Sepkowitz KA, Ahkami RN, Myskowski PL. **Regression of AIDS-related Kaposi's sarcoma following treatment with an HIV-1 protease inhibitor.** *AIDS* 1997; **11**:261–262.
3. Aboulafia DM. **Regression of acquired immunodeficiency syndrome-related pulmonary Kaposi's sarcoma after highly active antiretroviral therapy.** *Mayo Clin Proc* 1998; **73**:439–443.

4. Bower M, Weir J, Francis N, Newsom-Davis T, Powles S, Crook T, *et al*. **The effect of HAART in 254 consecutive patients with AIDS-related Kaposi's sarcoma.** *AIDS* 2009; **23**:1701–1706.
5. Pellet C, Chevret S, Blum L, Gauvillé C, Hurault M, Blanchard G, *et al*. **Virologic and immunologic parameters that predict clinical response of AIDS-associated Kaposi's sarcoma to highly active antiretroviral therapy.** *J Invest Dermatol* 2001; **117**:858–863.
6. Leidner RS, Aboulafia DM. **Recrudescent Kaposi's sarcoma after initiation of HAART: a manifestation of immune reconstitution syndrome.** *AIDS Patient Care STDS* 2005; **19**:635–644.
7. Bower M, Nelson M, Young AM, Thirlwell C, Newsom-Davis T, Mandalia S, *et al*. **Immune reconstitution inflammatory syndrome associated with Kaposi's sarcoma.** *J Clin Oncol* 2005; **23**:5224–5228.
8. Martín-Carbonero L, Palacios R, Valencia E, Saballs P, Sirera G, Santos I, *et al*. **Caelyx/Kaposi's Sarcoma Spanish Group. Long-term prognosis of HIV-infected patients with Kaposi sarcoma treated with pegylated liposomal doxorubicin.** *Clin Infect Dis* 2008; **47**:410–417.
9. Cannon TL, Lai DW, Hirsch D, Delacure M, Downey A, Kerr AR, *et al*. **Squamous cell carcinoma of the oral cavity in nonsmoking women: a new and unusual complication of chemotherapy for recurrent ovarian cancer?** *Oncologist* 2012; **17**:1541–1546.
10. Randon G, Nicoletto MO, Milite N, Muggia F, Conte P. **Squamous cell carcinoma of the oral cavity in a woman with a 9-year history of ovarian cancer: is exposure to pegylated liposomal Doxorubicin a factor?** *Oncologist* 2014; **19**:429.
11. Matsuo K, Blake EA, Yessaian AA, Roman LD. **Long-term pegylated liposomal doxorubicin use and oromaxillary squamous cell carcinoma in endometrial cancer.** *Oncologist* 2012; **17**:1598–1599.
12. Gu P, Wu J, Sheu M, Myssiorek D, Cohen R. **Aggressive squamous cell carcinoma of the oral tongue in a woman with metastatic giant cell tumor treated with pegylated liposomal doxorubicin.** *Oncologist* 2012; **17**:1596–1597.
13. Wieland U, Kreuter A, Pfister H. **Human papillomavirus and immunosuppression.** *Curr Probl Dermatol* 2014; **45**:154–165.
14. **Cancer burden in the HIV-infected population in the United States.** *J Natl Cancer Inst* 2011; **103**:753–762.

DOI:10.1097/QAD.0000000000000420

Virological and immunological outcomes of elvitegravir-based regimen in a treatment-naïve HIV-2-infected patient

Patients infected with HIV type 2 (HIV-2) have limited therapeutic options. HIV-2 is naturally resistant to most non-nucleoside reverse transcriptase inhibitors (NNRTIs) and the fusion inhibitor enfuvirtide [1]. Furthermore, several protease inhibitor-containing regimens showed weak inhibitory properties against HIV-2 *in vivo*. In this regard, the new integrase strand transfer inhibitors (INSTIs) may represent a valuable therapy for HIV-2 infection.

Approved antiretroviral drugs binding to the HIV-1 integrase catalytic site [raltegravir (RAL), elvitegravir (EVG), and dolutegravir (DTG)] act as inhibitors of the strand transfer reaction. They are all effective against HIV-2 in phenotypic assays despite only 60% sequence identity between HIV-1 and HIV-2 integrase genes. A limited number of HIV-2 case reports were reported for successful RAL treatment based on an optimized background regimen [2–6]. In 2012, EVG was approved as a component of the first integrase inhibitor-based single tablet regimen Stribild for once-daily treatment of HIV-1 infection in therapy-naïve adults. Stribild is a fixed-dose combination drug containing cobicistat-boosted EVG and emtricitabine (FTC) and tenofovir disoproxil fumarate (TDF). In two large phase III trials, this regimen showed efficient and sustained suppression of HIV-1 RNA levels to less than 50 copies/ml for up to 144 weeks [7,8].

Here, we describe the first case of EVG use in a treatment-naïve HIV-2 patient. A 57-year-old man originating from Ivory Coast was diagnosed with HIV-2 in 2008. Plasma HIV-2-RNA viral load and $CD4^+$ cell counts were recorded during 6 years before treatment initiation and 8 months thereafter (Fig. 1). The viral load was measured using a quantitative real-time PCR [9].

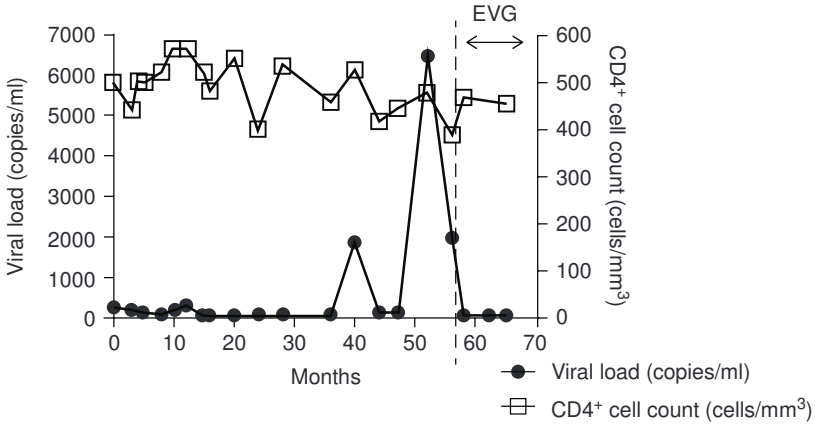


Fig. 1. $CD4^R$ cell count and viral load of the patient over time. $CD4^+$ cell count was measured in cells/ml. Viral load was measured using quantitative real-time PCR in copies/ml. Initiation of elvitegravir (EVG) therapy is indicated with a dashed line.

The reverse transcriptase, protease and integrase genes of the patient's plasma viruses were sequenced using an in-house method [10]. At the time of the diagnosis, the viral load was 230 copies/ml and the $CD4^+$ cell count was 500 cells/ml. The patient harboured a HIV-2 group B virus displaying no resistance mutation in the three genes. The $CD4^+$ cell count decreased to 386 cells/ml and the viral load was 1975 copies/ml in July 2013. The EVG-based regimen was given in August 2013 when Stribild became available in Luxembourg. One month after treatment initiation, as well as at months 5 and 8, the viral load was undetectable. $CD4^+$ cell count increased to 466 copies/ml at month 1 and remained further stable at 450 cells/ml at month 8.

This is the first report describing the successful treatment response of EVG in an HIV-2- infected patient. It demonstrates that EVG boosted by the pharmacokinetic enhancer cobicistat in combination with FTC and TDF was effective in suppressing the viral load and increasing the $CD4^+$ cell count in vivo. EVG, FTC and TDF exhibited potent antiviral efficacy against wild-type group A and group B HIV-2 in vitro [11–13]. RAL and EVG have a similar structure, show a low-to-moderate genetic barrier to resistance, but are highly potent in first-line HIV-1 antiretroviral therapy [14]. The genetic pathways leading to RAL and EVG resistance are similar, leading to extensive cross-resistance between the two INSTIs. According to all in-vitro data, the major mutations leading to the resistance to EVG in HIV-2 might be E92Q/Y143C, T97A/Y143C, Q148K, Q148R, G140S/Q148R, and N155H, but were not reported so far in the HIV-2-infected patient [15]. DTG is the second generation of INSTI; it conserves antiviral efficacy in RAL/EVG-resistant strains and offers a currently available robust second-line therapy [16].

