LIÈGE université
Gembloux
Agro-Bio Tech

An integrated approach to survey and manage the cassava brown streak disease (CBSD) epidemics and investigating the diversity of viruses infecting cassava (*Manihot esculenta*)

KWIBUKA BISIMWA Yves

UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

An integrated approach to survey and manage the cassava brown streak disease (CBSD) epidemics and investigating the diversity of viruses infecting cassava (*Manihot esculenta*)

KWIBUKA BISIMWA Yves

Original dissertation presented in the fulfilment of the requirements for the Philosophical Degree (PhD) *es* Agronomical sciences and Bio-engineering

Promotor :Prof. Sébastien MASSART

Co-Promotors :Prof Hervé VANDERSCHUREN
Prof. Espoir BISIMWA BASENGERE

Academic Year: 2022- 2023

© Kwibuka Yves – 05 Mai 2023

Abstract

Several milestones have been achieved since the beginning of research on the etiology and epidemiology of the two historically significant viral diseases of cassava in Africa: the cassava mosaic disease (CMD) and the cassava brown streak disease (CBSD). They were important because they enabled the design of diagnostic and control strategies/measures to mitigate both diseases. Unfortunately, the success of control measures reached a different level because the CMD is no longer critical to cassava cultivation in Africa while the CBSD is still advancing into new areas, putting the livelihoods of millions of people at risk of food insecurity.

As cuttings are essential for the spread of CBSD, establishing a clean and sustainable cassava seed system is essential to mitigate the spread and manage the threats of this disease. Throughout history, evidence has shown how major social, economic and cultural changes have shaped the epidemiology of various seed-borne plant diseases, raising concerns about the critical role that seed pathways play in driving their epidemiology.

Strengthening cassava seed systems against CBSD requires designing context-specific intervention strategies and accurate disease diagnostic procedures, among other possibilities. In the past, unfortunately, approaches to capture regions and community specificities while assessing the epidemiology of the CBSD was poorly documented and used. In addition, extensive survey methods providing only snapshots of the real epidemic situation at local levels have been privileged in assessing CBSD epidemiology and collecting samples necessary for accurately detecting the disease.

This thesis has been initiated to contribute to the establishment of a clean and improved cassava seed system in the D.R. Congo by prioritizing improved decision-making in the design of intervention strategies to mitigate CBSD-related threats, and by contributing to the knowledge of the viruses infecting cassava crops in the D.R. Congo.

A multidisciplinary assessment approach, integrating the traditional components of plant disease epidemiology as well as socio-cultural, geographical and agroecological components, has been designed and tested in the South-Kivu province (Uvira, D.R. Congo) for assessing CBSD epidemiology as well as to elucidate the role played by cuttings pathways in the dissemination of the CBSD. Cassava leaves and cuttings collected from farmer's fields were RT-PCR indexed to detect the presence of the viruses responsible for the CBSD. The High throughput sequencing technology using the Illumina platform and various library preparation strategies (total RNA, VANA and dsRNA) has been used as a powerful method to investigate the diversity of viruses infecting cassava crop, using samples collected following an intensive sampling approach. Bioinformatic analyses following the metagenomic approach have been used to analyse the reads produced from the HTS. The *de novo* assembly method using the rnaviralSPAdes algorithm has been used to reconstruct contigs. The homology of nucleotide and amino-acid sequences and the presence of protein motifs are used to

identify contigs of viral origin. Additional analyses (identification of genes and cleavage sites) were conducted to characterize the viruses detected. Reconstructed viral genome sequences were confirmed using RT-PCR and Sanger sequencing, and their corresponding 5' and 3' extremities were reconstructed using the Rapid Amplification cDNA Ends (RACE) system. Multiple sequence alignments and phylogeny inferences were undertaken using various models predicted for amino acid and nucleotide sequences.

The main results are summarized as follows:

Factorial analysis of epidemic parameters revealed that the spread of CBSD in the studied area is spatially heterogeneous and characterized by three zones/clusters, one of which, located at mild altitude, showed the lowest disease pressure, making it putatively suitable for cutting multiplication activities. Nine parameters related to the epidemiology of CBSD were highlighted as characteristic in describing this spatial trend, notably the type of infected virus, the type of foliar symptoms, the foliar symptom incidence, the plant age, the source of cuttings, the farming system, the type of land tenure, the number of whiteflies, as well as the presence of weeds in the cassava field. These results suggest that strategies for mitigating the spread of CBSD in the studied area should be multi-layer, considering the specificities of each described cluster.

The molecular detection of CBSD-causing viruses in the study area using the RT-PCR revealed three important findings: (i) the overall incidence rates of the CBSD-causing viruses was higher at the time of the study (31.3%, in 2019) compared to the rates reported earlier, (ii) among infected samples, the prevalence of single infection by the UCBSV was the highest (42.9%), followed by the CBSV infection (35.1%) while mixed infection showed the lowest prevalence (22.1%), (iii) single infections by the CBSV and UCBSV were most prevalent in the clusters 1 and 3 respectively, whereas the mixed infection was most prevalent within the cluster 2.

The prediction analysis revealed the multidisciplinary factors putatively influencing the spread of the CBSD into cuttings channels/pathways, including the factors related to (i) the farmer's knowledge and awareness, (ii) the geographical location of the farmer's fields, as well as (iii) the type of the pathway/channel used by the farmers to obtain cuttings.

The HTS allowed for broadening the diversity of viruses infecting the cassava crop in the studied area. Several isolates belonging to two putative viral species from the genus ampelovirus, the Manihot esculenta-associated virus 1 and 2 (MEaV-1 and 2), and isolates putatively belonging to a species from the cheravirus genus, the cassava Congo cheravirus, were identified and molecularly characterised for the first time.

A collaborative initiative between researchers allowed to extend the area of spread of these viruses. Indeed, ampelovirus species have also been detected in cassava samples from Rwanda, Madagascar, Reunion and Mayotte, whereas the cheravirus have been detected in samples from the Kongo-Central (former Bas-Congo) in the West of the D.R. Congo.

Pending validation, these viruses may also spread in Tanzania since mining the publicly available HTS data using the associated RdRp fingerprint recovered data from closely related genome sequences assembled.

The characterisation of the cassava Congo cheravirus genome brought a novelty in that it allowed to disentangle the existence of a previously unreported cleavage site delineating two genes (encoding the X1 and X2 proteins) upstream the NTB, a feature only restricted to nepovirus and sadwavirus until then but putatively present for cheravirus as well. Additionally, a gene encoding for a putative ITPase homolog (HAM1) has been identified in the c-terminal part of the polymerase-encoding gene. Only two out of three essential amino acid (aa) residues were conserved for the protein encoded by this gene, compared to the functionally characterized human ITPase, whereas all intermediate and dispensable aa were not conserved at all.

Results from this thesis opens interesting perspectives towards the improvement of approaches used to study the epidemiology of the CBSD to make it capable of providing useful results. These are necessary for supporting effective decision-making in the response against the spread of the CBSD through seed systems. Furthermore, it opens perspectives towards the characterization of the biological features for the newly identified viruses. There is an opportunity to adapt the approach herein described for the characterization of other important plant disease's epidemiology in Africa and, importantly, to investigate virus diversity in other ecosystems.

Key Words: CBSD, Cassava, Epidemiology, virus diversity, seed systems, D.R. Congo, ampelovirus, cheravirus

Résumé

Des réalisations importantes ont été atteintes depuis le début des recherches sur l'étiologie et l'épidémiologie des deux maladies virales historiquement les plus importantes du manioc en Afrique : la mosaïque Africain de manioc (MAM) et la striure brune du manioc (SBM). Ces réalisations ont été importantes car elles ont permis de mettre au point des stratégies/méthodes de diagnostic et de contrôle nécessaires pour contrôler la dissémination de ces maladies et atténuer les dégâts occasionnés. Ces mesures de contrôle n'ont pas atteint le même niveau de réussite puisque, actuellement, la MAM ne constitue plus une menace critique pour la culture du manioc en Afrique tandis que la SBM continue toujours de progresser dans des nouvelles régions, menaçant la sécurité alimentaire des millions des personnes.

La mise en place d'un système semencier de manioc sain et résiliant est essentiel pour atténuer la propagation de cette maladie et en gérer les dégâts étant donné que les boutures utilisées pour la replantation se sont révélées essentielles à la propagation de ladite maladie. Au cours de l'histoire, davantage d'évidences ont démontré comment des changements sociaux, économiques et culturels majeurs ont influencé l'épidémiologie de diverses maladies des plantes transmises par les semences, soulevant des inquiétudes quant au rôle joué par les voies d'obtention/dissémination des semences dans l'épidémiologie de ces maladies.

Le renforcement des systèmes semencier de manioc contre la CBSD nécessite, entre autres, la conception des stratégies d'intervention adaptées aux contextes locaux ainsi que l'utilisation des puissantes méthodes de diagnostic. Dans le passé, malheureusement, les approches permettant de saisir les spécificités des régions et des communautés lors de l'évaluation de l'épidémiologie de la CBSD n'étaient pas suffisamment documentés et utilisés. De plus, des méthodes d'investigation extensives, ne fournissant qu'une « capture » de la situation épidémique réelle à des échelles locales ont été privilégiées dans l'évaluation de l'épidémiologie de la CBSD et dans la collecte des échantillons nécessaires pour une détection précise des agents causal.

Cette thèse a été initiée pour contribuer à la mise en place d'un système semencier de manioc sain et résilient en R.D. Congo en donnant la priorité au renforcement de la prise de décisions dans la conception des stratégies d'intervention visant à atténuer les menaces liées à la CBSD, ainsi qu'en contribuant à une connaissance approfondie de la diversité des virus infectant la culture de manioc en R.D. Congo.

Une approche multidisciplinaire d'évaluation, intégrant des composantes traditionnelles de l'épidémiologie des maladies des plantes avec les composantes socio-culturelles, géographiques et agroécologiques, a été conçue et testée à travers une enquête conduite dans la province du Sud-Kivu (Uvira, D.R. Congo). Des analyses statistiques multifactorielles ainsi que des régressions logistiques ont été utilisées pour identifier les principales caractéristiques de l'épidémiologie de la CBSD ainsi que les facteurs de risque associés à la dissémination de cette maladie à travers les boutures. Des échantillons de feuilles et de tiges du manioc ont été collectées et

utilisés pour la détection par RT-PCR des virus responsables de cette maladie (CBSV et UCBSV). La technologie de séquençage à haut débit à travers la plate-forme Illumina et différentes stratégies de préparation de librairie (ARN total, VANA et ARN double brin) a été utilisée comme méthode avancée d'investigation de la diversité des virus infectant le manioc. Des échantillons collectés suivant une approche d'échantillonnage intensive ont été utilisés pour cette finalité. Des analyses bio-informatiques métagénomiques ont été utilisées pour reconstruire des contigs. L'approche de screening basée sur l'homologie de séquence nucléotidiques et protéiques et la présence de motifs protéiques a été utilisée. . Des analyses complémentaires (identification des gènes et des sites de clivage) ont été effectuées pour caractériser les virus détectés. L'identification des sites de clivages et la déduction des domaines fonctionnels a été réalisé à partir des alignements des séquences des acides aminés des polyprotéines. Les séquences des génomes viraux assemblées ont été par la suite confirmées par RT-PCR et par séquençage Sanger. Leurs extrémités 5' et 3' ont aussi été reconstruites par l'utilisation du système d'amplification rapide des extrémités de l'ADN complémentaire (RACE). Des alignements multiples des séquences ainsi que des analyses phylogénétiques selon divers méthodes prédites pour les séquences d'acides aminés et des nucléotides ont été effectuées.

Les principaux résultats obtenus peuvent être résumés de la manière suivante :

Les analyses factorielles des données épidémiologiques ont révélé que la propagation de la striure brune du manioc dans le milieu d'étude est spatialement hétérogène et caractérisée par l'existence de trois zones/clusters dont l'une, située en moyenne altitude, a montré la plus faible pression de la maladie. Ceci pourrait rendre cette zone propice aux activités de multiplication/production des boutures. Neuf paramètres ayant trait à l'épidémiologie de la striure brune du manioc se sont révélés importantes dans la caractérisation de cette répartition spatiale de la maladie à savoir : le type de virus impliqué dans l'infection, le type et l'incidence des symptômes foliaires, l'âge des plantes, l'origine des boutures, le système de culture, le statut foncier du champs, le nombre des mouches blanches ainsi que la présence des mauvaises herbes dans les champs de manioc. Ces résultats suggèrent que les stratégies de mitigation de la propagation de la striure brune dans la zone étudiée devraient être variables en tenant en compte des spécificités qui caractérisent chaque cluster.

La détection moléculaire par RT-PCR des virus impliqués dans la causation de la striure brune du manioc a révélé trois résultats importants : (i) le taux d'incidence global des virus de la striure brune était élevé au moment de l'étude (31.3% in 2019) comparativement aux taux d'incidence rapportés antérieurement, (ii) la prévalence de l'infection simple par UCBSV était la plus élevée (42.9%), suivie par l'infection simple par le CBSV (35.1%) tandis que l'infection mixte a montré une faible prévalence (22.1%) (iii) en fonction des cluster, les infections simples par le CBSV et l'UCBSV étaient les plus prévalents dans les clusters 1 et 3 respectivement, tandis que l'infection mixte CBSV+UCBSV était prévalent dans le cluster 2.

Les analyses prédictives ont révélé les facteurs qui pourraient influencer la propagation de la CBSD à travers les voies d'obtention des boutures, entre autres : (i) les facteurs en rapport avec les connaissances paysannes, (ii) les facteurs en rapport avec la location géographique des champs paysans, ainsi que (iii) le type de canal utilisé par l'agriculteur pour obtenir les boutures.

Le séquençage à haut débit a permis d'élargir la diversité des virus infectant le manioc dans la zone d'étude. En effet, plusieurs isolats formant potentiellement deux espèces de virus du genre ampelovirus (*Closteroviridae*), le Manihot esculenta-associated virus 1 and 2 (MEaV-1 et 2), ainsi que des isolats formant potentiellement une espèce de virus, le Cassava--Congo cheravirus (Ca-Coc) du genre cheravirus (*Secoviridae*), ont été identifiés pour la première fois.

Une initiative collaborative entre chercheurs a permis d'étendre les zones de présence de ces nouveaux virus et d'enrichir leur diversité. En effet, les deux espèces d'ampelovirus ont aussi été détectées dans les échantillons originaires du Rwanda, de Madagascar, de la Réunion et de Mayotte tandis que des isolats de cheravirus avaient déjà été détectés dans les échantillons originaires du Congo-Central (Bas-Congo) à l'Ouest de la R.D. Congo. En attendant la validation, ces nouveaux virus pourraient aussi être disséminés en Tanzanie puisque les recherches dans les bases des données RNAseq ont permis d'identifier des données à partir desquelles des génomes viraux proches de ceux des nouveaux virus ont été assemblés.

La caractérisation du génome du cassava-Congo cheravirus a apporté des nouvelles informations qui ont permis de révéler l'existence d'un site de clivage, autrefois inconnu, qui délinée deux gènes (codant pour deux protéines, X1 et X2) en amont de la NTB, une caractéristique jusque-là réservée aux virus du genre nepovirus et sadwavirus mais qui désormais serait présent pour les cheravirus notamment. En plus, un gène codant pour un homologue de l'ITPase (HAM1) a été identifié dans la partie c-terminale du gène codant pour la polymérase. Seulement deux des trois résidus d'acides aminés essentiels étaient conservés pour la protéine codée par ce gène, comparativement à l'ITPase fonctionnellement caractérisée des humains. Par contre, tous les acides aminés intermédiaires et indispensables n'étaient pas conservés du tout.

Les résultats de cette thèse ouvrent d'intéressantes perspectives vers l'amélioration des approches utilisées pour étudier l'épidémiologie de la striure brune du manioc afin de les rendre aptes à fournir des résultats utiles et nécessaires pour des prises de décisions efficaces dans la réponse contre la propagation de cette maladie à travers les systèmes semenciers. En outre, ils ouvrent des perspectives vers la caractérisation des aspects biologiques de ces virus nouvellement identifiés. Il y a une opportunité d'adapter l'approche esquissée dans ce travail pour la caractérisation de l'épidémiologie d'autres maladies des plantes importantes pour l'Afrique et, d'une manière importante, d'investiguer la diversité des virus dans d'autres écosystèmes.

Mots clés : Manioc, striure brune, épidémiologie, diversité virale, système semencier, ampelovirus, cheravirus, D.R. Congo.

Remerciements

Eternel mon Dieu, tout ceci n’est que grâce est le fruit de ton amour. Merci pour tout.

Les prémisses de mes remerciements sont naturellement adressés à la personne qui m’a suivi au jour le jour durant cette formation doctorale, mon promoteur, Professeur Sébastien Massart. Non seulement que j’ai eu la chance de croiser la route d’un être aux valeurs humaines impressionnantes, mais surtout j’ai passé 5 précieuses années à partager l’une de vos passions : « les virus des plantes ». Tu as fait émerger en moi des potentialités que peu de gens, on dirait autant moi-même, pouvaient identifier. Je garde de toi la passion, la rigueur, la discipline du travail et plus important, ta grande bienveillance. Merci de m’avoir accueilli, encadré, formé et surtout d’avoir ouvert ma vie à des grands enjeux.

Professeur Espoir : Tu as ouvert ma vie à la grandeur. Je me rappelle de cette matinée du 14 Novembre 2014 lorsque je suis venus te voir sur le balcon du campus de Karhale pour demander une lettre de recommandation afin de postuler pour une bourse de maitrise. Même après ces études de master, tu avais une idée plus claire que moi-même sur la voie professionnelle qu’il me fallait emprunter. Tu m’as rappelé de là où j’étais et tu m’as motivé encore et encore pour parachever ces études doctorales. Tu m’as aidé avec des conseils , surtout quand j’étais sur le point d’abandonner. Tu n'as pas hésité de m’amener sur d’autres champs d’apprentissage en dehors du domaine des recherches doctorales et tout cela, pour mon bien. Voici pour moi le moment de dire infiniment merci pour tout. Ma grande fierté n’est pas seulement de clôturer cette étape, mais surtout d’honorer la confiance que tu as placé en moi. Je garde de toi les leçon de sacrifice, de persévérance et de responsabilité.

Mes remerciements s’adressent également au Professeur Hervé Vanderschuren qui a assuré la gestion du projet I-CARE. Tu as répondu à mes multiples sollicitations au-delà d’avoir orienté et éclairé le déroulement de cette formation doctorale. Tes multiples interventions ont facilité l’aboutissement de ce travail et je dis infiniment merci. Je remercie infiniment les professeurs Claude Bragard et Ludivine Lassois pour leurs conseils et collaborations dans les multiples aspects de recherches effectuées.

Je suis infiniment reconnaissant au Dr Stephan Winter pour n’avoir pas hésité de soutenir les efforts de ma formation en disponibilisant, sans contrepartie quelconque, les résultats de ses recherches personnels et de m’avoir reçu dans son laboratoire pour un séjour de recherche.

Un grand merci particulier à tout le staff du laboratoire de phytopathologie intégrée et urbaine et à tous les chercheurs qui y ont séjourné durant mon passage. Vous m’avez accordé votre temps, votre attention et votre tolérance lorsque j’en avais besoin. Merci d’avoir accepté de partager votre espace: Gladys, Arnaud, Abdoul, Chadi, François, Lucie, Johan, Nuria, Coline, Angelo, Vanessa, Hajar je vous remercie infiniment. Je remercie également le staff du laboratoire de génétique des plantes dont Soklay, Kumar, Sara, Chantal et Jean Pierre.

Un grand remerciement est adressé à Welmoed, Frank et Esto Bahizire respectivement de l'institut de santé internationale-Charité-Universitätsmedizin Berlin et du Centre pour les maladies Tropicales et la santé globale (CTDH-UCB) pour avoir organisé un séjour de visite au sein de leur groupe de recherche.

Je voudrais exprimer ma profonde gratitude envers les personnes qui m'ont aidé à collecter les échantillons et effectuer les enquêtes de terrain : Patrick Safari Baluku, Olivier Bisimwa, Mwenda et aux memorands qui y ont participé : Lucet et Jovin. Un merci particulier à Jules, Benjamin, Léandre pour l'attention manifestée sans cesse à ce parcours, pour vos multiples implications et soutiens qui ont été très importants. Je suis très reconnaissant envers les membres de la Faculté d'agronomie ainsi que de la grande communauté de l'Université Catholique de Bukavu, vous avez été plus qu'une équipe professionnelle, vous avez été une famille. Merci pour la diversité des soutiens que vous m'avez apporté ainsi que l'accompagnement dont j'ai bénéficié de votre part au courant de ce périple.

Finalement et d'une manière importante, ma famille :

Augustin Bisimwa, mon regretté Papa. Tu nous a été arraché très tôt. Merci pour tes sacrifices incommensurables. J'honore ta mémoire à travers ce travail. Maman Hélène et tous mes frères et Sœurs, les BISIMWA : Jolie, Patrick, Yvonne, Wivine, Natacha, Olivier, Martine, Charles et Charlotte, merci pour avoir été là pour moi, pour votre amour, vos prières, vos encouragements et votre attention. Mes sincères remerciements sont adressés à toute la famille Bajope BALUKU Jean-Pierre, pour ses multiples soutiens et motivations témoignés.

Rosette Baluku, ma tendre épouse, tu as été d'accord que je puisse entamer ce long processus d'études. Tu as enduré seule mes longs moments d'absence, partagée entre le travail, les études et la famille. Tu es celle qui a porté les lourds tributs de cette thèse. Malgré cela, ton amour et ton soutien ont été indéfectibles et m'ont été d'une aide incommensurable pour tenir la route. Ce travail est le couronnement de tous les sacrifices que tu as accepté et un autre signe de la valeur que tu représentes dans ma vie. Puisses-tu trouver à travers ces lignes mes sincères hommages et l'expression de ma profonde gratitude.

Mes chers enfants Brial, Wise-Sébastien et Harmonie, je ne sais pas exactement ce que vous vous disiez quand vous me voyiez débarquer à la maison après des longs moment d'absence. Merci infiniment d'avoir enduré ces moments. A vous et à tous les autres qui arriveront, je dédie ce travail.

Il m'est impossible de lister toutes les personnes formidables que j'ai rencontré durant ces cinq années et qui ont contribué, parfois même sans s'en rendre compte, à cette aventure scientifique et surtout humaine qu'a été mon doctorat. Trouvez à travers ces lignes l'expression de mes sincères gratitudes.

Cette thèse n'aurait pas été possible sans le soutien financier de "l'Académie de Recherche et d'Enseignement Supérieure (ARES)" ainsi que la bourse complémentaire de l'Université de Liège.

Table of contents

Abstract I
Résumé IV
Remerciements VII
Table of contents X
List of Figures XV
List of Tables XX
Study context XXIII
Preliminaries 3
I.1. Origin, dissemination and Importance of cassava 4
I.1.1. Origin 4
I.1.2. Dissemination in Africa 5
I.1.3. Importance of cassava 7
Production of cassava 8
a. Food and economic importance 10
b. Uses and nutritional value of cassava products 12
c. Ecological requirements 15
d. Production systems 16
I.2. Viral pandemics associated with cassava in Africa 17
I.2.1. General consideration 17
I.2.2. The cassava mosaic disease (CMD) 17
a. History 17
b. Aetiology, diversity, vectors and transmission modes 17
c. Genome organization of CMD-associated viruses 23
d. Epidemiology 25
e. Economic importance 27
I.2.3. The cassava brown streak disease (CBSD) 28
a. Historical background 28
b. Research milestones from 2000's 30

c. Etiology, structure, properties, and diversity of the CBSD-associated viruses .31
d. Symptoms and diagnostic methods 39
e. Transmission modes, vectors, host range, and epidemiology 40
f. Economic importance 45
I.2.4. Control and management of CMD and CBSD 46
i. Management of CMD 46
ii. Management of CBSD 48
I.2.5. Other viruses 50
I.3. Diagnosis of viral plant diseases 52
I.3.1. General consideration 52
I.3.2. Plant virus diagnosis 52
I.3.3 Plant virus identification 54
I.3.4. Plant virus detection methods 56
a. Biological indexing 56
b. Electron-microscopy (EM) 57
c. Serological-based methods or immunoassays 59
d. Molecular or nucleic acid-based methods 64
I.4. Sequencing technologies 75
I.4.1. The advent of sequencing technologies 75
I.4.2. First generation of DNA sequencers 75
a. Sanger sequencing technology 75
b. The Maxam-Gilbert sequencing technology 77
I.4.3. Second generation of DNA sequencers or "High throughput sequencing" 77
1. The method used to detect the nucleotide sequences 78
2. The proximate source of the nucleotide: 79
3. The sequencing chemistry employed: 81
4. The length of the read produced: 81
I.4.4. Conclusion: comparison of various sequencing strategies and performances in the detection of plant viruses 82
I.5. Decision-making for selecting a diagnostic test 85

I.6. Cassava seed systems and mitigation of disease outbreaks in Africa 89
I.6.1. The seed systems paradigm and farmer's perceptions on the value of seeds in Africa 89
I.6.2. Cassava seed systems in D.R. Congo 91
I.6.3. Understanding interventions in cassava seed system in a context of viral diseases burden: a brief analyse of the Great Lakes Cassava Initiative (GLCI) experience. 98
a. Introduction 98
b. Organization of the project 99
c. Seed system innovations within the GLCI and its challenges 99
d. Learning from the GLCI project implementation and coordination 102
Chapter II. Research gap, objectives and thesis structure 110
II.1. Research gap 110
II.1.1. The epidemiology of plant diseases for seedborne pathogens 110
II.1.2. The role of seed systems in the epidemiology of viral diseases: the need for a new paradigm in Africa 111
II.2. Research hypothesis and objectives 114
II.3. Thesis structure 115
Preliminaries 118
III. Risk factors associated with cassava brown streak dissemination through seed pathways in Eastern D.R. Congo 119
III.1. Abstract 120
III.2. Introduction 121
III.3. Material and methods 123
III.3.1. Conceptual framework 123
III.3.2. Study area 124
III.3.3. Farmers and fields selection 124
III.3.4. Farmer interviews 126
III.3.5. Epidemic survey 126
III.3.6. Sample collection: leaves and stem cuttings 126
III.3.7. Molecular analysis 127

III.3.8. Data analysis 128
III.4. Results and interpretations 131
III.4.1. Identification of clusters within the study area 131
III.4.2. CBSVs detection and symptom incidence within the study area and the clusters 134
III.4.3. Pathways of cuttings used by farmers 137
III.4.4. Seed system risk factors associated with CBSD 139
III.5. Discussions 142
III.6. Supplementary material 146
IV. Novel ampeloviruses infecting cassava in Central Africa and the South-West Indian Ocean Islands 155
IV.1. Abstract 156
IV.2. Introduction 157
IV.3. Material and methods 158
IV.3.1. Origin of the analyzed cassava samples and high-throughput sequencing 158
IV.3.2. Bioinformatic analyses 160
IV.3.3. Confirmatory RT-PCR and sequencing 160
IV.3.4. Characterization of defective RNA(D-RNA molecules) 161
IV.4. Results 162
IV.4.1. Identification of *Closteroviridae members in cassava* 162
IV.4.2. Genome annotation 163
IV.4.3. Phylogenetic analysis of cassava isolates with members from the family *Closteroviridae* 166
IV.4.4. Diversity of the new isolates 168
IV.4.5. Recombination analysis 171
IV.4.6. Validation of defective RNAs 172
IV.5. Discussions 174
IV.6. Supplementary Materials: 176
Preliminaries 184

V. The Cassava-Congo cheravirus is characterized: a Maf/HAM1 motif is revealed and two domains (X1 and X2) are proposed upstream the NTB in the genus Cheravirus 185

V.1. Abstract 186

V.2. Introduction 187

V.3. Material and methods 188

V.3.1. Virus isolation from cassava collected in Bas-Congo 188

V.3.2. Study in South-Kivu province 190

V.3.3. Identification of the RNA1 and RNA2-encoded cleavage sites 192

V.3.4. Data mining 193

V.4. Results 194

V.4.1. Virus isolation, electron-microscopical examination, purification and genome reconstruction of a virus isolate from Bas-Congo 194

V.4.2. HTS analysis of cassava samples from South-Kivu Province and data mining of publicly available RNASeq data 195

V.4.3. Genome annotation and phylogenetic placement 197

V.4.4. Diversity of the new isolates 202

V.4.5. Identification of the RNA-1 and RNA-2 encoded proteins and cleavage sites 204

V.4.6. A putative Maf/HAM1 motif is present within the C-terminal part of the polymerase domain 210

V.5. Discussions and conclusions 212

V.6. Supplementary material 217

VI. General discussions and conclusion 231

BIBLIOGRAPHY 247

List of Figures

Figure I-1.Three putative areas of initial domestication of cassava in the Brazilian neotropics as suggested by various authors cited by (A. C. Allem 2001). Whether cassava was first domesticated in Northern Brazilian Amazonia, or, somewhere in Brazil's westernmost states of Rondônia and Mato Grosso is controversial 5

Figure I-2. Places and dates where cassava was first reported and hypothesized paths by which the crop was diffused in Africa (S. E. Carter et al. 1993)......................7

Figure I-3. Trends in global cassava production from 1960 to 2020. Figure elaborated using data from (FAOSTAT 2022) 8

Figure I-4. Cassava production: Ranking of leading countries in term of cassava production and yield. Figure elaborated using data from (FAOSTAT 2022). 9

Figure I-6**.** Regional use (in 2020) of cassava and products. Figure elaborated using data from (FAOSTAT 2022)..11

Figure I-7**.** Production of cassava in D.R. Congo on a per province basis (Computed with QGIS 2.14 Essen using the INS 2019 production data) (INS - RDC 2021).12

Figure I-8. Phylogenetic relationships among various representative species of the cassava mosaic geminivirus. Sequences were aligned using the ClustalW algorithm, and the tree was constructed by the Neighbor-Joining method. 22

Figure I-9. Typical genomic organization of cassava mosaic begomoviruses. 24

Figure I-10. Principal routes of dissemination of cassava from South America to Africa and India, and distribution of cassava cultivation in Africa. The hatched lines mark the northern and southern limits of cassava production, and the dates are the years African cassava mosaic disease was first described in each country (C. Fauquet, Fargette, and Munihor 1990). 26

Figure I-11. Various morphologies of CMD symptoms on leaves..........................26

Figure I-12. Distribution of cassava mosaic disease in Africa and Asia. Adapted from (CABI 2023)..27

Figure I-13. Literature analysis on (**A**) the number of scientific publications as well as (**B**) the main thematic covered by the 206 publications on cassava brown streak disease (or virus) retrieved through SCOPUS between 2000's and 2022. 30

Figure I-14. Key events in cassava brown streak disease (CBSD) research history (2000-2022). CP, coat protein; RT-PCR, reverse-transcription polymerase chain reaction, CBSD, cassava brown streak disease; CBSV, cassava brown streak virus; UCBSV, Ugandan cassava brown streak virus; RNAi, RNA interference; HTS, high throughput sequencing; CMD, cassava mosaic disease; NWG, new whole genome. 31

Figure I-15**.** Genome organization of representative species for various genus in the family *Potyviridae*. The figure highlight the unique features of the CBSV and UCBSV genome organization. This figure is a personal adaptation based on

genome organization from the ICTV website [https://ictv.global/report/chapter/potyviridae/potyviridae] accessed on December 12th, 2022).. 34
Figure I-16A. Phylogenetic trees based on the first five genes encoded by the CBSV and UCBSV and showing how these genes place these viruses into three clades. The trees were generated using best-fit model preselected in jModel Test. The number at each branch represents the bootstrap value (1000 replicates) and the scale bar represents nucleotide substitution per site. The figures were adapted from (Amisse et al. 2019).. 37
Figure I-16B. Phylogenetic trees based on the five additional genes encoded by the CBSV and UCBSV and showing how only the HAM and CP genes place the isolates into two distinct clades whereas three distinct clades could be seen using the 3 other genes as shown in figure 16A. The trees were generated using best-fit model preselected in jModel Test. The number at each branch represents the bootstrap value (1000 replicates) and the scale bar represents nucleotide substitution per site. The figures were adapted from (Amisse et al. 2019). 38
Figure I-17. Symptoms of CBSD in leaves (A-F), stem (G) and roots (H-L) of cassava (for leaves, symptoms in *N. benthamiana* are also inserted). 40
Figure I-18. Status of CBSD outbreak in African countries and Comoros until 2022. 44
Figure I-19**.** Various virus-like symptoms on cassava leaves and roots observed out of Africa under greenhouse conditions:... 51
Figure I-20**.** Evolution of virus classification pipelines in the past from 1970s to 1990s (blue arrows), currently (green arrows) and a classification proposed for enabling both metagenomic sequence data and conventionally derived virus sequences to be classified (red arrows). Figure adapted from (Simmonds et al. 2017) 55
Figure I-21: Schematic diagram of families an genera of viruses and of viroids that infect plants (Agrios 2005).. 58
Figure I-22. Schematic representation of the organizational framework of seed system in D.R. Congo and it particularity in relation to cassava multiplication scheme. 92
Figure I-23. Cassava variety maintenance scheme in D.R. Congo......................... 97
Figure I-24: The GLCI operating framework.. 99
Figure II-1. Various factors of host, pathogen and environment involved in epidemic (Haq, Khan, and Sarwar 2022).. 110
Figure II-2. The framework adopted in this study for investigating the epidemiology of a cutting-borne viral disease as is the case for the CBSD. 112
Figure II-3. General organization of the thesis 115
Figure III-1. Illustration of the conceptual framework adopted in this study 123
Figure III-2. Geographic map of the Study area showing subdivision into 5 sites. Red dots represent surveyed fields 124

Figure III-3. **(A)** Cluster dendogram showing the repartition of the data into three clusters using the Hierarchical Clustering on Principal Component method. **(B)** Mapping of the clusters identified by HCPC in the study area. Each color category is associated to a cluster: Green for cluster 1, yellow for cluster 2 and red for cluster 3. 131
Supplementary Figure 1 Schematic representation of a cassava field showing the selection strategy and process for field observation and sample collection. 146
Supplementary Figure 2. Results from molecular detection showing the agarose gel with amplicons: 147
Supplementary figure 3. Geographical mapping of results from molecular detection 148
Figure IV-1. Schematic representation of the genomic organization of representative isolates 164
Figure IV-2**.** Phylogenetic analysis of the aligned amino acid sequences of the *HP70h* (ORF3) of the thirteen isolates from cassava and of selected members of the family *Closteroviridae* 167
Figure IV-3**.** Phylogenetic trees reconstructed using the amino acid sequences of the following three taxonomically relevant proteins for the family *Closteroviridae*: 170
Figure IV-4**.** Geographic repartition of the two novel cassava ampelovirus species (MEaV-1 and MEaV-2) according to the country of origin of positive samples. 171
Figure IV-5**.** Recombination event detected among isolates from MeaV-2 by the RDP4 program....172
Figure IV-6**.** Gel red-stained 2% agarose gel showing RT-PCR products from: (**a**) D-RNA1 expected amplicon size 380 bp; (**b**) D-RNA2 and D-RNA3, expected amplicon size 380 bp and 400 bp, respectively; and (**c**) D-RNA4, expected amplicon size 500 bp. M1: GeneRuler 1 kb plus ladder (Thermo Scientific, SM1343). M2: fast DNA ladder 10 kb (NEB N3238S). 173
Figure S1: Multiple sequence alignment of D-RNA1 clone (referred to in the figure as D-RNA1_Sanger) with the defective reference from HTS (MEaV-1_RE-LJV-D-RNA-1)....176
Figure S2: Multiple sequence alignment of D-RNA2 clone (referred to in the figure as D-RNA2_Sanger) with the defective reference from HTS (MEaV-1_RE-LJV-D-RNA-2)....177
Figure S3: Multiple sequence alignment of D-RNA3 clones (two clones were sequenced and are referred to in the figure as D-RNA3a-Sanger and D-RNA3b-Sanger) with the defective reference from HTS (MEaV-2_RE-LJV-D-RNA3). 177

Figure S4: Multiple sequence alignment of D-RNA4 clone (referred to in the figure as D-RNA4-Sanger) with the defective reference from HTS (MEaV-2_MG-Mena-D-RNA-4)... 178

Figure V-1**.** Cassava plant collected in 2007 from stem cuttings and maintained in the DSMZ greenhouse at DSMZ. Red necrotic spots (red arrows) and severe mosaic symptoms indicate the presence of an additional virus or other pathogens....... 189

Figure V-2. Cassava landraces (a: Nabwilalanga, b: Nabwigoma, c and d: Rava) collected in 2019 as stem cuttings and grown in a greenhouse at Gembloux Agro Bio-tech showing mild to low leaf yellowing symptoms spreading from secondary veins.. 191

Figure V-3. (a) Virus symptoms in *N. benthamiana* starting with chlorotic mottle and leaf deformation of youngest leaves, resulting in an oak leaf pattern on old leaves and young leaves recovered from symptoms. (b) Electron microscopically examined Adsorption preparations of leaf homogenates showing VLp of icosahedral structure (red arrows), sometimes empty (black arrows). 194

Figure V-4. Electrophoretic separation of soluble leaf proteins from healthy *N. benthamiana* (lanes 2 to 6 and lanes 8 to 10) and from purified virions (lane 7) in a discontinuous SDS-PA gel (gradient 8-20%). 195

Figure V-5. Geographical map of the study area showing the localization of the provinces where samples originated and the spatial repartition of positive fields in the four pools from South-Kivu Province. 197

Figure V-6. Genome organization of the RNA1 and RNA2 segments for the reference isolate CGO-BC. The RNA genomes are drawn as black lines, and coloured rectangles represent the predicted open reading frames. A scale is presented above to size the length of the RNA genomes in nucleotides.................................... 198

Figure V-7**.** Phylogenetic analysis of the conserved amino acid region between the "CG" motif of the protease and the "GDD" motif of the polymerase (the Pro-Pol region) for isolates detected in cassava (strong green triangles and green frame in the tree) and of type isolates from recognized species in the family *Secoviridae* (detailed information on these viruses is provided in supplementary material 5). 201

Figure V-8. Putative consensus cleavage sites delineating putative functional domains predicted within RNA1 (P1) and RNA2 (P2)-encoded polyproteins. 204

Figure V-9**.** Computer-assisted prediction of transmembrane helices (TMH) in the X2-encoded domain of the RNA1 segment:.. 207

Figure V-10**.** Phylogenetic relationships of Maf/HAM1 motifs between viruses reported in this study (framed in blue) with those of representative members of different kingdoms of life (viruses are pointed in solid red circles)................. 211

Supplementary material 2. Accession numbers and abbreviations for selected *Secoviridae* members used to deduce cleavage sites and functional domains of the RNA1 and RNA2 encoded polyproteins (respectively P1 and P2) 217

Supplementary material 4. Phylogenetic analysis of the coat protein amino acid for the five isolates detected in cassava (strong green triangles and green frame) and of type isolates from recognized species in the family *Secoviridae* (detailed information on these viruses is provided in supplementary material 5)............220

Supplementary material 6. Phylogenetic trees reconstructed using **(A)** the amino acid sequences of the "Pro-Pol" region, **(B)** the amino acid sequences of the entire P1 polyprotein, **(C)** the complete/nearly nucleotide sequence of the RNA1 segment **(D)** the amino acid sequences of the coat protein block, **(E)** the amino acid sequences of the entire P2 polyprotein and **(F)** the nearly/complete nucleotide sequence of the RNA2 segment for the new *Secoviridae* members described in this study. 224

Supplementary material 7. Multiple amino acid sequence alignment of the region between the X1 and the X2 domains shows putative cleavage sites delineating the X1 and X2 domains (highlighted in yellow for viruses identified in this study and red for other cheraviruses previously characterized)..224

Supplementary material 8. Multiple amino acid sequence alignment of the protease domain localises the conserved Histidine of the substrate binding pocket region and the protease catalytic Cysteine. 225

Supplementary material 9. Multiple amino acid sequence alignment of the conserved "CG-GDD" motif shows the insertion of 36 amino acids (highlighted in red) upstream of the "GDD" motif in the polymerase domain**.** 225

Supplementary material 10: Multiple amino acid sequence alignment of the 5' extremity of the polymerase domain of viruses reported in this study, together with the arracacha virus A. The figure shows a putative cleavage site that could delineate a putative domain downstream of the polymerase.............................226

Supplementary material 11. Multiple amino acid sequence alignment of the RNA2-encoded polyproteins showing the insertion of 6 amino acids at the N-terminal extremity of the MP domain in the polyprotein of the P_Winter isolate. The alignment was trimmed to show the location of the insertion...........................226

Supplementary material 13: Multiple amino acid sequence alignment showing conserved sites of ITPases from representative members of various kingdoms of life. The alignment was built using Mview (Brown, Leroy, and Sander 1998) .228

List of Tables

Table I-1. Some of the common cassava-based products (adapted from Parmar 2017). 14
Table I-2. Viruses infecting cassava in various regions where cassava is cultivated. 19
Table I-3. Summary of CBSD history before 2000's (Legg and Thresh 2003; K. R. Tomlinson et al. 2018a; Vanderschuren and Rey 2017) 29
Table I-4_A**.** Four main events associated with the epidemiology of CBSD in Africa 42
Table I-4_B : Summary of the advantages, disadvantages and example application of mostly-used conventional approaches for virus detection 73
Table I-5: Advantages and drawbacks of different sample preparations for HTS-based detection of plant viruses. The type of protocols used to detect a specific nucleotide is used as the categorization criteria (Maclot 2021). 84
Table I-6. Farmer's perceptions of the value of seeds for traditional crops 90
Table I-7. Cassava varieties approved in the national catalog of varieties in 2019 (SENASEM 2019). For further varietal characteristics, refer to the catalog 95
Table I-8. Technical regulations for production, control and certification of seeds and seedlings of cassava. Edition 2019. Excerpt of the phytosanitary conditions regarding diseases and pests 96
Table I-9. Partners views on the importance of each others in the implementation of the project and institutional benefit gained through the GLCI project 105
Table I-10: The mechanism of the working relationship between local partners and other partners involved in the GLCI project. Assessment by local partners themselves 107
Table III-1. Characteristics of the 5 sites 125
Table III-2**.** Rules (**Rn**) for interpretation of odd ratios 129
Table III-7. Proportion of fields grown by types of cassava varieties from different pathways. 138
Table III-8: Prediction of risk factors associated with CBSD (based on RT-PCR detection) 139
Supplementary Table 1. Proportion of fields following pathways used to acquire cuttings and following means used to obtain cassava cuttings per clusters 149
Supplementary Table 2. Number of fields grown by improved as well as local cassava varieties identified in each cluster 149
Supplementary Table 3. Proportion of healthy and infected fields for each type of infection according to the pathways used to obtain cuttings 151
Table IV-1. Identity and origin of cassava landraces used and indication on the high-throughput sequencing (HTS) protocol performed on these samples 159

Table IV-2**.** RT-PCR primer pairs used for the confirmation of the viral sequences in the samples. 161
Table IV-3**.** RT-PCR primer pairs used for the characterization of the defective RNA molecules...162
Table IV-4**.** Lengths, number of reads integrated, percent of total reads and average coverage for Closteroviridae contigs from the various analyzed cassava samples. 163
Table IV-5**.** Genome length, position of ORFs and size of the encoded proteins for the representative isolates of the two species CG-Nmb (MEaV-1) and MG-Men-9 (MEaV-2). 164
Table IV-6**.** Amino acid sequence identity of the RdRp, HSP70h and CP of MEaV-2 (isolate MG-Men-9) with those of other selected Closteroviridae members. 168
Table IV-7**.** Intergroup and intragroup average pairwise divergence and standard deviation calculated for the three taxonomically relevant proteins (RdRP, HSP70h and CP) and for the nearly complete genomes. 170
Table S1: Length of ORF and molecular mass of 3 encoded proteins (RdRp, HSP70h and CP) of the 15 isolates of the cassava ampelovirus reported in this study. ...179
Table S2: Accession numbers and abbreviations for selected Closterovirus used in the phylogenetic reconstruction of the HSP70h nucleotide sequences. 180
Table S3: Estimates of evolutionary divergence between amino acid (aa) sequences for the HSP70h protein among cassava ampeloviruses isolates reported in this study. 180
Table S4. Estimates of evolutionary divergence between amino acid (aa) sequences for the HSP70h protein among cassava ampeloviruses isolates reported in this study. 181
Table S5. Estimates of evolutionary divergence between amino acid (aa) sequences for the coat protein among cassava ampeloviruses isolates reported in this study. 181
Table S6. Estimates of evolutionary divergence between nucleotide sequences for the whole genomes of cassava ampelovirus isolates reported in this study. 181
Table V-1. Genome organization of the 13 contigs reconstructed in this study 198
Table V-2**.** Variation ranges of the identity percentages for the conserved "Pro-Pol" region and the coat protein amino acid sequences between isolates described in this study and representative species of various genera in the family Secoviridae...202
Table V-3**.** Intergroup and intragroup average pairwise identity and standard deviation calculated for the "Pro-Pol" region, the CP block, the RNA1 and RNA2-encoded polyproteins (P1 and P2, respectively) and the nucleotide sequences of (nearly) complete genomes of the RNA1 and the RNA2 segments. 203
Table V-4: Location of putative cleavage sites predicted in the RNA1 and RNA2-encoded polyproteins for isolates identified in cassava plants in this study. 205

Table V-5: Characteristics of putative RNA1 and RNA2-encoded domains identified inside the P1 and P2 polyproteins of viruses reported in this study 205
Table V-6. Consensus cleavage sites of isolates from this study and those of previously-described viruses from the genus Cheravirus.................................. 209
Table V-7. Comparison of conserved amino acid residues in ITPases of representative members of various kingdoms of life (including viruses reported in this study) to essential amino acid residues that influence substrate specificity in human ITPase. 212
Supplementary material 1. Primers used for Confirmatory RT-PCR, RACE and Sanger sequencing for HAM1 motif confirmation... 217
Supplementary material 3. Lengths, number of reads integrated, percent of total reads and average coverage for the Secoviridae contigs from the various analyzed pools of cassava samples... 219
Supplementary material 5. Complete names of selected Secoviridae members used for the phylogenetic study. 221
Supplementary material 12: Complete names and accession numbers for representative members of various kingdoms of life used for phylogenetic study of the HAM1 motifs.. 227

Study context

Located in the eastern part of the country, the South Kivu is one of the 26 provinces of the D.R. Congo with a total land area of 65,070 km^2 and one of the most densely populated region, with ca 90 persons per square kilometre. With a bimodal rainfall pattern that allows a two-seasons crop cultivation calendar, this province is characterized by an altitude-driven ecology, comprised with high, mild and lowlands. Here, the cassava crop is continuously grown on small farms, with no to limited inputs, and frequently intercropped with legumes, maize etc. The yield is varying between 7 and 17 t ha^{-1} in farmer's fields (Munyahali et al. 2017).

The history of this thesis began in 2015 when I sought a research subject for a master's specialization that brought me to Professor Espoir Bisimwa, the head of phytopathology department at UCB, where the decision to work on the issue related to the burden of viral diseases of cassava in the Eastern D.R. Congo was taken.

At that time, the situation of cassava production in the Uvira territory was (and still is) seriously hampered by the spread of the CBSD: because of the loss extents, farmers referred to it as a "ruthless" disease, "sina huruma" in Kiswahili. Most of the cassava fields at that time were cropped with CMD-resistant varieties, massively provided to farmers as a result of the burden of the cassava mosaic disease (CMD) epidemics in the Great lakes region decades before. Nevertheless, these varieties are all susceptible to the CBSD. All farmers and actors involved in this question were at the rush for obtaining resistant varieties that were hard to obtain. The context was a relative institutional regulation for the movement of planting material, a lack of a clear national response strategy and a gap in the knowledge of the biology of viruses involved in the causation of this disease. Therefore, the hesitations in the implementation of efficient response to mitigate the spread of the disease became fatal to the cassava production and the livelihood of populations. In such a situation, it became difficult for farmers to find reliable sources of healthy planting material in part because it was definitely difficult to reliably diagnosing the disease.

The long collaboration history between Prof. Espoir Bisimwa and Prof. Claude Bragard that brought me first to the U.C. Louvain for initiations in the study of plant diseases was pivotal to what happened next. In fact, the few investigations conducted at that time on the etiology of viral diseases of cassava in the Uvira territory suggested that the CBSD-like symptoms morphologies in that area, were not associated to the detection of the causing viruses. An unquenched thirst to deeply investigate this etiological issue, together with a global reflection on the future of the CBSD diagnosis and mitigation strategies became critical to investigate for the good of improving the cassava production.

Definitely, in 2018, the opportunity to work on these aspects in a frame of a doctoral training came with the regional and ARES-funded project with the aim of generating tools and strategies necessary for the implementation of a robust and sustainable cassava seed system (ICARE – coordinated by Prof. Hervé Vanderschuren) for the mitigation of the CBSD. This partnership brought me to the Professor Sébastien

Massart's laboratory in Gembloux Agro-Bio Tech where the decision to investigate the etiological and epidemiological issues related to the CBSD and the use of the high throughput sequencing technologies was taken.

Chapter I

I. General introduction

“A diagnosis is not a label, it’s the beginning of a solution”

(Unknown author)

Preliminaries

This introductory chapter is focused on the review of the milestones attained in the study of the CBSD. A literature review has been undertaken to inquire on the state of the art in relation with (i) the cassava crop (origin, dissemination and importance in Africa), (ii) the diversity of viruses that infect this crop, especially those involved in the etiology of the two most important pandemic diseases in Africa (CMD and CBSD), (iii) the methods used in the diagnosis of viral diseases as well as the criteria to be considered for selecting a diagnostic test and (iv) a brief analysis of the cassava seed system context in the D.R. Congo as well as experience gained in the management of this disease in the Great Lakes region from the past.

I.1. Origin, dissemination and Importance of cassava

I.1.1. Origin

The cassava plant, *Manihot esculenta* Crantz, is a perennial shrub of the *Euphorbiaceae* (Dicotyledons) family also known as *manioc*. This name is a derivative of the term *maniot* in the *"Tupi"* language (an extinct language spoken by the native Tupi people in Brazil). The term "*cassava"* is believed to derive from the word *Casavi* or *Cazabi* which translates to the term "bread" in the *Arawak* language (language of the first indigenous people who lived in the greater Antilles and were in contact with Christopher Colombus) (Lebot 2009).

Investigations using both morphological criteria (Antonio C. Allem 1999; A. C. Allem 2001; A C Allem 2001) and molecular markers (SNPs and SSRs) (Olsen 2004) identified *Manihot esculenta ssp flabellifolia* as likely the wild progenitor of the modern cassava cultivars and proposed the Brazilian periphery of the Amazon forest as the geographical origin of this progenitor's ancient ancestors (Figure I-1).

As an agricultural crop, archaeological investigations showed that cassava domestication started between 5000 and 10,000 years ago, suggesting it is one of the oldest cultivated root and tuber crops (A C Allem 2001; Howeler, Lutaladio, and Thomas 2013; Parmar et al. 2017).

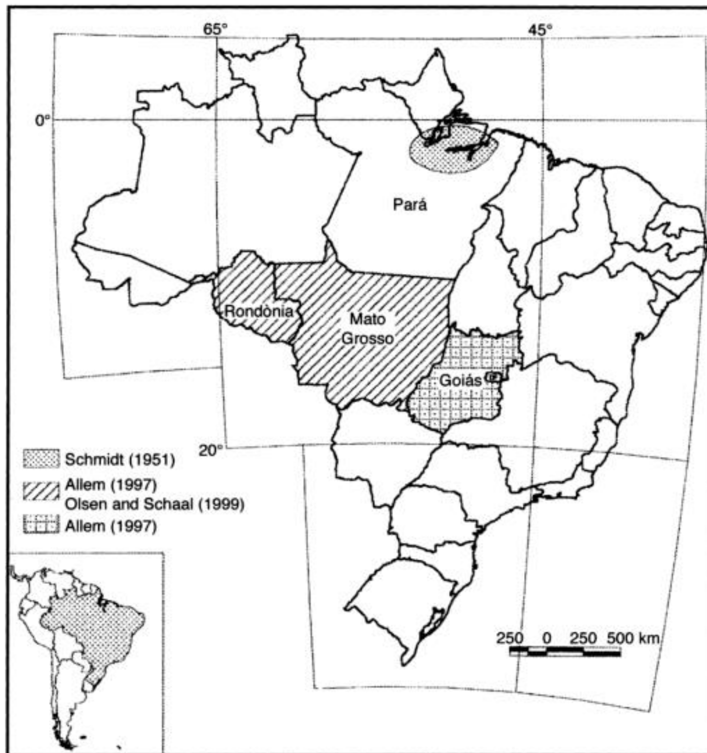


Figure I-1.Three putative areas of initial domestication of cassava in the Brazilian neotropics as suggested by various authors cited by (A. C. Allem 2001). Whether cassava was first domesticated in Northern Brazilian Amazonia, or, somewhere in Brazil's westernmost states of Rondônia and Mato Grosso is controversial

I.1.2. Dissemination in Africa

Portuguese were the first to bring cassava to Africa in the form of flour called "Farinha" by the *Tupinamba* Indians from Eastern Brazil. The first mention of cassava cultivation in Africa dates between 1558 and 1600, when the plant was cultivated to provide food for slave ships (S. E. Carter et al. 1993).

Multiple and simultaneous introductions took place at Portuguese trading stations of Fernando Po (Equatorial Guinea), Sao Tomé and Principe, Sierra Leone islands, and on the Angolan coast between Luanda and the mouth of the Congo River in the Kongo Kingdom (Ross 1975; W.O. Jones 1959).

Early dissemination of cassava to inland areas in precolonial D.R. Congo was carried out solely by Africans. Europeans entered the interior only a long period afterwards (the first was Stanley in 1877) since it was impossible to join the Congo river from the sea (R. W. Harms 1981).

Cassava plants were spread by following rivers throughout much of Central Africa during the 18th and 19th centuries. It was present on the Western shores of Lakes Nyasa (Malawi) and Tanganyika in the second half of the 19th century when Europeans first explored the area. In some regions, more widespread diffusion to the interfluves appears to have been a slower process, dependent upon trade and hindered by political relations and possibly warfare. Before the imposition of European administrations, migrations were probably the primary means of diffusion of cassava across watersheds

In West Africa, cassava was introduced at several points along the West African coast during the 17th century, from the Gambia river to Nigeria. However, unlike in Central Africa, the diffusion of cassava in West Africa was universally slow, and most of the crop's spread occurred during the late 19th and 20th centuries (W.O. Jones 1959).

Information on the diffusion of cassava in Eastern Africa is the most speculative, and there are no concrete details on the date of cassava's introduction. However, it was proposed that, as in Central and Western Africa, the plant could have been introduced in Eastern Africa at Portuguese trading stations: in Mozambique, Benguela, Sofala, Kilwa, Zanzibar, Pemba and Mombasa during the 17th or 18th centuries (S. E. Carter et al. 1993).

Given the successful introduction of the plant in Central Africa, it seems likely that cassava diffused along the eastern shores of lakes Malawi and Tanganyika to the highlands of Rwanda and Burundi. In addition, numerous travellers reported Cassava throughout the Great Lakes Regions in the mid-19th century (W.O. Jones 1959; Langlands 1966).

Cassava seems to have replaced millet, yam and plantain as the principal staple in most areas along the Congo river, resulting in a boosted trade of agricultural products. In 1698, cassava was already the staple food at the Stanley Pool (Malebo), near Kinshasa. From there, it spread upriver and inland. The main reasons for its introduction were its high caloric production per unit area and its resistance to spoilage once processed.

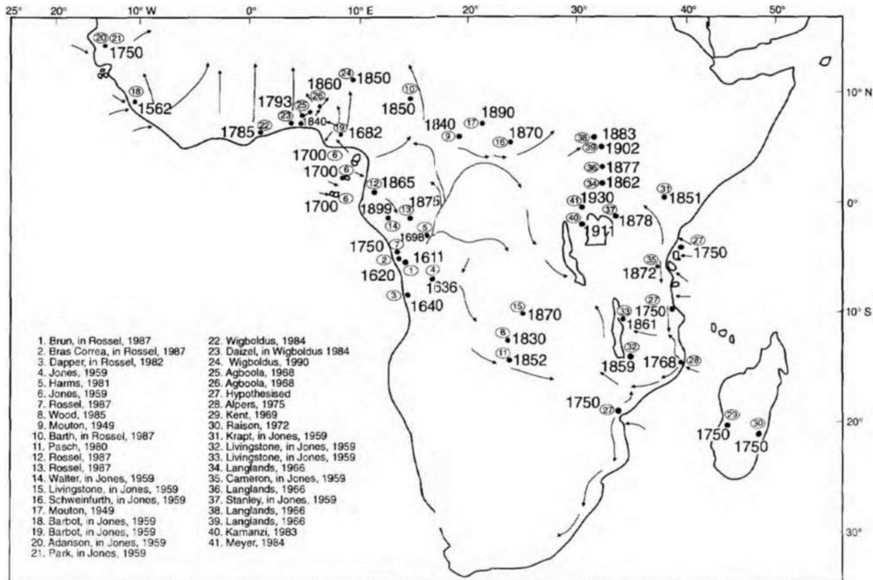


Figure I-2. Places and dates where cassava was first reported and hypothesized paths by which the crop was diffused in Africa (S. E. Carter et al. 1993).

I.1.3. Importance of cassava

Cassava consumption was widespread amongst the river people and was especially suited to take along on trips, presumably in processed form such as "chikwange", and constituted a balanced diet in combination with fish (R. Harms 1979).

Some pieces of evidence in Central and West Africa suggested that the cassava crop may have been also destined for the urban market at a very early stage of its introduction. This modifies the classical image of cassava as a traditional food staple of Africa and even the assertion that cassava is a "female crop" (S. E. Carter et al. 1993).

Whilst cassava was introduced as an anti-famine and anti-locust crop in the colonial era; it was also thought to promote laziness, soil depletion and malnutrition (W.O. Jones 1959; William O Jones 1957).

Such misconceptions/misunderstanding, half-truths and myths about the cassava crop has been the cause of its overlooking by policy makers and its underutilization in developing countries (Parmar, Sturm, and Hensel 2017). Nweke demonstrated that all of these myths and half-truths were based on misconceptions and out-of-date information.

Production of cassava

More than a hundred countries worldwide are involved in cassava production, with a total output of 302 million tons on approximately 28 million hectares (FAOSTAT, 2022 accessed 18/07/2022). Global cassava production is increasing after a sharp decline between 2014 and 2017. Production from Africa and, to a lesser extent, from Asia seems to follow this rising trend, whereas that from Latin America appears to be relatively constant (Figure I-3).

The current major producing countries are Nigeria, D.R. Congo and Thaïland. However, the main countries producing cassava are not necessarily the leaders in cassava yield. In fact, the five leading countries in terms of cassava yield per hectare are Guyana, Laos, India, Cambodia and Niger (Figure I-4). When ranking the most produced crops worldwide, cassava is in 9th place (303 million tons). However, cassava is the first most produced crop in Africa and contributes up to 64% (194 million tons) global production (Figure I-5).

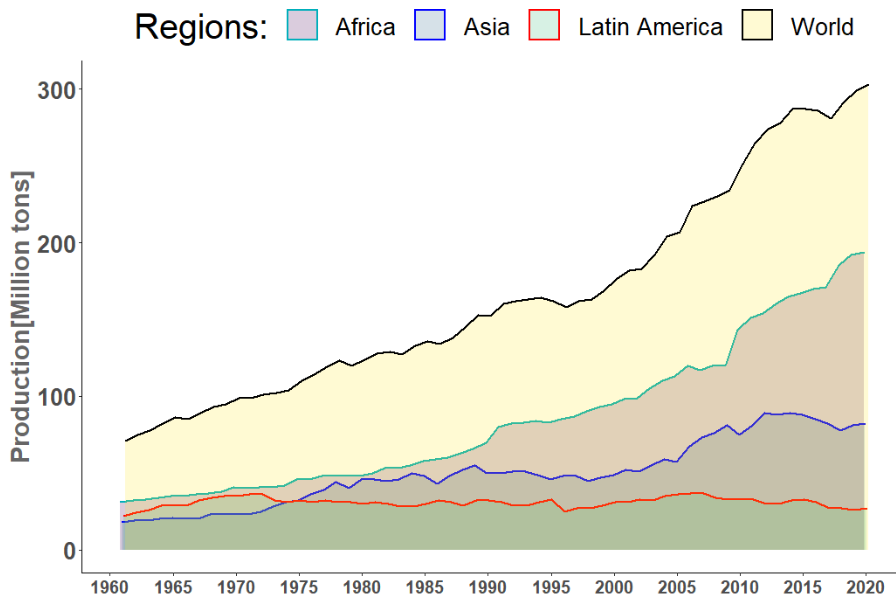


Figure I-3. Trends in global cassava production from 1960 to 2020. Figure elaborated using data from (FAOSTAT 2022)

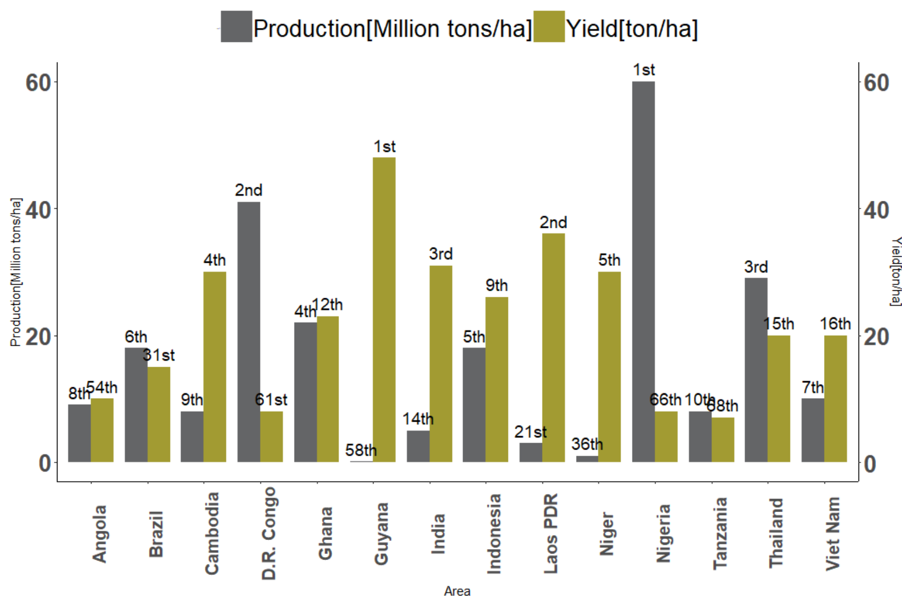


Figure I-4. Cassava production: Ranking of leading countries in term of cassava production and yield. Figure elaborated using data from (FAOSTAT 2022).

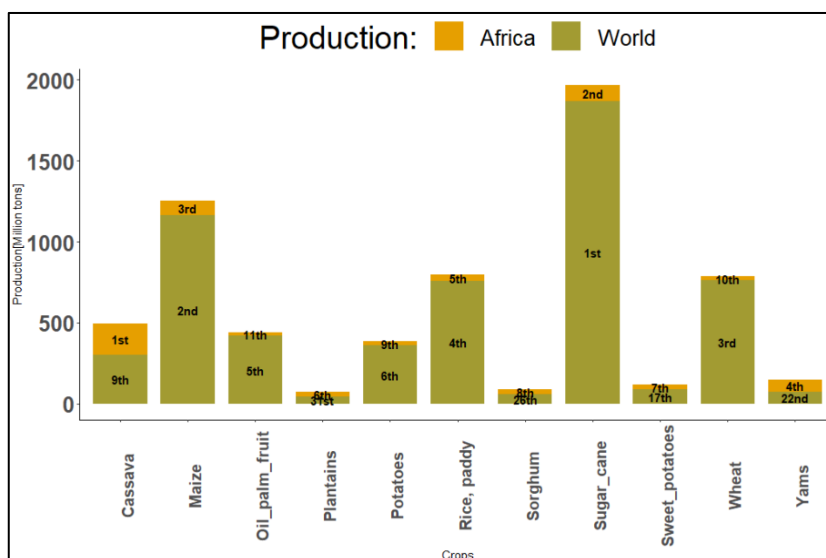


Figure I-5.Ranking of the most produced staple crops in the world and in Africa in 2020. Figure elaborated using data from (FAOSTAT 2022).

a. Food and economic importance

Cassava is a primary foodstuff for more than 800 million people, mostly living in tropical and sub-tropical regions of the World (Parmar, Sturm, and Hensel 2017). The reasons for the rapid spread of cassava cultivation in Africa include (Lebot 2009; Howeler, Lutaladio, and Thomas 2013):

- The ease of cultivation in shifting systems and the ease of propagation by stem cuttings: cassava can be planted at any time of the year, provided there is enough moisture for stem cuttings to take root.
- A flexible growth cycle and harvest: cassava roots can be left in the ground without being harvested for long periods of time ranging from six to forty-eight months after planting (Okigbo 1980).
- The resistance to locust attacks, pests and diseases, and the resistance to drought/water stress (except at planting time). These characteristics make the crop a good famine-reserve crop (S. E. Carter et al. 1993; Nweke, Spencer, and Lynam 2002).
- It adapts to poor soils or marginally fertile lands on which many other crops fail
- It has a relatively high yield and is a low-cost source of calories. It can produce more carbohydrates per hectare than any other food staple and can be harvested when needed.

The significance of cassava as a cash crop is increasing with enhanced usage in multiple industrial applications such as the production of paper, textiles, plywood, glue, biofuel, animal food and beverages (Kleih et al. 2013; Uchechukwu-Agua et al 2015).

According to the use of the crop, four stages were defined in the evolution of the cassava-based products valorisation (Nweke, Spencer, and Lynam 2002). This evolution ranged from a famine reserve crop (first stage), rural staple crop (second stage), cash crop for urban consumption (third stage) to a livestock feed and industrial raw material (fourth stage). In Sub-Saharan Africa (SSA), cassava is in a transformation phase from a rural staple to a cash crop for urban consumption and further towards an industrial raw material while in Asia and Latin America, cassava is an important industrial raw material and livestock feed for export. African countries such as Nigeria and Ghana have already entered the third phase, whereas in the D.R. Congo, Côte d'Ivoire and Uganda it is still mostly a rural staple crop. In Kenya, Malawi, Tanzania and Zambia the role of cassava is more as a famine reserve crop, owing to its drought tolerance, when the first staple (maize) fails due to erratic rainfalls. The fourth stage of cassava evolution has been realized by cassava producing countries in Asia and Latin America (Parmar, Sturm, and Hensel 2017).

The consumption pattern of cassava and its products by regions according to the FAO (2022) balance sheet is presented in Figure I-6. Currently the Oceania has the highest proportion of cassava domestic supply used as staple source of energy in human food (83%) followed by Africa (58%) and Americas (42%). In Europe, the essential part of cassava domestic supply is used for livestock feeding (52%) followed by Americas (43%), Asia (32%) and Africa (26%). There are also other non-food uses of cassava; such as starch, bioethanol etc.; for which Europe (46%) and increasingly Asia (25%) uses an essential part of their domestic supply. The maximum amount of postharvest losses and waste (food or feed loss) occur in Africa (11%), Oceania (9%) and both America and Asia (8% respectively).

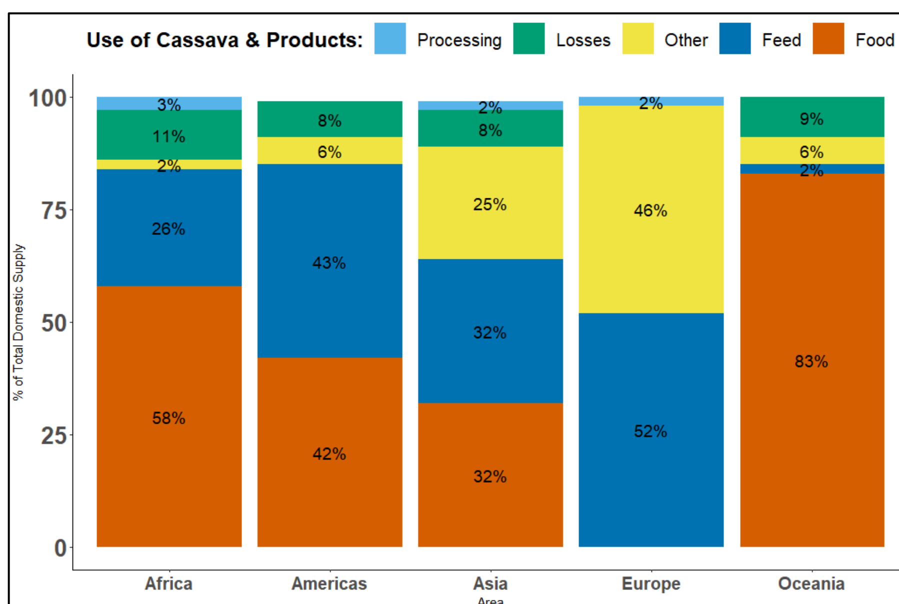


Figure I-6. Regional use (in 2020) of cassava and products. Figure elaborated using data from (FAOSTAT 2022)

According to the Observatory of Economic Complexity (OEC) (Simoes and Hidalgo 2011), the total trade in cassava was 2.56 billion USD in 2020. The major exporters of cassava and its products in this period were the Thailand (697 million USD), Cambodia (424 million USD), United States (223 million USD), Vietnam (207 million USD), and Laos (195 million USD). The top importers where Chine (879 millions), Thailand (439 millions), Vietnam (175 million), Netherlands (144 millions) and United States (127 millions).

In D.R. Congo, the cassava production has mobilized 5 million hectares of arable lands in 2020 (FAOSTAT 2022). According to the statistics from the (INS - RDC 2021), in 2019 The top provinces in term of cassava production were the Kwilu (5 million tons), the Kongo-Central (3 million tons), Haut Lomami (3 million ton) and

Tanganyika (2 million tons). In the meantime, the South-Kivu province was ranked the 13th with 1.4 million tons of cassava produced (Figure I-7).

Cassava is a very important source of food in D.R. Congo. In 2019, the average per capita consumption of cassava products were approximated to 353 kg per year, placing the country at the first rank of the African countries consuming the cassava (FAOSTAT 2022). The crop is also the first most important source of energy providing a large amount of daily calories (1,015 kcal/capita/day) to people compared to maize (190 kcal/capita/day), rice (122 kcal/capita/day) and plantains (111 kcal/capita/day) (FAOSTAT 2022).

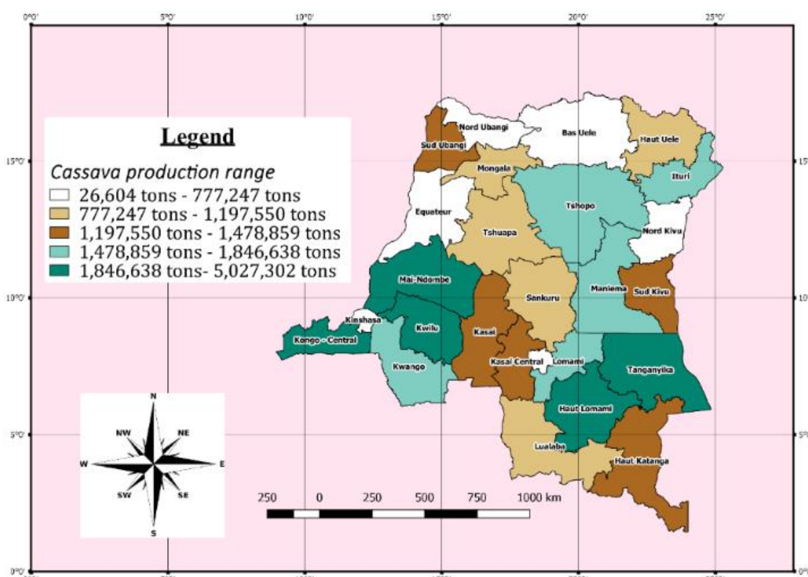


Figure I-7. Production of cassava in D.R. Congo on a per province basis (Computed with QGIS 2.14 Essen using the INS 2019 production data) (INS - RDC 2021).

b. Uses and nutritional value of cassava products

The principal product of the cassava plant is its starchy roots. However, the leaves are also used as a vegetable in various African, Asian and South American countries. The consumption of cassava leaves as a vegetable (source of protein) has been proposed to be a pure African invention (Lebot 2009).

A list of most common cassava-based human food, livestock feed and other industrial products is presented in table I-1. A huge diversity of cassava-based food specialties exists in various regions where cassava products is used for human consumption. This wide food diversity results from various treatments that cassava roots undergoes for principally reducing the rate of HCN or make the end-products suitable for human consumption.

Cassava meal for feeding livestock possesses excellent digestibility due to the presence of naturally occurring lactic acid bacteria and yeasts. In fact, micro-doses of

HCN present in cassava make enzymes like lactoperoxidase (secreted from mammary, mucosal and salivary glands) more efficient in their function as natural anti-bacterial agents (Howeler, Lutaladio, and Thomas 2013).The production of cassava-based products for livestock feeding is standard across most exporting countries such as Vietnam and Thailand (Parmar 2013).

The starch and flour from cassava are also used as raw material for various industrial products because of the easiness of the starch extraction process as well as the low-cost acquisition of cassava raw material.

Table I-1. Some of the common cassava-based products (adapted from Parmar 2017).

Category	Name/Type	Nature/Process	Region/Country
Food	Whole roots	Raw or boiled fresh starchy roots	Americas, Asia, and SSA
	Fufu, Dumby	Sticky dough	SSA
	Farina de mandioca	Course roasted flour	South America and West Indies Landang (cassava rice)
	Chikwangue	Fermented, retted and cooked cassava paste	D.R. Congo
	Landang (cassava rice)	Pelleted pulp of cassava in the form of rice.	Philippines
	Macaroni (cassava noodles)	A composite of cassava flour, ground nut flour and wheat semolina	Asia
	Gaplek	Dried cassava	Indonesia
	Gari, Kapok pogari	Toasted fermented course savoury flour.	West Africa
	Attieke	Fermented steamed cassava paste.	West Africa
	Chick-range, Lafun	Fermented sun or smoke-dried cassava	West Africa
	Lituma	Composite paste of cooked cassava roots and plantain	D.R. Congo
	Sago, tapioca pearls	Small ball shaped dessert.	Asia
	Cassava bread	Composite flour of cassava and wheat.	West Africa
	Undon	Cassava and wheat flour frozen undon noodles	Japan
	Mole	Cooked cassava roots	D.R. Congo
	HQCF	High-quality cassava flour (unfermented, pressed and	SSA
	Malemba	Cooked and gretted and washed cassava paste	D.R. Congo
	Cassava leaves	In various cooked forms as green leafy vegetable	SSA
Beverages	Mingao	Fermented starch drink	South America (Amazon region)
	Tapioca tea	Cold milk tea with tapioca pearls	Asia (Taiwan)
	Kasiri (Cassava beer)	Alcoholic drink after fermentation	America, West and South Africa Africa
	Alcohol	Distilled composite of macerated cassava and fermented	D.R. Congo
Feed	Fresh cassava leaves and roots	In limited quantities, fresh cassava leaves and roots can be fed to livestock.	SSA, Asia, America
	Cassava pellets	Pelleted cassava roots after sun-drying and steaming	Asia, Americas
	Leaves and root silage	Silage produced from cassava leaves and roots	Asia and South America
Industrial products	Adhesives (Gums with and	Cassava liquid starch processed by cooking with water.	Asia, America, SSA
	Dextrin	Cassava starch processed into dextrin by various chemical processes	Asia, America, SSA
	Liquid glucose, fructose, maltose.	Hydrolysis (acid or enzyme) of starch into glucose, fructose and maltose	Asia, America, SSA
	Biodegradable plastic	Starch based plastics.	Asia, America, SSA
	Bioethanol	Fermentation of cassava starches into alcohol.	Asia, America, SSA
	Citric and lactic acids	Citric acids using fungal strains, and lactic acids by bacterial fermentation of sugars.	China

The dry matter (DM) content of cassava roots is close to 40%, sufficiently higher compared to most other root crops but well lesser compared to other non-root crops such as wheat, rice and Maize (Lebot 2009). The cassava DM consists primarily of starch and sugars (about 90%). Its contribution to daily lipid and protein requirements is very small. This low levels of protein in cassava -based diets are a cause of concern if no supplementary protein sources (for instance legumes, leafy vegetables, and animal protein) are available (Parmar, Sturm, and Hensel 2017).

While cassava roots are a minor source of other vitamins and minerals, the young leaves however are classified as a decent source of protein and vitamins as suggested by Latif and Müller (2015). The amount of the various components are highly variable and depend on the age, cultivar, soil and ecological conditions. The HCN content of the leaves is approximately similar to roots which are degraded during the cooking or processing in the same way as in roots. Cassava leaves are an important source of Vitamin C, Vitamin A, riboflavin, phosphorus, magnesium, potassium and calcium. However, specific cooking methods results in a reduction of these vitamins (Achidi 2008).

c. Ecological requirements

Cassava cultivation is concentrated in tropical equatorial regions from 23.5°N (Tropic of cancer) to 23.5°S (Tropic of Capricorn) (Parmar et al, 2017).

Prior to European intervention, cassava was adopted by African smallholder farmers sometimes voluntarily or by force for its particular characteristics (William O Jones 1957; Nweke, Spencer, and Lynam 2002).

Cassava is cultivated throughout the year up to a maximum altitude of about 2,000m. It tolerates drought but grows best where annual rainfall reaches 1,000-2,000 mm. Cassava grows well on many types of soils, with the exception of hydromorphic soils which are unsuitable. It is propagated vegetatively through stem cuttings, and the growth cycle generally ranges between 10 and 30 months (C. Fauquet, Fargette, and Munihor 1990).Various physiological characteristics of cassava enables the crop to tolerates prolonged water deficit . As listed by (El-Sharkawy 2004), the most important are :

- High photosynthetic capacity
- Tight stomatal control that reduces moist loss
- Shedding and formation of small leaves
- Resumption of growth when water become available

The ability to extract deep soil water slowly which makes it tolerant of dry season and semi-arid environments.

d. Production systems

At the origin, cassava crop production by the Amerindians involved vegetative propagation by stem cuttings and typical aboriginal slash-and-burn cultivation, which were already in practice when the first Europeans sailed to the Amazon around 1540s. Since soils were rich in organic matter and free from compaction, zero or no tillage planting of cassava stalks became therefore the common practice (A. C. Allem 2001).

Actually, although cassava can be propagated with true botanical seeds, such practical prevail only for the crop improvement purposes and is almost not used for field production. The vegetative propagation using stem cuttings is still the most common method at a household level as well as for commercial production. such mode of propagation involves practical issues such as bulkiness, perishability and phyto-sanitation of cassava stem cuttings that results in a lack of high-quality planting material for smallholders (Nduwumuremyi et al. 2016).

Whereas a rapid multiplication method called the “mini-stem” method have been introduced to allow improving the amount of stem-cutting produced, the cassava is also challenged by its low multiplication rate. In fact, basically the cassava plant can produces only 10 stalks after growing one stalk during a year, in opposition for the maize which can produce 300 seeds from one maize seed in three months for example. This “mini-stem” method introduced by the IITA (Otoo 1994) can raise the multiplication rate up to 60-100 mini-stem cuttings per plant (Parmar et al 2017).

When root production is the objective, the optimal planting density is 10,000 plants per hectare (1 plant per square meter). Cassava stalks can be planted in different positions depending on the soil type and agro-ecological conditions. In fact, horizontal position is privileged in relatively dry areas (rainfall <1000 mm) while vertical planting is proposed in locations where there are sandy soils with plentiful rainfall. Planting vertically or inclined are believed to be associated with higher yields.

The depth of planting can vary from 5 to 15 cm depending on the soil type and climatic conditions. Although being a resilient crop, cassava cannot tolerate excessive soil moisture and weed competition in the early establishment phase.

I.2. Viral pandemics associated with cassava in Africa

I.2.1. General consideration

A virus is a non-cellular entity defined as "a set of one or more nucleic acid template molecules, either RNA or DNA, normally encased in a protective coat or coats of protein or lipoprotein, that is able to organize its own replication only within suitable host cells" (Hull 2009). The association between the nucleic acids and the protective protein capsid is termed the virus particle or virion. Plant viruses are obligate parasites that rely on host cellular machinery to reproduce by assembling pools of their structural components. Outside their hosts, virus particles cannot be active and are dormant (Hull 2009).

Following the introduction of cassava to Africa in the sixteenth century, it became infected by a set of viruses that were not known to infect the crop in its center of origin in South America. Until recently, about 15 virus species composed of diverse strains have been identified infecting cassava (Table I-2). Eleven of these species are responsible for the two most devastating diseases, namely CMD (nine species, collectively named cassava mosaic geminiviruses CMGs) and CBSD (two species, collectively named cassava brown streak ipomoviruses CBSIs). In the following sections, we will therefore consider these two main groups of viruses in details. The remaining virus species will only be touched on briefly.

I.2.2. The cassava mosaic disease (CMD)

a. History

The first description of CMD dates back to 1894 under the name *Krauselkrankheit* and has since been reported throughout Africa and in Madagascar, Zanzibar, Seychelles, India and Java. In East Africa, the disease was reported a century ago (E. F. Martin 1928). In West Africa, it was first recorded in the coastal areas of Nigeria, Sierra Leone, and Ghana in 1929 and has spread northward by 1945 (C. Fauquet, Fargette, and Munihor 1990).

CMD is considered to be of African origin since it is unknown in South America. However, similar symptoms are induced by an unrelated South American virus, called cassava common mosaic virus, which is a potexvirus having rod-shaped particles.

b. Aetiology, diversity, vectors and transmission modes

The viral aetiology of African cassava mosaic was first proposed by Storey in 1936 (Storey 1936), who demonstrated in Tanzania that the disease was transmissible and inferred that a virus was responsible. The causal agent of CMD was initially named as cassava latent virus but was subsequently characterized and renamed as African

cassava mosaic virus (ACMV, genus Begomovirus; family Geminiviridae) (Bock, Guthrie, and Figueireido 1981).

The confirmation that the transmission of pathogens causing CMD was done via whiteflies was established early in Central and Eastern Africa (Kufferath and Ghesquière 1932; Storey and Nichols 1938). Subsequent investigations on the characterization of this transmission involved the ACMV and suggested a persistent transmission manner where the virus circulated in the transstadial manner within the vector (Dubern 1994).

This circulative transmission manner involved a minimum of 3 hours for the vector to acquire the virus (acquisition access period, AAP) and a minimum of 10 minutes for inoculation (inoculation access period, IAP). A moderately long latent period of 3 hours minimum and a retention period of up to 9 days, were a logical consequence of the persistent and circulative transmission pathway of the CMGs. Storey showed that the whitefly vector could introduce virus into young leaves but not into mature leaves (Storey 1936; C. Fauquet, Fargette, and Munihor 1990).

Basically the long-term association between CMGs and whitefly vector suggested these viruses could be carried over long distances by the whitefly vectors. However, only circumstantial evidence obtained from the spread of the virus regionally suggested *B. tabaci* populations can carry CMGs over distances of up to 38 km in a year (Stansly et al. 2010).

Table I-2. Viruses infecting cassava in various regions where cassava is cultivated.