Sustained antiviral activity of EVG requires the presence of a supportive background and the combination with FTC and TDF demonstrated virological efficacy non-inferior to that of the once-daily combination with the NNRTI efavirenz (EFV) or ritonavir-boosted atazanavir in HIV-1 clinical trials [7,8]. EVG has a safety profile similar to RAL and presents the advantage of a single-tablet regimen co-formulation for once-daily dosing. Fixed-dose drug combinations result in higher treatment adherence, more persistent viral suppression and reduced resistance selection [17]. Co-formulated EFV/FTC/TDF once-daily therapy is among the preferred regimens of first-line therapy in all guidelines since its approval. Since NNRTI is not appropriate in HIV-2 therapy, the co-formulation based on EVG could represent an important niche for HIV-2-infected patients. In conclusion, this case report describes the first short-term immunological and virological efficacy of EVG, and needs to be supported by further studies. No randomized clinical trials have been launched for HIV-2 treatment. Stribild might be a drug of choice to be evaluated in resource-limited countries for the initial treatment of HIV-2 infection due to its combination with FTC/TDF [18] and its good tolerance.

Acknowledgements

Y.Z. and C.S.-D. wrote the manuscript. C.L. performed the sequencing of the PRO-RT and integrase genes of the patient. V.A. monitored the patient and prescribed elvitegravir to the patient, and wrote the manuscript.

The authors acknowledge the 'Fonds National de la Recherche' (FNR) of Luxembourg for the PhD grant of Y.Z. (PHD AFR1189522) and 'La Fondation recherche sur le SIDA' of Luxembourg for financial support.

Funding from other organization (Fondation Recherche sur le SIDA).

Conflicts of interest

There are no conflicts of interest.

Yue Zheng[a,b], Christine Lambert[a], Vic Arendt[a,c] and Carole Seguin-Devaux[a], *[a]CRP-Santé, Laboratory of Retrovirology, [b]CRP-Santé, Laboratory of Cellular and Molecular Oncology, and [c]Centre Hospitalier de Luxembourg, Department of Infectious Diseases, Luxembourg City, Luxembourg.*

Correspondence to Yue Zheng, CRP-Santé, L-1526, Luxembourg City, Luxembourg.
E-mail: yuezh15@gmail.com

Received: 4 July 2014; revised: 10 July 2014; accepted: 16 July 2014.

References

1. Menendez-Arias L, Alvarez M. **Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection.** *Antiviral Res* 2014; **102**:70–86.
2. Peterson K, Ruelle J, Vekemans M, Siegal FP, Deayton JR, Colebunders R. **The role of raltegravir in the treatment of HIV-2 infections: evidence from a case series.** *Antivir Ther* 2012; **17**:1097–1100.
3. Wandeler G, Furrer H, Rauch A. **Sustained virological response to a raltegravir-containing salvage therapy in an HIV-2-infected patient.** *AIDS* 2011; **25**:2306–2308.
4. Armstrong-James D, Stebbing J, Scourfield A, Smit E, Ferns B, Pillay D, *et al.* **Clinical outcome in resistant HIV-2 infection treated with raltegravir and maraviroc.** *Antiviral Res* 2010; **86**:224–226.
5. Damond F, Lariven S, Roquebert B, Males S, Peytavin G, Morau G, *et al.* **Virological and immunological response to HAART regimen containing integrase inhibitors in HIV-2-infected patients.** *AIDS* 2008; **22**:665–666.
6. Francisci D, Martinelli L, Weimer LE, Zazzi M, Floridia M, Masini G, *et al.* **HIV-2 infection, end-stage renal disease and protease inhibitor intolerance: which salvage regimen?** *Clin Drug Investig* 2011; **31**:345–349.

7. Clumeck N, Molina JM, Henry K, Gathe J, Rockstroh JK, DeJesus E, *et al.* **A randomized, double-blind comparison of single-tablet regimen elvitegravir/cobicistat/emtricitabine/tenofovir DF vs. ritonavir-boosted atazanavir plus emtricitabine/tenofovir DF for initial treatment of HIV-1 infection: analysis of week 144 results.** *J Acquir Immune Defic Syndr* 2014; **65**:e121–e124.
8. Wohl DA, Cohen C, Gallant JE, Mills A, Sax PE, Dejesus E, *et al.* **A randomized, double-blind comparison of single-tablet regimen elvitegravir/cobicistat/emtricitabine/tenofovir DF versus single-tablet regimen efavirenz/emtricitabine/tenofovir DF for initial treatment of HIV-1 infection: analysis of week 144 results.** *J Acquir Immune Defic Syndr* 2014; **65**:e118–120.
9. Ruelle J, Mukadi BK, Schutten M, Goubau P. **Quantitative real-time PCR on Lightcycler for the detection of human immunodeficiency virus type 2 (HIV-2).** *J Virol Methods* 2004; **117**:67–74.
10. Bercoff DP, Triqueneaux P, Lambert C, Oumar AA, Ternes AM, Dao S, *et al.* **Polymorphisms of HIV-2 integrase and selection of resistance to raltegravir.** *Retrovirology* 2010; **7**:98.
11. Andreatta K, Miller MD, White KL. **HIV-2 antiviral potency and selection of drug resistance mutations by the integrase strand transfer inhibitor elvitegravir and NRTIs emtricitabine and tenofovir in vitro.** *J Acquir Immune Defic Syndr* 2013; **62**:367–374.
12. Witvrouw M, Pannecouque C, Switzer WM, Folks TM, De Clercq E, Heneine W. **Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis.** *Antivir Ther* 2004; **9**:57–65.
13. Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, *et al.* **Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137).** *J Virol* 2008; **82**:764–774.
14. Geretti AM, Armenia D, Ceccherini-Silberstein F. **Emerging patterns and implications of HIV-1 integrase inhibitor resistance.** *Curr Opin Infect Dis* 2012; **25**:677–686.
15. Shimura K, Kodama EN. **Elvitegravir: a new HIV integrase inhibitor.** *Antivir Chem Chemother* 2009; **20**:79–85.
16. Castagna A, Maggiolo F, Penco G, Wright D, Mills A, Grossberg R, *et al.* **Dolutegravir in antiretroviral-experienced patients with raltegravir- and/or elvitegravir-resistant HIV-1: 24-week results of the phase III VIKING-3 study.** *J Infect Dis* 2014; **210**:354–362.
17. Llibre JM, Clotet B. **Once-daily single-tablet regimens: a long and winding road to excellence in antiretroviral treatment.** *AIDS Rev* 2012; **14**:168–178.
18. Peterson K, Jallow S, Rowland-Jones SL, de Silva TI. **Antiretroviral therapy for HIV-2 infection: recommendations for management in low-resource settings.** *AIDS Res Treat* 2011; **2011**:463704.

DOI:10.1097/QAD.0000000000000414

Etravirine and rilpivirine-specific mutations selected by efavirenz and nevirapine exposure in patients infected with HIV-1 non-B subtypes

We have previously published on the occurrence of drug resistance mutations typically associated with etravirine or ripilvirine exposure that appear to have been selected instead by efavirenz and nevirapine in non-B HIV-1 subtypes [1]. The original HIV-1 subtype assignment in our article was based upon an analysis using the Stanford HIV Drug Resistance Database (hivdb.stanford.org), which resulted in two of the sequences identified as CRF01_AE and one as a B subtype. As these HIV-1 sequences are not common to the region, a more rigorous analysis was performed on the sequences. Thus, subsequent to the acceptance and publication of our article, a detailed analysis of the sequences was performed by Dr Linda Jagodzinski (Walter Reed Army Institute of Research, US Military HIV-1 Research Program, Department of Laboratory Diagnostics and Monitoring, Silver Spring, Maryland, USA) to confirm HIV-1 subtype assignment. Protease/reverse transcriptase sequences were reviewed, aligned and subjected to phylogenetic tree analysis (DNASTAR Inc.; MegAlign; Madison, Wisconsin, USA) against reference sequences obtained from the HIV-1 Database (www.hiv.lanl.gov). In addition, each sequence was further analyzed using the HIV BLAST search analysis tool (www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic-blast.html). Based upon this analysis, the HIV-1 subtype assignment changed for four of the 14 sequences referred to in our article. A revised Table 1 is shown below.

Table 1. Resistance mutations detected at position 138 of the reverse transcriptase matched with viral load, HIV-1 subtype, regimen and the range of time on treatment.

No.	Regimen	Subtype	Time-on-treatment	HIV RNA (copies/ml)	Mutation	Genbank accession #
1	EFV + AZT + 3TC	A1	>24 months	708 134	E138G	KF781839
2	NVP + AZT + 3TC	A1	>24 months	734 322	E138A	KF781840
3	EFV + AZT + 3TC	A1	>24 months	74 858	E138Q	KF781841
4	NVP + TDF + 3TC	A1	13–24 months	1021	E138K	KF781842
5	EFV + D4T + 3TC	A1	>24 months	109 987	E138A	KF781843
6	EFV + D4T + 3TC	A1	>24 months	5463	E138A	KF781844
7	NVP + AZT + 3TC	A1	>24 months	4242	E138Q	KF781845
8	EFV + TDF + 3TC	A1	>24 months	3647	E138A	KF781846
9	EFV + TDF + 3TC	A1	>24 months	27 784	E138G	KF781847
10	EFV + AZT + 3TC	A1	6–12 months	21 114	E138A	KF781848
11	NVP + TDF + 3TC	D	13–24 months	313 281	E138G	KF781849
12	EFV + TDF + 3TC	A1	6–12 months	1700	E138K	KF781850
13	NVP + TDF + 3TC	A2D	6–12 months	1902	E138A	KF781851
14	EFV + AZT + 3TC	A1	>24 months	4428	E138A	KJ764862

3TC, lamivudine; AZT, zidovudine; EFV, efavirenz; NVP, nevirapine; TDF, tenofovir.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

Keith W. Crawford, *US Military HIV Research Program, Global Health Programs, Walter Reed Army Institute of Research, Silver Spring, and Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, Maryland, USA.*

Correspondence to Keith W. Crawford, PhD, RPh, US Military HIV Research Program/Walter Reed Army Institute of Research, Bethesda, Maryland, USA.
E-mail: kwcrawford1@gmail.com

Received: 13 May 2014; revised: 14 May 2014; accepted: 14 May 2014.

Reference

1. Crawford KW, Njeru D, Maswai J, Omondi M, Apollo D, Kimetto J, *et al*. **Occurrence of etravirine/rilpivirine-specific resistance mutations selected by efavirenz and nevirapine in Kenyan patients with non-B HIV-1 subtypes failing antiretroviral therapy.** *AIDS* 2014; **28**:442–445 .

DOI:10.1097/QAD.0000000000000350

Annexes II

Anti-HIV activity of fucoidans from three brown seaweed species

Carbohydrate Polymers 115 (2015) 122–128

Contents lists available at ScienceDirect

Carbohydrate Polymers

journal homepage: www.elsevier.com/locate/carbpol

Anti-HIV activity of fucoidans from three brown seaweed species

Thanh Thi Thu Thuy[a,*], Bui Minh Ly[b], Tran Thi Thanh Van[b], Ngo Van Quang[a], Ho Cam Tu[c], Yue Zheng[c], Carole Seguin-Devaux[c], Bilan Mi[d], Usov Ai[d]

[a] Institute of Chemistry, Vietnam Academy of Science and Technology, Viet Nam
[b] Nha Trang Institute of Technology Research and Application, Vietnam Academy of Science and Technology, Viet Nam
[c] Laboratory of Retrovirology, CRP-Santé, Luxembourg
[d] N.D.Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninskii Prospect 47, Moscow 119991, Russian Federation

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form 13 August 2014
Accepted 17 August 2014
Available online 2 September 2014

Keywords:
Fucoidan
Brown seaweed
Sargassum mcclurei
Sargassum polycystum
Turbinara ornata
Anti-HIV activity

ABSTRACT

Fucoidans are sulfated polysaccharides derived from marine brown algae. In the current work the anti-HIV activity of three fucoidans, extracted from three brown seaweeds *Sargassum mcclurei*, *Sargassum polycystum* and *Turbinara ornata* and collected from Nha Trang bay, Vietnam was investigated. Fucoidans extracted from the three species displayed similar antiviral activities with a mean IC_{50} ranging from 0.33 to 0.7 μg/ml while displaying no cell toxicity. Our results showed that the anti-HIV activity of fucoidans is not primarily linked to the sulfate content and the appropriate position of sulfate groups in the fucoidan backbones was also not associated with the antiviral activity. Fucoidans inhibited HIV-1 infection when they were pre-incubated with the virus but not with the cells, and not after infection, blocking the early steps of HIV entry into target cells. These data contribute to a better understanding of the influence of fucoidans structural characteristics on their biological activity.

1. Introduction

Fucoidans are sulfated polysaccharides derived from marine brown algae. A common structural feature of fucoidans is that they contain a large amount of α-L-fucose residues and sulfate groups, together with a minor amount of other monosaccharide residues including galactose, mannose, xylose, rhamnose, and glucuronic acid (Li, Lu, Wei, & Zhao, 2008; Usov & Bilan, 2009). Fucoidans were reported to exhibit a wide range of physiological and biological activities (Berteau & Mulloy, 2003; Kusaykin et al., 2008; Wijesinghe & Jeon, 2012), such as anti-inflammatory, anticoagulant, antithrombotic (Cumashi et al., 2007; Ustyuzhanina et al., 2013), antiviral including anti-HIV (Ghosh et al., 2009; Lee, Hayashi, Hashimoto, Nakano, & Hayashi, 2004; Trinchero et al., 2009), immunomodulatory (Raghavendran, Srinivasan, & Rekha, 2011), antioxidant (Wang, Zhang, Zhang, Song, & Li, 2010), and antitumor (Senthilkumar, Manivasagan, Venkatesan, & Kim, 2013; Synytsya et al., 2010). A first report on the HIV inhibitory property of sulfated polysaccharides from sea algae was published in 1987 (Nakashima et al., 1987). Queiroz et al. (2008) further reported that sulfated fucans from the seaweed species *Dictyota mertensii*, *Lobophora variegata*, *Spatoglossum schroederi* and *Fucus vesiculosus* inhibit HIV reverse transcriptase.

Despite an extensive investigation of the biological activity of fucoidans from marine seaweeds during the last decade, the overall relationship between the activity of polysaccharides and their structure remains unclear (Berteau & Mulloy, 2003). It is usually believed that biological activity of fucoidans is related to the content and position of sulfate groups, monosaccharide composition, molecular weight, and structure of the backbone and branches (Ale, Mikkelsen, & Meyer, 2011; Patankar, Oehninger, Barnett, Williams, & Clark, 1993; Pomin, 2009).

The current challenges of the current HIV therapy prompt the worldwide researchers to search for other natural and affordable anti-HIV compounds. Based on preliminary promising results, the structure–activity relationship of the SPs structural characters on HIV-1 inhibition need to be elucidated for a deeper understanding of their binding structure and their antiviral mechanisms.

The present paper is devoted to the investigation of the anti-HIV activity of fucoidans extracted from three brown seaweed species *Sargassum polycystum*, *Sargassum mcclurei* and *Turbinara ornata* collected at Nha Trang bay, Vietnam. Native polysaccharides and several polysaccharide fractions differing in sulfate content were tested to unravel the role of sulfate groups on the anti HIV-activity of fucoidans. Similarly to other fucoidans present

* Corresponding author at: 18 Hoang Quoc Viet, Cau giay, Hanoi, Viet Nam. Tel.: +84 4 37563788; fax: +84 4 38361283.
E-mail address: thuyttt@ich.vast.ac.vn (T.T.T. Thuy).

http://dx.doi.org/10.1016/j.carbpol.2014.08.068
0144-8617/

in algae belonging to the family Sargassaceae, these polysaccharides are mainly constituted by the highly sulfated galactofucans. Structural characteristics and fractionation of a fucoidan from *S. mcclurei* were described recently (Pham et al., 2013). The polysaccharide was shown to have complex monosaccharide composition. Its highly sulfated main fraction is a galactofucan containing a backbone built up of 3-linked α-L-fucose 2,4-disulfate residues. Structurally related, but less sulfated galactofucan was obtained from *T. ornata* (Thanh, Tran, Yuguchi, Bui, & Nguen, 2013). Fractionation of fucoidan from *S. polycystum* and structural analysis of its highly sulfated galactofucan fraction was also described recently in the literature (Bilan et al., 2013). Additional information on the position of sulfate groups in the polysaccharide samples used in this work was obtained by partial acid hydrolysis followed by ESI–MS of the derived fucose monosulfates.