Virus name	Genus/Family	References	Diagnostics[b]	Distribution
(a) Latin America				
Cassava common mosaic virus (CsCMV)	*Alphaflexiviridae/Potexvirus*	Costa (1940), Silva, Kitajima, and Oliveira (1963), and Kitajima et al. (1965)	ELISA/RT-PCR	Colombia, Brazil (isolated cases from Africa, Asia)
Cassava vein mosaic virus (CsVMV)	*Caulimauviridae/Cavemovirus*	Costa (1940), de Kochko et al. (1998)	PCR	Brazil
Cassava virus X (CsVX)	*Alphaflexiviridae/Potexvirus*	Lennon, Aiton, and Harrison (1986)	ELISA/RT-PCR	Colombia
Cassava new alphaflexivirus (CsNAV)	*Alphaflexiviridae/Potexvirus*	Carvajal-Yepes et al. (2014)	RT-PCR	Colombia
Cassava frogskin-associated virus (CsFSaV)	*Reoviridae/Oryzavirus*	Calvert et al. (2008)	RT-PCR	Colombia, Brazil, Costa Rica, Argentina
Cassava polero-like virus (CsPLV)	*Luteoviridae/Polerovirus*	Carvajal-Yepes et al. (2014)	RT-PCR	Colombia, Costa Rica
Cassava torrado-like virus (CsTLV)	*Secoviridae/Torradovirus*	Carvajal-Yepes et al. (2014)	RT-PCR	Colombia, Argentina
Cassava symptomless virus (CsSLV)	*Rhabdoviridae/Nucleorhabdovirus*	Kitajima and Costa (1979)	NA	Brazil
Cassava Carribean mosaic virus (CsCaMV)	*Alphaflexiviride/Potexvirus*	Lennon et al. (1986)	NA	Colombia
Cassava Colombian symptomless virus (CsCSLV)	*Alphaflexiviride/Potexvirus*	Lennon et al. (1986)	NA	Colombia
Cassava American latent virus (CsALV)	*Secoviridae/Nepovirus*	Walter, Ladeveze, Etienne, and Fuchs (1989)	NA	Brazil, Guyana
(b) Africa				
Cassava mosaic disease				
African cassava mosaic virus (ACMV)	*Begomovirus/Geminiviridae*	Morris et al. (1990)	PCR and Real-time PCR	SSA[d]
African cassava mosaic Burkina Faso virus (ACMBFV)	*Begomovirus/Geminiviridae*	Tiendrébéogo et al. (2012)	PCR and Real-time PCR	Burkina Faso
Cassava mosaic Madagascar virus (CMMGV)	*Begomovirus/Geminiviridae*	Harimalala et al. (2012)	PCR and Real-time PCR	Madagascar
East African cassava mosaic Cameroon virus (EACMCV)	*Begomovirus/Geminiviridae*	Fondong et al. (2000)	PCR and Real-time PCR	SSA and Comoros

Virus name	Genus/Family	References	Diagnostics[b]	Distribution
(b) Africa				
Cassava mosaic disease				
East African cassava mosaic Kenya virus (EACMKV)	*Begomovirus/Geminiviridae*	Bull et al. (2006)	PCR and Real-time PCR	East Africa, Madagascar, Seychelles, Comoros
Eas African cassava mosaic Malawi virus (EACMMV)	*Begomovirus/Geminiviridae*	Zhou, Robinson, and Harrison (1998)	PCR and Real-time PCR	Malawi
East African cassava mosaic virus (EACMV)	*Begomovirus/Geminiviridae*	Bull et al. (2006)	PCR and Real-time PCR	SSA
East African cassava mosaic virus-Ugandan variant (EACMV-UG)	*Begomovirus/Geminiviridae*	Pita et al. (2001)	PCR and Real-time PCR	SSA
East African cassava mosaic Zanzibar virus (EACMZV)	*Begomovirus/Geminiviridae*	Bull et al. (2006)	PCR and Real-time PCR	Zanzibar, Madagascar
South African cassava mosaic virus (SACMV)	*Begomovirus/Geminiviridae*	Berrie, Rybicki, and Rey (2001)	PCR and Real-time PCR	South Africa, Madagascar, Zimbabwe
Cassava brown streak disease				
Cassava brown streak virus (CBSV)	*Ipomovirus/Potyviridae*	Winter et al. (2010)	RT-PCR, Real-Time RT-PCR, RT-LAMP, and ELISA	Kenya, Mozambique, Uganda, Tanzania, Malawi, Rwanda, Burundi and DR Congo.
Ugandan cassava brown streak virus (UCBSV)	*Ipomovirus/Potyviridae*	Mbanzibwa et al. (2009)	RT-PCR, Real-Time RT-PCR, RT-LAMP, and ELISA	Kenya, Mozambique, Uganda, Tanzania, Malawi, Rwanda, Burundi and DR Congo.
(c) South Asia and minor viruses				
Cassava mosaic disease				
Indian cassava mosaic virus (ICMV)	*Begomovirus/Geminiviridae*	Malathi, Nair, and Shantha (1985), Hong, Robinson, and Harrison (1993)	PCR	Southern India and Sri Lanka
Sri-Lankan cassava mosaic virus (SLCMV)	*Begomovirus/Geminiviridae*	Saunders et al. (2002)	PCR	Southern India and Sri Lanka
Cassava viruses not linked to with any major disease [c]				
Cassava virus C (CsVC) (syn. Cassava Q virus)	*Ourmiavirus/Unassigned*	Calvert and Thresh (2002), Rastgou et al. (2009)	NA[e]	Ivory Coast
Cassava green mottle virus (CsGMV)	*Nepovirus/Comoviridae*	Lennon, Aiton, and Harrison (1987)	NA	Australia and Pacific Islands, Solomon Islands

Virus name	Genus/Family	References	Diagnostics[b]	Distribution
Cassava viruses not linked to with any major disease [c]				
Cassava ivorian bacilliform virus (CIBV)	*Anulavirus/Bromoviridae*	Fargette, Roberts, and Harrison (1991), Scott et al. (2014)	NA	Ivory Coast
Cassava Kumi viruses A and B	Uncharacterized	Calvert and Thresh (2002)	NA	Kumi district of Uganda

In Africa, nine different Begomovirus species have been identified in association with CMD in different regions of Africa (Table I-2) (Alabi, Kumar, and Naidu 2011; J. P. Legg and Fauquet 2004; Tiendrébéogo et al. 2012; Harimalala et al. 2012): ACMV, East African cassava mosaic virus (EACMV), East African cassava mosaic Malawi virus (EACMMV), South African cassava mosaic virus (SACMV), East African cassava mosaic Cameroon virus (EACMCV), East African cassava mosaic Zanzibar virus (EACMZV) and East African cassava mosaic Kenya virus (EACMKV), Cassava mosaic Madagascar virus (CMMGV) and African cassava mosaic Burkina Faso virus (ACMBFV). In addition, several strains of these viruses have been identified, with the most virulent being the East African cassava mosaic virus-Uganda (EACMV-UG) also known as the "Uganda variant" (Zhou et al. 1997). This strain was the first proof of recombination between two distinct begomovirus species, the EACMV and ACMV (Zhou et al. 1997).

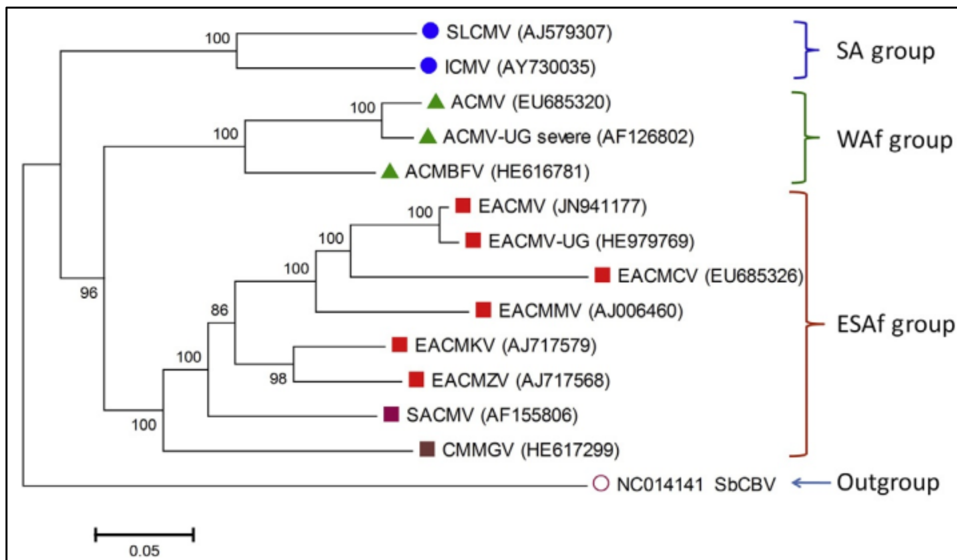


Figure I-8. Phylogenetic relationships among various representative species of the cassava mosaic geminivirus. Sequences were aligned using the ClustalW algorithm, and the tree was constructed by the Neighbor-Joining method.
The tree was rooted using Soybean chlorotic blotch virus (SbCBV) as an outgroup. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown at the branch nodes. The NCBI GenBank accession numbers of 14 DNA-A sequences are indicated in parenthesis. The tree was adapted from (James P. Legg et al. 2015).

Coinfection with more than one species or strains is a common feature in the aetiology of CMD in Africa. Mixed infections occur often and involve ACMV and one of the EACMV-like CMG species. This favors the advent of synergistic interactions between the species, resulting in an increased overall virus titer leading to more severe symptoms (Fondong et al. 2000; Harrison et al. 1997). Rapid regionwide spread of EACMV-UG and ACMV, frequently in mixed infections, was

an important feature of the African severe CMD pandemic that occurred in the 1988's (Harrison et al. 1997; Thresh et al. 1997).

CMGs infects mainly euphorbiaceous hosts including seven *Manihot* species and *Jatropha* multifida L. Two other species, *Hewittia sublobata* (L.F.) Kuntze (Convolvulaceae) and *Laportea aestuans* (L.) Chew (Urticaceae), are suspected to be natural hosts for ACMV in Kenya and West Africa, but the virus has not been transmitted from them back to cassava (C. Fauquet, Fargette, and Munihor 1990). Additionally, in Nigeria, Alabi et al., (2008) reported the presence of ACMV, EACMV and EACMCV isolates in leguminous plant species such as *Senna occidentalis* L. (Fabaceae), *Leucaena leucocephala* (Lam) De witt (Fabaceae), *Glycine max* L. (Fabaceae), *Combretum confertum* (Benth) M.A. Lawson (Combretaceae) *and Manihot glaziovii* Müll.arg (Euphorbiaceae) (Alabi et al. 2008).

It was also demonstrated that under high disease pressure conditions in Yangambi/D.R. Congo, alternate host plants such as weeds or plants intercroped within cassava fields may also be infected through whiteflies transmission (Monde et al. 2010).

The mechanical transmission of ACMV was first reported in 1978.

c. Genome organization of CMD-associated viruses

The genome of CMBs consist of two genomic components, referred to as DNA A and DNA B, of about 2.8 kb each, encapsidated in a 30 x 20-nm twinned icosahedral particles. The DNA A component can replicate autonomously and produce virions, but requires DNA B for systemic infection (C. M. Fauquet et al. 2005). DNA carries six ORFs, each encoding a specific protein whereas DNA-B has only two ORFs (Figure I-9).

The complementary-sense DNA synthesis to produce dsDNA occurs in the nucleus and depends solely on host factors because all CMBs, and to a wide extent, all geminiviruses, do not encode a DNA polymerase. Double-stranded circular molecules serve as templates for both transcription and replication. The replication for the synthesis of the virus ssDNA is initiated by cleavage of the virion-sense strand by Rep in a conserved 5'-TAATATTAC-3' sequence within the long intergenic region (LIR), and employs two strategies: the rolling-circle and the recombination-dependent mechanisms. The transcription is bi-directional and the resulting coding regions in both strands diverge from the LIR (Zerbini et al. 2017).

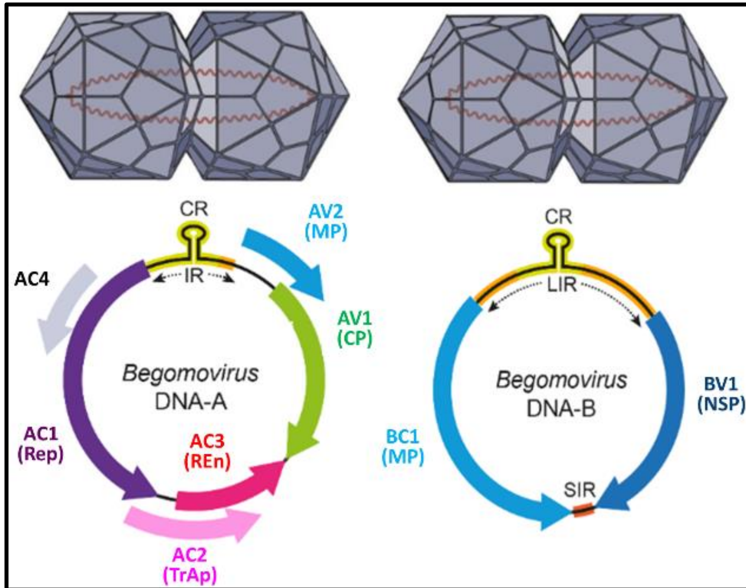


Figure I-9. Typical genomic organization of cassava mosaic begomoviruses.
The ORFs are denoted as being encoded on the virion-sense (V) or complementary-sense (C) strand, preceded by component designation (A or B) if bipartite. Corresponding protein products are indicated. The "common region" that is shared between the two genomic components is shown as light green boxes within the intergenic region (IR) / long intergenic region (LIR). The position of the stem-loop containing the conserved 5′-TAATATTAC-3′ sequence located in the intergenic region (IR) or LIR is shown. CP, coat protein; Rep, replication-associated protein; TrAP, transcriptional activator protein; REn, replication enhancer protein; MP, movement protein; NSP, nuclear shuttle protein; SIR, short intergenic region. The figure was adapted from (Fiallo-Olivé et al. 2021; Vanderschuren and Rey 2017; Hulo et al. 2011).

In general, Begomovirus genomes show an extreme plasticity due to frequent recombination and pseudo recombination events. The first report that documented the interspecies recombination between the ACMV and the EACMV and that lead to the emergence of the EACMV-UG recombinant, a more severe variant in Uganda, dates back to 1997’s (Zhou et al. 1997). This recombinant had two regions of its CP related to EACMV (99 and 98% identical respectively), and one region related to the ACMV (99% identical).

It was shown that these recombination events that involved A components of both species impacted their discrimination during detection by ELISA since most of the monoclonal antibodies (MAbs) used could not distinguish EACMV-UG from EACMV (Zhou et al. 1997). This provided the understanding of the severe form of the mosaic disease that has spread since 1988 from southwards to westwards of Uganda at a speed of approximately 20 Km per year (Zhou et al. 1997). Additionally, later in 2001, it was demonstrated that the genetic variation resulted not only from recombination, but also pseudo recombination between the A and B components of strains of the same as well as different species (EACMV-UG2 DNA-A and EACMV-UG3 DNA-B; ACMV and EACMV respectively) (Pita et al. 2001)

d. Epidemiology

Serious yield losses due to CMD were first observed on mainland East Africa in the 1920s (E. W. Hall 1928). Recorded epidemics of CMD later occurred in the 1930s, 1940s and from 1990s (Jameson 1964; Otim-Nape and ~et al 2000).

Initial observations conducted in the early years of the disease characterization suggested that the rate of spread of CMD was much greater at hot and moist low-altitude locations than it was at higher elevations where temperatures were cooler and the whiteflies less abundant (Storey 1936; 1938). Similar epidemiological characteristics were reported in Eastern D.R. Congo by (E. Bisimwa, Walungululu, and Bragard 2012). A most extensive review of key milestones achieved in developing knowledge of the CMD epidemiology is presented by (James P. Legg et al. 2015; Vanderschuren and Rey 2017)

The important resistance mechanisms that have been reported to influence the epidemiology of CMD are related to the recovery and reversion mechanisms observed in infected cassava plants. Recovery mechanism consisted in the disappearance of symptoms during crop growth whereas the reversion mechanism consisted into the production of virus-free cuttings by infected parent plants (Fargette, Thresh, and Otim-Nape 1994).

One of the most important epidemiological fact that has characterized the spread of CMD in Africa arose when an epidemic of unusually severe symptoms expanded its geographical range in Uganda in 1988's, to affect many parts of the country as well as other countries in the Great Lakes. Through various studies has helped explaining the causes of this rapid spread and the severity of symptoms: the mixed infection between ACMV and EACMV, the synergism between cassava geminiviruses and the massive increase in abundance of the *B. tabaci* whitefly (Fondong et al. 2000; Harrison et al. 1997; Ogbe et al. 2006).

CMGs species isolated from mosaic-affected cassava collected from different parts of Africa and from India are partitioned into three groups on the basis of serology and DNA hybridization (Harrison, Robinson, and Thomas 1986). Group A includes strains from Angola, Ivory coast, Nigeria, Congo, and Western Kenya and also defective strains that do not produce virus particles; group B strains are from coastal Kenya, Madagascar, and Malawi; and group C strains are from India and Sri Lanka. The distribution of these three groups is hypothesized to have resulted from the different routes of cassava dissemination (Figure I-10). As cassava was introduced into the various countries that now constitute its range, it may have become infected with different geminivirus variants that were endemic in the three geographic regions

Figure I-10. Principal routes of dissemination of cassava from South America to Africa and India, and distribution of cassava cultivation in Africa. The hatched lines mark the northern and southern limits of cassava production, and the dates are the years African cassava mosaic disease was first described in each country (C. Fauquet, Fargette, and Munihor 1990).

Figure I-11. Various morphologies of CMD symptoms on leaves

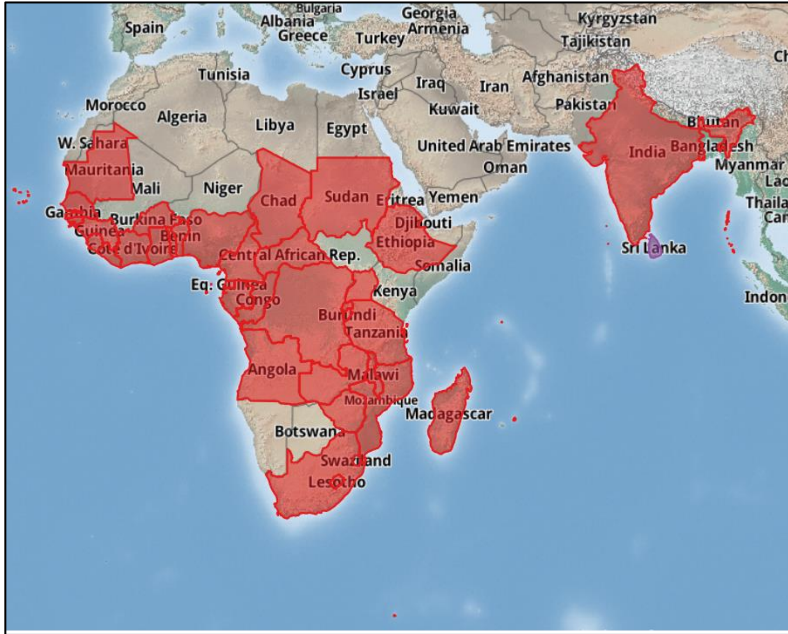


Figure I-12. Distribution of cassava mosaic disease in Africa and Asia. Adapted from (CABI 2023)

e. Economic importance

The symptoms of CMD in cassava are usually conspicuous. Much of the evidence on the occurrence, incidence and spread of this disease is based on visual observations. However, symptoms are sometimes indistinct, especially in dry conditions when vegetative growth is restricted, or when plants develop symptoms of mineral deficiency, or are severely attacked by cassava green mite (*Mononychellus* tanajoa) or cassava mealybug (*Phenacoccus manihoti*) (CABI 2023).

The most visible symptoms of the disease is the expression of the characteristic leaf mosaic , and young plants are more severely affected than old ones. Symptoms range from barely perceptible mosaic to stunting of the plant and extreme reduction of the leaf blades (Figure I-11). The severity of symptoms varies with the cultivar and increases with plant age until about 60 days after planting. Thereafter, symptoms are more moderate or lesser or do not develop, depending on the cultivar, climatic conditions, and season. Symptoms on fully expanded leaves do not change, however (C. Fauquet, Fargette, and Munihor 1990).

Results from several trials to assess the effect of CMD on cassava have shown that the percentage loss could be a function of various factors at play, among which the most important included the susceptibility of the variety, the stage of crop growth at which infection occurred, the severity of the virus or virus mixture causing the infection, and the abiotic environmental conditions (Fargette, Fauquet, and Thouvenel 1988; C. Fauquet, Fargette, and Munihor 1990; Seif 1982).

Yield losses ranging from 20 to 95% have been reported for individual cultivars in different countries (Thresh, Fargette, and Otim-Nape 1994). From an experimental estimation using a moderately susceptible variety, Fargette et al., 1988 suggested all plants would be infected and an average yield losses of 37% would be attained. If extended to the whole Africa and considering the actual production estimates of 192 million tons (FAO 2021a), such calculations would suggests losses equivalent to 71 million tons actually. However, losses are very variable and range from insignificant to almost total. The following generalizations have been mentioned elsewhere:

- Plants grown from infected cuttings sustain a greater yield loss than those of the same variety infected later by whiteflies, and plants infected at a late stage of crop growth are virtually unaffected (CABI 2023).
- There are big varietal differences in response to infection (Fargette, Fauquet, and Thouvenel 1988);
- Infected plants of varieties designated as resistant may sustain substantial yield losses (Seif 1982)
- There is a positive relation between the extent and severity of symptoms and yield loss (Thresh et al. 1997; Seif 1982);
- Competition and compensation effects are likely to be important and infected plants surrounded by uninfected ones are more seriously affected than those in groups (Otim-Nape, Thresh, and Shaw 1997);
- Effect on yield are influenced by crop duration (CABI 2023);
- From experience with other virus-host interactions, it is likely that soil fertility, seasonal factors, crop spacing and other cropping practices, virus strain, weed control and other pests/diseases influence the effects of CMD on growth and yield, although they have not been sufficiently investigated (CABI 2023).

I.2.3. The cassava brown streak disease (CBSD)

CBSD represents the most damageable threat to cassava productivity in sub-Saharan Africa. The disease was first reported during the early days of the Amani cassava research program in northeastern Tanzania in 1930s (Storey 1936; 1938). At this earliest record period, the disease was noted to occur widely in the coastal region of East Africa and was considered to have a viral etiology from the outset.

a. Historical background

Research milestones accomplished on cassava brown streak disease since its first report were previously summarized through various extensive reviews (Legg and Thresh 2003; K. R. Tomlinson et al. 2018a; Vanderschuren and Rey 2017) (Table I-3).

The historical facts related to CBSD started from the 1930s with the first report of the disease by the Amani research station team in Tanzania (Storey 1936). At that time, the disease was symptomatically described and distinguished from CMD. Since

no pathogen could be observed whereas the disease was graft-transmitted, it was assumed that the causal agent could be a virus.

Table I-3. Summary of CBSD history before 2000's (Legg and Thresh 2003; K. R. Tomlinson et al. 2018a; Vanderschuren and Rey 2017)

1936	First reported: Tanzania. Distinguished from CMD, Graft transmissible, assumed to be viral.
1939	(Storey 1939)
1940	Reported in Zanzibar
1950	Reported from Uganda (assumed to have been introduced from Amani research station)
1940s-1950s	Resistance breeding (Tanzania)
1950	Reported in Southern Malawi, detailed descriptions of symptoms, effect of temperature/altitude recognized.
1959	Successful sap transmissions to/from cassava and herbaceous hosts (Lister 1959)
1970s-1980s	Virus studied in Kenya (further sap transmissions) and UK (first electron microscopy)
1990	"rediscovered" in Tanzania
1992	"rediscovered" in Malawi
1994	"rediscovered" in Uganda
1993/1994	First countrywide survey (Tanzania)
1995-2000	Epidemiology studies in Tanzania, Mozambique (Nampula/Zambezia) and coastal Kenya
1996	Reported in Zambia

After the disease being also reported in Zanzibar, Malawi (along the lakeshore), and Uganda (it was thought to have been introduced in cutting material from Amani in 1934), subsequent efforts to study the disease through symptoms description and elucidation of the temperature/altitude effects were done. It was also noted that the disease was endemic throughout Tanzania, occurring right up to the borders with Kenya to the North and Mozambique to the south. Following this spread, Major campaigns were launched in Uganda and Tanzania to eradicate the disease (Legg and Thresh 2003).

Later on, Lister (1959) transmitted successfully the viral agents of CBSD to a range of herbaceous hosts through mechanical inoculation and then back-transmit them to cassava with sap from leaves of infected herbaceous indicator plants.

After the Amani research program, the next effort on CBSD started later through the ODA/Kenya Plant Virology Project as well as in the UK. The main outcome from these investigations consisted in the realization of the first electron microscopic study that revealed the presence of filamentous particles *c.*650 nm long (Bock 1994). However, it has been considered that these particles could possibly be carlaviruses, although various limitations complicated the final conclusions. This consideration has

been later supported by the fact that a positive serological reaction was obtained between the extracts from infected leaves and antibodies raised against the carlavirus *cowpea mild mottle virus* (Brunt 1996). However, the presence of pin-wheel inclusions structures, at the time only observed in association with potyvirus infections, was also noted (Lennon, Aiton, and Harrison 1987).

Globally, the 5 decades of sporadic search into CBSD resulted in the development of a substantial body of information on the disease. Unfortunately, several important aspects of the disease remained non elucidated: (i) the etiology was not fully determined, (ii) the causal viruses were not characterized, (iii) no vector was known, and (iv) there was little information on temporal or spatial patterns of the disease spread (Legg and Thresh 2003). Most of this important missing information were addressed during the following decades and are summarized below.

b. Research milestones from 2000's

In this section we updated these works by adding the main research accomplishments realized up to October 30th 2022. In this line, we conducted a literature review of various published references available through SCOPUS from 2000's up to October 30th 2022. We used the following research formula for this purpose: ***(TITLE-ABS-KEY (Cassava) AND TITLE-ABS-KEY ("brown streak virus") OR TITLE-ABS-KEY ("brown streak disease") AND PUBYEAR > 2000***.

In total we could retrieve 206 documents, most of which have been published in the 2016-2020 period (107 documents) (Figure I-13A). The main thematic covered by these publications deals with the aspects related to genetics and breeding for resistance (67 documents); disease epidemiology, surveillance and risk assessments (42 documents), disease diagnosis and detection (31), molecular biology and virus evolution; as well as insect science (28 documents respectively) (Figure I-13B).

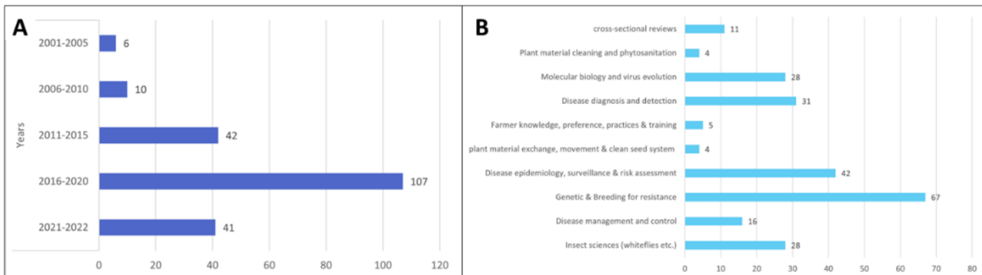


Figure I-13. Literature analysis on (**A**) the number of scientific publications as well as (**B**) the main thematic covered by the 206 publications on cassava brown streak disease (or virus) retrieved through SCOPUS between 2000's and 2022.

Through literature exploitation identified some of the research events on CBSD during the aforementioned period as summarized on the figure I-14 below.

Figure I-14. Key events in cassava brown streak disease (CBSD) research history (2000-2022). CP, coat protein; RT-PCR, reverse-transcription polymerase chain reaction, CBSD, cassava brown streak disease; CBSV, cassava brown streak virus; UCBSV, Ugandan cassava brown streak virus; RNAi, RNA interference; HTS, high throughput sequencing; CMD, cassava mosaic disease; NWG, new whole genome.

Further details and comments regarding this summary were compiled together with other information from the period before 2000's to generate brief descriptions presented in subsections c-*g* below.

c. Etiology, structure, properties, and diversity of the CBSD-associated viruses

One of the key milestone achieved in the early 2000's was that from the University of Bristol's (UK) team who adopted a molecular approach to resolve

the etiology of the CBSD-causing viruses by partially purifying the virus from CBSD-affected cassava material collected from Tanzania (W. a Monger et al. 2001). They successfully generated a 1114 bp long partial sequence of the coat protein and subsequently designed primers which were used for RT-PCR detection in extracts from CBSD-affected leaves. This first ever-made CBSV sequence finally resolved the identity of the virus to belong to the *Ipomovirus* genus since the closest identical sequence identified (43.2%) was the one from the *sweet potato mild mottle virus*, a unique member of the genus *Ipomovirus* at that time.

Based on the variation in the coat protein nucleotide sequences between different isolates collected from Mozambique and Tanzania (Monger et al. 2001b), and on variable symptoms observed in different cassava cultivars (R. J. Hillocks, Raya, and Thresh 1999; W. A. Monger, Seal, Cotton, et al. 2001) and experimental hosts such as *Nicotiana benthamiana* (W. A. Monger, Seal, Isaac, et al. 2001), the occurrence of different strains of CBSD-causing viruses was suspected (J. Legg, Kumar, and Kanju 2015)

The first complete characterization, and by the way, the first diversity analyze of CBSVs came in 2009 with very interesting information such as:

(i) A genome characterized by a positive-sense, single-stranded RNA genome consisting in 9,069 nt long, with a polyprotein expression strategy;
(ii) The genome is different from other *Potyvidae* members in that it lacks the HC-Pro but contained a single P1 serine protease involved in the RNA silencing suppression in the 5'-proximal part of the genome, and more importantly, contained an exceptional structure related to a Maf/HAM1-like sequences in the 3'-proximal part of the genome, recombined between the NIb and the CP (Figure I-15); and
(iii) two "genetically different and geographically separated populations" of CBSVs were involved in the aetiology of the disease: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) (genus *Ipomovirus*, family *Potyviridae*) (Mbanzibwa, Tian, Tugume, et al. 2009; Mbanzibwa, Tian, Mukasa, et al. 2009).

It is only in 2010 that the team from (Winter et al. 2010a) finally completed the aetiology of CBSD through the demonstration of the Koch's postulate and clarified by the way the existence of "at least two distinct species" considering differences observed both in the biological behavior and in genomic and protein sequences of CBSVs: the first species gathering isolates from Kenya, Uganda, Malawi, and the Tanzanian part of the lake Victoria basin in the North-western Tanzania (UCBSV) whilst the second species included isolates from coastal lowland regions of Mozambique and Tanzania (CBSV) and shared only 70% nucleotide sequence identities with the first species group (Winter et al. 2010a). If these different species could be geographically separated at that time, their geographic repartition were mixed further on.

Among other biological characteristics, it was importantly noted that the CBSV was more devastating than UCBSV since it could infect all cassava lines, susceptible or tolerant (Winter et al. 2010a). Later in 2016 with the advent of the HTS technology and the raise in the number of available complete genomes, (Titus Alicai et al. 2016a) demonstrated that CBSV had the capacities of outsmarting the cassava immune system by having a faster rate of evolution compared to UCBSV, and by having more predominant nonsynonymous substitutions than synonymous substitutions .

From an evolutionary perspective, the question of the origins of the CBSVs, as is the case for CMGs, remains until now not clarified. In fact, it was hypothesized that at its arrival in East Africa in the eighteenth century, the cassava must have been infected from a wild host somewhere in the region. However, no epidemiologically significant alternative host plants, other than the wild relative *Manihot glaziovii*, has been shown to be infected by these viruses (Mbanzibwa et al. 2011).

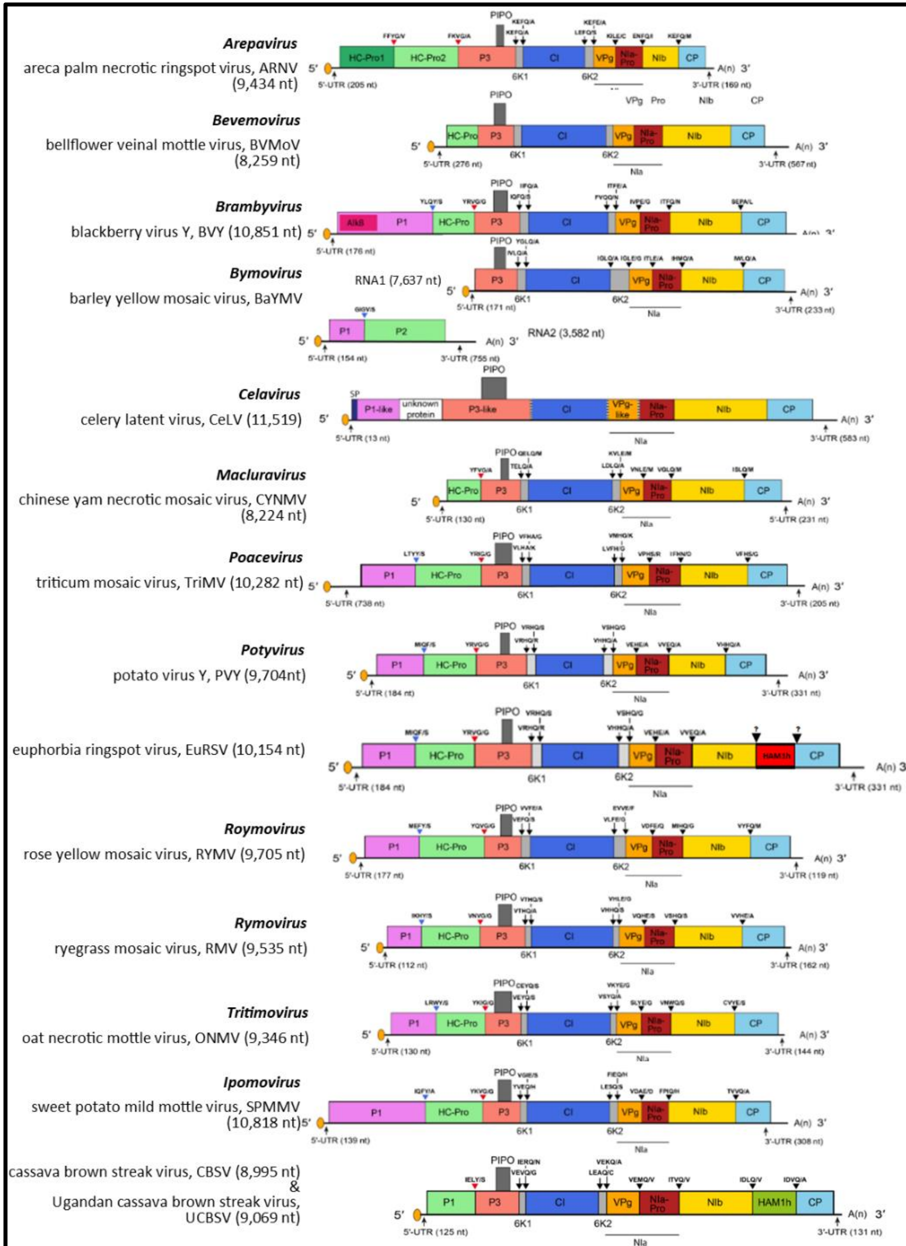


Figure I-15**.** Genome organization of representative species for various genus in the family *Potyviridae*. The figure highlight the unique features of the CBSV and UCBSV genome organization. This figure is a personal adaptation based on genome organization from the ICTV website [https://ictv.global/report/chapter/potyviridae/potyviridae] accessed on December 12th, 2022).

The positive sense single-stranded RNA genome of CBSV consists of size varying between 8,972 bp to 9,036 bp (Genbank accession numbers HG965221.1 and

MF975780.1 respectively) , whereas that of UCBSV is slightly longer, consisting of sizes varying between 9,069 bp and 9,097 bp (Genbank accession numbers FJ039520.1 and KX753357.1 respectively). Their genomes express a polyprotein of about 2902 amino acid residues encoding 11 proteins. From the 5' to the 3', these are: the P1 (protein1 protease), which is a protein with a serine proteolytic activity responsible for cleavage at typically Tyr/Phe-Ser and is also a suppressor of RNA silencing because of the lack of the P1b protein which plays that role in other potyvirus genomes; a P3 (protein 3); PIPO (pretty interesting potyviridae ORF), embedded in the P3 ORF and expressed as the *trans*-frame P3N-PIPO protein; 6K1 (6-kDa protein 1); CI (cylindrical inclusion); 6K2 (6-kDa protein 2); VPg (virus protein genome-linked), the genome-linked virus protein covalently attached to the 5-terminal nucleotide represented by an aval; NIa-Pro (the second of two functional regions of the NIa, nuclear inclusion a protease, susceptible to being proteolytically separated from the VPg region), a cysteine-like proteolytic activity responsible for cleavage at Gln/Glu-(Ser/Gly/Ala); NIb (nuclear inclusion b), RNA-directed RNA-polymerase; Maf/HAM1 (a putative pyrophosphatase); and the coat protein. The protein-encoding region is flanked at the 5' and 3' ends each by an untranslated region, and the 3' end finalizes with a *circa* 30 nucleotides poly-A tail (Mbanzibwa, Tian, Mukasa, et al. 2009; Winter et al. 2010b).

The arrangement of ORFS on the genome of CBSVs is particular compared to other potyvirid (Figure I-15): the ORFs encoding for proteins 6K1, CI, 6K2, VPg, NIa-Pro and NIb are similar to members of the family. However, at the 5' end, the P1 ORF is functionally equivalent to the P1b protein of other ipomoviruses (the *squash vein yellowing virus* and *Cucumber vein yellowing virus*) but absent in the genomes of CBSVs. Additionally, the Maf/HAM1 protein in the CBSV and UCBSV genomes is unique and previously identified only in *Euphorbia ringspot virus*, a member of the genus potyvirus, and in cassava *cassava torrado-like virus*, a member of the family *Secoviridae* (Mbanzibwa, Tian, Mukasa, et al. 2009; Leiva et al. 2022a). These recent years, this protein has been investigated to understand its role in the infection mechanisms and as a virulence determinant for CBSVs.

The Maf/HAM1 protein is a group of ubiquitous proteins expressed by all living organisms and involved in the reduction of mutagenic pools of non-canonical nucleotides (A. M. James et al. 2021). The relevance and role of this protein for viruses, and especially for CBSVs, is a matter of investigations since recent years. Investigations from (K. R. Tomlinson et al. 2019) suggested that in *N. benthamiana*, the Ham1 protein of CBSVs has a function of hydrolyzing mutagenic nucleotides as well as a determinant of necrosis. However, these functions have been shown to be distinct when it came to consider cassava, and more extensively plants from the *Euphorbiaceae* family, as the host.

In fact, since the 3' extremity of the ITPase from cassava had a particularity of being flanked with a nuclear localization signal, it has been demonstrated that cassava expressed its ITPase in the nucleus, leaving the cytoplasm a highly challenging environment for replication of RNA viruses due to a raise in the amount of non-canonical nucleotides. Some authors suggested this over-accumulation of non-canonical nucleotides might had a key role in the antiviral defense of the plant (Valli

et al. 2022). To overcome this situation, RNA viruses infecting the cassava were predicted to need this Maf/HAM1h protein (A. M. James et al. 2021).

Recently, it has been validated that, in fact, CBSVs required the ITPase activity for infecting the cassava, but not in the model organism *N. benthamiana* and that a fraction of the ITPase remained covalently bound to the viral RdRp to optimize viral fitness (Valli et al. 2022). The most important findings about the role of the HAM1h protein of CBSVs in the infection of cassava were brought by (Sheat and Winter 2022) who experimentally demonstrated that apart from being required for cassava infection, it was involved in the determination of pathogenicity, replication and movement, but also impacted on resistance and resistance breaking.

The phylogeny aspect of CBSVs have been commonly addressed using demarcation cutoffs suggested by the ICTV as follows: the thresholds for species demarcation using nucleotide identity values for the individual coding regions range from <58% for the P1 coding region to <74-78% for other regions. For the coat protein, the optimal species demarcation criterion is <76-77% nucleotide and <80% amino acid identity (M. J. Adams, Antoniw, and Fauquet 2005). These criteria have early suggested the existence of two viral species: the CBSV and the UCBSV.

However, with the advent of the HTS, the availability of much more complete genomes and the possibility of supercomputing have allowed a deep investigation and understanding of evolutionary mechanisms driving the CBSVs genomes. In that perspective, findings from (Titus Alicai et al. 2016a; Mbewe et al. 2017; Amisse et al. 2019) supported that among the 10 genes encoded by the CBSVs genome, eight could support the existence of two genetically distinct clades among CBSV, distinguishable from UCBSV; whereas only two genes could support the existence of the known CBSV and UCBSV species (CP and HAM1h genes) (Figure 16). Additional biological data are awaited to support the existence that putative divergent species within the CBSV species.

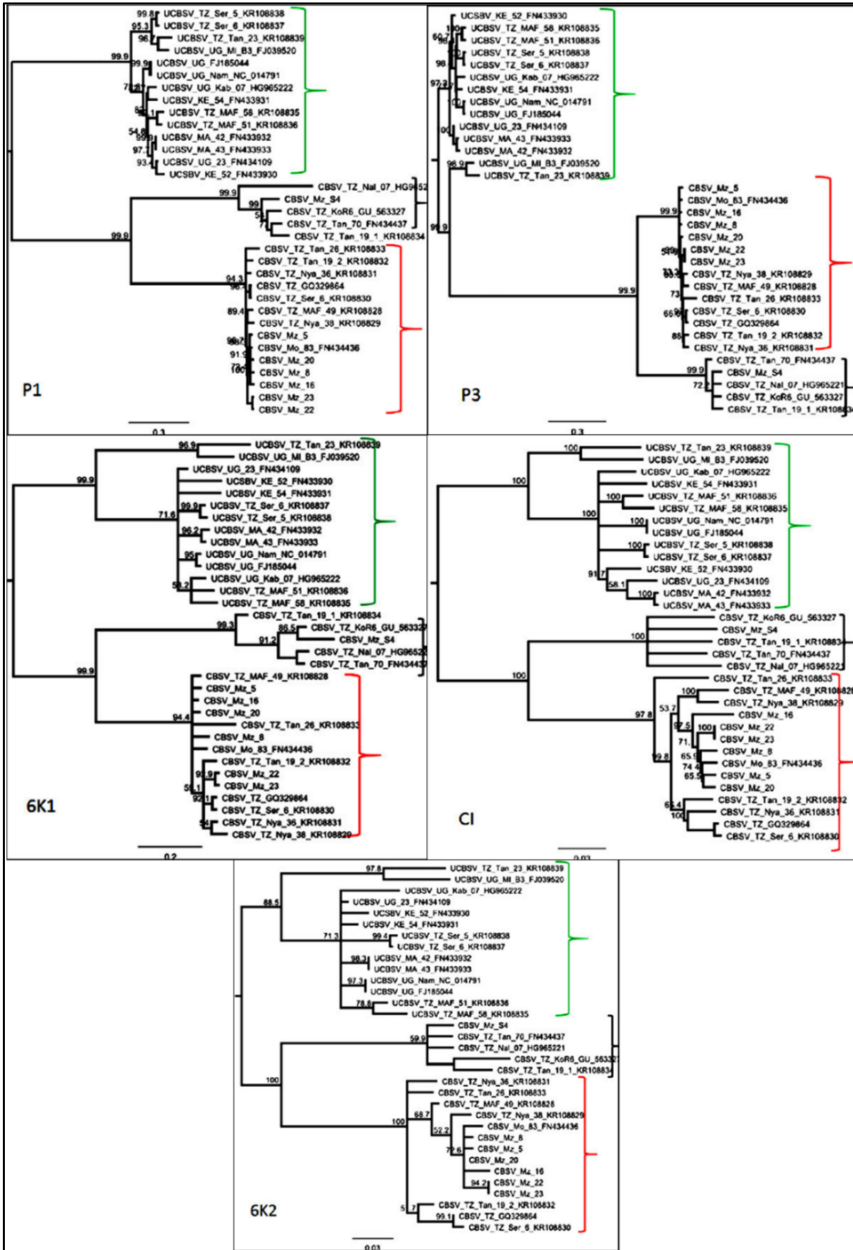


Figure I-16A. Phylogenetic trees based on the first five genes encoded by the CBSV and UCBSV and showing how these genes place these viruses into three clades. The trees were generated using best-fit model preselected in jModel Test. The number at each branch represents the bootstrap value (1000 replicates) and the scale bar represents nucleotide substitution per site. The figures were adapted from (Amisse et al. 2019).

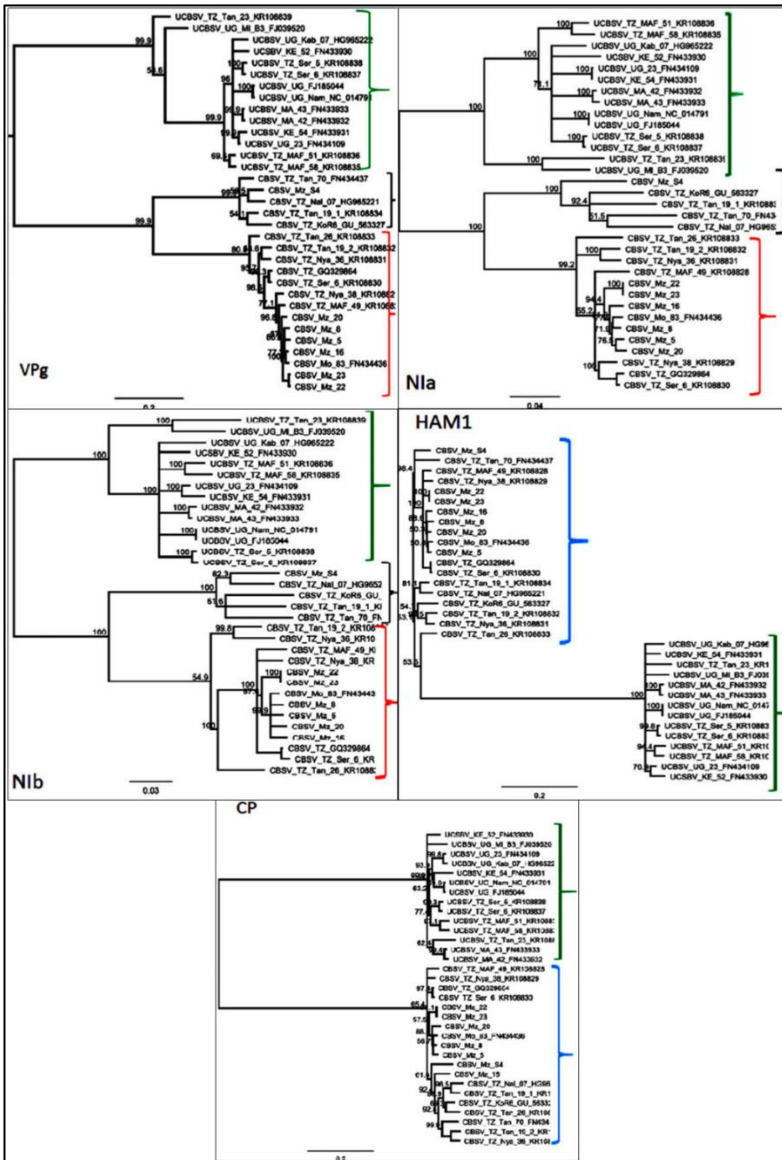


Figure I-16B. Phylogenetic trees based on the five additional genes encoded by the CBSV and UCBSV and showing how only the HAM and CP genes place the isolates into two distinct clades whereas three distinct clades could be seen using the 3 other genes as shown in figure 16A. The trees were generated using best-fit model preselected in jModel Test. The number at each branch represents the bootstrap value (1000 replicates) and the scale bar represents nucleotide substitution per site. The figures were adapted from (Amisse et al. 2019).

d. Symptoms and diagnostic methods

The first detailed description of the CBSD symptoms was provided by (Nichols 1950) from the Amani station in Tanzania. Already at this first description, a key facet of symptoms highlighted was their extreme variation, not only from one variety to another (depending on the susceptibility of the cultivar), but also from season to season (depending on environmental conditions), and the age of the plant. Although symptoms could be observed on the leaves, stems, fruits, and tuberous roots; it was noted that in sensitive varieties, symptoms may be present on all plant parts while for more tolerant varieties, there may be only one symptom, commonly on leaves.

On leaves, two types of symptoms have been recognized by Nichols (1950): a first type consisting in a feathery chlorosis associated with secondary and spreading to tertiary veins as infection progresses and coalescing to produce a more general blotchy chlorosis. A second symptom type, more common, consisting in a general blotchy chlorotic mottle lacking clear association with veins (Figure I-17). In both cases, these symptoms are more prominent on lower leaves, and the yellowing can be readily distinguished from senescence by the presence of patches of green that occur in symptomatic leaves of CBSD-affected plants. Distinctly to CMD, no distortion of the leaf lamina and no size reduction happens (Legg and Thresh 2003).

On the green portions of stems of sensitive varieties, brown necrotic streaks occurs and gave the name of the disease although sometimes not consistent considering that stem symptoms are less frequent than either root or leaf symptoms (Figure I-17). In sensitive varieties, shoot dieback may occur during periods of cool dry weather with the upper portion of the stem becoming necrotic and then drying out. The die-back symptoms were particularly severe when plants were generated from infected cuttings and in extreme cases, the entire plant was destroyed (Legg and Thresh 2003).

The development of dry, corky, and necrotic lesions in the root tissue is by far the most important symptom produced by the disease, additionally to malformations and constrictions that can be also observed (Figure I-17). These necrotic lesions enlarge and change color from yellow to sepia dark brown, becoming therefore more intense as the plant matures, particularly “beyond” the physiological maturity at about 12 months after plantation (Nichols 1950; Hillocks et al. 2001). Even though radial constrictions or longitudinal fissures may occasionally be caused, the dry root rot often occur in the absence of any obvious external signs (J. Legg, Kumar, and Kanju 2015). The time of onset of root symptoms and their overall severity as well as the relative importance of foliar versus root symptoms has been demonstrated to be highly variable between cultivars. Additionally, severely affected roots were shown to be susceptible to secondary infections by soil-inhabiting fungal or bacterial pathogens, leading ultimately the plant to the death (J. Legg, Kumar, and Kanju 2015).

All symptoms types are more apparent during periods of cooler weather and in more mature plants. Additionally, the link between foliar and root symptoms has been clarified as being uncertain for most varieties and situations (Hillocks et al. 1996).

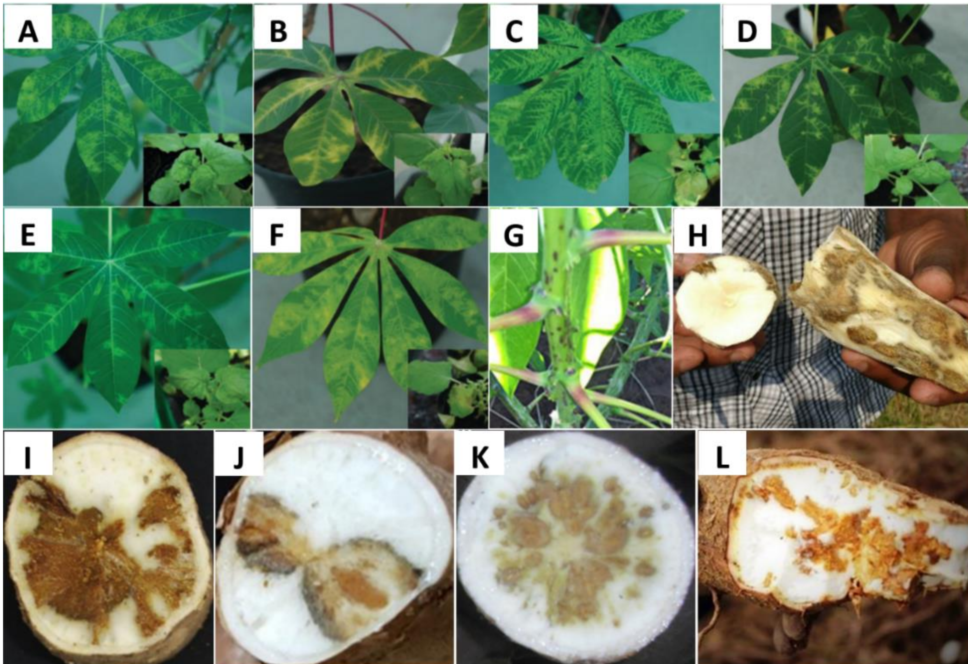


Figure I-17. Symptoms of CBSD in leaves (A-F), stem (G) and roots (H-L) of cassava (for leaves, symptoms in *N. benthamiana* are also inserted).
(A) Natural infection of CBSV Tan_70 in cassava collected from Mlandizi, Tanzania; (B) cassava land race Mbongo, D.R. Congo, infected mechanically with CBSV Ke_125; (C) TMS 96/0529 infected mechanically with CBSV Mo_81; (D) TMS 96/0160 infected mechanically with CBSV Ug_23; (E) TMS 96/0304 infected mechanically with CBSV Ma_43; (F) TME 4 infected mechanically with CBSV Mo_83 (Winter et al. 2010b). (G) Stem of CBSD-affected cassava plant expressing brown streak necrosis (Munganyinka et al. 2018); (H) Cassava root showing necrosis restricted on outer surface of the tuber in Uvira (D.R. Congo) (This study) ; (I) Necrotic root rot of cultivar Cho5_203 at Bunda in Tanzania (Rudolph R. Shirima et al. 2020); (J) Marked root necrosis of a cassava germplasm line from South-America to infections with CBSV-Mo83; (K) Roots of CBSD-affected plant showing uncommon circular yellowish-brown necrosis in Rwanda (Munganyinka et al. 2018); (L) Root necrosis of a cassava plant infected by the cassava root necrosis disease (CRND) in Congo central, D.R. Congo (Zeyimo et al. 2020).

This seasonality in the CBSD symptom expression combined with the fact that they are much less obvious, ephemeral and of variable nature make the visual inspection of plant not reliable and complicate the diagnosis of infection during field-based assessment of the disease. This necessitate more skills and attention on behalf of agricultural workers or researcher during field-based assessment of the disease (J. Legg, Kumar, and Kanju 2015). Various symptoms guides have been published in order to aid in the field diagnosis of symptoms (CABI, 2014; CRS, 2014).

e. Transmission modes, vectors, host range, and epidemiology

After its first report in the northeast Tanzania (Amani station), the CBSD was reported to be endemic in all East African coastal lowland from Kenya to the border

between Tanzania and Mozambique (Rj Hillocks and Jennings 2003; R. J. Hillocks, Raya, and Thresh 1999; R. J. Hillocks et al. 2002). Early reports mentioned Uganda, the southern Malawi (in the inland and shores of lake Malawi), and the North-Eastern region of Kenya (Nichols 1950; Rj Hillocks and Jennings 2003; Patil et al. 2015) as first regions where the disease has spread (Table I-4_A).

Therefore, it was known during that period that CBSD symptoms can be expressed at altitudes greater than 1,000 masl when infected cuttings have been planted. This is what occurred in Uganda in 1934 (in Bukalasa and Serere) when infected material was taken from Tanzania, fortunately the disease was controlled by destroying all plants showing symptoms (Jameson 1964).

Table I-4_A**.** Four main events associated with the epidemiology of CBSD in Africa

	Main epidemiological events	Periods	Details		References
1	First report and subsequent early reports	1936-1950	First reported in Northeast Tanzania. Endemic in coastal lowland areas of East Africa	Restricted to coastal East Africa	Storey, 1936
2	Abrupt shift from lowland to Mild altitude at 1000 masl	2004	Re-emergence in further inland areas around Lake Victoria		Alicai et al. 2007
3	Shift from low incidence to high incidence	2008	In Uganda, Tanzania and Kenya from c. 12% in 2008 to c. 27% in 2011		(Legg et al., 2011; Mware et al., 2009; Ntawuruhunga and Legg, 2007)
4	Shift from restricted geographical range to wide geographical range	2001-2019	Confirmed reports from Burundi, Rwanda, eastern D.R. Congo , South Sudan, Mayotte Island and Comoros	spread through many East and Central African countries	(Bigirimana et al., 2011), (FAO, 2011), (Mulimbi et al., 2012), (Roux-Cuvelier et al., 2014), (Scussel et al. 2019)

This restriction to lowland regions abruptly changed in the early 2000s, as new outbreaks re-emerged in mild-altitude areas above 1000 masl in central Uganda, western Kenya, and north-western Tanzania (Lake Victoria zone) (T Alicai et al. 2007). It was not clear what was the drivers of that outbreak but several hypotheses have been proposed including:

- The possibility that the disease might have been present for many years at a low level and was not noticed,
- The emergence of a new and more aggressive strain of CBSV: Since the diversity studies have identified two different virus species associated with the CBSD (the CBSV and UCBSV), and even though they have been initially detected in geographically separated regions (Tanzania and Uganda, respectively), the virulence studies later demonstrated that the CBSD outbreak at altitude higher than 1000 masl was not driven by the appearance of new virulent strains in the Great Lakes region (Mbanzibwa, Tian, Tugume, et al. 2009; Winter et al. 2010a)
- Introduction of infected material from the coastal areas of Kenya, Tanzania, Malawi, or Mozambique where CBSD is endemic: Before this outbreak, there was a general lack of scientific interest in the CBSD due, among other reasons, to the fact that the disease has a restricted localization to low altitude along the coastal eastern Africa. A greater priority was dedicated to the mitigation of the devastating impacts of CMD on food security and its prevalence across all cassava-growing areas of Africa. To help control the disease, CMD-resistant cultivars were distributed to areas severely affected (Legg and thresh, 2000). However, it was later shown that these cultivars were susceptible to one or both species of viruses causing the CBSD, and that the transportation of infected material to areas in which CBSD was previously absent has enabled the disease to spread over long distances from independent hot spots (Winter et al. 2010a; J. P. Legg et al. 2011).
- The coincidence of the CBSD re-emergence with a massive increases in the whitefly *Bemisia tabaci* populations (superabundant populations) observed in these periods and that could have aided the spread of the disease (J. P. Legg et al. 2011; T Alicai et al. 2007). This observation is relevant considering the fact that the spread of the recombinant variant, EACMV-UG, has been associated with severe CMD pandemic that spread from Uganda to neighboring countries between 1988 and 2007, and was linked to unusually high populations of *B. tabaci* (Deng et al. 1997; Gibson, Legg, and Otim-Nape 1996; J. P. Legg 1999; Otim-Nape, Alicai, and Thresh 2001). Considering the transmission mode of CBSVs that have been later demonstrated to be semi-persistent, the *B. tabaci* have been demonstrated to be involved in the transmission of CBSVs on short distances inside or on nearby fields.

Other driving forces might have been at stake in the dissemination of CBSVs as hypothesized by (Vanderschuren and Rey 2017) including the contribution of CBSVs infection to the attraction of *B. tabaci* by the cassava host and notably the role of minimum temperature on CBSD incidence (Vanderschuren and Rey 2017).

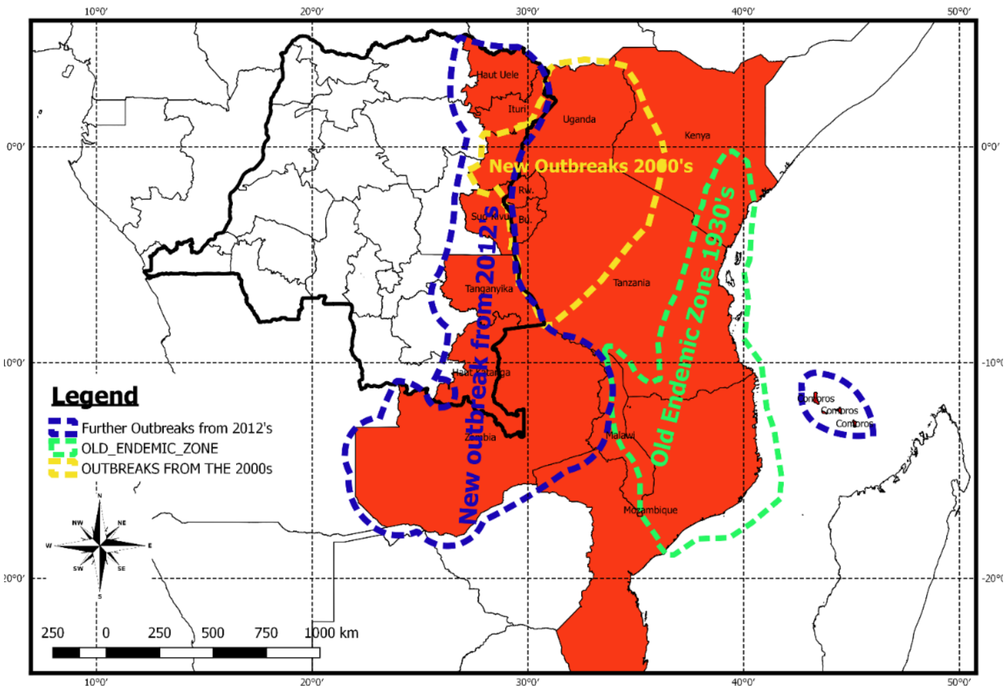


Figure I-18. Status of CBSD outbreak in African countries and Comoros until 2022.

Since the initial suspicions that *B. tabaci* could be the vector of viruses causing the CBSD (Storey 1935), experimental evidences waited until 2005 to validate the role of the whitefly in the transmission of CBSVs (Maruthi et al. 2005). In fact, that work suggested a low transmission efficiency (22%), a minimum of 5 and 30 min for AAP and IAP, respectively, and a maximum of 24h retention time. However, recent work indicated a transmission efficiency of up 60% using 20-25 viruliferous whiteflies given an AAP and IAP of 24 and 48 hours per plant respectively. Additionally, findings from this study suggested no difference between the transmission efficiency of CBSV and UCBSV, and importantly that the disease could spread in the field up to a distance of 17 m in a cropping season. These evidences definitely supported a semipersistent transmission by *B. tabaci* on short distances in the field, and importantly, the role of planting material movement in the spread of the disease over longer distances (Maruthi et al. 2017).

More than validating the semipersistent mode of transmission, Winter (2022) has recently demonstrated that *Bemisia tabaci* was very inefficient in transmitting U/CBSV and that consequently, high populations of whiteflies were necessary for the transmission to occur in laboratory conditions whereas in field condition this could only be attained during seasonal upsurge of whiteflies populations. These findings allowed to show that the whitefly vector was not a bottleneck for U/CBSV maintenance and more importantly that U/CBSV populations have shaped towards vegetative propagation and would not probably need the whitefly transmission factors to sustain their life (Winter, 2022).

In addition to cassava, the CBSD has been reported to naturally occur on *M. glaziovii* and the "tree" cassava, which is presumed to be a hybrid between *M. glaziovii* and *M. esculenta* (Storey 1936; Bock 1994) (Storey, 1936; Bock, 1994). However, following mechanical inoculation, various herbaceous species have been also experimentally infected. For instance, chlorotic or necrotic lesions were produced in *N. debneyi*, systemic and irregular chlorotic vein banding in *N. benthamiana* and *N. clevelandii*. Systemic chlorotic stippling with vein banding was produced in *D. stramonium* and *D. ferox*. Additionally, sap inoculation has been also used to reproduce symptoms in *Solanum nigrum* and *Chenopodium quinoa* (Bock, 1994).

f. Economic importance

- The various types of symptoms resulting from the CBSD infection in cassava causes economical losses in several ways (J. Legg, Kumar, and Kanju 2015):
- Reduction of root yield through impairment of the growth of the plant: a study conducted in Tanzania (Hillocks et al. 2001) suggested losses of up to 70% in the most sensitive cultivars due to the impairment of growth, although a more typical level of loss was considered to be 30% (Rj Hillocks and Jennings 2003).
- Spoilage of roots through the effects of severe root rot: loss due to root spoilage have been estimated to account for up to 24% in sensitive cultivars, and circa 17% on average (Hillocks et al. 2001)
- Premature harvesting by farmers in order to avoid spoilage of roots through rot: the reduced yield due to early harvesting has been recognized for many years as an important indirect consequence CBSD infection (Childs 1957). However its economic cost has not been quantified yet.
- Loss of planting material arising from premature harvesting: basically, cassava is commonly planted on a 12-month cycle. Therefore, stems from early harvested plants (7–10 months after planting) may be unusable as planting material by the time the normal planting season comes around.
- Increased labor costs associated with peeling partially rotten roots.

As highlighted above, it is clear that the main economically significant effect of the disease happens through the root necrosis symptoms as recognized since the earliest research on CBSD (Nichols 1950). The cited economic consequences are exacerbated by the fact that usually, farmers are not aware that their crop is affected until they cut open the tuberous root during harvest operations to realize it is late (J. Legg, Kumar, and Kanju 2015).

Various small-scale, and sometimes country-wide assessments of the impact of CBSD have been realized in different affected regions and allowed the estimation of yield losses. For example, it was demonstrated in Tanzania that sensitive varieties may lose up to 70% of fresh root yield due principally to the effects of die-back that can be compounded by the effects of necrosis on root quality (from 2 to 29 % additional losses), which prevent harvested roots from being marketed or encourage premature harvesting to avoid the most severe damage. This suggested that in most extreme situations, losses ranged from 72 % to 99 % (Hillocks et al. 2001).

A regional survey on the economic impact of CBSD conducted from East and Central Africa in 2009 (Burundi, eastern DRC, Kenya, Rwanda, Tanzania and Uganda) estimated losses of up to 1,600,000 tons (Manyong et al., 2012). Even tough it is not easier to obtain a truly accurate estimation of the economic damage caused by CBSD; an overall loss of *circa* US$750 million a year has been estimated across Kenya, Tanzania, Uganda and Malawi (R Hillocks and Maruthi 2015). Although CBSD has attracted an increasing amount of interest since the development of new outbreaks in the Great Lakes region, these loss figures, when compared to those of CMD (more than US$1.9 billion annually);

Additionally, it was also observed that the relationship between leaf and root symptoms was ill-defined since some varieties or plants with clear leaf symptoms failed to show root symptoms while others, not expressing leaf symptoms, produced root symptoms (Legg and Thresh 2003).

Symptoms of root necrosis and yield loss increase as the age of the crop increases and losses are particularly acute for local varieties in which root necrosis begins to increase from six months after planting, encouraging farmers to harvest prematurely (Nichols 1950; Hillocks et al. 2001)

I.2.4. Control and management of CMD and CBSD

Most of general components that are essential for the effective management of plant viral diseases are equally applicable to cassava. These components involve: (i) the detection and monitoring, (ii) the prevention of infection, and (iii) the control of infection. The control strategies that have been applied for CMD and CBSD have been extensively reviewed elsewhere (J.P. Legg et al. 2006; James P. Legg et al. 2015; Thresh and Cooter 2005; Vanderschuren and Rey 2017; E. B. Bisimwa 2012). Below I summarize and discuss some of the milestones regarding the breeding for resistant varieties.

i. Management of CMD

Two major approaches have been prioritized in the management of CMD in Africa. They include the maintenance of CMD-free crops through phytosanitation and the use of virus-resistant varieties (Thresh and Otim-Nape 1994). The phytosanitation involves the removal of diseased plants from the field (or roguing) to prevent further spread and/or the selection of CMD-free stems at the end of each growing cycle in order to plant new fields with "clean" material (Legg and Thresh 2003). In addition to virus-free planting material, the management has also included quarantine measures, crop hygiene , changes in cropping practices (e.g. intercropping) and the use of pesticides to control vectors (E. B. Bisimwa 2012; Thresh 2003)

The use of resistant or tolerant varieties has obvious advantages in seeking to decrease virus-induced losses. A number of research programs with a priority of identifying CMD resistance sources have been prioritized in Africa. The breeding activities against the cassava mosaic viruses have stared at the Amani program in 1930(Jennings 1957). The first resistant varieties, characterized by a resistance of CMD1 type, were obtained by crossing cassava with its wild-relative *Manihot*

glaziovii Muell.-Arg and subsequently through backcrosses to combine the resistance with good root yield (Jennings 1957; Hillocks and Jennings 2003). This CMD1 resistance is of a recovery type, recessive and polygenic. The following resistance type, the CMD2, was later identified within a West African cassava landrace TME 3 (Akano et al. 2002) and provided extreme resistance against all known viruses causing CMD (Ariyo et al. 2005; 2006; Atiri et al. 2004) and is currently the predominant cassava mosaic virus resistance source. CMD2 is monogenic, has a dominant inheritance, and is very stable and consistent in conferring virus resistance (Rabbi et al. 2014; Codjia et al. 2022; Okogbenin et al. 2012). Another type of resistance, the CMD3, described in the elite cultivar TMS 97/2205 that was selected from crosses of TMS 30572 (CMD1-resistant type) and TME 6 (CMD2-resistant type) was later characterized (Sheat et al. 2019). This resistance type was a quantitative trait locus and similarly confers very high resistance against the viruses causing CMD.

Multiple donor-funded research initiatives over the last 45 years took place to release, promote and disseminate planting material resistant against the CMD since the disease emerged as an epidemy in Africa.

From 1986, the International fund for Agriculture Development (IFAD) funding in Nigeria was used to multiply and disseminate five CMD-resistant and high-yielding varieties from the TMS line that were released by IITA in 1977. These varieties remained popular up to 35 years after their release: TMS 50395, 63397, 30555, 4(2)1425, and 30572.

Additionally, from 1991 to 1994, the IITA introduced in Malawi cassava varieties tolerant to CMD and CBSD, having high yield and drought-tolerant characteristics. Varieties such as CH92/082, TME 6 and BA95/070 were multiplied and disseminated through funding from the Agency for International Development (USAID) (Benesi et al. 2010).

As a response to the CMD-UG epidemic in Uganda, various initiatives have been initiated by several actors to promote multiplication and distribution of CMD-resistant TMS varieties.

The crop crisis control project (C3P), a regional intervention implemented in six countries from the Great Lakes Regions and aimed at responding to CMD and banana wilt disease, multiplied and disseminated four CMD-resistant cassava varieties in the Eastern D.R. Congo: Sukisa (MM96/1666), Sawasawa (MM96/3920), Mayombe (MM967752) and Liyayi (MM96/0287) (C. Legg 2008). Before that period, four CMD-resistant cassava varieties, resulting from a research collaboration between the IITA and the National Agricultural Research from D.R. Congo (INERA-Nvuasi station), were released. These varieties are locally-known as *Butamu, Disanka, M'vuazi* and *Nsasi* and are also widely cultivated.

The near-immunity of the CMD2 and CMD3 resistance locus in cassava conferred broad-spectrum resistance against all known cassava mosaic viruses. For that reason, currently the CMD is no more a critical threat to cassava cultivation in Africa.

ii. Management of CBSD

The expansion of the CBSD epidemic across the Great Lakes region of East and Central Africa has triggered the need of rapidly developing and implementing effective control strategies. Several important projects were initiated following CBSD re-emergence, which aimed at developing research, extension and policy capacity in the countries affected. Key targets have been to breed or genetically engineer resistant cultivars, provide certified virus-clean planting material and improve viral surveillance and diagnosis (Legg et al., 2014).

Research on breeding resistance initiated at the Amani station in the 1930's considered both CMD and CBSD. Following the success of the interspecific crossing technique using *M. glaziovii* to introduce CMD resistance to cultivated cassava, a similar approach was explored for CBSD, although in this case, using the other wild relative *M. malanobasis* Muell. Arg. (Legg and Thresh 2003).

However, the control of CBSD posed a serious challenge compared to that of the CMD: "all" African cassava varieties, improved against CMD or not, were susceptible to the two viruses causing the disease, and while there were less sensitive varieties, "all" cassava eventually succumbed to the disease (Sheat et al. 2019).

A number of interventions, reviewed in (Vanderschuren and Rey 2017), aiming at mitigating the damages due to CBSD through research and development projects have been undertaken in various African countries mostly as regional interventions. Early at the re-emergence of the disease, most of these interventions privileged the multiplication and distribution of elite cassava varieties, initially bread for resistance against CMD.

Since all of those varieties were susceptible or tolerant to CBSD and in a context of significant knowledge gap on the CBSD epidemiology, those interventions unintentionally exacerbated the risk of virus spread through susceptible germplasm exchanges between and within countries. The following statements, taken from a report resulting from a field visit of a cassava multiplication site intended for production of cuttings for dissemination to farmers in 2010 in Burundi (CRS 2012), illustrates the extent of the facts:

> "During the field visit, clear symptoms of Cassava Brown Streak Disease (CBSD) were observed in all sites and on all the clones. All of the samples collected from the visit confirmed CBSD during lab testing. The clones multiplied by FAO (MH96/5280 and MH96/0287) showed considerably more severe symptoms than that multiplied by GLCI (Mh96/7204). In fact, it was agreed in 2006 that, based on the experience with MH96/5280 in Kenya where this variety seriously succumbed to CBSD, this clone would no longer be multiplied. It was continued to be multiplied by the Burundi Institute of Agronomic Sciences (ISABU) and FAO, perhaps because there had been no lab evidence of CBSD in Burundi until this visit."

Breeding cassava is notoriously difficult because of the high heterozygosity and a challenging cross-pollination process (Ceballos et al., 2012). Additionally, breeding is also complicated by cultivars showing variation in CBSD resistance depending on the environment, which necessitates the testing of cultivars in different agro-ecological zones to ensure that their resistance is stable (Tumuhimbise et al., 2014).

Although tolerant cultivars develop reduced symptoms, they remain susceptible to CBSD viruses and thereby their adoption does not remove viral inoculum from the field. This is why considerable efforts have been made to screen and breed cassava cultivars for CBSD resistance, which are able to restrict CBSD viral replication and/or movement (K. R. Tomlinson et al. 2018b).

The lack of CBSD resistance in African cassava varieties prompted researchers to new sources of virus resistance within the diversity of South American cassava germplasm accessions. From a screening of 238 accessions using bud grafting against various CBSVs-infected cassava plants, two breeding lines known as DSC167 (COL40) and DSC118 (COL2182) were identified as having a broad resistance spectrum against CBSV and UCBSV (Sheat et al. 2019). Because the CMD resistance was essentially absent from those lines, introgression of CMD resistance was needed to complement CBSV resistance (Sheat, Zhang, and Winter 2022).

Crosses of African and South American source into hybrids with dual resistance were initiated in 2019 and hybrids with dual resistance U/CBSV and CMD provided evidence that, as with CMD2, resistance against U/CBSV is a dominant trait. Research are still underway and best results are expected from the combination of the best parental lines, COL40 and C33, combining immunity against all CBSV and CMD2-type resistance acting against all known CMVs respectively. The promising hybrid candidates have been identified and are being actually hardened for adaptation and virus resistance performance evaluation in field under high-pressure disease context (Sheat and Winter 2023).

The benefits of selecting CBSD-free stems when replanting have been clearly demonstrated (Hillocks et al. 2001). Additionally, It has been estimated that the net value of the release of CBSD-resistant cultivars will be US$436 million for Western Kenya and US$790 million for Uganda over a 35-year period starting in 2025 (Taylor et *al.,* 2016).

The lack of cultivars highly resistant to CBSV makes the existence of a clean seed system critical for the effective management of CBSD. Clean cassava seed systems are non-existent in most Eastern Africa countries where CBSD is a problem. Previous interventions for the mitigation of CBSD using routine interventions of distributing tolerant cultivars or cleaning infected material has shown its limits as illustrated by the Great Lakes Cassava Initiative project and discussed in subsection *1.6.3* below (CRS 2012).

I.2.5. Other viruses

Indian cassava mosaic virus (ICMV) was the first CMG to be recorded from South Asia (Malathi et al., 1985), followed by Sri Lankan cassava mosaic virus (SLCMV) several years later (Saunders et al. 2002). Although SLCMV was initially reported from Sri Lanka, it was subsequently shown to occur also in southern India, together with ICMV (Anitha, Makeshkumar, and Edison 2011; Patil et al. 2005).

Cassava green mottle virus, Cassava virus C and Cassava Kumi virus A & B have been reported infecting cassava but they were not well characterized, and their economic significance is not known. Cassava Ivorian bacilliform virus (CIBV) (genus Anulavirus; family Bromoviridae) has recently had its full genome characterized (Scott et al., 2014) but has only been recorded from Ivory Coast and has no known effect on cassava.

Viruses reported infecting cassava in the Latin America are diverse and belong to virus families *Alphaflexiviridae*, *Reoviridae*, *Rhabdoviridae*, *Secoviridae* and *Luteoviridae* among the RNA viruses, and to the family *Caulimoviridae* among the DNA viruses. Unlike for Africa, no geminivirid or potyvirid species have been reported associated with diseases in cassava in the Americas (Calvert, Cuervo, & Lozano, 2012).

The Cassava frogskin-associated virus (CsFSaV), associated with the CFSD, is the most economically important disease of cassava in Latin America followed by the cassava torrado-like virus (CsTLV). CFSD is characterized by the failure of the storage roots to accumulate starch and so affected roots develop a rough epidermis resembling the wart-like skin of toads or alligators, hence its name in Spanish (cuero de sapo) and in Portuguese (jacare). Only recently, it was found that complex virus infections are associated with root symptoms of CFSD, that several novel viruses are part of that complex, and that neither CsFSaV nor phytoplasma can induce CFSD symptoms or the associated leaf symptoms in single infections (Alvarez et al. 2009; Carvajal-Yepes et al. 2014).

Cassava vein mosaic virus (CsVMV), first reported from Brazil in 1940 (Costa, 1940), is the type species of the genus Cavemovirus (family Caulimoviridae) and is the only pararetrovirus of cassava so far only been reported from Brazil. The 35S promoter sequence derived from its genome is extensively used as a constitutive promoter in the genetic engineering of plants (Verdaguer et al. 1996).

A diverse range of mechanically transmitted potexviruses have been detected in symptomless cassava, including Cassava virus X (CsVX) and Cassava new alphaflexivirus (CsNAV) (Carvajal-Yepes et al. 2014; Lennon, Aiton, and Harrison 1987). However, because these viruses do not cause symptoms, it is difficult to determine their distribution or to evaluate their importance. In fact, at least four other potexviruses infecting cassava have been reported since the 1960s (Table I-2).

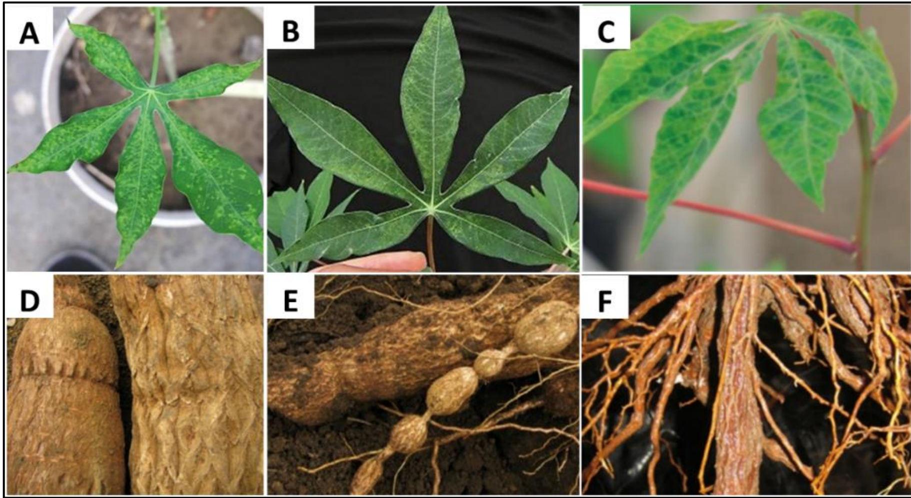


Figure I-19**.** Various virus-like symptoms on cassava leaves and roots observed out of Africa under greenhouse conditions:

(A) leaf mottle and ringspot symptoms induced by CsCMV, **(B)** leaf symptoms induced by CsTLV in the variety 'Secundina', **(C)** cassava variety "Secundina" with mixed infection by CsCMV, CsFSaV and CsTLV **(D)** CFSD-root symptoms showing two different severity of lip-shaped fissures, or **(E)** root with constriction zones and lip-shaped fissures, and **(F)** severely affected root of cassava variety 'Brasilera' with no accumulation of starch. Images adapted from (Carvajal-Yepes et al. 2014; Lennon, Aiton, and Harrison 1987; James P Legg et al. 2015)

I.3. Diagnosis of viral plant diseases

I.3.1. General consideration

Plant viruses are a significant agricultural problem and cause major losses through diseases that they induce in plants. In any disease situation, it is important to know : which virus is causing the problem, where it comes from, and how it spreads, before developing control measures (Hull 2009).

The scientific investigation on plant diseases known to be caused by viruses did not begin until the late nineteenth century. However, there are much earlier written and pictorial records of such diseases:

The earliest known written record describing what was almost certainly a virus disease is a poem in Japanese written by the Empress Koken in 752 AD and translated by T. Inouye. The plant, identified as *Eupatorium lindleyanum*, has been found to be susceptible to TLCV, which causes a yellowing disease (Saunders et al. 2003).

There are three basic situations in which the techniques for recognizing and identifying a virus are needed :

- Diagnosis of a viral infection in the field to determine whether it is caused by a known or unknown virus
- Characterization of a known virus, for instance by purification or by various manipulations.
- Detection of a known virus: usually in an epidemiological survey or for quarantine or certification purpose

I.3.2. Plant virus diagnosis

A question often asked to plant pathologist is : “what is wrong with my plant? ”; followed by: “what can I do to get rid of the problem?”. It may be too late to help the specific plant when the question is asked, but proper diagnosis may be extremely important in preventing the problem on other plants or in preventing the problem in the future (Riley, Williamson, and Maloy 2002).

A precise identification of the causal agent is the first step, and the most important for the management of viral diseases. As other plant diseases, accurate symptoms description are necessary to describe the disease. This is why plant pathologists often rely on symptoms for the identification of a disease problem. However, in many situations symptoms description alone are inadequate for causal agent identification since several unrelated viruses can cause similar symptoms and a single species can cause different symptoms on the same host or on different host species (Kumar, P. L. and Legg 2009).

Control measures depend on proper identification of diseases and of the causal agents. Therefore, diagnosis is one of the most important aspects of a plant pathologist's training. Without proper identification of the disease and the disease-causing agent, disease control measures can be a waste of time and money and can lead to further plant losses (Riley, Williamson, and Maloy 2002).

The terms “diagnosis” and “detection” are often used interchangeably. However, they do not always have the same meaning. According to the EPPO standards, the diagnosis is the process of detection and identification of a pest i.e. the interpretation of the result of a diagnostic process (EPPO 2018).

The diagnosis can be seen as a process during which the characteristics of a disease are investigated and described. It involves the steps ranging from clinical observation (careful examination to determine underlying cause of the disease with unknown or unclear etiology), the description of the disease etiology (causal agent: detection, isolation, characterization), as well as investigations of the disease causation.

Therefore, the detection is a step of the diagnosis that consists in the elucidation of the presence or absence of the virus, even in the absence of visible symptoms. For example, cassava mosaic disease is “diagnosed” as due to cassava mosaic Geminiviruses (CMGs) whereas methods such as PCR are employed to “detect” CMGs in diseased plants. Detection of a virus in a diseased plant not necessarily is a proof that its cause the disease.

The diagnostic process of a disease include (but is not limited to) the following steps (Bos 1976):

1. Observation of the disease in the field to determine affected plant species and cultivars, disease incidence and distribution within field (random, clustering, peripheral, uniform).

2. Recording of symptoms and comparing with the literature for any similar descriptions on the same host in various locations.

3. Studying infectivity and transmission tests by grafting; mechanical inoculation; transmission through vectors (insects, mites, nematodes or fungi).

4. Inoculation (using plant sap, by grafting or vector) to a range of test plants and back inoculate to a parallel range of test plants to check possible multiple infections and to determine host range and symptoms. This step allow to select systemically infected host for virus propagation in the frame of purification purpose and virus assays.

5. Determination of the persistence of infectivity in sap extracts (dilution end point, thermal inactivation point, stability and retention of infectivity upon storage at various temperatures and lengths of time).

6. Examination of leaf dip preparations under electron microscope to detect any virus particles.

7. Isolation of the virus and purify thereafter to determine the physico-chemical properties

8. Study of the cytopathology for virus inclusions and cytological changes in affected cells.

9. Development of a diagnostic test for virus detection

10. Assessment of the virus serological relationships using antiserum (less practiced), and inter-relationships from nucleotide sequence information to determine the virus genomic properties, expression strategies and virus taxonomic status

11. Fulfillment of the Koch's postulates, especially using purified virus or isolated virus cultures especially if purified virus preparation loses infectivity.

Depending on the virus type, previous knowledge on virus or knowledge gained from during experimentation, laboratory facilities and expertise, the order of steps described can be changed or some few steps can be passed over.

Throughout time, some steps involved in this framework have been prioritized, scientifically discussed and improved or sometimes changed. This is the case for instance the necessity of the fulfillment of the Koch's postulate for plant viruses, especially in the era of High Throughput Sequencing technological revolution (A. Fox 2020; Massart et al. 2017).

I.3.3 Plant virus identification

The properties elucidated during the course of isolation, purification and diagnosis of the viral disease determines the virus relationships with the previously characterized viruses and forms a basis to identify it as a new species / an isolate of a known virus species / a new strain of a known virus species, and to place it into an appropriate taxonomic group following the actual ICTV plant virus classification.

Through time, the criteria used by the ICTV to classify new viruses have been evolving (Figure I-19).

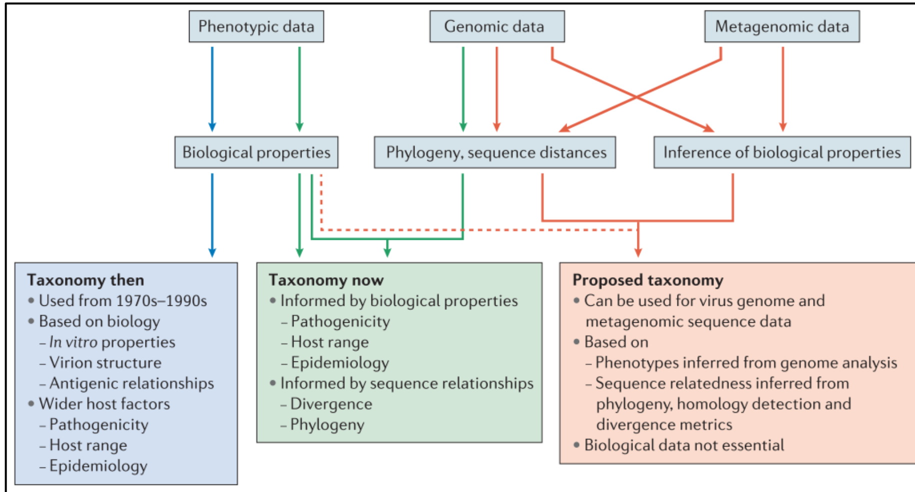


Figure I-20. Evolution of virus classification pipelines in the past from 1970s to 1990s (blue arrows), currently (green arrows) and a classification proposed for enabling both metagenomic sequence data and conventionally derived virus sequences to be classified (red arrows). Figure adapted from (Simmonds et al. 2017)

In the past, commonly used characters for virus identification emphasized biological characters, physico-chemical properties, morphological characters and inter-relationships (Kumar, P. L. and Legg 2009):

- Biological characters : They consist in transmission characters, host range, symptomatology, cross-protection and in vitro properties.