2. Experimental

2.1. Seaweed collection

The algae *S. polycystum*, *S. mcclurei* and *T. ornata* were harvested from Nha Trang bay, Vietnam in June, 2010. A voucher specimen of each species is deposited in Nha Trang Institute of Technology Research and Application. The collected seaweeds were washed with tap water in order to remove salt, epiphytes, and sand attached to the surface of the samples and then dried by air in the shade. The dried seaweeds were crushed, grounded into a powder form, passed through a 40-mesh sieve and then stored at room temperature.

2.2. Extraction of fucoidan

The extraction was followed by the method described previously (Bilan et al., 2002). 200 g of dried seaweed was treated at room temperature with a 4:2:1 MeOH–$CHCl_3$–water mixture to remove colored matter, filtered and vacuum dried to get defatted algal biomass. This material was extracted with 2% aqueous $CaCl_2$ solution under mechanical stirring at 85 °C for 8 h. An aqueous hexadecyltrimethylammonium bromide solution (10%) was added to the extract. The precipitate formed was centrifuged, washed with water, stirred with 20% ethanolic NaI solution for 2–3 days at room temperature, washed with ethanol, and dissolved in water. The solution was dialyzed, concentrated and lyophilized to give fucoidan as sodium salt. Fucoidans obtained from *S. mcclurei*, *S. polycystum* and *T. ornata* were named F_{SM}, F_{SP} and F_{TO}, respectively. The yields of fucoidans were calculated based on dried seaweed weight.

2.3. Chemical analyses

Neutral monosaccharide compositions were elucidated by GLC (Bilan et al., 2002). Alditol acetates were prepared by hydrolysis of fucoidan samples in 2 M CF_3COOH (TFA), 8 h at 100 °C, followed by sodium borohydride reduction and acetylation, and analyzed by 17AAFW Shimadzu GC-FID.

Sulfate content was estimated using gelatin/$BaCl_2$ method (Dodgson, 1961) after hydrolysis of fucoidan in 2 M TFA as described above.

2.4. Partial acid hydrolysis

Mild acid hydrolysis of fucoidans was carried out using trifluoroacetic acid (0.75 M, 1 h, 60 °C), the resulting solutions were evaporated several times with methanol.

2.5. ESI–MS spectra

ESI–MS experiments were performed on a Xevo TQ MS, Waters-USA. The analyses were carried out in negative mode. Dried partial acid hydrolysis products of fucoidans were dissolved in 1:1 MeOH–water and introduced into the mass spectrometer.

2.6. Anti-HIV activity assays

2.6.1. Cytotoxicity and viral infection assays

Protection of fucoidans on U373-CD4-CXCR4 cells infected by luciferase-tagged viral X4 pseudotype particles was evaluated. Pseudotype particles were produced by co-transfecting a luciferase-tagged pNL4-3-derived vector deleted of the env ectodomain, pNL4-3Δenv (Baatz et al., 2011), and a pCMV-derived HXB2 full env plasmid using lipofectamine (Invitrogen) into HEK293T cells. Viral supernatant produced by HEK293T cells was collected 48 h after transfection. P24 antigen concentration was measured using P24 ELISA kit (Perkin Elmer). U373-CD4-CXCR4 cells were infected with 200 pg of pNL4-3Δenv-HXB2 pseudotype particles. All infections were conducted in triplicate wells. Infection was synchronized by spinoculation at 1200 × *g* for 2 h at 25 °C and incubation for 1 h at 37 °C before supernatants were replaced with fresh medium. Serial dilutions of fucoidans were added during and after spinoculation. Infection was monitored after 48 h by measuring bioluminescence activity in the cell lysates (Promega luciferase assay system) using a Tecan Genios Pro Luminometer. Percentage of protection was expressed as the percentage of inhibition of the relative light units (RLUs) of U373-CD4-CXCR4 cells infected without fucoidans (100%) and in the presence of increased concentrations of fucoidans. IC_{50} values were calculated by non-linear regression using GraphPad Prism v5.01 (GraphPad Software). Statistical analyses were performed using GraphPad Prism v 5.01. Mean IC_{50} of fucoidans were compared using *t*-tests. A p value was considered significant when lower than 0.05.

Cell viability was assessed by flow cytometry using the LIVE/DEAD Fixable Dead Cell Stain Kit (Invitrogen) in U373-CD4-CXCR4 cells incubated with serial dilutions of fucoidans for 48 h according to manufacturer's protocol. Percentage of cell viability was expressed as the percentage of viable U373-CD4-CXCR4 cells without fucoidans (100%) and in the presence of increased concentrations of fucoidans.

To decipher if the antiviral activity was due to direct interactions of the compounds with either the virus, the cell membrane surface or the viral life cycle, pre-incubation of crude extracts with U373-CD4-CXCR4 cells or with the pNL4-3Δenv-HXB2 pseudotype particles as well as post-incubation of crude extracts with infected U373-CD4-CXCR4 cells were tested.

2.6.2. Cell lines

HEK293T (ATCC) cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS, L-glutamine (2 mM), penicillin (50 U/mL) and streptomycin (50 μg/mL) (all from Gibco). U373-CD4-CXCR4 cells (AIDS Research and Reference Reagent Program, NIAID) were maintained in DMEM supplemented with 15% FBS, L-glutamine (2 mM), penicillin (100 U/mL), streptomycin (100 μg/mL), geneticin (300 ng/mL) (Gibco) and puromycin (1 ng/mL) (Sigma–Aldrich).

3. Results and discussion

Three fucoidans were extracted from three brown seaweeds *S. mcclurei*, *S. polycystum* and *T. ornate*. The yields, sulfate content and sugar composition of three crude fucoidans are given in Table 1. Like other marine fucoidans, these polysaccharides contain several neutral monosaccharide components, namely, fucose, mannose,

Table 1
Yield and chemical analysis of crude fucoidans.

Crude fucoidan	Yield (% dried seaweed)	Neutral sugar composition (mol %)					Sulfate content (mass %)
		Fuc	Man	Gal	Xyl	Glc	
F_{SP}	2.75	20.30	1.90	13.70	2.60	1.10	23.40
F_{TO}	2.50	30.30	Trace	25.60	Trace	Trace	25.60
F_{SM}	2.10	40.00	10.70	20.10	6.20	10.04	30.50

galactose, xylose and glucose with different molar ratios. F_{SP} and F_{TO} were mainly composed of fucose, a high proportion of galactose and very small amount of glucose, xylose and mannose. The highest fucose content was detected in F_{SM}.

Fucoidans were subjected to mild hydrolysis as described in experimental part to obtain fragments suitable for mass spectrometric investigation. Mass spectra may be used for determination of position of sulfate groups in fucose monosulfates (Tissot, Salpin, Martinez, Gaigeot, & Daniel, 2006) according to specific fragmentation patterns of molecular anions $[FucSO_3]^-$. Signal at *m*/*z* 183 indicates sulfate group at C-4, signal at *m*/*z* 139 corresponds to sulfate group at C-2, and signal at *m*/*z* 169 was assigned to sulfate group at C-3 of fucose monosulfates.

Mass spectra of partial hydrolysis products of all three fucoidans (data not shown) contain a signal at *m*/*z* 243 corresponding to the deprotonated molecule of monosulfated fucose $[FucSO_3]^-$. The MS^2 spectra of this daughter ion at *m*/*z* 243 of three fucoidans F_{SM}, F_{SP} and F_{TO} were shown in Figs. 1–3, respectively.

In Fig. 1, ESI–MS/MS of monosulfated fucose $[FucSO_3]^-$ at *m*/*z* 243 was similar to that observed for *Fucus evanescens* reported by Anastyuk, Shevchenko, Nazarenko, Dmitrenok, and Zvyagintseva (2009), where signals from fucose 4-sulfate (*m*/*z* 183) and fucose 2-sulfate (*m*/*z* 139) have approximately equal intensity. No evidence on sulfation at C-3 was obtained, since the corresponding fragment ion at *m*/*z* 169 was not observed. Hence, the partial hydrolyzate of F_{SM} contained approximately equal amounts of 2- and 4-sulfated fucose. In contrast, Fig. 2 shows that the ion at *m*/*z* 183 was the major fragment, and the high abundance of this fragment indicates that the fucosyl units of F_{SP} are mainly sulfated at position 4. In Fig. 3, signals from fucose 4-sulfate (*m*/*z* 183) and fucose 2-sulfate (*m*/*z* 139) were detected, but the ion at *m*/*z* 139 was the major fragment indicating that fucosyl units of F_{TO} are mainly sulfated at position 2. Similar result was reported for a fucoidan from *Laminaria cichorioides* (Anastyuk et al., 2010). Thus, according to mass-spectrometric evidence, all three fucoidans have sulfate groups attached to C-2 and C-4 of fucose residues, but in different proportions.