- Physico-chemical properties consists of: they consist in the number of virus components, number and molecular weight of the structural proteins (coat protein and nucleoprotein), type of nucleic acid, number and molecular weight of the virus genome, sedimentation coefficient, particle buoyant density

- Morphological properties: size and shape, special futures such as lipid membranes

- Inter-relationships: They consists in serology-based (serological relationships or relationships by western immune-blotting) and nucleic acid-based (genome organization and expression, amino acid composition, percent nucleic acid homology)

These criteria evolved later to importantly accommodate information related to the genome or gene sequences (divergence and phylogeny).

However, the advent of the HTS technology and the development of metagenomic study approaches have exponentially raised the number of discovered viruses to the extent that the overwhelming majority of these newly described viruses could not be directly linked to biological attributes (hosts, symptoms, etc.). In this context, it has been proposed that fundamental changes

had to be brought to the way new virus species had to be approved by the ICTV so as to include also sequences obtained through metagenomic studies (Simmonds et al. 2017).

The concern that was formulated against integrating sequence assemblies resulting from metagenomic studies in the classification of viruses was that the lack of biological properties would result into a taxonomy of sequence rather than that of viruses (Van Regenmortel 2016). The argument raised against this concern consisted in reminding that the properties of a virus are largely, or entirely, encoded by its genome and that the classification of viruses based on sequence information alone is not limited primarily by the absence of biological attributes but by the inability to accurately extrapolate the biological information and robustly infer enzymatic functions from the virus genome.

The new classification pipeline that has been proposed differed from the previous in that it considered that inferred biological properties obtained by bioinformatic analysis of virus sequences could be used together with other information from the previous classification pipelines.

I.3.4. Plant virus detection methods

Since the first steps of virology as a discipline, a number of methods have been used to detect the presence of viral particles (nucleic acids, proteins or virions) in plant extracts. Various types, specificities and applications of these methods have been extensively described and discussed elsewhere. In this subsection, only some mostly used methods are summarized in the frame of briefly showing their diversity and particularity.

a. Biological indexing

Biological indexing have been the first methods used for virus diagnosis at early stages of virology. This was done through the use of host symptomatology, sap/graft transmission, and indicator plants (Harrison 2009).

Since the beginning of the last century, plant virologists have used host range as a criterion for attempting to identify and classify viruses. In a typical experiment, the virus under study would be inoculated by mechanical means to a range of plant species that would then be observed for the development of virus-like disease symptoms.

Most of the known virus hosts are plants used in agriculture or horticulture or are weed species that grow in cultivated areas. This is due to the fact that plant virologists working primarily on diseases of crop plants have been primarily concerned with viruses that cause economic losses in cultivated plants or other plant species that might act as reservoirs for the virus or its vector that affect a cultivated species (Hull 2009).

The genera and species chosen for a host range study may not form a taxonomically balanced selection since this selection is to a great extent governed

by those species that are easy to grow in glasshouses and to handle for mechanical and insect vector inoculation.

Many good indicator species, reacting to a wide range of viruses have been found in the genera Nicotiana, Solanum, Chenopodium, Cucumis, Phaseolus, Vicia, and Brassica. Testing large numbers of samples using indicator hosts requires glasshouse facilities to keep inoculated test plants under suitable containment to minimize escape of the virus(es) to the outside environment.

When symptomless infection was produced, back-inoculation to a host known to develop disease was envisaged. However the meaning of the term "infection" on which this oversimplified view of the problem was perceived was subsequently refined to take into account the molecular mechanisms that might make a plant a host or a nonhost for a particular virus. In such context, the terms "local host" and "systemic host" were preferred in place of the term "host" to designate respectively plant species in which the virus is restrained to the inoculated leaf (subliminal infection), as well as plant species in which the virus spreads from the inoculated leaf to other (but not necessarily all) parts of the plant.

Viruses may vary widely in their host range with at one extreme some viruses that are confined on only one host in the nature (e.g. BSMV) to viruses that can infect a considerable number of species from more than 100 botanical families (CMV and TSWV). Plant viral host ranges are controlled by various factors, including the presence/absence of virus-plant host cell receptor recognition system and the little host specificity in the initial uncoating process; mechanisms that block virus replication in the cells where the virus first gained entry; the ability of the virus to move from the first infected cell; and the host defense system.

Biological indexing methods are much more time-consuming than most other methods now available. Nevertheless, they remain especially very important in phytopathology since they are require for determining the phytosanitary risks of newly discovered viruses. For diagnosis, in most circumstances only inoculation to an appropriate host species can determine whether a particular virus isolate causes severe or mild disease. However, this group of methods does have a major drawback, especially when used on a new virus or in quarantine situations. It raises the possibility that the production of infected plants might be a source of infection for local crops even in the face of the strictest containment conditions.

b. Electron-microscopy (EM)

In the 20th century, the advent of electron-microscopical technology have revolutionized the study and understanding of viruses. The electronic microscope allowed revealing the size, shape, and any surface features of the virus particles. These are parameters have been considered for long-time as a basic requirement for virus identification and have been used for virus classification (Almeida 1963) (Figure 19).

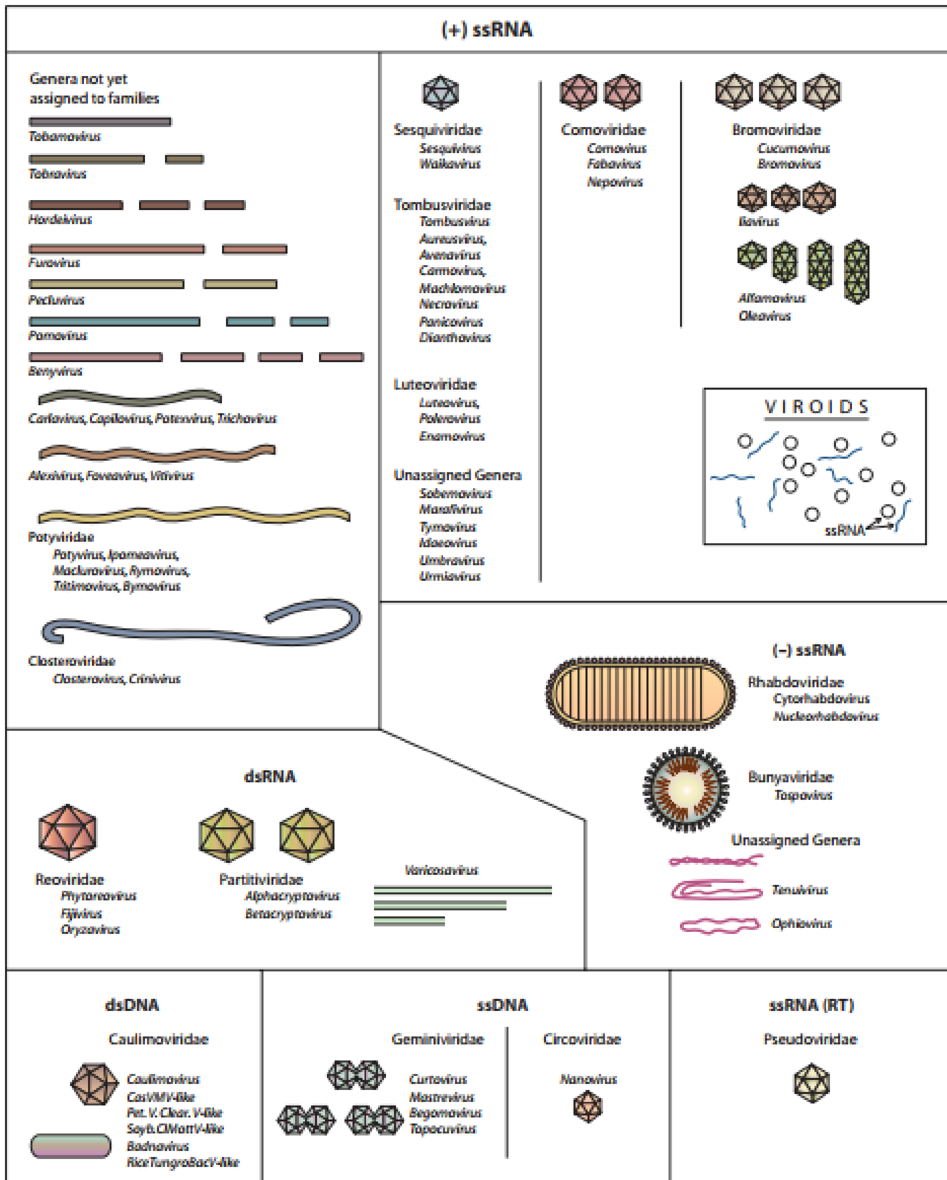


Figure I-21: Schematic diagram of families an genera of viruses and of viroids that infect plants (Agrios 2005)

Negative staining of crude extracts or purified preparations using stains like uranyl acetate, ammonium molybdate or sodium phosphotungstate is the procedure universally used. For stable viruses, EM can give rapid results using negative staining techniques. However, viruses occurring in low concentration are not easily seen. In such case sap from test material needs to be concentrated prior to observation or particles from sap can be trapped using antibody-coated grids

(Immuno-Sorbent Electron Microscopy, ISEM) to improve the detection efficiency (Kumar, P. L. and Legg 2009).

In fact, various terms have been used to describe this process : immunosorbent electron microscopy (ISEM), serologically specific electron microscopy (SSEM), solid-phase immune electron microscopy, and electron microscope serology. Basically, virus particles can be trapped onto an electron microscope grid previously coated with an antiserum and revealed by two approaches:

- The particles can be negatively stained and reacted with another antiserum that decorate particles of the virus against which the antiserum is directed.
- After being adsorbed onto the EM grid, the virus particle are coated with a virus-specific antibody on which gold particles have been attached.

Depending on size and morphology, a virus identified using this procedure may be tentatively assigned to a particular taxonomic group. However, two drawbacks or limitations arise from this procedure:

- Some small icosahedral viruses cannot be distinguished from members of unrelated groups based on morphology alone
- Differentiating virus particles from normal cell constituent when trying to detect virus within cells and tissues can sometimes be difficult. only large, enveloped viruses, plant reoviruses, and the rod-shaped viruses can usually be readily distinguished using this procedure because their appearance in thin sections generally differs from that of any normal structures.

c. Serological-based methods or immunoassays

Immunoassays constitute one of the most important and widely used methods for detection, and diagnosis of viruses. They rely on antibodies recognizing most often the surface properties of viral protein(s) that make up the viral coat. Different procedures may use the protein on the intact virus, the protein subunits from disrupted virus, or proteins expressed from cloned cDNA or DNA in a system such as E. coli or insect cells. However, non-structural proteins coded for by a virus can also be used in diagnosis.

All methods that use the viral protein properties as detection strategy are based on the serological principle of interaction between pathogen protein(s) (termed antigen) with antibodies raised against them (termed antibody) in various possible systems (mammalian or insect cell lines, prokaryotic systems (E. coli), yeast, plant-based systems) (Buyel, Twyman, and Fischer 2017; Carvalho et al. 2017; Hiatt, Cafferkey, and Bowdish 1989; Lizak et al. 2011; Spadiut et al. 2014). This explains the fact that these methods are often designed as serological methods. Depending on which question the diagnostician is addressing, three kinds of antibodies can be used (Ma and O’Kennedy 2015):

- Polyclonal antibodies : they binds multiple different structures or epitopes on the target organism/molecule i.e. they are polyspecific and useful in situations

where the desired outcome is indiscriminate detection of all viral strains or when the goal is to capture all strains of a virus from a sample.

- Monoclonal antibodies: present singular-epitope specificity, hence, they are monospecific. They are much more specific than polyclonal antisera and can be used to differentiate several strains of many pathogens. This specificity can be a disadvantage as one or several variant(s) of the pathogen may not be detected. In virus diagnosis, this is useful for differentiation between different isolates or genotypes of a single virus, or between similar viral species within a genus (Cassedy, Parle-McDermott, and O'Kennedy 2021)
- Recombinant antibodies: in the recombinant technology, monoclonal antibodies are produced as fragments of the whole antibody protein using expression hosts such as yeasts and bacteria in lieu of mammalian cell-culture (Jeong, Jang, and Velmurugan 2011). In situations where the targets are highly similar, for example similar strains of a given virus, isolating non-cross-reactive antibodies can be challenging. To overcome this, antibodies can be isolated based on their ability to detect short peptide sequences. These peptides correspond to unique regions on the target protein, providing the capacity to detect strain-specific regions on an entire protein (Cassedy, Parle-McDermott, and O'Kennedy 2021). Targeting highly specific peptides, rather than an entire protein with multiple epitope regions, could result in antibodies which are less likely to cross-react with other viruses or isolates of the viral species.

The initial development of pathogen-targeting antibodies requires time to identify a valid disease-specific antigen, isolate suitable antibodies and validate the resulting antibodies and assays. However, once generated, the production of the antibody proteins is relatively fast. Furthermore, the advent of recombinant antibody technology, has made the generation of antibodies more rapid and cost-effective due to the use of cheaper expression hosts as stated above.

A wide variety of methods have been developed for demonstrating and estimating combination between antibodies and antigens including :

C.1. Enzyme immunotests procedures [enzyme-linked immunosorbent assay (ELISA), and enzyme immunoassay (EIA) and] :

All the techniques where enzymes are employed to show antigen-antibody reactions are generally referred to as enzymatic immunoassay method (Aydin 2015).

The basic principles of ELISA techniques dates back to 1941 (Coons, Creech, and Jones 1941). However the first practical application of the ELISA method was used in 1971 by (Engvall and Perlmann 1971; Van Weemen and Schuurs 1971). The ELISA Assay has been the most widely used method and many variations of the basic procedure have been described with the objective of optimizing the test for particular purposes: direct and indirect ELISA, sandwich, competitive ELISA.

The ELISA offers some advantages (Hull 2009; Aydin 2015):

- It is very economical in the use of reactants and readily adapted to quantitative measurement.
- It can be applied to viruses of various morphological types in both purified preparations and crude extracts
- It is particularly convenient when large numbers of tests are needed
- It is very sensitive, detecting concentrations as low as 1-10 ng/ml.

The main factor limiting the number of tests that can be handled using ELISA procedures is the preparation of tissue extracts.

Direct and indirect formats are similar: initially, the antigen is coated to the surface of the ELISA well through passive adsorption, or it may be chemically linked. For direct detection, a labelled, anti-target primary antibody is applied to the well, and subsequently detected. For indirect detection, the primary antibody is unlabeled, and is detected by the addition of a labelled, secondary antibody. These formats are suitable for measuring an antibody response against a given antigen. However, there are some drawbacks to direct and indirect formats when transitioning to detecting viruses in crude samples. In fact, for both formats, the suspect sample may first have to be coated onto the ELISA test well. Since for most biological samples, the proteins of target is always combined with other proteins, they will compete for the adsorption to the assay well surface, meaning that the target-antigen may not be accurately represented in the sample. This will make the quantitation challenging (He 2013). To overcome this situation, formats such as competitive and sandwich are used.

Competitive assays are effective in the detection of viral-specific proteins from a given source. Competitive ELISA first involves the immobilization of the target antigen to an ELISA well, followed by the simultaneous application of the antibody and the suspect sample to the same well. Target-antigen in the sample competes for antibody-binding, leaving less antibody available for binding to the immobilized antigen. Therefore, in this assay format, the signal is inversely proportional to the amount of analyte in the sample. Competitive analysis is widely used for the detection of small antigens, where the binding of multiple antibodies may not be permitted (He 2013). This assay format is useful when considering the rapid-development of viral detection assays as it only requires one target-specific antibody. However, the use of only one antibody also has consequences in the form of reduced assay sensitivity and specificity if any cross-reactivity with the sample matrix occurs (Cassedy, Parle-McDermott, and O’Kennedy 2021).

The employment of a sandwich formats overcomes most issues associated with cross-reactivity as it employs two antibodies, both specific for different regions (epitopes) on the target molecule. The first of these is immobilized to the well surface. Thereafter, the sample is applied, and specific antigen is captured by the immobilized antibody. Unbound entities are washed away, and the second antibody is applied. This second antibody can be labelled directly or detected with a labelled-secondary antibody. This format offers high specificity as it requires the

binding of two antibodies to produce a signal. This could be helpful for virus detection where there is a risk of cross-reactivity occurring between similar strains of virus. It is also considered the most sensitive and robust ELISA format (Ecker et al. 2013; He 2013). Sandwich assays are less-prone to cross-reactivity. Even if one of the antibodies in the pair has some cross-reactivity with the sample matrix, the likelihood is that the same cross-reactivity will not be observed with the alternate binding antibody. This makes this assay suitable for the detection of viruses directly from their sources, such as tissue samples where the matrix can be very complex (L. Zhang et al. 2018).

A drawback to the sandwich assays is the complexity of design. Both antibodies needs to bind, unhindered by one another, to the target molecule. Finding a suitable binding pair can be difficult, and thorough assay validation is required to ensure no non-specific interaction occurs between the antibodies within the assay. However, the use of antibody-discovery technology, such as phage display, has led to the development of methods which facilitate the identification of binding pairs of antibodies in a rapid manner, negating the time-consuming screening associated with finding two separate binding pairs (Gorman et al. 2017). Alternatively, the previously described strategy of peptide-targeting can be employed to design antibodies which can detect distinct regions on the whole antigen molecule, facilitating sandwich detection.

ELISA continues to remain a staple platform in virus detection due to its sensitivity and robustness. However, there are drawbacks to its use. These include the risk of cross-reactivity of antibodies to other co-infecting viruses, resulting in false positive or inaccurate quantification. This highlights the need to fully validate virus assays prior to use. In contrast to a potential for false-positives, ELISA, and other immuno-assays, may also be at risk of presenting false negative results. This generally occurs when testing is performed at the early window of infection. At this stage, the quantities of viral antigen may be present at low concentrations, challenging the sensitivity of the ELISA format, potentially leading to erroneous results (Alexander 2016).

Further to this, ELISA procedures can take several hours to perform. This limits the capacity for rapid testing and results turnaround. However, the multi-well layout of ELISA plates facilitates multiple simultaneous tests, which is ideal for scenarios where high volume testing is required, such as epidemic or pandemic scenarios (Cassedy, Parle-McDermott, and O’Kennedy 2021)

C.2. Immunoblotting or blotting techniques:

They involve generally a detection of antigen on the surface of a membrane. Sap or insect extracts are spotted onto a membrane, allowed to dry and the membrane is then probed with an anti-virus antibody. Blotting techniques are useful for cheap detection of multiple samples, as many samples can be blotted on the same membrane. However, the time required to produce results is a considerable drawback, as the turnaround may take several hours. There are several variations of blotting techniques, but the most commonly used ones are:

- Dot immuno-blotting assay (DIA) or dot blot: its principle is similar to that of ELISA, except that it is performed on nitrocellulose membranes and precipitable substrates are used for development of positive reaction at the site of reaction. Chemiluminescent or radioactive substrates are also used, but in this case, energy (light or radiation) emitted is captured by exposing it to x-ray film. Additionally, this method is cost-effective and offers advantages such as the requirement for a small sample volume, the ability to detect virus directly from the sample and the fact that blotted membranes can be stored for a number of days before testing, facilitating the testing of numerous membranes simultaneously (Cassedy, Parle-McDermott, and O'Kennedy 2021). DIA is as sensitive as ELISA, but it requires optimization and it is not suitable for testing plant tissues which contain high amount of polyphenols that gives background reaction. Additionally, the results are mostly qualitative, requiring further instrumentation to perform quantitative analysis which may increase costs.

- Tissue printing (or tissue blot or tissue print immuno-blotting or immuno-tissue printing) : is similar to DIA, but instead of sap extracts, the whole infected tissue is pressed onto the nitrocellulose membrane to facilitate virus antigen immobilization. Subsequent detection is similar to that of DIA. This methods is very useful in determining virus in the tissues (Kumar, P. L. and Legg 2009).

- Western immune-blotting (WIB): is another variation of DIA. In this case, proteins are first separated by electrophoresis on a polyacrylamide gel in a medium containing sodium dodecyl sulfate (SDS-PAGE). Then the proteins are transferred onto nitrocellulose membrane and detected using antibodies. This assay is commonly used to differentiate virus strains, epitope mapping and also for accurate detection of virus from total protein extracts (Kumar, P. L. and Legg 2009).

- Protein dot blots: resemble Western blotting except that the constituent proteins in the sample are not separated. The samples are simply spotted onto a matrix such as a nylon or nitrocellulose and probed with an antiserum to which a reporter group has been attached. Reporter group can be either enzymes that give a colour reaction with a specific chemical or a fluorescent compound. Unlike ELISA, the colour product must be insoluble. Several commercially available kits in which the antibody is immobilized on a strip and the sample is diffused along the strip have been commercially made available for routine diagnostic of Covid-19 test, pregnancy testing, malaria testing, etc. (Tangpukdee et al. 2009; Jacofsky, Jacofsky, and Jacofsky 2020).

C.3. Lateral-flow immunoassay (LFIA) or immuno-chromatographic test (ICT):

Unlike immuno-enzymatic tests, immuno-chromatographic tests are a combination of chromatography (separation of components of a sample based on differences in their movement on a solid adsorbent pad by capillary flow) and immunological reactions (I. D. Watson 2000). LFIAs are extensively used for

detecting virus-associated protein directly from source. The assays are typically formatted to be read by way of a color change at a test line as a result of either sandwich, competitive or multiplex immunological reactions: In the sandwich format, a colour change at this line corresponds to a positive result. Whereas in competitive format, no colour change is indicative of a positive result. The multiplex format is used for detection of more than one target species and assay is performed over the strip containing test lines equal to number of target species to be analysed (J. a Tomlinson, Dickinson, and Boonham 2010; J. a. Tomlinson et al. 2013). Control lines are incorporated to ensure test validity. The color change is facilitated by the labelling of antibodies with various molecules. Various labels such as gold nanoparticles, coloured latex beads, carbon nanoparticles and magnetic particles facilitate one-step results where a colour change can be observed with a naked eye, facilitating rapid diagnosis. Other labels such as fluorophores, quantum dots and enzymatic labels can also be employed, however these may require additional equipment or steps to demonstrate results which may be qualitative or quantitative (Mak, Beni, and Turner 2016).

For virus detection, LFIA offer an appealing alternative to methods such as ELISA or blotting methods. One primary advantage is that LFIAs have a turnaround time of minutes, rather than hours. LFIAs are also cost-effective, requiring minimal sample volumes to prepare and facilitating diagnosis with naked eye. The limit of this is that it offers only the possibility of qualitative data. To achieve quantitative results, several commercially available optical readers that are based on reflectance or fluorescence principles have been developed for lateral flow tests (Mak, Beni, and Turner 2016).

Another advantage of LFIA is their ease-to-use, which makes the technology accessible to a wide range of users, even those less-skilled (Byzova et al. 2018). This is helpful in scenarios where a high volume of testing is required, as non-specialists can be quickly trained to perform assays. While they may not always offers the same sensitivity as nucleic acid detection methods, particularly at early stages of infection, they come with the advantage of a reduced cost, reduced complexity and higher utility for use by untrained personnel. These factors are exemplified by the LFIA which can be produced cheaply, has no need for additional equipment and requires little to no training to use (Cassedy, Parle-McDermott, and O'Kennedy 2021).

d. Molecular or nucleic acid-based methods

The sequence of a viral nucleic acid, whether double or single-stranded DNA or RNA in a single or several segments is fundamental for allocating an unknown virus to a particular family or group and to determine its species, representing an interesting target for specific detection. For detection and diagnosis, nucleic acids that make-up the genomes of viruses can be analyzed following four approaches:

1. The cleavage pattern of DNA: The procedures are based on the principle that any DNA or cDNA of RNA genomes can be cleaved by restriction enzymes to produce fragments of various sizes that can be used for differentiating viruses of

a particular group (Saraswathy and Ramalingam 2011). This procedure is known as the **restriction fragment length polymorphism** (RFLP), invented in 1984 by Alec jeffreys (Chaudhary and Maurya 2019) and widely used for instance to distinguish isolates of cassava begomoviruses (Borah and Dasgupta 2012).

2. Hybridization procedures: All molecular hybridization procedures mimics the interaction of base-pairing principle that involve hydrogen bonds between complementary bases, resulting in the double-stranded structure of the DNA (WATSON and CRICK 1953). For diagnosis of viruses by hybridization, a known nucleic acid termed “probe” comprising sequences complementary to an unknown nucleic acid termed the “target” and to which a reporter system is attached to reveal when hybridization has taken place, are used. Various hybridization formats exists and include for instance:

- Southern blotting: Based on the principle of separation of DNA fragments, generally obtained through RFLP. The gel electrophoresis is used to separate those fragments, followed by the revelation of the target-fragment using labelled probe hybridization after a transfer on a nylon membrane (Southern 1975).
- Dot blot hybridization (DBH) : resemble the southern blotting except that the specific sequence of DNA or RNA is not separated by electrophoresis. One version of the DBH is known as the NASH (virus-specific nucleic acid spot hybridization :NASH), and consist in spoting the sap extracts directly on membrane where the hybridization and revelation are directly conducted. This method has the advantage of testing several samples from different sources in a single run and can simplify virus detection and eliminates the drudgery of sample collection and nucleic acid extraction. However, it has been demonstrated that some unexpected cross-hybridization can occur. The NASH has been tested particularly for the detection of CBSVs in cassava tissues but , unfortunately proved unsuccessful due to interference of latex in cassava tissue, secondary structures or low virus titer (CRS 2012).
- Tissue print hybridization: the basic principle of tissue printing is that most of the cell contents, especially on the surface of a freshly cut tissue section, can be transferred to an adhesive or absorptive surface with little or no diffusion, by simple contact, with excellent preservation of anatomical details (Ding et al. 2017). It is similar to the immune-tissue printing, where a cut surface of a leaf or stem (or any other plant organ) is applied to a nitrocellulose membrane and the presence of target nucleotide sequence is revealed using a labeled nucleic acid probe. This method is practical in studying the distribution and localization of target nucleotide sequences in tissue sections of organisms and has been of special use for detecting viroïds that do not have proteins that cannot be detected immunologically (Kaponi, Sano, and Kyriakopoulou 2022).
- In situ hybridization (ISH) : In this technique, a biological sample consisting of tissue sections, cells or chromosomes from an individual is affixed to a

glass slide and then exposed to a labelled probe. The location of the bound probe can then be seen with the use of a microscope (Baumgart, Schad, and Grabenbauer 2001).

3. Polymerase chain reaction (PCR) : formulated by (Mullis et al. 1986), it is an in vitro enzymatic method of amplifying DNA sequences, spanning between two synthetic complementary oligonucleotide primers that hybridize to a target genome region, by repetitive synthesis using a heat-stable DNA polymerase. PCR utilize multiple stepwise temperature cycles comprised with 3 main steps: melting, annealing and extension. During these steps, the sequences between the primers are doubled so that after n cycles, a 2n amplification should be obtained. The wide applicability of the PCR has made it a powerful tool from which revolution in the detection and diagnosis of plant pathogens, increasingly viruses, have been accomplished. Several variants of this technique exists, each one adapted for a specific context. For instance:

- Reverse-transcription PCR (RT-PCR): for amplifying RNA sequences from an intermediate complementary DNA (cDNA) synthesized through a preliminary step of reverse transcription before the PCR.
- Immuno capture PCR (IC-PCR or IC-RT-PCR): the PCR is coupled with an initial step of immune-capturing viral particles using antibodies (Ding et al. 2017). This technique is widely used for diagnosing episomal infection of certain viruses, notably the banana streak virus (BSV) (Harper et al. 1999).
- Multiplex PCR: in which primers are designed to distinguish between strains, isolates or sometimes species of one organism or several species of an organism in the same reaction tube.

Gel electrophoresis is considered as a standard method of analyzing amplified DNA products. However, there are other nucleic acid detection methods that uses alternative post-amplification strategies and that deserves special attention. These methods are the following:

a. *Real-time quantitative* **(RT)-PCR** (**qPCR**): this method monitor the amplification process using fluorescently-labelled probes or DNA intercalating dyes such as SYBR®green. To use these dyes, the primers must be highly optimized and produce no non-specific amplicons as these non-specific products will also produce signal, skewing amplification. The level of fluorescence in the PCR sample is directly proportional to the initial concentration of the target in the sample and is measured to allow its quantification. One advantage of the qPCR is the promise of high specificity when using probes such as hydrolysis or hybridization probes, as both the primers and the probe must bind the target sequence to achieve signal. When the time-pressure is an issue, an assay which produces highly specific, quantitative results over the course of a few hours is very desirable. Another desirable trait of this method is its capacity of high throughput testing owned to the fact that it is typically used in a 96-well assay format (Eigner et al. 2019). This high-throughput

feasibility is complemented by the possibility of performing multiplexed reactions where multiple pairs of specific primers and suitable probes, targeting different viruses which may be found in the same sample, facilitates testing for panels of viruses (Ou et al. 2020). However, sample preparation, which is generally required, can be a delaying factor. Several commercially-available kits can facilitate a rapid isolation of DNA or RNA, however they come with the disadvantage of added costs (Clark, Zhang, and Anderson 2016). This in conjunction with the initial cost of the qPCR equipment itself, could make detection methods like qPCR inaccessible or difficult to run on full potential in developing regions. Another drawback to qPCR include the high cost of equipment and reagents, the need for strictly controlled temperature cycling, and the necessity for trained staff to run testing (Cassedy, Parle-McDermott, and O'Kennedy 2021).

b. *Loop-Mediated Isothermal amplification* **(LAMP):** developed by (Notomi et al. 2000), this method rely on the use of multiple primers, at minimum four, to initiate the polymerase-driven extension of the gene sequence. The amplification is facilitated by the formation of stem-loop structures created by the primers. The sequence is then amplified by a strand-displacing polymerase to generate a copious amounts of dsDNA following a repetitive polymerization. Unlike PCR techniques that uses a stepwise temperature cycles, the LAMP method is performed at a stable temperature that can be comprised between 57°C and 67°C with an optimum of 65°C (Francois et al. 2011). The reaction can be performed on a range of time varying between 10 to 60 minutes according to several possibilities of optimization of the reaction conditions (Estrela et al. 2019). When the target is RNA-based, a reverse transcriptase can be incorporated directly into the LAMP reaction (RT-LAMP). The amplicons generated from the LAMP reaction can be detected through various possibilities:

    1. Use of intercalating dyes : dyes such as SYBR®green can be either added to an end-point reaction (for a rapid qualitative result), or added prior to the reaction to measure the real-time increase in fluorescence (X. Zhang, Lowe, and Gooding 2014). However it is recommended to assess the most suitable dye before validating a new assay since some dyes may have inhibitory effects on the LAMP reaction (Quyen et al. 2019).

    2. Use of fluorescently labelled probes :due to the use of a strand-displacing polymerase in the LAMP assay, the probes that are used in qPCR cannot be used. Alternative probe-based methods are proposed, including quenching of unincorporated amplification signal reporters (QUASR), detection of amplification by release of quenching (DARQ) or mediator displacement probes that makes it possible also for multiplexing

reactions (Tanner, Zhang, and Evans 2012; Ball et al. 2016; Becherer et al. 2018).

3. Use of gold nanoparticles (Au-NPs)-labelled with DNA complementary to a sequence on the target (Arunrut et al. 2016).

4. Measure of the reaction turbidity : Here the turbidity of a reaction is measured real time using a turbidimeter which facilitates quantitative analysis of target DNA by comparing an unknown sample to a standard curve. A successful LAMP reaction generate large amounts of magnesium pyrophosphate precipitate as a by-product (and can be qualitatively observed by naked eyes) (X. Zhang, Lowe, and Gooding 2014)

LAMP offers a highly sensitive and rapid alternative to classic PCR, or qPCR since the method does not require sophisticated apparatus as the reaction is maintained at a single temperature. This temperature, however, does means that some heating apparatus is required, and various types have been proposed (Papadakis et al. 2022; Uddin et al. 2021; García-Bernalt Diego et al. 2022; DHSC 2020).

The use of multiple primers make the LAMP a highly specific detection method since the primers must bind to six regions on the sequence in order for the assay to function. However, even aided with available software, design of LAMP primers is complicated since they must be designed to have specific distance between their target regions and amplify a 300 bp size of product. This is further exacerbated by the variation in viral gene sequences across different strains, which may make the identification of conserved and suitable primer-binding sites difficult (Parida et al. 2008). Additionally, as a consequence of its high sensitivity, the LAMP is faced with a challenge of carryover contamination, usually caused by products from previous experiments carrying over via environment, researcher's clothing, or lab equipment (Hsieh et al. 2014).

LAMP can be combined with a lateral flow device for onsite detection. This is made possible by labelling loop (FP and BP) or inner primers (FIP and BIP) at their 5′ end with biotin/digoxigenin (H. E. James et al. 2010; Bhat, Aman, and Mahfouz 2022).

The LAMP, followed by detection using LFD, has been successfully used for detection of cassava viruses, notably the cassava brown streak ipomoviruses, the African cassava mosaic virus and the cassava witches' broom phytoplasma (J. a. Tomlinson et al. 2013; Uke et al. 2022; Okereke et al. 2013; Vu et al. 2016; J. a Tomlinson, Dickinson, and Boonham 2010).

c.*Recombinase-polymerase amplification* (**RPA**): Proposed by (Piepenburg et al. 2006), this method is an isothermal procedure that requires a single temperature for target amplification and an enzyme (the recombinase) to separate the strands of double-stranded (ds)DNA to achieve primer binding at the target region on the template. At the

beginning of the RPA reaction, the recombinase protein integrates with the primers to form a recombinase-primer complex that identifies the complementary sequence on the template and allows the primer to anneal after the separation of the dsDNA strands (X. Zhang, Lowe, and Gooding 2014). After the primer annealing, the recombinase separates from the complex, leaving the 3′ end available to the DNA polymerase to extend the chain. The separated DNA strands are then stabilized by single-stranded binding protein (SSB) as the DNA polymerase extends the chain, forming a new dsDNA that acts as a template for further amplification (Bhat, Aman, and Mahfouz 2022). Although manufacturers recommended the used of primers 30-35 nt long, it has been demonstrated that standard PCR primers can also be successfully used (Lobato and O'Sullivan 2018)(Lobato and O'Sullivan 2018). The template for RPA can be dsDNA, single-stranded ssDNA, or cDNA. However, amplification of RNA can be achieved via the addition of a reverse transcriptase directly to the RPA mixture (RT-RPA), making the reaction time-friendly (Cassedy, Parle-McDermott, and O'Kennedy 2021). Alternatively, the RNA template can be converted into cDNA separately and then used for RPA (Bhat, Aman, and Mahfouz 2022).

RPA offers a range of advantages : a simple primer design process with no need to consider the annealing temperature (Boyle et al. 2014), the possibility to perform it on crude samples and a less sensibility to contamination compared to conventional PCR assays (Rojas et al. 2017). Additionally, RPA can be performed in a simple incubator, dry bath, or at room temperature without the need for a thermocycler and reagents can be supplied in lyophilized form without the need for a cold chain substantiation. These features make RPA suitable for use onsite or in a laboratory with minimum facilities.

RPA can be combined with a lateral flow immunoassay, through a sandwich format termed "nucleic acid lateral flow immunoassay (NAFLIA)". During this process, the amplicon is labelled with two different antigenic labels throughout the amplification. These labels are then captured by label-specific antibodies on a lateral flow strip making this approach suitable for the onsite detection of viruses in only 15–30 min (Powell et al. 2018).

RPA and RT-RPA has been widely used for the diagnosis of RNA and DNA viruses, as well as viroïds as illustrated by (Bhat, Aman, and Mahfouz 2022).

d. *Helicase-dependent amplification* (**HDA**): employs a DNA-helicase to unwind double-stranded DNA rather than heat-separation as is the case for standard PCR-based reactions. Flanking primers (of 30 bp long) anneal to the strands, where a polymerase amplifies the target region. The process is repeated to exponentially amplify the target (Cassedy, Parle-McDermott, and O'Kennedy 2021). This method has undergone improvement that allowed to raise the working temperature initially set to 37°C, to higher temperatures between 60 and 65°C (Teo et al. 2015). An RT enzyme can be added directly to the HDA reaction in a one-step

protocol to facilitate the direct amplification of RNA viruses (Goldmeyer, Kong, and Tang 2007). The amplified product can be detected through the use of fluorescent dyes and lateral flow-based detection (Kolm et al. 2019). Molecular probes commonly used in qPCR are also usable for HDA, indicating a possibility to transit from qPCR assays to HAD easily, with the advantage of providing a low-cost isothermal alternative to traditional PCR.

Since HDA only requires two primers and a few number of proteins to function, it offers a more simplistic assay design when compared to other isothermal reactions such as LAMP. However, this simplicity of design does come at the cost of reduced specificity, as issues with non-specific primer annealing can occur (Barreda-García et al. 2018). This non-specificity can however be overcome by the application of molecular probes. As for other isothermal assays, only a simple heating apparatus is required for running HDA, and it can also be used alongside crude samples, without the need for DNA or RNA extraction (Jevšnik et al. 2020).

e.*Rolling-circle amplification* (**RCA**): Its mimics the natural process by which circular DNA is replicated, and is therefore adapted for use in the amplification of circular DNA sequences, making it unique from previously described methods which amplify linear DNA. Four various formats of RCA exists:

- The standard, basic format requiring only one primer from which a strand-displacing polymerase initiates replication, providing a linear amplicon product that consists of repeat sequences of the circular DNA template. However, this method does not facilitate exponential (Mohsen and Kool 2016).

- To facilitate a rapid exponential amplification, the ***hyper-branched RCA*** is proposed, and consists in the use of two primers: the first binds to the circular target-DNA, while the second is complementary for a separate region on the single-stranded product. AS the circular DNA is replicated multiple times, this second primer continuously binds repeat regions, creating a branched amplicon (X.-H. Li et al. 2019).

- The possibility of using multiple primer through a ***multiply-primed RCA***: each primer anneals to a specific on the circular DNA template and elongation is initiated. Amplified strands are displaced by the polymerase, and a secondary primer-binding can occur on the displaced strands, creating large quantities of amplified product (Fux et al. 2018).

- Through the use of specific probes, the ***padlock probe-based RCA*** provide the possibility to act when the desired target is a single-stranded linear DNA. The padlock probe features two regions that are complementary to two regions on the target. Upon hybridization, the probe and the target are joined via a ligase-

mediated reaction, creating a circular template on which RCA can act (Mezger et al. 2014).

The RCA reaction can be performed at a range of temperatures between 30 and 65°C and generate products that can be detected in real-time through the use of intercalating dyes, fluorescently-labelled probes, or molecular beacons. Lateral flow technology can also be used to detect labelled, amplified products (Liu et al. 2019).

RCA is a rapid, sensitive and specific technique that offer advantages considering its simplicity since no target denaturation is required apart from when linear targets are to be amplified through padlock probes (Cassedy, Parle-McDermott, and O'Kennedy 2021).

f. *CRISPR-Cas-based nucleic acid detection*: Cas proteins are a category of effector proteins which play an integral role in the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system (Cassedy, Parle-McDermott, and O'Kennedy 2021). This category of proteins have the ability to cleave DNA or RNA, by the help of guide-RNAs which direct the Cas proteins to the correct cutting site on the target nucleic acid. However, some Cas proteins have the possibility to also cause an off-target cleavage, known as collateral cleavage, while having a normal cleaving activity. The applicability of the Cas system to nucleic acid detection lies in the possibility of modifying the guide RNA to facilitate targeting of specific sequence (Bhattacharyya, Thakku, and Hung 2018). Additionally, the naturally-occurring collateral cleaving activity of some Cas proteins can be also repurposed for nucleic acid detection.

In fact, to facilitate Cas-mediated nucleic acid detection, the target nucleic acid is first amplified via an isothermal reaction, usually RPA. Thereafter, the CRISPR-Cas system is introduced, featuring a guide RNA specific for a location on the target DNA. Binding of the guide RNA induces target cleavage, and at the same time, collateral cleavage also occurs. The addition of fluorophore probes which are cleaved by the collateral activity of the Cas protein provide a measurable fluorescence signal of the assay (Batista and Pacheco 2018). Commonly used fluorescent measurement apparatus can be valued for the detection of amplified and cleaved products, or, for a simpler alternative, a lateral flow strip can also be used (Gootenberg et al. 2018; Myhrvold et al. 2018; Broughton et al. 2020).

Three primary Cas systems are mostly proposed for nucleic acid detection, named SHERLOCK (specific high-sensitivity enzymatic reporter unlocking), DETECTR (DNA endonuclease targeted CRISPR trans reporter) and HOLMES (one hour low-cost multipurpose highly efficient system). SHERLOCK uses Cas13a to detect RNA, while DETECTR and HOLMES uses the Cas12a for DNA recognition (Gootenberg et al. 2018; S.-Y. Li et al. 2018; Chen et al. 2018).

The dual use of isothermal amplification and CRISPR-Cas proteins provides a rapid assay with an added specificity safety net (Cassedy, Parle-McDermott, and O'Kennedy 2021). This is due to the two-step processes of product amplification through the binding of primer sets in the initial phase, followed by the signal generation through the detection of a specific sequence on the generated products by the guide RNA. This dual mechanism can help alleviating issues associated with the non-specific primer binding sometimes observed in isothermal amplification. Additionally, the fact that both amplification and Cas-cleavage can be performed in a single reaction helps reducing the risk of contamination and requires less hands-on time (Cassedy, Parle-McDermott, and O'Kennedy 2021).

4. DNA microarray or DNA chips : The principle is hybridization of fluorescently labeled target sequences to probe sequences spotted onto a solid surface, usually a glass microscope slide. Total RNA from the infected plant is converted to DNA by RT-PCR, and the cDNA is labeled by reaction with a fluorescent dye. Probes to many different viruses or variants can be spotted onto the glass slide, and hybridization can reveal joint infections with more than one virus (Govindarajan et al. 2012).

Among the various techniques detailed above, the ELISA and PCR have been the most popular in the detection of plant viruses in general and particularly for viruses infecting the cassava plant.

Table I-4_B : Summary of the advantages, disadvantages and example application of mostly-used conventional approaches for virus detection

Method	Advantages	Disadvantage	Detection method	Detection range	Virus and host plant	References
PCR/RT-PCR	more sensitive than ELISA, as even a few copies of the viral nucleic acid present in the test sample can be amplified and detected	requires sophisticated laboratory equipment and skilled personnel. PCR-based methods are cumbersome because they require the initial isolation of nucleic acids, followed by amplification of the target sequence and analysis of the products on an agarose gel	Agarose gel electrophoresis (AGE)	<1 fg	CBSV, UCBSV, ACMV, an EACMV-UG. In cassava	(Freddy *et al.* , 2015)
qPCR	Quantitative, highly specific when using well-designed primers, highly sensitive, amenable to high-volume testing	Specialised equipment and trained staff required, relatively long runtime	Fluorescence signal with melting curves	4-10 fg	CBSV, UCBSV, ACMV, an EACMV. In cassava	Otti et al. 2015
LAMP	Rapid (30-60 min), sensitive, multiple primer requirements aid in specificity, can be quantitative or qualitative	Require higher run temperature of 65°C, primer design can be challenging, DNA sample neede denaturation prior to amplification, appear particularly prone to carryover contamination, low operating temperature can reduce specificity	Fluorescence signal on LFD	2ex-2 to 2ex-3	CBSV and UCBSV	Tomlinson et al. 2013
RPA	Rapid (20 min), can operate at room or body temperature, simple primer design, no requirements for DNA denaturation prior to amplification	Low operating temperature can reduce specificity	AGE	Equal to ELISA	Bean golden yellow mosaic virus (BGYMV) (Begomoviurs; Geminiviridae), Beans	Londono et al. 2016, Kapoor et al. 2017
HDA	Relatively simple primer design, no requirement for DNA denaturation prior to amplification	Cannot operate at ambient temperatures, may be prone to non-specific amplification due to lower operating temperature	AGE	4 pg	Tomato spotted wilt virus (orthotospovirus, tospoviridae), Tomato	Barreda-garcia et al. 2018
RCA	Multiple formats available (linear, hyper-branched, multiply-primed, padlock probe), at minimum requires only one primer	Product yield can be low depending on the amplification method, additional sample manipulation may be required when using padlock-probes in RCA	AGE	10ex-5 to 10exp-7	Cassava mosaic geminiviruses	Kathurima et al. 2016

Method	Advantages	Disadvantage	Detection method	Detection range	Virus and host plant	References
CRISPR-Cas Nucleic Acid Detection	Can be used standalone or in conjunction with already established amplification methods, may offer additional specificity as the target must be recognised by both the initial amplification method and guide RNA, relatively rapid (~60min)	Generally requires pre-amplification of target, may require manipulation of sample post-amplification, increasing carry-over contamination risk	CRISPR-Cas12a Reporter Assay with fluorescence signal	0.1 pM	Beet necrotic yellow vein virus (BNYVV) (Benyvirus; Benyviridae)), Apple	Ramachandran et al. (2021)
ELISA	Multiple formats available (direct, indirect, competit	Long runtime (hours to days), can be affected by sample matrix, different antibodies may exhibit different specificities for target, false negatives may occur in early infection window	-	1-10 ng/ml	ACMV, an EACMV. In cassava	(Freddy *et al.* , 2015)
Hybridization procedures	Cheap, can be used with crude samples blotted onto membrane, blots are relatively stable which facilitates multiple sample collection/testing	Only qualitative/semi-quantitative, results turnaround of hours-days, limited sensitivity	-	<1 pg	-	-
LFIA	Rapid (~10 min), easy to interpret results, resistant to sample matrix effects, user-friendly	Results generally limited to quantitative/semi-quantitative, does not typically achieve sensitivity comparable to nucleic acid detection since the method is only suitable for detecting viruses occurring in high titre.	-	More than PCR	Little cherry virus 1 (Ampelovirus, Closteroviridae), sweet cherry	Mekuria et al. 2014

I.4. Sequencing technologies

I.4.1. The advent of sequencing technologies

Since the discovery that DNA was the genetic material in which information is stored and that controls the characteristics of all organisms, the race for unravelling its structure and composition became a challenge of highest importance for the science because it would result in better understanding of life's secrets (Masoudi-Nejad, Narimani, and Hosseinkhan 2013).

After several scientific findings have paved the way to the discovery of the DNA structure (Demidov 2003; Tsongalis and Silverman 2006; WATSON and CRICK 1953), Frederick Sanger and Allan Maxam and Walter Gilbert facilitated a historic breakthrough by independently inventing the first technologies that allowed to know about the order (sequence) of nucleotides in DNA (Sanger et al. 1977; Maxam and Gilbert 1977).

Knowing about the genome of viruses was particularly of high importance because of their atypical morphology lacking normal cellular-type structure and the concentration of all information necessary for their functioning onto the genome. As more information regarding genomes of viruses were obtained using the sequencing technologies, their detection, characterization, diversity, classification, functioning and evolution were progressively resolved. Over years, the genome sequencing of living organisms and particularly viruses, become one of the essential tool for biological researches.

I.4.2. First generation of DNA sequencers

These technologies allowed to sequence the first organism, which was a virus : the bacteriophage Phi X174 in 1977; and the first plant viruses: the DNA of the CaMV in 1980 and the RNA of the TMV in 1982 (Sanger et al. 1977; Hull 2009).

a.Sanger sequencing technology

Also termed the “chain termination” or the “dideoxy method”, the principle used in Sanger technology is known as the sequencing by synthesis method. The sequencing reaction is performed in the presence of the single-stranded DNA template, DNA primers, DNA polymerase, four normal DNA nucleotides, and four fluorescently labelled nucleotides that have been modified by removing the 3’-OH group required for extension (ddATP, ddTTP, ddGTP, ddCTP). The DNA template is initially divided into four separate sequencing reactions containing primers, polymerase, normal nucleotides, and a small amount of only one of four modified nucleotides which will randomly incorporate into the growing strand when the reaction is started, terminating therefore the elongation and resulting in DNA fragments of various lengths.

Each of the four reaction run is then charged into individual lanes of a polyacrylamide gel and a high-resolution capillary electrophoresis is then launched to

separate the DNA fragments according to their sizes. DNA bands corresponding to DNA fragments with differing lengths are then visualized using UV light or X-ray autoradiography, and the order of nucleotides can be determined according to the relative positions of DNA bands among four different lanes (Masoudi-Nejad, Narimani, and Hosseinkhan 2013).

This method was gradually improved by the replacement of radio-labelling with fluorescent labeling of nucleotides (Smith et al. 1986), slab gel with capillary electrophoresis systems (Marsh et al. 1997) and later automated (in 1987 the first sequencing machine was the Applied Biosystem's AB370 machine) to constitute the method used in the first generation of DNA sequencers. It could deliver 96 or 384 sample sequences per instrument with a read length ranging from 600 to over 1,000 nucleotides (nt). As a result of its lesser technical complexity and lesser amount of toxic chemical used, the Sanger sequencing technology has been the most popular method for DNA sequencing for approximately 15 years and therefore used for large-scale sequencing projects, including the sequencing of the human genome and plant viruses. The method is still actually in use for specific applications.

The main limitations of the sanger technique has been (Masoudi-Nejad, Narimani, and Hosseinkhan 2013) :

- The short size of fragments to be sequenced: only DNA fragments of sizes varying between 100 and 1,000 base pairs could be produced due to limitations in the power of discrimination during capillary electrophoresis. Larger sequences must first be fragmented into smaller pieces and amplified to obtain large number of copies for each individual fragment that must be reassembled into the original sequence after sequencing.
- The needs for amplification and fragment assembly steps: the generation of DNA fragment templates was carried out through bacterial cloning steps (both in the map-based sequencing and in the shotgun sequencing) which were tedious and time-consuming since high coverage of each fragment was required to obtain high-quality sequence. The algorithms and computing resources required for the assembly of DNA fragments into larger contigs, as well as the subsequent filtering and gap-filling steps were not advanced enough to address these problems and to produce good results in the expected amount of time.
- Problems with parallelization: even though the automated sequencing through parallelization was a significant advance made at that period, the number of samples that could be processed remained of great limitation for massive sequencing projects such as those involved in the sequencing of mammals (±3.2x109 base pairs for the human genome) which would be too time-consuming.
- The cost: After three decade of gradual reductions in cost, the costs of the Sanger sequencing method remained too high to be practical for many important projects. Costs of up to 0.5$ could be generated per kilobase (Canadian Agency for Drugs and Technologies in Health 2014).

- The need for complete automation: the bacterial cloning steps used initially to generate populations of DNA fragments to be sequenced could not be done automatically. An improved method to significantly advance automation was required.

b.The Maxam-Gilbert sequencing technology

This technique is based on the degradation of DNA by successive cleaving of nucleotides by chemicals. The procedure determines the nucleotide sequence of a terminally labelled DNA molecule (the 3' normal phosphate backbone is replaced by the radioactive 32P using the alkaline phosphatase and the polynucleotide kinase) by breaking it using chemical reagents having specific cleavage activity (the dimethyl sulfate and the hydrazine as purine and pyrimidine-specific reagents respectively, then diluted acid and piperidine to separate A to G and T to C respectively). This process is run in four separate successive reactions then the PAGE is used to resolve the fragments. An autoradiograph allows to read the sizes of radioactive molecules cleaved in each of the four reaction mixtures (Maxam and Gilbert 1977).

As a consequence of its less ergonomical feasibility and the extensive use of toxic chemicals, the Maxam-Gilbert method has become obsolete (Masoudi-Nejad, Narimani, and Hosseinkhan 2013).

I.4.3. Second generation of DNA sequencers or "High throughput sequencing"

The first-generation of DNA technology underwent an unprecedent changes, paving the way for the arrival of the second and third-generation sequencing technologies, which are generally referred to as "High throughput sequencing (HTS)" technology, a term used to designate methodologies used to generate millions to trillions of nucleotide sequences in a single instrument run (Villamor et al. 2019).

The need for a considerable reduction in the cost of sequencing methods was reflected by the National Human Genome Research Institute (NHGRI) in 2004 through a prize to be granted to a group or individuals who could "significantly advances automated DNA sequencing" (Masoudi-Nejad, Narimani, and Hosseinkhan 2013). Since then, several HTS technologies, here designated as HTS platforms, have been released.

These platforms shares three common steps in the first process of generating DNA template for sequencing : (i) DNA fragmentation to create the library, (ii) addition/ligation of synthetic DNA adapters to individual fragments (for transforming the sample nucleic acids into a format compatible with the sequencing platform) and (iii) sequencing of (millions) individual fragments.

The complexity of additional steps vary not only according to platforms types but also according to the type of starting material used. For instance, when RNA is used as starting material, RNA selection is performed prior to library construction and the

fragmented RNA is first reverse transcribed to create a cDNA library. Three commonly used selection methods are: (i) ribosomal RNA depletion, (ii) enrichment for polyadenylated RNAs, and the small RNA selection.

The limitation of host nucleic acid in the HTS technologies is critical to boosting virus signals toward the detection threshold since extremely low titer viruses may be missed (Kreuze et al. 2009). For that purpose, some strategies allowing the enrichment in viral particles are privileged before sequencing. For instance, the ribosomal RNA depletion is the most commonly employed selection method as it allows the detection of virtually all viruses present in the sample whereas polyadenylated RNA selection is particularly suited if the target virus contains of a poly-A tail in the 3' terminal end of its genome (Villamor et al. 2019). The small RNA selection method on the other hand, has been also widely used since the demonstration that the virus or viroïds-derived small-interfering RNAs (siRNAs) produced by the plant silencing machinery during the infection process could be used to detect both known and previously uncharacterized plant viruses and viroïds within infected plants (Kreuze et al. 2009; Donaire et al. 2009; Massart et al. 2019). This method starts with total RNA isolation followed by the isolation or separation of the low molecular weight RNA species (LMW RNA), containing the sRNAs (Rosas-Cárdenas et al. 2011).

Four parameters can be used to categorize the different existing HTS platforms, notably (i) the method used to detect the nucleotide sequences, (ii) the proximate source of the nucleotide (the mostly used criteria for the categorization of sequencing strategies), (iii) the sequencing chemistry employed, and (iv) the length of the read produced (Levy and Myers 2016; Villamor et al. 2019; Masoudi-Nejad, Narimani, and Hosseinkhan 2013). To allow focusing on sequencing technologies and protocols that have been centered on the study of plant viruses and their detection, and more notably to avoid referring to sequencing technologies that no longer exist, the above-mentioned parameters will be detailed only for the Illumina (the most popular) and the Oxford Nanopore Technology (ONT, in expansion) that have shown a real potential for plant viruses study. The ONT is provided by the consortium Oxford Nanopore (Oxford, UK), Quantapore Inc. (San Francisco, CA, USA) and Stratos (Roche, USA)

1. The method used to detect the nucleotide sequences

The Illumina uses an optical fluorescence detection method for identifying the base that have been incorporated or hybridized. In most Illumina platforms, each dNTP is bound to a single fluorophore that is specific to that base type and therefore requires four different imaging channels, however in other Illumina platforms like the NextSeq and Mini-Seq systems, a two-fluorophore system is used (Quer et al. 2022).

The ONT on the other side detects directly the DNA composition of a native ssDNA molecule. In this platform a motor protein is used to translocate a DNA template through a protein pore in which a current is also concomitantly passed. Shifts in voltage resulting from the blockade of the current when a particular nucleotide occupies the pore is temporally traced (squiggle space) and is used to characterize the particular DNA sequence in the pore, which can then be interpreted as a K-mer (Quer

et al. 2022). The ONT MinION uses a leader-hairpin library structure that allows the forward DNA strand to pass through the pore, followed by a hairpin that links the two strands, and finally, the reverse strand.

2. The proximate source of the nucleotide:

This parameter is the mostly used for the categorization of sequencing strategies. In fact, having many thousands of identical copies of a nucleic acids fragment in a defined area of a sequencing platform ensures that the signal can be distinguished from background noise. The Illumina platform realize this by performing a clonal amplification of DNA to generate a library of reads for sequencing. A solid-state amplification following a Bridge strategy is employed by the Illumina to realize this clonal amplification. A forward and reverse primers covalently bound to a slide surface, either randomly or on a patterned slide, provide complementary ends to which ssDNA templates cab binds. The distal ends of hybridized templates interact with nearby primers by forming a kind of bridge between these primers, then amplification can take place.

The ONT on the other side uses reads originating directly from a single molecule previously prepared for sequencing.

According to the purpose of the sequencing, various library preparation strategies, adapted to the type of starting material (nucleic acid template) can be used to realize the clonal amplification in the frame of the HTS. Most of these approaches has been borrowed from the metagenomic studies and particularly adapted to be suitable for studies of viruses . Most important strategies of library preparation used for the study of viruses includes (but is not restricted to):

- Total RNA: This approach has been privileged in recent years for the characterization of CBSVs diversity, distribution and evolution (Titus Alicai et al. 2016b; Ndunguru et al. 2015; Amisse et al. 2019).
- Small interfering RNA (siRNA): Whenever dsRNA occurs within a plant cell, such as occurs during the replication of RNA viruses, it is processed by Dicer-like proteins to produce siRNAs of sizes varying between 21 and 24 nucleotides. This mechanism is derived from the antiviral adaptive immune system of plants, a component of a more general sequence-dependent system that allows controlling gene expression (RNA silencing) (Baulcombe 2004). Various studies have used the virus-derived siRNAs to detect a variety of plant-infecting RNA viruses (Potyviridae, Closteroviridae, Tymoviridae, Luteoviridae, Betaflexiviridae, and Alphaflexiviridae); and surprisingly circular dsDNA viruses (Geminiviridae, and Caulimoviridae) (Roossinck 2015).

The reason why circular dsDNA viruses can also be detected using this strategy could be due to the fact that both coding and non-coding regions of these viruses are transcribed in both orientations in the nucleus to generate dsRNA precursors of viral siRNAs (Seguin et al. 2014). Apart from its sensitivity in detecting known and unknown viruses within single plants, this approach can also be used to detect

integrated endogenous elements if transcribed. The main disadvantages of this approach is that siRNA extraction might involve cumbersome Trizol and CTAB-based protocols, and may be problematic when used to detect viruses that either do not trigger silencing responses or that produce silencing suppressors. This is the case for some viruses that keeps their dsRNA genome encapsidated and never expose it to the cytoplasm of the cell. These viruses only extrude a ssRNA from the capsid and used it as mRNA and pregenomic RNA that is packaged prior to second strand replication (Safari and Roossinck 2014). This mechanism is used by many persistent plant viruses that are found in meristem of plants where most other viruses are controlled by silencing (Baulcombe 2004).

- Double-stranded RNA (dsRNA): RNA viruses often form dsRNA either as their encapsidated genome or during the process of replication. Although dsRNA is largely unique to viruses, they can be abundantly found in plants. For the purpose of dsRNA analyses, several protocols exists. For some of them, the nucleic acids are first purified through a phenol:chloroform extraction and the resulting mix of total DNA and RNA is enriched for dsRNA either by LiCl fractionation (Diaz-ruiz and Kaper 1978) or the use of long polymer cellulose chromatography (Dodd et al. 1984). The use of this approach allowed the discovery of fungal viruses (M.J. Roossink 2014) and allowed also to show that in uncultivated plants, persistent viruses are the dominant types of plant viruses (Marilyn J Roossink 2012). The main shortcomings of this approach are that it is labor intensive, it is not effective for the detection of (-) sense ssRNA viruses (they do not accumulate large dsRNA amounts during replication), and it cannot be used for the detection of DNA viruses (Roossinck, Martin, and Roumagnac 2015). Various studies using this approach allowed to identify for the first time viruses from the family Secoviridae, Tombusviridae, and Tymoviridae (Thapa, 2012; Scheets 2011; Min 2012).

- Virion-associated nucleic acids (VANA): this approach is based on the enrichment of virus-like particles (VLPs) by removing the host nucleic acids to maximize the proportions of virus sequence reads obtained from NGS runs. Up to four steps are generally used to achieve this: (i) VLP isolation through CsCl gradient or filtration-based techniques, (ii) RNAse and DNAse treatments to remove nonencapsidated nucleic acids after virion purification, (iii) ultracentrifugation for harvesting any virions present in the sample, and (iv) extraction of VLPs protected within capsids followed by the sequencing (R. J. Hall et al. 2014). The advantage of this approach is that it allows a simultaneous detection of both RNA and DNA viruses due to the fact that it combines a reverse-transcriptase priming and a Klenow fragment polymerization in initial step (Candresse et al. 2014). The main drawback of this approach is that it most of the times reported as cumbersome, it may not detect unencapsidated viruses or those with unstable particles. Additionally, this approach cannot effectively recover viruses from plants having high levels of either phenolic compounds or highly viscous polysaccharides (Roossinck, Martin, and Roumagnac 2015).

Two nucleic acid preparation approaches can also be combined together to allow bridging the detection gaps of each approach and to solve ancient or emerging etiologic enigma as illustrated by studies from (Candresse et al. 2014; Wylie et al. 2014), who have combined siRNA with VANA and total RNA with dsRNA respectively.

3. The sequencing chemistry employed:

The Illumina platform utilize the polymerase to drive a sequencing by synthesis reaction (SBS) following a cyclic reversible termination (CRT) approach. The CRT approach is defined by its use of a terminator molecule that is similar to that used in Sanger sequencing, in which the ribose 3' -OH group is blocked, thus preventing elongation (Guo et al. 2008). This strategy is the one used by the Illumina CRT system, accounting for the largest market share for sequencing instruments.

On the other side, the ONT directly determine DNA sequence as it passes through a nanopore (Quer et al. 2022).

4. The length of the read produced:

The Illumina generate short reads of sizes that can vary up to 300 bp. This read size can be suitable for genome sequencing or the transcriptomic applications through RNA sequencing (RNA-seq). For this reason, the Illumina is categorized as a Short read sequencing (SRS) platform.

The Illumina dominates the short-read sequencing industry owing to its high level of cross-platform compatibility and its wide range of platforms primarily resulting from its maturity as a technology (Goodwin, McPherson, and McCombie 2016). In addition, platforms relying on the CRT approach are much more precise in resolving homopolymer regions than those using the single nucleotide addition (SNA) approach. The standard limitation of platforms using the SRS strategy is their incapacity to resolve the long repetitive elements of genomes. Additional shortcomings of SRS platforms are the following :

- They restrict the possibility of identifying structural variants,
- The large deletions or insertions settle repetitive elements and/or sequences with extreme guanine-cytosine (GC) content,
- Identifying whether different mutations are associated inside the same genome is difficult. On the other side, even though third-generation sequencing could provide large and extremely large reads in one single read (60 K-2 M bp), their main drawbacks are the systematic higher sequencing error rates. These errors are the consequences of problems on the consistency of the electric signals, difficulties in obtaining multiple overlapping reads, the large amount of genetic material required for direct sequencing, and the low throughput (Quer et al. 2022).

On the opposite, the ONT platform is categorized as a Long read sequencing technology due to its capacity of delivering reads in excess of several kilobases, allowing for the resolution of large structural futures among those constituted by long

repetitive elements for which SRS technology is insufficient to resolve. The approach used by the ONT for constructing the long read is often called a "Single-molecule long-read sequencing" approach in the opposite of other approaches known as "synthetic long-read sequencing" approaches. The main difference is that single-molecule methods do not rely on a clonal population of amplified DNA fragments to generate a detectable signal, nor do they require chemical cycling for each dNTP added. In contrast, the synthetic approaches generate long-reads following an in silico computational assembly of a larger fragment after using the barcode approach for library preparation (Goodwin, McPherson, and McCombie 2016).

One of the major achievements of these technologies has been to resolve the 20-year-old barrier that has hidden the 8% of the human genome and to finally release the complete human genome, telomere-to-telomere (T2T) (Nurk et al. 2022).

Within its series of instruments, the ONT offers an advantageous option in term of practicability: the ONT MinION; a small USB-based device that runs on a personal computer, giving it the smallest footprint of any current sequencing platform. Although the device require substantial adjunct equipment for library preparation (thermocycler for instance), it is highly portable, cost effective, provide real-time results if no large datasets are needed, and is very useful for diagnosis in isolated locations.

However, ONT MinION has some limitations, namely : large error rate (5-25%), errors due to modified bases (they alter the typical voltage shift for a given k-mer), and indel errors in the homopolymers causing frameshift errors during gene calling (M. Watson and Warr 2019; Delahaye and Nicolas 2021; Jain et al. 2015). However, the high error rate have been recently significantly decreased to 0.0041% by using an approach copied from the synthetic long-read sequencing approach, of Metabarcoding: the unique molecular identifiers (UMI) (Karst et al. 2021). In this approach, each molecule within a given sample library is labelled with a unique barcode before PCR amplification so that bioinformatics software can filter out duplicate reads and PCR errors with a high level of accuracy (Goodwin, McPherson, and McCombie 2016).

I.4.4. Conclusion: comparison of various sequencing strategies and performances in the detection of plant viruses

While the HTS technologies have the theoretical capacity to target any viral nucleic acid in any host plant or vector, the available protocols present distinct advantages and limitations as reviewed previously (Roossinck, Martin, and Roumagnac 2015; Maclot 2021) (Table I-5).

The selection of the target nucleic acid s population is critical as it defines the types of viral sequences that will be detected, including total RNA/DNA, RNA-Seq, Virion-Associated Nucleic Acids (VANA), dsRNA, small RNA, Circular ssDNA or amplification of targeted PCR products using generic primers.

The various limitations mentioned suggests a careful a priori evaluation of the viruses and viroids potentially infecting the plants in the study area. However, neither the viruses, nor their (genomic) characteristics are a priori all known. Therefore protocol selection may be strongly influenced by information about the virus that are common in the study area or are most interesting given the question(s) addressed. The research question and the objectives of the study are therefore key for protocol selection, along with a proper sampling strategy (Maclot 2021).

In the sequencing phase, proper read length selection is of key consideration. For small RNA sequencing for example, the very short length of small RNAs (21-24 nt) complicates sequence assembly and annotation, and makes genome reconstruction and strain identification more difficult (Massart et al. 2019; 2017).

Sequencing depth is directly correlated with an improved ability to detect viruses present at low concentrations/abundance as previously demonstrated (Massart et al. 2019). However, a higher sequencing depth increases the probability of false positive detection due to potential contamination problems (i.e. the likelihood of sequencing contaminants is higher) and increases the sequencing price per sample (Massart et al. 2014)

Protocols allowing an enrichment of viral sequences, such as VANA or dsRNA will significantly improve the sensitivity of virus detection for a given sequencing depth (Arnaud G. Blouin et al. 2016; Essowè Palanga et al. 2016).

Protocols that necessitate intensive labor work or cumbersome steps such as dsRNA and small RNA sometimes favor contaminations that can affect seriously sequencing results.

Table I-5: Advantages and drawbacks of different sample preparations for HTS-based detection of plant viruses. The type of protocols used to detect a specific nucleotide is used as the categorization criteria (Maclot 2021).

Nucleic acids	Total RNA	dsRNA	Small RNA (21, 22, 24 nt)	VANA
Advantages	Detection of any RNA or DNA virus and viroids	All RNA viruses including viroïds	Screen any kind of virus and viroïd targeted by silencing mechanism	All viral particles, detection of DNA and RNA viruses
	For individual plants and pooled samples	Enrichment of viral sequences in the data	For individual plant samples	Enrichment of viral sequences in the data.
		For individual plants and pooled samples		For individual plants and pooled samples
Drawbacks	High sequencing depth is needed as there is a high background of rDNA (even with depletion of ribosomal RNAs)	Labor intensive	Cumbersome extraction methods (Trizol and CTAB-based).	In theory, no detection of viroids or virus nucleic acids not encapsidated r with unstable particles. But endornaviruses were demonstrated to be detected with this technique.
	No enrichment of viral sequences	Limited or no detection of DNA viruses	Difficult annotation of sequences and genome reconstruction due to the small size of the sequences.	Introduction of technical bias (enrichment) for quantification of variants and species
	Limited sensitivity to detect viruses in low concentration	Introduction of technical bias (enrichment) for quantification of variants and species	Only detecte actively replicating agents targeted by plant silencing.	Highly variable in recovery of viruses.
			Many viruses can have very low sRNA titer in woody crops.	
			Complicated assembly requires high sequencing depth in order to be able to assemble and identify viruses	
			No enrichment of viral sequences	
			Limited sensitivity to detect viruses in low concentration.	

I.5. Decision-making for selecting a diagnostic test

From a historical perspective, new virus detection has been often the output of etiological investigations (virus identification and characterization) of economically important virus and virus-like diseases of agricultural crops (Villamor et al. 2019). These etiological investigations are considered as the spearhead for the development of reliable detection methods necessary for the application of a successful disease management program.

The wide possibility of applying biological, physical and molecular techniques has given a large “tool-kit” for the diagnosis of plant viruses. In addition, the range of available techniques been further expanded these last years as a consequence of technological innovation with the advent of high throughput sequencing technologies. Despite this innovation, two key elements to consider when it comes to choose which method(s) is/are the most suitable for diagnosing or detecting viruses are important to define : (i) ***which question is to be addressed*** *?* (i.e. what is the scope/purpose of the test ?) , and (ii) ***which performance criteria*** to use for identifying the best-suited method ?.

For the first question, the purposes of a test include the good knowledge of the organism, the matrix, and the method, e.g. detection and/or identification of organism X in matrix Y by method Z (EPPO 2021). The purpose of a test can consist in (Massart et al. 2022) :

- Determining the presence of plant virus in a pre- or post-entry quarantine, phytosanitary certification and overall movement of planting materials across the globe.
- Verifying the suspected presence of a virus on a diseased plant
- A targeted routine surveillance of plant pathogens, whether in epidemiological purposes

Fulfill a specific research purpose: for instance studying viral genetic diversity, investigating the unknown etiology of symptoms, characterizing the virome or the microbiome of an entity, breeding for resistance, study of gene expression etc. Because each specific test has a broad range of applications, the importance of defining clearly the scope of the test lies in the fact that it will allow targeting the suited application(s) necessary for addressing correctly the investigated question (Massart et al. 2022).

In certain contexts, ones does not necessarily need a sophisticated technique that identifies strains of a virus while in other contexts it is very useful to have an understanding of the range of variation of the virus, the population of viruses or interestingly the microbiome of an ecosystem.

International guidelines, drawn up by expert panels to assist in the choice of the safest and simplest tests for ensuring that plant propagules do not contain the viruses that are of quarantine purpose, have been elaborated previously and used

extensively for the diagnostic using herein described traditional methods (Diekmann et al. 1994; IPPC 2021; Kumar, P. L. and Legg 2009). However, the growing interest brought about by the application of HTS technologies in diagnosing plant health has triggered the development of new guidelines and/or the adaptation of previously existing rules and procedures among states and countries to reflect the changes that HTS can brings into the certification pipelines (Benedicte Lebas et al. 2022; B Lebas et al. 2020; Massart et al. 2014; P. L. Kumar et al. 2021; Olmos et al. 2018).

On the other hand, the various possible tests does not have the same performances. The important performance criteria to consider while assessing for the suited test are the following (Massart et al. 2022; EPPO 2021; 2018; Hull 2009):

- Analytical sensitivity : correspond to the smallest amount of target that can be detected reliably (it is sometimes referred to as “limit of detection”).
- Analytical specificity: Comprises inclusivity and exclusivity:
    - Inclusivity: Correspond to the performance of a test with a range of target organisms covering genetic diversity, different geographical origin and hosts.
    - Exclusivity: Correspond to the performance of a test with regards to cross-reaction with a range of non-targets (e.g. closely related organisms, contaminants).
- Selectivity: Correspond to the extent to which variations in the matrix affect the test performance (matrix effect).
- Repeatability: is defined as the level of agreement between replicates of a sample tested under the same conditions.
- Reproducibility: Is the ability of a test to provide consistent results when applied to aliquots of the same sample tested under different conditions (e.g. time, persons, equipment, location).
- Diagnostic sensitivity: Correspond to the proportion of infected/infested samples testing positive compared with results from an alternative test (or combination of tests). It is calculated as : Diagnostic sensitivity= True positives/(True positives + False negatives).
- Diagnostic specificity and false discovery rate (FDR): The diagnostic specificity correspond to the proportion of of uninfected/uninfested samples (true negatives) testing negative compared with results from an alternative test (or combination of tests). ***Diagnostic specificity = true negative/(True negatives + false positives)***. For HTS tests, the FDR is the most suitable way to compute the diagnostic specificity as the number of potential targets is often unknown, making the determination of true negatives questionable (Massart et al. 2022)

Overall, for making the decision when selecting the best or appropriated diagnostic technique, the following points have to be taken into account:

- The sampling protocol and the reliability of the technique:

The choice of the diagnostic test can be significantly affected by the sampling plan and sampling procedure that are intended to be used in an assay. They will be important in determining the speed of operations, the cost of equipment, the consumable supply and the labor needed.

The sampling procedure as well as the samples handling are therefore important when it come to specify which test will be suitable because they will clearly detail the:

- Matrix: plant, seed, cuttings, leaf, fruit, flower, stem, roots, purified cultures, soil, water, insects, etc.
- Type of material: which zone or part (eg. lower leaves) ? age ? depth?
- Type of sample: random ? pooled ? symptomatic or not ?
- Minimum amount for each sample
- Number of samples: per field, per location, per batch, per grid
- Season of sampling
- Handling, sample preservation, labelling, packaging, storage and transportation procedures and conditions
- The target substance necessary for the test: crude extracts ? purified extracts ? genome ?

These information can impact significantly the reliability of a test in two ways: causing false negative results if for instance the part of the plant sampled did not contained the virus, the inhibition of the reaction by a plant constituent or the limitation in the reagents being used (e.g. primers for PCR etc.); or false positives results that can be due to poor traceability of samples, cross-contamination or low specificity of the test.

Additionally, this information can guide in choosing which test is suitable according to the entity of target during the reaction: viral particles, viral proteins, viral genome. In this perspective, the methods involving the use of crude or/and slightly purified plant extracts will necessitate less steps than methods targeting the viral genome since the latter will require additional steps necessary for the isolation of good quality nucleic acid (without inhibitors) and in sufficient quantity required by the test (sensitivity).

In certain contexts, samples collected in certain matrix can naturally result in a low yield of the target genome and would necessitate the use of an extra-step for raising its concentration so that the sensitivity of the technic could be guaranteed. This is done through enrichment of viral particles for which protocols varies

according to the type of target genome (ssRNA, ds-RNA, total RNA, circular DNA), its physical properties (naked RNA of viroïds, encapsidated RNA/DNA) and the type of matrix considered (plants, soils, etc.) (EPPO standard PM7/98 2021). For instance, the ultracentrifugation, the ribosomal RNA depletion, the cellulose affinity chromatography or the rolling-circle amplification are applied for improving the detection of plant viruses by HTS technologies.

- The performance criteria of the test

Diagnostic tests have different performances for detecting a virus. These performances in return, can be influenced by various contextual factors that are highly recommended to consider before taking a decision on the test to use. Generally, assessing diagnostic tests includes taking into consideration their capacity to deliver reliable test results. The (EPPO standard PM7/98 2021) suggested criteria that need to be considered while assessing the performance of a particular test as presented above. The inherent capacity of a test to perform better an analysis (analytical specificity and analytical sensitivity) should be envisaged together with its capacity to guarantee a better assay or diagnosis (the selectivity: diagnostic specificity and diagnostic sensitivity) (Saah and Hoover 1997).

Analytical sensitivity represents the smallest amount of substance in a sample that can accurately be measured by an assay while analytical specificity refers to the ability of an assay to measure one particular organism or substance, rather than others, in a sample. On the other hand, the diagnostic sensitivity is the percentage of samples known as infected by a particular disease and who are diagnosed by the assay as positive to that disease while the diagnosis specificity is the percentage of samples that are not infected by a disease who are identified by the assay as negative for that disease. Other criteria such as the repeatability, the reproducibility and the robustness need also to be considered.

There have been much discussion about the sensitivity of various detection tests. However, the sensitivity of many of the serological and nucleic acid-based tests is adequate for most purposes. Generally, the ease, the speed of operation, and the low cost favours procedures based on immunological or nucleotide hybridization techniques such as LAMP, LFD, ELISA, dot blots (J. A. Tomlinson et al. 2013). However, the sensitivity of the PCR supersedes that of ELISA or blotting. Additionally, the specificity of the PCR test depend on the used primers. This fact offers another advantage to the PCR over immunological tests since primers can be easily improved whereas antibodies are difficult to change.

- Laboratory facilities, equipment and information technology infrastructure

The importance of the laboratory facilities in generating reliable results is increasingly its capacity to avoid contaminations during diagnosis. For simpler diagnostic procedures such as immunological or nucleotide hybridization techniques, the level of requirements make them adapted to a wide range of conditions. However, for complex methods involving multiple handling steps and the use of many different reagents, as is the case for some molecular and HTS techniques, the possibility of contamination become a serious problem. The

sources of contaminations and how they can be avoided/reduced have been previously discussed and insightful recommendations could be accessed through (EPPO standard PM7/98 2021; Massart et al. 2019). For instance, the possibility to apply a forward workflow with dedicated areas for non-compatible steps such as nucleic acid extraction and amplification (Benedicte Lebas et al. 2022).

The HTS technology requires significant investment in terms of information technology for both data storage (number of samples, volume of data per sample, data backup and maintenance etc.), and computing capacity (operating system environment, computing power or server, level of expertise etc.) that makes it suitable only in particular environments (Lebas et al. 2022). For instance, all sequencing platforms are regularly updated and one would consider monitoring these updates and envisage the capacity of the laboratory to access them and evaluate their potential impacts on the results.

- Expertise available and personnel requirements

Most of the basic diagnostic techniques (ELISA, PCR and blotting techniques) are relatively easy to learn, but it is important that they are learned properly so that potential sources of error can be recognized. Several resources and protocols have been optimized previously for the diagnosis of cassava viral disease and are available (Kumar, P. L. and Legg 2009; Girma et al. 2017; Diekmann et al. 1994).

Some diagnostic techniques such as those described for immunological and hybridization can be conducted by basic technicians well trained. Others, such as molecular diagnostic tests, require that only qualified and trained personnel should process the samples. Requirements becomes of great concern for HTS techniques for which additional expertise in IT bioinformatic for sequence analysis, as well as a relevant scientific expertise is needed for the choice of biologically specific settings and parameters, for appropriate interpretation of the data and for evaluating their biological relevance are required. The capacity to run bioinformatic pipelines includes installation, development, validation, routine use and regular update of the software and database (Lebas et al. 2022).

Even though some specific expertise can be outsourced when considering adopting the HTS diagnosis, routine methods for diagnosis using immunological and hybridization methods are less demanding in term of personnel expertise.

I.6. Cassava seed systems and mitigation of disease outbreaks in Africa

I.6.1. The seed systems paradigm and farmer's perceptions on the value of seeds in Africa

The term 'seed system' is understood in various ways. It is often understood as referring to the organized, formal mechanisms through which farmers obtain seed and

through which seed quality can be guaranteed. This "formal" conception draws seed systems as a chain of interlinked activities, starting from genetic resource management, breeding research and crop improvement, through seed multiplication, marketing and distribution, to the use of the seed by farmers (Louwaars, Le Coent, and Osborn 2010).

In the reality, and parallel to this formal approach, seeds can travel through distinct routes, especially when they are produced and/or exchanged by farmers following pathways that they entertain themselves (local), sometimes since a long time ago (traditional). The importance of these 'informal' systems varies according to the type of crops and environments, being most prominent for Root Tubers, Banana (RTB) and other vegetatively propagated crops (VPCs), in developing countries (Louwaars, Le Coent, and Osborn 2010).

The basic expectations from (or purpose of) seed systems, no matter how formal or informal they are, are to better address farmers' demand for varietal traits (good quality planting material) and improve the distribution schemes to reach farmers better (accessibility) (Bentley et al. 2017). The advantages and limits of each of the formal and informal seed systems have been discussed previously (McGuire and Sperling 2016) and most arguments advocated for an approach favouring the integration of both systems so that the advantages of one can alleviate the limitations of the other.

More than consisting in formal and informal sectors, seed systems can operate on different geographical scales and levels, and integrate sub-systems like project-based seed systems (Gibson 2013; Rachkara et al. 2017).

Many studies on seed systems distinguish between formal and informal systems, with the former following more or less the model of an industrial supply chain, while the latter entail a range of mostly farmer-managed activities, for example in saving, using, exchanging and selling seed in local networks and markets (McGuire and Sperling 2016). However, the division of seed system actors and components into formal and informal categories appears problematic since there is a growing degree of overlap between both systems allowing, for example, varieties originally developed by the formal sector to enter the informal and *vice versa* (Christinck, Rattunde, Kergna, Mulinge, and Weltzien 2018).

Niekerk and Wynberg (2017) elucidated the various reasons, perceptions and attributes that farmers dedicated to seeds in South African context (Table I-6). He showed that these attributes could be summarized in Physical, financial, social and cultural (Table I-6).

Table I-6. Farmer's perceptions of the value of seeds for traditional crops (van Niekerk and Wynberg 2017)

Attribute	Farmer's perceptions
Physical	Crops grown from traditional seeds are hardier than shop-bought crops
	Traditional seeds last a long time in storage without chemical applications
Financial	Saving seed means seed does not have to be bought every season
	Exchanging seed precludes the need to pay for seed
Social	Historical ties – seeds are passed on from one generation to the next
	Seed saving encourages self-sufficiency
	Sharing traditional seed enhances social cohesion
Cultural	Seeds were being taken care of on behalf of the ancestors
	Seed from the household of a person in mourning cannot be removed until certain rituals have been performed and a specific period of time has passed
	Seeds are associated with marriage rituals: seed either accompanied newly married women to their husband's family home (Lindizwe) or were not allowed to travel with a newly married woman to her in-laws' homestead

I.6.2. Cassava seed systems in D.R. Congo

As is the case in many African countries, the formal and the informal seed systems characterize the seed system in D.R. Congo (Figure I-22). The informal system, where the farmer acquires and multiplies seeds himself, is the most predominant. On the other side there is the formal system, where both public and private institutions intervene, and that is based on: (i) varieties establishment by the research institutions, (ii) seed multiplication and dissemination following established regulatory conditions, (i) strict control of multiplication and dissemination activities by a public institution.

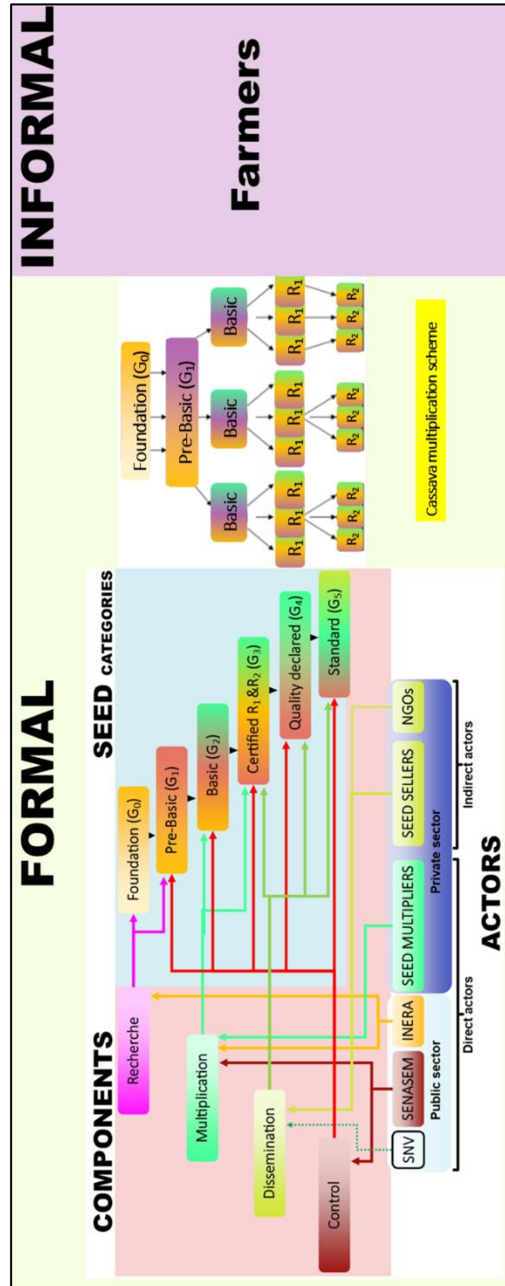


Figure I-22. Schematic representation of the organizational framework of seed system in D.R. Congo and it particularity in relation to cassava multiplication scheme.
This figure was elaborated using information collected from (A. N. Frangoie, Bidiaka, and Mahungu 2019b; Mahungu et al. 2014; N. Frangoie, Bidiaka, and Mahungu 2012).

Four main components, led by five main actors, are involved in the production of six categories of seeds that flows inside the formal seed sector. These actors can

be separated into public or private actors, but importantly their intervention can be categorized into direct or indirect:

Direct actors are those intervening in the:

- **Research**: This component is at the first level of the seed sector and is in charge of developing novel and improved varieties designated as "foundation seeds" (pre-basic, G0) that serves as the starting point for subsequent categories of seeds. Before being approved and added to the national catalog of varieties, its quality is first assessed following the DHS (Distinction, Homogeneity and Stability) and VATE (Agronomic, Technological and Environmental Value) tests. If evaluations of the foundation seed are successful, it will be added to the national varietal catalog that will detail information regarding the morphological, agronomic and technological traits of approved varieties as well as the regions where they can be grown. This national catalog of varieties serves therefore as a working document tool for seed producers as well as for the seed inspectors. The last version of the national variety catalog (2019) contained 214 approved varieties among which 35 of cassava (see Table I-6). The research also assumes the multiplication of the foundation seed into basic seeds (G1), using the *In vitro propagation* technic for cassava, and ensure their varietal identity is maintained throughout various cycles of multiplication and dissemination. Over the years, the research component has been exclusively assumed by the INERA but the whole formal seed sector has benefited from several supports made up of donor-assisted projects. In that way, since 1972, the INERA has been strongly supported by the IITA and CIAT, especially in the cassava seed sector where it has been appointed by the D.R. Congo government to mitigate the spread of cassava bacterial blight disease. Since then, the IITA has initiated various projects in collaboration with the INERA in various sectorial programs of the ministry of agriculture such as the "Programme national Maïs (PNM)", the "Programme National Legumineuses (PNL)", and importantly the "Programme National Manioc (PRONAM)". These various programs resulted in the creation of the project entitled the "Projet de Recherche Agronomique Appliqué et Vulgarisation (RAV)" which resulted in the development of improved and resistant varieties including cassava varieties resistant to cassava mosaic disease (Table I-6).
- **Multiplication**: Once approved and included in the national seed catalog, the G1 varieties go through a process of multiplication in order to generate large seed quantities necessary for satisfying farmer's needs. The stages consist in multiplying pre-basic seeds into basic seeds (G2), then into two generations of certified seed (R1 & R2 seeds, both the G3 stage), quality declared as well as standard seeds. Various actors intervene in this component, starting from the INERA again, which act as a true seed enterprise, and extend to individual seed multipliers often regrouped into cooperatives of seed multipliers. For cassava, the multiplication scheme is done following a "cascade" model that ends at the second generation of certified seeds (R2 stage). The various stages in the multiplication of cassava

consists into the primary (Pre-basic to basic), secondary (basic to certified R1), and tertiary multiplications (certified R1 to certified R2). Certified R2 seeds are those dedicated for use by farmers.

- **Control** : The aim of the control is to guarantee that the seed quality is maintained through the various multiplication and dissemination process so that the seeds that have been admitted to the catalog can reach the end users (farmers) with their fully declared potential. This control process is disrupted for a variety when it is removed from the varietal catalogue. The control component has been long time under the responsibility of the ministry of agriculture through its specialized official control service termed "SENASEM" (Service National des Semences). Through its network of inspectors and analysts, this service conduct a routine monitoring and control of all seed multiplication activities starting from the pre-basic to the standard seeds. This control is conducted on various levels of the seed sector, starting from the field (during the crop vegetative cycle) to laboratories (tests analyzes are conducted on samples collected on various products in storage pending commercialization or distribution). They deliver a "certificate of conformity" to seeds that have been multiplied, packaged, stored and disseminated under required conditions. This service is competent in controlling and certifying seeds from the pre-basic level (G1) to the standard level (G5). These conditions are described in a special document as described below.

On the other side, indirect actors serve as the support to the seed sector and are constituted by :

- ❖ **Dissemination**: Various local and international organizations specialized in development or extension: they support seed multipliers and various public sector in their activities of multiplication and control.
- ❖ **Commercialization organizations**: they are specialized in the packaging, the storage, the distribution and the sale of seeds.

The research and development of new cultivars is a core process necessary for maintaining and improving the crop productivity. There have been nearly 30 years of varietal releases from the National Cassava Program (PRONAM) managed by the INERA with strong linkages to IITA and CIAT. PRONAM'S main cassava breeding program is located at MVUAZI in the Bas-Congo with a major secondary station at Kiyaka in Bandundu. This program provided breeder level and certified clonal planting material for bulking by a range of international and national NGOs, FAO, multilateral, and donors projects, certified seed multipliers, and commercial farms. The core process is the breeding and release of higher productivity clones with good pests and disease tolerance that are adapted to the domestic market.

Within the 35 cassava varieties actually approved for multiplication and dissemination in the national catalog of varieties, most of them are of sweet type (25) but we can also note biofortified varieties (4) breeded in collaboration between the INERA with the HARVESTPLUS and IITA. The oldest introduced varieties that are still maintained in the catalog date from the 1998 (MVUAZI and

RAV), whereas the recently introduced one date from 2019 (MUKOLESHI). Most of the varieties were locally bred by the INERA (18) whereas introductions of some varieties from Uganda (6) and Nigeria (9) by the IITA are also registered. The variety maintenance is done by the various INERA stations that are disseminated throughout the regions where the variety is widespread. It seems like in D.R. Congo, the cassava breeding program has been importantly influenced by the cassava mosaic disease as most varieties (26) are identified as having a "very good" level of resistance against the disease. This is not the case for the CBSD as only certain varieties are reported as having a "very good" level of resistance (4 varieties) or "tolerant" (9 varieties) (SENASEM 2019) (see Table I-7. Below).

Table I-7. Cassava varieties approved in the national catalog of varieties in 2019 (SENASEM 2019). *For further varietal characteristics, refer to the catalog.*

N°	Code	Name[1]	Year of introduction	Geographic origin	Genetic origin	Resistance to diseases
1	M98/115	BOMENGO	2013	D.R. Congo	IITA	Good against CMD, CBB and CBSD; tolerant against CAD
2	MVZ 99/0395	BUTAMU	2004	D.R. Congo	INERA M'VUAZI	Tolerant against CMD, CAD and CBB
3	MVZ 2008/223	GIMBI	2016	D.R. Congo	-	Very good against CMD, CBB and CAD
4	GKA 006/033	KAMANA NZALA	2006	D.R. Congo	INERA NDANDAJIKA	Very good against CMD, tolerant against CBB and CAD
5	-	KIZIMBANI	2014	Tanzania	-	Mild to CMD, CBB, CAD, and CBSD
6	MVZ 2006/114	LITOY	2013	D.R. Congo	IITA	Very good against CMD, CBB and CBSD; tolerant against CAD
7	ARNET, MM96/028	LIYAYI	2001	Ouganda	IITA	Mild against CAD and CMD
8	MVZ 2008/0305	LONGO-LONGO	2006	D.R. Congo	-	Very good against CMD, CBB and CAD
9	1004/024	LUBILANJI	2001	D.R. Congo	INERA M'VUAZI	Very good against CMD, CBB, CBSD and CAD
10	MVZ 92/298	MAHUNGU	2000	D.R. Congo	INERA M'VUAZI	Tolerant to CBB and CMD, mild against CAD
11	MM 96/7752	MAYOMBE	2004	Ouganda	IITA	Tolerant to CMD, CBB and CAD
12	I92/0067	MBAKANA	2008	Nigeria	INERA M'VUAZI	Very good against CMD, CBB and CAD
13	MM 97/2015	MUGOLI	2008	Ouganda	IITA	Very good against CMD, tolerant to CBSD, CBB and CAD
14	I 92/326	MUTIENE	2013	Nigeria	IITA	Very good against CMD, tolerant to CBB, CAD and CBSD
15	2006/073	MUZURI	2013	R.D. Congo	IITA	Very good against CMD, CBB and CBSD; tolerant to CAD
16	MM 96/4653	NABANA	2008	Ouganda	IITA	Very good against CMD, tolerant to CBSD, CBB and CAD
17	MM 96/7204	NAMALE	2008	Ouganda	IITA	Very good against CMD, tolerant to CBSD, CBB and CAD
18	MVZ 99/150	NGANDAJIKA	2008	D.R. Congo	INERA M'VUAZI	Very good against CMD, tolerant to CBSD, CBB and CAD
19	I96/0160	NSANSI	2000	Nigeria	IITA	Tolerant to CMD and CAD
20	MVZ 2010/061	NZOAZUZU	2017	D.R. Congo	INERA M'VUAZI	Very good against CMD, CBB and CAD
21	TME 419	OBAMA	2008	Nigeria	INERA M'VUAZI	Very good against CMD, tolerant to CBB and CAD
22	2001/014	OBAMA 2	2015	Nigeria	INERA M'VUAZI	Very good against CMD, CBB, CAD and CBSD
23	MVZ 85/297	RAV	1998	D.R. Congo	INERA M'VUAZI	Tolerant to CAD and CBB
24	MM 96/3920	SAWASAWA	2003	Ouganda	INERA MULUNGU	mild to CMD and CAD
25	MVZ 99/038	ZIZILA	2003	D.R. Congo	-	Very good against CMD
26	I 96/0211	DISANKA	2004	Nigeria	IITA	Very good against CMD, CBB and CAD
27	MVZ 2007/126	ILONA	2013	D.R. Congo	IITA	Very good against CMD, tolerant to CBB, CAD and CBSD
28	MVZ 2007/102	KANSAKAKO	2015	D.R. Congo	IITA	Very good against CMD, CBB,CAD and CBSD
29	-	MVUAZI	1998	-	IITA (I95/0528)	Very good against CMD, CBB and CAD
30	I 94/0330	VUVU	2008	Nigeria	INERA M'VUAZI	Very good against CMD, CBB and CAD; tolerant to CBSD
31	2001/1229	WINA	2008	Nigeria	IITA	Very good against CMD, and CBB, tolerant to CAD and CBSD
32	GKA (2012/149)	LUMONU	2012	D.R. Congo	INERA NGANDAJIKA	Very good against CMD
33	MVZ 2011B/0360	VIMPI	2019	D.R. Congo	INERA M'VUAZI	Very good against CMD, tolerant to CBB and CAD
34	I 2001/1661	KINDISA	2008	Nigeria	IITA	Very good against CMD, CBB and CAD; tolerant to CBSD
35	GKA 2011/274	MUKOLESHI	2019	D.R. Congo	INERA NGANDAJIKA	Very good against CMD, tolerant to CBB and CAD

[1]The name correspond to that used by extension services. Sweet varieties are highlighted in orange, bitter varieties in green, and biofortified varieties in blue. IITA: International Institute of Tropical Agriculture, CMD: Cassava mosaic disease, CBSD: cassava brown streak disease, CBB: cassava bacterial blight, CAD: cassava anthracnose disease.

The standards and conditions that govern the production and the control activities within the seed system in D.R. Congo are contained within a special document: the "règlement technique de la production, du contrôle et de la certification des semences et plants" (translated as "the technical regulations for production, control and certification of seeds and seedlings"). Four different versions of this document have been issued since 1997, and the last one date from 2019. It is only within this last version that the phytosanitary standards related to the CBSD and the CRND have been integrated for consideration among the conditions to be fulfilled by a multiplication field (Table I-8).

Table I-8. Technical regulations for production, control and certification of seeds and seedlings of cassava. *Edition 2019. Excerpt of the phytosanitary conditions regarding diseases and pests*

Phytosanitary status for diseases (maximum incidence % allowed)	**Seeds categories**		
	Pre-basic	***Basic***	***Certified (R1 & R2)***
9 Cassava mosaic disease (CMD)	0.1	0.1	0.5
10 Cassava bacterial blight	0.1	0.1	0.5
11 Cassava anthracnose disease (CAD)	10	10	10
12 Cassava brown streak disease (CBSD)	4	4	10
13 Cassava root necrosis disease (CRND)	10	10	20
Phytosanitary status for pests (maximum incidence % and damage severity=3)			
14 Cassava green mites (CGM)	10	10	10
15 Cassava mealybugs (CM)	10	10	10
16 Cassava roots and tubers mealybug (CRTM)	10	10	10
17 Thrips	10	10	10

In fact, it is suggested that following the inspection of 100 plants within a multiplication field, the limit in the incidence of root necrosis symptoms due to CBSD should not exceed 4% for the pre-basic and basic seeds and 10% for the first and the second generations of certified seeds (R1 and R2). There is no criterion related to the diagnosis of CBSD symptoms on above-ground parts (stems and leaves), suggesting that multiplication fields should stay longtime enough in the field until roots are developed to allow symptoms diagnosis.

As the phytosanitary status has to be controlled not only during the seed multiplication process, but also during the varietal maintenance, the *In vitro culture* strategy is the sanitation method envisaged for cassava especially for regenerating plant material previously infected by viral disease (Figure I-23). Apart from the visual diagnosis of CBSD symptoms on below-ground parts of the plant, the technical regulation scheme does not mention an additional diagnostic or detection method to be further conducted to ensure the phytosanitary status of multiplied plants is safe of viral diseases.

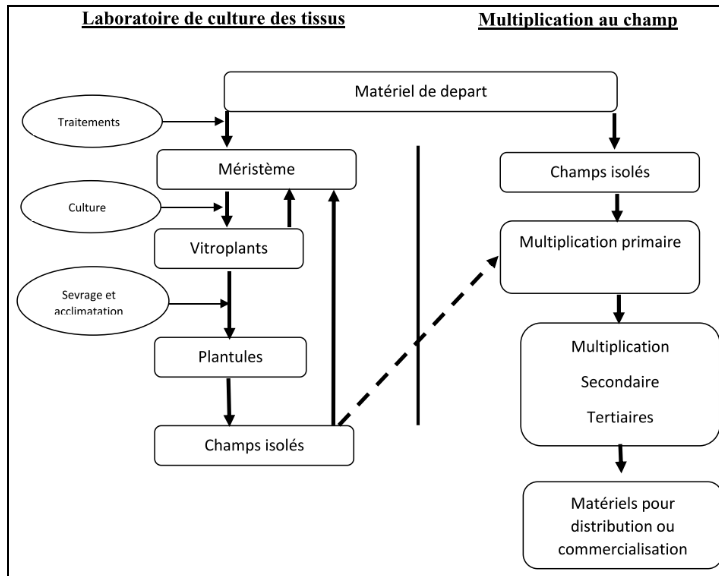


Figure I-23. Cassava variety maintenance scheme in D.R. Congo

Additionally, the technical regulation scheme suggests a number of operations to conduct in case putative plants are detected as being not conform to phytosanitary standards. These operation are collectively designated as "normes d'épuration du manioc". They are highlighted as below:

- ❖ Uprooting of plants infected by viral, bacterial and fungal diseases as soon as the symptoms appear;
- ❖ The uprooting must be complete and the plant debris immediately removed from the field;
- ❖ Efficiently getting rid of weeds, especially during the first four months of plant growth;
- ❖ Discard diseased plants as well as those not well developed within the mini-cutting nursery

The necessary steps for the approval and the registration of novel varieties in the national catalog of varieties are detailed within a special manual: the "Manuel des procédures administratives et techniques de prestation des services du SENASEM" (translated as: "manual of administrative and technical procedures for the SENASEM service provisions") edited in three volumes. This manual was adapted from that used by the Southern African Development Community (SADC) and allow that varieties approved in D.R. Congo can be also easily approved within the SADC countries.

We have presented above a brief summary of the formal seed system as organized by the D.R. Congo regulations. However, as highlighted in the introduction of this subsection, in the reality the formal seed industry in the country coexist in parallel with the informal seed system. Whereas the formal seed sector

is focused on the selection and evaluation of improved varieties as well as the production and trade of certified seeds, the informal seed sector is thought to be governed by traditional knowledge and norms and little information is available about its performance. The importance of this informal seed sector lies in the fact that a vast majority of farmers depends on it for various reasons including a limited knowledge, a restricted varietal diversity, a lack of resources to afford certified seeds as well as the lack of accessibility.

Farmer self-propagation and exchange of cassava cuttings is the primary form of cassava propagation by smallholders and commercial farmers in the informal seed sector. However, many projects and programs have worked with farmer demonstration groups to train them in careful selection of vigorous, disease-free planting material. Some producers groups and NGOs have cassava multiplication fields and serve as independent or contract growers producing planting material.

I.6.3. Understanding interventions in cassava seed system in a context of viral diseases burden: a brief analyse of the Great Lakes Cassava Initiative (GLCI) experience.

Various types of interventions have been directed for improving the functionality of seed systems of various crops across the world. For RTB particularly, several authors have suggested that most of the interventions directed to seed systems of these crops did not reach their full potential for various reasons. The most prominent one was shown to be a neglected rigor when it came to focusing on understanding issues resulting from the adequation between the type of the work or intervention, the type of actors, the specificities of the environment (where) and the underlining reason or motivation supporting these interventions (Almekinders et al. 2019)

In the past, an interesting intervention specific to the seed system of cassava has been regionally implemented to mitigate the burden of viral disease (CMD and CBSD) in the African context. This project, known as the GLCI, has generated a number of lessons that we considered here as important to document in the specific frame of understanding the challenges posed by cassava seed system interventions in the context of disease burden. This simple analytical case study was carried out by exploiting available documentation related to the project itself and only some key features, considered as relevant in the context of this work, will be presented.

a. Introduction

Maybe one of the most important interventions ever realized so far on cassava seed systems in the context of an existing and emerging viral diseases (CMD and CBSD respectively), the Great Lakes Cassava Initiative was a project aiming at distributing clean planting material of disease-tolerant or resistant varieties to 1.15 million farmers across six countries in the great lakes region: Burundi, D.R. Congo, Kenya, Rwanda, Tanzania and Uganda.

Implemented for a period of 57 months (from 12/1/2009 to 8/31/2012), with a 23,876,316 U.S. $ grant from the Bill and Melinda Gates foundation (BMGF), this project was implemented by the Catholic Relief Service (CRS) with research and technical support from the Africa-based International Institute of Tropical Agriculture (IITA) and the UK-based Food and Environment Research Agency (FERA).

Before the GLCI, the CRS had already gained experience in managing regional interventions in seed systems of RTB in the Great Lakes Region of Africa. Several important results from its two years-USAID funded C3P project (Crop Crisis Control Project, 2006-2008), multiplied and disseminated CMD resistant varieties and developed various strategies and tools to intervene in seed systems: the use of vouchers for dissemination of planting material, the development and use of a quality management protocol, and dissemination planning to promote participatory and transparent dissemination at community multiplication sites.

b. Organization of the project

The sustainability of interventions in cassava seed system was envisaged at three levels (Figure I-24): (i) research, (ii) development and (iii) program management. At the research level, the project conducted various epidemiological studies to identify the best growing places as well as the best mechanisms of delivery of clean seeds. In addition, the research aspect of the project provided tools necessary for diagnosing the disease and detecting viruses so that best decisions on the phytosanitary status of seeds could be taken.

The development aspect of the project was organized around the identification of disease resistant clones, their multiplication and dissemination, as well as the awareness raising about the disease.

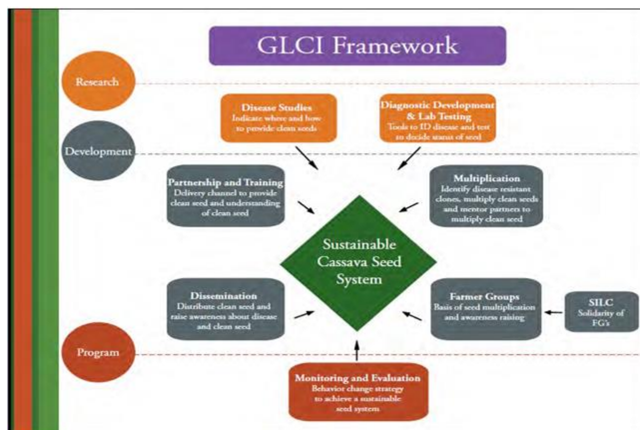


Figure I-24: The GLCI operating framework

c. Seed system innovations within the GLCI and its challenges

The GLCI multiplication and dissemination system was organized in a hierarchical manner: small number of primary sites feeding a larger number of

district-level secondary sites and providing planting material for thousands of community level multiplication sites, referred to as tertiary sites.

With the advent of the GLCI project, CRS capitalized on the achievements from the C3P and build up a set of four inter-dependent innovative approaches to intervene in cassava seed system: (i) decentralized production and dissemination, (ii) quality management protocols, (3) targeted dissemination and traceability in the seed system, and (iv) mitigation against CBSD through surveillance, sampling and testing.

- The decentralized production and dissemination approach consisted in establishing many small multiplication plots (from ¼ to 1 ha), as opposed to fewer larger plots, spatially allocated in target areas (10 km apart in the intervention zone). Additionally, the approach involved allocating small amounts of planting material (20-25 full stems corresponding to 100-150 cuttings, each having 3 to 5 nodes) per farmer, enough to have a demonstration effect at farm level.

  This approach, also called the "***small is beautiful***" had the main benefit of: effective strategy in moving planting material quickly into target zones and to gain geo-spatial saturation, making access easier for farmers, effectivity in supporting training and having the production site serving as a classroom, thus raising the demonstration effect of planting material, reducing the loss of planting material which occurs between harvest and planting, and reducing production costs at multiplication sites. Additionally, smaller more decentralized fields promoted the transport of planting material in full stems on farmers heads in a single bundle as opposed to cuttings or mini-stems in bags loaded on vehicles.

  The main challenge of this approach was that it was predicted that most material was disease resistant (against CMD, the most prevalent disease at the time of project inception). Nevertheless, the germplasm resistance to CBSD was not clear and, in CBSD endemic zones, this approach became inappropriate as multiplication sites would require field isolation and quarantine against CBSD viruses. Additionally, this decentralization increased cost of monitoring production sites, bringing the cost per direct stem recipient to 16,80 US $ during the period of project implementation.
- Quality Management Protocol (QMP): The aim of QMP was to apply a seed quality standard for cassava which was relevant to farmer conditions and could be conducted at field level visually and at low cost. It represented an opportunity for public sector engagement in setting of standard. QMP is based on visual assessment of varietal purity, scoring for pest and disease, and estimating the total stems to be harvested. One hundred plants per field were involved in the inspection and a subsample of 10 plants are taken for root symptoms assessment

  In the context of GLCI, the utility if the QMP have been limited and challenged by the knowledge gaps in the epidemiology of CBSD. In addition, various factors prevented the application of the QMP approach at large scale: multipliers not wanting to sacrifice ten plants for CBSD assessment, GLCI partners not being trained in QMP, GLCI multipliers not being required to do QMP, and a general low level of sensitization and buy-in from local government.

Additionally to this, a central challenge in all GLCI countries remained the capacity of NPPOs to promote and coordinate the application of minimal phytosanitary standards since they lack the resources, staff and capacity to support even a rudimentary framework to oversee inspections of cassava stem production fields. This economic basis and the costs to implement certification and quality control should be significantly lower than the cost of no standard and the benefits of functioning standards. Even with a strong economic argument to justify public sector investment in QMP, it remained highly unlikely to continue in any form without significant donor funding..

- Targeted dissemination and traceability: This approach was required since an efficient and well targeted dissemination of disease tolerant planting material necessitated a special care because disease tolerant cassava planting material was highly sought in areas impacted by disease. Additionally, given the role of cassava as a food crop for vulnerable farmers, explicit efforts to reach disease-affected vulnerable farmers with tolerant planting material deserved to be made. Therefore, systematic processes for identifying communities and farmers to be served with planting material and having a seed supply chain that recorded the origin of material was developed.

The dissemination process was underpinned with three values:

i. Accountability: included being capable of demonstrating to partners, governments and donors who received planting materials, when, how much and from where the material is received and how far the recipients travel to receive the material.

ii. Transparency: referred to having clear and documented processes for allocating material and ensuring that the processes are well understood and practiced at partner level and with all sites of multiplication.

iii. Participation : referred to having strong community engagement in decisions on allocating planting material and involving local government, civil society leaders and the multipliers in discussions on who Is targeted to receive planting material.

Dissemination plans were used for allowing a transparent discussion with local authorities on the amount of planting material to be disseminated, the target villages, and criteria for beneficiary selection. Furthermore, dissemination reports documented the extent to which marginal groups were served by the project. The use of a voucher system allowed farmers to access to cuttings in an ordered manner during seed dissemination.

The main challenges of this seed dissemination and traceability strategy relied on the uncertainty of its adaptability since to replicate them by governments or other donors require a certain level of awareness of some best practices and approaches.

- Mitigating against CBSD using surveillance, sampling and testing: The main threat for the GLCI multiplication system was the spread of viruses that cause CBSD. The CBSD risk assessment mitigation in GLCI involved surveillance and testing which was donor mandated and not based on farmer cost-benefit analysis. To minimize the risk of spreading CBSVs through the GLCI multiplication

scheme, primary sites and most secondary sites in CBSD-threatened areas were tested using RT-PCR tests. CBSD threatened areas were determined by surveillance activities undertaken through disease annual surveys conducted by IITA. The most known “zero tolerance approach” consisting in using only source sites having no positive virus test results for secondary or tertiary sites was used.

While the process of linking surveillance to testing was unique and had a number of benefits, the investment cost of the GLCI CBSD risk mitigation process was not linked to any demonstratable economic benefits. For example, the estimated total cost of all the 45,000 RT-PCR reactions conducted throughout the project was $300-400,000. Additionally, the annual disease surveys did not provide results fast enough to support the early decision-making on CBSD-threatened or epidemic zones. Even with the advent of cheap and accurate diagnostic testing suitable for field application (LAMP, NASH), the necessity of having a systematic process to apply these detection methods remained the most important limiting factor. This systematic process had to combine some surveillance or disease reconnaissance to justify the testing scope, some analysis of the cost-benefit ratio to ensure that seed producers and seed consumers have economic incentives to support the system, and some buy in and backstopping from public sector actors in each country (NPPOs and National Seed Certification Associations).

d. Learning from the GLCI project implementation and coordination

The implementation of the project was envisaged around five main axes : the partnership, the disease, the training, the seed and the farmer groups.

Owing to its character of a regionally implemented project (6 countries), the partnership strategy adopted by the GLCI could be seen as a “wide machinery” considering the number and the complexity of stakeholders and partners involved more than (200 partner’s staff globally). This partnership involved local (55 local partners), national (NARS and NPPOs), regional (IITA), and international (FERA) actors.

The Figure I-25 shows the typical organizational model through which the system was coordinated and the respective roles and responsibilities attributed to each partner. Through various events (meetings, workshops and conferences, trainings, exchange visits and end-of-project events), these partners were bring together for sharing information, comparing experiences, reviewing and planning activities.

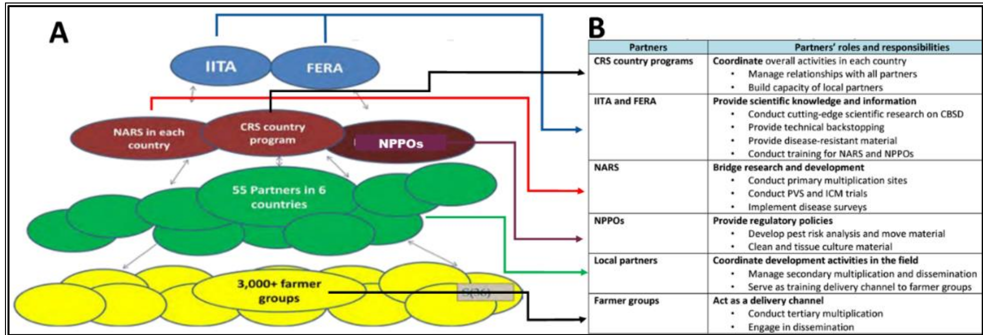


Figure I-25: Horizontal and vertical organizational web of GLCI partners (A) linked to their corresponding roles and responsibilities. Source: GLCI Final report, 2012.

In that model, crucial research studies as well as disease diagnostics were conducted by IITA and FERA research institutions at the top of the organizational web. Results and recommendations generated through their studies were channelled to NARS and NPPOs via the CRS regional team and country programs to assist them in making decisions on multiplication, cleaning and disseminating planting materials. Research institutions were also responsible of conducting training for NARS and NPPOS. IITA collaborated with NARS by providing with them necessary training and technical backstopping. In return, IITA relied on NARS to carry out the work in the field, particularly for implementing annual surveys and lab-testing of samples. FERA coordinated the NPPOs to develop the regional Pest Risk Assessment (PRA) necessary for developing individual country-based PRAs.

The roles of NPPOs were not as broad as those of the NARS since the first were limited to developing PRA and the associated guidelines for planting material movements within the framework of risk analysis while the NARS were considered as a bridge between research and development. In fact, apart from the collaborations with the IITA, the NARS had the responsibility of (i) coordinating, training and supervising partners and farmers groups to conduct participatory Varietal Selection (PVS) and integrated Crop Management (ICM) trials, and (ii) providing and multiplying foundation material from their primary fields to start the multiplication chain.

The local implementing partners (55) were considered as the engine of the project implementation in the field and the key to success. They were responsible of (i) conducting secondary multiplication sites, (ii)implementing PVS and ICM trials, (iii) establishing and training farmers groups, (iv) supervising tertiary sites managed by the farmers groups, (v) organizing dissemination of planting material, and (vi) collecting and entering data.

The farmers groups (3,048) were considered as the delivery channel and the pivotal point to reaching scale. They managed the 5,500 tertiary multiplication sites necessary for reach the 1.35 million beneficiaries.

Significance of the GLCI partnership strategy to reaching its goal: how partners evaluated the roles of each others within the partnership set up by GLCI

Through the project implementation, all actors interacted together to reach the goal assigned to the project. They are therefore the key informers who can reflect on the consistency of the partnership strategy adopted through the perceptions that they had on the roles played by each partner in the consortium. The importance of this evaluation lies the fact that there can be demarcation between what the project has assigned as role before the implementation and what really happened during the implementation. The summary of the views that various GLCI's partners had about their mutual relationships within the partnership consortium as well as the challenges encountered and recommendations for moving forward is presented in the table I-9.

Table I-9. Partners views on the importance of each others in the implementation of the project and institutional benefit gained through the GLCI project

		International Institutions		**National institutions**		**Local partners**	Institutional benefit from GLCI
		IITA	FERA	NARS	NPPOs		
International Institutions	IITA			This partnership was "productive and worthwhile", IITA supported NARS and provided capacity building "smoothly and effectively"	The work was essential	The benefits were mainly related to the local partner's interface with the farmers: "they implement the science, guidelines, and policies set by us", "they reach farmers/beneficiaries in the way we cannot", "they increase the relevance and quality of our work", and ""as the interface with farmers, they facilitate testing and provide feedback on our science, guidelines, policies"	"The partnership broadened IITA's partnership for cassava research and development"
	FERA			GLCI provided a good basis on which to further the capacity of the NARS in CBSV research	GLCI "provided a good basis on which to further the capacity of NPPOs in CBSV phytosanitation and quarantine		GLCI has furthered its "reputation as an internation centre for research and risk evaluation
National institutions	NARS	Appreciated IITA's "technical support particularly on annual survey and PVS activities" as well as the training, guidance, and "building capacity of NARS in disease diagnosis, especially in lab implementation and training"			Do not seem to see much point of the role of the NPPOs: some deemed the relationship as "not necessary" while other "disregard them"	Local partners "need some training to understand the techniques they are administering" or "partnership with local partners was very challenging. Most of the partners on the ground were not aware of agricultural research activities before GLCI"	(i) Capacity building through trainings (virus cleaning, lab. detection, and field diagnosis), lab and office equipment; (ii) technical material development (improved cassava varieties, seed source for partners and farmer groups, enhanced germplasm exchange in the region).
	NPPOs		"Building our capacity in the testing of CBSV using molecular techniques	Appreciated more the collaboration with the NARS by considering them as a "useful partner"		Local partners strenghthened their role in "monitoring and inspection of seed fields f cassava cuttings to be healthy to provide farm households"	
Local partners		Gaining scientific knowledge		Gaining scientific knowledge and help orient work to national context (66 and 48 % respectively)			

Briefly, the information provided within the table I-9 shows that globally partners assessed the presence of other partners as important within the consortium. A sort of complementarity between the actors emerged as the main reason supporting the existence of the partnership. At one side there are international institutions considered as knowledge/technology providers but they needed implementors at national level and ultimately local partners as interface with farmers. The value of the national institutions was perceived as more related to the national contexts in that the NARS identified the national priority while the NPPOs provided the policy and regulatory environment. However, this complementarity was not fully supported at national level since it appeared that NARS institutions disregarded the role of NPPOs by judging their role as “not necessary” and by being more critical regarding the local partner’s technical ability to conduct participatory varietal selection (PVS) and integrated crop management (ICM) trials.

Additionally, international institutions have seen in the GLCI’s partnership an opportunity to broader their interventions and to further their reputation in cassava research development and risk evaluation. Local partners on their side, have gained through the GLCI the opportunity of capacity building as well as technical material development.

The following comments from the partners sum up the significance of the GLCI partnership to reaching its goal : ”These partnerships provided a platform for integrated research and development, from which many innovations were developed. These innovations established a rigorous seed system within the context of diseases, built a model to manage the large scale of field staff and farmer groups, and developed ana ambitious field-based M&E system”.

Constraints of the GLCI’s partnership strategy in the management of disease

The complexity of the GLCI project, owed particularly to its regional character, prompted it to collaborate with a consistent number of actors. As a consequence, the disease management approach adopted by the project was also regional. Surely this approach had its merits and advantages in addressing the disease issues “across the borders”. However, still this approach had some repercussions on the project performance.

The important role of local partners within the partnership, as recognized by other actors, was related to their direct contact with farmers. We, therefore, considered them as the best point to assess relevancy of the GLCI partnership strategy. We found within the project document reports, a table that supports our arguments (Table I-10).

Table I-10: The mechanism of the working relationship between local partners and other partners involved in the GLCI project. Assessment by local partners themselves

Mechanism of the relationship	IITA	FERA	NARS	NPPOs
Working directly with them within GLCI	46	8	26	18
Using results from their work	44	50	36	24
Provide feedback and inputs to them	4	12	40	46
Do not use anything from them	4	28	4	10

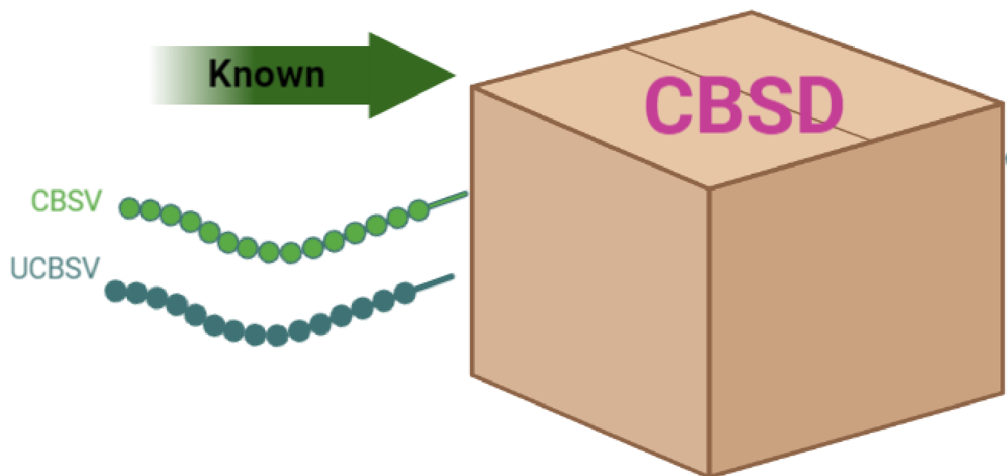
Results presented in that table reflect two aspects:

- ✓ National institutions did not played their roles as stated in the framework: a very low percentage of local partners suggested they worked directly with NARS and NPPOs (26% and 18% respectively).We see a kind of contradiction between what the project has planned as attributions and what have been implemented. This situation pointed to putative issues in the coordination of the partnerships which resulted into unclear definition of roles and responsibilities.
- ✓ An overall "lower" percentage of local partners could "use" the results obtained from their collaboration with various partners, increasingly with national institutions.

This information suggested that the partnership was not sufficiently effective in strengthening relationships at national level, including linkages between NARS and NPPOs, and between NARS and local partners. This is reflected by some of these facts retrieved in the final document report: "Though the local partners are more directly linked to the NARS, they seem to look more on IITA for scientific guidance". Additionally, the low usage of results could have originated from the combined weaker synergy between national actors and the lack of clarity in the definition of roles and responsibilities. This is exemplified by the following comment retrieved through the project documentation : "Some partners felt that there was no meaningful follow up on the PVS trials, which indicated a lack of clear definition of whose responsibility it was to follow up"

Chapter II

II. Objectives and thesis structure



Chapter II. Research gap, objectives and thesis structure

II.1. Research gap

II.1.1. The epidemiology of plant diseases for seedborne pathogens

Pathogens, hosts and their environment have been the three traditional components used by plant pathologists to investigate plant diseases (Agrios 2005). The interaction of these components, initially known as the "disease triangle", allowed the understanding of complex causes underlining the existence of plant diseases in ecosystems (Figure II-1) (Horsfall and Dimond 1960).

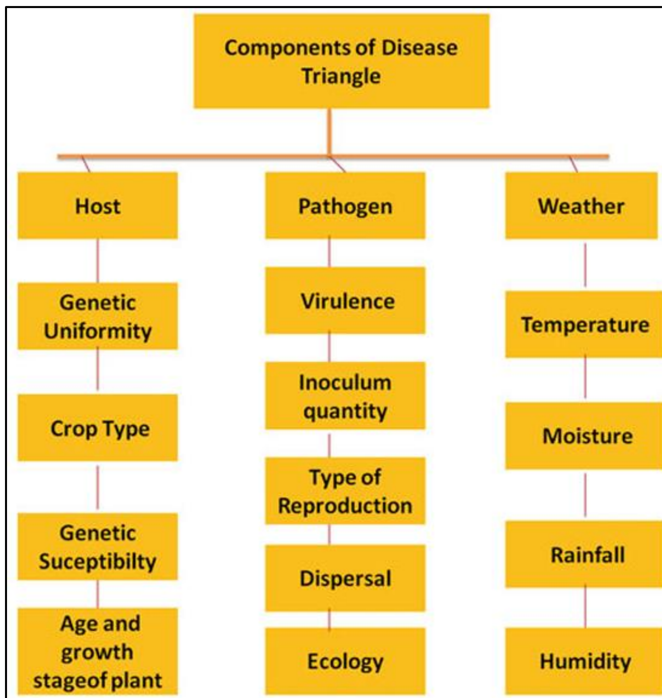


Figure II-1. Various factors of host, pathogen and environment involved in epidemic (Haq, Khan, and Sarwar 2022)

This "disease triangle" concept was adapted in an epidemiological perspective to timeline those interrelationships by adding a fourth component, the time. The concept of "disease tetrahedron" or "disease pyramid" therefore emerged (Agrios 2005; Francl 2001) (Figure II-2).

In fact, the progress in the understanding of plant diseases has revealed that interrelationships between the various disease components are variable according to the specificities of each component involved. For example, since the first elucidation that plant viruses could be "transmitted" by the intermediate of ***vectors***, notably insects (Takami 1901), their role in the epidemiology of viral diseases was recognized and as to accommodate the traditional "disease triangle", the concept "ecological trinity" of viruses, hosts, and vectors also termed the "virus disease triangle" mostly used in plant virus ecology and epidemiology emerged (W. Carter 1939).

Epidemiology and ecology of plant viruses are two important concepts in plant virology that have been confused for a long time as they were both based on the ecological trinity concept: virus epidemiology investigates diseases and factors influencing their spread and population dynamics, whereas virus ecology extends the focus to include understanding patterns of virus distribution and dynamics within a given environment, their effects on community and ecosystem properties, and the reciprocal effects of the environment on virus dynamics and evolution (Lefeuvre et al. 2019; Maclot 2021)

The changing social and environmental context that have marked the beginning of the 20th century has been accompanied by several issues that were scientifically not addressed until then. In fact, The increasing industrialization, the increase of demography, world trade, globalization; the movement of populations, global warming are practical examples of what became the new challenges for the plant epidemiology in the new era.

As the practical consequence of this changing environment, many reports of new outbreaks of diseases caused by **fungi** (*Alternaria, Colletotrichum, Erysiphe, Fusarium, Rhizoctonia, Septoria and Verticillium*), **bacteria** (*Clavibacter, Xanthomonas*), along with **viral** disorders emerged in the early 1900's (Smith 1909; Allard 1915; Kreitlow et al. 1961).

These reports triggered the concern over the role of seeds in the transmission of plant pathogens since all these pathogens were demonstrated to be seedborne and the corresponding outbreaks coincided with: (i) the birth and growth of the seed industry, (ii) the increasing seed movement worldwide and (iii) the recognition of the potential role that seeds play in epidemics of plant diseases (Elmer 2001).

II.1.2. The role of seed systems in the epidemiology of viral diseases: the need for a new paradigm in Africa

The association between seeds and pathogens is presumed to be an ancient one since for as long as humans have collected and moved seeds to new sites, they have unknowingly transported plant pathogens. In such a way, this association helped the dissemination of plant pathogens and importantly the regulation of plant pathogen populations (Elmer 2001).

We used these various theories borrowed from the scientific community to propose the fact that human factors and by far the socio-cultural system that shapes his

behavior, have the potential to impact the epidemiology of plant diseases (Figure II-2).

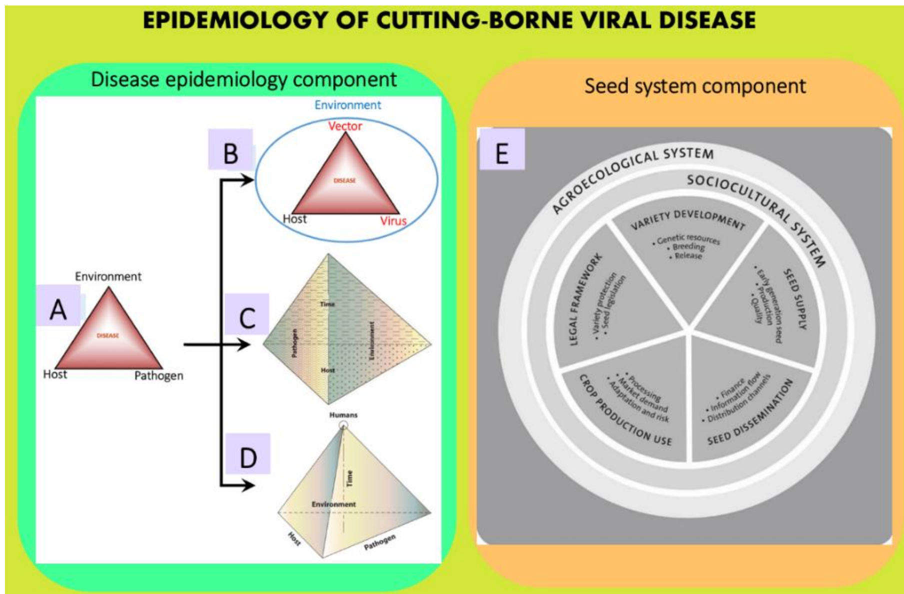


Figure II-2. The framework adopted in this study for investigating the epidemiology of a cutting-borne viral disease as is the case for the CBSD.

(A) Disease triangle suggesting the disease as resulting from the impact of three factors (Horsfall and Dimond 1960). **(B)** The viral disease triangle in the concept of the "ecological trinity" (W. Carter 1939; Thresh and Fargette 2001). **(C)** The disease tetrahedron involving time as a fourth component (Agrios 2005; Francl 2001). **(D)** The disease tetrahedron involving humans as the fifth factor: host, pathogen, and environment are each represented by one of the sides of the triangle. Time is represented as the perpendicular line arising from the center of the triangle and humans as the peak of the tetrahedron whose base is the triangle and height is the length of time. In this way, humans interact with and influence each of the other four components of an epidemic, thereby increasing or decreasing the magnitude of the epidemic (Agrios 2005). **(E)** Five basic seed system functions (center), embedded in a larger socio-cultural and agro-ecological contexts as proposed by (Christinck, Rattunde, Kergna, Mulinge, Weltzien, et al. 2018).

We supports that all the mechanisms, processes or schemes for accessing and disseminating seeds, known as seed systems, could be of key importance in the mitigation of the spread of seedborne pathogens responsible of major epidemies at various scales as demonstrated by potato seed systems in Europe and America (Vanderschuren and Rey 2017).

For this to be effective, we considered the following postulates as underlining our study context:

1. Seed systems are conceptualized as a human activity system, established and maintained by human actors motivated by individual or collective purposes and who have relationships with- and contribution to- the

larger environment in which they are embedded (Banathy 1996; Christinck, Rattunde, Kergna, Mulinge, and Weltzien 2018; Checkland 1981).

2. Many studies on seed systems distinguish between formal and informal systems, with the former following more or less the model of an industrial supply chain, while the latter entail a range of mostly farmer-managed activities, for example in saving, using, exchanging and selling seed in local networks and markets (McGuire and Sperling 2016). However, the division of seed system actors and components into formal and informal categories appears problematic since there is a growing degree of overlap between both systems allowing, for example, varieties originally developed by the formal sector to enter the informal and *vice versa* (Christinck, Rattunde, Kergna, Mulinge, and Weltzien 2018).

3. Despite the importance of cassava in developing countries, the majority of casava growers are subsistence or small-scale farmers and the crop has received relatively little attention in the period preceding the last twenty years when compared with maize, rice and wheat (Varshney et al. 2012). This favored informal seed system to be the most predominant, and, as a consequence, favored the accumulation and dissemination of viruses since phytosanitary measures were rarely applied. It appeared therefore that informal seed systems of cassava could be particularly prone to the spread of cutting-borne viral diseases than formal seed system.

4. To date, CBSD viruses are only found in East and Central Africa. However, the risk for being disseminated to other cassava-growing areas of Africa, Latin America and Asia could increase if infected planting material is introduced in those areas, which would result in huge economic losses and food insecurity (J. Legg et al. 2014). Additionally, the Western Africa is a belt which has globally leading cassava producing countries and fortunately, CBSD is not yet there. The D.R.Congo is the Western frontier of CBSD spread, and plays a key role in the risk of CBSD spread in West Africa which could result in great concern for global cassava production and food insecurity. Unfortunately, recent reports on the epidemiology of this disease in D.R. Congo showed a global westward expansion at rapid speed, moving from one province in which the disease was reported in 2012 to 5 provinces actually (Muhindo et al. 2020; Mulimbi, Phemba, and Assumani 2012). Based on this, we argue that the research methods that are used to investigate the main factors underpinning the epidemiology of CBSD in the seed system are not yielding context-specific information necessary for a tailored strategy for mitigating the disease spread.

5. Early at the revolution of the HTS methods, various protocols for library preparation and for processing and analyzing mass-generated data have been developed. They are still constantly evolving with new tools regularly created for the purpose of improving the detection and the characterization of viral sequences present in the generated data. Currently, some of these tools, when used to re-analyze data that have been generated since several years, allow identifying some previously unidentified and/unknown viruses.

6. Despite the availability of this technology, various authors highlighted that the current knowledge about plant virus diversity is underestimated and biased : underestimated because a significant proportion of virus-derived sequences have no similarity to anything in public databases (dark matter) (Roossinck 2015), and biased because most studies about virus diversity focused on pathogenic viruses in cultivated plants and ornamentals, putting aside the asymptomatic wild vegetation (Roossinck, Martin, and Roumagnac 2015)

II.2. Research hypothesis and objectives

In line with what has been depicted above, we align the hypothesis of this work as follows:

I. We first suggest that research methods for investigating the epidemiology of the CBSD, a seed-borne disease, need to take into account the intrinsic characteristics of the African agriculture and the specificities of seed systems of cassava, a vegetatively-propagated crop, and we call for a more inclusive assessment method that consider both the ecological as well as socio-cultural aspects in the study of disease epidemiology. We support this based on the concept brought by (Banathy 1996; Christinck, Rattunde, Kergna, Mulinge, and Weltzien 2018; Christinck, Rattunde, Kergna, Mulinge, Weltzien, et al. 2018; Checkland 1981). We therefore propose that **intensive surveys of epidemic areas following (using) multidisciplinary approaches could be suitable (adapted) in identifying context-specific drivers of the epidemiology of cassava brown streak disease instead of routine extensive surveys that are focused only on traditional disease components, providing only a snapshot of the real context necessary for the implementation of sustainable and clean seed system.**

II. Many different management strategies, mainly borrowed from the long experience gained with the CMD, to tackle the dissemination of CBSD in Africa have been both proposed and imposed as to control the disease early at its re-emergence and used by a number of interventions aiming at mitigating its spread. As the epidemiology of the CBSD was only partially elucidated, yet not now it is fully, the success of the proposed strategies met with mixed success (McQuaid et al. 2016) and sometimes resulted in unexpected non intentional spread of the disease. As investigations were pursued to fully elucidate the infection mechanisms and virulence determinants of CBSD these last 10 years, various reports of uncommon root symptoms features and negative molecular detection of cassava samples showing CBSD-like symptoms emerged (Munganyinka et al. 2018; Zeyimo et al. 2020). Even though this situation could be due to various factors, we anticipate that **the viral diversity of viruses infecting cassava, involved or not in the CBSD etiology, could be underestimated and a better sampling strategy linked to the use of a cutting-edge detection technology could allow contributing to the elucidation of this situation**.

In order to investigate on these hypothesis, we have drawn the following objectives :

i. **Objective 1**: Proposing an integrated, multidisciplinary framework h for assessing the epidemiology of CBSD that consider both the socio-cultural, geographical, agroecological contexts together with the traditional components of disease epidemiology.
ii. **Objective 2**: Testing the efficiency of this approach at a local level by surveying a high disease-pressure region in order to capturing the farmer's seed sourcing practices that can act as potential factors impacting the dissemination of CBSD through various seed pathways. These factors will be considered as metrics useful for seed actors and policy makers to enable informed decisions and interventions to tackle the dissemination of CBSD in the local cassava seed systems.
iii. **Objective 3**: Building an efficient sampling framework that, combined to a cutting-edge technology of virus detection (HTS), can allow elucidating the diversity of viruses infecting cassava in the surveyed area.

II.3. Thesis structure

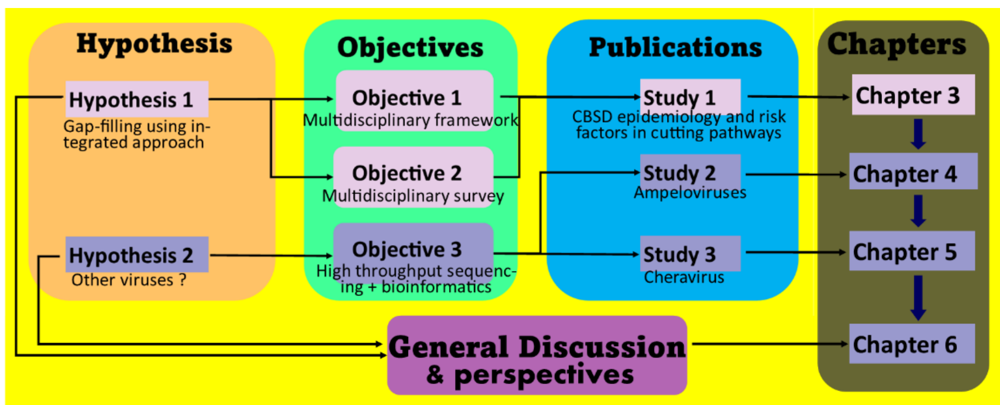


Figure II-3. General organization of the thesis

After the first two chapters presenting the state of the art and the objectives in relation with this study, the third chapter will be dedicated on investigations related to the first and the second objectives while the fourth and the fifth chapters will be dedicated on investigations related to the third objective. In the last chapter of this thesis (sixth), the various results obtained are discussed in the view of the established hypotheses and global perspectives are established. Chapters 3, 4 and 5 have been organized as manuscripts and have been published or submitted in peer-reviewed journals (the chapter 5 is still under review).

Chapter III

III. Risk factors associated with cassava brown streak disease dissemination through seed pathways in Eastern D.R. Congo

"*Most of the issues that vex humanity daily,…, can be solved only by integrating knowledge from the natural sciences with that from the social sciences and humanities*" (Wilson 1998)

Preliminaries

The problem of cassava viral diseases is a concern of the whole agricultural system and it necessitates the intervention of various key stakeholders to be efficiently addressed. For the effectiveness of this intervention, these key stakeholders need to collaboratively work together for accurately assessing the critical issues on which specific actions are needed. This collaboration implies the use of a tool that bring these partners on the same ground, in order to assess the problem with the same objective but according to various angles of view. Unfortunately, such tool that allows this multidisciplinary work to be effective and that allows decision makers to definitely design and coordinate intervention strategies suited to local contexts, is not documented.

In this chapter, reflexions and investigations have been undertaken to design a framework that could serve as a tool to generate accurate, context-specific and action-driven information regarding the epidemiology of the CBSD with the aim of strengthening the resilience of the cassava seed production system against the CBSD. Additionally, the designed framework was tested within a strategic pilot area, the Uvira territory, in order to assess its relevance especially in the view of highlighting the role played by farmer's seed-sourcing practices in the CBSD dissemination. For example, a specific question that is investigated consisted in knowing which cutting pathway had less risk of delivering infected planting material? This central question and additional ones are investigated within this chapter 3.

Chapter 3

This chapter is based on the article published in:
"*Frontiers in plant sciences", 22 July 2022*
Section Plant Biotechnology
Volume 13-2022
https://doi.org/10.3389/fpls.2022.803980
Only text formatting was operated to accommodate the published version

III. Risk factors associated with cassava brown streak dissemination through seed pathways in Eastern D.R. Congo

Yves KWIBUKA[1, 3 *], Chantal NYIRAKANANI[2] Jean Pierre BIZIMANA[2, 4], Espoir BISIMWA[3*], Yves BROSTAUX[5], Ludivine LASSOIS[2], Hervé VANDERSCHUREN[2,6] and Sébastien MASSART[1*]

[1] Plant Pathology Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[2] Plant Genetics Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[3] Faculté des sciences Agronomiques, Université Catholique de Bukavu, Democratic Republic of Congo.

[4] Rwanda Agriculture Board (RAB), Department of Research, Kigali, Rwanda

[5] Applied Statistics, Computer science and Modeling Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[6] Laboratory of Tropical Crop Improvement, Division of Crop Biotechnics, Biosystems Department, KU Leuven, Leuven, Belgium.

***Correspondence:**
Yves KWIBUKA, Sébastien MASSART, Espoir BISIMWA
bisimwayves@gmail.com, sebastien.massart@uliege.be, ebisimwa@yahoo.com

Keywords: Cassava, viral diseases, epidemiology, seed system, risk factors, molecular diagnostic

III.1. Abstract

Vegetatively propagated crops are particularly prone to disease dissemination through their seed systems. Strict phytosanitary measures are important to limit the impact of diseases as illustrated by the potato seed system in Europe. Cassava brown streak disease (CBSD) is a devastating disease caused by two viral species collectively named cassava brown streak viruses (CBSVs). CBSD can cause substantial root yield losses of up to 100 % in the worst affected areas and is easily transmitted through stem cuttings. In Eastern and Central Africa, the epidemiology of CBSVs in the local socio-economical context of production remains poorly known while a better understanding would be an asset to properly manage the disease. This lack of information explains partially the limited efficiency of current regulatory schemes in increasing the availability of quality seed to smallholders and mitigating the spread of pests and diseases. This study surveyed the epidemiology of CBSVs in Uvira territory, Eastern D.R. Congo, and its drivers using a multivariate approach combining farmer's interview, field observation, sampling and molecular detection of CBSVs. Investigation on the epidemiology of CBSD revealed that three clusters in the study area could be identified using five most significant factors: (i) symptoms incidence, (ii) number of whiteflies, (iii) types of foliar symptoms, (iv) cutting's pathways and (v) plant age. Among the three clusters identified, one proved to be potentially interesting for seed multiplication activities since the disease pressure was the lowest. Through risk assessment, we also identified several key socio-economic determinants on disease epidemy: (i) factors related to farmer's knowledge and awareness (knowledge of cassava pests and diseases, knowledge of management practices, support from extension services and management strategies applied), (ii) factors related to the geographical location of farmer's fields (proximity to borders, proximity to town, distance to acquire cuttings), as well as (iii) the pathways used to acquire cuttings.

III.2. Introduction

Cassava (*Manihot esculenta* Crantz) is the tenth most important crop in the world in terms of global annual production (303 Million tons) (FAO 2020). In Africa, its importance in the livelihood of populations has long been demonstrated: it is ranked the first most important food crop in terms of global annual production (192 Million tons in 2019), the first in terms of source of food (76 Million tons) and the fourth in terms of source of calories (167 Kcal/person/day) after wheat, maize and rice (FAO 2021a). It is the most important food crop (occupying approximately 40% of agricultural land dedicated to food crops) and the largest non-cereal carbohydrate source for more than 70% of people in D.R. Congo (Mahungu et al. 2014). According to statistics from 2019, this country is ranked the second in Africa in term of production (40 Million tons) after Nigeria (59 Million tons) (FAO 2021a).

One of the crucial factors to increase agricultural productivity is the planting material: in this paper stem cuttings used for the propagation of cassava are referred to as *seed* (McEwan et al. 2021). Farmers often use different approaches to obtain seeds of a crop. Because of their variability and local specificity to needs and preferences, local approaches (e.g. household stocks, markets and social exchange networks) provide usually most of the seeds that small farmers use (Sperling and Cooper 2003). Common figures suggest that the local seed system provides 80-90 percent of the planting material to farmers (DANAGRO 1988; Rabobank 1994). On the other hand, under established, controlled and optimized seed systems, farmers can access affordable and high-quality seeds (Buddenhagen et al. 2017).

Seed systems of root, tuber, banana and other vegetatively propagated crops (VPC) are predominantly informal or managed at local levels by farmers themselves without major public or private sector involvement in the production, supply, or quality control of planting materials. The quality of seeds is often signaled through trust and reputation while the vegetative mode of seed multiplication increase the risk of pathogens, including viruses, building up over multiple cycles of propagation (Jarvis et al. 2012; Campo, Hyman, and Bellotti 2011; Almekinders et al. 2019; McEwan et al. 2021). In practice, most farmers in low-income countries save seed from the previous season for replanting (Devaux, Kromann, and Ortiz 2014).

Cassava cultivation is suffering significant losses due to biotic stresses (Vanderschuren and Rey 2017; Lozano and Booth 1974; Reddy 2015b; Kwibuka et al. 2022) among which two viral diseases, cassava brown streak disease (CBSD) and cassava mosaic disease (CMD) are of major economic importance in sub-Saharan Africa (Vanderschuren and Rey 2017). In Africa, previous estimates indicated overall incidences of 50% to 60% with estimated annual losses of $1.2-2.4 billion for CMD (Thresh et al. 1997; J.P. Legg et al. 2006), while annual economic losses of up to US $ 726 million were associated to CBSD with incidences of up to 100% being recorded (Maruthi 2015). The situation is far from being controlled as future CBSD pressure is projected to increase by at least 2% in D.R. Congo by 2030 (Jarvis et al. 2012).

CBSD is associated with two *Ipomovirus* species (collectively named CBSVs), named cassava brown streak virus (CBSV) and Uganda cassava brown streak virus (UCBV). (Lozano and Booth 1974; Reddy 2015a)(E. B. Bisimwa 2012)This disease is considered as endemic to low-altitude and coastal zones of Kenya, Tanzania and Mozambique (Storey 1936; Nichols 1950). New outbreaks into areas 1200 meters above sea level of Uganda and D.R. Congo (Mulimbi, Phemba, and Assumani 2012; T Alicai et al. 2007; K. R. Tomlinson et al. 2018a) indicated later a significant shift in its epidemiology and a westward progression to areas previously not at risk.

Recently, several studies tried to elucidate the components of pathogens and arthropod pest invasion risk. Factors such as climatic conditions (Kroschel, Sporleder, and Carhuapoma 2016), structure of trade routes (Bebber, Holmes, and Gurr 2014) as well as habitat (cropland) connectivity (Xing et al. 2020) have been reported to play a major role in pest and disease dissemination. However, these studies have addressed these aspects from a global point of view and specific analysis of cassava crop is still missing. Additionally, network studies (Shaw and Pautasso 2014) improved the understanding of the pathways of pathogens spreads. Such study can facilitate the reduction of the disease inoculum flow in exchanged plant materials by identifying likely ways to find the best sites to monitor as warning sites. The risk that pathogens can move through particular pathways of a seed system network is a key component of disease risk, along with other risk factors such as potential transmission by vectors or wind dispersal (Buddenhagen et al. 2017).

Pathways (sources) by which farmers obtain cassava seeds (cuttings) are of key importance in the mitigation of plant diseases. Accumulation and spread of viruses in planting material of vegetatively propagated crops provide the primary inoculum on field and are a key factor in the development of disease epidemics. This impact is particularly important when cuttings of susceptible cultivars used by farmers come from pathways without any sanitation measure (A. N. Frangoie, Bidiaka, and Mahungu 2019a). Therefore, understanding in a timely manner which cutting pathway/source has a high risk of pathogen dissemination in a seed network is a milestone for mitigation measures to be implemented.

Previous research has identified four factors impacting the spread of cassava viral diseases in most of affected areas in Africa: education and access to information, lack of plant health and extension services, weak access to improved varieties (FAO 2021b) and social factors, exemplified by kinship systems (Delêtre et al. 2021).

The international Plant Protection Convention (IPPC) has proposed a standardized approach for assessing and managing the phytosanitary risk of a pest and/or of a pathway [(Pest Risk Assessment (PRA)] (IPPC 2021). This study focuses on a component of PRA related to the probability of spread of an established pest (CBSV) through pathways used by farmers to access planting material of a vegetatively propagated crop. It aimed at (i) establishing the epidemic profile of CBSD in Uvira territory by identifying the viruses occurring and describing their spatial distribution and (ii) identifying multidisciplinary factors underpinning the spread of CBSD and that can be useful in generating necessary knowledge for seed quality assurance, clean seed use and ultimately better CBSD control.

III.3. Material and methods

III.3.1. Conceptual framework

Two sets of factors that can play a role in the outcome of CBSV infection were included within the model adopted in this study (Figure III-1). The first set (considered here as independent variables) is related to cutting pathways (diversity and characteristics) used by farmers and the second set (considered as intermediate variables) is related to human-mediated (Bebber, Holmes, and Gurr 2014; Delêtre et al. 2021; FAO 2021b) and environmental factors (Kroschel, Sporleder, and Carhuapoma 2016; Xing et al. 2020; Buddenhagen et al. 2017). Risk factors are those associated positively and significantly to the increase of the number of diseased plants and therefore to a significant increase in the probability of CBSD infection.

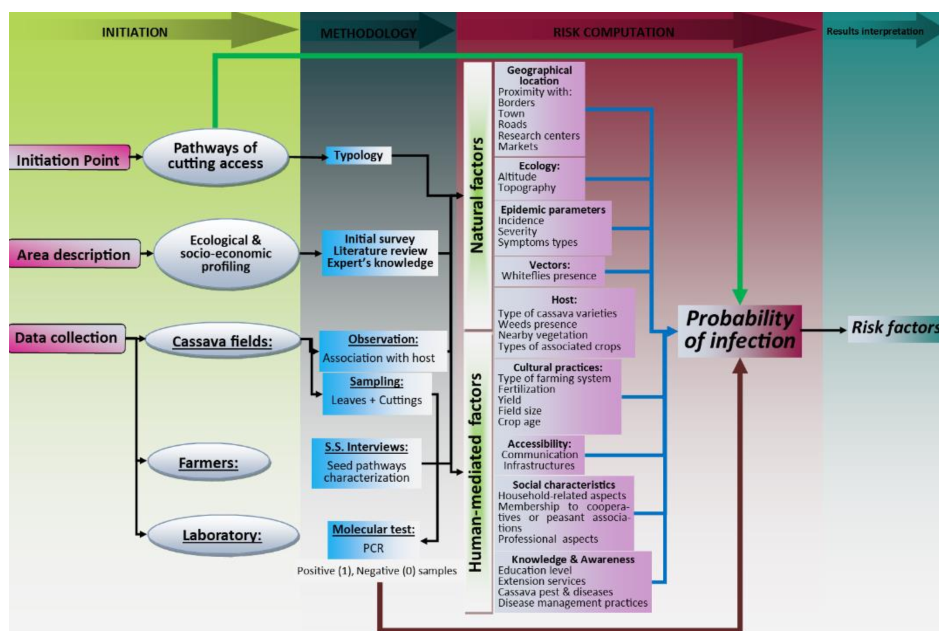


Figure III-1. Illustration of the conceptual framework adopted in this study

Among other factors, human-mediated factors included (i) farmer's demographic factors (sex, age, marital status, size of household, household headship and relation to the head of household), (ii) knowledge based factors (education level, access to training, experience in cassava farming, knowledge of cassava pests and diseases, knowledge of management practices against cassava pests, access to extension services, membership to cooperatives), (iii) economic-based factors (land size, land ownership, livestock ownership, access to inputs, main sources of income, labor type and availability) and, (iv) farming practices (type of farming system, field hygiene and rotation, planting and harvesting periods, use of disease-free planting material, use of resistant/improved varieties). Environmental factors included the (i) geographical location of fields (proximity

with country borders, roads, main cities, markets, research centers, crop diversity around cassava field etc.), (ii) ecological factors (altitude, topography) and (iii) epidemiology (incidence, severity, type of symptoms, vectors, hosts etc.). etc.

III.3.2. Study area

This study was conducted in Uvira territory, one of the eight administrative entities composing the South Kivu province in the Eastern D.R. Congo (Figure 2). This territory is composed of two ecological regions located in tropical zone of low altitude (climate type $AW_{1\text{-}3}$, altitude lower than 1000 m, rainfall < 1300 mm/year and annual mean temperature > 24°C) and in tropical zone of mild and high altitude (climate type Am, altitude between 1000 and 1800 m, rainfall < 1600 mm/year and annual mean temperature < 23°C) respectively (E. B. Bisimwa 2012). This territory is ranked among the top cassava producers of the province (IPA 2020).

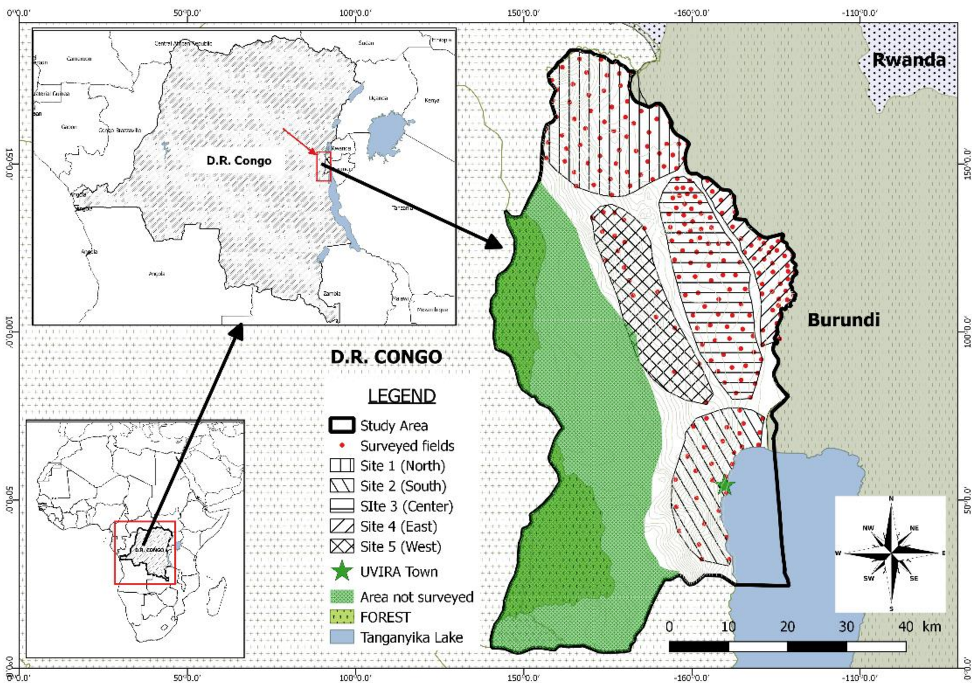


Figure III-2. Geographic map of the Study area showing subdivision into 5 sites. Red dots represent surveyed fields

III.3.3. Farmers and fields selection

The design adopted in this study consisted into a multistage approach to select farmers and fields to survey. In the first stage, a purposive sampling strategy was used to select villages to be surveyed. In fact, from an initial sampling frame constituted by the total number of villages in the territory (226), sixty-five villages were retained after discarding villages where cassava productivity was marginal. A preliminary

survey as well as literature review allowed to document the main socio-economic and ecological parameters susceptible to play a role in the local epidemiology of CBSD and impacting the way farmers access cuttings. However, some parameters were voluntary ignored because considered as potentially leading to conflicts or inappropriate by the local administrators (village chiefs, local agricultural officers). According to these harmonized parameters (table III-1), the 65 villages were clustered into 5 study sites: the North (site1), the South (site 2), the center (site 3), the East (site 4) and the West (site 5) (Figure III-2). Using official production statistics from 2019 (IPA 2020) related to the amount of cassava produced, a typology of villages was established and consisted of villages with high intensity of production (annual cassava production > 1,000 tons) as well as those of low intensity (annual cassava production < 1,000 tons).

In the second stage, a simple random sampling strategy was used to select farmers and fields in each village. Four farmers and their corresponding fields were surveyed within each village belonging to the first category while at least 2 fields surveyed within villages belonging to the second category. These numbers of fields were determined according to the time and resources assigned to this work. Random numbers generated in Microsoft excel and assigned to the list of cassava producers for each village allowed to select fields to visit. These fields were identified under the lead of local agricultural officers. Fields were distant of at least 2 km and sown with cassava plants of more than 6 months old. An official authorization letter to conduct this study was delivered by the authorities of the Université Catholique de Bukavu and served as the official communication to local administrative chiefs who, in return granted authorization to undertake the study in entities under their responsibility.

Table III-1. Characteristics of the 5 sites

Denomination	Location	Main villages	Characteristics
Site 1	North	Kamanyola, Luvungi, Bwegera, Kiringye, Katogota, Ndolera, Lubarika	Share border with both Rwandan and Burundian Republics Distant from the administrative seat of the territory (Uvira, 70 Km) The topography is mixed (plain and mountains)
Site 2	South	Rutemba, Muhungu, Kavimvira, Kalungwe, Sango	Close to the Uvira city (the administrative main town of the territory) The topography is dominated by mountains
Site 3	Center	Kitemesho, Luberizi, Mutarule, Nyakabere, Sange, Runingu	Most of villages are close to the main national road NR1 Located entirely in the low altitude zone (uniform topography) Villages are easily accessible
Site 4	East	Rwenena, Ndunda, Rusabagi, Sasira, Kigurwe, Rurimbi, Ruzia, Mwaba	Located on the border close to Republic of Burundi Distant from the national road 1 crossing the territory. Located entirely in the low altitude zone (uniform topography).
Site 5	West	Rubanga, Langala, Lemera, Mushegereza, Mulenge, Lusheke, Mugaja, Kanga	Located completely in mid or high altitude and dominated by mountains The area is poorly accessible Population density is lower compared to other sites Most agro-ecological characteristics differs from other sites

III.3.4. Farmer interviews

The survey consisted into semi-structured interviews using mobile-recorded questionnaires using the Open Data Kit platform (Hartung et al. 2010) and was conducted directly on field site (except for some cases when the farmer could not be present on field). This strategy allowed to perform field inspections directly after interviews and to record information regarding the epidemiology of CBSD. Before starting interviews and performing sampling, a voluntary agreement of farmers was required through an explanation of the purposes of this study. In return, the farmer had to give an oral consent and additional explanations regarding confidentiality of collected data were also provided if required by the farmer.

III.3.5. Epidemic survey

Epidemic parameters that were observed and recorded consisted in symptoms incidence, symptom types, symptoms severity on leaves, stems and roots as well as whitefly number. These observations were conducted on randomly selected plants, symptomatic or not. (Rj Hillocks and Jennings 2003)(R. J. Hillocks, Raya, and Thresh 1999)(Munthali 1992)(G. M. Rwegasira and Rey 2012)(Rj Hillocks and Jennings 2003)(R Hillocks and Thresh 2000)Symptoms incidence observation was conducted on 30 cassava plants encountered following diagonals and medians through the field (including 5 plants on each of the four outer sides and 5 plants on each of the two inner diagonals) (R. J. Hillocks 2004; G. Rwegasira and Rey 2012). This parameter was then recorded as the number of plants that showed CBSD-like symptoms out of the 30 observed. Symptoms severity on aerial plant parts was assessed using the 5 levels scoring-scale from (Titus Alicai et al. 2016b). (Kaweesi et al. 2014)The number of whiteflies was counted on the top five youngest leaves of cassava plants selected for observation. The type of symptoms were categorized based on distribution of leaf chlorosis and stem lesions on the plant; systemic and on the whole plant (SW), systemic on leaf or stem parts but localized (SL), only on lower leaves (LL) (Titus Alicai et al. 2016b).

III.3.6. Sample collection: leaves and stem cuttings

In each field, among the 30 plants inspected, 10 plants were randomly selected for sample collection irrespective to symptoms presence. From the top of the plant canopy downwards, 400 - 600 mg of the third fully expanded leaf on the major stem were collected, silica gel-dried and stored into sealed plastic tubes pending RNA extraction. Young leaves were preferred as content of polysaccharides and polyphenols, interfering with the molecular detection of the viruses, are lower (Heikrujam, Kishor, and Behari 2020; Orek 2018; J. Zhang and Stewart 2016; Shankar et al. 2010). A total of 2,400 leaf samples were therefore collected from cassava fields of at least 6 months old. Two stem cuttings having at least 6 node buds were also sampled in the middle part of each selected plant using a pair of shear. The stem cuttings were labelled and stored pending re-plantation in an experimental field. A total of 480 stem cuttings were therefore collected.

III.3.7. Molecular analysis

RNA extraction

Total nucleic acid was extracted from all of the silica gel-dried leaf samples shipped to Belgium using a modified CTAB protocol (Chang, Puryear, and Cairney 1993; Moreno, Gruissem, and Vanderschuren 2011).

The 10 separate tubes containing silica gel-dried leaf-tissues from unique fields were pooled (50 mg for each sample) and transferred into a thick-gauged plastic grinding bag. 2ml of CTAB extraction buffer (2% CTAB, 2% PVP, 100mM Tris-HCl pH 8.0, 25 mM EDTA, 2M NaCl and 2% β-Mercaptoethanol added before use) was added and the leaf tissues were thoroughly grounded using a hand-held ball bearing sample grinder. 1 ml of the lysate was transferred to a 2 ml Eppendorf tube, homogenized and incubated at 60°C for 30 min with periodic vortexing at 10 min interval. 600 µl of chloroform: IAA (24:1) was added and the mix homogenized by inverting the tube. Phases were separated by centrifugation at maximum speed for 10 min at 4°C and the supernatant (upper aqueous phase) thoroughly transferred to a 1.5 ml Eppendorf tube. This Chloroform-IAA treatment was repeated once and 0.6 volumes of ice-cold Isopropanol (-20 °C) was then added. Samples were allowed to stand for 2 hours at -20°C then centrifuged for 30 min at 4°C to pellet the nucleic acid. Supernatant was removed and 0.5 ml of 70% Ethanol added to wash the pellet by centrifugation at 4°C for 5 min at maximum speed. Supernatant was removed, the pellet air-dried, re-suspended in 100 µl of TE buffer and stored on ice. Samples were DNase-treated using Amplification Grade DNASE I (Invitrogen®, USA) according to the manufacturer's instructions and the quality of RNA tested using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Samples were aliquoted and stored at -80 °C prior to testing.

CBSV and UCBSV detection by RT-PCR

All samples were tested for CBSV and UCBSV using a two-step RT-PCR assay. cDNA synthesis was carried out with Tetro™ reverse transcriptase (Meridian Bioscience®) according to manufacturer's instructions. Random hexamers primers were used for generating the first strand cDNA. Amplification of cDNA was done using Mangotaq™ DNA polymerase (Meridian Bioscience®).

Degenerated primer pair targeting the coat protein genes of both CBSV and UCBSV was used [CBSVs-F (5´-CCTCCATCWCATGCTATAGACA-3´) and CBSDD-R (5´-GGATATGGAGAAAGRKCTCC-3´)] (Anjanappa et al. 2016).

These primer pair amplified a product of ~703 bp in the presence of CBSV and a product of ~800 bp in the presence of UCBSV. A 10 µl PCR mixture containing 5.8 µl nuclease free water, 2 µl PCR buffer (5X), 0.40 µl $MgCl_2$ (50 mM), 0.20 µl dNTPs (10 mM), 0.20 µl of each primer (10 mM), 0.4 µl Mango taq DNA polymerase (Meridian Bioscience®) and 1.0 µl of cDNA. The temperature profile of PCR consisted of 95°C for 2 min followed by 30 cycles of 94°C (30 s), 56°C (30 s) and

72°C (50 s) for denaturation, annealing and extension, respectively. A final elongation of 72°C for 5 min was also included to terminate the amplification.

The cassava *PP2A* gene was used as internal control gene in parallel reactions with the following primer pair (PP2A-F: 5´-TGCAAGGCTCACACTTTCATC-3’ and PP2A-R: 5´-CTGAGCGTAAAGCAGGGAAG-3´) (Moreno, Gruissem, and Vanderschuren 2011). The *PP2A* primer pair generates an ~187-bp amplicon from cassava cDNA samples. PCR products were analyzed by electrophoresis in TAE buffer (1X) on a 1% agarose gel stained with Gel red® (Biotium), visualized under UV light and photographed using a gel documentation system (E-Box CX5 Edge, Vilber/Fisher Biotech).

III.3.8. Data analysis

Data collected from surveys were CSV-formatted and resulting files were loaded into R Software version 4.1.1 for analysis.

Descriptive statistics

Factors related to sites and clusters served as grouping factors to average the epidemic parameters (field symptom incidence, molecular detection incidence and severity score). The averaging of symptoms incidence and severity in cassava fields included both symptomatic and asymptomatic plants.

To get information about the CBSD incidence with regards to clusters, a generalized linear regression with logit link was applied (Dodge 2008; Agresti 2002), namely logit(p)=ln(p/(1-p)). The regression equation used was therefore written as:

$$logit(p) = \ln(p/(1-p)) = \beta_0 + \beta_1 * Cluster + Error$$

Where p is the probability of the dependent variable [0;1], β_n the regression coefficients, and “Cluster” an explanatory variable with 3 levels. The model parameters were estimated using the maximum likelihood method (Everitt 2006), with Chi-squared test for significance, and the least-squared means comparison by the “lsmeans” package (Lenth 2016) of the R software (when statistical difference was significant).(de Mendiburu and Yassen 2020). Additional univariate and bivariate descriptive statistics included the calculation of frequencies and percentages, standard deviation, minimum and maximum values and were used to describe the main socioeconomic characteristics as well as the parameters for disease epidemics in the field.

Multivariate statistics

A multivariate analysis was performed using a factorial method (Factor Analysis of Mixed Data) to analyze the association between epidemic parameters, detection results and the main seed systems parameters. Hierarchical Clustering on Principal Component (HCPC) method was used to identify groups (or clusters) of fields showing similar characteristics within the study area according to relevant epidemic parameters, detection results and seed systems parameters. For the analysis, the Ward

clustering algorithm and the Euclidian distance were used (Hartigan 1975; Kassambara 2017).

The "Test Value" criterion (VT) was used to select variables considered as relevant for the characterization of groups/clusters from HCPC analysis. For continuous variables, the VT was used to rank/sort variables according to their relevance in order to distinguish the variables that play an essential role in the interpretation of the groups. For discrete variables, the VT was used to highlight the category which characterizes the better the group/cluster of observations (Lebart, Morineau, and Piron 1995; TANAGRA 2009)

Risk factors/determinants of CBSVs

Binary logistic regression (Hosmer and Lemeshow 2000; Collet 2003) was used to identify seed system risk factors (determinants) associated with the absence (negative detection) of CBSVs infection in cassava fields.

The probability $p(x)$ of a field to be infected by a particular form of CBSVs (Y=1) given a modality x_1 of the predictor variable X (1) was estimated by comparing the Odd Ratio (OR) of that modality x_1 to the reference x_0 (2). The possible outcomes of this comparison were interpreted as indicated in the table III-2.

$$p(x) = P(Y = 1|X = x). (1)$$

$$(x_1, x_0) = \frac{odds(x_1)}{odds(x_0)} = \frac{\frac{p(x_1)}{1-p(x_1)}}{\frac{p(x_0)}{1-p(x_0)}}. (2)$$

Rules for interpretation of OR (*Rn*) are presented in the table III-2.

Table III-2**.** Rules (**Rn**) for interpretation of odd ratios

R1	$OR(x_1, x_0) > 1 \Rightarrow p(x_1) > p(x_0)$
R2	$OR(x_1, x_0) = 1 \Rightarrow p(x_1) = p(x_0)$
R3	$OR(x_1, x_0) < 1 \Rightarrow p(x_1) < p(x_0)$

They could be summarized as follow:

R1: When OR>1, the risk of having an infected field is higher (more probable) in the case X= x_1 compared to the reference case X= x_0.

R2: When OR=1, the risk of having an infected field in the case X= x_1 is equal to that of the reference case X= x_0.

R3: When OR<1, the risk of having an infected field is lower (less probable) in the case X= x_1 compared to the reference case X= x_0.

Model building process

Upon completion of the bivariate analysis (univariable model), all variables whose univariate test had a p-value <0.05 were included in the initial multivariate model along with all variables that were relevant for explaining the studied questions (Mickey and Greenland 1989; Hosmer and Lemeshow 2000).

Subsequent models were fitted by sequentially including or excluding variables from that initial model based on statistical criteria using forward and backward stepwise procedures (Chambers 1992; Venables and Ripley 2002). Before running the stepwise algorithm, multicollinear variables were visually identified using correlation matrix for predictor variables and ensured they were effectively suppressed from the fitted model after running the stepwise procedure. The likelihood ratio test (LRT or deviance adequation test) was used to assess the fitness of these second-order models compared to the initial one (Hastie and Pregibon 1992; Hosmer and Lemeshow 2000). Additionally, all candidate models were assessed for their goodness-of-fit using the Hosmer and Lemeshow test implemented using the “ResourceSelection” package in R and for their stability, sensitivity and specificity using the mplot. The AIC (Akaike Informative Criterion) was used to select the optimal model among the best fitted. All else being equal, the model with the lower AIC was considered as optimal and the corresponding significant predictors were considered as determinant risk factors (Sakamoto, Ishiguro, and Kitagawa 1986).

The VIF (Variance Inflation Factor) was used to detect the presence of multicollinearity within the predictors for the best-fitted model and therefore to assess the suitability of the coefficients for the interpretation (J. Fox and Weisberg 2019) . Odd Ratios and confidence intervals were computed using Wald method.

Geographical maps were elaborated using Quantum GIS Software version 2.14.0 Essen by interpolating values of disease incidence using the Inverse Distance Weight method (IDW) (QGIS Development Team 2021).

III.4. Results and interpretations

III.4.1. Identification of clusters within the study area

An HCPC analysis was carried out based on the results of the survey (farmer interview and symptoms observation) (see the questionnaires in supplementary materials 1, 2 and 3) and of the molecular detection of both viruses (see next chapter). The nine most important parameters that optimally characterize the studied area were identified with the HCPC analysis. Using those selected parameters (see table III-3 for categorical variables and table III-4 for quantitative variables), the study area was separated into three clusters each with a similar number of fields (Figure III-3)

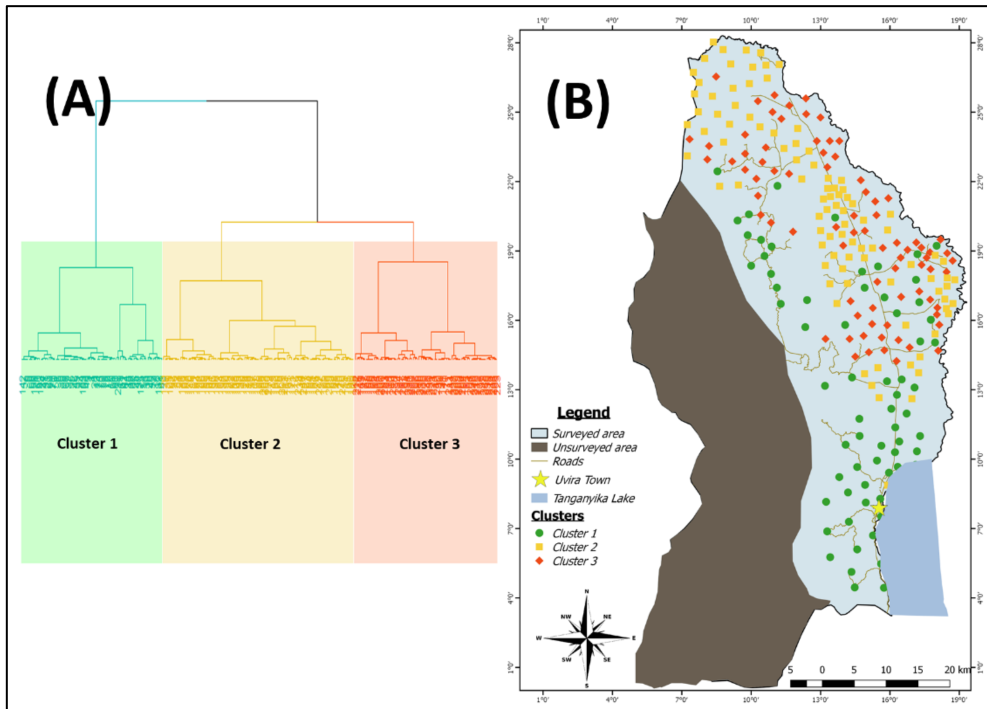


Figure III-3. **(A)** Cluster dendogram showing the repartition of the data into three clusters using the Hierarchical Clustering on Principal Component method. **(B)** Mapping of the clusters identified by HCPC in the study area. Each color category is associated to a cluster: Green for cluster 1, yellow for cluster 2 and red for cluster 3.