Table 2
Position and content of sulfate groups and anti-HIV activity of three crude fucoidans.

Crude fucoidan	Sulfate group position	Sulfate content (mass %)	Mean IC_{50}, μg/ml (SD)[a]
F_{SP}	C2 < C4	23.40	0.34 ± 0.12
F_{TO}	C2 > C4	25.60	0.39 ± 0.18
F_{SM}	C2 = C4	30.50	0.96 ± 0.59

[a] Mean IC_{50} was calculated from three independent experiments. SD, standard deviation.

These data on the position of sulfate groups in three crude fucoidans were summarized in Table 2 together with their sulfate content and their anti-HIV activity (mean of IC_{50}).

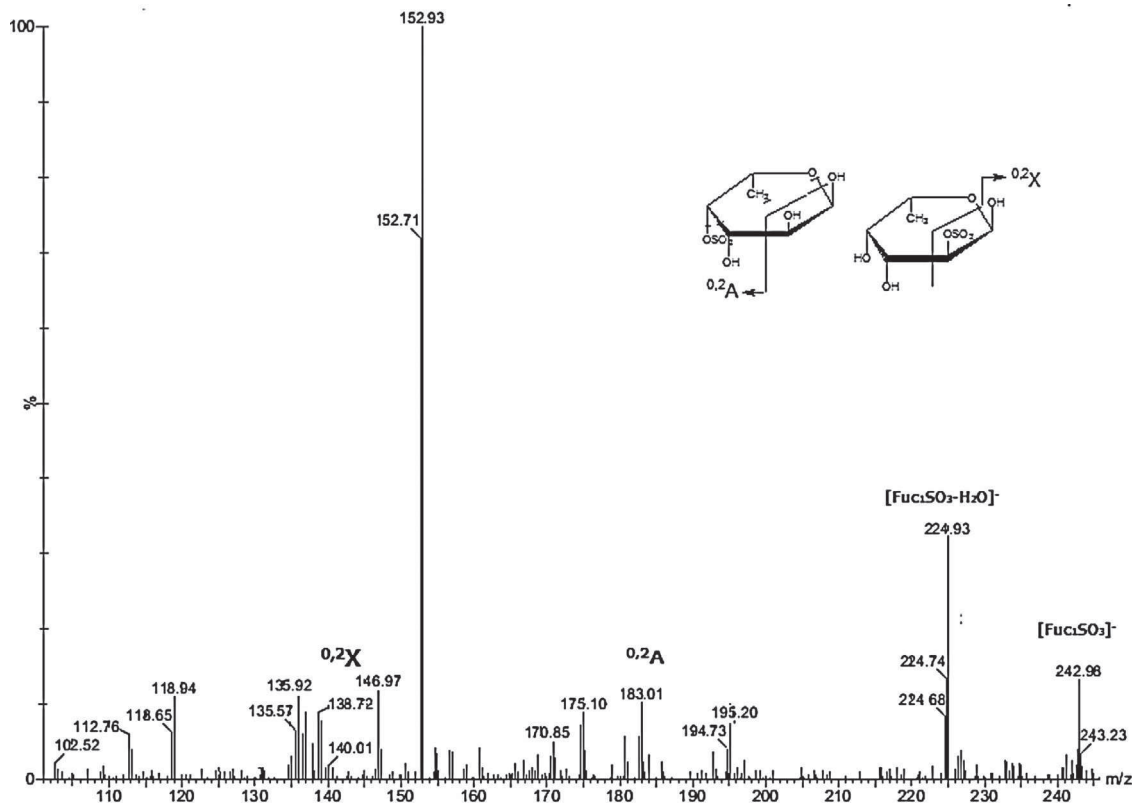


Fig. 1. Negative ESI–MS/MS spectra of the monosulfated fucose ion $[FucSO_3]^-$ at *m*/*z* 243 of F_{SM}.

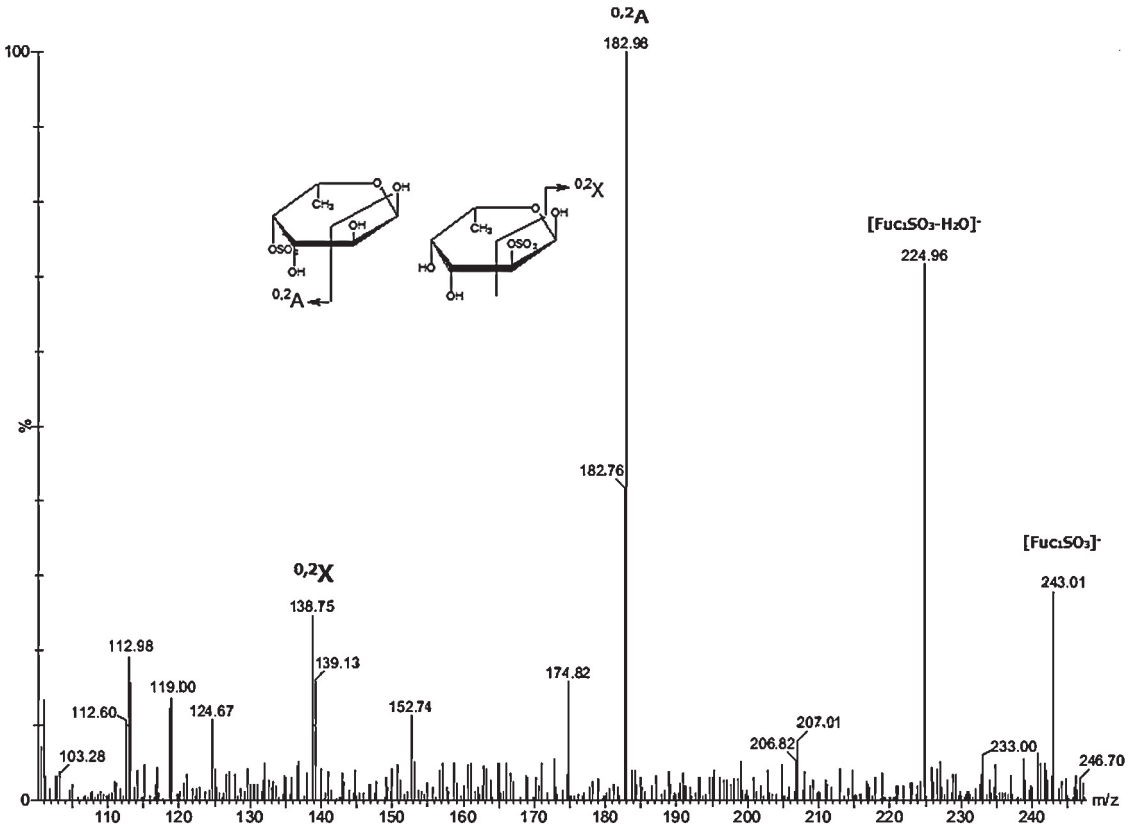


Fig. 2. Negative ESI–MS/MS spectra of the monosulfated fucose ion $[FucSO_3]^-$ at m/z 243 of F_{SP}.