Table III-3**.** Categorical variables associated with the description of clusters from HCPC analysis

Cluster 1 (n=80; 33%)	Cla.Mod[1]	Mod.Cla[2]	Global[3]	p.value[4]	v.test[5]
Cutting pathways=Local fields, Communautary groups	91	39	14	0	7,66
Types of foliar symptoms=No symptoms	69	51	24	0	6,73
Farming system=Monocropping + polycropping	100	15	5	0	4,94
Number of whiteflies="No whiteflies"	53	38	23	0	3,58
Presence of weeds=No	38	70	60	0,02	2,27
Cluster 2 (n=85; 35%)					
Number of whiteflies= "1-10"	62	86	48	0	8,89
Land tenure=rented	65	69	37	0	7,61
Types of foliar symptoms=Systemic and localized	67	66	34	0	7,55
Cutting pathways=Neighbour countries	100	29	10	0	7,35
Cluster 3 (n=81, 33%)					
Types of foliar symptoms=Systemic and on the whole plant	96	63	22	0	11,16
Land tenure=Owner	50	96	63	0	8,26
Cutting pathways=Local fields	42	86	67	0	4,67
Number of whiteflies="11-20"	56	33	20	0	3,69
Presence of weeds=yes	43	53	40	0	2,84
Infection status=UCBSV	52	21	13	0,02	2,35
Farming system=Monocropping	37	81	72	0,02	2,26

[1]: Percentage of individuals showing the characteristic (variable=modality) who belongs to the cluster.

[2]: Percentage of individuals of from that cluster showing the characteristic (variable=modality).

[3]: Percentage of individuals showing the characteristic (variable=modality) in the whole population (n=246).

[4]: Pearson’s Chi squared test; Fisher’s exact test. It assess the strength of the link between a modality and a cluster. The p-value of a modality is less than 5% when that modality is significantly linked to the cluster that is being interpreted. Only modalities with p-values less than 5% are shown.

[5]: Test value: transformation of the p-value into a quantile of the normal law. When the V-test is negative, it means that the modality is significantly less present (under-expressed) in that cluster compared to the presence of this modality in the whole dataset (these modalities were not included in the table). However, if the v-test is positive, the corresponding modality is significantly more present (over-expressed) in that cluster (Husson, Lê, and Pagès 2011).

The cluster 1 is characterized by the presence of a consistent number of fields (91%) that were grown using cuttings that farmers obtained from 2 pathways (i.e. their own fields and the social organizations to which they belong to). In these fields, farmers combined monocropping and crop association farming systems to grow cassava and majority of the plants from this cluster (69%) did not display CBSD symptoms. Fields

from cluster 1 were 11-month-old at the time of survey and had CBSD mean symptoms incidence of 25%, significantly lower than the other clusters (table III-4). This cluster also includes high number of fields characterized by the absence of weeds (38%) or whiteflies (53%).

Table III-4. Continuous (quantitative) variables associated with the description of clusters from HCPC analysis

	In cluster[1]		**Overall[2]**		**p. value**	**v. test[4]**
	Mean	sd[3]	Mean	Sd[3]		
Cluster 1 **(80 fields)**						
Plant age [months]	11.2	9.6	3	2.7	2.7e-08	5.6
Mean symptoms incidence (in %)	25	48	14	28	2.1e-15	7.9
Cluster 2 **(85 fields)**						
Mean symptoms incidence (in %)	42	48	22	28	0.012	2.5
Cluster 3 **(81 fields)**						
Mean symptoms incidence (in %)	74	48	22	28	1.3e-24	10.2
Plant age [months]	8.6	9.7	3	2.7	6.8e-06	4.5

[1]: Statistics of continuous variables in the cluster [2]: Statistics of continuous variables in the whole subpopulation

[3]: Standard deviation

[4]: Test Value: transformation of the p-value into a quantile of the normal law. When the V-test is negative, it means that the modality is significantly less present (under-expressed) in that cluster compared to the presence of this modality in the whole dataset (these modalities were not included in the table). However, if the v-test is positive, the corresponding modality is significantly more present (over-expressed) in that cluster (Husson, Lê, and Pagès 2011).

The ten cassava fields in which farmers used cuttings obtained from neighbor countries are grouped in the cluster 2. These fields are mainly held by the farmers

in a tenancy mode (65%). The abundance of whitefly varied between 1 and 10 in most of the fields (62%). Most cassava plants from cluster 2 (67%) displayed CBSD symptoms localized on leaves and/or on the stem.

The cluster 3 is characterized by the presence of CBSD symptoms on both leaves and stems in nearly all fields (96%). Additionally, around 50% of cassava fields are held by farmers in a private ownership. Cassava is grown in monocropping in 37% of the field using cuttings that farmers obtained exclusively from their own fields. The absence of weed was recorded in 57% of the fields while the number

of observed whiteflies was higher (11-20) than cluster 2 in 56% of the fields. The average age of cassava plants at the time of survey was 8.5 months. Most of fields from this cluster (52%) were found to be infected by UCBSV and had a mean symptoms incidence higher than what was reported in the two previous clusters (74%).

The spatial arrangement of the above-described clusters is shown on the figure III-3. Fields initially classified as belonging to the site 2 (South) are entirely included in the cluster 1 as it is the case for the majority of fields initially defined as belonging to the site 5 (West). However, there is not a clear spatial demarcation between areas occupied by clusters 2 and 3 despite the fact that much more fields belonging to the cluster 3 presented a tendency to aggregate on the northern part of the border with Burundi. Additionally, the clusters 2 and 3 occupied the areas initially defined as sites 1 (North), 3 (Center) and 4 (East).

III.4.2. CBSVs detection and symptom incidence within the study area and the clusters

The molecular detection of CBSVs in the region and per cluster is shown in the table III-5. Overall, samples from 77 fields (31.3 %) out of the 246 were tested positive for CBSVs (supplementary material 4). The geographical localization is shown in supplementary material 5. Fields infected by CBSV alone were evenly distributed along the survey area while fields infected by UCBSV or by both species were concentrated in the Northern and central parts or in the Eastern part of the surveyed area respectively. Among infected samples, UCBSV in single infection was the most prevalent (42.9%) followed by single infection with CBSV (35.1%) and mixed infection with CBSV and UCBSV (22,1%).

Table III-5**.** Percentages of CBSVs detection and of symptom incidence according to clusters of the study area

Characteristic	**Cluster 1**, [80][1]	**Cluster 2**, [85]	**Cluster 3**, [81]	**Overall**, [246]	**p-value**[2]
Infection status (molecular incidence)					0.021
CBSV	15%	7%	11%	11%	
CBSV+UCBSV	6%	8%	6%	7%	
UCBSV	5%	14%	21%	13%	
Negative	74%	71%	62%	68%	
Mean symptoms incidence (%)	25 c	42 b	74 a	47	
SD	15	22	22	20	
Min	3	7	17	9	
Max	67	83	100	83	

[1] [n] : Numbers in brackets represents the number of fields
[2] Pearson's Chi-squared test

In addition, the CBSVs' detection percentage varied significantly across clusters (P-value = 0,021. Infection by CBSV was most prevalent within the cluster 1 (15%) while the mixed infection was most present within the cluster 2 (8%). The cluster 3 was characterized by the higher prevalence of UCBSV (21%).

According to the symptoms incidence results (table III-5), the first cluster was characterized by an average incidence (25%) significantly lower than the cluster 2 (42%) which was also significantly lower than cluster 3 (74%). The number of fields for each level of symptoms severity score across the three clusters as well as the number of fields according to symptom types is shown in the table III-6

.

Table III-6**.** Types and severity of foliar symptoms observed on surveyed plants across clusters.

Characteristic	Clusters			Overall ($N = 246$[1])	Value of p[3]
	Cluster 1 [$N = 80$][1]	Cluster 2 [$N = 85$]	Cluster 3 [$N = 81$]		
Foliar symptoms types[2]					<0.001
LL	51%	14%	7%	24%	
NO	31%	66%	4%	34%	
SL	15%	20%	26%	20%	
SW	3%	n.a.[4]	63%	22%	
Severity score[5]					<0.001
1	37%	26%	2%	20%	
2	45%	39%	25%	35%	
3	9%	26%	40%	26%	
4	4%	0%	22%	0.1%	
5	5%	9%	11%	0.1%	

[1][n]: Numbers in brackets or parentheses represents the number of fields.

[2]Types of foliar CBSD symptoms based on distribution of leaf chlorosis and stem lesions on the plant: systemic and on the whole plant (SW), systemic on leaf or stem parts but localized (SL), only on lower leaves (LL).

[3]Pearson's Chi-squared test.

[4]Not applicable. It means that the modality related to this infection type was not observed.

[5]Foliar symptom severity score based on 1–5 scale (Alicai et al., 2016): 1 = No visible symptoms (not shown in Table 4), 2 = mild vein yellowing or chlorotic blotches on some leaves, 3 = pronounced/extensive vein yellowing or chlorotic blotches on leaves but no lesions or streaks on stems, 4 = pronounced/extensive vein yellowing or chlorotic blotches on leaves and mild lesions or streaks on stems, 5 = pronounced/extensive vein yellowing or chlorotic blotches on leaves and severe lesions or streaks on stems, defoliation and dieback.

Results suggested that most fields belonging to the cluster 1 had typical CBSD symptoms located on lower leaves (41 plants out of 80). This was not the case for the two other clusters where symptoms were rather absent in most fields (cluster 2: 56 fields out of 85) or scattered on the whole plant (systemic-like symptoms, on both leaves and stems) in the cluster 3. Results further suggests that differences in proportions of fields for each modality of foliar symptoms across clusters are statistically significant (p-value Pearson's Chi-squared test = 0,001). Symptom severity score of 1, 2 and 3 included most fields in the survey (82%) with significant differences between clusters. For example, eighty two percent of the fields presented a severity score of 1 or 2 for the cluster 1 while they represented 63% and 27% of the fields for cluster 2 and 3 respectively. Cluster 3 was characterized by a higher proportion of field with severe symptoms (4 and 5 scores): 33% compared to 10% and 9 % for cluster 1 and 2 respectively.

III.4.3. Pathways of cuttings used by farmers

Cuttings used for planting cassava fields were obtained from diverse pathways (table III 7). Pathways that required no or less charges and located in the closest farmer's environment (representing no or low charges, without travelling long distances and involving actors closely related to the farmer) were the most used. These corresponded to cuttings provided by the farmer himself (obtained from own fields grown with cassava or fields from the neighbors) as well as cuttings obtained from cooperatives to which the farmer belongs or could access (associations, cooperatives, Non-governmental Organization-NGO-, peasant local associations etc.). Unless the fact that some farmers used cuttings from only one pathway (cuttings originated rather exclusively from farmer's own production: 22% or exclusively from cooperatives: 3%), most of farmers used a combination of different pathways to access cuttings (75%). In fact, the situation in the surveyed area suggested that most of farmers used a combination of 2 or 3 pathways to obtain cuttings

Table III-7. Proportion of fields grown by types of cassava varieties from different pathways.

Characteristic	**Local varieties** [1][1]	**Improved varieties** [126][3]	**Both** [119]	**Overall** [246][1]	**p-value**[2]
Cutting pathways					0.4
Farmers (F)	-	23% [27]	23% [27]	23% [54]	
F+Cooperatives (C)	100% [1]	44% [53]	41% [49]	43% [103]	
F+C+Market	-	15% [18]	8% [10]	12% [28]	
F+C+Multiplier	-	17% [20]	26% [31]	21% [51]	
F+Neighbor countries	-	1.7% [2]	2% [2]	2% [4]	

[1] [n]: Numbers in brackets represents the number of fields

[2] Pearson's Chi-squared test

-: The modality related to this cutting pathway was absent.

[3]: No data on the pathways used to obtain cuttings of improved varieties grown in 6 fields could be obtained.

The table presented in supplementary material 6 summarizes the proportion of fields grown with cuttings obtained from each category of pathways identified as well as the means used by farmers to obtain cuttings across clusters. Globally, 103 surveyed fields were grown with cuttings that farmers obtained from a dual pathway source: from their own fields (and fields from neighboring farmers) and from social organizations they belong to. Fifty-four fields were also identified as being grown exclusively with cuttings that farmers obtained from their own fields (or fields of neighboring farmers). Cuttings originating from seed multipliers were found to be grown in 51 fields, always in combination with cuttings originating from farmer's own fields and Cooperatives. Majority of fields grown with cuttings from seed multipliers were in cluster 2 (39 fields). Few fields (4) were found to be grown with cutting originating from neighbor countries.

Cassava fields sown by cuttings that farmers obtained exclusively from their previously grown fields were mostly located in the cluster 1 (26 fields out of 54) while cuttings resourced from Cooperatives pathways are mostly found in the cluster 3 (44 fields out of 103). Cutting obtained from Market (18 fields out of 28) and from Seed multiplier (20 fields out of 51) were mostly grown in the cluster 2. Results further suggested that differences observed in proportions of fields grown by cuttings from different pathways across sites were not different (Pearson's Chi-squared test = 0.04).

Results further suggested that all farmers (100%) have used cuttings obtained for free while nearly half of them (48%) have paid for cuttings. Other sources of cuttings

were obtained by working in the field of other farmers (6%) or by sharing production after harvest (10%) (supplementary material 5).

The table III-7 also shows the proportion of fields grown by different types of cassava varieties from each of the pathways described above. Generic names of improved varieties are most of the times changed by farmers during the adoption process to adapt them to local dialects. During the survey, the challenge was to identify a variety as local or improved despite the local name assigned by farmers.

Physical traits or appearance of cassava plants were mainly used to determine if a variety was local or an improved one. This identification strategy was rendered efficient by including into the survey team local agronomist officers able to identify varieties in the field. Local names of all the varieties (improved as well as local) identified during this work are shown in the table presented in supplementary material 7. Results suggested that most of fields were grown either with improved varieties only (51%), either with a mixture of improved and local varieties (48%) (table III-9). A single field was sown exclusively by local varieties.

The table presented in supplementary material 8 summarizes the proportion of each type of infection according to cutting pathways. Results suggested that among fields that tested positive to UCBSV infection (34 fields), more than the half (19 fields) were grown by cuttings originated from the dual source Cooperatives + farmers. Among fields that tested positive to CBSV infection (28 fields) and to mixed CBSV+UCBSV infection (18 fields), majority of them were grown with cutting originating from a dual (farmer's + Cooperatives) and exclusively from farmer's pathways (respectively 17 fields out of 28 for CBCSV infection, and 13 fields out of 18 for mixed CBSV+UCBSV infection). On the other side, the overall proportion of fields free from infections is higher (78%) when they are grown using cuttings issued from seed multipliers.

III.4.4. Seed system risk factors associated with CBSD

After fitting an initial model containing 37 candidate predictors (most of which were studied in previous sections), a final model containing 9 predictors, all statistically significant, was optimized using a combined forward + backward stepwise procedure (Chambers 1992; Venables and Ripley 2002). The results obtained are summarized in the table III-8.

Table III-8: Prediction of risk factors associated with CBSD (based on RT-PCR detection).

Characteristic	Bivariate statistics				Prediction		
	Absence of infection [98][1]	Presence of infection [50]	Overall [148]	p-value[2]	OR[3]	95% CI[3]	p-value
Assistance/ support by extension services				n.s.[4]			**0.05**
No	62% [43]	38% [26]	100% [69]		1.00	*Reference*	
Yes	70% [55]	30% [24]	100% [79]		0.32	0.08, 1.03	0.041
Knowledge of cassava pests and diseases				n.s.[4]			**0.002**
No	43% [3]	57% [4]	100% [7]		1.00	*Reference*	
Yes	67% [95]	33% [46]	100% [141]		29.1	3.23, 355	0.004
Knowledge of management practices				**0.064**			**0.008**
Yes	77% [34]	23% [10]	100% [44]		1.00	*Reference*	
No	62% [64]	39% [40]	100% [104]		0.14	0.02, 0.62	0.016
Which distance to acquire cuttings ?				0.5			**0.001**
Very close (<1 Km)	71% [49]	29% [20]	100% [69]		1.00	*Reference*	
Close (1-5 Km)	60% [12]	40% [8]	100% [20]		0.96	0.23, 4.22	n.s.[4]
Far (5-10 Km)	59% [16]	41% [11]	100% [27]		0.3	0.66, 2	n.s.[4]
Very far (>10 Km)	66% [21]	34% [11]	100% [32]		0.08	0.02, 0.33	0.001
Proximity to town (Uvira)				0.5			**0.036**
Very Close (<1Km)	75% [6]	25% [2]	100% [8]		1.00	*Reference*	
Close (1-5Km)	78% [18]	22% [5]	100% [23]		0.59	0.03, 7.51	n.s.[4]
Far (5-10Km)	66% [23]	34% [12]	100% [35]		0.12	0.01, 1.26	n.s.[4]
Very Far (>10Km)	62% [51]	38% [31]	100% [82]		0.09	0.00, 0.85	0.061
Proximity to borders				n.s.[4]			**0.05**
Very Close (<1Km)	68% [39]	32% [18]	100% [57]		1.00	*Reference*	
Close (1-5Km)	67% [28]	33% [14]	100% [42]		1.16	0.56, 2.41	n.s.[4]
Far (5-10Km)	65% [20]	36% [11]	100% [31]		2.07	0.82, 5.31	n.s.[4]
Very Far (>10Km)	61% [11]	39% [7]	100% [18]		4.45	1.30, 17.4	0.023
Methods used to manage CBSD				**0.027**			**0.001**
Use cuttings from symptomless plants	76% [34]	25% [11]	100% [45]		0.43	*Reference*	
Use local varieties	85% [29]	15% [5]	100% [34]		1.00	0.97, 5.86	n.s.[4]
Use certified varieties	53% [23]	45% [20]	100% [43]		2.25	0.89, 5.89	0.001
Cutting pathways				n.s.[4]			**0.001**
Farmers (F)	57% [20]	43% [15]	100% [35]		1.00	*Reference*	
F+Cooperatives (C)	64% [41]	36% [23]	100% [64]		2.06	0.55, 7.81	n.s.[4]
F+C+Market	67% [4]	33% [2]	100% [6]		10.7	0.56, 272	n.s.[4]
F+C+Multiplier	75% [30]	25% [10]	100% [40]		7.96	1.55, 53.1	0.019
F+Neighbor Country	100% [4]	0% [0]	100% [4]		6.051	0.00, NA	n.s.[4]
(Intercept)					17	0.49, 700	0.12

[1] [n]: numbers in brackets represents the number of fields

[2] Pearson's Chi-squared test; Fisher's exact test

[3] OR = Odds Ratio, CI = Confidence Interval
[4]n.s.= the p-value is > 0.05

Results showed that cassava fields owned by farmers who received supports (training, advising or field visit) and had a certain knowledge of cassava pests and diseases as well as on management practices against CBSD, are significantly more likely to be free from CBSVs infection compared to fields belonging to the other farmers. Beyond these aspects related to farmer's awareness, factors related to the distance location of cassava fields were also found to significantly impact the outcome of CBSD infections. In fact, cassava fields located very far (more than 10 Km) from the borders as well as from Uvira town were significantly and highly associated with the absence of CBSD compared to fields that were very close (less than 1 km) toborders and Uvira (P-Value < 0.05). Also, fields grown with cuttings obtained by farmers from very far locations (more than 10 km) appeared to be significantly less prone to CBSVs infections compared to fields grown with cuttings that farmers obtained in nearby locations (less than 1 Km) (P-Value =0.05). Additionally, fields in which farmers were using certified varieties to grow cassava were significantly and highly associated to the absence of CBSD compared to fields grown with cuttings taken from asymptotic plants (P-value < 0.005). Results further suggest that when farmers envisage the option of using cuttings from seed multipliers pathways to grow cassava, the risk of having infection with CBSVs in their fields is significantly lower (less probable) (P-value <0.05) compared to situations where fields where grown with cuttings taken exclusively from farmer's own fields.

III.5. Discussions

The preliminary description of the study area based on socioeconomic, ecological and agronomic parameters supported the existence of various sites (Table I1 and Figure III-2). However, to highlight the role of seed cutting pathways in the epidemiology of CBSD, this information should be completed by plant disease observation and virus detection. Such phytopathological information opened the possibility of associating CBSD-infected plants obtained from a particular seed pathway to a set of explanatory parameters depicting the environmental context that could explain the outcome of infection. Even tough symptoms presence on a plant is a proxy of infection, they can be misleading in certain circumstances particularly when they are not specific as it is the case for the CBSD. It appeared therefore rationale to use molecular diagnosis so that errors due to the misidentification of the causal agent could be significantly lowered. This integration of several multidisciplinary data facilitated an in-depth description of the study area regarding the objectives of the study and allowed designing local-adapted approaches to act at the formal-informal interface of the cassava seed system and to mitigate the dissemination of CBSD.

Before discussing the results, it is important to state that, despite the multidisciplinary approach used in this study, our survey, as any survey, presented some limitations that might have introduced bias on the conclusion drawn. For example, the use of some pre-selected questions in the questionnaire, the random sampling and observation on field (impacting sample representativeness) as well as the limited number of observations.

The molecular diagnostic revealed the presence of the two viral species known to cause the disease in single or mixed infection. The prevalence of both species was similar, 18% for CBSV and 20% for UCBSV. Nevertheless, the analysis of prevalence for each cluster revealed contrasted situations: CBSV prevalence was between 17% (cluster 3) and 21% (cluster 1) while UCBSV prevalence was more heterogeneous ranging from 11% (cluster 1) to 20% (cluster 3). UCBSV tended to be restricted in the lower altitude zones where it showed a spot-like distribution.

Epidemiologically, the identification of three clusters made sense. Indeed, the homogeneous cluster 1 gathered most of villages located in the high altitude zone where the fields presented the lowest average of symptom incidence rate and virus detection (with CBSV the most prevalent virus) while the whiteflies were very rare. On the other side, a heterogeneous zone in the low altitude area (cluster 2 and 3) was characterized by the higher prevalence of UCBSV-infected fields but with distinct symptoms incidence rates and cutting pathways. In the cluster 2, most diseased fields showed systemic symptoms that were localized either on leaves or stems, presented an average symptom incidence rate higher than in the cluster 1. In addition, among fields grown by cuttings originating from seed multipliers pathways, a consistent number of them were found in the cluster 2. In the cluster 3, most cassava fields were grown in monoculture, most of them were colonized by weeds and presented the highest number of whiteflies, virus infection as well as the highest symptoms

incidence rate while presenting typical systemic-like symptoms on the plant. Previous studies have shown that growing conditions (temperature, rainfall and altitude) can induce variation in the expression of foliar symptoms of CBSD (Mohammed et al. 2012) as well as in the dynamic of whiteflies (E. Bisimwa, Walungululu, and Bragard 2012) thus globally reducing disease incidence as for the cluster 1. The presence of CBSD-like symptoms on lower leaves is often misleading because symptoms can be confused with those due to normal leaf senescence, and the prevalence of symptoms type restricted on lower leaves could not be necessarily considered as indicative of virus infection. The high prevalence of infected fields in the low altitude areas (clusters 2 and 3) as well as their high symptoms incidence and severity is in agreement with previous findings (E. Bisimwa, Walungululu, and Bragard 2012) suggesting that the disease pressure was decreasing with the rise of altitude.

Furthermore, it is worthful to note that in the clusters 2 and 3 the level of symptom incidence was higher compared to the level of molecular incidence (percentage of fields tested positive to RT-PCR). This situation suggested that there were a consistent number of plants with symptoms but which were negative when tested by RT-PCR. This situation could be explained by the non-specificity of CBSD symptoms because they are reported to be an inconsistent way of identification of CBSVs (G. Rwegasira and Rey 2012) but also, the inclusivity of primers used for indexing could be pointed out. While knowing that CBSV and UCBSV genomes are reported to evolve rapidly under high disease pressure (Titus Alicai et al. 2016b; Ndunguru et al. 2015) it is possible that there could be different CBSVs isolates circulating in these clusters. These findings underpins an urgent need to unveil full genomes of CBSVs strains in this area to support designing adapted primers necessary for an efficient diagnostic (I. P. Adams et al. 2018). High throughput sequencing technology (Massart et al. 2014) can contribute to decipher this problem and will help designing new primers required for increasing the inclusivity of molecular tests. In addition, the viral species complex causing symptoms on cassava might not be fully understood so far. For example, two new viral species belonging to the genus Ampelovirus (MEaV1 and MEaV2*)* were previously described for the first time infecting cassava plants and one of these species was detected in the area where this study was conducted (Uvira territory) (Kwibuka et al. 2021) . Nevertheless, the association of these viruses with disease symptom is not yet studied, so the phytosanitary risk posed by these new species is not known yet.

As suggested in this study, cutting pathways would be determinant in the outcome of CBSVs infections through the fact that when farmers took the option of using cuttings provided from seed multipliers, additionally to cuttings taken from other pathways, the prevalence of CBSVs infection lowered significantly. However, additive effects, resulting from the fact that several pathways were used at the same time by most farmers, might have been obscuring the real role of each cutting pathway.

Improved varieties are widely used. This is consecutive to previous interventions for mitigating devastations caused by the CMD pandemic in this area. Local varieties were progressively abandoned before the outbreak of CBSD against which most of improved CMD-resistant varieties became susceptible. Currently, the research for developing CBSD-tolerant varieties is ongoing (Manze et al. 2021; Sheat et al. 2019).

Nevertheless, due to a scarcity of healthy cuttings and disappointments often encountered after using cuttings expected to be healthy, farmers return to traditional varieties that they have been using for years. This scarcity of healthy cuttings is therefore primarily due to the lack of multiplication of phytosanitized tolerant varieties that are available as well as to the lack of varieties with dual resistance to CBSD and CMD for multiplication and dissemination."

Factors related to farmer's awareness, especially the assistance from extension services, the knowledge of cassava pests and disease, the knowledge of management practices of CBSD as well as the use of certified varieties, are key determinants to limit CBSVs infection. This is absolutely relevant because being aware of CBSD, its symptoms as well as management practices helps farmers in identifying diseased plants and taking appropriate decisions for managing and mitigating the disease. This is supported by the fact that using cuttings from seed multipliers lowered significantly the risk of CBSD. Here, it is very important to emphasize that seed multipliers benefited more from extension services supports and seemed to develop more skills and knowledge than ordinary farmers. Therefore, intensifying actions aiming at raising farmers awareness on CBSD control would constitute an effective option to mitigate the disease.

It was also very interesting to found that parameters related to the geographical location of fields, particularly in relation with the national border, were significantly associated with the outcome of CBSD infection. At this stage we do not know if this association is causative or spurious, border proximity hiding other factors. In addition, only 4 farmers mentioned neighboring countries as pathways of seeds. Nevertheless, it would be very important to mention that the proportion of fields effectively grown with cuttings originating from neighbor countries might be higher than what was reported in this study because the unauthorized transboundary movement of planting material is forbidden, therefore despite guarantees of non-divulgation of information, respondent might not explicitly admit to be engaged in such kind of exchanges. Such phenomenon of underestimation of transboundary movement of planting material has been already suggested (P. L. Kumar et al. 2021). Therefore, if this association is further demonstrated as causative, farmers living along borders between D.R. Congo, Burundi and Rwanda should be particularly targeted for awareness raising on importance of avoiding moving cassava planting material across borders without following proper international regulations (IPPC 2021).

Mapping epidemiological aspects of a disease in a range of environments can identify locations where investments in extension and farmer support are most likely to be effective (Buddenhagen et al. 2020). Results provided in this study suggested that efforts to promote a clean seed system in the study region could therefore target the areas covered by the cluster 1 and use it as multiplication site due to its low disease pressure and low vector population density. Additionally, in the same analysis, efforts of extension work should be focused on raising farmer awareness of CBSD to sustain the effectiveness of control strategies. Such extension efforts must target both local and national organizations involved in the farmer's Cooperatives and would put much attention on areas covered by clusters 2 and 3.

The use of improved varieties in the previous years did not guarantee effective protection against CBSD as observed in the study area because most of these varieties were tolerant to CMD only; viruses kept multiplying inside their vegetative tissues thus leading to increased virus load over years (Manze et al. 2021). Therefore, there is an urgent need to insist on the application of a rigorous phytosanitation program to ensure planting material will be subjected to a cleaning process that will lower the virus load. This will necessitate a good expertise in plant virus diagnostic to ensure that sensitive and inclusive tests are applied on elite materials before they could be multiplied and supplied to farmers. These interventions must also include local (traditional) varieties and could contribute to reduce the disease pressure observed into clusters 2 and 3. A key element on which an adapted program of integrated pest and disease management (IPDM) is based for reducing the disease severity and preventing a disease from further spread include the use of healthy/resistant varieties. Within the surveyed country, existing organizations involved/ in charge of developing and supplying healthy seeds and regulating activities within the formal seed sectors include local, national as well as international organizations/institutions. The predominance of informal actors within this seed system (private seed sellers, village seed multipliers etc.) allows disseminating seeds where some bigger institutions and even public sector does not reach out. However, this capacity of delivering seeds on the last mile of the territory is undermined by the lack of access to healthy and resistant varieties.

Results from this study further suggested that it would be important to empower and to promote local cassava-seed multipliers, particularly in the cluster 1, as they have been shown to be more reliable in delivering disease-free materials. However, it was shown that farmers had to travel more than 10 Km to access less risky cuttings. This could mean that the number of actors susceptible of delivering good-quality planting material to farmer's is still too low. This could represent an opportunity to draw farmer's attention to business opportunities offered through cassava seed system activities, thus giving a scope to turn this activity commercially attractive in this area.

In this context, promoting the cassava seed system will undoubtedly raise the need of reinforcing the mechanisms of controlling seed multiplication fields. This could be done by rigorous inspections by well-trained inspectors as well as by testing elite cassava materials used for propagation using sensitive molecular techniques preferably on-site (J. a. Tomlinson et al. 2013). However, this will raise the question of sustainability due to the high costs involved in running such techniques. Fortunately, sensitive as well as easy-to-use kits [such as Lateral Flow devices (LFDs) or RT-LAMP] that can be implemented on field and that require reasonable consumables, resources and instruments have been developed (J. a. Tomlinson et al. 2013) and, after proper training and validation, could allow local seed multipliers to directly identify healthy mother plants candidate for multiplication and dissemination. This could be an additional aspect where expertise from local Universities would be crucial in supporting the establishment of an efficient cassava seed system.

Acknowledgements: Members of the Integrated and Urban Plant Pathology Laboratory (Angelo Locicero, Nuria Fontdevila and Coline Temple) are sincerely acknowledged for their help in molecular analysis. Members from the agricultural faculty of the UCB are acknowledged for their help during the survey (Safari Baluku) and data analysis (Jules Masimane).

III.6. Supplementary material

Supplementary material 1. Questionnaire used for the epidemiological survey in cassava farmer's fields. Available on the following link:

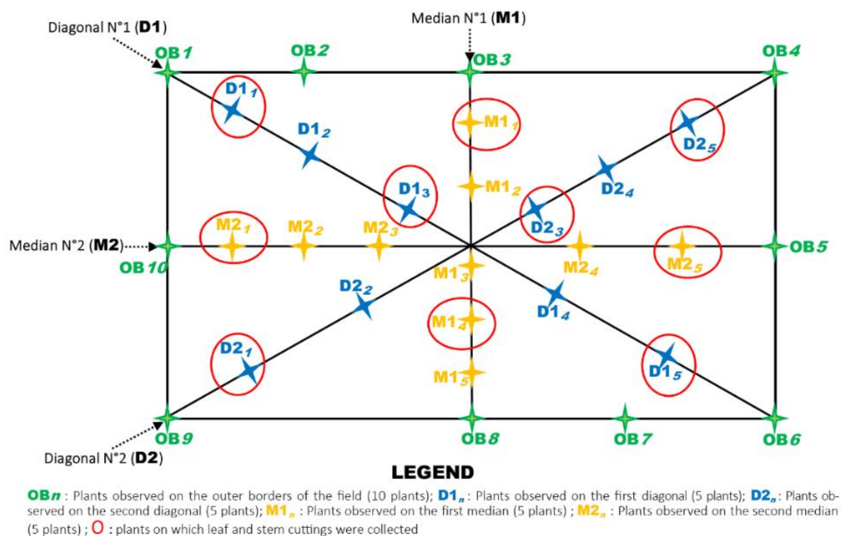
https://www.frontiersin.org/articles/10.3389/fpls.2022.803980/full#supplementary-material

Supplementary material 2. Questionnaire used for farmers interviews on seed systems. Available on the following link:

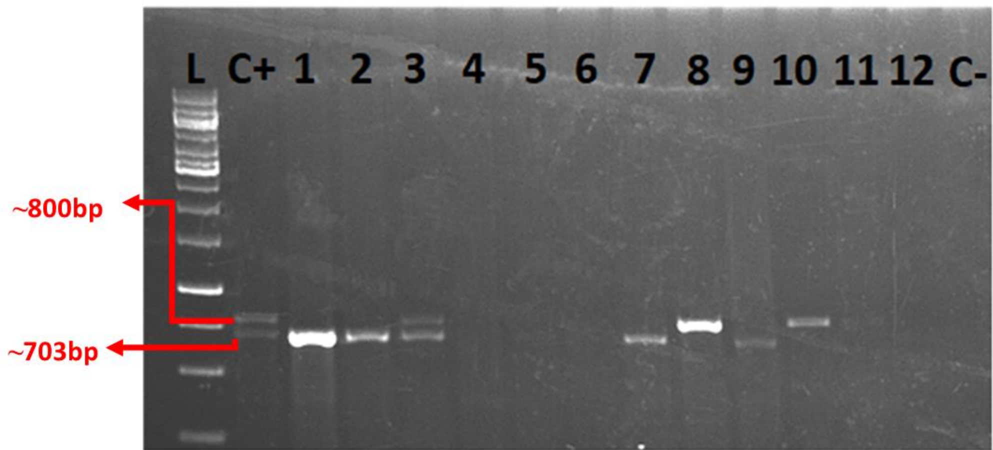
https://www.frontiersin.org/articles/10.3389/fpls.2022.803980/full#supplementary-material

Supplementary material 3. Questionnaire used for interviewing farmers specialized in cassava seed multiplication. Available on the following link:

https://www.frontiersin.org/articles/10.3389/fpls.2022.803980/full#supplementary-material

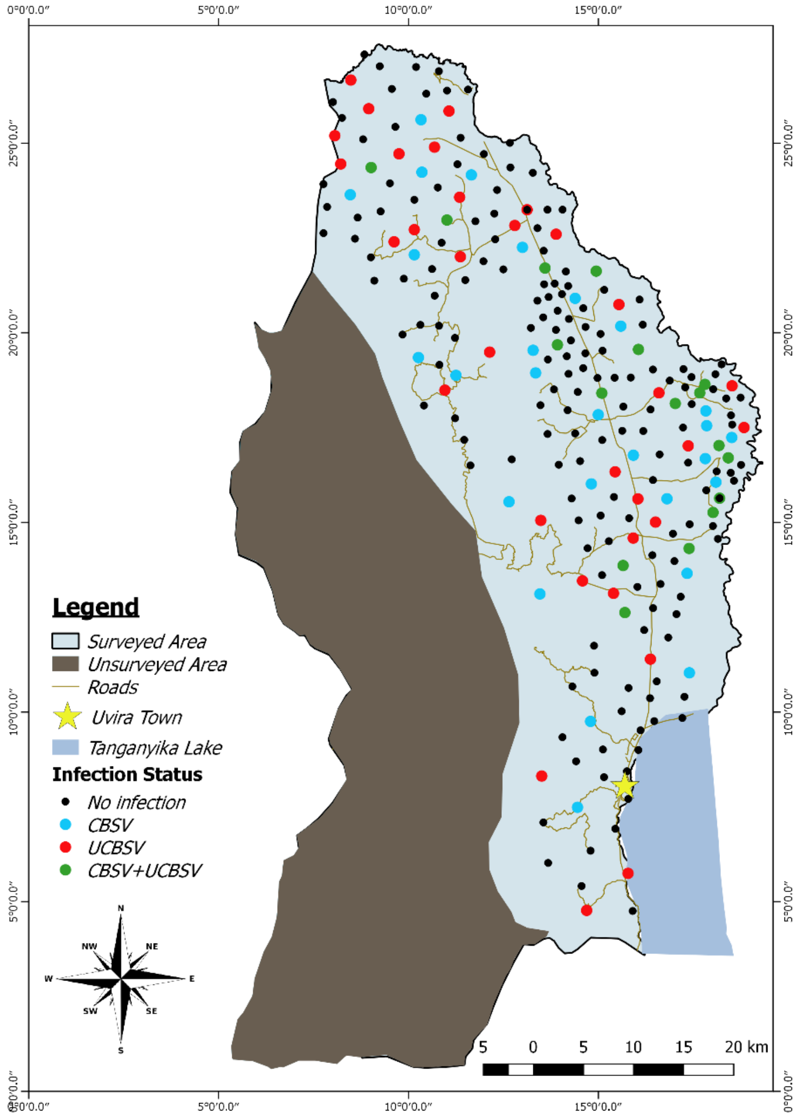


Supplementary Figure 1 Schematic representation of a cassava field showing the selection strategy and process for field observation and sample collection.



Supplementary Figure 2. Results from molecular detection showing the agarose gel with amplicons:

Detection of CBSV (~703 bp) and UCBSV (~800 bp) in samples from the site 1 (1 and 6), site 2 (2 and 7), site 3 (3 and 8), site 4 (4, 5 and 9) and site 5 (10, 11 and 12). 1kb molecular ladder (L) is shown on the left side of the gel. Positive (C+) and negative controls (C-). Numbers 1 to 12 are representative samples retained for the illustration.



Supplementary figure 3. Geographical mapping of results from molecular detection

Supplementary Table 1. Proportion of fields following pathways used to acquire cuttings and following means used to obtain cassava cuttings per clusters.

Characteristics	**Cluster 1** [80][1]	**Cluster 2** [85]	**Cluster 3** [81][2]	**Overall** [246]	**p-valu**
Cutting pathways					0.04
Farmers (F)	48% [26]	32% [17]	20% [11]	100% [54]	
F+Cooperatives (C)	30% [31]	27% [28]	43% [44]	100% [103]	
F+C+Market	18% [5]	64% [18]	18% [5]	100% [28]	
F+C+Seed Multipliers	31% [16]	39% [20]	29% [15]	100% [51]	
F+ Neighbour Country	50% [2]	50% [2]	-	100% [4]	
Means of cutting obtention					0.001
Free	50% [40]	28% [24]	56% [45]	44% [109]	
+Money	28% [22]	55% [47]	41% [33]	42% [102]	
+Work/Yield	23% [18]	17% [14]	4% [3]	14% [35]	

[1] [n]: Numbers in brackets represents the number of fields grown by cuttings obtained from the corresponding pathways/pathways

[2]: No data on the pathways used to obtain cuttings in 6 fields of the cluster 3 could be noticed.

-: The modality related to that cutting pathway was absent in the corresponding cluster.

[2] Pearson's Chi-squared test

Supplementary Table 2. Number of fields grown by improved as well as local cassava varieties identified in each cluster

Improved varieties				
Local names	***Cluster 1***	***Cluster 2***	***Cluster 3***	***Overall***

Dorothea	7	7	29	43
V8	24	1	14	39
Sawa sawa	3	24	10	37
Magouringware	8	8	11	27
Mabwaki	1	7	3	11
Nabwigoma	7	0	1	8
Bwika	3	3	0	6
Kantintima	0	5	1	6
Namuzungu	4	1	1	6
Rava	0	3	2	5
Nabwilalanga	1	0	3	4
Liyayi	0	0	3	3
Mayombe	3	0	0	3
Mvuama	0	0	3	3
Nakahegere	1	0	2	3
Mukombe	0	2	0	2
Naluvuzi	0	0	2	2
Ndunda	0	2	0	2
Butamu	0	1	0	1
Kihonya	0	0	1	1
Mabwilalanga	0	0	1	1
Migera	1	0	0	1
Nakaronda	0	1	0	1
Local Varieties				
Local names	***Cluster 1***	***Cluster 2***	***Cluster 3***	***Overall***
Nahunde	6	7	2	15
Nakarasi	1	4	2	7
Kahungu	1	0	2	3

Supplementary Table 3. Proportion of healthy and infected fields for each type of infection according to the pathways used to obtain cuttings

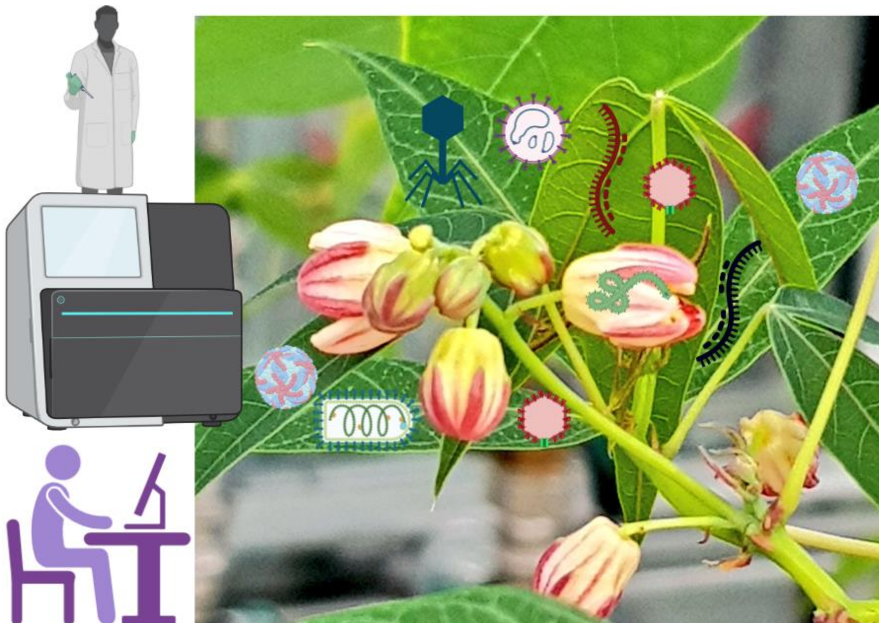
	CBSV INFECTION			UCBSV INFECTION			CBSV+UCBSV INFECTION			OVERALL (Infected	
Characteristic	**Absence** (N=218[1])	**Presence** (N=28[1])	**Overall**, N=246[1]	**Absence** **(N=212[1])**	**Presence** (N=34[1])	**Overall**, N=246[1]	**Absence** **(N=228[1])**	**Presence** (N=18[1])	**Overall**, N=246[1]	**Presence** (N=80[1])	**Absence** (N=166[1])
Cutting channels											
Farmers (F)	87% [47]	13% [7]	100% [54]	85% [46]	15% [8]	100% [54]	87% [47]	13% [7]	100% [54]	41% [22]	59% [32]
F+Cooperatives (C)	90% [93]	10% [10]	100% [103]	82% [84]	18% [19]	100% [103]	94% [97]	6% [6]	100% [103]	34% [35]	66% [68]
F+C+Market (M)	86% [24]	14% [4]	100% [28]	89% [25]	11% [3]	100% [28]	89% [25]	11% [3]	100% [28]	36% [10]	64% [18]
F+C+ Seeds Multipliers	88% [45]	12% [6]	100% [51]	94% [48]	6% [3]	100% [51]	96% [49]	4% [2]	100% [51]	22% [11]	78% [40]
F+Neighbor Countries	100% [4]	-	100% [4]	100% [4]	-	100% [4]	100% [4]	-	100% [4]	-	100% [4]
p-value2		**0.9**			**0.3**			**0.4**			**0.2**
Missing values	5	1	6	5	1	6	6	0	6	2	4

[1] % [n]

[2] Fisher's exact test

Chapter IV

IV. Novel ampeloviruses infecting cassava in Central Africa and the South-West Indian ocean Islands



"Sequence locally, think globally…"

Harris Lewin, the Darwin tree of life. 2021

Preliminaries

The results obtained from investigations conducted in the chapter three showed that the proposed framework allowed a proper characterisation of the CBSD epidemiology by identifying locally-adapted options for managing the disease. If a robust cassava seed system is a priority for the mitigation of the CBSD, it is also important that, at all stages of the system, healthy cassava plants can be selected for cuttings generation. This reinforces the need of reliable diagnostic test and a proper understanding of disease etiology. Unfortunately, the chapter three showed that there is still a recurrent issue in the diagnosis of the CBSD : several leaf samples collected from field-growing cassava plants showing CBSD-like symptoms on leaves are suggested as negative from the RT-PCR tests. Is it due to limits in the detection method used, or rather is it an etiological issue ? In the next chapters (four and five) we used the high throughput sequencing technologies to explore the diversity of viruses infecting the cassava samples collected in the study area.

ARTICLE 2

This chapter is a based on the article published in :
Viruses, 29 May 2021
Volume 13; Issue 6, 1030
https://doi.org/10.3390/v13061030.

Only text formatting has been operated to accommodate the published version

IV. Novel ampeloviruses infecting cassava in Central Africa and the South-West Indian Ocean Islands

Yves Kwibuka [1,2,*], Espoir Bisimwa [2], Arnaud G. Blouin [1], Claude Bragard [3], Thierry Candresse [4], Chantal Faure [4], Denis Filloux [5,6], Jean-Michel Lett [7], François Maclot [1], Armelle Marais [4], Santatra Ravelomanantsoa [8], Sara Shakir [9], Hervé Vanderschuren [9,10] and Sébastien Massart [1,*]

[1]Plant Pathology Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés, 2, 5030 Gembloux, Belgium;

[2]Faculté des Sciences Agronomiques, Université Catholique de Bukavu, BP 285 Bukavu, Democratic Republic of Congo; ebisimwa@yahoo.com

[3]Earth and Life Institute, Applied Microbiology-Phytopathology, UCLouvain, 1348 Louvain-la-Neuve, Belgium.

[4]Univ. Bordeaux, INRAE, UMR BFP, CS20032, CEDEX, 33882 Villenave d'Ornon, France.

[5]CIRAD, UMR PHIM, 34090 Montpellier, France.

[6]PHIM Plant Health Institute, Univ. Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France

[7] CIRAD, UMR PVBMT, Pôle de Protection des Plantes, Saint-Pierre, F-97410, Ile de la Reunion, France.

[8]FOFIFA-CENRADERU, Laboratoire de Pathologie Végétale, BP 1444 Ambatobe, Madagascar.

[9]Plant Genetics Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés, 2, 5030 Gembloux, Belgium.

[10]Laboratory of Tropical Crop Improvement, Division of Crop Biotechnics, Biosystems Department, KU Leuven, 3000 Leuven, Belgium

***Corresponding authors**: bisimwayves@gmail.com (Y.K.);
sebastien.massart@uliege.be (S.M.)

IV.1. Abstract

Cassava is one of the most important staple crops in Africa and its production is seriously damaged by viral diseases. In this study, we identify for the first time and characterize the genome organization of novel ampeloviruses infecting cassava plants in diverse geographical locations using three high-throughput sequencing protocols [Virion-Associated Nucleotide Acid (VANA), dsRNA and total RNA], and we provide a first analysis of the diversity of these agents and of the evolutionary forces acting on them. Thirteen new *Closteroviridae* isolates were characterized in field-grown cassava plants from the Democratic Republic of Congo (DR Congo), Madagascar, Mayotte, and Reunion islands. The analysis of the sequences of the corresponding contigs (ranging between 10,417 and 13,752 nucleotides in length) revealed seven open reading frames. The replication-associated polyproteins have three expected functional domains: methyltransferase , helicase, and RNA-dependent RNA polymerase (*RdRp*). Additional open reading frames code for a small transmembrane protein, a heat-shock protein 70 homolog (*HSP70h*), a heat shock protein 90 homolog (*HSP90h*), and a major and a minor coat protein (*CP* and *CPd* respectively). Defective genomic variants were also identified in some cassava accessions originating from Madagascar and Reunion. The isolates were found to belong to two species tentatively named Manihot esculenta-associated virus 1 and 2 (MEaV-1 and MEaV-2). Phylogenetic analyses showed that MEaV-1 and MEaV-2 belong to the genus *Ampelovirus*, in particular to its subgroup II. MEaV-1 was found in all of the countries of study, while MEaV-2 was only detected in Madagascar and Mayotte. Recombination analysis provided evidence of intraspecies recombination occurring between the isolates from Madagascar and Mayotte. No clear association with visual symptoms in the cassava host could be identified.

Keywords: *Manihot esculenta*; *Ampelovirus*; high-throughput sequencing; Central Africa; Indian Ocean islands; *Closteroviridae*

IV.2. Introduction

Approximately 40% of the world's area dedicated to the cultivation of roots and tubers is planted with cassava. In Africa, 56% of the total production of roots and tubers comes from cassava, 51% of which are from Nigeria and the Democratic Republic of Congo (DR Congo) (Food and Agriculture Organization of the United Nations 2017). Cassava is one of the most important staple crops, as it ranks fourth as a source of calories for human consumption (Food and Agriculture Organization of the United Nations 2017). Many of the poorest farmers and most undernourished households in Africa depend on cassava as a principal nutrition source. It also constitutes an important income source in rural and marginal areas and has multiple uses, most notably as a food security and regular food crop (Burns et al. 2010).

Due to its remarkable adaptability to a wide range of soil and environmental conditions, cassava is bestowed with resilience to global warming and climate change, and with a potential for better return under adverse soil and weather conditions (Lobell et al. 2008). The exceptional soil carbon sequestration properties of cassava makes it a potential crop for improving the green revolution fatigue (Jarvis et al. 2012).

Despite the large contribution of Africa to global cassava production (60%), its performance in terms of yield is the lowest (9 tons/ha on average) (Food and Agriculture Organization of the United Nations 2017). In most African cassava-producing areas, the yield is far below the potential (Fermont et al. 2009). In DR Congo, yields of approximately 8 tons/ha can be obtained under farmers' fields conditions, but yields are highly variable depending on the local pedo-climatic conditions.

Cassava suffers from many pests and diseases, including viruses, which can seriously affect the quality and quantity of the harvest as well as the quality of the planting materials. Two viral diseases are of major economic importance in sub-Saharan Africa, namely, cassava brown streak disease (CBSD) and cassava mosaic disease (CMD) (Vanderschuren and Rey 2017). CBSD is associated with two *Ipomovirus* species, collectively named cassava brown streak viruses (CBSVs), and CMD is associated with nine *Begomovirus* species, collectively named cassava mosaic geminiviruses (CMGs) (Vanderschuren and Rey 2017). Yield reduction due to CMD may be severe and losses up to 82% have been reported, especially in cassava plants dually infected with African cassava mosaic virus (ACMV) and the Ugandan strain of East African cassava mosaic virus (EACMV-UG) (Owor et al. 2004).

The family *Closteroviridae* is a large and diverse group of filamentous plant viruses (particles of 650−2,200 nm in length) with single-stranded RNA genomes. Some species can be transmitted semi-persistently by aphids, whiteflies, mealybugs or soft scales insects [8,9]. The members of this family are known to affect several crops of major economic importance, such as sugar beet, citrus, tomato, lettuce, potato, sweet potato, grapevine, pineapple, cherry, and some ornamentals (Karasev 2000; A G Blouin et al. 2013; Dey et al. 2019; G P Martelli et al. 2012; Verdin et al. 2017).

To date, 56 virus species have been classified as definitive or tentative members of the family *Closteroviridae* (Fuchs et al. 2020). They are grouped into the following

four genera: *Ampelovirus* (monopartite genome, mealybugs and soft scale insect vectors), *Closterovirus* (monopartite genome, aphid vectors), *Velarivirus* (monopartite genome, no known vectors) and *Crinivirus* (bipartite genome, whitefly vectors) [14,15].

The accepted demarcation criteria between the species in the *Closteroviridae* family [9,16] include the following aspects: (i) particle size, (ii) size of *CP*, as determined by the deduced amino acid sequence data, (iii) genome structure and organization (number and relative location of the open reading frames (ORFs)), (iv) amino acid sequence of the relevant gene products (*CP, RdRp, HSP70h*), differing by more than 25%, (v) vector species and specificity, (vi) magnitude and specificity of the natural and experimental host range, and (vii) cytopathological features (aspect of inclusion bodies and origin of cytoplasmic vesicles) (Fuchs et al. 2020).

Specifically, the genus *Ampelovirus* comprises species with linear particles of 1,400−2,000 nm long. Their genome is a monopartite positive sense, single-stranded RNA of 13.7−18.5 kb with a number of ORFs varying between 7 and 12 (G P Martelli et al. 2012). The members of this genus are divided into two subgroups accommodating, respectively, seven species with large (15,000 to over 18,000 nt) and complex (9 to 12 ORFs) genomes (subgroup I), and five species with smaller (13,000−14,000 nt) and simpler (7 ORFs) genomes (subgroup II) [9,16].

The majority of the *Ampelovirus* species have been identified from woody hosts (grapevine, *Prunus* sp., figs) and pineapple. Although their pathogenicity sometimes remains unclear, several members are reported to induce a diverse range of symptoms, while others are reported to have no association with symptoms. Natural vectors are *Pseudococcidae* mealybugs and soft scale insects, which transmit viruses in a semipersistent manner [9,12]. None of the known ampeloviruses are transmitted through seeds or mechanically. All ampeloviruses persist in plant parts used for vegetative propagation and are disseminated with them over long distances. The geographical distribution is therefore usually wide (G P Martelli et al. 2012).

In this study, we identify for the first time and characterize the genome organization of novel ampeloviruses infecting cassava plants originating from diverse geographical locations in central Africa and the southwestern Indian Ocean islands. We also provide a first analysis of the diversity of these agents and of the evolutionary forces acting on them.

IV.3. Material and methods

IV.3.1. Origin of the analyzed cassava samples and high-throughput sequencing

Field surveys were conducted from March to May 2016 in the South Kivu province (DR Congo). Cuttings were collected on cassava landraces plants showing foliar CBSD-like symptoms. These cuttings did not express the foliar symptoms anymore after growing in the greenhouse for six months. In addition, leaf samples from plants of different landraces were collected from the germplasm collections of the Conseil

Départemental (DARTM in Mayotte), of CIRAD (Reunion) and of FOFIFA (Madagascar) between 2015 and 2017 (Table IV-1).

Table IV-1. Identity and origin of cassava landraces used and indication on the high-throughput sequencing (HTS) protocol performed on these samples.

Cassava Landraces	Country of Origin	Collection Date DD/MM/YY	Symptoms	Sequencing Strategy and Target	Isolates Detected	Other Virus Detected
Menatana	Madagascar	18/10/2017	Asymptomatic in greenhouse	VANA	MG-Men-241, MG-Men-9	-
Miandrazaka	Madagascar	18/10/2017	Asymptomatic in greenhouse	VANA	MG-Mia-10, MG-Mia-362, MG-Mia-403, MG-Mian 2-2	-
Long Java	Reunion	25/02/2015	Asymptomatic in field	dsRNAs	RE-Ljv	-
Reunion	Mayotte	26/03/2015	Symptoms of CMD and CBSD	dsRNAs	MY-Ren-5, MY-Ren-8	EACMV+UCBSV
6 Mois Blanc	Mayotte	26/03/2015	Symptoms of CMD	dsRNAs	MY−6mb-4, MY−6mb-6	EACMV
Nambiyombiyo	DR Congo	13/01/2016	Asymptomatic in greenhouse	Total RNA	CG-Nmb	CBSV
Kahunde	DR Congo	13/01/2016	Asymptomatic in greenhouse	Total RNA	CG-Kah	-

[1]VANA: virion-associated nucleic acid; dsRNAs: double-stranded RNAs; EACMV: East African cassava mosaic virus; UCBSV: Ugandan cassava brown streak virus; CBSV: cassava brown streak virus; CMD: cassava mosaic disease; CBSD: cassava brown streak disease.

For the two samples from DR Congo, total RNA was extracted from cassava leaves using the RNeasy Plus Plant Mini Kit (QIAGEN®, Hilden, Germany) according to the manufacturer's instruction and quality tested using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). DNase treatment was applied using Amplification grade DNASE I (Life Technologies, California, USA) according to the manufacturer's instructions. Ribosomal RNAs were depleted using a RiboZero plant leaf kit for RNA-seq (Life Technologies Limited, Paisley, UK), and libraries were prepared following the manufacturer's instructions using a TrueSeq stranded total RNA kit (Illumina, New York, USA). The RNA libraries were sequenced on a Nextseq 500 sequencing machine at the University of Liege (Belgium), with a read length of 2 × 75 nt.

For the three samples from Mayotte and Reunion, double-stranded RNAs (dsRNAs) were extracted from leaf samples using the procedure of Marais et al. (Marais et al. 2018). Purified dsRNAs were converted to cDNA and amplified using a random whole-genome amplification procedure (Marais et al. 2018) and finally sequenced using Illumina Myseq technology (2 × 250 nt paired reads) on the Genotoul INRAE platform (Toulouse, France).

For the two samples from Madagascar, HTS was performed using a virion-associated nucleic acid (VANA)-based metagenomics approach as described by Palanga et al. (Essow Palanga et al. 2016).

IV.3.2. Bioinformatic analyses

Samples from Madagascar, Mayotte and Reunion: Following demultiplexing and quality trimming, reads were assembled into contig using CLC Genomic Workbench 8.5.1 and following versions. Contigs were annotated by BLASTx analysis against the GenBank protein database. Contigs were then manually assembled into scaffolds, which were consolidated and extended by several rounds of mapping of reads to yield the finalized scaffolds sequences. For two of the isolates (RE-Ljv and MY-Ren), reads corresponding to deletion events generating defective molecules were identified during this assembly process and scaffolds corresponding to these DI RNAs were therefore also reconstructed.

Samples from DR Congo: Sequence analyses were done using the Geneious 11.1.3 environment (www.geneious.com and embedded plugins). Data were submitted to a pre-processing step consisting of setting paired reads, and trimming adaptor sequences and low-quality reads (by BBDUK). Unique sequences were generated as clean reads by merging paired reads (by BBMerge) and removing duplicate reads (Dedupe). SPADES was used for de novo assembly. Reconstructed contigs were screened against the Refseq viral database retrieved from NCBI (October, 12th 2018) using BLASTn and BLASTx searches and annotated through tBLASTx. Viral contigs were further analyzed directly on the NCBI site using a BLASTn and BLASTx search with standard parameters.

Identification of functional gene domains for all samples was done by submitting predicted open reading frames (ORFs) directly on the NCBI's conserved domain search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (accessed from October to December 2018).

Multiple sequence alignments were built with ClustalW embedded in Mega X. Maximum likelihood phylogenetic trees were reconstructed in Mega X using the GTR+GI model for the alignment of nucleotide sequences and a Poisson model with uniform distribution for amino acid sequence alignments. Bootstrap analysis (100 replicates) was performed in order to evaluate the stability and significance of branches. Phylogenetic trees of amino acid sequences were reconstructed using only isolates for which the amino acid sequence of the analyzed protein was complete or nearly complete.

Putative recombination events were detected and evaluated using the RDP4 program (version 4.99) (D. P. Martin et al. 2015) and a ClustalW-built multiple alignment of the complete genome sequences.

IV.3.3. Confirmatory RT-PCR and sequencing

The presence of the detected viral contigs of MEaV-1 (isolates CG-Nmb and CG-Kah) in the corresponding plants was confirmed by RT-PCR using specific primer pairs (Table IV-2) designed according to the consensus sequence of the reconstructed

viral contigs. This also allowed to confirm the sequence of the contigs of isolates mentioned.

Table IV-2**.** RT-PCR primer pairs used for the confirmation of the viral sequences in the samples.

PRIMER	TARGET ORF	SEQUENCES (5′==> 3′)	Tm
12.495 F cp	CP ORF5	AATTTGGGAGGAGTGCGACC	60.3
12.715 R cp	CP ORF5	AGACCGACTTGTGCTACTCTTG	60.0
11.735 F hsp90	HSP90 ORF4	GTCCCGGCCTACATGCAATT	60.8
11.959 R hsp90	HSP90 ORF4	ACCGCCAGTCAACTCTCGTA	60.9
9.021 F hsp70	HSP70 ORF4	TATGGTTGTTGCTCGCGACT	60.0
9.292 R hsp70	HSP70 ORF4	CTGACAAACCAGCAGCAGTTG	60.3
7.244 F rdrp	RdRp ORF1B	GGACAACCTCCGAAACCGTAT	60.1
7.544 R rdrp	RdRp ORF1B	CTTTCGCTGCCATTGGTGTC	60.1

IV.3.4. Characterization of defective RNA(D-RNA molecules)

RNA isolation, sDNA synthesis and amplification of D-RNAs

Total RNA extracted from cassava accessions in which defective RNA molecules have been identified (RE-Ljv and MG-Mena-242) was purified and used for cDNA synthesis using the ProtoScript II First Strand cDNA Synthesis Kit (NEB) following the manufacturer’s instructions. Briefly, 0.5 µg of RNA was mixed with 2 µL of random hexamers primers (60 µM), heated at 65 °C for 5 min, and immediately cooled on ice for 5 min. The cooled solution was mixed with 10 µL of 2x ProtoScript II reaction mix and 2 µL of 10x ProtoScript II enzyme mix, mixed well and centrifuged for 5 s. This reaction mixture was incubated at 25 °C for 5 min and 42 °C for 1 h. The reaction was later terminated by incubation at 80 °C for 5 min.

Specific primer pairs designed on genome regions flanking the deletion in each of the four defective RNA molecules (Table IV-3) were used to amplify a region spanning the deletion of the D-RNAs. PCR amplification was performed in 20 µL total volume with 10 µL of Q5 High-Fidelity Master Mix (NEB), 2 µL of cDNA template, 0.5 µL of 10 µM of each primer and nuclease-free water for the remaining volume. The amplification program was set using 98 °C for 3 min of initial denaturation, followed by 35 cycles of 98 °C for 10 s, 57 °C for 10 s, 72 °C for 30 s and a final extension at 72 °C for 5 min. The amplified PCR products were analyzed on 1% agarose gel to validate the presence of amplicons of the expected size. The amplicons were gel-purified using the Monarch DNA Gel Extraction Kit (NEB) and cloned into pJET1.2 (Thermo Fisher Scientific).

Table IV-3**.** RT-PCR primer pairs used for the characterization of the defective RNA molecules.

Primers Name	Sequences (5′==>3′)	Expected Amplicon Size
MEaV-1-RE-Ljv-D-RNA1-F	ACATCTAAATGCTAACGAACGAAGAG	380 bp
MEaV-1-RE-Ljv-D-RNA1-R	CAACGCCAGAATCTTCGTACA	
MEaV-1-RE- Ljv-D-RNA2-F	AGGCTTTCGACAGTGAAGAAGTG	330 bp
MEaV-1-RE- Ljv-D-RNA2-R	CGTAGCCATACTGAAGGATAGCA	
MEaV-2-RE- Ljv-D-RNA3-F	TGGAAGCCGCTGGTAAACTACA	400 bp
MEaV-2-RE-Ljv-D-RNA3-R	CAAGCACGTTCAATATTAGGAATAGTAC	
MEaV-2-MG-Mena-D-RNA4-F	AGACATATGAAAGAGTTGCATTGGTG	500 bp
MEaV-2-MG-Mena-D-RNA4-R	ACCTACAAATAATTTCGCTCGTCTG	

Sequence analysis and alignments

The amplicons clones were Sanger-sequenced using pJET1.2 forward and reverse primers. Two clones were sequenced for each defective RNA molecule and no sequence variation was observed among them. The obtained sequences were aligned to reference parental full-length segments (RE-Ljv, MG-Mena-9) using the Clustal Omega multiple sequence alignment tool (online version, October 2020), allowing to confirm the existence of molecules containing the expected deletions.

IV.4. Results

IV.4.1. Identification of *Closteroviridae members in cassava*

Thirteen contigs or scaffolds of more than 10 kilobases of *Closteroviridae* isolates could be reconstructed using the sequencing reads from the seven analyzed landraces (Table IV-4). The lengths of the contigs, which represent large parts of the corresponding genomes, ranged from 10,417 to 13,770 nucleotides (nt). Some of the analyzed plants contained multiple contigs corresponding to different isolates (Table IV-4). Thorough screening of these reconstructed contigs against viral reference databases revealed nucleotide and amino acid sequence homologies with members of the family *Closteroviridae* and, in particular, with the genus *Ampelovirus*. In addition to these previously unknown viruses, the results further indicated the presence of CBSVs and/or CMGs in three cassava samples (Table IV-1). Cassava brown streak virus (CBSV, isolate KOR6, GU563327) was detected in one of the plants from the DR Congo, while Uganda cassava brown streak virus (UCBSV, isolate KM) (Scussel et al. 2019) and/or EACMV were detected in samples from Mayotte.

Table IV-4**.** Lengths, number of reads integrated, percent of total reads and average coverage for Closteroviridae contigs from the various analyzed cassava samples.

Cassava landraces	Isolate Name [1]	Accession Number (Genbank)	Virus Name	Contig Length [Nt]	Number of Reads Mapped	Percent Of Total Reads	Average Genome Coverage	Internal Undetermined Nucleotides	Internal Gaps
Menatana	MG-Men-241	MT773584	MEaV-1	13,528	38,475	12.1	506.8X	0	0
	MG-Men-9	MT773591	MEaV-2	13,752	10,945	4.10	153.6X	0	0
Miandrazaka	MG-Mia-10	MT773592	MEaV-2	13,770	26,987	4.8	368X	377	1
	MG-Mia-362	MT773594	MEaV-2	12,007	19,421	8.24	322X	0	0
	MG-Mia-403	MT773590	MEaV-1	10,518	26,917	11.4	480.9X	0	0
	MG-Mian 2-2	MT773596	MEaV-2	10,417	7,676	1.4	136.4X	843	1
Long java	RE-Ljv	MT773586	MEaV-1	13,640	21,294	18.5	312.5X	0	0
Reunion	MY-Ren-5	MT773589	MEaV-1	13,172	1,976	3.4	29.6X	72	1
	MY-Ren-8	MT773595	MEaV-2	13,762	377	0.7	5.4X	2,889	14
6 Mois blanc	MY-6mb-4	MT773585	MEaV-1	13,615	4,992	12.7	73.3X	288	4
	MY-6mb-6	MT773593	MEaV-2	13,764	2,631	6.7	39.2X	990	9
Nambiyombiyo	CG-Nmb	MT773587	MEaV-1	13,612	7,502	0.12	42X	0	0
Kahunde	CG-Kah	MT773588	MEaV-1	13,616	15,351	0.23	75X	11	1

[1] Isolates were named considering two letters of the code of the country where the landraces were collected—three letters of the landrace short acronym—and a number (when multiple contigs from the same landrace were obtained)

IV.4.2. Genome annotation

The sequences comparisons between the various isolates identified two clearly separated groups (Figure IV-1). The nearly complete genome of a representative isolate was selected for each group (respectively, CG-Nmb for the MEaV-1 and MG-Men-9 for MEaV-2) and was analyzed in detail to determine the genome organization. As shown in Figure IV-1, this analysis predicted in both cases seven open reading frames encoding for polypeptides with a mass ranging from 6 to 258 kDa (Table IV-5). The genomic organization, the size, and the number of ORFs identified revealed a similarity with the subgroup II ampeloviruses. The recognized members of this subgroup are grapevine leafroll-associated virus 4 (GLRaV-4), pineapple mealybug wilt-associated virus 1 and 3 (PMWaV-1 and -3), air potato Ampelovirus 1 (AiPoV1), and plum bark necrosis and stem pitting-associated virus (PBNSPaV) (Candresse and Fuchs 2020).

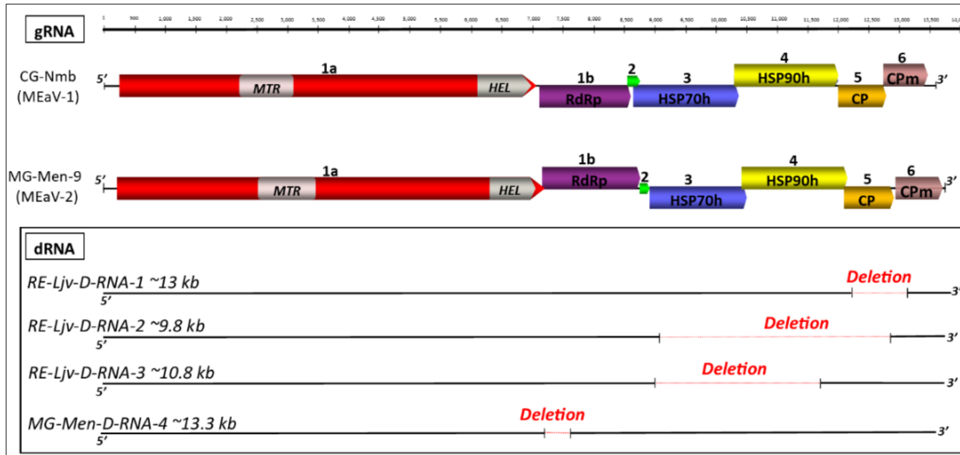


Figure IV-1. Schematic representation of the genomic organization of representative isolates CG-Nmb (MEaV-1) and MG-Men-9 (MEaV-2) (top) and structure of the defective variants (dRNA) identified (bottom). The RNA genome is drawn as a black line and the predicted open reading frames (ORFs) are represented by colored rectangles. Annotations and ORF numbers are given inside and above rectangles, respectively. Abbreviations: *MTR*—methyltransferase, *HEL*—helicase, *RdRp*—RNA-dependent RNA polymerase, *HSP70h*—heat shock protein 70 homolog, *HSP90h*— heat shock protein 90 homolog, *CP*—coat protein, *CPd*—minor coat protein.

Table IV-5**.** Genome length, position of ORFs and size of the encoded proteins for the representative isolates of the two species CG-Nmb (MEaV-1) and MG-Men-9 (MEaV-2).

Isolates	**Genome Length (Nt)**	**Length of Open Reading Frames (Orfs) (Nt) and Molecular Mass of Encoded Proteins (Kda)**						
		1a	*1b*	*P5-P7*	*HSP70h*	*HSP90h*	*CP*	ORF6
CG-Nmb (MEaV-1)	13,616 nt	6789 nt	1503 nt	204 nt	1707 nt	1686 nt	774 nt	717 nt
		254.6 kDa	56.6 kDa	7.6 kDa	62.2 kDa	63.4 kDa	27.7 kDa	27.1 kDa
MG-Men-9 (MEaV-2)	13,752 nt	6975 nt	1593 nt	147 nt	1608 nt	1698 nt	816 nt	639 nt
		258.8 kDa	60.4 kDa	5.3 kDa	58.4 kDa	64.3 kDa	29.3 kDa	24 kDa

The ORFs *1a* and *1b* encode the replication-associated proteins (Figure 1). As expected, a conserved domain search identified the following two replication-associated domains in the 1a protein: a methyltransferase motif (*MTR*; pfam01660) in the N-terminal part, and a viral helicase (*HEL*; superfamily 1, pfam01443) in the C-proximal part. The RNA methyltransferase domain was found in a wide range of ssRNA viruses and is known to be involved in mRNA capping.

The 1b ORF encodes the RNA-dependent RNA polymerase (*RdRp* 2 superfamily, pfam00978) expressed through a +1 ribosomal frameshift, as is seen in other members of the *Closteroviridae* family. Together with the *MTR* domain located at the 1a protein N-terminus, the *RdRp* forms the defining unique feature of the alpha-like ensemble of viruses (Koonin, Dolja, and Morris 1993).

Downstream of the polymerase, a small ORF2 encodes a hypothetical protein with a predicted molecular mass varying from 5 to 7 kDa depending on the isolate, and which is lacking known functional domains. The numbers and the sizes of the hypothetical proteins occurring downstream of the polymerase in the family *Closteroviridae* are highly variable. For example, Actinidia virus 1 (AcV-1) has three hypothetical proteins of predicted molecular masses of 13.6, 25.4 and 5.7 kDa (A. G. Blouin et al. 2017). However, in subgroup II ampeloviruses, only one hypothetical protein between 5 and 6 kDa is reported, sometimes overlapping with the 5′-proximal region of ORF3. The small size and the high proportion of the hydrophobic amino acids of the proteins encoded by the MG-Men-9 and CG-Nmb isolates (36/48 aa and 42/67 aa, respectively) could indicate that they are similar to the transmembrane proteins often present in the *Closteroviridae* members, where they have been proposed to function as cell-to-cell movement proteins (Peremyslov, Pan, and Dolja 2004).

ORF3 encodes a heat shock protein 70 homolog (*HSP70h*; cd10170) from the NBD–sugar–kinase–HSP70–actin superfamily. The functions postulated for *HSP70h* are as follows: the mediation of cell-to-cell movement through the plasmodesmata, involvement in the assembly of multisubunit complexes for genome replication and/or subgenomic RNAs synthesis, and the assembly of viral particles (G P Martelli et al. 2012). The size of this ORF is variable, being 62.2 kDa for MEaV-1 and 58.49 kDa for MEaV-2, as shown in the Supplementary Material (**Table S1**).

ORF4 codes for a heat shock protein 90 homolog (*HSP90h*, pfam03225), which is found to partially overlap with the 3′-proximal region of ORF3 and the 5′-proximal region of ORF5 as is the case for subgroup II ampeloviruses.

The coat protein (Closter coat superfamily, pfam01785) is encoded by ORF5.

It has been shown that downstream of the coat protein, some members of the family *Closteroviridae* have a variable number of ORFs encoding accessory proteins, some of which may have a functional conserved domain (Koloniuk et al. 2018), while others show no or only very limited identity to any other proteins. The translated amino acid sequences of the protein encoded by the ORF6 have shown similarity with the p24 protein of the pineapple mealybug wilt-associated virus-1 (25.1%) and the minor coat protein (*CPd*) of the grapevine leafroll-associated virus-5 (20.8%) from the TrEMBL database of UniProt. This suggests that, although they share <20% identity with their respective, the proteins coded by the ORF6 could be considered as minor coat proteins. This ORF6 overlaps partially with the 3′-proximal region of ORF5.

All of the isolates sequenced share the same genome organization, although the contigs obtained for the isolates MG-Mia-403 and MG-Mian-2-2 were incomplete, missing ORF5 and ORF6.

The reads belonging to four defective RNA (D-RNAs) molecules were identified (Figure 1). Three of them were detected in the long Java accession (RE-Ljv) originating from the Reunion island (two for MEaV-1 (RE-Ljv-D-RNA-1 and RE-

Ljv-D-RNA-2, accession numbers MW306827 and MW306828, respectively), one for MEaV-2 (RE-Ljv-D-RNA-3, accession number MW306829) and the remaining one in the Menatana accession originating from Madagascar (MG-Men-D-RNA-4 belonging to MEaV-2, accession number MW306830)). The deleted zones cover different genome regions and consist of most of the ORF5 and the ORF6 for the RE-Ljv-D-RNA-1, most of ORF3, ORF4, ORF5 and the start of ORF6 for the RE-Ljv-D-RNA-2, and most of the ORF3 and ORF4 for the RE-Ljv-D-RNA-3. In the case of the MG-Men-D-RNA-4, the deletion is at the beginning of the 1b ORF and does not alter the coding frame. The presence of molecules bearing these deletions in the original plant material was validated by the sequencing of amplicons spanning each of the individual deletions (see below). The proportion of deleted RNAs, as compared to the non-deleted genomic RNA, was estimated by comparing the read counts for the deletion borders and for the corresponding regions of the undeleted genome. This proportion was found to be variable, corresponding to about 4%, 37%, 27 % and 3.8% for RE-Ljv-D-RNA-1, RE-Ljv-D-RNA-2, RE-Ljv-D-RNA-3 and MG-Men-D-RNA-4, respectively.

IV.4.3. Phylogenetic analysis of cassava isolates with members from the family *Closteroviridae*

A multiple alignment of the *HSP70h* protein of the thirteen isolates, and of the corresponding protein of known members of the family *Closteroviridae*, was used to perform pairwise amino acid comparisons and to generate a phylogenetic tree that allowed to first address the taxonomic position of the cassava isolates in the family *Closteroviridae*.

The resulting tree (Figure IV-2) placed them with strong bootstraps support in a clade with members of the genus *Ampelovirus*, specifically in the subgroup II clade adjacent to GLRaV-4 (NC_016416) and PMWaV-1 and 3 (NC_010178 and DQ_399259), confirming the taxonomical relatedness of these cassava isolates to the subgroup II of the *Ampelovirus* genus.

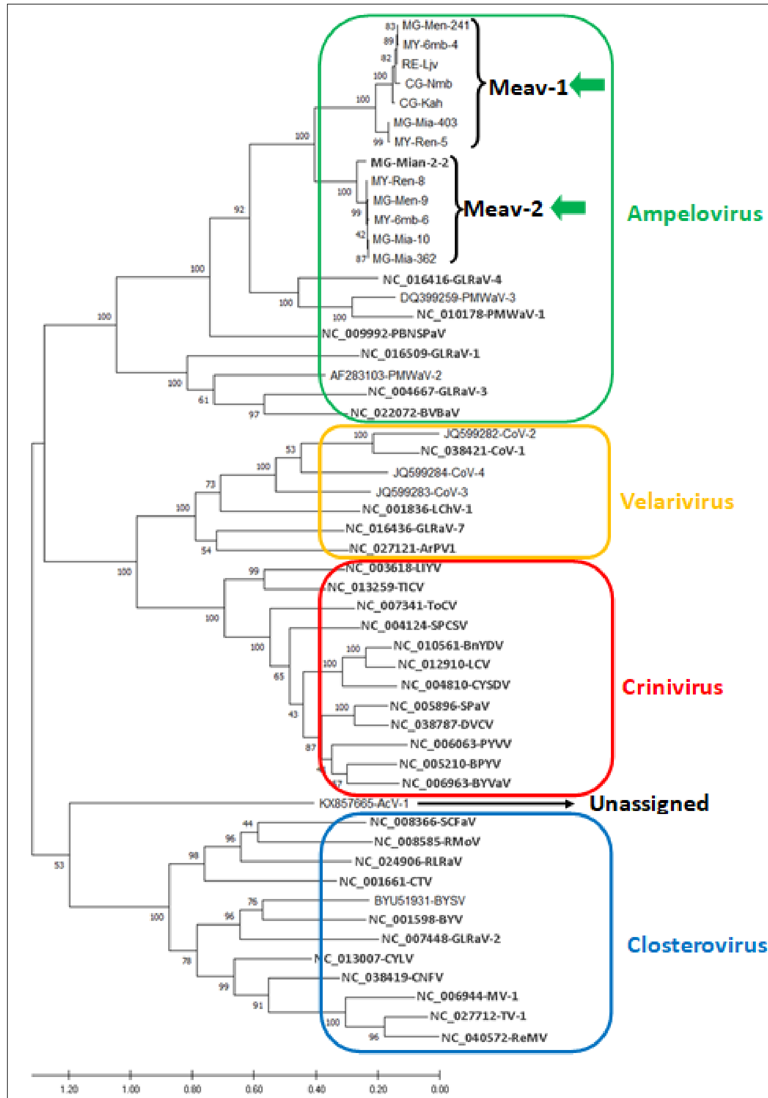


Figure IV-2**.** Phylogenetic analysis of the aligned amino acid sequences of the *HP70h* (ORF3) of the thirteen isolates from cassava and of selected members of the family *Closteroviridae*

(see the Supplementary Material Table S2 for detailed information on these viruses). Green arrows indicate isolate sequences obtained in this study. Maximum likelihood phylogenetic tree was reconstructed in Mega X using Poisson model with uniform distribution for amino acid sequence alignments. Bootstrap values are indicated at the main branch nodes. The bar represents the number of amino acid substitution per site.

The molecular identity at the amino acid level for the *RdRp*, *HSP70h* and *CP* of the cassava isolates with members of the family *Closteroviridae* is systematically lower than 55% (Table IV-6), well below the 75% molecular species demarcation criterion

for these taxonomically relevant genes in this family (G P Martelli et al. 2012). This shows that the isolates from cassava presented in this study are novel agents that do not belong to a known species in the family *Closteroviridae.*

Table IV-6. Amino acid sequence identity of the RdRp, HSP70h and CP of MEaV-2 (isolate MG-Men-9) with those of other selected Closteroviridae members.

Genus	Representative Members	Proteins		
		RdRp	***HSP70H***	***CP***
Ampelovirus	MEaV-1 (isolate CG-Nmb)	66.5%	69.9%	80.1%
Ampelovirus	Grapevine leafroll-associated virus 4 (NC_016416)	41.7%	51.6%	47.6%
Closterovirus	Citrus tristeza virus (NC_001661)	27.2%	25%	12.5%
Crinivirus	Potato yellow vein virus (NC_006062 and NC_006063)	25.7%	26.5%	17.2%
Velarivirus	Grapevine leafroll-associated virus 7 (NC_016436)	27.5%	27.7%	13.6%

IV.4.4. Diversity of the new isolates

To investigate the phylogenetic relationships between the various reconstructed *Closteroviridae* genomic sequences from cassava, multiple alignments of the near-complete genomic sequences and of the amino acid sequences of the three taxonomically relevant proteins were used to reconstruct phylogenetic trees (Figure IV-3a–d).

From these analyses, it is clear that the cassava isolates form two well-separated phylogenetic groups (MEaV-1 and MEaV-2) that show 41.4 ± 0.3% average nucleotide divergence based on the whole genome and, respectively, 32.3 ± 1.7% (*RdRp*), 30.8 ± 2.0% (*HSP70h*) and 19 ± 2.2% (*CP*) average amino acid divergence (Table IV-6 and see the Supplementary Material Tables S3–S6 for additional information).

The first group (MEaV-1) clusters two isolates identified in two cassava landraces from Madagascar (MG-Men-241 and MG-Mia-403), one isolate identified in an accession from Reunion (RE-Ljv*),* two isolates identified in accessions from Mayotte (MY-6mb-4 and MY-Ren-5) and two isolates from DR Congo (CG-Nmb and CG-Kah). The following two isolates from this group are divergent from the others and form a distinct subgroup: MY-Ren-5 and MG-Mia-403. Indeed, the MEaV-1 average pairwise divergence is found to be 17.1 ± 0.2% based on the whole genome, and, respectively, 7.1 ± 0.7% (*RdRp*), 6.0 ± 0.6% (*HSP70h*) and 7.4 ± 1.0% (*CP*) for the various proteins (Table IV-6). However, if excluding the divergent isolates MG-Mia-

403 and MY-Ren-5, these values fall down to 7.9 ± 0.2% for the whole genome and to 0.8 ± 0.4%(*RdRp*), 2.3 ± 0.4% (*HSP70h*) and 3.2 ± 0.6% (*CP*) for the various proteins, respectively. The average pairwise distance between these two divergent isolates, MG-Mia-403 and MY-Ren-5, is only 2.3% for the whole genome (Table IV-6).

The second group (MEaV-2) is composed of four isolates identified in two landraces from Madagascar (MG-Men-9, MG-Mia-2-2, MG-Mia-10 and MG-Mia-362) and two isolates identified in two landraces from Mayotte (MY-6mb-6 and MY-Ren-8). The isolate MG-Mian-2-2 forms a sub-cluster of its own. The MEaV-2 average pairwise divergence is 12.3 ± 0.1% for the whole genome, and, respectively, 7.3 ± 0.7% (*RdRp*) and 2.8 ± 0.4% (*HSP70h*) for the various genes (Table IV-7) (the average *CP* divergence was not computed because the *CP* gene is not covered by the MG-Mian-2-2 contig). If excluding the divergent MG-Mian-2-2 isolate, the average pairwise divergence for the whole genome falls to 3.8 ± 0.1%, and to 1.1 ± 0.3% (*RdRp*), 1.0 ± 0.3% (*HSP70h*) and 3.1 ± 0.8% (*CP*) for the various proteins, respectively (Table IV-7).

Comparisons across the trees based on the *RdRp, HSP70h* and *CP* proteins reveal variations in the branching topology that point to recombination events.

Considering the taxonomic differentiation between these two isolate groups (MEaV-1 and MEaV-2), largely supported by the phylogenetic trees, and the average divergence rates well over the 25% species demarcation criterion [14,15] for two of the three proteins (Table IV-6), it can be considered that the analyzed isolates form two distinct species, tentatively named “Manihot esculenta-associated ampelovirus 1 and 2” (MEaV-1 and MEaV-2), respectively.

MEaV-1 is detected from the four regions included in the study, while MEaV-2 is observed only from Madagascar and Mayotte (Figure IV-4).

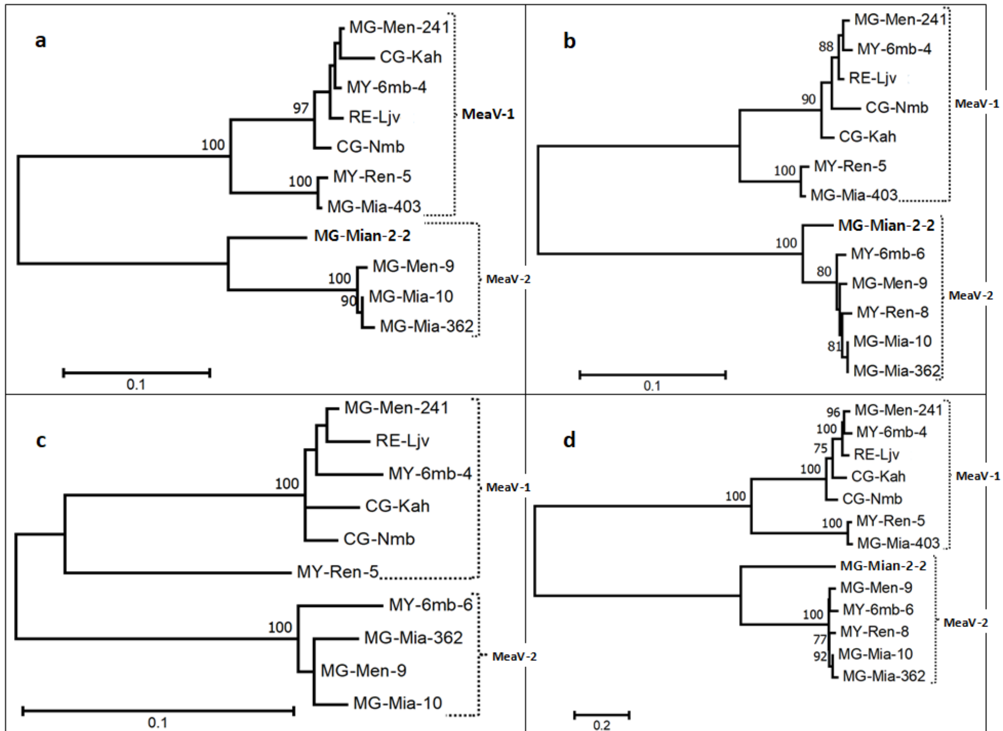


Figure IV-3. Phylogenetic trees reconstructed using the amino acid sequences of the following three taxonomically relevant proteins for the family *Closteroviridae*:
(**a**) *RdRp*; (**b**) HSP70h; (**c**) *CP*; and (**d**) the whole genome nucleotide sequences. Maximum likelihood phylogenetic trees were reconstructed in Mega X using the GTR+GI model for nucleotide sequences alignments and a Poisson model with uniform distribution for amino acid sequence alignments. Only bootstrap values of more than 70% are mentioned on nodes. The scales provide branch distance for the given number of substitutions per site.

Table IV-7. Intergroup and intragroup average pairwise divergence and standard deviation calculated for the three taxonomically relevant proteins (RdRP, HSP70h and CP) and for the nearly complete genomes.

		Within MEaV-2 with Mian 2-2	Within MEaV-2 without Mian2-2	Within MEaV-1 with Divergents	Within MEaV-1 without Divergents	Within MEaV-1 Divergents	Between MEaV-1 and MEaV-2
RdRp	aa divergences	7.3 +/− 0.7%	1.1 +/− 0.3%	7.1 +/− 0.7%	0.8 +/− 0.4%	0.8%	32.3 +/− 1.7%
HSP70h	aa divergences	2.8 +/− 0.4%	1.0 +/− 0.3%	6.0 +/− 0.6%	2.3 +/− 0.4%	0.7%	30.8 +/− 2.0%
CP	aa divergences	na	3.1 +/− 0.8%	7.4 +/− 1.0%	3.2 +/− 0.6%	na	19 +/− 2.2%
Genome	nt divergences	12.3 +/− 0.1%	3.8 +/− 0.1%	17.1 +/− 0.2%	7.9 +/− 0.2%	2.3%	41.4 +/− 0.3%

MEaV-1 divergents: MY-Ren-5 and MG-Mia-403. na: information not available as the contig/scaffold does not extend to the region concerned.

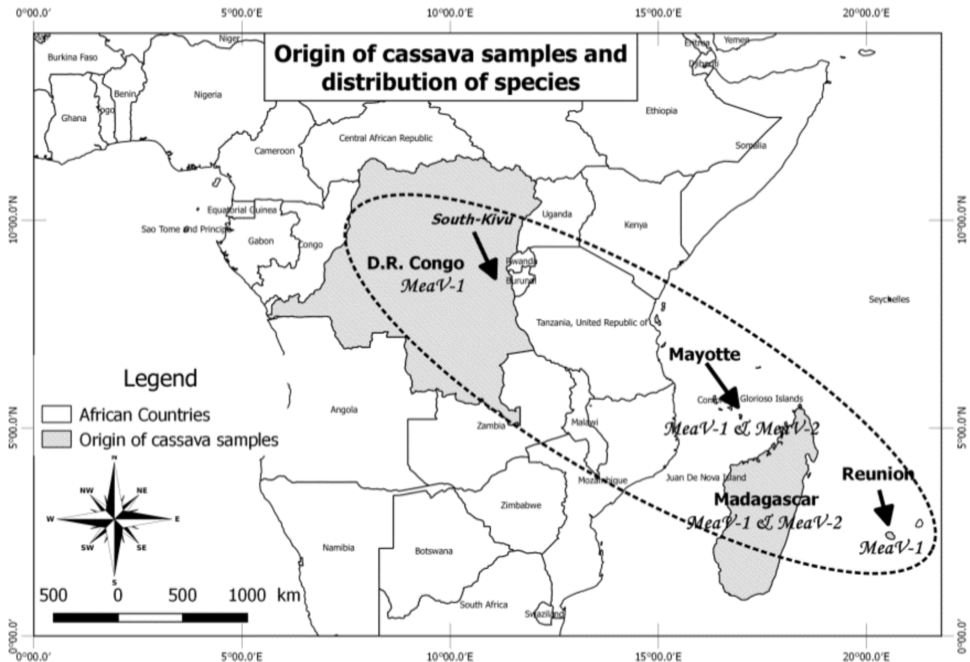


Figure IV-4**.** Geographic repartition of the two novel cassava ampelovirus species (MEaV-1 and MEaV-2) according to the country of origin of positive samples.

IV.4.5. Recombination analysis

Intraspecific and interspecific recombination events were investigated among the cassava isolates, and the recombination events supported that more than four of the nine algorithms integrated in RDP4 were considered as possible events (D. P. Martin et al. 2015).

Only one such recombination event was reliably detected using this criterion (Figure IV-5). This recombination event involves only the MEaV-2 isolates (two isolates MG-Mia-362 and MG-Mia-10 from Madagascar, and one isolate MY-6mb-6 from Mayotte) and is detected by seven out of the nine algorithms, with a strong probability (p-value = 2.75×10^{-91}). The predicted recombination breakpoints are identified within ORF1a (nucleotide position 906 of the isolate MG-Mia-362) and within ORF4 (nucleotide position 8997 of the isolate MG-Mia-362).

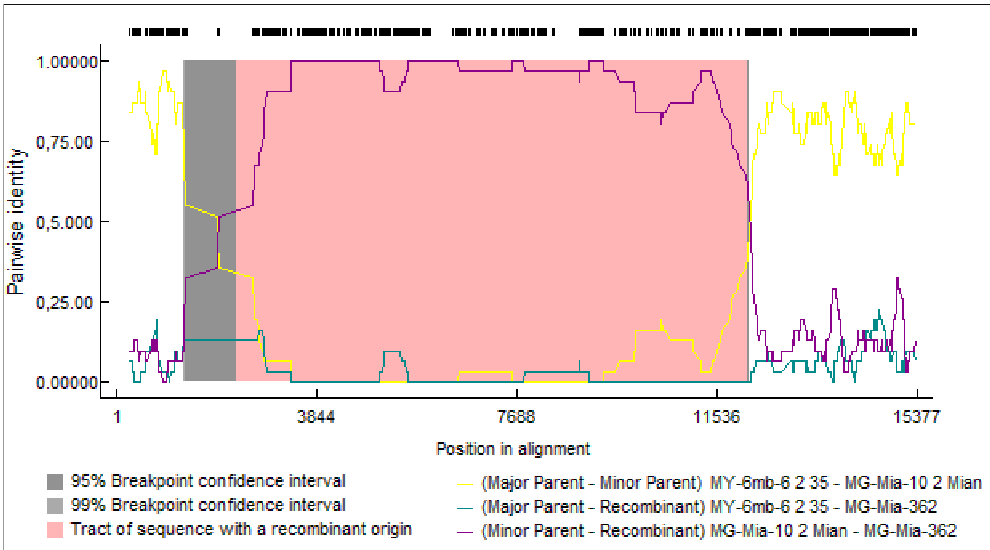


Figure IV-5**.** Recombination event detected among isolates from MeaV-2 by the RDP4 program

IV.4.6. Validation of defective RNAs

The presence of defective molecules was validated by the sequencing of RT-PCR amplicons spanning each of the identified deletions (Figure IV-6). The alignment of the four defective clones to their HTS-defective references, as shown in the Supplementary Material (Figures S7–S10), confirmed the presence of deletion zones at their predicted location. On the reference parental genomes (RE-Ljv and MG-Men-9), the deleted zones were found to be, respectively, 568 nt, 3761 nt, 648 nt and 444 nt for the D-RNA1, 2, 3 and 4.

These findings confirm that the D-RNAs reported here are not artifacts from high-throughput sequencing analysis, but continuous sequences that exist in the analyzed cassava samples.

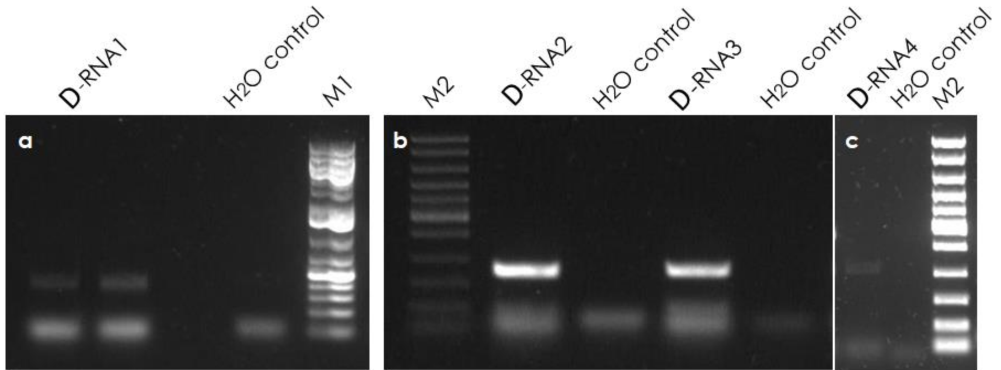


Figure IV-6**.** Gel red-stained 2% agarose gel showing RT-PCR products from: (**a**) D-RNA1 expected amplicon size 380 bp; (**b**) D-RNA2 and D-RNA3, expected amplicon size 380 bp and 400 bp, respectively; and (**c**) D-RNA4, expected amplicon size 500 bp. M1: GeneRuler 1 kb plus ladder (Thermo Scientific, SM1343). M2: fast DNA ladder 10 kb (NEB N3238S).

IV.5. Discussions

We have reported here the genome characterization from field-grown cassava plants from the DR Congo, Madagascar, Mayotte and Reunion of thirteen new *Closteroviridae* isolates, belonging to two potentially new ampelovirus species.

The analysis of the sequences of the corresponding contigs (ranging between 10,417 and 13,752 nucleotides in length) revealed seven open reading frames. In subgroup II, the length of the untranslated regions (UTRs) varies between 218 and 353 nt at the 5′ end, and between 125 and 132 nt at the 3′ end (Adiputra, Jarugula, and Naidu 2019; Dey et al. 2018; Yu et al. 2015; Melzer et al. 2008; Green et al. 2020; Sether et al. 2009). The UTR sizes reported here for the various genomes vary between 210 and 253 nt for the 5′end, and between 157 and 166 nt for the 3′ end. The comparison of these values, and the unambiguous assignation of MEaV-1 and MEaV-2 to the subgroup II of ampelovirus, show that the genomes reported here are unlikely to miss much more than a hundred nucleotides at their 5′ end and probably even less at the 3′ end, since some of the reported 3′ UTRs are already longer than the longest ones reported to date in the subgroup. This is confirmed when comparing the complete genome sizes; GLRaV-4 (isolate Man086, KJ810572) and PMWaV-1 (isolate HN, KJ872494) are the two subgroup II ampeloviruses with, respectively, the longest and the shortest genome (13,858 nt and 13,069 nt, respectively) [25,27]. The longest contigs for the representative isolates of MEaV-1 and MEaV-2 are, respectively, 13,616 nt (CG-Nmb) and 13,752 nt (MG-Men-9) long, suggesting that near-complete genomes, comprising the totality of the coding sequences and missing limited terminal non-coding nucleotide sequences, have been obtained, although the extremities of the genomes were not determined by RACE.

Based on the divergence of their *CPs*, which was determined to be below the species threshold (19 ± 2.2%), MEaV-1 and MEaV-2 could be considered as a single species. However, the official formulation of the taxonomic criterion for species demarcation in the family *Closteroviridae* and in ampelovirus genus is as follows: “Amino acid sequence of relevant gene products (polymerase, *CP* and *HSP70h*) differing by more than 25%” [14,15]. Therefore, there is an ambiguity on the fact of whether this criterion should be met for at least one of the three proteins or by all three simultaneously. Such a situation has previously been reported for Rehmannia virus 1 (ReV1) (Jung et al. 2018), for which the *RdRp* shows only 11% divergence with that of tobacco virus 1, while the *HSP70h* and the *CP* show 26% and 38% divergence, respectively. ReV1 was nevertheless accepted as a valid new *Closterovirus* species, suggesting that, as no recombination event was identified between them, MEaV-1 and MEaV-2 could similarly be considered as two distinct *Ampelovirus* species. This notion is further supported by the observation that all of the proteins diverge by more than 25%, between MEaV-1 and MEaV-2, except for the CP (and for a few comparisons involving one specific isolate, for the *HSP90h*).

Recombination is one of the mechanisms by which viruses evolve. Several studies have reported recombination events in *Closteroviridae* members [24,32,33]. In this

study, we report a recombination event involving MEaV-2 isolates. No evidence of recombination between MEaV-1 and MEaV-2 was detected.

In most cases, the analyzed cassava plants contained complex mixed infections involving either several isolates from a single MEaV species, or isolates belonging to the two species. Such complex infection patterns are probably due to the vegetative propagation practices used for cassava cultivation. In two plants with these complex mixed infections, defective molecules were identified and confirmed by RT-PCR, cloning and sequencing. Defective RNAs belong to the category of virus-associated molecules that are not required for normal virus propagation, but can sometimes affect the accumulation of the helper virus and symptoms expression (X. H. Li et al. 1989; Bar-Joseph and Mawassi 2013; White and Morris 1995). The presence of defective RNAs is reported in three genera of the *Closteroviridae* family [28,37–40], and is therefore not unexpected here. The considerable advances in virus characterization by HTS-based approaches are now revealing that, in addition to genomic and subgenomic RNAs, plants infected with viruses from the family *Closteroviridae* may contain several different subviral defective RNAs whose role is not known (Bar-Joseph and Mawassi 2013). Further studies are needed to assess their impact on epidemiology and pathogenicity.

The sampling campaign conducted in the DR Congo collected stems/cuttings from cassava plants showing typical symptoms of CBSD on their leaves and stems. These symptoms consisted of a yellow blotchy pattern on mature lower leaves and brown–black marks (‘streaks’) on green stem portions. The collected cuttings were planted in a glasshouse and, after approximately six months, two of the planted cuttings that have previously expressed foliar symptoms in a farmer’s field remained asymptomatic in the glasshouse and were used for the present study. The HTS results have confirmed the presence of CBSV in one of the samples, but not in the second one. Furthermore, the typical symptoms of CBSD and/or CMD that were observed on the samples from Mayotte are to be connected, in both cases, to the identification of the corresponding causal virus(es). No virus-like symptoms were observed on the samples from Reunion or from Madagascar, the Reunion sample being even selected in this study as a healthy control plant. Additional experiments of artificial inoculation and larger epidemiological studies are needed to clarify the symptomatology of MEaV-1 and MEaV-2, and to estimate the synergistic interaction between coinfecting viruses such as CBSVs and/or CMGs. The symptoms mentioned in this study consist only of foliar symptoms, while no observation was made on below-ground organs.

There is a need to develop a diagnostic test in order to be able to evaluate the distribution and prevalence of these new viral agents in other regions of the world, and to evaluate their impact on the yield. This evaluation of the phytosanitary risk should further be completed through wider surveys on symptomatic and asymptomatic plants in both agricultural and natural ecosystems to gain insight into the genetic variability of these new viruses, and their biological significance and impact (Massart et al. 2017).

Data Availability Statement: The datasets of genome sequences generated and analyzed during this study are available in Genbank repository under the following accession numbers: MT773584, MT773585, MT773586, MT773587, MT773588, MT773589, MT773590, MT773591, MT773592, MT773593, MT773594, MT773595, MT773596, MW306827, MW306828, MW306829, MW306830.

Acknowledgments: The Genotoul Platform (INRAE, Toulouse, France) is gratefully acknowledged for Illumina Myseq sequencing. The authors thank Dr. Marc Fuchs (Cornell University, USA) for a valuable exchange on *Closteroviridae* taxonomy and species demarcation criteria.

IV.6. Supplementary Materials:

```
Meav-1_RE-LJV-D-RNA1    ACATGCAATTCGATTTTGCGGTGGCCTTGGATCCTCTACATCTAAATGCTAACGAACGAA  11813
D-RNA1_Sanger           --------------------------------------------AATGCTAACGAACGAA  61
                                                                    ****************

Meav-1_RE-LJV-D-RNA1    GAGCTTTGACTTCATTGTTAGCTAGATTTAGAACCAGGAGAACACCTGTCAGAGGGCAAG  11873
D-RNA1_Sanger           GAGCTTTGACTTCATTGTTAGCTAGATTTAGAACCAGGAGAACACCTGTCAGAGGGCAAG  121
                        ************************************************************

Meav-1_RE-LJV-D-RNA1    CCTTAGGACAACGTGCTAGCAATCCACTAGATGATGTGGTTGAATCTATAGTTCCAGCGA  11933
D-RNA1_Sanger           CCTTAGGACAACGTGCTAGCAATCCACTAGATGATGTGGTTGAATCTATAGTTCCAGCGA  181
                        ************************************************************

Meav-1_RE-LJV-D-RNA1    CTTCTCGAGGAATGTTACGAGAGTTGACTGGCGGTGCCAGTACGCTTGGGTATAATAGAA  11993
D-RNA1_Sanger           CTTCTCGAGGAATGTTACGAGAGTTGACTGGCGGTGCCAGTACGCTTGGGTATAATAGAA  241
                        ************************************************************

Meav-1_RE-LJV-D-RNA1    GAAGACGCTACTAATTCAGCACCATGGCAACACCAACACCTGGAACACCAAACTCTACTC  12053
D-RNA1_Sanger           GAAGACGCTACTAATTCAGCACCATGGCAACACCAACACCTGGAACACCAAACTCTACTC  301
                        ************************************************************

Meav-1_RE-LJV-D-RNA1    CTCCTGATGCAAACAGTACTAATACCAACAGTCAAGTTGTGGCTGCTGTCGTACCTGGAG  12113
D-RNA1_Sanger           CTCCTGATGCAAACAGTACTAATACCAACAGTCAAGTTGTGGCTGCTGTCGTACCTGGAG  361
                        ************************************************************

Meav-1_RE-LJV-D-RNA1    GTCGTCAGACCGTTGDELETIONZONEAATAAACAAGAGTAGCACAAGTCGGTCCAACCC  12173
D-RNA1_Sanger           GTCGTCAGACCGTTG------------AATAAACAAGAGTAGCACAAGTCGGTCCAACCC  409
                        ***************            *********************************

Meav-1_RE-LJV-D-RNA1    TCACAATATGTACGAAGATTCTGGCGTTGCTCCAGAAATTTTCATGGAGGACGACGTTG  12229
D-RNA1_Sanger           TCACAATATGTACGAAGATTCTGGCGTTG------------------------------  465
                        *****************************
```

Figure S1: Multiple sequence alignment of D-RNA1 clone (referred to in the figure as D-RNA1_Sanger) with the defective reference from HTS (MEaV-1_RE-LJV-D-RNA-1). The term “DELETIONZONE” has been inserted into the HTS-defective reference to identify the zone where the deletion is located. This zone consists of 568 nucleotide sequences and is indicated by the rectangle.

```
Meav-1_RE-LJV-D-RNA2      GAAGGAGCGTGCAGTAAAACTCTAATGCGTGAAAAATGGGTGTCTTTCAATGACATTACT  8400
D-RNA2_Sanger             ------------------------------------------------------------  7

Meav-1_RE-LJV-D-RNA2      CAGGCTTTCGACAGTGAAGAAGTGTGTTGTCGTCTAGCCGCGTTGTGTGAGGAGAAATAC  8460
D-RNA2_Sanger             ------TTCGACAGTGAAGAAGTGTGTTGTCGTCTAGCCGCGTTGTGTGAGGAGAAATAC  67
                                ******************************************************

Meav-1_RE-LJV-D-RNA2      GGTCCTAGTCCGTGGTGGTACGCTGCGATAGCGACAATTCATTGTATCAGGGCCAATCCG  8520
D-RNA2_Sanger             GGTCCTAGTCCGTGGTGGTACGCTGCGATAGCGACAATTCATTGTATCAGGGCCAATCCG  127
                          ************************************************************

Meav-1_RE-LJV-D-RNA2      CAGCAATTTGAGAGGTGTTGGAGTGTAATATCTTCTCCTAGTATAAATGAAGCAAAGAAG  8580
D-RNA2_Sanger             CAGCAATTTGAGAGGTGTTGGAGTGTAATATCTTCTCCTAGTATAAATGAAGCAAAGAAG  187
                          ************************************************************

Meav-1_RE-LJV-D-RNA2      AGAAGCAAGGTCGTCTCCGAAGGAGGACCTGCGAGTTGTACAAATGACCAATTTATATTT  8640
D-RNA2_Sanger             AGAAGCAAGGTCGTCTCCGAAGGAGGACCTGCGAGTTGTACAAATGACCAATTTATATTT  247
                          ************************************************************

Meav-1_RE-LJV-D-RNA2      TGATTTTAAGTGGGCTATAACCGTTTDELETIONZONEACCTACAATTATGATGAAATAA  8700
D-RNA2_Sanger             TGATTTTAAGTGGGCTATAACCGTTT------------ACCTACAATTATGATGAAATAA  295
                          **************************            **********************

Meav-1_RE-LJV-D-RNA2      AAACGACCATTGATAATGCTATCCTTCAGTATGGCTACGAAAACCCCGTTCGCCAATTCG  8760
D-RNA2_Sanger             AAACGACCATTGATAATGCTATCCTTCAGTATGGCTAC----------------------  337
                          **************************************

Meav-1_RE-LJV-D-RNA2      GGAGGAGTGCGACCAGTTTCATAATTCAAGCATTAGCAAATGGACTCATGGAACCCAATG  8820
D-RNA2_Sanger             ------------------------------------------------------------  337
```

Figure S2: Multiple sequence alignment of D-RNA2 clone (referred to in the figure as D-RNA2_Sanger) with the defective reference from HTS (MEaV-1_RE-LJV-D-RNA-2).

The term "DELETIONZONE" has been inserted into the HTS-defective reference to identify the zone where the deletion is located. This zone consists of 3,761 nucleotide sequences and is indicated by the rectangle.

```
Meav-2_RE-LJV-D-RNA3      CATTCCACTCCGTACTTCTGTTCGAAGTTCTTGGTGGAAGCCGCTGGTAAACTACATTAC  8460
D_RNA3a-Sanger            --------------------------------------------------ACTACATTAC  10
D_RNA3b-Sanger            --------------------------------------------------ACTACATTAC  10
                                                                            **********

Meav-2_RE-LJV-D-RNA3      GTACCCGACCCCTTCAAAATTTTTATAAAATTTGGAGCGGAAGGTGCACCGACAAAGCCA  8520
D_RNA3a-Sanger            GTACCCGACCCCTTCAAAATTTTTATAAAATTTGGAGCGGAAGGTGCACCGACAAAGCCA  70
D_RNA3b-Sanger            GTACCCGACCCCTTCAAAATTTTTATAAAATTTGGAGCGGAAGGTGCACCGACAAAGCCA  70
                          ************************************************************

Meav-2_RE-LJV-D-RNA3      TTATTAAAAGAGAAGTGGGCTTCATTTTACGACATAACAACTGCTTATGACAGTGAGGAA  8580
D_RNA3a-Sanger            TTATTAAAAGAGAAGTGGGCTTCATTTTACGACATAACAACTGCTTATGACAGTGAGGAA  130
D_RNA3b-Sanger            TTATTAAAAGAGAAGTGGGCTTCATTTTACGACATAACAACTGCTTATGACAGTGAGGAA  130
                          ************************************************************

Meav-2_RE-LJV-D-RNA3      GTCTGCATCAAGTTGGCCCATCTATGCGCTGAGAAATACGGTGCATCGAATTGGTGGTAC  8640
D_RNA3a-Sanger            GTCTGCATCAAGTTGGCCCATCTATGCGCTGAGAAATACGGTGCATCGAATTGGTGGTAC  190
D_RNA3b-Sanger            GTCTGCATCAAGTTGGCCCATCTATGCGCTGAGAAATACGGTGCATCGAATTGGTGGTAC  190
                          ************************************************************

Meav-2_RE-LJV-D-RNA3      GCTGCTATAGCTACAATCCATTGTATTCGTGCAAACCCAACACAATTTGCACGATGTTGG  8700
D_RNA3a-Sanger            GCTGCTATAGCTACAATCCATTGTATTCGTGCAAACCCAACACAATTTGCACGATGTTGG  250
D_RNA3b-Sanger            GCTGCTATAGCTACAATCCATTGTATTCGTGCAAACCCAACACAATTTGCACGATGTTGG  250
                          ************************************************************

Meav-2_RE-LJV-D-RNA3      GAGACAGTTGATTGTGAGAGCATACGTTCTCGTAAGTCGDELETIONZONECAAACAAAC  8760
D_RNA3a-Sanger            GAGACAGTTGATTGTGAGAGCATACGTTCTCGTAAGTCG------------CAAACAAAC  298
D_RNA3b-Sanger            GAGACAGTTGATTGTGAGAGCATACGTTCTCGTAAGTCG------------CAAACAAAC  298
                          ***************************************            *********

Meav-2_RE-LJV-D-RNA3      TAGTGTGGCAGTGGGGTGACAAAACCGTTACTACTACTTATAATAGTATTTCCGCTATGT  8820
D_RNA3a-Sanger            TAGTGTGGCAGTGGGGTGACAAAACCGTTACTACTACTTATAATAGTATTTCCGCTATGT  358
D_RNA3b-Sanger            TAGTGTGGCAGTGGGGTGACAAAACCGTTACTACTACTTATAATAGTATTTCCGCTATGT  358
                          ************************************************************

Meav-2_RE-LJV-D-RNA3      TTGCGAAACATGGAGGTACTATTCCTAATATTGAACGTGCTTGGTGCAATCCACTAGCTT  8880
D_RNA3a-Sanger            TTGCGAAACATGGAGGTACTATTCCTAATATTGAACGTGCTTGATCTT------------  406
D_RNA3b-Sanger            TTGCGAAACATGGAGGTACTATTCCTAATATTGAACGTGCTTGATCTT------------  406
                          ******************************************* *
```

Figure S3: Multiple sequence alignment of D-RNA3 clones (two clones were sequenced and are referred to in the figure as D-RNA3a-Sanger and D-RNA3b-Sanger) with the defective reference from HTS (MEaV-2_RE-LJV-D-RNA3).

The term “DELETIONZONE” has been inserted into the reference HTS-defective to identify the zone where the deletion has been identified. This zone consists of 648 nucleotide sequences and is indicated by the rectangle.

```
Meav-2_MG-Mena-D-RNA-4    CGAGTGGCCACTGTGCATGAGTGTCAAGGGAAGACATATGAAAGAGTTGCATTGGTGCGA    5100
D-RNA4_Sanger             ------------------------------------------------------------    0

Meav-2_MG-Mena-D-RNA-4    ATCAAGCCCGCACAAGACGAAGTCTTCACGAGCGCTCCACATAGATTAGTGGCTCTCACT    5160
D-RNA4_Sanger             -------AGCCGCAAGACGAAGTCTTCACGAGCGCTCCACATAGATTAGTGGCTCTCACT    53
                                    * ************************************************

Meav-2_MG-Mena-D-RNA-4    CGACACACTACGAGTTTAGACTTCTATTGCATAAGAAATCGAATGGACCAGGGGATTGGT    5220
D-RNA4_Sanger             CGACACACTACGAGTTTAGACTTCTATTGCATAAGAAATCGAATGGACCAGGGGATTGGT    113
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    CGAGATGTCGAACTGATTGAGAAAATAACGGAACACGTCGCGAGGACTTTCTTAATTGAG    5280
D-RNA4_Sanger             CGAGATGTCGAACTGATTGAGAAAATAACGGAACACGTCGCGAGGACTTTCTTAATTGAG    173
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    CAGTGTTCCTAACAATTATTACGTDELETIONZONETGATTTAGCAGAATGGACCATGTC    5340
D-RNA4_Sanger             CAGTGTTCCTAACAATTATTACGT------------TGATTTAGCAGAATGGACCATGTC    221
                          ************************            ************************

Meav-2_MG-Mena-D-RNA-4    CAGAACGACATCGGGTAAGGCAGGTTTGTTACAAAAACTGCAAATACCTTCACTTCCCCA    5400
D-RNA4_Sanger             CAGAACGACATCGGGTAAGGCAGGTTTGTTACAAAAACTGCAAATACCTTCACTTCCCCA    281
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    TACTCAACTGGATGAATTCAAACTTATGGTGAAAGGGGATTTGAAACCAAAACTAGATGA    5460
D-RNA4_Sanger             TACTCAACTGGATGAATTCAAACTTATGGTGAAAGGGGATTTGAAACCAAAACTAGATGA    341
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    CACTTGCTTATCCGACATTCCGAGTGGTCAAAATATAGTTTATCACGAACGAGCAGTTTG    5520
D-RNA4_Sanger             CACTTGCTTATCCGACATTCCGAGTGGTCAAAATATAGTTTATCACGAACGAGCAGTTTG    401
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    TGCTTGTTTTTCTGTCTGTTTCATCCAAATGACAAAAAGATTGAAAGCCATATTAAACAG    5580
D-RNA4_Sanger             TGCTTGTTTTTCTGTCTGTTTCATCCAAATGACAAAAAGATTGAAAGCCATATTAAACAG    461
                          ************************************************************

Meav-2_MG-Mena-D-RNA-4    ACGAGCGAAATTATTTGTAGGTGTAGACCTCTCTGAATTTGCTCTCGATATTCAGCTCAC    5640
D-RNA4_Sanger             ACGAGCGAAATTATTTGTAGGTAA------------------------------------    485
                          **********************
```

Figure S4: Multiple sequence alignment of D-RNA4 clone (referred to in the figure as D-RNA4-Sanger) with the defective reference from HTS (MEaV-2_MG-Mena-D-RNA-4).

The term “DELETIONZONE” has been inserted into the reference HTS-defective to identify the zone where the deletion is located. This zone consists of 444 nucleotide sequences and is indicated by the rectangle.

Table S1: Length of ORF and molecular mass of 3 encoded proteins (RdRp, HSP70h and CP) of the 15 isolates of the cassava ampelovirus reported in this study.

Species	Isolates	Genome Length (nt)	Length of ORFs for predicted proteins (nt)/Molecular Mass of encoded proteins (kDa)		
			RdRp	HSP70h	CP
MeaV-1	MG-Men-241	13,528	1,521/57.4	1,707/62.1	774/27.8
	RE-LJV	13,640	1,521/57.3	1,707/62.7	774/27.7
	MG-Mia-403	**na**	1,653/60	1,671/58.6	**na**
	MY-6mb-4	13,615	1,500/56.5	1,707/60.6	774/26.9
	CG-Nmb	13,612	1,503/56.6	1,707/62.2	774/27.7
	CG-Kah	13,616	1,521/57.4	1,707/61.7	774/27.7
	MY-Reu-5	13,636	1,614[a]/57.8	1,707/62.3	768/27.6
MeaV-2	MG-Mian-2-2	**na**	16,53/60.3	16,71/58.2	**na**
	MG-Men-9	13,752	15,93/60.4	16,08/58.5	816/27.9
	MG-Mia-10	13,770	15,93/60.3	16,08/58.5	816/27.9
	MY-6mb-6	13,764	15,54/42.3	16,08/58.4	816/27.9
	MY-Reu-8	**na**	15,93/49.3	16,08/55.8	**na**
	MG-Mia-362	12,007	15,93/60.3	16,08/58.4	768/27.7

na: information not available as the contig/scaffold does not extend to the region concerned. [a]: The ORF is whether truncated or has gaps.

Table S2: Accession numbers and abbreviations for selected Closterovirus used in the phylogenetic reconstruction of the HSP70h nucleotide sequences.

Genus	Species Names	Assigned abbreviation	Sequence accession N° (GenBank)
Closterovirus	beet yellow stunt virus	(BYSV)	[BYU51931]
	beet yellow virus	(BYV)	[NC_001598]
	Citrus tristeza Virus	(CTV)	[NC_001661]
	mint virus 1	(MV-1)	[NC_006944]
	grapevine leafroll-associated virus 2	(GLRaV-2)	[NC_007448]
	strawberry chlorotic fleck associated virus	(SCFaV)	[NC_008366]
	raspberry mottle virus	(RMV)	[NC_008585]
	carrot yellow leaf virus	(CYLV)	[NC_013007]
	rose leaf rosette-associated virus	(RLRaV)	[NC_024906]
	tobacco virus 1	(ToV-1)	[NC_027712]
	carnation necrotic fleck virus	(CNFV)	[NC_038419]
	Rehmania virus 1	(ReV1)	[NC_040572]
Velarivirus	little cherry virus 1	(LChV-1)	[NC_001836]
	grapevine leafroll-associated virus 7	(GLRaV-7)	[NC_016436]
	Areca palm velarivirus 1	(APV1)	[NC_027121]
	Cordyline virus 1	(CoV-1)	[NC_038421]
	Cordyline Virus 2	(CoV-2)	[JQ599282]
	Cordyline virus 3	(Cov-3)	[JQ599283]
	Cordyline virus 4	(CoV-4)	[JQ599284]
Crinivirus	lettuce infectious yellows virus	(LIYV)	[NC_003618]
	sweet potato chlorotic stunt virus	(SPCSV)	[NC_004124]
	cucurbit yellow stunting disorder virus	(CYSDV)	[NC_004810]
	beet pseudo-yellows virus	(BPYV)	[NC_005210]
	strawberry pallidosis-associated virus	(SpaV)	[NC_005896]
	potato yellow vein virus	(PYVV)	[NC_006063]
	blackberry yellow vein-associated virus	(BYVBaV)	[NC_006963]
	tomato chlorosis virus	(ToCV)	[NC_007341]
	bean yellow disorder virus	(BnYDV)	[NC_010561]
	lettuce chlorosis virus	(LCV)	[NC_012910]
	tomato infectious chlorosis virus	(TICV)	[NC_013259]
	Diodia vein chlorosis virus	(DVCV)	[NC_038787]
	tetterwort vein chlorosis virus	(TEVCVO)	[NC_038789]
Ampelovirus	pineapple mealybug wilt associated virus-2	(PMWaV-2)	[AF283103]
	pineapple mealybug wilt associated virus-3	(PMWaV-3)	[DQ399259]
	grapevine leafroll-associated virus 3	(GLRaV-3)	[NC_004667]
	little cherry virus 2	(LChV-2)	[NC_005065]
	plum bark necrosis and stem pitting-associated virus	(PBNSPaV)	[NC_009992]
	pineapple mealybug wilt-associated virus 1	(PMWaV-1)	[NC_010178]
	grapevine leafroll-associated virus 4	(GVLRaV-4)	[NC_016416]
	grapevine leafroll associated virus 1	(GLRaV-1)	[NC_016509]
	blackberry vein banding associated virus	(BVBaV)	[NC_022072]
	pineapple mealybug wilt associated virus-2	(PMWaV-2)	[AF283103]
Unassigned	Actinidia virus 1	(AcV-1)	[KX857665]

Table S3: Estimates of evolutionary divergence between amino acid (aa) sequences for the HSP70h protein among cassava ampeloviruses isolates reported in this study.

MG-Men-241										
MY-6mb-4	0,9%									
RE-Ljv-x	1,5%	1,5%								
CG-Kah	2,5%	3,0%	3,0%							
CG-Nmb	3,0%	2,8%	2,8%	3,4%						
MY-Ren-5	12,3%	12,7%	12,7%	12,9%	12,1%					
MG-Mia-403	11,7%	12,3%	12,3%	12,5%	11,9%	0,8%				
MG-Men-9	32,4%	32,6%	32,4%	32,1%	32,6%	32,8%	32,6%			
MG-Mia-362	33,1%	33,3%	33,1%	32,8%	33,3%	33,6%	33,1%	1,7%		
MG-Mia-10	32,6%	32,8%	32,6%	32,2%	32,8%	33,2%	32,8%	1,0%	0,8%	
Mian-2-2	31,0%	31,2%	31,4%	30,3%	31,0%	31,5%	30,8%	13,4%	13,7%	13,2%

The number of amino acid differences per site from between sequences are shown. The analysis involved 11 amino acid sequences. All ambiguous positions were removed for

each sequence pair. There were a total of 535 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

Table S4. Estimates of evolutionary divergence between amino acid (aa) sequences for the HSP70h protein among cassava ampeloviruses isolates reported in this study.

MG-Men-241												
MY-6mb-4	1,0%											
RE-Ljv-x	0,9%	1,1%										
CG-Kah	2,6%	2,7%	2,2%									
CG-Nmb	3,0%	3,3%	2,6%	3,2%								
MY-Ren-5	9,9%	9,8%	10,1%	10,1%	11,0%							
MG-Mia-403	10,1%	9,8%	10,3%	10,1%	11,2%	0,7%						
MG-Men-9	31,0%	30,6%	31,0%	31,0%	30,1%	30,3%	30,1%					
MY-6mb-6	31,0%	30,6%	31,0%	31,0%	30,1%	30,7%	30,5%	1,1%				
MG-Mia-362	30,8%	30,6%	30,8%	30,8%	29,9%	30,5%	30,3%	0,9%	1,3%			
MG-Mia-10	30,8%	30,6%	30,8%	30,8%	29,9%	30,5%	30,3%	0,9%	1,3%	0,0%		
MY-Ren-8	31,6%	31,2%	31,6%	31,6%	30,6%	31,0%	30,8%	1,0%	1,2%	1,0%	1,0%	
Mian-2-2	31,0%	30,6%	31,2%	31,2%	30,8%	30,8%	30,7%	6,5%	6,4%	6,2%	6,2%	6,3%

The number of amino acid differences per site from between sequences are shown. The analysis involved 13 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 535 positions in the final dataset. Evolutionary analyses were conducted in MEGAX.

Table S5. Estimates of evolutionary divergence between amino acid (aa) sequences for the coat protein among cassava ampeloviruses isolates reported in this study.

MG-Men-241									
MY-6mb-4	3,2%								
RE-Ljv-x	1,9%	3,2%							
CG-Kah	2,3%	4,8%	4,3%						
CG-Nmb	2,3%	4,0%	3,1%	2,7%					
MY-Ren-5	15,7%	16,6%	15,7%	16,5%	15,3%				
MG-Men-9	18,7%	18,9%	18,3%	19,1%	17,9%	16,5%			
MY-6mb-6	20,3%	20,6%	19,9%	20,7%	19,5%	18,5%	3,6%		
MG-Mia-362	19,1%	19,8%	18,7%	20,3%	19,1%	17,3%	1,6%	4,8%	
MG-Mia-10	19,1%	19,3%	18,7%	19,5%	18,3%	16,9%	1,2%	4,8%	2,8%

The number of amino acid differences per site from between sequences are shown. The analysis involved 10 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 257 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

Table S6. Estimates of evolutionary divergence between nucleotide sequences for the whole genomes of cassava ampelovirus isolates reported in this study.

MG-Men-241												
MY-6mb-4	4,2%											
RE-Ljv-x	4,3%	4,8%										
CG-Kah	8,9%	9,3%	8,9%									
CG-Nmb	9,6%	9,9%	9,6%	9,2%								
MY-Ren-5	28,0%	28,2%	28,1%	28,0%	27,8%							
MG-Mia-403	27,5%	27,6%	27,6%	27,3%	27,5%	2,3%						
MG-Men-9	41,4%	41,2%	41,2%	41,2%	41,2%	41,8%	40,6%					
MG-Mia-10	40,6%	40,5%	40,3%	40,4%	40,6%	41,2%	40,8%	3,3%				
MY-6mb-6	41,6%	41,3%	41,4%	41,3%	41,4%	42,2%	41,1%	4,3%	4,1%			
MG-Mia-362	39,4%	39,5%	39,4%	39,4%	39,5%	40,2%	40,8%	4,4%	1,7%	5,2%		
MY-Ren-8	42,7%	42,4%	42,4%	42,4%	42,6%	43,0%	41,1%	4,0%	2,9%	4,5%	3,5%	
Mian-2-2	43,1%	43,2%	42,9%	42,9%	42,9%	43,2%	42,0%	29,5%	29,4%	29,8%	28,2%	29,1%

The number of base differences per site from between sequences are shown. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair. There were a total of 14121 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 X.

Chapter V

V. The Cassava-Congo cheravirus genome is characterized: a Maf-HAM1 motif is revealed and two domains (X1 and X2) are proposed upstream the NTB

Preliminaries

In this chapter, we report on the identification of an additional virus in cassava apart from the ampeloviruses. This virus, tentatively named the cassava-Congo cheravirus (Ca-Coc), is proposed as a new member from cheravirus genus in the family Secoviridae. The genome characteristics of this virus, already identified in Bas-Congo in the western part of the D.R. Congo, are provided. Its particular features that suggests an improved understanding of the genome organization in the family *Secoviridae* are also provided.

ARTICLE 3

(In preparation)

V. The Cassava-Congo cheravirus is characterized: a Maf/HAM1 motif is revealed and two domains (X1 and X2) are proposed upstream the NTB in the genus Cheravirus

Yves KWIBUKA[1, 2 *], Espoir BISIMWA[2], Vasudevan KUMAR[3, 4], Hélène SANFAÇON[5], Hervé VANDERSCHUREN[3, 4], Stephan WINTER[6] and Sébastien MASSART[1*]

[1] Plant Pathology Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[2] Faculté des sciences Agronomiques, Université Catholique de Bukavu, Democratic Republic of Congo.

[3] Plant Genetics Laboratory, TERRA-Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[4] Laboratory of Tropical Crop Improvement, Division of Crop Biotechnics, Biosystems Department, KU Leuven, Leuven, Belgium.

[5] Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, Canada.

[6] Plant Virus Department, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

V.1. Abstract

This study has combined complementary approaches to identify and characterize the complete genome sequences of an uncharacterized bipartite secovirid, the cassava Seco-Congo virus (**Ca-Coc**). Sequences were retrieved from field-grown cassava plants in D.R. Congo, and data mining of sequence read archive (SRA). Leaf extracts from a cassava sample with severe mosaic symptoms comprised of CMD collected in 2007 were mechanically inoculated in *Nicotiana benthamiana*, resulting in chlorotic mottle and deformation of young leaves followed by an oak leaf pattern symptoms.

Electron microscope observations revealed the presence of icosahedral virus-like particles, and western-blot analyses suggested the presence of three coat-proteins. Through cloning, 3' and 5' RACEs and subsequent Gibson assembly approach, two complete genome sequences of 7,405 and 3,460 nucleotides, corresponding to the RNA1 and RNA2 genomes, were obtained from a first isolate sampled in Bas Congo province. Four additional isolates were obtained from cassava plants in South-Kivu province in the Eastern D.R. Congo. Finally, an additional putative isolate was retrieved from transcriptomic data of a Tanzanian sample. RNA1 and RNA2 segments of all these isolates encoded a single polyprotein of various sizes, between 2369 and 2482 amino acids (aa) and between 1067 and 1069 aa, respectively. Results suggested an evolutionary relatedness with members from the cheravirus genus with distant identities.

Furthermore, they supported the existence of two species: one heterogeneous gathering isolates from D.R. Congo and another represented by the putative isolate from Tanzania. Further biological data would be necessary to address these findings. Experimental investigation using western blot validated the existence of three coat proteins of closer sizes as proposed by the results from the alignment-based prediction of cleavage peptides within the RNA2-encoded polyprotein.

Additional alignment-based predictions of cleavage sites within the RNA1-encoded polyproteins of genomes reconstructed in this study, together with genomes of other secovirids for which cleavage sites have been experimentally characterized previously, have also been conducted. Results suggested six putative cleavage sites, delineating seven putative domains. One unusual putative cleavage site could delineate two domains, X1 and X2, upstream the NTB for cheraviruses as previously demonstrated for viruses in the genus nepovirus and sadwavirus. A new ITPase protein homolog (HAM1), only reported for a few viruses infecting euphorbiaceous hosts (including cassava), was also detected.

V.2. Introduction

In several countries worldwide, the daily caloric demands of millions of people are fulfilled using cassava-based food. The importance of cassava for food security and income crop is high in sub-Saharan Africa due to its ability to produce reasonable yields in poor soils and with minimal inputs (Parmar, Sturm, and Hensel 2017). The threats due to climate change are now multiplying the magnitude of the daily challenges for the African agriculture commodities' production, food security and nutrition needs (DODO, 2020). In this particular context, cassava is mentioned as one of the crops on which African farmers could rely in the future compared to other leading crops in sub-Saharan Africa (SSA) (Jarvis et al. 2012). The main factors on which this reputation of cassava as a "food bank for the poor" is built include its capacity to successfully grow under environmentally challenging conditions (drought, fertility, temperatures etc.) (Bokanga 1999, Lebot, 2009, Rees et al. 2012, Howeler 2013).

One of the most important limitations for cassava production in Africa these decades has been associated with pests and diseases (Otun et al. 2022). In fact, like any other major crop in the world, cassava is susceptible to various pathogens, arthropods and mammal pests. Most have been characterized for many years, but the list of potentially threatening pathogens has been extended by discovering new viruses infecting cassava in Africa (Kwibuka et al. 2021). The main factors that have favoured disease spread are related to the vegetative propagation nature of the crop and vectors. In contrast, the plant's semi-perennial nature helped maintain pathogens and pest populations on farms. From the surprisingly long list of known pathogens and pests affecting cassava (Lebot 2009; Kwibuka et al. 2022), only some can significantly reduce yields and are therefore of economic importance: Xanthomonas axonopodis pv. manihoti, cassava brown streak viruses (CBSVs) and cassava mosaic virus (CMV) (Maruthi 2015; Parmar, Sturm, and Hensel 2017; Thresh et al. 1997; J.P. Legg et al. 2006). More specifically, in East and Central Africa, the two species associated with CBSD and the nine species associated with CMD are the most important pathogens causing economically essential diseases.

The Maf1/ham1-like proteins (HAM1) are ubiquitous proteins among all living organisms, involved in the pyrophosphatase (ITPase, also termed the "house-cleaning" activity) of the harmful non-canonical nucleotides: inosine triphosphate (ITP), xanthosine triphosphate (XTP) and their deoxy analogues (dITP/dXTP) (Xie et al. 2019; Parker, Hessler, and Cui 2022; Zamzami 2022). Until recently, only three viruses, all from the Potyviridae family, were known to encode a HAM1 protein homolog to ITPases: CBSV and UCBSV (genus ipomovirus) (Mbanzibwa, Tian, Mukasa, et al. 2009) and Euphorbia ringspot virus (EuRSV, genus potyvirus) (Knierim, Menzel, and Winter 2017). Very recently, another virus, the cassava torrado-like virus (CsTLV, genus torradovirus) from the Secoviridae family was shown to encode also a HAM1-like protein (Petrzik et al. 2016; Leiva et al. 2022b). Surprisingly, all these viruses were isolated from euphorbiaceous host plants (cassava and the ornamental plant Euphorbia milii), for which previous studies have hypothesized higher concentrations of ITP/XTP in the cytoplasm due to the relocation

of the plant HAM1 into the nucleus (Xie et al. 2019; Mbanzibwa, Tian, Mukasa, et al. 2009; A. M. James et al. 2021; Lopez et al. 2022). This high concentration of cytoplasmic pools of ITP act as selection pressure, particularly over cytoplasm-localized viruses, and promote the acquisition of the HAM1 gene to allow their survival into euphorbiaceous hosts (K. R. Tomlinson et al. 2019).

The Secoviridae is the only family of viruses from the order Picornavirales whose members infect plants (Sanfaçon et al. 2009; Walker et al. 2020). Viruses assigned to this family shares common properties with other members of the order: (a) a linear positive-strand RNA genome which can be monopartite or bipartite, with each RNA flanked at its 5' end by a small viral protein, the "VPg"; (b) the presence of a single ORF that encodes a single large polyprotein cleaved by a 3C-like cysteine protease; (c) a conserved "replication bloc" within the RNA1-encoded polyprotein that includes a type III helicase, the VPg, a cysteine protease and a type I RNA-dependent RNA-polymerase; the capsid, composed of coat protein (CP) subunits, that can be divided into a single large CP, or into two or three smaller CPs depending on the genus (Thompson et al. 2017; Sanfaçon et al. 2020; Zell et al. 2017). To date, 106 virus species belonging to one subfamily, nine genera and three subgenera have been assigned as members of the family Secoviridae. The genus Cheravirus in particular, is composed of viruses with a bipartite genome, encoding three capsid proteins on RNA-2. RNA-1 carries all the information required for replication and can replicate in individual cells without RNA-2, although no virus particles are produced. They are transmitted by nematodes and seeds (Karasev et al. 2017). Most of them are also transmissible experimentally by mechanical inoculation (Thompson et al. 2017).

Five criteria, not to be simultaneously met, are used to separate genera within the family Secoviridae, whereas seven are helpful in the demarcation of species: a CP amino acid sequence with less than 75% identity, a conserved Pro-pol region amino acid sequence with less than 80% identity, differences in antigenic reactions, distinct host range, distinct vector specificity, absence of cross-protection and absence of re-assortment between RNA1 and RNA2 for viruses with a bipartite genome (Thompson et al. 2017)

This publication presents preliminary biological and molecular characteristics of a bipartite virus from the Secoviridae family, detected in field-grown cassava plants from Central Africa (D.R. Congo). The diversity, phylogenetic and geographic distribution of these novel viruses were also investigated.

V.3. Material and methods

V.3.1. Virus isolation from cassava collected in Bas-Congo

In 2007 during field studies on the diversity of CMD conducted in D.R. Congo (Bas-Congo), samples (stem cuttings) were collected from cassava plants. Cuttings were rooted, and plants were established and maintained under greenhouse conditions in the Plant virus Collection at the Leibniz-Institute German collection of microorganisms and cell cultures (DSMZ). Leaves from virus-infected cuttings developed with typical mosaic symptoms; however, one cassava plant showed severe

mosaic symptoms, leaf blistering and deformation and conspicuous red necrotic spots indicating further viruses or other pathogens (Figure V-1).

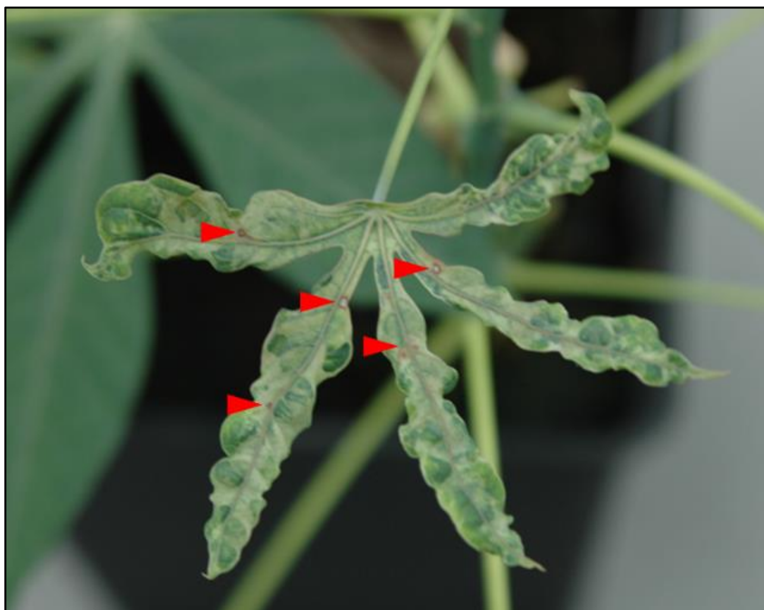


Figure V-1**.** Cassava plant collected in 2007 from stem cuttings and maintained in the DSMZ greenhouse at DSMZ. Red necrotic spots (red arrows) and severe mosaic symptoms indicate the presence of an additional virus or other pathogens.

Leaf extracts were prepared from leaves of infected plants and mechanically inoculated to a set of herbaceous virus indicator hosts from the Cucurbitaceae, Solanaceae and Chenopodiaceae, including Nicotiana occidentalis and N. benthamiana leave.

Adsorption preparates of N. benthamiana leaf extracts were prepared, negatively stained with uranyl acetate and subjected to electron microscopical (EM) examinations.

N. occidentalis was mechanically inoculated for virus propagation with homogenates of symptomatic leaves. The virus was purified ten days after infection using a modified purification protocol for tomato ringspot virus (Stace-Smith 1966). Purified virus particles were used to prepare an antiserum against the virus isolate that was subsequently modified for ELISA.

Soluble leaf proteins from virus-infected N. benthamiana plants and purified particle preparations were subjected to a discontinuous sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE, gradient 8-20%) to determine the size of the purified viral coat protein. Following electrophoresis and western-blot transfer to nitrocellulose membranes, the viral proteins were verified by incubating membranes with antiserum prepared against virion preparations.

cDNA from purified virus preparations was synthesized using random-primed cDNA synthesis followed by 2nd strand synthesis and cloning into a pBluescript SK- cloning vector. Positive clones from blue-white screening were sequenced with vector-specific primers, and contigs were assembled using Vector NTI. PCRs spanning missing genome sequences and 5' and 3' RACEs were conducted to reconstruct the complete viral RNA 1 and RNA2 genome. Genome sequences obtained for this virus isolate, PV-0896, were included in the molecular analysis and diversity study of cassava virus samples collected in South Kivu Province analyzed by HTS.

V.3.2. Study in South-Kivu province

a. Origin of the analysed samples, total RNA extraction and High-Throughput sequencing

From April to October 2019, during investigations on the diversity of viruses infecting in South-Kivu/Eastern D.R. Congo, samples were collected in the Uvira territory as previously indicated (Kwibuka et al. 2022). Surveyed fields in each village were randomly selected following transects oriented within main cassava-growing villages with a minimal distance of 2 km between fields. Two hundred forty-two fields were sampled in total. In each field, a portion of the third fully expanded young leaf on a shoot was collected on ten cassava plants selected following diagonals. Samples from individual plants, symptomatic or not, were conserved in tubes dried with silica gel. All leaf samples from the same field were pooled together to make a unique field-core sample. Stem cuttings corresponding to each sampled plant were also collected and grown in an experimental field in Bukavu, while some of them were shipped to Belgium and grown in a glasshouse at 27°C. Total RNA was extracted from these pooled leaf samples using a modified CTAB protocol as previously described (Kwibuka et al. 2021). Extracted RNA was further tested for CBSV and UCBSV infection by RT-PCR as previously described (Kwibuka et al. 2022).

All samples that tested negative to this primer-based screening (71 field-core samples representing 710 individual plants collected into 71 farmer fields) were regrouped into 12 equimolar HTS pools (P1 to P12) that were subjected to high throughput sequencing (HTS). DNase treatment was applied using Amplification grade DNase I according to the manufacturer's instructions (Life Technologies, California, USA). In addition, the RiboZero plant leaf kit for RNA-seq (Life Technologies Limited, Paisley, UK) was used for depleting ribosomal RNAs, and a TrueSeq stranded total RNA kit (Illumina, NY, USA) was used for library preparation following the manufacturer's instruction.

The RNA libraries were subjected to HTS on a Nextseq 500 sequencing machine at the University of Liege GIGA facilities (Liège, Belgium), with a read length of 2 X 150 nt.

Cassava plants regenerated through stem cuttings, kept in a glasshouse (27 °C), and in which the herein described virus isolates were detected in a single infection had characteristic leaf symptoms consisting of leaf yellowing spreading from secondary

veins. These symptoms were absent on newly expanded leaves but later developed as the plant grew. No leaf deformation could be observed.

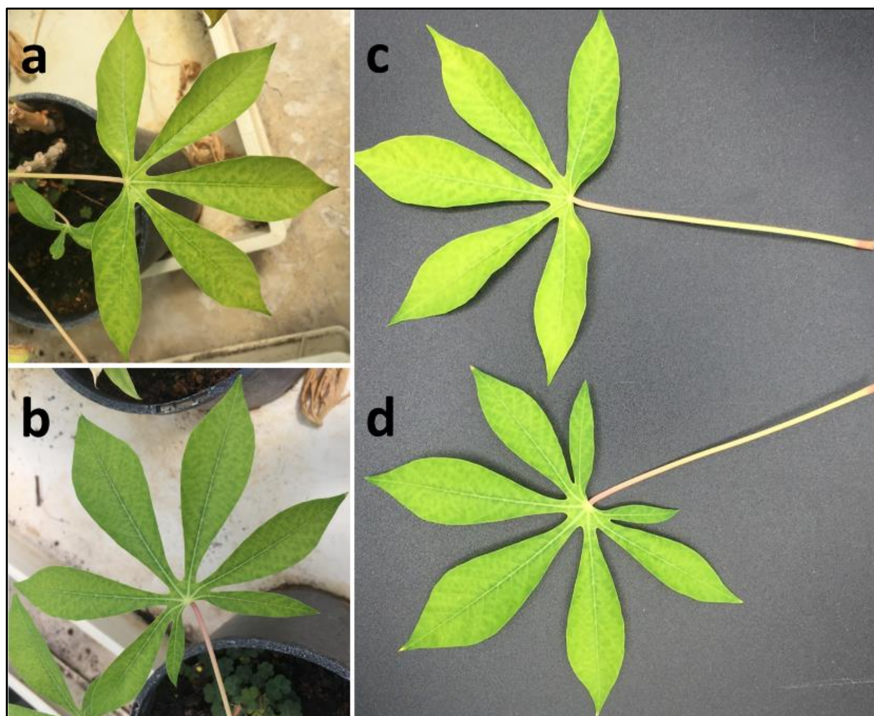


Figure V-2. Cassava landraces (a: Nabwilalanga, b: Nabwigoma, c and d: Rava) collected in 2019 as stem cuttings and grown in a greenhouse at Gembloux Agro Bio-tech showing mild to low leaf yellowing symptoms spreading from secondary veins.

b. Bioinformatic analyses

Sequence reconstruction and analysis were performed using Geneious 2021.1.1 software (www.geneious.com) and embedded plugins. Following demultiplexing and quality trimming (BBDUCK, BBMerge and Dedupe plugins), reads were de novo assembled into contigs using RNA-SPADES (embedded in Geneious). Reconstructed contigs were screened against the RefSeq non-redundant viral database (release 202, 8 September 2020, retrieved from NCBI) using BLASTn and tBLASTx searches (Sayers et al. 2022). Early screening of reconstructed contigs from 4 HTS pools (P3, P4, P5, P12) against the RefSeq Viral database revealed nucleotide and amino acid sequence homologies with members of the family Secoviridae. Twenty-five rounds of iterative read mapping extended the size of corresponding contigs to yield the final sequences.

Multiple sequence alignments were built with ClustalW embedded in Mega X (S. Kumar et al. 2018). Maximum likelihood phylogenetic trees were reconstructed in Mega X after predicting the suitable model on amino acid or nucleotide sequence

alignments. The Le Gascuel (LG) matrix-based model of amino acid substitution using a discrete Gamma distribution with invariant sites (+G+I) was used on the alignment of the conserved amino acid "Pro-Pol" region (in the RNA1). In contrast, the same substitution matrix (LG), but with frequencies and a discrete Gamma distribution (+G+F), was used for the alignment of the coat protein amino acid sequences (in the RNA2). These trees were built using translated amino acid sequences from nucleotide sequences of the RNA1 and RNA2-encoded polyproteins (respectively P1 and P2) for genomes reconstructed in this study and all available genomes of Secoviridae members retrieved from GenBank. Supplementary Table V-1 shows the names of these viruses and accession numbers of the corresponding RNA1 and RNA2 genomes. Bootstrap analysis (1000 replicates) was performed to evaluate the stability and significance of branches.

c. Confirmatory RT-PCR, rapid amplification of cDNA Ends (RACE) and Sanger sequencing

Primers were designed to confirm the presence of the virus sequences in pools of several fields and samples of an individual field. Primers were also designed to reconstruct the 5' and the 3' ends of the assembled genomes. Additional primers were also designed to confirm the presence of the identified HAM1 motif in the RNA1 polyprotein by amplifying the region spanning the motif. All the primers are shown in supplementary material 1.

For confirmation of virus presence in pools of individual fields, RNA was extracted using the RNEasy Plus® Kit (QIAGEN®, Hilden, Germany) following the manufacturer's recommendations. For the RT-PCR confirmation, the first strand cDNA synthesis was performed using Tetro™ Reverse Transcriptase (meridian Bioscience) and random hexamers, following the manufacturer's instructions. The PCR was conducted using MangoTaq™ DNA polymerase (meridian Bioscience). Primer pairs were designed to amplify both extremities (5´ and 3´) of RNA1 and RNA2 (Table V-1). A 10 µl PCR mixture containing 5.8 µl nuclease-free water, 2 µl PCR buffer (5X), 0.40 µl MgCl2 (50 mM), 0.20 µl dNTPs (10 mM), 0.20 µl of each primer (10 mM), 0.4 µl Mango Taq DNA polymerase (Meridian Bioscience®) and 1.0 µl of cDNA. PCR products were analyzed by electrophoresis in TAE buffer (1X) on a 1% agarose gel stained with Gel red® (Biotium), visualized under UV light and photographed using a gel documentation system (E-Box CX5 Edge, Vilber/Fisher Biotech).

Reconstruction of the 5' and 3' ends of genomes was conducted using the SMARTer® RACE 5'/3' Kit (Takara Bio USA, Inc) following the manufacturer's instructions. Obtained RACE products were sanger-sequenced using virus-specific primers (supplementary material 1).

V.3.3. Identification of the RNA1 and RNA2-encoded cleavage sites

Nucleotide sequences corresponding to the RNA-1 and RNA-2 encoded polyproteins of genomes reconstructed in this study, together with those of representative members of the Secoviridae family for which cleavage sites have been

previously characterized (Mann, Walker, and Sanfaçon 2017; A. Wang and Sanfacon 2000; Wetzel et al. 2008) (Supplementary material 2), were translated into amino acid sequences and aligned using Clustal Omega (Madeira et al. 2019). The alignment files obtained (P1 and P2 respectively for polyproteins of the RNA1 and RNA2) were used to predict putative cleavage sites and therefore deduce functional domains based on sequence homologies with related viruses. In addition, to determine if predicted cleavage sites of the RNA1 and RNA-encoded polyproteins were conserved among other viruses from the family Secoviridae, alignments were extended to other previously characterized viruses in this family (Supplementary material 2).

V.3.4. Data mining

To screen for a potential presence of the newly identified Secoviridae members in publicly available RNASeq data, the nucleotide sequences corresponding to the polymerase domain encoded into the RNA1 segment of the secovirids detected in field-grow cassava plants were extracted and screened against the SRA (Leinonen, Sugawara, and Shumway 2011) using an RdRp-based search. This search uses the core of the polymerase palm subdomain (Babaian and Edgar 2021). This search was implemented by the Serratus cloud computing infrastructure (Edgar et al. 2022). Raw reads data identified by the Serratus search as containing palm-id motifs closely related to the RdRp of the novel Secoviridae members were retrieved from the SRA using the “Faster Download and Extract Reads in FASTQ format from NCBI SRA” tool integrated into GALAXY Europe platform (Afgan et al. 2018). These reads were processed following the bioinformatic pipeline and tools used as stated above and the contigs showing homologies with the Secoviridae were generated and included in the subsequent analyzes.

V.4. Results

V.4.1. Virus isolation, electron-microscopical examination, purification and genome reconstruction of a virus isolate from Bas-Congo.

N. benthamiana plants infected by mechanical inoculation started showing chlorotic mottle and deformation of youngest leaves, resulting in an oak leaf pattern on old and young leaves recovered from symptoms (Figure V-2a). Additionally, electron microscopical examinations revealed virus-like particles (VLps) of icosahedral structure (Figure V-3b, red arrows), some of which could be empty (Figure V-3b, black arrows).

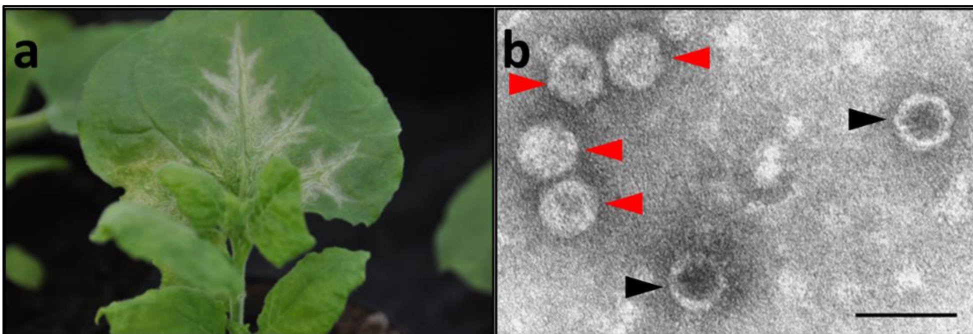


Figure V-3. (a) Virus symptoms in *N. benthamiana* starting with chlorotic mottle and leaf deformation of youngest leaves, resulting in an oak leaf pattern on old leaves and young leaves recovered from symptoms. (b) Electron microscopically examined Adsorption preparations of leaf homogenates showing VLp of icosahedral structure (red arrows), sometimes empty (black arrows).

Results from the SDS-PAGE separation revealed the presence of three virus particles, pointing to the presence of three putative coat protein units of sizes varying between 20 and 30 kDa (Figure V-3).

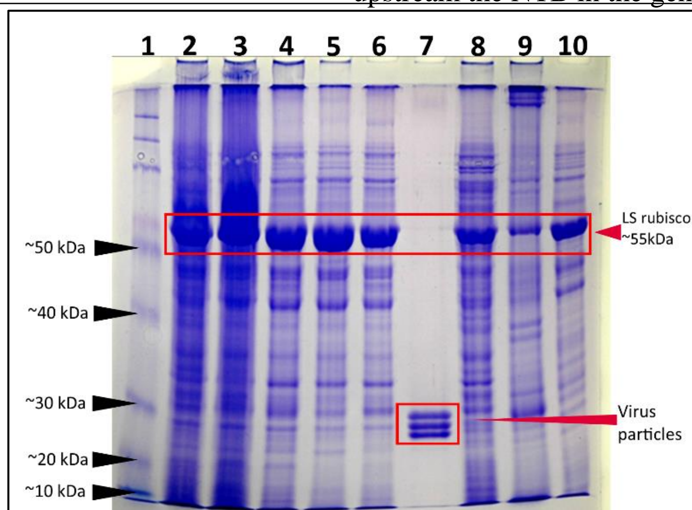


Figure V-4. Electrophoretic separation of soluble leaf proteins from healthy *N. benthamiana* (lanes 2 to 6 and lanes 8 to 10) and from purified virions (lane 7) in a discontinuous SDS-PA gel (gradient 8-20%).

Results from the complete genome assembly revealed sizes of 7,405 and 3,460 nucleotide sequences, respectively, for RNA1 and RNA2 (Table V-1). Unless the lack of phylogenetic investigations, the presence of 3 coat proteins prompted the assignation of this virus to the genus cheravirus and represented, until then, the only species from this genus known to infect the cassava plant. These sequences were deposited in GenBank and are available under the accession number PV-0896. In addition, double-Sandwich ELISA antibodies for detecting this virus have been produced and are available at DSMZ under the reference number RT-0896. They have been used by (Badamasi et al. 2020) to investigate the distribution of viruses infecting cassava in Nigeria unless the virus has not been confirmed.

Additionally, to this previously unknown virus, the presence of the East African cassava mosaic virus (EACMV) has been validated as co-infected. In this study, the denomination CGO-BC (indicative of the D.R. Congo country and the Bas-Congo Province) has been used for this isolate and served as a reference when included in further analyses together with other isolates retrieved from the HTS as described below.

V.4.2. HTS analysis of cassava samples from South-Kivu Province and data mining of publicly available RNASeq data

The screening of the contigs assembled from field-grown cassava plants against the NCBI BLAST nt/nr databases (Sayers et al. 2022) identified ten contigs from four of the twelve sequenced HTS pools (Supplementary Material 3). They showed high nucleotide and amino acid sequence homologies with members of the genus cheravirus within the family Secoviridae.

To uncover other Secoviridae genome sequences from the publicly available RNASeq data, the polymerase sequences of the CGO-BC isolate were used for the RdRp search using the Serratus tool. This data screening identified 80% identity with an RdRp palm_id reference u138608 from the RNASeq ERR996012 generated in 2015 by the team of Ndunguru while assembling the CBSV and UCBSV genomes from Tanzania (Ndunguru et al. 2015). The corresponding RNASeq dataset was downloaded, and, following demultiplexing, de novo assembly and BLAST searches as specified above, three contigs showing homologies with secovirids could be retrieved.

The sequence comparison revealed that in total, from all the 13 reconstructed contigs, two groups could be defined according to their sizes: one group of six long contigs with sizes varying between 7,405 nt and 7,795 nt, showing similarities with the RNA1 of the stocky prune virus (YP_009665965.1) and another group of seven shorter contigs with sizes varying between 3,460 and 3,496 nt showing homologies with RNA2 of the same virus (YP_009665964.1) (Table V-1).

These isolates were named as follows:

- For the isolates detected in South-Kivu province: the CGO-KV initials (indicative of the D.R. Congo country and the Kivu region) plus a number (indicative of the pooled sample in which the isolate was detected)
- For the isolates detected in RNASeq data from Tanzania: the TZ initials (indicative of the Tanzania country) plus a number (indicative of the reference number corresponding to the contig assembled from the de novo process).

Of the thirteen contigs assembled, eight belonged to four HTS pools (P3, P4, P5 and P12) from 23 fields located in South-Kivu province (Uvira territory) in the Eastern part of D.R. Congo, two other contigs were detected within samples originating from fields located in the Bas-Congo province (Mbanza-Ngungu) in the Western part of the country (Figure V-1 and Table V-1), and the remaining three contigs were reconstructed from the publicly available RNASeq data generated from cassava samples collected in Tanzania in 2015 (SRA reference number ERR996012) (Ndunguru et al. 2015). Analysis of each field-core sample from each positive HTS pool from South Kivu (ten samples of individual plants in each of the 23 fields) revealed that, in total, 130 plants out of the 230 were positive.

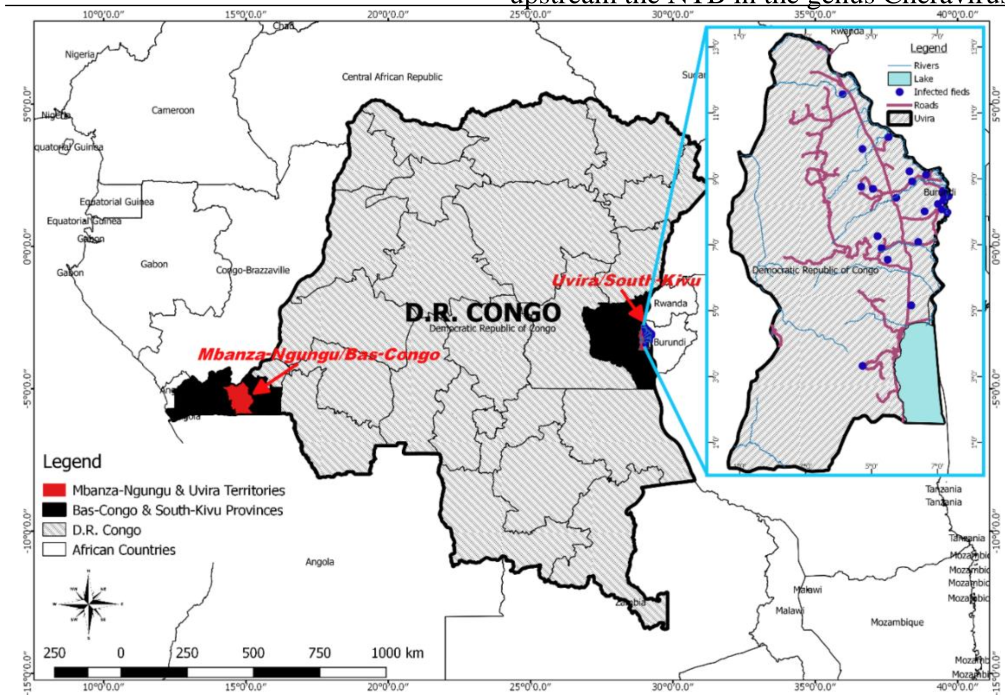


Figure V-5. Geographical map of the study area showing the localization of the provinces where samples originated and the spatial repartition of positive fields in the four pools from South-Kivu Province.

Additionally to these viruses, results further indicated the presence of the following viruses:

- From the field-grown plants, results indicated the presence of begomovirus-associated satellites II and III in one pool (P4) (accession number NC_006957.1, NC_006956.1) and UCBSV in two pools (P4 and P5) (accession numbers MW961202.1, MZ486425.1, MW961214.1, MW961221.1).
- From the publicly available RNASeq data, results allowed to additionally identify both CBSV (KR108830.1, 100% id) and UCBSV (KR108837.1, 100% id) in the same read run where the secovirid contigs were identified (ERR996012). In contrast, in a distinct run of the same sequencing project (ERR996014), analyses detected the presence of the UCBSV (KR108839.1, 100% id) and the Manihot esculenta-associated ampelovirus-1 (MeAV-1, MT7735841 96% id).

V.4.3. Genome annotation and phylogenetic placement

a. Genome annotation

The genomes of all reconstructed contigs from the HTS and the complete genome reconstructed through the Gibson assembly method were included in the annotation process by predicting the ORFs and identifying the corresponding genome organization. Results showed that all genomes included in the analysis had the same organization: They had a single open reading frame encoding a polyprotein of various

sizes. For instance, polyproteins encoded by RNAs-1 (P1) had sizes varying between 2369 aa and 2483 aa, while those encoded by RNAs-2 (P2) presented quite close sizes varying between 1067 and 1069 aa (Table V-1). Therefore, the genome of the CGO-BC isolate has been used as a reference for the annotation presented in Figure V-5.

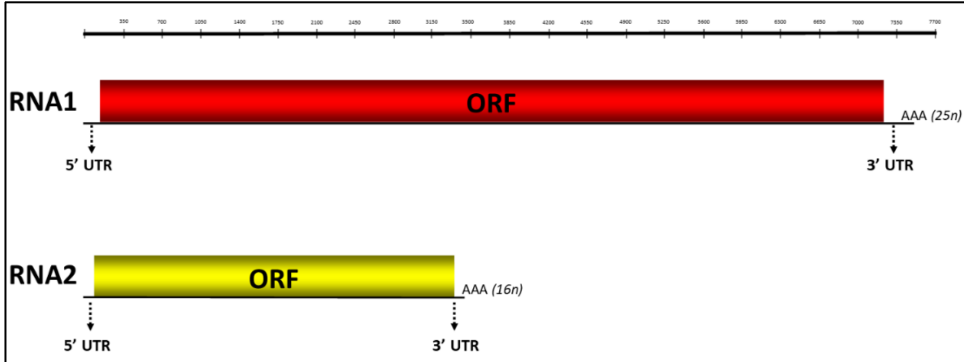


Figure V-6. Genome organization of the RNA1 and RNA2 segments for the reference isolate CGO-BC. The RNA genomes are drawn as black lines, and coloured rectangles represent the predicted open reading frames. A scale is presented above to size the length of the RNA genomes in nucleotides.

Table V-1. Genome organization of the 13 contigs reconstructed in this study

RNA1

Contig names	Genome length	5' UTR	ORF1		3'UTR	Poly-A tail
	nt	nt	nt	aa	nt	nt
CGO-BC (reference)	7,405	131	7110	2,369	139	25
CGO-SK3	7,731	171*(6)	7380	2460	153	29
CGO-SK4	7,61	37*(144)	7,391	2,463	153	29
CGO-SK5	7,737	166*(10)	7,389	2,462	153	29
CGO-SK12	7,795	163*(10)	7,449	2,482	154	29
TZ	7,637	171*(5)	7,347	2,449	120	-

RNA2

Contig names	Genome length	5' UTR	ORF1		3'UTR	Poly-A tail
	nt	nt	nt	aa	nt	nt
CGO-BC (reference)	3460	95	3210	1069	139	16
CGO-SK3	3479	110*(8)	3204	1067	150	30
CGO-SK4	3487	108*(10)	3204	1067	150	30
CGO-SK5	3475	109*(9)	3204	1067	152	30
CGO-SK12	3460	103*(8)	3204	1067	151	30
TZ-55	3457	109*	3195	1065	153	-
TZ-56	3449	96*(6)	3195	1065	158	-

*: The 5’ ends of these contigs were not fully reconstructed. The number of missing nucleotides for each genome compared to the fully reconstructed one is shown in brackets. – information regarding the poly-A tail is unavailable as the 3’ extremity was not experimentally checked.

Genomes for which the 5’ ends of the RNA1 and RNA2 were not experimentally reconstructed were aligned to the fully reconstructed one (CGO-BC) using ClustalW. The comparison suggested that for most of them, less than 11 nucleotides were missing in the 5’ extremities to achieve the whole genome (except for the CGO-SK4 genome, for which 144 nt were missing in the 5’ end of the RNA1). On the other side, the 3’ ends of all remaining contigs were fully retrieved using 3’RACE except for those reconstructed through RNASeq data from the SRA database. The alignment of RNA2 sequences to the CGO-BC reference suggested the entire 5’ extremity could be retrieved for the TZ-55 contig reconstructed through RNASeq data, although it was not experimentally checked.

This genomic organization, size, number of ORFs, and genome segments are typical of viruses from the genus cheravirus in the Secoviridae family.

b. Phylogenetic analysis

Two multiple amino acid sequence alignments were used to perform pairwise amino acid comparisons and generate phylogenetic trees that early address the taxonomic position of these novel isolates in the family Secoviridae. In addition, as full-length nucleotide sequences corresponding to the RNA1 and RNA2-encoded polyproteins for all reconstructed contigs were obtained, phylogenetic analyses included them.

The first multiple sequence alignment was built using the conserved amino acid region between the “CG” motif of the protease and the “GDD” motif of the polymerase (the Pro-Pol region) in the RNA1-encoded polyprotein of isolates characterized in this study together with various genus representative species from the family Secoviridae. The second multiple sequence alignment was built using the coat protein amino acid sequences of isolates characterized in this study and those of selected type isolates from the recognized genus in the family Secoviridae. Phylogenetic trees were built for both alignments, following the procedures mentioned above. As both trees provided approximately identical results, only the phylogenetic tree resulting from the Pro-Pol region is shown in Figure V-6. The tree resulting from the CP block could be accessed on supplementary material 4.

These phylogenetic investigations placed the isolates herein described into the same clade with members of the genus cheravirus with strong bootstrap support, adjacent to the stocky prune virus (StPV), arracacha virus B (AVB), currant latent virus (CuLV), cherry rasp leaf virus (CRLV) and apple latent spherical virus (ALSV). This suggested a phylogenetic evolutionary relatedness of these isolates to members from the genus cheravirus.

The molecular identity at the amino acid level for the “Pro-Pol” and the CP regions between isolates described here and representative species of each genus from the

family Secoviridae was systematically lower than 47% and 13%, respectively (Table V-2), far below the 80% and 75% identity cut-offs required to demarcate species within the Secoviridae family. This suggests that isolates from this study are novel species in the family Secoviridae and could be phylogenetically related to the cheravirus genus.

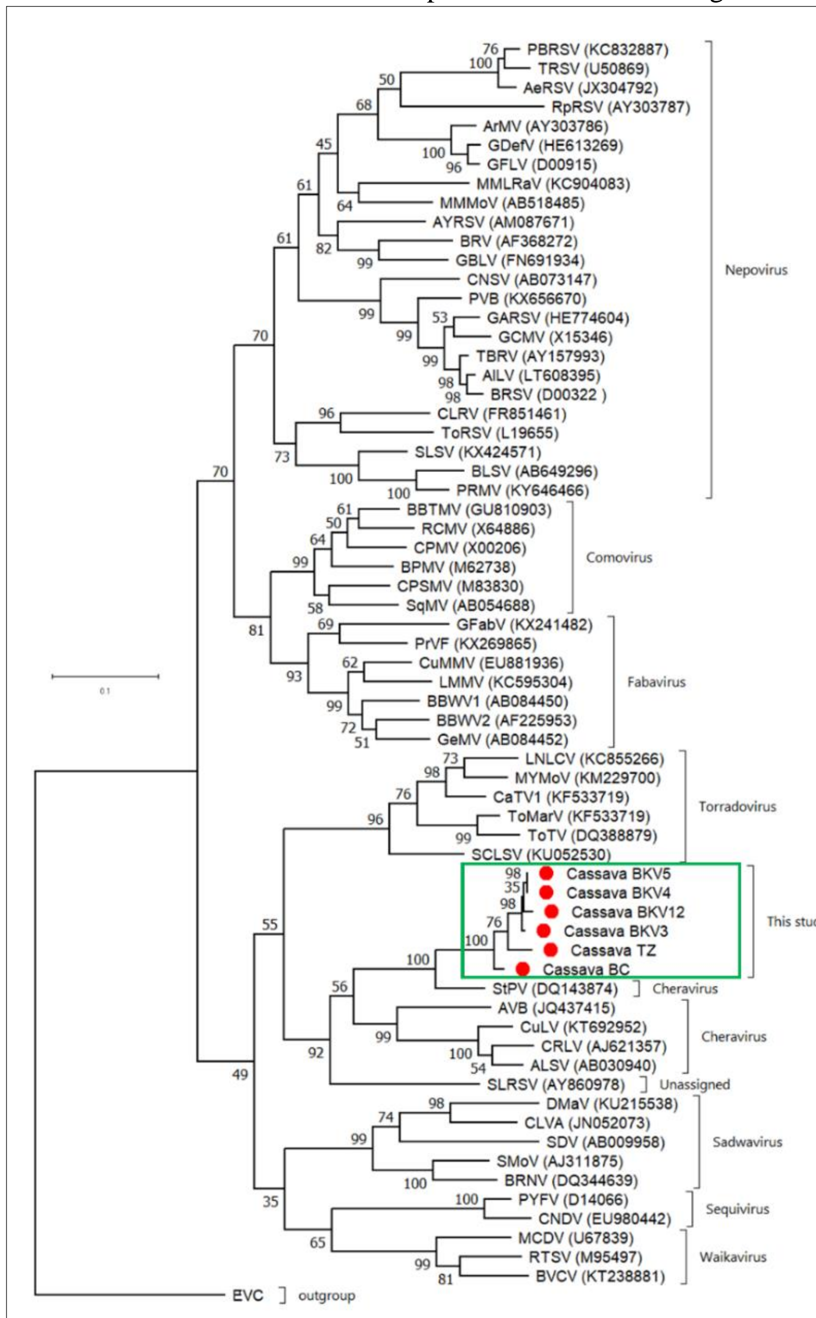


Figure V-7**.** Phylogenetic analysis of the conserved amino acid region between the "CG" motif of the protease and the "GDD" motif of the polymerase (the Pro-Pol region) for isolates detected in cassava (strong green triangles and green frame in the tree) and of

type isolates from recognized species in the family *Secoviridae* (detailed information on these viruses is provided in supplementary material 5).

For each Secoviridae species, the amino acid sequence of the "Pro-Pol" region was deduced from the nucleotide sequence of the corresponding genomic RNA from the type isolate. The alignment was generated using ClustalW integrated into MEGA X. The Maximum likelihood phylogenetic tree was also reconstructed in Mega X using the Poisson model with uniform distribution for amino acid sequence alignments. Bootstrap values are indicated at the primary branch nodes (1000 replicates). The bar represents the number of amino acid substitutions per site. The tree was rooted using the Pro-Pol sequence of poliovirus (EVC, species Enterovirus C, genus Enterovirus, family Picornaviridae).

Table V-2. Variation ranges of the identity percentages for the conserved "Pro-Pol" region and the coat protein amino acid sequences between isolates described in this study and representative species of various genera in the family Secoviridae.