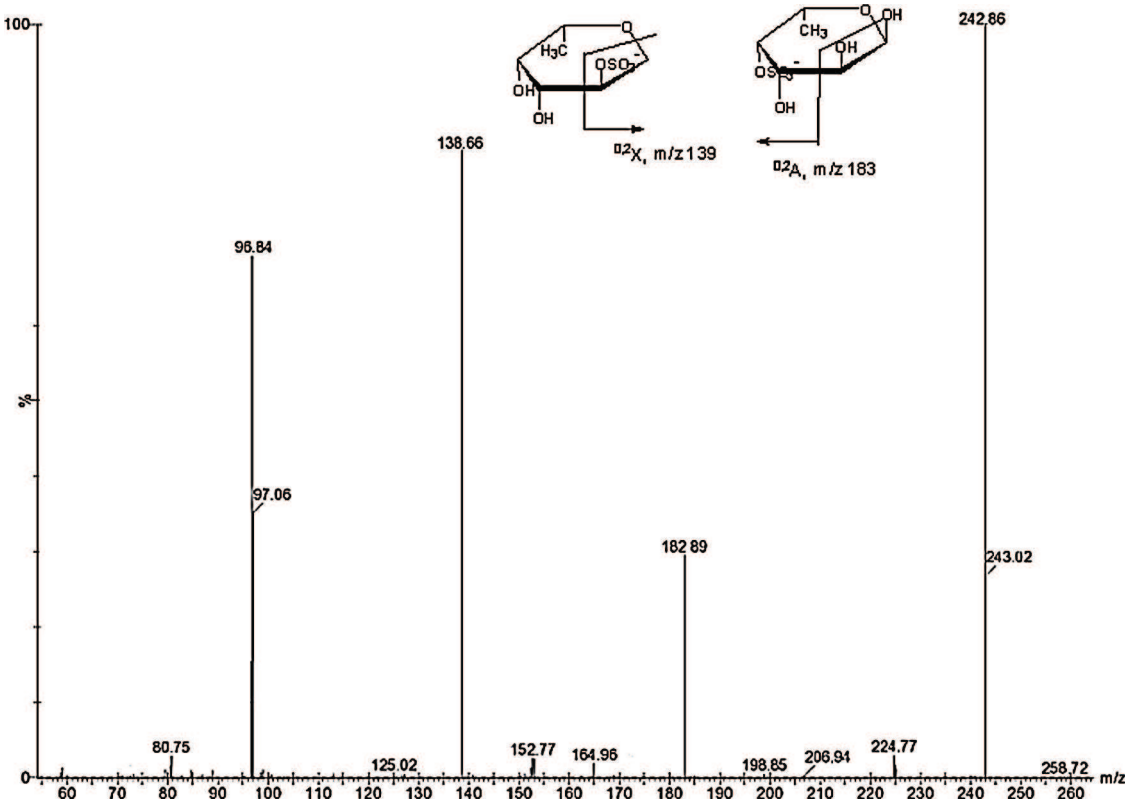


Fig. 3. Negative ESI–MS/MS spectra of the monosulfated fucose ion $[FucSO_3]^-$ at m/z 243 of F_{TO}.

Table 3
Yield and chemical analysis of fractions from F_{SM}.

Fraction (M NaCl)	Yield (% of F_{SM})	Monosaccharide composition (mol %)					SO_3Na (mass %)
		Fuc	Man	Gal	Xyl	Glc	
F_{SM1} (0.5)	8.5	30.4	21.7	20.7	9.20	18.0	17.60
F_{SM2} (1.0)	19.2	46.2	8.1	30.11	6.31	12.90	24.71
F_{SM3} (1.5)	28.9	60.5	Trace	38.40	Trace	0.70	33.03

Table 4
Yield and chemical analysis of fractions from F_{SP}.

Fraction (M NaCl)	Yield (% of F_{SP})	Monosaccharide composition, mol %					SO_3Na (mass %)
		Fuc	Man	Gal	Xyl	Glc	
F_{SP1} (0.5)	6.0	8.8	5.8	12.8	8.7	3.6	4.4
F_{SP2} (1.0)	18.0	12.7	3.3	20.3	5.8	1.5	19.3
F_{SP3} (1.5)	17.0	29.0	1.0	12.1	2.4	Trace	26.6
F_{SP4} (2.0)	17.0	34.4	–	21.1	Trace	–	27.4

The protection against HIV-1 was further estimated on U373-CD4-CXCR4 cells infected with pseudotype viral particles tagged with luciferase to measure HIV infection (Table 2; Fig. 4A) as well as cell viability in presence of fucoidans by flow cytometry (Fig. 4B). All fucoidans exhibited similar anti-HIV activity with a mean IC_{50} ranging from 0.33 to 0.7 μg/ml while displaying no cytotoxicity up to 2 μg/ml for F_{TO} and F_{SP} and 20 μg/ml for F_{SM}, respectively. The specific CXCR4 inhibitor AMD3100 showed an IC_{50} of 21.6 nM whereas the non-sulfated polysaccharide hydroxyethil starch had no antiviral activity.

Several authors reported that the sulfate content of sulfated polysaccharides is one of the most important factors for its biological effects (Haroun-Bouhedja, Ellouali, Sinquin, & Boisson-Vidal, 2000; Nishino & Nagumo, 1991). As shown in Table 2, the sulfate content was the highest in F_{SM} as compared to F_{SP} and F_{TO}, however their antiviral activity was not significantly different. In addition, the amount of 2- and 4-sulfated fucose observed in these fucoidans were all different suggesting that they do not play a role in the anti-HIV activity.

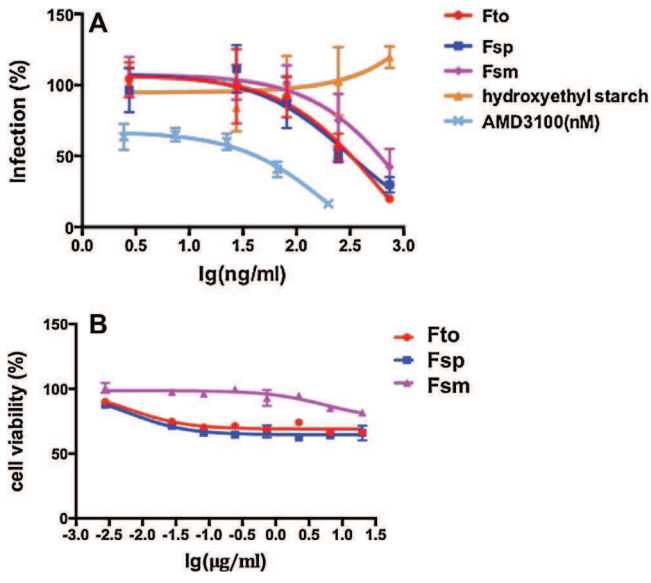


Fig. 4. (A) Infection of U373-CD4-CXCR4 cells by X4 pseudotype viral particles in presence of the fucoidans F_{SP}, F_{TO} and F_{SM}. One representative experiment of three is shown. Each concentration of the three crude fucoidans were tested in triplicates in each independent experiment. The CXCR4 inhibitor AMD3100 was used as a positive control and hydroxyethyl starch was used as a negative control as a non-sulfated polysaccharide. (B) Viability of U373-CD4-CXCR4 cells in the presence of the fucoidans F_{SP}, F_{TO} and F_{SM}. One representative experiment of two is shown. Each concentration of the three crude fucoidans were tested in duplicates in each independent experiment.

To further investigate this relationship crude F_{SM} and F_{SP} fucoidans (containing respectively the highest and lowest content of sulfate) were fractionated into three (F_{SM1}, F_{SM2}, F_{SM3}) and four fractions (F_{SP1}, F_{SP2}, F_{SP3}, F_{SP4}) by ion-exchange chromatography on DEAE-Sephadex column using aqueous sodium chloride of increasing concentration as eluant. The chemical analysis of the fractions was reported in Tables 3 and 4. All fractions were examined for their anti-HIV activity, and the content and position of sulfate groups were determined. The results were summarized in Tables 5 and 6.

Although the fraction F_{SM3} was containing the highest content of sulfate (comparable to the sulfate content of F_{SM}), they did not exhibit a significantly higher antiviral activity as compared to F_{SM1} or F_{SM2}.

The fraction F_{SP1}, having a very low sulfate content, did not show any antiviral activity. Nevertheless, the derived fractions F_{SP2}, F_{SP3} and F_{SP4} showed a similar antiviral activity comparable to F_{SP}. Despite the purification of fucoidans by ion-exchange chromatography, the antiviral activity was not improved and remains similar between all fractions. These data do not support any role of the sulfate content on the anti-HIV activity of fucoidans but rather the role of a specific fucoidan that might have been lost during the purification of F_{SP1} and shared between all other fractions. The direct dependence of biological activity on the position of sulfate groups was also not observed.

Table 5
Position and content of sulfate groups and anti-HIV activity of three fractions of F_{SM}.

Fractions	Sulfate group position	Sulfate content (mass %)	Mean IC_{50}, μg/ml (SD)[a]
F_{SM1}	Mainly on C4	17.60	0.75 (±0.6)
F_{SM2}	Mainly on C4	24.71	0.39 (±0.27)
F_{SM3}	C2 > > C4	33.03	0.18 (±0.16)

[a] Mean IC_{50} was calculated from three independent experiments. SD, standard deviation.

Table 6
Position and content of sulfate groups and anti-HIV activity of four fractions of F_{SP}.