Genus	**Representative species**	**Pro-Pol region**		**CP region**	
		Lowest value	Highest value	Lowest value	Highest value
Cheravirus	Cherry rasp leaf virus	45%	46%	8%	9%
Torradovirus	Tomato torrado virus	37%	38%	10%	11%
Waikavirus	Rice tungro spherical virus	32%	33%	11%	12%
Sequivirus	Parnship yellow fleck virus	37%	37%	10%	11%
Sadwavirus	Satsuma dwarf virus	30%	30%	10%	11%
Nepovirus_C	Tomato ringspot virus	30%	32%	8%	10%
Nepovirus_A	Grapevine fanleaf virus	34%	34%	10%	10%
Nepovirus_B	Beet ringspot virus	32%	33%	8%	10%
Comovirus	Cowpea mosaic virus	33%	34%	8%	10%
Fabavirus	Broadbean wilt virus 2	35%	36%	7%	8%

V.4.4. Diversity of the new isolates

Intergroup and intragroup amino acid and nucleotide sequences' pairwise identities were computed to investigate the molecular diversity between these various reconstructed Secoviridae members originating from various locations. The computation over the "Pro-Pol" region, the coat protein block and the RNA1 and RNA2-encoded polyproteins included amino acid pairwise comparisons. In contrast, the computation over the complete and nearly complete genome sequences of RNA1 and RNA2 involved the pairwise nucleotide comparison. Phylogenetic trees of corresponding aspects were also reconstructed and appended in supplementary material 6.

From all phylogenetic reconstructions performed, two separated clusters could be identified: one cluster gathering the isolates from South-Kivu (CGO-KV3, CGO-KV4, CGO-KV5 and CGO-KV12) and another cluster of isolates from Tanzania (TZ55 and TZ56). However, the phylogenetic assignation of the isolate from Bas-Congo (CGO-BC) was quite challenging to determine since its topology appeared to be quite variable: sometimes associated with the cluster from South-Kivu (for the trees involving the Pro-Pol region and the P1 polyprotein), sometimes associated to the cluster from Tanzania (for the trees involving the RNA1 nucleotide sequences and CP block) and sometimes clustering distinctly from both clusters (for the trees involving the P2 polyprotein and the RNA2 nucleotide sequences).

The amino acid sequence identity comparison between isolates from South Kivu and isolates from Tanzania suggested 86% and 70% for the Pro-Pol region and the CP block, respectively (Table V-4). It is worth noting that all other comparisons, including the RNA1 segments of isolates from these two clusters (entire P1 polyprotein and complete RNA1 nucleotide sequences), showed identity percentages below the 80% demarcation level but not the Pro-Pol region. These values suggested only the CP cut-off is met for the demarcation of these two groups as distinct species (CP<75%). On the other side, the isolate from Bas-Congo appeared to be closer to isolates from South-Kivu (87% and 81% identical respectively for the Pro-Pol region and the CP block) than to isolates from Tanzania (84% and 75% id. For the Pro-Pol and the CP block respectively). Species demarcation cut-offs between Bas-Congo and Tanzania isolates are not met for the Pro-Pol region and fall quite borderline for the CP block. All isolates from South Kivu belonged to the same group as isolates from Tanzania (Pro-Pol and CP identity levels below the demarcation cut-offs) (Table V-3).

Table V-3**.** Intergroup and intragroup average pairwise identity and standard deviation calculated for the “Pro-Pol” region, the CP block, the RNA1 and RNA2-encoded polyproteins (P1 and P2, respectively) and the nucleotide sequences of (nearly) complete genomes of the RNA1 and the RNA2 segments.

		Within Kivu isolates	**Within Tanzania isolates**	**Between Kivu & Tanzania isolates**	**Between Kivu & Bas-Congo isolates**	**Between Bas-Congo & Tanzania isolates**
"Pro-Pol" region	aa id	97% ± 0.5%	-	86% ± 1.5%	87% ± 1.5%	84% ± 1.7%
CP block	aa id	98% ± 0.4%	97% ± 0.8%	70% ± 2%	81% ± 2%	75% ± 2%
P1 polyprotein	aa id	94% ± 0.4%	-	79% ± 0.8%	82% ± 0.8%	80% ± 0.8%
P2 polyprotein	aa id	97% ± 0.4%	95% ± 0.6%	74% ± 1%	71% ± 1%	73% ± 1%
RNA1 genome	nt id	90% ± 0.2%	-	75% ± 0.5%	75% ± 0.5%	75% ± 0.5%
RNA2 genome	nt id	95% ± 0.3%	89% ± 0.6%	72% ± 0.7%	70% ± 0.8%	73% ± 0.8%

Considering the taxonomic differentiation between the various groups [South-Kivu (CGO-SK), Bas-Congo (CGO-BC) and Tanzania(TZ)] as illustrated by the analyses conducted herein, it is pretty challenging to address their diversity at the species level as both demarcation criteria suggested by the ICTV could not be concomitantly met

for each comparison. Additional information about the biology of these various isolates (host range, vector specificity, possibility of re-assortment between RNAs) is therefore required to shed light on this as recommended by the ICTV (Karasev et al. 2017).

V.4.5. Identification of the RNA-1 and RNA-2 encoded proteins and cleavage sites

a. Proteins and cleavage sites inside the RNA1-encoded polyprotein

Putative cleavage sites were predicted based on alignments of translated amino acids from the RNA-1 and RNA-2 encoded polyproteins (P1 and P2, respectively) for isolates identified in this study and those of related secovirids for which cleavage sites have been experimentally characterized previously (A. Wang and Sanfacon 2000; Wetzel et al. 2008) (Figure V-7). The locations of these cleavage sites within P1 and P2 polyproteins for the thirteen RNA segments are presented in Table V-4, and the characteristics of the corresponding domains are also shown in Table V-5.

As shown in Figure V-4, in RNA1, six putative cleavage sites delineating seven putative domains could be identified. One of these putative domains showing similarities with HAM1 proteins was identified in the 3'ends of the P1 polyproteins.

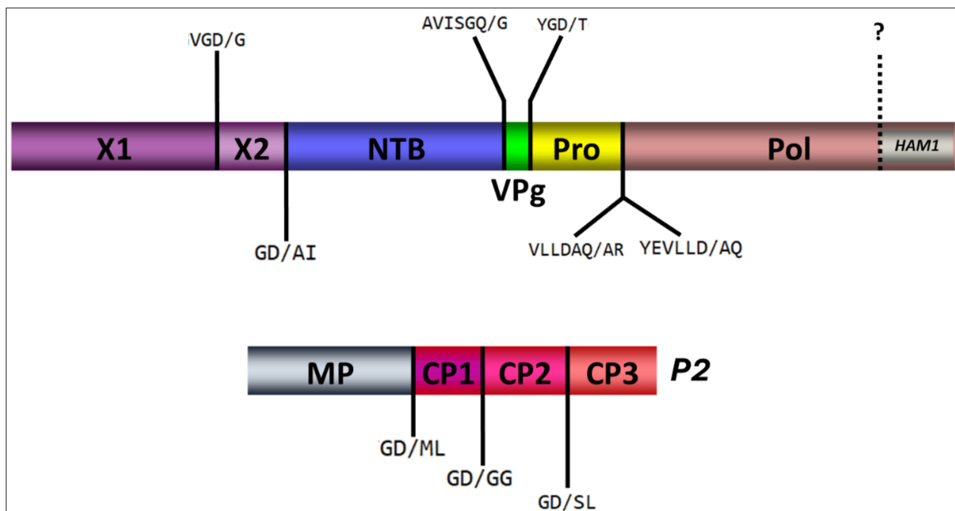


Figure V-8. Putative consensus cleavage sites delineating putative functional domains predicted within RNA1 (P1) and RNA2 (P2)-encoded polyproteins.

Predicted cleavage sites are shown as short vertical lines above and behind the full-length polyproteins, along with the predicted cleaved peptide. Deduced functional domains based on sequence homologies with related viruses are presented as follows: X1; X2; NTB, nucleoside triphosphate binding protein or putative helicase; VPg, viral genome-linked protein; Pro, protease; Pol, polymerase; HAM1, putative inosine triphosphate pyrophosphatase; MP, movement protein; CP1, CP2, CP3, capsid protein 1,2 and 3 respectively.

Table V-4: Location of putative cleavage sites predicted in the RNA1 and RNA2-encoded polyproteins for isolates identified in cassava plants in this study.

RNA1 Polyprotein						
Domains	**CGO-SK3**	**CGO-SK4**	**CGO-SK12**	**CGO-BC**	**TZ**	**Consensus**
X1-X2	LVGVGD^{565}/GL	LVGVGD^{514}/GL	LVGVGD^{533}/GM	LMVGD^{419}/GF	LMVGD^{493}/GF	VGD/G
X2-NTB	EHMKGD^{730}/AI	I*MKGD^{681}/AI	ECMRGD^{698}/AI	EKMKGD^{585}/AI	GMOGD^{657}/AI	GD/AI
NTB-VPg	AVISGQ^{1,310}/GD	AVISGQ^{1,261}/GD	AVISGQ^{1,314}/GD	AVISGQ^{1165}/GD	AVISGQ^{1237}/GD	AVISGQ/GD
VPg-Pro	ADLYGD^{1,365}/TQ	ADLYGD^{1,316}/TQ	ADLYGD^{1,369}/TQ	ADMYGD^{1220}/TQ	ADMYGD^{1292}/TQ	YGD/TQ
Pro-Pol	VLLDAQ^{1,609}/AR	VLLDAQ^{1,560}/AR	VLLDAQ^{1,613}/AR	VLLDSQ^{1463}/AR	VLLDAQ^{1535}/AQ	VLLDAQ/AR
	YEVLLD^{1,607}/AQ	YEVLLD^{1,558}/AQ	YEVLLD^{1,611}/AQ	YEVLLD^{1461}/SQ	YEVLLD^{1533}/AQ	YEVLLD/AQ
RNA2 Polyprotein						
Domains	**CGO-SK3**	**CGO-SK4**	**CGO-SK12**	**CGO-BC**	**TZ**	**Consensus**
MP-CP1	ESGTGD^{447}/ML	ESGTGD^{447}/ML	ELGTGD^{456}/ML	ETGFGD^{447}/ML	ESAKGD^{447}/ME	GD/ML
CP1-CP2	KPSFGD^{623}/GG	KPSFGD^{623}/GG	KPSFGD^{632}/GG	KPSFGE^{624}/GG	KPSFGD^{618}/GG	GD/GG
CP2-CP3	LLPSGD^{868}/SL	LLPSGD^{868}/SL	LLPSGD^{877}/SL	MMAQGD^{869}/AL	LAQGD^{863}/AL	GD/SL

*Numbering corresponds to the amino acid position starting from the beginning of the polyprotein. All cheraviruses proteases have Histidine in their substrate binding pockets, suggesting that they should recognize cleavage sites with Q, E or D at the -1 position. The red highlight is filled when this consensus is met by isolates described in this study.

Table V-5: Characteristics of putative RNA1 and RNA2-encoded domains identified inside the P1 and P2 polyproteins of viruses reported in this study

POLYPROTEINS	**PROTEINS**						
	Names	**Reference genome**		**Shortest genome**		**Largest genome**	
		Contig id	size (aa) & MW (kDa)	Contig id	size (aa) & MW (kDa)	Contig id	size (aa) & MW (kDa)
P1	**X1**	CGO-BC (RNA1)	418(46)	CGO-BC	418(46)	CGO-SK12	531(57.5)
	X2		166(18.6)	TZ	164(18)	CGO-BC & SK4	166(18.6)
	NTB		580(66.6)	*All**	580(66.6)	*All**	580(66.6)
	VPg		55(6)	*All**	55(6)	*All**	55(6)
	Pro		243(27)	TZ	241(27)	CGO-SK3,4,5 & 12	244(27)
	Pol		907(102)	CGO-SK5	906(102)	TZ	916(103)
P2	**MP**	CGO-BC (RNA2)	447(50)	CGO-SK3,4 & 5	446(50)	CGO-SK12	455(51)
	CP1		176(19)	TZ-55 & TZ56	171(18)	CGO-SK3,4,5 & 12	176(19)
	CP2		245(27)	*All**	245(27)	*All**	245(27)
	CP3		201(23)	CGO-SK3,12	200(23)	TZ-55 & TZ-56	202(23)

*The corresponding protein domain has the same size and molecular weight across polyproteins encoded by all reconstructed contigs. Therefore, numbers not in parentheses

represent the sizes in amino acids (aa), while those in parentheses represent molecular weights in kDa.

Two domains (X1 and X2) were previously identified in the N-terminal region of the RNA1-encoded polyprotein of nepovirus (Wang and Sanfacon 2000; Wetzel et al. 2008) and sadwavirus (subgenus stramovirus) (Mann, Walker, and Sanfaçon 2017). The existence of such an unusual cleavage site delineating these two domains upstream of the NTB had not been suggested for viruses in the genus cheravirus. Only a more extended NTB domain was so far known as delineating the 5' end of the RNA1-encoded polyprotein in this genus. However, when looking up amino acid sequence alignment of the P1 polyproteins, including those of viruses reported in this study, those of all known cheraviruses and other Secoviridae members whose the X1 and X2 domains have been experimentally characterized (Supplementary material 7), a putative cleavage site can also be found upstream the NTB. This suggests that cheraviruses also could have two protein domains upstream of the NTB. This cleavage site is characterized by the presence of an aspartic acid (D) in the -1 position [(for other cheraviruses it is a glutamic acid (E) or a glutamine (Q)] and a glycine (G) in the +1 position (as it is the case for other cheraviruses except for Arachacha virus B which has an alanine (A) at this position). Thorough comparison of the X1 domain against Blast protein search identified 45% similarity with the replicase of trillium govanianum cheravirus (accession number DAF42463.1). The X1 protein is more variable among nepoviruses and cheravirus and its function is still unknown. The X1 domain of the isolate CGO-BC has the shortest size (418 aa) while the largest size is located inside the CGO-SK12 contig genome (Table V-5).

The function of the X2 domain has been previously investigated for the tomato ringspot nepovirus (TRNV). It crossed the membrane twice since it possessed two transmembrane helices (G. Zhang and Sanfaçon 2006). Using the algorithm of Kyte and Doolittle with a window size of 17 amino acids (Kyte and Doolittle 1982), a computer-assisted prediction of transmembrane helices (TMH) inside the corresponding regions of the X2 protein for viruses reported in this study (Figure 8) proposed the existence of two domains with high values of hydrophobicity. Therefore, these highly hydrophobic domains could act as putative transmembrane helices, conferring the X2 protein the role of a transmembrane protein. The most similar protein identified using the Blast-protein search engine was associated with the replicase of trillium govanianum cheravirus with 48% similarity. Viruses identified in this study had sizes varying between 164 and 166 amino acids for the X2 protein (Table V-5).

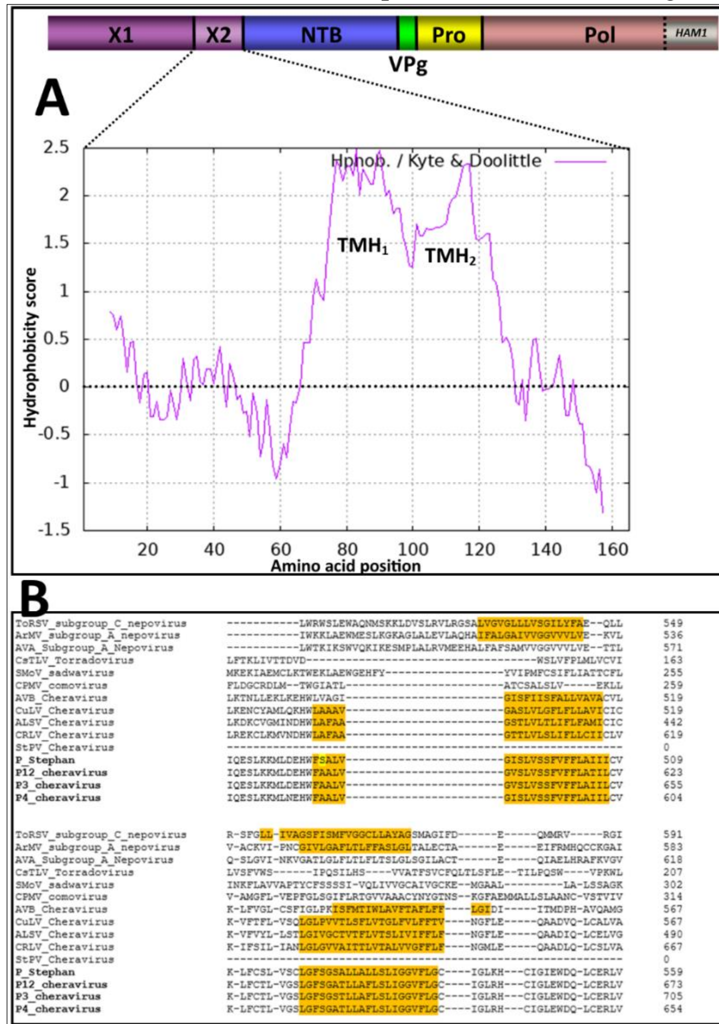


Figure V-9**.** Computer-assisted prediction of transmembrane helices (TMH) in the X2-encoded domain of the RNA1 segment:

(A) The hydrophobicity plot of X2, calculated using the algorithm of Kyte and Doolittle with a window size of 17 amino acids (Kyte and Doolittle 1982). The RNA1-encoded polyprotein is shown at the panel's top with the indicated individual protein domains. Vertical lines represent the putative cleavage sites recognized by the RNA1-encoded 3C-like proteinase of viruses reported in this study. (B) Multiple amino acid sequence alignment of the X2 domain showing amino acid residues associated with the putative double trans-membrane helices for a 2-pass transmembrane X2 domain of viruses reported in this study similar to that of nepoviruses (orange highlighting in alignment). Amino acids are numbered from the first amino acid of the P1 polyprotein.

The dipeptide D/A delineated the X2 and the NTB domains. Based on amino acid sequence alignment, this cleavage site aligned well with cleavage sites previously

reported for other cheraviruses but differed in composition (it was a Q/G for isolates described in this study). The putative NTP binding protein (NTB) domains identified for viruses reported in this study had the size of 580 amino acids for the four isolates described here. They had a molecular weight of 66.5 kDa. Proteins containing the NTB domain have been previously shown to be transmembrane proteins associated with the endoplasmic reticulum (ER) -derived membranes that anchor the replication complex to the ER (Aiming Wang, Han, and Sanfaçon 2004).

The putative NTB-VPg cleavage site predicted for isolates described in this study was the dipeptide Q/G (Table 4). This cleavage site was the same as ALSV and CuLV and aligned well with cleavage sites predicted for other cheraviruses using the complete RNA1-encoded polyprotein sequence alignment. The putative VPg domains of all isolates identified in this study had the size of 55 amino acids and six kDa molecular weight (Table V-5). No significant similarity was identified through blast searches.

The VPg-Pro cleavage site was variable in location and composition among all cheraviruses involved in this analysis. However, this site could be putatively cleaved by the dipeptide D/T for isolates discovered in this study. The upstream region delineated by this cleavage site is the 3C-like cysteine protease responsible for the proteolytic processing of the RNA1 and RNA2 - encoded polyproteins. The conserved Histidine of the protease substrate-binding pocket for the novel viruses reported in this study and those of other viruses included in the analysis are shown in supplementary material 8.

The size of the protease domains across herein described isolates varied between 241 and 244 aa while their corresponding molecular weights were equivalent (27 kDa) (Table V-5).

The protease and the polymerase domains could be cleaved at one of the two possible cleavage sites suggested in Table V-5. These two cleavage sites were not distant; they were located at just two amino acids. With sizes varying between 906 and 916 amino acids for the six isolates, the polymerase was the most prominent protein of the RNA1-encoded polyprotein. The conserved “CG-GDD” motif between the protease and the polymerase domains that is characteristic of viruses in the family Secoviridae (used for phylogenetic classification) was quite “particular” for the viruses described in this study. They had an insertion of 36 additional amino acids, an uncommon feature compared to the sequences of all other included viruses (Supplementary material 9).

Comparing the complete amino acid sequence of the polymerase domain for the new viruses reported in this study toward the blast database identified 54% identity with the stocky prune virus RNA1 polyprotein. However, looking closely at the 5’ extremity of the polymerase domain, Except for the cassava torrado-like virus, it seemed longer than expected (Supplementary material 10). This pattern is investigated on point IV.5. below.

b. Proteins and cleavage sites inside the RNA2-encoded polyprotein

Within the RNA2-encoded polyprotein (P2), three putative cleavage sites delineating four putative domains could be identified (Figure V-7, Tables V-4 and V-5).

A first putative cleavage site GD/ML (Table V-4) delineating a putative movement protein (MP) in the N-terminal region of the RNA2-encoded polyprotein was predicted. This cleavage site was followed downstream by two additional putative cleavage sites, GD/GG and GD/SL, delineating three putative Coat protein subunits (CP1-CP2 and CP2-CP3). The sizes and molecular weights of movement protein domains for viruses reported in this study were variable, with the CGO-SK12 MP showing the largest size and molecular weight (455 aa and 51 kDa, respectively). Through NCBI, blast protein identified the similarity of these putative MP to the apple latent spherical virus (ALSV, NP_733989.1). However, the MP of the CGO-BC isolate showed a particularity. It had an insertion of 6 amino acids in its N terminal extremity (6 amino acids upstream of the cleavage site delineating the MP-CP1 domains) (Supplementary material 11). These additional amino acids contained a KRK motif (highlighted in red) that could be similar to the potential nuclear or membrane localization signal (A. M. James et al. 2021).

The three putative coat protein blocks identified within the RNA2-encoded cleavage site did not have the exact sizes and molecular weights. However, the CP2 had the largest size and molecular weight (245 aa and 27 kDa), while the CP1 had the smallest one (176 aa and 19 kDa) (Table V-5).

Table V-6. Consensus cleavage sites of isolates from this study and those of previously-described viruses from the genus Cheravirus

RNA1-Polyprotein						
Sites	**This study**	**ALSV**	**CRLV**	**CuLV**	**AVB**	**StPV**
X1-X2	VGD/GL	NTKEGQ/GP	NCKVGE/GP	EGKRGE/GP	DRYVGE/AG	-
X2-NTB	GD/AI	EALRGQ/GL	DDLRGQ/GV	DAKKGQ/GI	SSATGQ/GP	-
NTB-VPg	AVISGQ/GD	SSLSAQ/GP	SNLSGD/GA	SNLSGQ/GP	AFCAFK/GE	-
VPg-Pro	YGD/TQ	IPLWGQ/GP	ADFFGE/GP	SDYEGQ/GP	ADFYGE/GP	-
Pro-Pol	VLLDAQ/AR YEVLLD/AQ	SEKVGQ/GP	LSDKGQ/GP	GTLVGE/GP	TLQDIE/GA DFCAGE/VA	GCSLPE/VE
RNA2 Polyprotein						
Sites	**This study**	**ALSV**	**CRLV**	**CuLV**	**AVB**	**StPV**
MP-CP1	GD/ML	NLLEGQ/GP	NLLEGQ/GP	NLLEGQ/GP	IAGVGE/GP	AGSVGD/VT
CP1-CP2	GD/GG	FYNIGQ/GA	VYNLGQ/GQ	FYNLGE/SN	SEYHGN/AT	-
CP2-CP3	GD/SL	GPLVGE/GS	PILAAE/GP	LSLEGQ/GP	GFSLGE/AN	-

The comparison between the consensus cleavage site of viruses reported in this study and those of other cheraviruses previously described, it seemed that there is a strong preference for an aspartic acid (D) in the -1 position of viruses reported in this study (Table V-6). This preference is quite unusual but possible.

V.4.6. A putative Maf/HAM1 motif is present within the C-terminal part of the polymerase domain

The screening of the 5' extremity of the polymerase domain against the NCBI blast protein (BLASTp) (Sayers et al. 2022), the Uniprot (Swiss-Prot and TrEMBL) (The UniProt Consortium 2021), and the InterPro databases (Blum et al. 2021) revealed similarities with the ITPase superfamily of proteins ubiquitous in prokaryotic and eukaryotic organisms as illustrated in supplementary material 10. This sequence of 232 amino acids could be putatively delineated in the N-terminal extremity by peptides similar to those described for upstream domains (FMSAD2288/GL)(Supplementary material 10).

Investigations on the phylogenetic relatedness between these ITPase-like proteins identified within the RNA1-encoded polyprotein of viruses reported in this study with those of representative organisms from various kingdoms of life are shown in Figure V-9.

Results showed that ITPase proteins of viruses reported in this study formed a clade phylogenetically adjacent to ITPases from the CsTLV and EuRSV, suggesting their possible phylogenetical relatedness. Furthermore, this clade seemed to derive from the clade formed by bacterial ITPases.

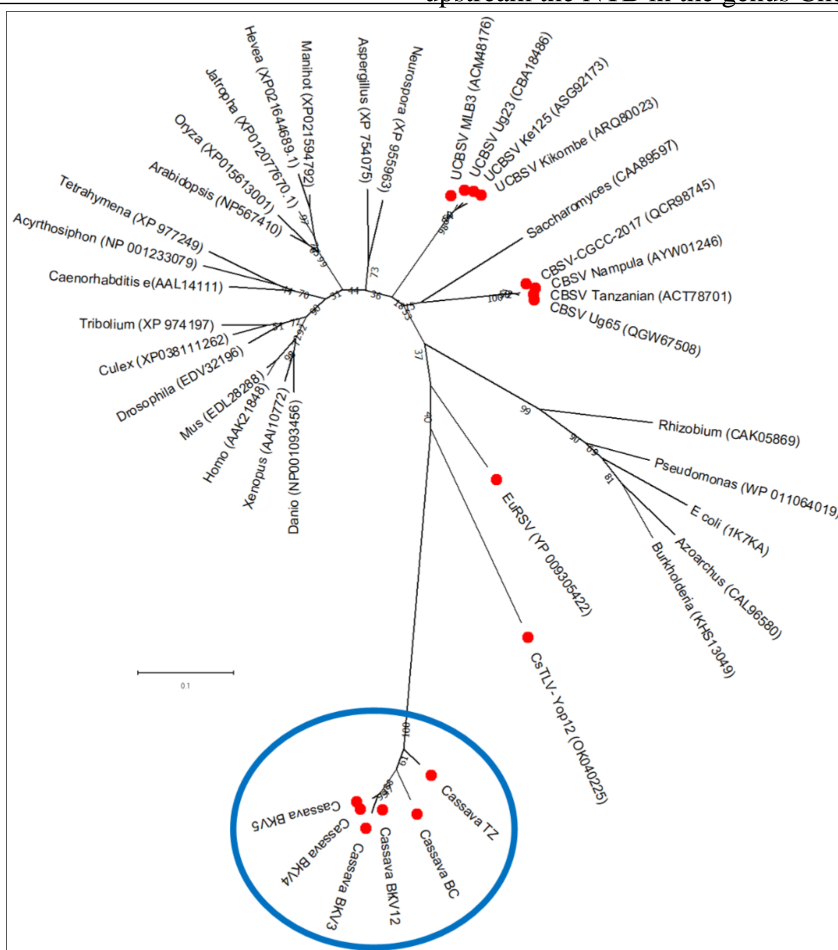


Figure V-10**.** Phylogenetic relationships of Maf/HAM1 motifs between viruses reported in this study (framed in blue) with those of representative members of different kingdoms of life (viruses are pointed in solid red circles).

The alignment was generated using ClustalW integrated into MEGA X, and the Maximum likelihood phylogenetic tree was also reconstructed in Mega X using the LG+G model for amino acid sequence alignments. Detailed information on these sequences (complete names) is provided in supplementary material 12.

The exact function of the viral ITPase has not been broadly characterized apart from investigations of (K. R. Tomlinson et al. 2019), who proposed that the presence of Ham1 proteins with highly conserved ITPase motifs for the U/CBSV served as a necrosis determinant in Nicotiana benthamiana and would serve highly selectable functions during infections of cassava. However, (Gall et al. 2013) categorized key residues/motifs that influence ITPase substrate specificity while investigating the biochemical model for human ITPase. They identified three essential, four intermediates and one dispensable amino acid residues. Active sites of an enzyme, determining its substrate specificity, are a critical element that can provide

information on its function. To investigate whether crucial residues from the human ITPase active sites could be conserved within ITPases from other organisms, including viruses reported in this study, we performed an amino acid alignment shown in supplementary material 13. Results from this alignment could be synthesized in Table V-7.

Table V-7. Comparison of conserved amino acid residues in ITPases of representative members of various kingdoms of life (including viruses reported in this study) to essential amino acid residues that influence substrate specificity in human ITPase.

N°	Essential amino acid residues in human ITPase	Correspondent amino acid residues in alignment with :					
		E. coli	*S. cerevisiae*	CsTLV	EuRSV	CBSV & UCBSV	Viruses in this study
	Essentials						
1	Glutamic acid (E_22)	E	E	E	E	E	D
2	Tryptophan (W_151)	Y	W	F	W	W	W
3	Arginine (R_178)	R	R	R	R	R	R
	Intermediates						
4	Phenylalanine (F_149)	F	F	F	F	F	Y (tyrosine)
5	Aspartic acid (D_152)	D	D	D	D	D	E
6	Lysine (K_152)	K	K	K	K	K	S
7	Serine (S_176)	S	S	G (glycine)	S	S	C (cysteine)
	Dispensable						
8	Histidine (H_177)	H	H	L (leucine)	H	H	A (alanine)

Table cells are filled in green when the residue is conserved when referring to the human ITPase; otherwise, the orange fill is highlighted.

Results suggested that from the three essential amino acids of the human ITPase suggested by (Gall et al. 2013), only two were conserved in viruses reported in this study (arginine and Tryptophan), while all intermediates and dispensable amino acids were not. For the CsTLV, the previously reported secovirid with the HAM1 motif, 2 out of the three essential amino acids were conserved (glutamic acid and arginine) together with three intermediate amino acids (phenylalanine, aspartic acid and lysine). However, the three viral species from the Potyviridae family (EuRSV, CBSV and UCBSV) had all eight amino acid residues conserved. The Arginine residue was the only one to be conserved across all the taxa in the alignment.

V.5. Discussions and conclusions

This study reported the genome characterization of six novel viral isolates from field-grown cassava plants collected in the D.R. Congo and from mining publicly available RNAseq data. All the isolates belonged to the Secoviridae family and could potentially represent a novel Secoviridae species only distantly related to other cheravirus species.

Thirteen contigs have been assembled, aggregated in two groups according to their sizes and corresponding to RNA1 and RNA2 segments. Analysis of the genome organization revealed a single open reading frame encoding a unique polyprotein of various sizes for each reconstructed genome. The 3’ extremities of all segments collected from field-grown plants have been determined by RACE. In contrast, only one contig from each RNA1 and RNA2 segment was included in the 5’ extremity determination. This suggested that RNA1 and RNA2 segments corresponding to one

isolate were fully reconstructed and served as the reference for further analyses. For the remaining genome segments, the comparison with the fully reconstructed suggested less than eight nucleotides were missing in the 5' extremity of the RNA2 segments, while less than 11 nucleotides were missing in the 5' extremity of RNA1 segments except for one isolate, which had 144 missing nucleotides. The genomic organization, the size, the number of ORF, and the number of genome segments were typical of viruses from the Cheravirus genus in the family Secoviridae.

The replication of many positive-strand RNA viruses is associated with the endoplasmic reticulum (E.R.) membrane, where they induce the formation of a viral replication complex (VRC) to protect from RNA degradation by the host during the replication process (Sanfaçon 2013). The binding/anchoring of the VRC to the E.R. membranes has been reported to be mediated by two E.R.-targeting sequences within the NTP-binding protein (NTB) of the tomato ringspot Nepovirus (Aiming Wang, Han, and Sanfaçon 2004). The NTB protein identified for the viruses reported in this study could also play this role. Additionally, a second possible protein, the X2, located directly upstream of the NTB domain, has been proposed to act as the second membrane anchor for the VRC of the same reference virus (G. Zhang and Sanfaçon 2006). This X2 protein was reported to have multiple E.R.-targeting domains, including two C-terminal transmembrane helices and a less well-defined amphipathic domain upstream. These characteristics previously described for the X2 protein could also be shared for viruses described in this study since results obtained from the computer-assisted prediction of transmembrane helices supported the existence of 2 domains with higher hydrophobicity values equivalent to transmembrane helices, within their X2 protein.

The type of amino acid present within the substrate-binding pocket of a viral protease determines the type of amino acid that can be recognized at the -1 position while cleaving the polyprotein. Consequently, this is the critical characteristic determining the specificity of the cleaving activity of the enzyme. For instance, the proteinases of subgroup A and B nepoviruses have a leucine in their substrate-binding pockets (Margis and Pinck 1992), and they cleave after lysine or arginine (for subgroup B) or after a cysteine, glycine or arginine (for subgroup A). However, all cheravirus proteases have histidine in their substrate binding pockets. This amino acid was identified in the substrate-binding pocket of viruses reported in this study. This would suggest that, like cheraviruses, viruses herein reported should recognize cleavage sites with glutamine (Q), glutamic acid or an aspartic acid (D) at the -1 position (Asparagine is also possible, although less frequent). As expected, the D and the Q were identified within the RNA1 and RNA2-encoded polyproteins, with a preference for a D. However, typically, torradoviruses protease also has the H in their substrate binding pocket (the tomato torrado virus). Still, surprisingly, the previously-reported CsTLV had glutamine instead of histidine in the protease substrate binding pocket, which could affect its specificity. The +1 position usually has a relatively small amino acid, but this position has some flexibility. So, the amino acids usually found in this position (for secovirids, potyvirids and picornaviruses) are glycine (G), serine (S), methionine (M), threonine (T) or alanine (A). Asparagine is also occasionally seen in this position. As expected, all these amino acids were present at

the specified position for viruses reported in this study and suggesting it is believable they are possible.

The X1-X2 cleavage sites were predicted with reasonable confidence for all cheraviruses, including the one reported in this study. This suggested that, like nepoviruses, cheraviruses would have two protein domains in front of the NTB domain. Two alternative cleavage peptides were proposed for the pro-pol cleavage site for the viruses reported in this study, but neither had a G at the -2 position. An alanine or leucine in either cleavage site may be acceptable, but we are not sure which cleavage peptide is the correct one. Therefore, the putative cleavage site consensus for the viruses reported in this study would be (G, a) (D, Q) / (G, A, t, s, m). While comparing this cleavage consensus to what was observed for other cheraviruses, the strong preference for a G at -2 position is conserved for all cheraviruses, including the ones reported in this study. The strong preference for a D at -1 position for viruses reported in this study is somewhat unusual but possible. However, some cheraviruses also sometimes had an aspartic acid (D) at that cleavage site position (Table V-4). If considering findings from (Petrzik et al. 2016), who previously compared cleavage sites of three characterized cheraviruses, the putative cleavage site consensus for viruses in the cheravirus genus could be summarized as (G, a) (Q, E, D) / (G, s). Since the SDS-PAGE suggested the presence of three virus particles of sizes lesser than 30 kDa, it is believable that alignment-based predictions of cleavage sites inside the RNA2-encoded polyprotein are also possible. This also makes it believable that the alignment-based predictions made for the RNA1-encoded polyprotein could be possible. However, only experiments would confirm if they are valid.

We looked for a possible cleavage site delineating a putative HAM1 motif downstream of the polymerase. Still, it wasn't easy to know where it could precisely be located according to the consensus cleavage site described. The one we have proposed is possible but not common because it has an alanine instead of a glycine at the -2 position. This amino acid was only identified in the -2 position of the NTB-VPg cleavage site of the apple latent spherical virus (ALSV). Without more information, it is difficult to predict a correct peptide, although we could be right. However, the evidence seems to support the absence of a cleavage site between these two proteins since their connection would allow the HAM1 enzyme to directly dephosphorylate non-canonical nucleotides once integrated within the RNA by the polymerase (Lopez et al. 2022).

Viruses reported in this study are the fifth virus species (and the second for the family Secoviridae) to encode an ITPase-like protein. The fact that these novel viruses are also found infecting a euphorbiaceous host reinforces one of the theories suggested by (A. M. James et al. 2021) that RNA viruses would require this protein to infect cassava plant because of an increased accumulation of cytosolic pools of ITP, which could result in a higher mutation frequency to the infecting viral genome. When increased mutation rates exceed a critical threshold within a viral population, they can cause a decrease in the specific infectivity and lead to its extinction (Cases-gonzález et al. 2008). Similar to the ITPase, a strategy to counteract deleterious mutations have been documented for replicases from the Flexiviridae and Closteroviridae families carrying an insertion from cellular origin: an AlkB domain that is active in repairing

methylation damage of nucleic acids by oxidative demethylation (Bratlie and Drabløs 2005). This domain was hypothesized to be advantageous for the stability of viruses since many of them were reported to infect perennial and woody host plants (Giovanni P. Martelli et al. 2008). However, until now, the function of viral ITPase has only been partially elucidated (K. R. Tomlinson et al. 2019), and the significance of its presence within genomes of only euphorbiaceous-infecting viruses is still under investigation (Lopez et al. 2022). However, the fact that the previously-reported ampeloviruses (Kwibuka et al. 2021) were identified in a single infection in cassava (only in certain situations) raise the question about which strategies they used to deal with high levels of cytoplasmic pools of ITP/XTP in cassava since they do not have a HAM1 gene: do they recruit plant HAM1 or instead do they RdRp less sensible? This aspect deserves more investigation for a better understanding.

Only ITPase from Escherichia coli (RdGb) (Burgis, Brucker, and Cunningham 2003; Ji et al. 2002; Bhatnagar, Bullions, and Bessman 1991), human (ITPA) (Gall et al. 2013; Lin et al. 2001) and yeast (HAM1) (Noskov et al. 1996) have been biochemically and structurally characterized. Results from the phylogenetic comparisons suggested that HAM1 proteins identified for viruses reported in this study could be evolutionarily related to those identified for the EuRSV and the CsTLV, while those from U/CBSV were distant. If the ITPases of the viruses reported in this study had the "house-cleaning" function similar to humans, at least the key residues associated with their active sites would be conserved. However, results suggested that only two of the three essential residues (tryptophan and arginine) described for the human ITPase were conserved for viruses reported in this study, while all intermediates and dispensable residues were not. (Gall et al. 2013) suggested an absolute requirement for arginine at position 178 of the substrate binding pocket, as observed for the ITPases of viruses from this study, for an ITPase ortholog to be active. This arginine has been shown to play a significant role in the recruitment of d(ITP) to the substrate binding pocket and to be a significant determinant of nucleotide discrimination between canonical and non-canonical (d)NTPs while the tryptophan, together with the arginine again, were shown to be the major determinants of non-canonical purine selectivity in ITPase by allowing a tight packing of the nucleobase into the substrate specificity pocket (Gall et al. 2013). CBSV, UCBSV and EuRSV had all of the eight key amino acid residues from the human ITPase substrate-binding pocket conserved, suggesting it would be possible that they assume the same function. However, it is difficult to keep such hypothesis for viruses reported in this study since we don't know if the conservation of only two key amino acid residues would be sufficient to maintain the "house-cleaning" function. As suggested by (A. M. James et al. 2021) and (Mbanzibwa, Tian, Mukasa, et al. 2009); the benefits for viruses infecting euphorbiaceous hosts to incorporate this HAM1 protein would be very important to elucidate.

The phylogenetic investigation using the conserved "pro-pol" region and the coat protein strongly supported the relatedness of the viruses reported in this study to members of the Secoviridae family. However, pairwise comparisons for each of both regions and involving representative species from each Secoviridae genus suggested the "Pro-Pol" regions of viruses reported here were much closer to the secovirids

(from 30% identical to the nepovirus C and the sadwavirus to 46% identical to the cheravirus). In contrast, their CP appeared much distant (from 7% identical to the fabavirus to 12% identical to the waikavirus). Whatever the region included in the comparison, identity percentages were far less behind the ICTV cut-off. Additionally, recombination analyses (not shown in the study) could not detect putative re-assortment events between RNA1 and RNA2 genome segments. The question at this level is whether described data support the demarcation of the viruses described here at the genus or simply at the species level. None of the five genus criteria proposed by the ICTV could be met for the viruses reported here except one: "the number of protein domains and/or processing sites within the polyprotein (s)". This means that as soon as an extra domain is identified within a novel genome, delineating a processing site is not mandatory. Therefore, the presence of the HAM1 domain in genomes of viruses reported in this study makes them satisfying to only one criterion out of the five at the genus level. This seemed insufficient to consider them as forming a distinct genus within the family. However, the ICTV was not clear while stating that "not all criteria may need to be met simultaneously" since they did not clarify "at least how many criteria would be required" or "which criteria are determinant" as it was clarified for the species level criteria. The ICTV-species criteria specify that "Not all criteria need to be met simultaneously. In some cases, sequence information alone can be a good indicator of a distinct species (i.e., when the percentage of sequence identity in both the Pro-Pol and CP(s) regions is well below the proposed cut-off)". This information supports that as the identity rate for the "pro-pol" and the coat protein were largely below the cut-off, viruses reported here could form a novel species in the genus cheravirus even tough information about the antigenic reactions, host range and vector specificity could not be provided. However, these missing data are still mandatory to address the diversity of the various isolates since sequence information regarding the "Pro-Pol" region always felt above the ICTV cut-off while comparing different isolate groups (South-Kivu, Bas-Congo and Tanzania).

The results presented in this study were generated from datasets collected in contexts that are spatially and temporally distinct but interestingly in harmony. On one side, there is an isolate detected and biologically characterized in samples collected in the Bas-Congo province in the Western part of the country during the period before the advent of the HTS technology. On the other side, we have five isolates detected and molecularly characterized in samples collected in the South-Kivu province in the Eastern D.R. Congo during the HTS era. These two provinces are approximately 1700 kilometers distant and separated by a dense equatorial forest barrier that could not favor a natural spread. However, if it comes to be experimentally confirmed, the fact that these viruses are also spread in Tanzania could indicate a wide distribution likely favored by their latent symptomology.

Historically, the Bas-Congo province has been among the leaders in D.R. Congo's cassava production and has held a reference hub for the cassava research program in D.R. Congo (INERA Nvuazi) for years. Since 2002, CBSD-like symptoms of root necrosis have been reported on cassava plants in this province, whereas no etiological evidence could be obtained until now (Zeyimo et al. 2020). However, since one of the isolates described in this study originated from samples collected in this region, we

could not establish a possible association between this isolate and these root necrotic symptoms because data related to the below-ground observations are still awaited.

Interestingly, This study highlighted how the initiative of establishing a publicly-accessible RNASeq database could further valorize data collected for years and deepen the understanding of scientific findings related to plant health. The data mining of the SRA repository allowed us to identify that these novel cheraviruses reported in this study were already captured several years ago and co-infected with U/CBSVs and CMGs.

V.6. Supplementary material

Supplementary material 1. Primers used for Confirmatory RT-PCR, RACE and Sanger sequencing for HAM1 motif confirmation.

Target genome segment	Segment ends	lab manipulation	Primer Name	Sequence 5'-3'	Length	Annealing T°	Product
RNA1	3' end	RACE	5658F	TGCGCTCCTTGCTGATGATGGCCTCA	26nt		
			546F	ATAAGGCGGCAGGGAAGGTGGGTATT	26nt		
		Confirmatory RT-PCR	7539R3	GCGGATGCGATTATTGACTTTCA	23nt	60,2	813bp
			8456F	CTGATGGATTGTCTCAGGCCATTC	24nt		
	5' End	RACE	2947R	AGCCTCTCGATCAGCAGCCTCAGTGT	26nt		
			4722R	AGGCAATCACCTCGGGCTTGAGCATC	26nt		
		Confirmatory RT-PCR	129F5	TCTGCAAAGATCCTTTCCTCTGGT	24nt	61,7	454bp
			583R5	CCGGCAGTAACATTAGCAGCAAA	23nt		
RNA2	3' end	RACE	815F	GGGCCAAGGCTCACCAGTCAAATGCA	26nt		
			855F	GACCATCGGTGGGAAGGTGACTGGAGA	27nt		
		Confirmatory RT-PCR	2388F	GCGGAGGAAGAGGGTATTGTTAT	23nt	60,9	396bp
			2784R	CCAACCACATCCTATCATCCCAT	23nt		
	5' End	RACE	2176R	GGAGCACGCCATTGGGAGTCCAGTT	25nt		
			1680R	TCTGATAGGCACTACCAGCATCACTTGA	28nt		
		Confirmatory RT-PCR	34F	CTTCTGCTCTCTGCTCTTGATCA	23nt	60,5	576bp
			610R	GTATCCAGAAACCTCCCAGTGAA	23nt		
HAM1 motif confirmation			8456F	CTGATGGATTGTCTCAGGCCATTC	24nt	65,2	664bp
			9120R	ATGCGGACTTGTTCTCCACAGTGA	24nt		

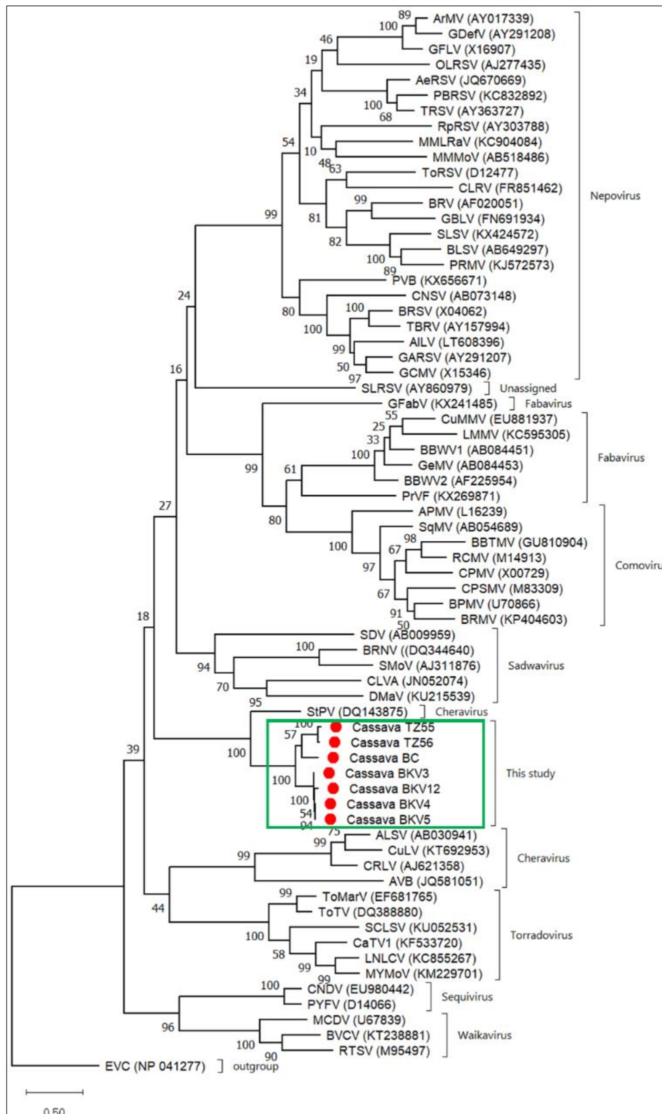
Supplementary material 2. Accession numbers and abbreviations for selected *Secoviridae* members used to deduce cleavage sites and functional domains of the RNA1 and RNA2 encoded polyproteins (respectively P1 and P2)

Species names	Genus	Abbreviation	Sequence accession numbers (GenBank)	
			RNA1	RNA2
Apple latent spherical virus	Cheravirus	ALSV	NC_003787.1	NC_003788.1
Cherry rasp leaf virus	Cheravirus	CRLV	NC_006271.1	NC_006272.1
Currant latent virus	Cheravirus	CuLV	NC_029038.1	NC_029036.1
Arracacha virus B	Cheravirus	AVB	NC_020898.1	NC_020897.1
Stocky prune virus	Cheravirus	StPV	NC_043388.1	NC_043387.1
Cowpea mosaic virus	Comovirus	CPMV	NC_003549.1	-
Arabis mosaic virus	Nepovirus	ArMV	NC_006057.1	-
Tomato ringspot virus	Nepovirus	ToRSV	NC_003840.1	-
Strawberry mottle virus	Sadwavirus	SMoV	NC_003445.1	-
Cassava Torrado-like virus	Torradovirus	CsTMV	MF449522.1	-
Black raspberry necrosis virus	Sadwavirus	BRNV	FN908128	FN908129
Chocolate lily virus A	Sadwavirus	CLVA	NC_016443	NC_016444
Dioscorea mosaic-associated virus	Sadwavirus	DMaV	KU215538	KU215539

Supplementary material 3. Lengths, number of reads integrated, percent of total reads and average coverage for the Secoviridae contigs from the various analyzed pools of cassava samples.

Origin of samples		Pools/samples	Contig name	Number of cassava fields/pool	Number of positive fields/pool	Positive individual plants/field [a]	Genome segment	Isolate name	Accession Number (GenBank)	Length of reconstructed contig [nt]	Number of reads mapped	Average genome coverage
Country	Province											
Congo DR	South-Kivu		CGO-SK3	7	5	27	RNA1[b]			7,710	57,769	1,064.6
							RNA2[b]			3,502	18,723	765.6
			CGO-SK4	8	7	33	RNA1[b]			7,610	54,876	1,038.9
							RNA2[b]			2,871	37,045	1,886.4
			CGO-SK5	7	5	31	RNA1[b]			7,718	64,395	1,098.8
							RNA2[b]			3,681	18,121	705.9
			CGO-SK12	11	6	39	RNA1[b]			7,793	58,827	1,236.6
							RNA2[b]			3,477	18,545	882.7
	Bas-Congo		CGO-BC	-	-	-	RNA1[c]			7,405	-	-
							RNA2[c]			3,460	-	-
Tanzania			TZ	-	-	-	RNA1[d]			7,637	179,438	3,007
							RNA2i[d]			3,533	59,753	1,869
							RNA2ii[d]			3,477	86,555	2,548

a = Not all individual plants could be included in the RT-PCR test as the stock of dried leaf samples run out and could no longer be retrieved through the experimental field (some stem cuttings failed to grow in the experimental field while others died during the experiment). **B** = only 3' extremity could be completed by RACE experiments. **C** = both 3' and 5' extremities were completed through RACE. **D** = not included in the RACE experiment. Contigs were named considering initial letters of the code of the country where the samples were collected-initial letters of the province of origin (when available)-and a number indicating the reference of the HTS pool (when available).



Supplementary material 4. Phylogenetic analysis of the coat protein amino acid for the five isolates detected in cassava (strong green triangles and green frame) and of type isolates from recognized species in the family *Secoviridae* (detailed information on these viruses is provided in supplementary material 5).

For each Secoviridae species, the amino acid sequence of the CP(s) was deduced from the nucleotide sequence of the corresponding genomic RNA from the type isolate. The alignment was generated using ClustalW integrated into MEGA X. The Maximum likelihood phylogenetic tree was also reconstructed in Mega X using the Poisson model with uniform distribution for amino acid sequence alignments. Bootstrap values are indicated at the main branch nodes (1000 replicates). The bar represents the number of amino acid substitutions per

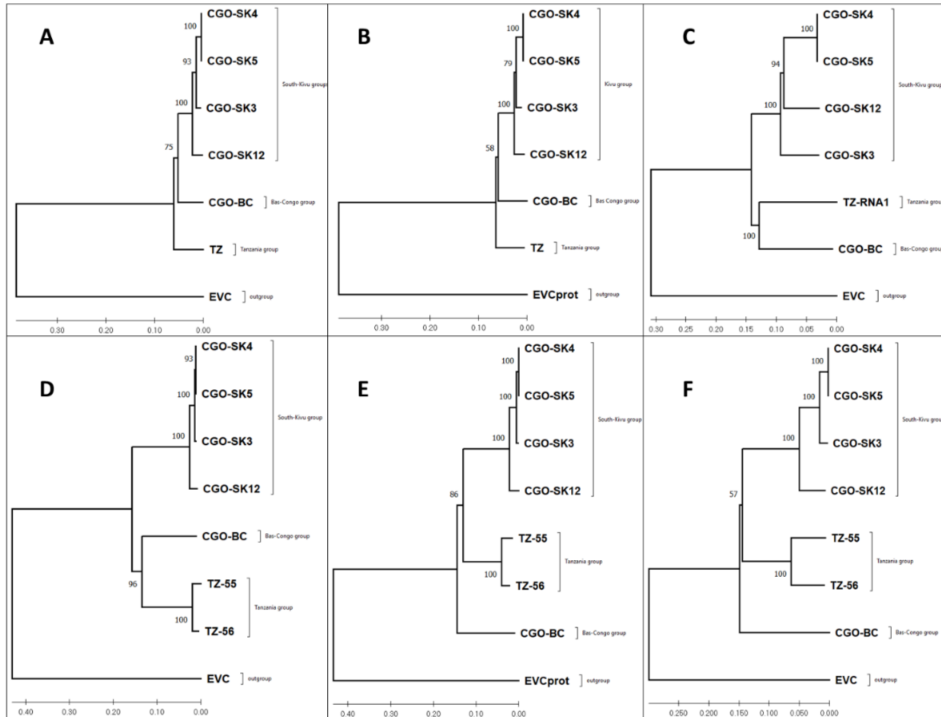
site. The tree was rooted using the combined sequence of the three CPs from poliovirus (EVC, species Enterovirus C, genus Enterovirus, family Picornaviridae).

Supplementary material 5. Complete names of selected Secoviridae members used for the phylogenetic study.

Genus	Species names	Abbreviation	Accession numbers	
			RNA1 (Pro-pol)	***RNA2 (CP)***
Cheravirus	Arracacha virus B	AVB	JQ437415	JQ581051
Cheravirus	Apple latent spherical virus	ALSV	AB030940	AB030941
Cheravirus	Cherry rasp leaf virus	CRLV	AJ621357	AJ621358
Cheravirus	Currant latent virus	CuLV	KT692952	KT692953
Cheravirus	Stocky prune virus	StPV	DQ143874	DQ143875
Comovirus	Bean pod mottle virus	BPMV	M62738	U70866
Comovirus	Cowpea mosaic virus	CPMV	X00206	X00729
Comovirus	Cowpea severe mosaic virus	CPSMV	M83830	M83309
Comovirus	Radish mosaic virus	RaMV	AB295643	-
Comovirus	Red clover mottle virus	RCMV	X64886	M14913
Comovirus	squash mosaic virus	SqMV	AB054688	AB054689
Comovirus	Broad bean true mosaic virus	BBTMV	GU810903	GU810904
Comovirus	bean rugose mosaic virus	BRMV	-	KP404603
Comovirus	Andean potato mottle virus	APMV	-	L16239
Fabavirus	Broad bean wilt virus 1	BBWV1	AB084450	AB084451
Fabavirus	Broad bean wilt virus 2	BBWV2	AF225953	AF225954
Fabavirus	Cucurbit mild mosaic virus	CuMMV	EU881936	EU881937
Fabavirus	Grapevine fabavirus	GFabV	KX241482	KX241485
Fabavirus	Prunus virus F	PrVF	KX269865	KX269871
Fabavirus	lamium mild mosaic virus	LMMV	KC595304	KC595305
Fabavirus	Gentian mosaic virus	GeMV	AB084452	AB084453
Nepovirus	Aeonium ringspot virus	AeRSV	JX304792	JQ670669

Nepovirus	Arabis mosaic virus	ArMV	AY303786	AY017339
Nepovirus	Artichoke yellow ringspot virus	AYRSV	AM087671	-
Nepovirus	blueberry latent spherical virus	BLSV	AB649296	AB649297
Nepovirus	Beet ringspot virus	BRSV	D00322	X04062
Nepovirus	blackcurrant reversion virus	BRV	AF368272	AF020051
Nepovirus	cherry leaf roll virus	CLRV	FR851461	FR851462
Nepovirus	cycas necrotic stunt virus	CNSV	AB073147	AB073148
Nepovirus	Grapevine Bulgarian latent virus	GBLV	FN691934	FN691935
Nepovirus	Grapevine chrome mosaic virus	GCMV	X15346	X15163
Nepovirus	Grapevine fanleaf virus	GFLV	D00915	X16907
Nepovirus	Melon mild mottle virus	MMMoV	AB518485	AB518486
Nepovirus	Mulberry mosaic leaf roll-associated virus	MMLRaV	KC904083	KC904084
Nepovirus	Peach rosette mosaic virus	PRMV	KY646466	KJ572573
Nepovirus	Raspberry ringspot virus	RpRSV	AY303787	AY303788
Nepovirus	tomato black ring virus	TBRV	AY157993	AY157994
Nepovirus	Tomato ringspot virus	ToRSV	L19655	D12477
Nepovirus	Potato black ringspot nepovirus	PBRSV	KC832887	KC832892
Nepovirus	Potato virus B	PVB	KX656670	KX656671
Nepovirus	Grapevine anatolian ringspot virus	GARSV	HE774604	AY291207
Nepovirus	Artichoke Italian latent virus	AILV	LT608395	LT608396
Nepovirus	Soybean latent spherical virus	SLSV	KX424571	KX424572
Nepovirus	Grapevine deformation virus	GDefV	HE613269	AY291208
Nepovirus	Tobacco ringspot virus	TRSV	U50869	AY363727
Nepovirus	Olive latent ringspot virus	OLRSV	-	AJ277435
Sadwavirus	Black raspberry necrosis virus	BRNV	DQ344639	DQ344640
Sadwavirus	Chocolate lily virus A	CLVA	JN052073	JN052074
Sadwavirus	Dioscorea mosaic-associated virus	DMaV	KU215538	KU215539
Sadwavirus	satsuma dwarf virus	SDV	AB009958	AB009959

Sadwavirus	Strawberry mottle virus	SMoV	AJ311875	AJ311876
Sequivirus	Carrot necrotic dieback virus	CNDV	EU980442	EU980442
Sequivirus	Parsnip yellow fleck virus	PYFV	D14066	D14066
Torradovirus	Carrot torradovirus 1	CaTV1	KF533719	KF533720
Torradovirus	Lettuce necrotic leaf curl virus	LNLCV	KC855266	KC855267
Torradovirus	Motherwort yellow mottle virus	MYMoV	KM229700	KM229701
Torradovirus	squash chlorotic leaf spot virus	SCLSV	KU052530	KU052531
Torradovirus	Tomato marchitez virus	ToMarV	EF681764	EF681765
Torradovirus	tomato torrado virus	ToTV	DQ388879	DQ388880
Unassigned	Strawberry latent ringspot virus	SLRSV	AY860978	AY860979
Waikavirus	bellflower vein chlorosis virus	BVCV	KT238881	KT238881
Waikavirus	maize chlorotic dwarf virus	MCDV	U67839	U67839
Waikavirus	Rice tungro spherical virus	RTSV	M95497	M95497
Enterovirus	Enterovirus C	EVC	NP_041277	



Supplementary material 6. Phylogenetic trees reconstructed using **(A)** the amino acid sequences of the "Pro-Pol" region, **(B)** the amino acid sequences of the entire P1 polyprotein, **(C)** the complete/nearly nucleotide sequence of the RNA1 segment **(D)** the amino acid sequences of the coat protein block, **(E)** the amino acid sequences of the entire P2 polyprotein and **(F)** the nearly/complete nucleotide sequence of the RNA2 segment for the new *Secoviridae* members described in this study.

The alignment was generated using ClustalW integrated into MEGA X using the Poisson model for amino acids and the GTR+GI model for nucleotide sequences. The neighbour-joining method was used for phylogenetic reconstruction in Mega X. Bootstrap values are indicated at the main branch nodes (1000 replicates). The bar represents the number of amino acid substitutions per site.

```
ArMV_subgroup_A_nepovirus     NIELAELPTDSTTCGQFN---DWASAAKKMAKGVGSIVGDFARM----------SGAGVL  447
ToRSV_subgroup_C_nepovirus    SSHAAQGFGNFLSRGKSA--------AINLASGLSSFVGEKVVS----------GANHVV  459
CsTVL_Torradovirus            VF---NATMSKFTKGV-------------------ETVESDVLP-------SLEGAASSF  86
SMoV_sadwavirus               QG--------GTSVGG------IQEAITSRIKRSFGWVGKAVESVN-----NMASATNLL  186
CPMV_comovirus                V------EGDAVAQGVSQLLYKMVTWVPTFVRGAVDWSVDAIL--------------VSF  188
CuLV_(Cheravirus)             KS---QPNWEFLDSF----------YEGKRGEGPWQFVKDKVRC-------AVDGTYRAF  452
ALSV_Cheravirus               RD---SKYWDAFIALR---------TNTKEGQGPIDWVKAGYRK-------CKEYYDNTF  375
CRLV_Cheravirus               RA---DKAWDAYIELN---------KNCKVGEGPLEYLKAAKDR-------CVEFF-SPF  552
AVB_(Cheravirus)              -----QNQRFAFTCGN---------IDRYVGEAGFDWGITRLKKFAASWGLTSDTIAQGA  452
StPV_(Cheravirus)             ------------------------------------------------------------  0
P12_cheravirus                NFESTLASDKDALVGV---------GDGMSLMTPLSWLIDACK---------MSWASNIY  556
P3_cheravirus                 NKESTLTTDKDFLVGV---------GDGLSLMTPLSWLVDACK---------MTWASNVY  588
P4_cheravirus                 NKESTLAADKDLLVGV---------GDGLSLMTPLSWLMDACK---------MSWASNIY  537
```

Supplementary material 7. Multiple amino acid sequence alignment of the region between the X1 and the X2 domains shows putative cleavage sites delineating the X1 and X2

domains (highlighted in yellow for viruses identified in this study and red for other cheraviruses previously characterized).

```
                                            Protease catalytic Cys     Protease substrate-binding pocket
ToRSV_subgroup_C_nepovirus     ---NYSEGGDYSNDLPTSIISEYVNSPEDCGALLVAHLEG-----------GYKIIGMHV      1452
ArMV_subgroup_A_nepovirus      ---DEGGSAVYQNKIRRYIIYAHEAKRNDCGAIAVAEIQR-----------TPKVLAMLV      1438
AVA_Subgroup_A_Nepovirus       ---RTNGGYTYERILNKFIRVDGLAQDDDCGTLVATLIGG-----------QPRIVGMLV      1470
CsTLV_Torradovirus             GPIKYEGADGFIFSSTHSLRIRHTGMNGECGSVLFAPNLE---------NKQPVIVGIQC      1083
SMoV_sadwavirus                -------KGRVVWQANNLLAAPLYHQVGHCGRLLLARDEA----------KCLKIVGIHV      1192
CPMV_comovirus                 ---------NYVNKVSRYLEYEAPTIPEDCGSLVIAHIGG-----------KHKIVGVHV      1132
AVB_Cheravirus                 ---------GL-LPLKKAYKYEMVTSPGFCGDVLLQVCSS-----------GVKILGMHT      1424
CuLV_Cheravirus                ---------GS-LRMPPCYSYTFDTYPGLCTSPLICMSG-------G----RCILLGLHV      1424
ALSV_Cheravirus                ---------GS-KPMPACYSYTFDTFAGLCTSPLISMDG-------G----RCVLLGLHV      1341
CRLV_Cheravirus                ---------GS-KQMPACYSYVFETYAGLCTSPLIAQEG-------G----RCIILGLHV      1514
StPV_Cheravirus                ------------------------MISGDCGVISFAPGGSTLEGSGVVPPKVCIM---HD      33
P_Stephan                      ---------KY-VC-SPGIGYKGHFGAGDCGVVLFSPTKV------GQPPLVCIM---HD      1427
P12_cheravirus                 ---------KY-IC-SPGIGYKGHFGAGDCGVVLFTPTKV------GQPPLVCIM---HD      1576
P3_cheravirus                  ---------KY-IC-SPGIGYKGHFGAGDCGVVLFTPTKV------GQPPLVCVM---HD      1573
P4_cheravirus                  ---------KY-IC-SPGIGYKGHFGAGDCGVVLFTPTKV------GQPPLVCVM---HD      1523
                                                            *                        :
```

Supplementary material 8. Multiple amino acid sequence alignment of the protease domain localises the conserved Histidine of the substrate binding pocket region and the protease catalytic Cysteine.

```
ArMV_subgroup_A_nepovirus  YRE------------------------------------------CFDRCVVLITYGDD        1838
ToRSV_subgroup_C_nepovirus LVN------------------------------------------NFKQEVCLIVYGDD        1875
CsTVL_Torradovirus         SLR----------------------------------SFTVDCTSSDFERLFVAVYGDD        1523
SMoV_sadwavirus            DPG-------------------------------------LVPYRVMRSHCTFSVYGDD        1615
CPMV_comovirus             QAP-------------------------------------ELMVQSFDKLIGLVTYGDD        1546
CuLV_(Cheravirus)          NDM------------------------------------DLYPLYSFKQLVSYAVYGDD        1855
ALSV_Cheravirus            VHK------------------------------------ALYPLYSFRTLVSYAVYGDD        1770
CRLV_Cheravirus            ELN------------------------------------SLYPLHSFRQMVAYATYGDD        1943
AVB_(Cheravirus)           LKS----------------------------------------HFWFTKNVAFAVYGDD        1868
StPV_(Cheravirus)          LGE------------------------------------AFVHRSIMDREVHFAVYGDD        471
P_Stephan                  LYREASIIALVEERVRYLKDFLQDDDALILQKARQIVTQPMVSKSVMDDSVTIAVYGDD        1895
P12_cheravirus             LRTEASVVALAEEKFRYLRDFSKEDDDFLRKKAFTIVTQPMVSKELMDSNVAIAVYGDD        2044
P3_cheravirus              LRTEASIVALAEERFKYLRDFSKEDEELLRKRAFTIVTQPMVSKEIMDSNVAIAVYGDD        2041
P4_cheravirus              LRTEASVVALAEERFKYLRDFSKEDEELLRSRAFTIVTQPMVSKKLMDSNVAIAVYGDD        1991
```

Supplementary material 9. Multiple amino acid sequence alignment of the conserved "CG-GDD" motif shows the insertion of 36 amino acids (highlighted in red) upstream of the "GDD" motif in the polymerase domain**.**

```
arrachacha_virus_A      EDSLNCVFVKQESDLPSAT-WVQSFGSPLNPCVARMRAAHSAGKK---LVFRG---HAPY  2141
P12_Cassava             -----QLLKKLDYNRPIIFMSADGLSQAIPPTLLCLRRIHGPGSKYERVILEKTTDSGKV  2323
P3_Cassava              -----QLLKKLDYNRPIIFMSADGLSQAIPPALLCLRRIHGPGSKYERVILEKTTDSGKV  2320
P4_Cassava              -----QLLKKLDYNRPIIFMSADGLSQAIPPTLLCLRRIHGSGSNYERVILEKTTDSGKV  2270
                              :: * : : *     .:.:.. : * :  :*  *. *.:   :::.     .

arrachacha_virus_A      MACWLAMMKFCISAGICDQDSLLALFYNLKGNKQTDLSSYFSKFDREKRMGHRILPRPIH  2201
P12_Cassava             NPYWLDII---------------------SGNETL----------------------TLN  2340
P3_Cassava              NPYWLDII---------------------SGNETL----------------------TLN  2337
P4_Cassava              NPYWLDII---------------------SGNETL----------------------TLN  2287
                           ** ::                     .**:                      ::

arrachacha_virus_A      FATTNSKKLEQSDSIF-----HGVVQVDISCRELDDQPSVKLVIGDKALKLKEKSRVFPL  2256
P12_Cassava             LVTSNETKFNDFKDIVSQIQGFDLKRVSIDLEELDNQPSCSKVMEGKLADACLKHACFPL  2400
P3_Cassava              LVTSNEKKFNDFKEIVSQTQGFDLDRVSMELEELDNQPSCSKVMEGKLADACLKHARFPI  2397
P4_Cassava              LVTSNEIKFNDFKAIVSQIQGFDLERVSMDLEELDNQPSCSKVMEGKLADACLKHARFPL  2347
                        :.*:*. *::: . *.      ..: :*.:. .***:*** . *: .*  .    *   **:

arrachacha_virus_A      LVEDTGFSILSEGEEPGARIKHVLQHEPGFWEQHKGKCAVVTSYAALYCDAKCPGHIHGK  2316
P12_Cassava             LVEDSGLY--TKTGRPGAMIKHWLKEKPSFWKAHCGEQVRIVSCVGVKCHSRCPSHIVEK  2458
P3_Cassava              MVEDSGLY--TSTGRPGAMIKHWLKEKPFFWKSHCGQQVRIVSCVGVKCHSKCPSHIVEK  2455
P4_Cassava              LVEDSGLY--TKTGRPGAMIKHWLKEKPFFWKAHCGEQVRIVSCTGVKCHPKCPSHIVEK  2405
                        :***:*:   :.  .*** *** *:.:* **: * *: . :.* ..: *. :**.**  *

arrachacha_virus_A      SSGCRIIDSQDMYRDDLHGFEKFITFDGMYADDSKLKLSKANPRLAVFDHLKRSNCFRIK  2376
P12_Cassava             TVVAT-VCQPDQYRDDLYGWEKFCLFDGKFSDKFCRTSEVLCARGKALRSVLSQTCAMGS  2517
P3_Cassava              AVVAT-VCQPDQYRDDLYGWEKFCLFDGKFSDKFSRASDSICARGKALRSILSQTCAMGS  2514
P4_Cassava              TVVAT-VCQPDQYRDDLYGWEKFCLFDGKFSDKFCRTSEVLCARGKALRSVLSQTCAMGS  2464
                        :  .  : . * *****:*:***  *** ::*.     .    *  .:  :  ..*   .

arrachacha_virus_A      --   2376
P12_Cassava             F-   2518
P3_Cassava              F*   2515
P4_Cassava              F*   2465
```

Supplementary material 10: Multiple amino acid sequence alignment of the 5' extremity of the polymerase domain of viruses reported in this study, together with the arracacha virus A. The figure shows a putative cleavage site that could delineate a putative domain downstream of the polymerase.

```
                                                       Cleavage site MP-CP1
P_Winter     SIRRRAKPRNSIDKYYQDVVEKRKDKDLHCESCSD-ETGFGDMLGEFPSGATSSIVPTPQ   465
P12          NIRRRAPPRHSIDRYYDKV------KDKHCDSCADEELGTGDMLPEIKSGQQGTSVLIAE   473
P3           NIRRKAPPRHSIDRYYDRG------KDKHCDSCADDESGTGDMLPEINSGQHGTSTLVAE   464
P4           NIRRKAPPRHSIDRYYDRS------KDKHCDSCADDESGTGDMLSEINSGQHGTSTLVAE   464
P_5          NIRRKAPPRHSIDRYYDRS------KDKHCDSCADDESGTGDMLSEINSGQHGTSTLVAE   464
             .***:* **:***:**:         ** **:**:* * * **** *: **  .: .    :
```

Supplementary material 11. Multiple amino acid sequence alignment of the RNA2-encoded polyproteins showing the insertion of 6 amino acids at the N-terminal extremity of the MP domain in the polyprotein of the P_Winter isolate. The alignment was trimmed to show the location of the insertion.

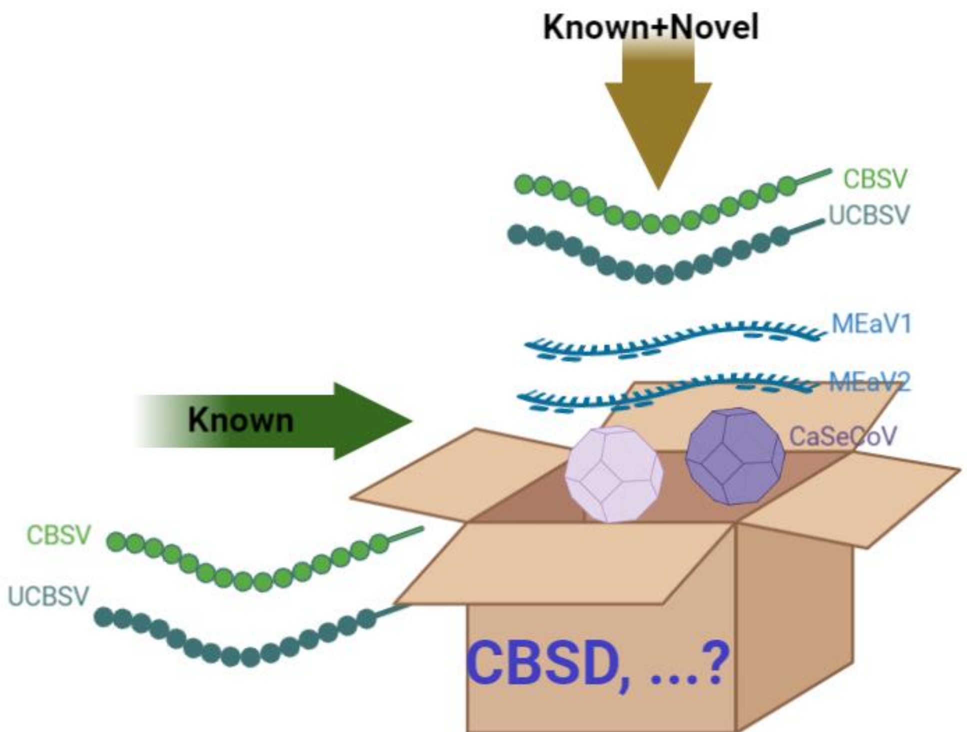
Supplementary material 12: Complete names and accession numbers for representative members of various kingdoms of life used for phylogenetic study of the HAM1 motifs

Kingdoms	Complete names and abbreviations (when available) of organisms	Accession number
Viruses	Euphorbia ringspot virus (EuRSV-PV-0902)	YP_009305422
	Cassava brown streak virus- isolate Nampula (CBSV-Nampula)	AYW01246
	Cassava brown streak virus-isolate Tan_Z (CBSV-Tan_Z)	ACT78701
	Cassava brown streak virus-isolate Ug65 (CBSV-Ug65)	QGW67508
	Cassava brown streak virus-isolate CGCC-2017 (CBSV-CGCC-2017)	QCR98745
	Ugandan cassava brown streak virus-isolate kikombe (UCBSV-Kikombe)	ARQ80023
	Ugandan cassava brown streak virus-isolate MLB3 (UCBSV-MLB3)	ACM48176
	Ugandan cassava brown streak virus-isolate Ug_23 (UCBSV-Ug_23)	CBA18486
	Ugandan cassava brown streak virus-isolate Ke_125 (UCBSV-Ke_125)	ASG92173
	Cassava Torrado-like virus-Isolate Yop12 (CsTLV- Yop12)	OK040225
Bacteria	*Escherichia coli*	1K7K_A
	Pseudomonas	WP_011064019
	Burkholderia multivorans	KHS13049
	Azoarchus olearius	CAL96580
	Rhizobium leguminosarum bv. Viciae 3841	CAK05869
Fungi	*Saccharomyces cerevisiae*	CAA89597
	Aspergillus fumigatus Af293	XP_754075
	Neurospora crassa OR74A	XP_955963
Plantae	*Arabidopsis thaliana*	NP_567410
	Oryza sativa japonica group	XP_015613001
	Manihot esculenta	XP_021594792
	Jatropha curcas	XP_012077670.1
	Hevea brasiliensis	XP_021644689.1
Animalia	*Caenorhabditis elegans*	AAL14111
	Acyrthosiphon pisum	NP_001233079
	Drosophila ananassae	EDV32196
	Culex quinquefasciatus	XP_038111262
	Tribolium castaneum	XP_974197
	Xenopus laevis	AAI10772
	Danio rerio	NP_001093456
	Mus musculus	EDL28288
	Homo sapiens	AAK21848
	Tetrahymena thermophila SB210	XP_977249

Supplementary material 13: Multiple amino acid sequence alignment showing conserved sites of ITPases from representative members of various kingdoms of life. The alignment was built using Mview (Brown, Leroy, and Sander 1998)

Chapter VI

VI. General discussions and conclusions



VI. General discussions and conclusion

After approximately 100 years of research on CBSD since its first report in Tanzania and large-scale interventions, the disease still represents one of Africa's major threats to food security. For decades, it has been confined into Eastern Africa but actually, a westward outbreak is putting the major cassava producing regions at high risk of spread, raising concerns over food security of people. Although the lack of resistant varieties has weakened the efficiency of the management strategies implemented so far, it is acknowledged that the epidemiology of this disease still needs further investigations to be fully elucidated.

Through time, investigations on the epidemiology of plant diseases have established rigorous approaches to study and to monitor the development and spread of diseases. . These approaches have evolved depending on circumstances and scientific knowledge evaluation. This has been exemplified for seed-borne diseases early in the 1900's whereby plant pathogens have been demonstrated to be transmitted through seeds, triggering the recognition of the role that seed systems plays in the epidemiology of diseases (Elmer 2001).

However, practices regarding seed selection, conservation, and importantly seed exchange in Africa are at the heart of traditional agricultural systems for millions of African farmers, and they are shaped by social and cultural values and perceptions, as previously demonstrated (van Niekerk and Wynberg 2017). We used this fact to clarify the important role that humans play in the dispersal of plant propagative material, and consequently in influencing the epidemiology of seed-borne diseases by inadvertently or deliberately dispersing infected seeds.

According to this, we first proposed a multidisciplinary approach as having the potential of studying problems related to the epidemiology of seed-borne/cutting-borne diseases in African context, more specifically in South Kivu of Democratic Republic of Congo, characterized by a pattern of social and ecological complexity. Additionally, we proposed a sampling approach capable of feeding the HTS detection technologies with high-quality and representative samples for accurately investigating the diversity of viruses infecting the cassava crop.

a) A multidisciplinary approach to capture the complexity of the CBSD epidemiology

We have proposed that analyzing the CBSD epidemiology in African contexts through interdisciplinary lens could allow capturing determinant factors necessary to efficiently support the decision making for disease mitigation in the cassava seed system.

The used framework has mobilized various disciplines related to sociology, ecology, geography, cropping systems and practices etc. It relies on various steps, each one including various aspects that required specific methods and strategies of collecting and analyzing data. Globally, the conception of this framework was made

to promote the flexibility of the boundaries between various disciplines, allowing them to be mixed and matched and facilitating their combination.

The approach allowed describing the epidemiology pattern of CBSD in the studied area and it went beyond by associating these epidemiological trends to parameters from associated disciplines. It therefore provided results that can be aggregated into three main points, namely: (i) epidemiology, (ii) clusters, and (iii) risk analysis.

i. **Epidemiology**. The viruses associated with CBSD were confirmed to be both CBSV and UCBSV as previously elucidated by (Casinga et al. 2021). We found that the molecular incidence of the CBSD associated viruses in the studied area was 31%, corresponding to 13% of UCBSV alone, 11% of CBSV alone and 7% of mixed CBSV-UCBSV infections.

The study from (Casinga et al. 2021) is the only study, before ours, that has proposed a two-years monitoring of the CBSD status in the Eastern D.R. Congo, including the area of this study (Uvira territory). Overall, it showed that in in 2016, only the UCBSV could be detected at incidence rate of 10.6% . In 2018, both CBSV and UCBSV were detected at incidence rate of 14% using other detection protocol (primers from (Rudolph R Shirima et al. 2017) instead of (W. A. Monger, Seal, Cotton, et al. 2001)). The work that this team have conducted was important in that they allowed establishing a general picture on the status of the CBSD spread in Eastern D.R. Congo, a wide region where little was known, until then, on the epidemiology of this disease. However, the approach used in that study were less informative for decision-makers to initiate locally-tailored interventions for mitigating the disease spread. Additionally, the study was less clear about how collected samples were selected for molecular detection, stating for example that in Uvira territory (the area of this study), only one field was positive out of the 9 CBSD-affected in 2016, whereas 4 fields out of 8 affected were positive in 2018. Whatever the situation, reporting on the epidemiology of a viral disease in an area of 3,146 km^2 by testing only 9 fields is a large extrapolation.

On this aspect, the approach that we used in our study, consisting in "intensively" surveying the Uvira territory on 246 fields rather than conducting an "extensive" approach on a wide region with few samples had the potential to provide accurate and representative results. In addition, our survey strategy covered most locations where cassava crop is cultivated rather than following motorable roads only as done previously (Casinga et al. 2021). It allowed unraveling that the incidence rate of the CBSD was more than the double of that reported by the previous studies, but importantly it revealed also a local spatial pattern in its epidemiology.

ii. **Clusters**: The second advantage of the approach used in this study consisted in identifying the factors related to crop management practices (farming systems, type of land tenure, presence of weeds), and notably to the pathways used to obtain cassava cuttings as influencing the epidemiology pattern of the CBSD in the surveyed area. Apart from this factor, the other ones were proposed for the first time by (Agrios 2005) under the term "human factors", as an additional component of the disease triangle. However, before that, several evidences from elsewhere have already shown that cropping practices had the potential of influencing the incidences and severities of viral diseases (Thresh 1982; 1991). Nevertheless, in Africa, the situation has been

shown to be particularly complex because of: (i) the high diversity of crops grown, (ii) the wide range of agroecological zones utilized, and (iii) the utilization of both traditional cropping practices together with modern, and at a certain level, industrial techniques (Thresh 1991).

This complexity puzzles the assessment of disease epidemiology and their effective control since measures shown to be suitable for some groups of farmers and their cropping systems could be inappropriate for others (Thresh and Fargette 2001).

A previous study that has surveyed the epidemiology of the CMD in the Eastern D.R. Congo has reported the distribution of this disease following agro-ecological stratification and assigned the Uvira territory as a high pressure zone of virus diseases (E. B. Bisimwa 2012). Additionally, that study was the first to highlight the complex nature of the CMD pathosystem, driven by the specificities of each region and has called for an adapted and integrated approach for controlling the disease. Our study shows that this ecological effect has also the potential of driving the epidemiology of the CBSD in the context of the Uvira territory. We showed that the resulting stratification of disease pressure for CBSD uncovered the existence of a low-pressure disease zone that could be privileged for seed multiplication activities.

(Agrios 2005) has conceptualized the human factors as consisting into cropping practices used by farmers. This perception have been criticized by some authors who argued that domesticated plants had already their identity intimately intertwined with husbandry, making humans already implicitly represented in the disease triangle configuration (Francl 2001). However, in our study, the concept " human factors" has been used in reference to the socio-cultural, socio-economic, geographical and socio-ecological factors that characterize human activities.

iii. **Risk analysis**: The notion of risk in the management of viral plant diseases is very important because of the lack of curative measures (Tatineni and Hein 2023). In this way, risk-reducing measures that are specific to local contexts are very important to identify. The ultimate advantage of the approach adopted in this study rely in its relative capacity of regrouping the various determinants involved in the study, to identify the prominent ones that are putatively associated with the outcome of CBSVs infection in the study area and to use them in the frame of the diseases management.

Some of the identified factors were already reported elsewhere as key in the management of various diseases and are mainly related to farmer's knowledge and awareness: knowledge of cassava pests and diseases, knowledge of management practices, assistance and support from extension services and the adoption of management strategies. Additionally to those factors already reported elsewhere, the approach allowed revealing some context-specific factors putatively acting as drivers of the CBSVs epidemiology. These factors are related to geographical location (proximity to towns, proximity to borders, and distance travelled to acquire cuttings), as well as the pathways used to acquire cuttings.

Overall, the fact that the approach proposed in this study has been capable to recall some drivers already proposed as relevant in the management of the CBSD elsewhere, together with drivers specific to the surveyed area, further demonstrated its potential in the frame of the contextualization of the CBSD epidemic survey.

b) The approach sounds great ! but there are challenges and the way is still relatively long for the approach to be validated

Although our approach was interesting for generating new knowledge and insights, a number of challenges and limitations are still present:

- The first challenge is related to the complexity of the approach as the assessment with multiple features from different categories generated too much factors that could not be easily interpreted . The consequences of this fact precludes that the results presented will need additional experiments and investigations to be fully elucidated and validated. For instance, we showed that cutting pathways would be determinant in predicting the outcome of CBSVs infections. However, we could not measure the additive effect resulting from the fact that multiple cutting sources are used at the same time by farmers. Additionally, it appeared also that the motivations that supports the decision for selecting pathways for acquiring cuttings, not surveyed, would help explaining the diversity or variability of cutting sourcing practices
- Another challenge comes from the fact that, basically, the multidisciplinary approach requires the constitution of a team involving specialists from each of the disciplines covered by the research. It also uses a set of methods and techniques of data collection and analysis that are specific for each discipline but definitely "combine or integrate" the various results obtained into a whole. (Wilson 1998) referred this to what he called the "concilience" defined as "linking facts and theories across disciplines by using science methods to create a common groundwork of explanation".

Due to various reasons, principally the limited resources and time, the various study aspects could not be sufficiently investigated because the team in charge of the study was not sufficiently diversified. Consequently, the data collection and analysis could have been improved. For instance, the risk computation suggested that a number of parameters could be associated to the outcome of CBSVs infection. However, for most of them, we could not clearly demonstrate that these associations were truly causative, neutral or simply spurious. Therefore, we strongly call for social experiments to validate/verify the conclusions drawn at this point.

Two particular cases support this observation: first we showed that farmers had to travel more than 10 Km for accessing less risky cuttings. However, it could have been more informative if these various destinations would also have been elucidated and mapped so that interactions among various actors, the movement of transactions as well as their flux could have been better characterized. Network analyzes would be useful for this step as shown elsewhere (Garrett et al. 2018; Andersen et al. 2019; Buddenhagen et al. 2017). Additionally, in the perspective of establishing a sustainable seed system of cassava, cuttings multiplication activities should be economically viable and attractive for business. However the state of the demand and offer of cuttings in the area, together with the attractiveness level of this sector of activities would be very important to investigate following economical and profitability assessment approaches.