Fractions	Sulfate group position	Sulfate content (mass %)	Mean IC_{50} (μg/ml) (SD)[a]
F_{SP1}	Mainly on C4	4.4	NA
F_{SP2}	Mainly on C4	19.3	0.37 (±0.17)
F_{SP3}	C2 > C4	26.6	0.23 (±0.4)
F_{SP4}	C2 > C4	27.4	0.15 (±0.12)

[a] Mean IC_{50} was calculated from three independent experiments. SD, standard deviation.

Finally, to investigate the antiviral mechanisms of fucoidans on HIV-1, F_{SP} were incubated 2 h before infection either with U373-CD4-CXCR4 cells or with viral pseudotype particles and 2 h post-infection. Fucoidans inhibited HIV-1 infection when they were pre-incubated with the virus (mean $IC_{50} = 0.38 \pm 0.14$ μg/ml) but not with the cells and not at the time of post-infection. These data indicate that fucoidans bind to the virus and inhibit HIV entry into target cells.

4. Conclusion

We have demonstrated that fucoidans isolated from the three species *S. mcclurei*, *S. polycystum* and *T. ornata* have significant anti-HIV activity. Although highly sulfated preparation of crude extracts was obtained by anion-exchange chromatography, they did not demonstrate any increased anti-HIV activity. Neither sulfate content nor position of sulfate groups was related to the anti-HIV activity of fucoidans suggesting the involvement of other structural parameters like molecular weight, the type of glycosidic linkage or even a unique fucoidan sequence. Although the presence of sulfo groups seems to be necessary for the anti-HIV activity (Schaeffer & Krylov, 2000), our data do not support random sulfation as the main antiviral factor. Fucoidans are structurally complex, known for their antiviral activities due to their direct interactions with either enveloped viruses or with cell membrane surface (Ghosh et al., 2009). In this study we showed that the antiviral activity of these fucoidans is due to their binding with HIV-1 blocking the early steps of HIV entry. These macromolecules might exert their anti-HIV-1 activity by shielding off the positively charged amino acids present in the viral envelope glycoprotein gp120 (Moulard et al., 2000) or by the strong binding with a specific sulfation motif (de Parseval et al., 2005). This former hypothesis is in agreement with the specific interactions of sulfated polysaccharides with proteins leading to their biological activities such as the binding of heparin with antithrombin III or chondroitin 4 sulfate with plasmodium falciparum or the lyme disease spirochete (Andrews, Klatt, Adams, Mischnick, & Schwartz-Albiez, 2005; Mulloy, 2005).

Acknowledgements

This work was financially supported by the National Foundation for Science and Technology Development (NAFOSTED) of Vietnam (code 104.01-2010.43 and 104.01-59.09) and Fondation Recherche sur le SIDA in Luxembourg.

References

Ale, M. T., Mikkelsen, J. D., & Meyer, A. S. (2011). Important determinants for fucoidan bioactivity: A critical review of structure–function relations and extraction methods for fucose-containing sulfated polysaccharides from brown seaweeds. *Marine Drugs*, *9*, 2106–2130.

Anastyuk, S. D., Shevchenko, N. M., Nazarenko, E. L., Dmitrenok, P. S., & Zvyagintseva, T. N. (2009). Structural analysis of a fucoidan from the brown alga *Fucus evanescens* by MALDI-TOF and tandem ESI mass spectrometry. *Carbohydrate Research*, *344*, 779–787.

Anastyuk, S. D., Shevchenko, N. M., Nazarenko, E. L., Imbs, T. I., Gorbach, V. I., Dmitrenok, P. S., et al. (2010). Structural analysis of a highly sulfated fucan from the brown alga *Laminaria cichorioides* by tandem MALDI and ESI mass spectrometry. *Carbohydrate Research*, *345*, 2206–2212.

Andrews, K. T., Klatt, N., Adams, Y., Mischnick, P., & Schwartz-Albiez, R. (2005). Inhibition of chondroitin-4-sulfate-specific adhesion of *Plasmodium falciparum*-infected erythrocytes by sulfated polysaccharides. *Infection and Immunity*, *73*(7), 4288–4294.

Baatz, F., Nijhuis, M., Lemaire, M., Riedijk, M., Wensing, A. M., Servais, J. Y., et al. (2011). Impact of the HIV-1 env genetic context outside HR1–HR2 on resistance to the fusion inhibitor enfuvirtide and viral infectivity in clinical isolates. *PLoS ONE*, *6*(7), 21535. http://dx.doi.org/10.1371/journal.pone.0021535

Berteau, O., & Mulloy, B. (2003). Sulfated fucans, fresh perspectives: Structures, functions, and biological properties of sulfated fucans and an overview of enzymes active toward this class of polysaccharide. *Glycobiology*, *13*, 29R–40R.

Bilan, M. I., Grachev, A. A., Shashkov, A. S., Thanh, T. T. T., Tran, T. T. V., Bui, M. L., et al. (2013). Preliminary investigation of a highly sulphated galactofucan fraction isolated from the brown alga *Sargassum polycystum*. *Carbohydrate Research*, *377*, 48–57.

Bilan, M. I., Grachev, A. A., Ustuzhanina, N. E., Shashkov, A. S., Nifantiev, N. E., & Usov, A. I. (2002). Structure of a fucoidan from the brown seaweed *Fucus evanescens* C.Ag. *Carbohydrate Research*, *337*, 719–730.

Cumashi, A., Ushakova, N. A., Preobrazhenskaya, M. E., D'Incecco, A., Piccoli, A., Totani, L., et al. (2007). A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic, and antiadhesive activities of nine different fucoidans from brown seaweeds. *Glycobiology*, *17*, 541–552.

de Parseval, A., Bobardt, M. D., Chatterji, A., Chatterji, U., Elder, J. H., David, G., et al. (2005). A highly conserved arginine in gp120 governs HIV-1 binding to both syndecans and CCR5 via sulfated motifs. *The Journal of Biological Chemistry*, *280*(47), 39493–39504.

Dodgson, K. S. (1961). Determination of inorganic sulphate in studies on the enzymic and non-enzymic hydrolysis of carbohydrate and other sulphate esters. *Biochemical Journal*, *78*, 312–319.

Ghosh, T., Chattopadhyay, K., Marschall, M., Karmakar, P., Mandal, P., & Ray, B. (2009). Focus on antivirally active sulphated polysaccharides: From structure-activity analysis to clinical evaluation. *Glycobiology*, *19*, 2–15.

Haroun-Bouhedja, F., Ellouali, M., Sinquin, C., & Boisson-Vidal, C. (2000). Relationship between sulfate groups and biological activities of fucans. *Thrombosis Research*, *100*, 453–459.

Kusaykin, M., Bakunina, I., Sova, V., Ermakova, S., Kuznetsova, T., Besednova, N., et al. (2008). Structure, biological activity, and enzymatic transformation of fucoidans from the brown seaweeds. *Biotechnology Journal*, *3*, 904–915.

Lee, J. B., Hayashi, K., Hashimoto, M., Nakano, T., & Hayashi, T. (2004). Novel antiviral fucoidan from sporophyll of *Undaria pinnatifida* (Mekabu). *Chemical and Pharmaceutical Bulletin*, *52*, 1091–1094.

Li, B., Lu, F., Wei, X., & Zhao, R. (2008). Fucoidan: Structure and bioactivity. *Molecules*, *13*, 1671–1695.

Moulard, M., Lortat-Jacob, H., Mondor, I., Roca, G., Wyatt, R., Sodroski, J., et al. (2000). Selective interactions of polyanions with basic surfaces on human immunodeficiency virus type 1 gp120. *Journal of Virology*, *74*, 1948–1960.

Mulloy, B. (2005). The specificity of interactions between proteins and sulfated polysaccharides. *Anais da Academia Brasileira de Ciencias*, *77*(4), 651–664.

Nakashima, H., Kido, Y., Kobayashi, N., Motoki, Y., Neushul, M., & Yamamoto, N. (1987). Purification and characterization of an avian myeloblastosis and human immunodeficiency virus reverse transcriptase inhibitor, sulfated polysaccharides extracted from sea algae. *Antimicrobial Agents and Chemotherapy*, *31*(10), 1524–1528.

Nishino, T., & Nagumo, T. (1991). The sulfate-content dependence of the anticoagulant activity of a fucan sulfate from the brown seaweed *Ecklonia kurome*. *Carbohydrate Research*, *214*, 193–197.

Patankar, M. S., Oehninger, S., Barnett, T., Williams, R. L., & Clark, G. F. (1993). A revised structure for fucoidan may explain some of its biological activities. *Journal of Biological Chemistry*, *268*, 21770–21776.

Pham, D. T., Menshova, R. V., Ermakova, S. P., Anastyuk, S. D., Bui, M. L., & Zvyagintseva, T. N. (2013). Structural characteristics and anticancer activity of fucoidan from the brown alga *Sargassum mcclurei*. *Marine Drugs*, *11*, 1456–1476.

Pomin, V. H. (2009). Review: An overview about the structure–function relationship of marine sulfated homopolysaccharides with regular chemical structures. *Biopolymers*, *91*, 601–609.

Queiroz, K. C. S., Medeiros, V. P., Queiroz, L. S., Abreu, L. R. D., Rocha, H. A. O., Ferreira, C. V., et al. (2008). Inhibition of reverse transcriptase activity of HIV by polysaccharides of brown algae. *Biomedicine and Pharmacotherapy*, *62*, 303–309.

Raghavendran, H. R., Srinivasan, P., & Rekha, S. (2011). Immunomodulatory activity of fucoidan against aspirin-induced gastric mucosal damage in rats. *International Immunopharmacology*, *11*, 157–163.

Schaeffer, D. J., & Krylov, V. S. (2000). Anti-HIV activity of extracts and compounds from algae and cyanobacteria. *Ecotoxicology and Environment Safety*, *45*(3), 208–227.

Senthilkumar, K., Manivasagan, P., Venkatesan, J., & Kim, S.-K. (2013). Brown seaweed fucoidan: Biological activity and apoptosis, growth signaling mechanism in cancer. *International Journal of Biological Macromolecules*, *60*, 366–374.

Synytsya, A., Kim, W. J., Kim, S. M., Pohl, R., Synytsya, A., Kvasnicka, F., et al. (2010). Structure and antitumour activity of fucoidan isolated from sporophyll of Korean brown seaweed *Undaria pinnatifida*. *Carbohydrate Polymers*, *81*, 41–48.

Thanh, T. T. T., Tran, V. T. T., Yuguchi, Y., Bui, L. M., & Nguen, T. T. (2013). Structure of fucoidan from brown seaweed *Turbinaria ornata* as studied by electrospray ionization mass spectrometry (ESIMS) and small angle X-ray scattering (SAXS) techniques. *Marine Drugs*, *11*, 2431–2443.

Tissot, B. S., Salpin, J.-Y., Martinez, M., Gaigeot, M.-P., & Daniel, R. (2006). Differentiation of the fucoidan sulfated L-fucose isomer constituents by CE-ESIMS and molecular modeling. *Carbohydrate Research*, *341*, 598–609.

Trinchero, J., Ponce, N. M. A., Cordoba, O. L., Flores, M. L., Pampuro, S., Stortz, C. A., et al. (2009). Antiretroviral activity of fucoidans extracted from the brown seaweed *Adenocystis utricularis*. *Phytotherapy Research*, *23*, 707–712.

Usov, A. I., & Bilan, M. I. (2009). Fucoidans – Sulfated polysaccharides of brown algae. *Russian Chemical Reviews*, *78*, 785–799.

Ustyuzhanina, N. E., Ushakova, N. A., Zyuzina, K. A., Bilan, M. I., Elizarova, A. L., Somonova, O. V., et al. (2013). Influence of fucoidans on hemostatic system. *Marine Drugs*, *11*, 2444–2458.

Wang, J., Zhang, Q., Zhang, Z., Song, H., & Li, P. (2010). Potential antioxidant and anticoagulant capacity of low molecular weight fucoidan fractions extracted from *Laminaria japonica*. *International Journal of Biological Macromolecules*, *46*, 6–12.

Wijesinghe, W. A. J. P., & Jeon, Y.-J. (2012). Biological activities and potential industrial applications of fucose rich sulfated polysaccharides and fucoidans isolated from brown seaweeds: A review. *Carbohydrate Polymers*, *88*, 13–20.

Résumé de thèse

Mon travail de thèse porte sur l'identification d'un arbre médicinal africain par le DNA barcoding et l'étude de composés à activité anti-VIH de cet arbre. Une première analyse de la séquence du marqueur ITS2 déterminée à partir d'ADN extrait de copeaux de bois a suggéré que la plante pourrait appartenir au genre Cassia ou au genre apparenté Senna. En analysant la séquence de ce marqueur ITS2 et aussi celle du trnH-psbA spacer d'une cinquantaine d'espèces des genres Cassia et Senna j'ai pu identifier la plante comme étant *Cassia abbreviata*. L'alignement de ces séquences m'a permis d'identifier, pour les deux marqueurs, des structures particulières spécifiques aux espèces du genre Cassia, permettant donc de les différencier des espèces du genre Senna. J'ai utilisé ces alignements pour effectuer une étude phylogenetique qui illustre que, pour les deux marqueurs, les Cassia forment en effet un clade bien séparé du clade des Senna qui peut être divisé en plusieurs sous-clades. Dans un deuxième temps j'ai étudié les effets anti-VIH de l'extrait alcoolique ainsi que de 57 composés purifiés obtenus au laboratoire. L'extrait brut ainsi qu'un des composés purifiés, le piceatannol, ont montré un grand spectre d'activités antivirales pour le VIH et le virus de l'herpès. Ils inhibent, à un stade précoce, l'infection par le VIH de lignées cellulaires de référence et d'isolats cliniques, ceci indépendamment de l'utilisation du co-récepteur (IC50: 10.47-40.77 µg/ml, CC50>1000 µg/ml; IC50: 8.04-47.46 µM, CC50>300 µM, respectivement). Ni l'un ni l'autre n'a d'effet sur CD4 et CCR5/CXCR4. L'extrait brut, mais pas le piceatannol, bloque l'interaction CD4-gp120, suggérant que l'extrait brut cible gp120 alors que le piceatannol agit comme un inhibiteur non-spécifique d'attachement du virus. Aussi, dans un modèle *in vitro* de tract génital femelle, le piceatannol inhibe l'infection de cellules cibles TZM-Bl par le VIH et n'active pas les cellules PBMCs, suggérant qu'il pourrait être un bon candidat comme microbicide. D'autres composés à activité anti-VIH dans l'extrait comportent l'acide oleanolique, l'acide palmitique, la taxifoline, ainsi que trois composés dont la structure est en train d'être élucidée.

Mots clés : *Cassia abbreviata*, Senna, identification d'espèces, DNA barcoding, analyse phylogénétique, antiviraux, VIH, composés, piceatannol, microbicide, entrée virale, attachement viral, gp120

Summary

My thesis project deals with the identification, by DNA barcoding, of an African medicinal plant and the study of anti-HIV compounds from this plant. A first analysis of the ITS2 marker sequence determined from DNA extracted from the wood suggested that the plant could belong to the Cassia or the related Senna genus. A further analysis of ITS2 as well as of trnH-psbA spacer sequences from about 50 species of the two genera allowed me to identify the plant as *Cassia abbreviata*. The sequence alignments, which reveal unique features present in the Cassia but not the Senna sequences, were used to construct phylogenetic trees showing the clear separation of the species of the Cassia and the Senna genus. The two markers therefore allow a quick discrimination between the species of the Cassia and the Senna genus and appear to be excellent barcode markers for identification of these species. Following the identification of the plant I have tested the crude ethanol extract as well as 57 purified compounds from the plant for an anti-HIV activity. The extract, as well as one of the compounds, namely piceatannol, showed a broad spectrum of antiviral activities for HIV and HSV. They inhibited HIV-1 infection at the early stage against various reference strains and resistant clinical isolates independent of the co-receptor usage (IC50: 10.47-40.77 µg/ml, CC50>1000 µg/ml; IC50: 8.04-47.46 µM, CC50>300 µM, respectively). Neither the crude extract nor piceatannol had an effect on CD4 and CCR5/CXCR4. The crude extract blocked CD4-gp120 interaction while piceatannol did not, indicating that CE may target gp120 and piceatannol may act as a non-specific viral attachment inhibitor. Moreover, piceatannol inhibited HIV infection of TZM-Bl target cells in an *in vitro* female genital tract model and did not activate PBMCs, suggesting that it may represent a good candidate as microbicide. Other anti-HIV compounds found in *Cassia abbreviata* include oleanolic acid, palmitic acid, taxifolin and three other compounds the structure of which is presently being elucidated.

Keywords: *Cassia abbreviata,* Senna, species identification, DNA barcoding, phylogenetic analysis, antiviral, HIV, compounds, piceatannol, microbicide, viral entry, viral attachment, gp120