Definitely, as in this study we have tested this approach in a pilot area, the important question that deserve now be discussed consists in clarifying "how this approach can be transposed to other regions that differ in socio-cultural

dimensions" ?. In the previous lines we have discussed how important the knowledge of the area to be studied is for the approach to be successful. So, the key aspect that deserve to be focused on when it comes to adapt this framework to other regions is related to a good description and understanding of the local socio-economic and ecologic dynamics. This include the use of methodological strategies that allows a close collaboration with key local actors and experts, observations and literature reviews. This is the main focus of the first and the second steps mentioned in the framework (the initiation and the methodology definition stages)

c) A CBSD epidemiology feature in the studied area: field diagnosis and molecular detection suggests disparate results. ***Stop privileging symptoms during field assessment of CBSD***

Generally, it is known that plant disease diagnosis using symptoms has strong limitations due to the lack of specificity. For instance, the possibility for disease symptoms to be confused with plant reaction to various environmental stress, the possibility for one pathogen to trigger various symptom types on the same host or alternatively, the possibility for various pathogens to triggers symptoms of the same characteristics on hosts, the problem related to the length of the incubation phase (between the infection and the apparition of symptoms), tolerant plants etc.

All these cases have been well documented for plant viruses, and it has been especially shown that the CBSD had some of these features. In fact, from its very first description, the disease was reported to be characterized by the variability of symptoms according to cassava varieties, plant age, environmental factors and types of virus species infecting the cassava representative of the disease triangle. Managing plant diseases that have such kinds of pattern require detecting the presence of the causal agent, and therefore discriminating infecting plants from healthy ones. In that way, virus detection is a cornerstone step in the plant virus management.

The context that has been established in the first chapter of this thesis is important because it depict the magnitude of the problem posed in the diagnosis of CBSD and the detection of its associated viruses. The issue is related to the fact that we are dealing with a pathogen that spreads through planting material within an environment characterized by important variability of agro-ecological conditions and a complexity of socio-cultural and socio-economic conditions. These factors complicate the epidemiology.

At this step, the question that deserves to be asked is why the CBSD symptoms diagnosis on field continues to be emphasized, sometimes too much as shown in D.R. Congo, in the epidemiology assessment ? It is true that at a certain level, information collected using the symptoms diagnosis are important to serve as a proxy of the presence of a disease. This is only relevant when we are sure that the pathogen assessed is the only biotic or abiotic stress of the crop causing these symptoms. Meanwhile, when it comes for field assessment of crops for which certainty on the exclusive presence of a pathogen is unknown, caution is required for interpreting the resulting data. This is exemplified by the situation that happened in Zanzibar in 1994 where surveys concluded that the CBSD was absent from the island (J. P. Legg and

Raya 1998) although it has been observed there in the 1950s, and subsequently found prevalent in 1998 when the island was surveyed again (Thresh and Mbwana 1998).

We have shown that the phytosanitary standards that have been in force for a longtime in D.R. Congo for the management of CMD have privileged symptoms diagnosis, in part due to their much more specificity. It was not fair to do the same for the CBSD diagnostic considering the issues related to the lack of analytical specificity (inclusivity and exclusivity) during symptoms diagnosis. currenty, the symptoms diagnosis of the CBSD via artificial intelligence-assisted imagery methods are emerging. These methods can facilitate the diagnosis because even not – or less trained people cans easily run the test. In such a way, these methods could reduce the human-induced bias and could have the potential of improving the symptoms diagnosis once their algorithms will be sufficiently trained on the recognition of the various symptoms morphologies of the CBSD. However this could happen only if progress are made in the epidemiological characterization of this disease to accurately differentiate which symptoms are related to CBSD and which one are not. Additionally and definitely, these methods will still including the bias due to the delay in symptoms appearance on cassava plants. Their application for CBSD diagnosis for farmers, seed multipliers or sellers could be much appreciated. However for seed certification or research purposes (SENASEM, INRA, IITA etc.) this approach should be done with precaution by always combining it with a molecular detection unit or HTS.

Among the detection methods that have been extensively used for CBSVs, and now it stills, figures the molecular methods including the RT-PCR and qRT-PCR due to their highest analytical sensitivity. However, their efficiency is highly dependent on the inclusivity and exclusivity of the primers used, both criteria determined by the representativity of viral sequences included in their design.

For years, especially during the first interventions for mitigating the spread of the CBSD after its re-emergence (between 2000's and 2010's), 95 complete and partial genome sequences of CBSVs, corresponding to 86 CBSV and 9 UCBSV, were available. Most probably, these sequences were not sufficiently representative to allow designing inclusive primers at that time. However, since then, much more researches on the CBSD have allowed to sequence much more genomes (complete or partial) of the U/CBSV, to reach approximately 640 sequences with 373 orf CBSV and 267 of UCBSV from diverse locations where the CBSD is endemic. However, even with this consistent number of sequences that basically could now allow designing more inclusive primers, the detection of U/CBSVs is still challenging as elucidated in the first manuscript of the thesis (Table V-1), and by other studies elsewhere.

In fact, field diagnosis has identified symptoms in 116 of the 246 surveyed fields, whereas the RT-PCR has confirmed the U/CBSV in only 76 fields of the 246, including fields in which no symptoms was identified during field assessments. Here again we showed the additional importance to have identified CBSD patterns through clusters because the magnitude of the gap between results from the field symptoms diagnosis and the molecular diagnosis are not the same across clusters. In fact, as

shown within the Table V-1, in the cluster 1 the molecular and symptoms incidences have closer rates (26 and 25 % respectively), whereas in clusters 2 and 3 these parameters become more distant (29 and 42 % in cluster 2; 38 and 74 % in cluster 3). Such results confirmed the observations made a decade before (G. Rwegasira and Rey 2012) showing that symptoms-based diagnosis for CBSV infections was unreliable as some symptomatic plants were free from CBSVs and other symptomless plants were CBSVs-infected.

In any case, at this stage, the increased number of symptomatic plants that could not be confirmed by molecular indexing was pivotal to investigate. We have discussed the various possibilities that could explain such situation including the non-specificity of CBSD symptoms, the type of samples used (leaf rather than roots), the non-inclusiveness of primers used (i.e. the existence of divergent strains at primer binding sites), and the existence of other viral complexes.

Even if it was evident that using leave samples rather than samples from roots could somehow induce some false negative due to possible recovery of some infected plants, we did not envisaged this in the explanation of the magnitude of the problem since other studies that reported similar discrepancies have used root samples for molecular detection. Despite the primers used were designed using various representative sequences of both CBSV and UCBSV, still we retained the possibility for unknown strains and/or viral complexes to occur.

Globally, more than 10 years after the demonstration of the Koch postulate for the CBSD, the link between the presence of causing viruses and the characteristic symptoms of the disease they cause remain vague despite remarkable technological progress.

d) High throughput sequencing, data mining of publicly available RNAseq data, and bioinformatics as tools for investigating the diversity of viruses infecting cassava crop.

Since the emergence and the revolution of high throughput sequencing technologies, a decade from now, special interest was dedicated on their use to allow investigating the shadows hanging over the diversity of viruses infecting the cassava crop and their presence in symptomatic and healthy samples. The HTS technologies have been already used for the cassava crop and have allowed very important achievements ranging from the reconstruction of the first full genome of cassava (Prochnik et al. 2012), the genome sequencing of a number of viruses infecting cassava as well as the study of their evolutionary pattern (Ndunguru et al. 2015; Titus Alicai et al. 2016b; Manani et al. 2017; Amisse et al. 2019). However, before the research conducted through this thesis, there was no study that have reported a previously unknown viral species infecting the cassava plant by the HTS technologies.

For this thesis, the decision of using the HTS to address the problems described above has delivered 6 major results:

1. The reconstruction of nearly full genomes of CBSV and UCBSV occurring in the study area: the reconstructed genomes were 98% and 99% identical respectively

to CBSV genomes (MZ486428.1) and UCBSV genomes (MZ486425.1) already sequenced in 2017 and 2021 in the same area. We additionally took advantage of mapping the primers used for the molecular screening against these reconstructed genomes and no primers mismatches could be noticed. However, some cassava plants that were negatively indexed during the molecular test were shown to have viral reads related to one or both viruses and subsequently re-confirmed using additional RT-PCR. Therefore, the possibility for a distinct strain of CBSV or UCBSV occurring in the area was definitely dropped out. At this stage, it was important to understand the origin of the false negative detection initially obtained. The possible explanation of this situation can be related to a period during which the virus copy number is lower enough to show symptom or to be detected by RT-PCR and/or is unevenly distributed in the plant, the virus being absent of the sampled tissue while progressing in the plant. This possible "cryptic phase" would consequently lead to the absence of detection and the absence of symptoms and would therefore allow infected plants to escape both from visual symptoms diagnosis and from molecular detection. This situation has also been mentioned elsewhere (Winter, 2022. Personal communication) and (R.R. Shirima et al. 2022; G. Rwegasira and Rey 2012). The existence of a putative cryptic phase can seriously hamper the distribution of disease-free planting material as cuttings can be collected from a plant that have been tested negative, disseminated to farmers but later become diseased. A special care should be dedicated on investigated this hypothesis because such CBSD patterns could make it complicated the mitigation of this disease within cassava seed systems. If it comes to be demonstrated, the application of a more powerful detection method such as the HTS for indexing cassava plants before cuttings are collected and distributed to farmers would be necessary.

2. Importantly, we can validate the hypothesis drawn on the occurrence of previously unknown viral species infecting cassava crop including:
    i. Several isolates representing two species of ampeloviruses, called MEaAV-1 and MEaV-2, from the family closteroviridae
    ii. Several isolates of the cassava-Congo cheravirus, from the family Secoviridae:
3. A geographical distribution that goes beyond that of the CBSVs, but globally remains not fully elucidated. Even though these new viral species were associated with known viruses of cassava, two patterns can be observed at this level: the ampeloviruses have been detected in the region spanning from the Eastern part of the D.R. Congo, via Rwanda and Tanzania, until the Madagascar island in the Indian Ocean where both CMD and CBSD are present. Whereas the CaCoc has been detected in the region spanning between the western part of the D.R. Congo (where only the CMD and the CRND are spread) via the eastern part of the D.R. Congo and Tanzania where again CMD and CBSD are known to occur.

We therefore have two zones that overlaps in the eastern part of the D.R. Congo but the geographical spread of these news viruses is not known with certainty. Additionally, it is impossible at this level to know if these novel viruses were previously presents in each of these area but were unnoticed, or if they rather have spread there as a result of recent introductions (from which place and from which host?). Testing herbarium samples collected from these areas could allow shedding light on this question as done for other viruses (A. Fox et al. 2022).

4. These novel agents (MEaV-1 and CaCoc) were already present in sequencing data generated from sample collected in Tanzania and Malawi in 2015: As already discussed elsewhere, the HTS generate a large amount of information, of which only a small proportion is known/exploited, leaving a significant amount of information as a “dark matter”. Among other causes, there is the fact that researchers were focused on the plant and its transcriptome without analyzing virus presence, but also the difficulty for non-virologists to screen sequencing reads for viruses that have not yet been characterized. Currently, the sequencing costs are becoming more affordable, making the HTS technology being used by more and more studies across the world and generating an exponentially growing quantity of data that can be freely accessed.

The homology to the database of known viruses through different version of BLAST (**BLASTn**, **BLASTx**, **tBLASTx**) has been the mostly used procedure to identify the sequences of viral origin from the *de novo* contigs. The limit of this procedure lies in the fact that the sequences of an unknown virus (not listed in the database, with different or distant features) might be missed from the large batch of information and homologies obtained from the numerous contigs. Actually, advances in the computing knowledge has brought out novelty both in the deep understanding of the taxonomy of viruses as well as in the prediction of the genomic features from viral origin in RNAseq data. Novel methods that use machine learning approaches (**VirFinder**, etc.), and better, deep learning convolutional neural network approaches (**DeepVirFinder**, **VirHunter**) are becoming widely used and more performant to detect novel and very divergent viruses in assemblies of RNAseq data. Additionally, the previous experience gained from the pandemic of the SARS-CoV-2 virus have triggered in the scientific community the need for characterizing the planetary diversity of viruses in the purpose of establishing a preparedness for mitigating future pandemics. In that way, platforms like **Serratus** (https://serratus.io/) (Edgar et al. 2022) (Edgar et al., 2022), RVMT (www.riboviria.org) (Neri et al. 2022) and ViroidDB (www.viroids.org) or **Galaxy** (https://usegalaxy.eu/) are becoming essential in that they offers facilities for freely accessing and/or analyzing DNA and RNAseq data from millions of biologically diverse samples across the world.

We took advantage of these modern tools to reanalyze RNAseq data generated from the cassava crop using the palmprint-barcode from the RdRps of the ampeloviruses and the cheraviruses identified for the first time in this study. This allowed to broaden the geographical spread of these viruses since we reconstructed their nearly full genomes from the data generated in 2015 in Tanzania as highlighted.

5. Taxonomy evolution and genomic similarities :

a. The revisitation of the known genomic organization of viruses from the genus cheravirus in the family *Secoviridae* : the prediction of cleavage sites within the RNA1-encoded polyprotein of viruses reported here have allowed to show a feature that existed within polyproteins of other known cheraviruses but that was not investigated until then. As experimental evidences are still awaited, the alignment-based evidence predicted with good confidence the existence of a cleavage site upstream of the NTB, a feature shared with nepovirus and sadwavirus that delineate the X1 and X2 proteins.

b. A putative first link between viruses infecting cassava in America and Africa, at the family level, and that share a common genomic pattern that seems to be specific

for viruses infecting cassava , including the CBSVs. Until recently, it was known that none of the viruses infecting cassava in Africa were known to infect it in its geographical origin (America), and therefore the question of their origin was of concern for the scientific community. Additionally, there has been these other information related to the infection mechanisms and virulence of the CBSVs that took too much time to be elucidated: the pathogenicity, replication and movement as well as the resistance behavior (Sheat and Winter 2022).

The CaCoc is the first *Secoviridae* that infects cassava in Africa. However, other *Secoviridae* have been reported to infect this crop in America including the Cassava torrado-like virus and the cassava American latent virus from the torradovirus and the nepovirus genus (Leiva et al. 2022b).

Especially, the CaCoc has shown a close similarity with the CsTLV in that they both encoded a Maf/HAM1 protein, only recently characterized for its roles in the infection mechanism as well as a virulence determinant in the disease causation. Before the CsTLV, no virus infecting cassava in the Latin America was known to have such HAM1 motif, that was thought to be restricted to the CBSVs that occur only in Africa. So, as now a virus from the natural origin of cassava has also this motif, perhaps it could open up new research avenues regarding the origin of viruses infecting cassava in Africa.

6. The biological characteristics of these viruses are still unclear and consequently their biological risk is not known:

a. For the ampelovirus: we have discussed the challenge resulting from the fact that in some contexts, symptoms could be observed on some plants, but not on others (subsequently disappeared, or only absent). Two additional challenges going in that way are due to the fact that these ampeloviruses were sometimes found co-infected with CBSVs or CMVs, and most of the times they were accompanied with defective molecules for which the role in the infection process is unknown. With such patterns and lack of biological information, it is impossible to determine the significance of the biological risk they really pose.

b. For the cassava-Congo cheravirus: we have reported that symptoms were of milder severity but these observations are from only one isolate and on only one cassava variety. There is no additional information.

Additionally, these new viruses are spread in areas where CMD and CBSD are endemic and as illustrated for some cases in this study, mixed infections can occur. Indeed, co-infection between MEaV-1 and both CBSV and UCBSV, MEaV-2 and EACMV, and the CaCoc with both EACMV, CBSV and UCBSV have been described. The impact of the occurrence of these virus mixtures on symptomatology deserves to be investigated carefully so that putative effects resulting from their interactions could be elucidated. Also, the few observations that have been made regarding the symptoms, deals with the aerial part of the plant, mainly leaves. Nothing is known about the effects in the below-ground part where the most economically important effect need to be characterized.

e) Conclusion and future prospects

In the last 25 years, CBSD has become a major cause of food insecurity across East Africa and a part of Central Africa. This disease continues to pose a growing threat

since it is moving westward and could affect the world's largest cassava production basin in West Africa. This growing threat reflects the limits of previous interventions for mitigating the disease spread, limits than can be due to various causes including (-but not restricted to): (i) the disease investigation approaches that did not delivered key information on "how context-specific interventions" should be implemented to take into account the "complexity" of the African context; (ii) the use of epidemiological information based on the diagnosis of misleading symptoms that have created confusion during field assessment of the disease (Zeyimo et al. 2020; J. P. Legg and Raya 1998; Thresh and Mbwana 1998; Thresh and Fargette 2001).

Considering this situation, this thesis has investigated the epidemiology of the CBSD following an integrated and multidisciplinary approach and has underlined its potential benefits in providing locally-adapted solutions for mitigating its spread. However further studies are required to fine-tuning the framework proposed and validating it for its routine application. Such studies are important to undertake and require close collaboration of plant pathologists with agronomists, socioeconomics, plant breeders, psychologist etc. so that accurate information could be captured and analyzed to allow decision makers design interventions strategies that are efficient in developing stable and resilient seed systems.

The multidisciplinary approach involving not only plant virologists, but also other specialists as proposed above, has already shown its merits in studying other important African diseases including cocoa swollen shoot, maize streak and groundnut rosette (Naidu et al. 1998; Rose 1978; Thresh 1988), but for restricted periods of studies. However, there are challenges in assembling and sustaining multidisciplinary teams as they require substantial commitment of funds over a prolonged period in order to overcome complexities of field experimentation and surveys as well as the need for studies in different agroecological areas and in contrasting seasons to take into account climatic variations and other variables (Thresh and Fargette 2001).

National programs seldom in D.R. Congo, and by far in many of the African countries, does not have resources required to undertake such studies because of a limited support of research programs by national and international institutions. So, most of programs directed on the mitigation of diseases have been limited, of sporadic duration as they were mostly funded by outside donors and have involved international agricultural research centers.

In such context, with some few exceptions, the main emphasis in plant virology has been on diagnosis (virus characterization and etiology) and on supporting the resistance breeding projects (biochemical and biotechnological aspects). It is easier for African scientists to collaborate with those in Europe and North America on these topics than on field-based studies. A major change of attitude is required if progress is to be made in solving the cassava disease problems that are becoming complex for mitigation in seed system of this crop.

Results obtained in this study in relation with the CBSD suggest there could be a risk that inadequate information on the prevalence and distribution of this disease exists as in most cases, this disease has been longtime assessed by privileging

symptoms that are less specific and misleading, despite the demonstrated Koch postulate several years ago. Additionally, as now novel viral complexes are shown to occur in cassava and for the most in association with CMBVs and CBSVs, we cannot exclude the possibility for them to interfere with the infection process, virulence determinants, transmission aptitudes, and definitely the symptomatology of the viruses already known to infect cassava. This is further supported by the fact that most of these functions have been very recently predicted to involve the HAM1 protein of the CBSVs for which a homolog has been just identified to be encoded within the genome of the CaCoc reported herein.

As now the virome of cassava is better described, it is becoming important to investigate the complex interactions between the novel agents reported here with the already known viruses, their associated vectors (whiteflies for CMGs and CBSVs as experimentally validated. Mealybugs and various homopters for the ne ampeloviruses and nematodes for the new cheravirus are prime candidates but these hypotheses need experimental confirmation) the cassava cultivars and environmental conditions need to be investigated. Also, the influence of these interactions on the severity and spread of CBSD should be evaluated.

Pending the determination of their biological characteristics for risk evaluation, it is important to already include the newly discovered viruses within regular pest surveillance activities to track their spread and periodic changes. Institutions in charge of surveillance of regulated pests need also to be alerted so that already, decisions might be envisaged to restrict their spread through seeds. Criteria to be investigated for a proper biological characterization and risk evaluation have been proposed (Massart et al. 2017). Interestingly, only some key steps of this process can be privileged (Fontdevila et al., submitted) when it comes to work into resource-limited contexts or remote areas such as some African countries. For instance, data gaps that would deserve to be filled for the viruses discovered in this study are related to the transmission mode, symptom causality, host range, severity, incidence and an elaborated geographic distribution. This information could be generated through greenhouse assays and field surveys requiring significant resources.

Very little is known about the viral populations from cassava plant in other ecosystems in D.R. Congo, within other agriculturally important crops, as well as within wild hosts plants. It is important to envisage investigating them since they can potentially serve as important sources of viral inoculum and give raise to unpredicted emergence of viral diseases. This could shed light on viral evolution and the contribution of these various wild hosts to the global pathosystem of cassava. It may also help to identify potential unknown viral diseases, against which pre-emptive control could be taken in anticipation of emerging diseases (Newbery, Qi, and Fitt 2016). These various approaches that are actually accessible and used in this study, starting to the use the HTS technology and various advanced bioinformatic tools would facilitate these investigations.

The outcome of virus infection into plants does not necessarily results in symptoms or yield loss. In a long or midterm, the fitness of a virus population can impact the survival of the host plant and putatively trigger its degeneration (Gibson and Kreuze

2015). This is critical in the context of sustaining the cassava seed system against the damage of viral diseases since the crop is vegetatively-propagated. We would gain more in adopting an ecological perspective when it comes to investigate the diversity of viruses or when it's about assessing the viruses' impact or risks. The ecological perspective privilege an assessment in term of fitness in a mid or long term. This could help understanding the real biological impact that the growing diversity of cassava viruses could have. This avenue has not already been explored.

The COVID pandemic illustrated that international cooperation is essential to achieve effective surveillance and interventions that are fit for purpose in diverse geographic locations. We have shown in this study that through collaboration with other scientific researchers, the spread of the novel viruses reported in this study has been broadened. The merits of this lies in the fact that a rapid idea on the spread of these novel agents will allow an early planning for their surveillance as well as deeper investigations on their biological significance

There is a great need for improving or updating the D.R. Congo's seed certification standards not only to make it more efficient in accurately diagnosing the CBSD, but also to improve its fields inspections activities in the studied zone. This study is one of the multitude examples that demonstrate the importance of adopting the HTS technology for virus testing in seed systems. The adoption of this technology in Africa, specifically for addressing the complicated epidemiology of the CBSD or to ensure clean starting material before first field multiplications could be an important asset in sustaining a clean cassava seed system. However, considering some characteristics of this technology (price, reagents, maintenance, IT equipment and technical skills), an important question to address consists in its operationalization for seed testing in the African context. However, definitely, initiatives needs be undertaken by local stakeholders (universities or research centers in D.R. Congo) to adopt these facilities considering the various opportunities offered by some sequencing platforms and the advantages offered by various cloud computing facilities for bioinformatic analyzes. The experience

In conclusion, considering the various results obtained from this study, the following summary of recommendations can be suggested for an efficient management of the CBSD and for raising the preparedness of the cassava seed system for future pandemics:

- Promotion of exclusion strategies through improved communication and sensibilization of stakeholders involved in the seed production, certification and quarantine : the main actors involved in the regulation of the seed sector needs to be informed on the challenges and the need to upgrade and strengthen the seed production and certification standards to make them efficient in the diagnosis of viruses and to facilitate the accessibility of farmers to certified and clean planting material. Raising awareness of quarantine and agricultural inspectors on the importance of a strict implementation of quarantine and certification measures to avoid the movement of infected material.

- Improved strategy for surveillance of viruses actually known as infecting cassava within the entire seed production system: from the establishment of novel varieties, through their various stages of multiplication, until the deployment in various regions for their use by farmers; the seed system grows in scale. And in the same way, the risk for virus build-up in planting material increases (not at the same extent but according to the CBSD prevalence in each of the areas). A trade-off in the testing strategies to be used at each step must be considered in term of technology and sampling strategies :
    - In term of technology: at the G1 level, the HTS could be more indicated. From G1 to G4, PCR, LAMP or ELISA could be suitable. Improved methods of symptoms recognition could be used at G4 and by farmers.
    - In term of sampling strategy: It could be important to consider the importance of an ecological approach that extended investigations on other crops and ecosystems around the cassava crop.
- Preparation of an efficient strategy to deploy and distribute CBSD-resistant varieties that should be available very soon : special care must be dedicated to this deployment to avoid the repetition of the previous mistakes where CMD resistant varieties deployment played a role in the dissemination of CBSVs. As actually we have additional viruses that seemed to trigger less noticeable symptoms, the risk of geographical expansion mediated by the deployment of those new CBSD-resistant varieties could be very high. The following aspects should be considered in the definition of their deployment by taking into consideration :
    - The level of CBSD pressure in each of the deployment areas : the approach developed in this study could be useful in accurately assessing the epidemiology of the CBSD and defining the disease pressure level. In the zones with no disease or low disease pressure (and with low risk of diseases dissemination), variety diversity should be a priority, and traditional varieties should be also included. These zones should be also privileged as central sites for the production of cuttings. In the zones of low disease pressure but with high risk of disease dissemination, variety diversity could be also considered but in this way, active monitoring and phytosanitation will be required, together with the deployment of resistant varieties. In the zones with high disease pressure, the focus should be dedicated on inoculum reduction by privileging the use of resistant varieties but afterwards considering reintroducing progressively other varieties when the disease has been significantly reduced.
    - The level of virus diversity in each zone
    - The strategic geographical location of various zones (on the borders, etc)
- Adapted strategies for controlling plant material movements and facilitating the delivery of clean seeds to farmers: because the main identified limits in cassava seed system was not only related to the availability of improved planting material but also in term of their accessibility and use by farmers, identifying best strategies to boost the use of high-quality planting material by farmers would be

key in establishing resilient seed systems. In such a way some of the key actors such as local seed multipliers or traders who has the advantage of accessing remote areas could be suitable to involve within the strategy to be promoted.

Definitely, some actions that can introduce immediate changes for the benefit of farmers in this context of CBSD burden would be focused on measures that allows reducing the infection and propagation risks. Measures such as (i) avoidance of infections through cultural practices such as the removal of green bridges (weeds or other crops that can serve as alternate hosts), (ii) eradication of infections through the use of cuttings from healthy plants, elimination (rouging and incineration) of infected plants so that they cannot be used as source of cuttings the next season. These measures will be strongly supported by the application of efficient detection methods as proposed above, and the multiplication of healthy cuttings.

For that purpose , the nurseries where seed multiplication activities take place need to be strengthened and promoted to make them not only efficient in mitigating the spread of viruses, but also economically viable and attractive for people to get interested in. Apart from targeting zones of low disease pressure for installing these nurseries, investigations should be conducted to identify which cost-effective methods of boosting the natural immunity of cassava plants against viral diseases, or cost-effective methods of boosting plant growth or stem branching could contribute to raising the efficiency of the seed multiplication activities as well as its economical attractiveness.

Bibliography

BIBLIOGRAPHY

Adams, Ian P., Adrian Fox, Neil Boonham, Sébastien Massart, and Kris De Jonghe. 2018. “The Impact of High Throughput Sequencing on Plant Health Diagnostics.” *European Journal of Plant Pathology* 152 (4): 909–19. https://doi.org/10.1007/s10658-018-1570-0.

Adams, M. J., J. F. Antoniw, and C. M. Fauquet. 2005. “Molecular Criteria for Genus and Species Discrimination within the Family Potyviridae.” *Archives of Virology* 150 (3): 459–79. https://doi.org/10.1007/s00705-004-0440-6.

Adiputra, Jati, Sridhar Jarugula, and Rayapati A. Naidu. 2019. “Intra-Species Recombination among Strains of the Ampelovirus Grapevine Leafroll-Associated Virus 4.” *Virology Journal* 16 (1): 139. https://doi.org/10.1186/s12985-019-1243-4.

Afgan, Enis, Dannon Baker, Bérénice Batut, Marius Van Den Beek, Dave Bouvier, Martin Ech, John Chilton, et al. 2018. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update.” *Nucleic Acids Research* 46 (W1): W537–44. https://doi.org/10.1093/nar/gky379.

Agresti, Alan. 2002. *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471249688.

Agrios, George N. 2005. *Plant Pathology, Fifth Edition*. Edited by K.D. Sonnack. Elsevier A. Burlington, USA: Dana Dreibelbis. https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

Akano, A. O., A. G.O. Dixon, C. Mba, E. Barrera, and M. Fregene. 2002. “Genetic Mapping of a Dominant Gene Conferring Resistance to Cassava Mosaic Disease.” *Theoretical and Applied Genetics* 105 (4): 521–25. https://doi.org/10.1007/s00122-002-0891-7.

Alabi, Olufemi J., Francis O. Ogbe, Ranajit Bandyopadhyay, P. Lava Kumar, Alfred G O Dixon, Jaqueline D’A. Hughes, and Rayapati A. Naidu. 2008. “Alternate Hosts of African Cassava Mosaic Virus and East African Cassava Mosaic Cameroon Virus in Nigeria.” *Archives of Virology* 153 (9): 1743–47. https://doi.org/10.1007/s00705-008-0169-8.

Alabi, Olufemi J, P Lava Kumar, and Rayapati A Naidu. 2011. “Cassava Mosaic Disease: A Curse to Food Security in Sub-Saharan Africa.” *Online. APSnet Features. Doi* 10.

Alexander, Thomas S. 2016. “Human Immunodeficiency Virus Diagnostic Testing: 30 Years of Evolution.” *Clinical and Vaccine Immunology* 23 (4): 249–53. https://doi.org/10.1128/CVI.00053-16.

Alicai, T, C A Omongo, M N Maruthi, R J Hillocks, Y Baguma, R Kawuki, A Bua, G.W. Otim-Nape, and J. Colvin. 2007. "Re-Emergence of Cassava Brown Streak Disease in Uganda." *Plant Disease* 91 (1): 24–29. https://apsjournals.apsnet.org/doi/pdf/10.1094/PD-91-0024.

Alicai, Titus, Joseph Ndunguru, Peter Sseruwagi, Fred Tairo, Geoffrey Okao-Okuja, Resty Nanvubya, Lilliane Kiiza, Laura Kubatko, Monica A. Kehoe, and Laura M. Boykin. 2016a. "Cassava Brown Streak Virus Has a Rapidly Evolving Genome: Implications for Virus Speciation, Variability, Diagnosis and Host Resistance." *Scientific Reports* 6 (1): 36164. https://doi.org/10.1038/srep36164.

Alicai, Titus, Joseph Ndunguru, Peter Sseruwagi, Fred Tairo, Geoffrey Okao-Okuja, Resty Nanvubya, Lilliane Kiiza, Laura Kubatko, Monica A Kehoe, and Laura M Boykin. 2016b. "Cassava Brown Streak Virus Has a Rapidly Evolving Genome: Implications for Virus Speciation, Variability, Diagnosis and Host Resistance." *Nature Publishing Group*. https://doi.org/10.1038/srep36164.

Allem, A. C. 2001. "The Origins and Taxonomy of Cassava." In *Cassava: Biology, Production and Utilization*, edited by R.J. Hillocks, J.M. Thresh, and A Belloti, 1–16. UK: CABI Publishing. https://doi.org/10.1079/9780851995243.0001.

Allem, A C. 2001. "Three New Infraspecific Taxa of Manihot (Euphorbiaceae) from the Brazilian Neotropics." *Novon* 11: 157–65. http://www.biodiversitylibrary.org/part/12775.

Allem, Antonio C. 1999. "The Closest Wild Relatives of Cassava (Manihot Esculenta Crantz)." *Euphytica* 107 (2): 123–33. https://doi.org/10.1023/A:1026422229054.

Almekinders, Authors C J M, Steve Walsh, Kim Jacobs, Jorge Andrade, Margaret Mcewan, and Stef De Haan. 2019. "Why Interventions in the Seed Systems of Roots , Tubers and Bananas Crops Do Not Reach Their Full Potential : A Reflection Based on Literature and Thirteen Case Studies." *Food Security*, no. January: 1–20. https://doi.org/10.1007/s12571-018-0874-4.

Alvarez, Elizabeth, Juan F Mejía, Germán A Llano, John B Loke, Alberto Calari, Bojan Duduk, and Assunta Bertaccini. 2009. "Characterization of a Phytoplasma Associated with Frogskin Disease in Cassava." *Plant Disease* 93 (11): 1139 – 1145. https://doi.org/10.1094/PDIS-93-11-1139.

Amisse, J. J.G., J. Ndunguru, F. Tairo, E. Ateka, L. M. Boykin, M. A. Kehoe, N. Cossa, C. Rey, and P. Sseruwagi. 2019. "Analyses of Seven New Whole Genome Sequences of Cassava Brown Streak Viruses in Mozambique Reveals Two Distinct Clades: Evidence for New Species." *Plant Pathology* 68 (5): 1007–18. https://doi.org/10.1111/ppa.13001.

Andersen, K. F., C. E. Buddenhagen, P. Rachkara, R. Gibson, S. Kalule, D. Phillips, and K. A. Garrett. 2019. "Modeling Epidemics in Seed Systems and Landscapes

to Guide Management Strategies: The Case of Sweet Potato in Northern Uganda." *Phytopathology* 109 (9): 1519–32. https://doi.org/10.1094/PHYTO-03-18-0072-R.

Anitha, J, T Makeshkumar, and S Edison. 2011. "Survey of Cassava Mosaic Disease in Kerala." *Journal of Root Crops* 37: 41 – 47. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84922401937&partnerID=40&md5=4f425725e44df67180fef0918c4b7495.

Anjanappa, Ravi B, Devang Mehta, M N Maruthi, Edward Kanju, Wilhelm Gruissem, and Hervé Vanderschuren. 2016. "Characterization of Brown Streak Virus–Resistant Cassava." *Molecular Plant-Microbe Interactions* 29 (7): 527–34. https://doi.org/10.1094/MPMI-01-16-0027-R.

Ariyo, O A, G I Atiri, A G O Dixon, and S Winter. 2006. "The Use of Biolistic Inoculation of Cassava Mosaic Begomoviruses in Screening Cassava for Resistance to Cassava Mosaic Disease." *Journal of Virological Methods* 137 (1): 43–50. https://doi.org/https://doi.org/10.1016/j.jviromet.2006.05.031.

Ariyo, O A, M Koerbler, A G O Dixon, G I Atiri, and S Winter. 2005. "Molecular Variability and Distribution of Cassava Mosaic Begomoviruses in Nigeria." *Journal of Phytopathology* 153 (4): 226–31. https://doi.org/10.1111/j.1439-0434.2005.00958.x.

Atiri, G I, F O Ogbe, A G O Dixon, S Winter, and O Ariyo PhD. 2004. "Status of Cassava Mosaic Virus Diseases and Cassava Begomoviruses in Sub-Saharan Africa." *Journal of Sustainable Agriculture* 24 (3): 5–35. https://doi.org/10.1300/J064v24n03_03.

Aydin, Suleyman. 2015. "A Short History, Principles, and Types of ELISA, and Our Laboratory Experience with Peptide/Protein Analyses Using ELISA." *Peptides* 72: 4–15. https://doi.org/10.1016/j.peptides.2015.04.012.

Babaian, Artem, and Robert C. Edgar. 2021. "Ribovirus Classification by a Polymerase Barcode Sequence." *BioRxiv*, 2021.03.02.433648. https://www.biorxiv.org/content/10.1101/2021.03.02.433648v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.03.02.433648v1.abstract.

Badamasi, H., M. D. Alegbejo, B. D. Kashina, and O. O. Banwo. 2020. "Incidence and Distribution of Viruses Infecting Cassava in Kaduna and Sokoto States, NIGERIA." *FUDMA JOURNAL OF SCIENCES* 4 (2): 325–36. https://doi.org/10.33003/fjs-2020-0402-221.

Ball, Cameron S., Yooli K. Light, Chung Yan Koh, Sarah S. Wheeler, Lark L. Coffey, and Robert J. Meagher. 2016. "Quenching of Unincorporated Amplification Signal Reporters in Reverse-Transcription Loop-Mediated Isothermal Amplification Enabling Bright, Single-Step, Closed-Tube, and Multiplexed Detection of RNA Viruses." *Analytical Chemistry* 88 (7): 3562–68.

https://doi.org/10.1021/acs.analchem.5b04054.

Banathy, Bela. 1996. “A Taste of Systemics, Why a System View ?” Edited by International Society for the Systems Sciences (ISSS) and International Institute for Systemic Inquiry and Integration (IISII). The First International Electronic Seminar on Wholeness. [Online]. 1996. http://www.newciv.org/ISSS_Primer/seminar.html.

Bar-Joseph, Moshe, and Munir Mawassi. 2013. “The Defective RNAs of Closteroviridae.” *Frontiers in Microbiology* 4HSP70h (MAY): 1–6. https://doi.org/10.3389/fmicb.2013.00132.

Barreda-García, Susana, Rebeca Miranda-Castro, Noemí de-los-Santos-Álvarez, Arturo J. Miranda-Ordieres, and María Jesús Lobo-Castañón. 2018. “Helicase-Dependent Isothermal Amplification: A Novel Tool in the Development of Molecular-Based Analytical Systems for Rapid Pathogen Detection.” *Analytical and Bioanalytical Chemistry* 410 (3): 679–93. https://doi.org/10.1007/s00216-017-0620-3.

Batista, Angelo C., and Luis G.C. Pacheco. 2018. “Detecting Pathogens with Zinc-Finger, TALE and CRISPR- Based Programmable Nucleic Acid Binding Proteins.” *Journal of Microbiological Methods* 152 (July): 98–104. https://doi.org/10.1016/j.mimet.2018.07.024.

Baulcombe, David. 2004. “RNA Silencing in Plants.” *Nature* 431 (7006): 356–63. https://doi.org/10.1038/nature02874.

Baumgart, Eveline, Arno Schad, and Markus Grabenbauer. 2001. “In Situ Hybridization: General Principles and Application of Digoxigenin-Labeled CRNA for the Detection of MRNAs.” In *Immunocytochemistry and In Situ Hybridization in the Biomedical Sciences*, edited by Julian E Beesley, 108–37. Boston, MA: Birkhäuser Boston. https://doi.org/10.1007/978-1-4612-0139-7_6.

Bebber, DP, T Holmes, and SJ Gurr. 2014. “The Global Spread of Crop Pests and Pathogens.” *Global Ecology and Biogeography* 23: 1398–1407.

Becherer, Lisa, Mohammed Bakheit, Sieghard Frischmann, Silvina Stinco, Nadine Borst, Roland Zengerle, and Felix Von Stetten. 2018. “Simplified Real-Time Multiplex Detection of Loop-Mediated Isothermal Amplification Using Novel Mediator Displacement Probes with Universal Reporters.” *Analytical Chemistry* 90 (7): 4741–48. https://doi.org/10.1021/acs.analchem.7b05371.

Benesi, I.R.M., T Mkandawire, F Chlpungu, P Pamkomera, J Mkumbira, N M Mahungu, C C Moyo, A.P. Ntonga, and S. Chilungo. 2010. “CH92/O82 And TME 6 Cassava Clones Officially Released for Farmer Growing in Malawi.” *Roots Newsletter* 12 (2): 4.

Bentley, J, A Olanrewaju, T Madu, O Olaosebikan, T Abdoulaye, T Assfaw Wossen, V Manyong, et al. 2017. *Cassava Farmers’ Preferences for Varieties and Seed*

Dissemination System in Nigeria: Gender and Regional Perspectives.

Bhat, Alangar I., Rashid Aman, and Magdy Mahfouz. 2022. “Onsite Detection of Plant Viruses Using Isothermal Amplification Assays.” *Plant Biotechnology Journal* 20 (10): 1859–73. https://doi.org/10.1111/pbi.13871.

Bhatnagar, S. K., L. C. Bullions, and M. J. Bessman. 1991. “Characterization of the MutT Nucleoside Triphosphatase of Escherichia Coli.” *Journal of Biological Chemistry* 266 (14): 9050–54. https://doi.org/10.1016/s0021-9258(18)31550-3.

Bhattacharyya, Roby P, Sri Gowtham Thakku, and Deborah T Hung. 2018. “Harnessing CRISPR Effectors for Infectious Disease Diagnostics.” *ACS Infectious Diseases* 4 (9): 1278–82. https://doi.org/10.1021/acsinfecdis.8b00170.

Bisimwa, Espoir B. 2012. “Epidémiologie, Diversité Génétique, Distribution et Contrôle Des Virus de La Mosaïque Africaine de Manioc et de Son Vecteur (Bemisia Tabaci) Dans La Région Du Sud Kivu En République Démocratique Du Congo.” Université Catholique de Louvain. https://doi.org/10.3923/ajppaj.2012.1.12.

Bisimwa, Espoir, Jean Walungululu, and Claude Bragard. 2012. “Occurence and Distribution of Cassava Mosaic Begomoviruses Related to Agro-Ecosystems in the Sud-Kivu Province, Democratic Republic of Congo.” *Asian Journal of Plant Pathology* 6 (1): 1–12. https://doi.org/10.3923/ajppaj.2012.1.12.

Blouin, A. G., R. Biccheri, M. E. Khalifa, M N Pearson, C Poggi Pollini, C Hamiaux, D Cohen, and C Ratti. 2017. “Characterization of Actinidia Virus 1 , a New Member of the Family Closteroviridae Encoding a Thaumatin - like Protein.” *Archives of Virology*, 1–6. https://doi.org/10.1007/s00705-017-3610-z.

Blouin, A G, M N Pearson, R R Chavan, E N Y Woo, B S M Lebas, S Veerakone, C Ratti, R. Biccheri, R.M MacDiarmid, and D Cohen. 2013. “Viruses of Kiwifruit (Actinidia Species).” *Journal of Plant Pathology* 95 (2): 221–35.

Blouin, Arnaud G., Howard A. Ross, Jody Hobson-Peters, Caitlin A. O’Brien, Ben Warren, and Robin MacDiarmid. 2016. “A New Virus Discovered by Immunocapture of Double-Stranded RNA, a Rapid Method for Virus Enrichment in Metagenomic Studies.” *Molecular Ecology Resources* 16 (5): 1255–63. https://doi.org/10.1111/1755-0998.12525.

Blum, Matthias, Hsin Yu Chang, Sara Chuguransky, Tiago Grego, Swaathi Kandasaamy, Alex Mitchell, Gift Nuka, et al. 2021. “The InterPro Protein Families and Domains Database: 20 Years On.” *Nucleic Acids Research* 49 (D1): D344–54. https://doi.org/10.1093/nar/gkaa977.

Bock, K. R., E. J. Guthrie, and Gina Figueireido. 1981. “A Strain of Cassava Latent Virus Occurring in Coastal Districts of Kenya.” *Annals of Applied Biology* 99 (2): 151–59. https://doi.org/10.1111/j.1744-7348.1981.tb05142.x.

Bock, K R. 1994. “Studies on Cassava Brown Streak Virus Disease in Kenya.” *Tropical Science* 34 (1): 134–45.

Borah, Basanta Kumar, and Indranil Dasgupta. 2012. “PCR-RFLP Analysis Indicates That Recombination Might Be a Common Occurrence among the Cassava Infecting Begomoviruses in India.” *Virus Genes* 45 (2): 327–32. https://doi.org/10.1007/s11262-012-0770-5.

Bos, L. 1976. “Problems and Prospects in Plant Virus Identification.” *EPPO Bulletin* 6 (2): 63–86. https://doi.org/10.1111/j.1365-2338.1976.tb02020.x.

Boyle, David S., Ruth McNerney, Hwee Teng Low, Brandon Troy Leader, Ailyn C. Pérez-Osorio, Jessica C. Meyer, Denise M. O’Sullivan, David G. Brooks, Olaf Piepenburg, and Matthew S. Forrest. 2014. “Rapid Detection of Mycobacterium Tuberculosis by Recombinase Polymerase Amplification.” *PLoS ONE* 9 (8): 1–9. https://doi.org/10.1371/journal.pone.0103091.

Bratlie, Marit S., and Finn Drabløs. 2005. “Bioinformatic Mapping of AlkB Homology Domains in Viruses.” *BMC Genomics* 6 (Ii): 1–15. https://doi.org/10.1186/1471-2164-6-1.

Broughton, James P, Xianding Deng, Guixia Yu, Clare L Fasching, Venice Servellita, Jasmeet Singh, Xin Miao, et al. 2020. “CRISPR–Cas12-Based Detection of SARS-CoV-2.” *Nature Biotechnology* 38 (7): 870–74. https://doi.org/10.1038/s41587-020-0513-4.

Brown, N P, C Leroy, and C Sander. 1998. “MView: A Web-Compatible Database Search or Multiple Alignment Viewer.” *Bioinformatics* 14 (4): 380–81. https://doi.org/10.1093/bioinformatics/14.4.380.

Brunt, AA. 1996. “Cassava Brown Streak-associated(?) Carlavirus, Cassava Brown Streak Potyvirus.” In *Plant Viruses Online: Descriptions and Lists from the VIDE Database*, edited by K Crabtree, MJ Dallwitz, AJ Gibbs, L Watson, and EJ Zurcher.

Buddenhagen, C. E., J. F. Hernandez Nopsa, K. F. Andersen, J. Andrade-Piedra, G. A. Forbes, P. Kromann, S. Thomas-Sharma, P. Useche, and K. A. Garrett. 2017. “Epidemic Network Analysis for Mitigation of Invasive Pathogens in Seed Systems: Potato in Ecuador.” *Phytopathology*, no. September: PHYTO-03-17-010. https://doi.org/10.1094/PHYTO-03-17-0108-FI.

Burgis, Nicholas E., Jason J. Brucker, and Richard P. Cunningham. 2003. “Repair System for Noncanonical Purines in Escherichia Coli.” *Journal of Bacteriology* 185 (10): 3101–10. https://doi.org/10.1128/JB.185.10.3101-3110.2003.

Burns, Anna, Roslyn Gleadow, Julie Cliff, Anabela Zacarias, and Timothy Cavagnaro. 2010. “Cassava: The Drought, War and Famine Crop in a Changing World.” *Sustainability* 2: 3572–3607. https://doi.org/10.3390/su2113572.

Buyel, J. F., R. M. Twyman, and R. Fischer. 2017. “Very-Large-Scale Production of

Antibodies in Plants: The Biologization of Manufacturing." *Biotechnology Advances* 35 (4): 458–65. https://doi.org/10.1016/j.biotechadv.2017.03.011.

Byzova, Nadezhda A., Svetlana V. Vinogradova, Elena V. Porotikova, Uliana D. Terekhova, Anatoly V. Zherdev, and Boris B. Dzantiev. 2018. "Lateral Flow Immunoassay for Rapid Detection of Grapevine Leafroll-Associated Virus." *Biosensors* 8 (4): 1–14. https://doi.org/10.3390/bios8040111.

CABI. 2023. "Cassava Mosaic Disease." In *CABI Compendium*, CAB Intern. Wallingford, UK. https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.2747.

Campo, Beatriz Vanessa Herrera, Glenn Hyman, and Anthony Bellotti. 2011. "Threats to Cassava Production: Known and Potential Geographic Distribution of Four Key Biotic Constraints." *Food Security* 3 (3): 329–45. https://doi.org/10.1007/s12571-011-0141-4.

Canadian Agency for Drugs and Technologies in Health. 2014. *Next Generation DNA Sequencing: A Review of the Cost Effectiveness and Guidelines*. Ottawa (ON), Canada.

Candresse, Thierry, Denis Filloux, Brejnev Muhire, Charlotte Julian, Serge Galzi, Guillaume Fort, Pauline Bernardo, et al. 2014. "Appearances Can Be Deceptive: Revealing a Hidden Viral Infection with Deep Sequencing in a Plant Quarantine Context." *PLoS ONE* 9 (7). https://doi.org/10.1371/journal.pone.0102945.

Candresse, Thierry, and Marc Fuchs. 2020. "Closteroviridae." In *Encyclopedia of Life Science*, Wiley. in press.

Carter, S E, L O Fresco, P G Jones, and J N Fairbairn. 1993. "Introduction and Diffusion of Cassava in Africa." *IITA Research Guide 49*. Ibadan, Nigeria: IITA. http://old.iita.org/cms/details/trn_mat/ irg49/irg49.htm.

Carter, Walter. 1939. "Populations of Thrips Tabaci, with Special Reference to Virus Transmission." *The Journal of Animal Ecology* 8 (2): 261–76. https://doi.org/10.2307/1234.

Carvajal-Yepes, Monica, Cristian Olaya, Ivan Lozano, Maritza Cuervo, Mauricio Castaño, and Wilmer J. Cuellar. 2014. "Unraveling Complex Viral Infections in Cassava (Manihot Esculenta Crantz) from Colombia." *Virus Research* 186: 76–86. https://doi.org/10.1016/j.virusres.2013.12.011.

Carvalho, Lucas Silva, Otávio Bravim da Silva, Gabriela Carneiro da Almeida, Juliana Davies de Oliveira, Nadia Skorupa Parachin, and Talita Souza Carmo. 2017. "Production Processes for Monoclonal Antibodies." *Fermentation Processes*. https://doi.org/10.5772/64263.

Cases-gonzález, Clara, María Arribas, Esteban Domingo, and Ester Lázaro. 2008. "Beneficial Effects of Population Bottlenecks in an RNA Virus Evolving at

Increased Error Rate," 1120–29. https://doi.org/10.1016/j.jmb.2008.10.014.

Casinga, C. M., R. R. Shirima, N. M. Mahungu, W. Tata-Hangy, K. B. Bashizi, C. M. Munyerenkana, H. Ughento, et al. 2021. "Expansion of the Cassava Brown Streak Disease Epidemic in Eastern Democratic Republic of Congo." *Plant Disease* 105 (8): 2177–88. https://doi.org/10.1094/PDIS-05-20-1135-RE.

Cassedy, A., A. Parle-McDermott, and R. O'Kennedy. 2021. "Virus Detection: A Review of the Current and Emerging Molecular and Immunological Methods." *Frontiers in Molecular Biosciences* 8 (April): 1–21. https://doi.org/10.3389/fmolb.2021.637559.

Chambers, J.M. 1992. "Linear Models." In *Statistical Models*, edited by J.M. Chambers and T.J. Hastie, Wadsworth. Cole.

Chang, Shujun, Jeff Puryear, and John Cairney. 1993. "A Simple and Efficient Method for Isolating RNA from Pine Trees." *Plant Molecular Biology Reporter* 11 (2): 113–16. https://doi.org/10.1007/BF02670468.

Chaudhary, Reema, and Ganesh Kumar Maurya. 2019. "Restriction Fragment Length Polymorphism." In *Encyclopedia of Animal Cognition and Behavior*, edited by Jennifer Vonk and Todd Shackelford, 1–3. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47829-6_175-1.

Checkland, P. 1981. *Systems Thinking, Systems Practice*. Chichester, UK: John Wiley. https://doi.org/10.1002/1099-1743(200011)17:1+<::AID-SRES383>3.0.CO;2-Q.

Chen, Janice S, Enbo Ma, Lucas B Harrington, Maria Da Costa, Xinran Tian, Joel M Palefsky, and Jennifer A Doudna. 2018. "CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity." *Science* 360 (6387): 436–39. https://doi.org/10.1126/science.aar6245.

Christinck, Anja, Fred Rattunde, Alpha Kergna, Wellington Mulinge, Ea Weltzien, Y O U Can, and T Grow Alone. 2018. "'You Cannot Grow Alone' — Prioritized Sustainable Seed System Development Options for Staple Food Crops in Sub-Saharan Africa: Cases of Kenya and Mali." Bonn. http://research4agrinnovation.org/publication/seed_systems_mali_kenya.

Christinck, Anja, Fred Rattunde, Alpha Kergna, Wellington Mulinge, and Eva Weltzien. 2018. "Identifying Options for the Development of Sustainable Seed Systems: Insights from Kenya and Mali." *ZEF Working Paper Series* 165: 56. http://hdl.handle.net/10419/187471.

Clark, Kevin D., Cheng Zhang, and Jared L. Anderson. 2016. "Sample Preparation for Bioanalytical and Pharmaceutical Analysis." *Analytical Chemistry* 88 (23): 11262–70. https://doi.org/10.1021/acs.analchem.6b02935.

Codjia, Esperance D., Bunmi Olasanmi, Paterne A. Agre, Ruth Uwugiaren, Adenike D. Ige, and Ismail Y. Rabbi. 2022. "Selection for Resistance to Cassava Mosaic

Disease in African Cassava Germplasm Using Single Nucleotide Polymorphism Markers." *South African Journal of Science* 118 (112): 1–7. https://doi.org/10.17159/sajs.2022/11607.

Collet, D. 2003. *Modelling Binary Data*. Chapman &.

Coons, Albert H, Hugh J Creech, and R Norman Jones. 1941. "Immunological Properties of an Antibody Containing a Fluorescent Group." *Proceedings of the Society for Experimental Biology and Medicine* 47 (2): 200–202. https://doi.org/10.3181/00379727-47-13084P.

CRS. 2012. "Final Report on the Great Lakes Cassava Initiatives." https://www.crs.org/sites/default/files/tools-research/final-report-great-lakes-cassava-initiative.pdf.

DANAGRO. 1988. "SADCC The Regional Seed Production and Supply Project: Main Report (Vol 1A) and Country Reports (Vol. 2A-2J)." GLostrup, Danemark.

Delahaye, Clara, and Jacques Nicolas. 2021. "Sequencing DNA with Nanopores: Troubles and Biases." *PLoS ONE* 16 (10 October). https://doi.org/10.1371/journal.pone.0257521.

Delêtre, Marc, Jean-Michel Lett, Ronan Sulpice, and Charles Spillane. 2021. "Kinship Networks of Seed Exchange Shape Spatial Patterns of Plant Virus Diversity." *Nature Communications* 12 (1): 1–10. https://doi.org/10.1038/s41467-021-24720-6.

Demidov, Vadim V. 2003. "DNA Diagnostics in the Fifty-Year Retrospect." *Expert Review of Molecular Diagnostics* 3 (2): 121–24. https://doi.org/10.1586/14737159.3.2.121.

Deng, D, W G Otim-Nape, A Sangare, S Ogwal, R N Beachy, and C M Fauquet. 1997. "Presence of a New Virus Closely Related to East African Cassava Mosaic Geminivirus, Associated with Cassava Mosaic Outbreak in Uganda." *African Journal of Root and Tuber Crops* 2 (1/2): 23–28.

Devaux, André, Peter Kromann, and Oscar Ortiz. 2014. "Potatoes for Sustainable Global Food Security." *Potato Research* 57 (3–4): 185–99. https://doi.org/10.1007/s11540-014-9265-1.

Dey, Kishore K., James C. Green, Michael Melzer, Wayne Borth, and John S. Hu. 2018. "Mealybug Wilt of Pineapple and Associated Viruses." *Horticulturae* 4 (4). https://doi.org/10.3390/horticulturae4040052.

Dey, Kishore K, Jaylinn Sugikawa, Christopher Kerr, and Michael J Melzer. 2019. "Air Potato (Dioscorea Bulbifera) Plants Displaying Virus-like Symptoms Are Co-Infected with a Novel Potyvirus and a Novel Ampelovirus." *Virus Genes* 55 (1): 117–21. https://doi.org/10.1007/s11262-018-1616-6.

DHSC. 2020. *Technical Validation of OptiGene RT LAMP Assay (Direct and RNA Formats).* Department of Health and Social Care. https://books.google.be/books?id=cXg6zgEACAAJ.

Diaz-ruiz, J R, and J M Kaper. 1978. "Isolation of Viral Double-Stranded RNAs Using a LiCl Fractionation Procedure." *Preparative Biochemistry* 8 (1): 1–17. https://doi.org/10.1080/00327487808068215.

Diekmann, M D, Food, Agriculture Organization of the United Nations, and International Plant Genetic Resources Institute. 1994. *FAO/IPGRI Technical Guidelines for the Safe Movement of Germplasm*. Food and Agriculture Organization of the United Nations. https://books.google.be/books?id=r-MkzQEACAAJ.

Ding, Fang, Cristina Paul, Ron Brlansky, and John S. Hartung. 2017. "Immune Tissue Print and Immune Capture-PCR for Diagnosis and Detection of Candidatus Liberibacter Asiaticus." *Scientific Reports* 7 (1): 46467. https://doi.org/10.1038/srep46467.

Dodge, Y. 2008. *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-32833-1.

Dolja, Valerian V, Alexander V Karasev, and Eugene V Koonin. 1994. "Molecular Biology and Evolution of Closterovirus: Sophisticated Build-up of Large RNA Genomes." *Annual Review of Phytopathology* 32: 261–85.

Donaire, Livia, Yu Wang, Daniel Gonzalez-Ibeas, Klaus F. Mayer, Miguel A. Aranda, and César Llave. 2009. "Deep-Sequencing of Plant Viral Small RNAs Reveals Effective and Widespread Targeting of Viral Genomes." *Virology* 392 (2): 203–14. https://doi.org/10.1016/j.virol.2009.07.005.

Dubern, J. 1994. "Transmission of African Cassava Mosaic Geminivirus by the Whitefly (Bemisia Tabaci)." *Trop. Sci.* 34 (1): 82 – 91. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0001925432&partnerID=40&md5=3c3fe33b00926e6b7d37f99e4cb0714a.

Ecker, Christina, Anna Ertl, Walter Pulverer, Albert Nemes, Pal Szekely, Angelika Petrasch, Gertrud Linsberger-Martin, and Margit Cichna-Markl. 2013. "Validation and Comparison of a Sandwich ELISA, Two Competitive ELISAs and a Real-Time PCR Method for the Detection of Lupine in Food." *Food Chemistry* 141 (1): 407–18. https://doi.org/10.1016/j.foodchem.2013.02.091.

Edgar, Robert C., Jeff Taylor, Victor Lin, Tomer Altman, Pierre Barbera, Dmitry Meleshko, Dan Lohr, et al. 2022. "Petabase-Scale Sequence Alignment Catalyses Viral Discovery." *Nature* 602 (7895): 142–47. https://doi.org/10.1038/s41586-021-04332-2.

Eigner, Ulrich, Svenja Reucher, Nadine Hefner, Sandrine Staffa-Peichl, Melissa Kolb, Ulrike Betz, Martin Holfelder, Gene Spier, Susanne Pfefferle, and Marc

Lütgehetmann. 2019. “Clinical Evaluation of Multiplex RT-PCR Assays for the Detection of Influenza A/B and Respiratory Syncytial Virus Using a High Throughput System.” *Journal of Virological Methods* 269 (December 2018): 49–54. https://doi.org/10.1016/j.jviromet.2019.03.015.

El-Sharkawy, Mabrouk A. 2004. “Cassava Biology and Physiology.” *Plant Molecular Biology* 56 (4): 481–501. https://doi.org/10.1007/s11103-005-2270-7.

Eliasco, E., I. C. Livieratos, G. Müller, M. Guzman, L. F. Salazar, and R. H.A. Coutts. 2006. “Sequences of Defective RNAs Associated with Potato Yellow Vein Virus.” *Archives of Virology* 151 (1): 201–4. https://doi.org/10.1007/s00705-005-0625-7.

Elmer, Wade H. 2001. “Seeds as Vehicles for Pathogen Importation.” *Biological Invasions* 3 (3): 263–71. https://doi.org/10.1023/A:1015217308477.

Engvall, Eva, and Peter Perlmann. 1971. “Enzyme-Linked Immunosorbent Assay (ELISA) Quantitative Assay of Immunoglobulin G.” *Immunochemistry.* 8 (9).

EPPO. 2018. “PM 7/76 (5) Use of EPPO Diagnostic Standards.” *EPPO Bulletin* 48 (3): 373–77. https://doi.org/10.1111/epp.12506.

———. 2021. “PM 7/98 (5) Specific Requirements for Laboratories Preparing Accreditation for a Plant Pest Diagnostic Activity.” *EPPO Bulletin* 51 (3): 468–98. https://doi.org/10.1111/epp.12780.

EPPO standard PM7/98, . 2021. “Specific Requirements for Laboratories Preparing Accreditation for a Plant Pest Diagnostic Activity.” *EPPO Bulletin* 51 (3): 468–98. https://doi.org/10.1111/epp.12780.

Estrela, Paulo Felipe Neves, Geovana de Melo Mendes, Kézia Gomes de Oliveira, Alexandre Melo Bailão, Célia Maria de Almeida Soares, Nilson Antônio Assunção, and Gabriela Rodrigues Mendes Duarte. 2019. “Ten-Minute Direct Detection of Zika Virus in Serum Samples by RT-LAMP.” *Journal of Virological Methods* 271 (May): 113675. https://doi.org/10.1016/j.jviromet.2019.113675.

Everitt, B S. 2006. *The Cambridge Dictionary of Statistics*. The Edinburgh Building, Cambridge cb2 2ru, UK.: Cambridge University Press. https://books.google.co.ke/books?id=m4B5QgAACAAJ.

FAO. 2020. *World Food and Agriculture - Statistical Yearbook 2020*. Rome: FAO. https://doi.org/10.4060/cb1329en.

———. 2021a. “Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database.” FAO. Rome. 2021. https://search.library.wisc.edu/catalog/999890171702121.

———. 2021b. “Plant Production and Protection Division: Cassava-Related Diseases.” FAO. Rome: Food and Agriculture Organization of the United

Natios. 2021. www.fao.org/agriculture/crops/thematic-sitemap/theme/pests/cassava-related-diseases/en/.

Fargette, D., C. Fauquet, and J-C. Thouvenel. 1988. "Yield Losses Induced by African Cassava Mosaic Virus in Relation to the Mode and the Date of Infection." *Tropical Pest Management* 34 (1): 89–91. https://doi.org/10.1080/09670878809371216.

Fargette, D, J M Thresh, and G William Otim-Nape. 1994. "The Epidemiology of African Cassava Mosaic Geminivirus: Reversion and the Concept of Equilibrium." *TROPICAL SCIENCE-LONDON-* 34: 123.

Fauquet, C.M., M.A. Mayo, J. Maniloff, U. Desselberger, and L.A. Ball. 2005. *Virus Taxonomy: Classification and Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press*. https://doi.org/10.1007/s13398-014-0173-7.2.

Fauquet, Claude, Denis Fargette, and Cassava Munihor. 1990. "African Cassava Mosaic Virus: Etiology, Epidemiology, and Control." *Plant Disease* 74 (6): 404. https://doi.org/10.1094/PD-74-0404.

Fermont, A M, P J A [van Asten], P Tittonell, M T [van Wijk], and K E Giller. 2009. "Closing the Cassava Yield Gap: An Analysis from Smallholder Farms in East Africa." *Field Crops Research* 112 (1): 24–36. https://doi.org/https://doi.org/10.1016/j.fcr.2009.01.009.

Fiallo-Olivé, Elvira, Jean Michel Lett, Darren P. Martin, Philippe Roumagnac, Arvind Varsani, F. Murilo Zerbini, and Jesús Navas-Castillo. 2021. "ICTV Virus Taxonomy Profile: Geminiviridae 2021." *Journal of General Virology* 102 (12): 1–2. https://doi.org/10.1099/jgv.0.001696.

Fondong, V. N., J. S. Pita, M. E C Rey, A. De Kochko, R. N. Beachy, and C. M. Fauquet. 2000. "Evidence of Synergism between African Cassava Mosaic Virus and a New Double-Recombinant Geminivirus Infecting Cassava in Cameroon." *Journal of General Virology* 81 (1): 287–97.

Food and Agriculture Organization of the United Nations. 2017. "FAOSTAT Statistical Database." Rome: [Rome] : FAO, c2017-. https://search.library.wisc.edu/catalog/999890171702121.

Fox, Adrian. 2020. "Reconsidering Causal Association in Plant Virology." *Plant Pathology* 69 (6): 956–61. https://doi.org/10.1111/ppa.13199.

Fox, Adrian, Adrian J. Gibbs, Aimee R. Fowkes, Hollie Pufal, Sam McGreig, Roger A.C. Jones, Neil Boonham, and Ian P. Adams. 2022. "Enhanced Apiaceous Potyvirus Phylogeny, Novel Viruses, and New Country and Host Records from Sequencing Apiaceae Samples." *Plants* 11 (15). https://doi.org/10.3390/plants11151951.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*.

Thousand Oaks CA: Sage. SAGE. California, USA: SAGE. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Francl, Leonard J. 2001. "The Disease Triangle: A Plant Pathological Paradigm Revisited." The Plant Health Instructor. 2001. https://doi.org/10.1094/PHI-T-2001-0517-01.

Francois, Patrice, Manuela Tangomo, Jonathan Hibbs, Eve-Julie Bonetti, Catharina C Boehme, Tsugunori Notomi, Mark D Perkins, and Jacques Schrenzel. 2011. "Robustness of a Loop-Mediated Isothermal Amplification Reaction for Diagnostic Applications." *FEMS Immunology & Medical Microbiology* 62 (1): 41–48. https://doi.org/10.1111/j.1574-695X.2011.00785.x.

Frangoie, Antoine N., S M Bidiaka, and N M Mahungu. 2019a. "La Filière Semencière En R.D. Congo." Informations Destinées Aux Acteurs et Parténaires Du Sous-Secteur Semences En R.D. C. Kinshasa/R.D. Congo.

———. 2019b. "Multiplication de Matériels de Plantation, Normes de Production et de Recolte & Maintenance Variétale." Kinshasa/R.D. Congo.

Frangoie, N, M Bidiaka, and N M Mahungu. 2012. "La Production Des Semences Certifiées. Module II. Partenariat IITA-PANA/ASA."

Fuchs, Marc, Moshe Bar-Joseph, Thierry Candresse, Hans J Maree, Giovanni P Martelli, Michael J Melzer, Wulf Menzel, Angelantonio Minafra, Sead Sabanadzovic, and ICTV Report Consortium. 2020. "ICTV Virus Taxonomy Profile: Closteroviridae." *Journal of General Virology*, 2. https://doi.org/https://doi.org/10.1099/jgv.0.001397.

Fuchs, Marc, Moshe Bar-joseph, Thierry Candresse, Giovanni P Martelli, Hano Maree, Michael J Melzer, Wulf Menzel, Angelantonio Minafra, and Sead Sabanadzovic. 2018. "Closteroviridae 2018.024 Approved Proposal." In *ICTV Online 10th Report*, 1–8.

Fux, Robert, Christina Söckler, Ellen Kathrin Link, Christine Renken, Roman Krejci, Gerd Sutter, Mathias Ritzmann, and Matthias Eddicks. 2018. "Full Genome Characterization of Porcine Circovirus Type 3 Isolates Reveals the Existence of Two Distinct Groups of Virus Strains." *Virology Journal* 15 (1): 1–9. https://doi.org/10.1186/s12985-018-0929-3.

Gall, Amanda D., Anthony Gall, Ashley C. Moore, Martin K. Aune, Steven Heid, Ayaka Mori, and Nicholas E. Burgis. 2013. "Analysis of Human ITPase Nucleobase Specificity by Site-Directed Mutagenesis." *Biochimie* 95 (9): 1711–21. https://doi.org/10.1016/j.biochi.2013.05.016.

García-Bernalt Diego, Juan, Pedro Fernández-Soto, Sergio Márquez-Sánchez, Daniel Santos Santos, Begoña Febrer-Sendra, Beatriz Crego-Vicente, Juan Luis Muñoz-Bellido, Moncef Belhassen-García, Juan M. Corchado Rodríguez, and Antonio Muro. 2022. "SMART-LAMP: A Smartphone-Operated Handheld

Device for Real-Time Colorimetric Point-of-Care Diagnosis of Infectious Diseases via Loop-Mediated Isothermal Amplification." *Biosensors* 12 (6). https://doi.org/10.3390/bios12060424.

Garrett, K. A., R. I. Alcalá-Briseño, K. F. Andersen, C. E. Buddenhagen, R. A. Choudhury, J. C. Fulton, J. F. Hernandez Nopsa, R. Poudel, and Y. Xing. 2018. "Network Analysis: A Systems Framework to Address Grand Challenges in Plant Pathology." *Annual Review of Phytopathology* 56: 559–80. https://doi.org/10.1146/annurev-phyto-080516-035326.

Gibson, R. W., and J. F. Kreuze. 2015. "Degeneration in Sweetpotato Due to Viruses, Virus-Cleaned Planting Material and Reversion: A Review." *Plant Pathology* 64 (1): 1–15. https://doi.org/10.1111/ppa.12273.

Gibson, R W, J P Legg, and G W Otim-Nape. 1996. "Unusually Severe Symptoms Are a Characteristic of the Current Epidemic of Mosaic Virus Disease of Cassava in Uganda." *Annals of Applied Biology* 128 (3): 479–90. https://doi.org/10.1111/j.1744-7348.1996.tb07108.x.

Girma, Gezahegn, Ismail Rabbi, Adetunji Olanrewaju, Oluwafemi Alaba, Peter Kulakow, Tunrayo Alabi, Bamikole Ayedun, et al. 2017. *A Manual for Large-Scale Sample Collection, Preservation, Tracking, DNA Extraction, and Variety Identification Analysis*. Ibadan, Nigeria: IITA. http://www.iita.org/wp-content/uploads/2017/A_manual_for_large_scale.pdf.

Goldmeyer, James, Huimin Kong, and Wen Tang. 2007. "Development of a Novel One-Tube Isothermal Reverse Transcription Thermophilic Helicase-Dependent Amplification Platform for Rapid RNA Detection." *The Journal of Molecular Diagnostics* 9 (5): 639–44. https://doi.org/https://doi.org/10.2353/jmoldx.2007.070012.

Gootenberg, Jonathan S, Omar O Abudayyeh, Max J Kellner, Julia Joung, James J Collins, and Feng Zhang. 2018. "Multiplexed and Portable Nucleic Acid Detection Platform with Cas13, Cas12a, and Csm6." *Science* 360 (6387): 439–44. https://doi.org/10.1126/science.aaq0179.

Gorman, Kevin T, Lauren C Roby, Allison Giuffre, Renhua Huang, and Brian K Kay. 2017. "Tandem Phage-Display for the Identification of Non-Overlapping Binding Pairs of Recombinant Affinity Reagents." *Nucleic Acids Research* 45 (18): e158–e158. https://doi.org/10.1093/nar/gkx688.

Govindarajan, Rajeshwar, Jeyapradha Duraiyan, Karunakaran Kaliyappan, and Murugesan Palanisamy. 2012. "Microarray and Its Applications." *Journal of Pharmacy & Bioallied Sciences* 4 (Suppl 2): S310–12. https://doi.org/10.4103/0975-7406.100283.

Green, James C., Maher A. Rwahnih, Alejandro Olmedo-Velarde, Michael J. Melzer, Islam Hamim, Wayne B. Borth, Tracie M. Brower, Marisa Wall, and John S.

Hu. 2020. “Further Genomic Characterization of Pineapple Mealybug Wilt-Associated Viruses Using High-Throughput Sequencing.” *Tropical Plant Pathology*. https://doi.org/10.1007/s40858-019-00330-y.

Guo, Jia, Ning Xu, Zengmin Li, Shenglong Zhang, Jian Wu, Hyun Kim Dae, Sano Marma Mong, et al. 2008. “Four-Color DNA Sequencing with 3′-O-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (27): 9145–50. https://doi.org/10.1073/pnas.0804023105.

Hall, E.W. 1928. “Annual Report of the Department of Agriculture, Uganda.” Entebbe.

Hall, Richard J, Jing Wang, Angela K Todd, Ange B Bissielo, Seiha Yen, Hugo Strydom, Nicole E Moore, et al. 2014. “Evaluation of Rapid and Simple Techniques for the Enrichment of Viruses Prior to Metagenomic Virus Discovery.” *Journal of Virological Methods* 195 (January): 194–204. https://doi.org/10.1016/j.jviromet.2013.08.035.

Haq, Imran Ul, Nabeeha Aslam Khan, and Muhammad Kaleem Sarwar. 2022. “Predictive Models for Plant Disease Assessment.” In *Trends in Plant Disease Assessment*, 225–39. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5896-0_12.

Harimalala, Mireille, Pierre Lefeuvre, Alexandre de Bruyn, Fidèle Tiendrébéogo, Murielle Hoareau, Julie Villemot, Sahondramalala Ranomenjanahary, Alice Andrianjaka, Bernard Reynaud, and Jean Michel Lett. 2012. “A Novel Cassava-Infecting Begomovirus from Madagascar: Cassava Mosaic Madagascar Virus.” *Archives of Virology* 157 (10): 2027–30. https://doi.org/10.1007/s00705-012-1399-3.

Harms, R.W. 1981. *River of Wealth, River of Sorrow: The Central Zaire Basin in the Era of the Slave and Ivory Trade, 1500–1891*. New Haven, Connecticut: Yale University Press. https://doi.org/10.1017/S0022050700029971.

Harms, Robert. 1979. “Fish and Cassava: The Changing Equation.” *African Economic History* 7: 113–16. https://www.jstor.org/stable/3601206.

Harper, Glyn, Ganesh Dahal, George Thottappilly, and R. Hull. 1999. “Detection of Episomal Banana Streak Badnavirus by IC-PCR.” *Journal of Virological Methods* 79 (1): 1–8. https://doi.org/10.1016/S0166-0934(98)00158-X.

Harrison, B.D. 2009. “A Brief Outline of the Development of Plant Virology in the 20th Century.” *Journal of Plant Pathology* 91 (3): 509–20.

Harrison, B D, X Zhou, G W Otim Nape, Y Liu, and D J Robinson. 1997. “Role of a Novel Type of Double Infection in the Geminivirus-Induced Epidemic of Severe Cassava Mosaic in Uganda.” *Annals of Applied Biology* 131 (3): 437–48.

https://doi.org/10.1111/j.1744-7348.1997.tb05171.x.

Hartigan, John A. 1975. *Clustering Algorithms*. 99th ed. USA: John Wiley & Sons, Inc.

Hartung, Carl, Adam Lerer, Yaw Anokwa, Clint Tseng, Waylon Brunette, and Gaetano Borriello. 2010. “Open Data Kit: Tools to Build Information Services for Developing Regions.” In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*. ICTD ’10. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2369220.2369236.

Hastie, T. J., and D. Pregibon. 1992. “Generalized Linear Models.” In *Statistical Models*, edited by J.M. Chambers and T. J. Hastie, Wadsworth. Cole.

He, Jianwen. 2013. “Practical Guide to ELISA Development.” In *The Immunoassay Handbook*, 381–93. Elsevier. https://doi.org/10.1016/B978-0-08-097037-0.00025-7.

Heikrujam, J., R. Kishor, and M.P. Behari. 2020. “The Chemistry Behind Plant DNA Isolation Protocols.” In *Biochemical Analysis Tools - Methods for Bio-Molecules Studies*, edited by Oana-Maria Boldura, Cornel Baltă, and Nasser Sayed Awwad. Rijeka: IntechOpen. https://doi.org/10.5772/intechopen.92206.

Hiatt, A, R Cafferkey, and K Bowdish. 1989. “Production of Antibodies in Transgenic Plants.” *Nature* 342 (6245): 76—78. https://doi.org/10.1038/342076a0.

Hillocks, and D. L. Jennings. 2003. “Cassava Brown Streak Disease: A Review of Present Knowledge and Research Needs.” *International Journal of Pest Management* 49 (3): 225–34. https://doi.org/10.1080/0967087031000101061.

Hillocks, R. J., M. D. Raya, and J. M. Thresh. 1999. “Distribution and Symptom Expression of Cassava Brown Streak Disease in Southern Tanzania.” *African Journal of Root and Tuber Crops* 3: 57–61.

Hillocks, R. J., J. M. Thresh, J. Tomas, M. Botao, R. Macia, and R. Zavier. 2002. “Cassava Brown Streak Disease in Northern Mozambique.” *International Journal of Pest Management* 48 (3): 178–81. https://doi.org/10.1080/09670870110087376.

Hillocks, R J. 2004. “Research Protocols for Cassava Borwn Streak Disease.” Chatham Maritime, Kent ME4 4TB, UK. http://www.fao.org/docs/eims/upload/agrotech/1985/r8227_protocols.pdf.

Hillocks, R, and M Maruthi. 2015. “Post-Harvest Impact of Cassava Brown Streak Disease in Four Countries in Eastern Africa.” *Food Chain* 5 (1–2): 116–22. https://doi.org/10.3362/2046-1887.2015.008.

Hillocks, R, and J M Thresh. 2000. “CASSAVA MOSAIC AND CASSAVA BROWN STREAK VIRUS DISEASES IN AFRICA: A Comparative Guide to

Symptoms and Aetiologies." *Roots* 7 (1): 8.

Hillocks, M. D. Raya, K. Mtunda, and H. Kiozia. 2001. "Effects of Brown Streak Virus Disease on Yield and Quality of Cassava in Tanzania." *Journal of Phytopathology* 149 (7–8): 389–94. https://doi.org/10.1046/j.1439-0434.2001.00641.x.

Hillocks, Rj, and Dl Jennings. 2003. "Cassava Brown Streak Disease: A Review of Present Knowledge and Research Needs." *International Journal of Pest Management* 49 (3): 225–34. https://doi.org/10.1080/0967087031000101061.

Horsfall, J., and A.E. Dimond. 1960. *Plant Pathology: An Advanced Treatrise. Volume II: The Pathogen*. Edited by J Horsfall and A.E. Dimond. Academic P. New York and London. https://www.cabdirect.org/cabdirect/abstract/19601603373.

Hosmer, David W., and Stanley Lemeshow. 2000. *Applied Logistic Regression. Second Edition*. Edited by Noel A.C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J.B. Kadane, David W. Scott, Bernard W. Silverman, Adrian F.M. Smith, et al. *Journal of Environmental Health*. JOHN WILEY. Vol. 70. New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

Howeler, R, N Lutaladio, and G Thomas. 2013. "Save and Grow: Cassava. A Guide to Sustainable Production Intensification." FAO.

Hsieh, Kuangwen, Peter L. Mage, Andrew T. Csordas, Michael Eisenstein, and H. Tom Soh. 2014. "Simultaneous Elimination of Carryover Contamination and Detection of DNA with Uracil-DNA-Glycosylase-Supplemented Loop-Mediated Isothermal Amplification (UDG-LAMP)." *Chemical Communications* 50 (28): 3747–49. https://doi.org/10.1039/c4cc00540f.

Hull, R. 2009. *Comparative Plant Virology*. Elsevier Science. https://books.google.be/books?id=UPCP-mVYj8wC.

Hulo, Chantal, Edouard de Castro, Patrick Masson, Lydie Bougueleret, Amos Bairoch, Ioannis Xenarios, and Philippe Le Mercier. 2011. "ViralZone: A Knowledge Resource to Understand Virus Diversity." *Nucleic Acids Research* 39 (Database issue): D576-82. https://doi.org/10.1093/nar/gkq901.

Husson, François, Sébastien Lê, and Jérôme Pagès. 2011. *Exploratory Multivariate Analysis by Example Using R*. *Computer Science & Data Analysis*. CRC Press. London, UK. https://doi.org/10.1080/02664763.2012.657409.

INS - RDC. 2021. "Annuaire Statistique RDC 2020." Kinshasa.

IPA. 2020. "Inspection Provinciale de L'Agriculture: Rapport Annuel." Sud-Kivu, D. R. of Congo.

IPPC. 2021. "International Standards for Phytosanitary Measures. ISPM 02: Framework for Pest Risk Analysis." Rome.

https://www.ippc.int/fr/publications/592/.

Jacofsky, David, Emilia M. Jacofsky, and Marc Jacofsky. 2020. "Understanding Antibody Testing for COVID-19." *The Journal of Arthroplasty* 35 (7): S74–81. https://doi.org/10.1016/j.arth.2020.04.055.

Jain, Miten, Ian T Fiddes, Karen H Miga, Hugh E Olsen, Benedict Paten, and Mark Akeson. 2015. "Improved Data Analysis for the MinION Nanopore Sequencer." *Nature Methods* 12 (4): 351–56. https://doi.org/10.1038/nmeth.3290.

James, Amy M., Susan E. Seal, Andy M. Bailey, and Gary D. Foster. 2021. "Viral Inosine Triphosphatase: A Mysterious Enzyme with Typical Activity, but an Atypical Function." *Molecular Plant Pathology*, no. October 2020: 382–89. https://doi.org/10.1111/mpp.13021.

James, Heather E., K. Ebert, R. McGonigle, Scott M. Reid, Neil Boonham, Jennifer A. Tomlinson, Geoffrey H. Hutchings, et al. 2010. "Detection of African Swine Fever Virus by Loop-Mediated Isothermal Amplification." *Journal of Virological Methods* 164 (1–2): 68–74. https://doi.org/10.1016/j.jviromet.2009.11.034.

Jameson, J D. 1964. "Cassava Mosaic Disease in Uganda." *East African Agricultural and Forestry Journal* 29 (3): 208–13. https://doi.org/10.1080/00128325.1964.11661927.

Jarvis, Andy, Julian Ramirez-Villegas, Beatriz Vanessa Herrera Campo, and Carlos Navarro-Racines. 2012. "Is Cassava the Answer to African Climate Change Adaptation?" *Tropical Plant Biology* 5 (1): 9–29. https://doi.org/10.1007/s12042-012-9096-7.

Jennings, D L. 1957. "Further Studies in Breeding Cassava for Virus Resistance." *The East African Agricultural Journal* 22 (4): 213–19.

Jeong, Ki Jun, Seung Hoon Jang, and Natarajan Velmurugan. 2011. "Recombinant Antibodies: Engineering and Production in Yeast and Bacterial Hosts." *Biotechnology Journal* 6 (1): 16–27. https://doi.org/10.1002/biot.201000381.

Jevšnik, Monika, Lara Lusa, Tina Uršič, Urška Glinšek Biškup, and Miroslav Petrovec. 2020. "Detection of Herpes Simplex and Varicella-Zoster Virus from Skin Lesions: Comparison of RT-PCR and Isothermal Amplification for Rapid Identification." *Diagnostic Microbiology and Infectious Disease* 97 (2): 115015. https://doi.org/https://doi.org/10.1016/j.diagmicrobio.2020.115015.

Ji, Hyung Chung, Hyun Young Park, Ho Lee Jong, and Yangsoo Jang. 2002. "Identification of the DITP- and XTP-Hydrolyzing Protein from Escherichia Coli." *Journal of Biochemistry and Molecular Biology* 35 (4): 403–8. https://doi.org/10.5483/bmbrep.2002.35.4.403.

Jones, W.O. 1959. *Manioc in Africa.* Edited by Stanford University Food Research Institute Studies in Tropical Development. Stanford U. Stanford California:

Stanford University Press.

Jones, William O. 1957. "Manioc : An Example of Innovation in African Economies." *Economic Development and Cultural Change* 5 (2): 97–117.

Jung, Sun, Kwon Meilan, Jin In, Sook Cho, Ju Yeon, Yoon Gug, Seoun Choi, T Ag, and A Ag. 2018. "Identification of Rehmannia Virus 1 , a Novel Putative Member of the Genus Closterovirus , from Rehmannia Glutinosa." *Archives of Virology*, 2–7. https://doi.org/10.1007/s00705-018-4014-4.

Kaponi, Maria S, Teruo Sano, and Panayota E Kyriakopoulou. 2022. "Tissue Print Hybridization for Detection and Characterization of Viroids." In *Viroids: Methods and Protocols*, edited by Ayala L N Rao, Irene Lavagi-Craddock, and Georgios Vidalakis, 111–31. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-1464-8_11.

Karasev, Alexander V. 2000. "Genetic Diversity and Evolution of Closteroviruses." *Annual Review of Phytopathology* 38: 293–324.

Karasev, Alexander V, Indranil Dasgupta, Marc Fuchs, Toru Iwanami, Karel Petrzik, Hélène Sanfaçon, Jeremy R. Thompson, et al. 2017. "ICTV Virus Taxonomy Profile: Secoviridae." *Journal of General Virology* 98: 529–31.

Karst, Søren M., Ryan M. Ziels, Rasmus H. Kirkegaard, Emil A. Sørensen, Daniel McDonald, Qiyun Zhu, Rob Knight, and Mads Albertsen. 2021. "High-Accuracy Long-Read Amplicon Sequences Using Unique Molecular Identifiers with Nanopore or PacBio Sequencing." *Nature Methods* 18 (2): 165–69. https://doi.org/10.1038/s41592-020-01041-y.

Kassambara, Alboukadel. 2017. *Multivariate Analysis II: Practical Guide to Principal Component Methods in R*. STHDA. http://www.sthda.com.

Kaweesi, Tadeo, Robert Kawuki, Vincent Kyaligonza, Yona Baguma, Geoffrey Tusiime, and Morag E Ferguson. 2014. "Field Evaluation of Selected Cassava Genotypes for Cassava Brown Streak Disease Based on Symptom Expression and Virus Load." *Virology Journal* 11: 216. https://doi.org/10.1186/s12985-014-0216-x.

Knierim, Dennis, Wulf Menzel, and Stephan Winter. 2017. "Analysis of the Complete Genome Sequence of Euphorbia Ringspot Virus, an Atypical Member of the Genus Potyvirus." *Archives of Virology* 162 (1): 291–93. https://doi.org/10.1007/s00705-016-3087-1.

Kolm, Claudia, Roland Martzy, Manuela Führer, Robert L Mach, Rudolf Krska, Sabine Baumgartner, Andreas H Farnleitner, and Georg H Reischer. 2019. "Detection of a Microbial Source Tracking Marker by Isothermal Helicase-Dependent Amplification and a Nucleic Acid Lateral-Flow Strip Test." *Scientific Reports* 9 (1): 393. https://doi.org/10.1038/s41598-018-36749-7.

Koloniuk, Igor, Thanuja Thekke-Veetil, Jean Sébastien Reynard, Irena Mavrič Pleško, Jaroslava Přibylová, Justine Brodard, Isabelle Kellenberger, et al. 2018. “Molecular Characterization of Divergent Closterovirus Isolates Infecting Ribes Species.” *Viruses* 10 (7). https://doi.org/10.3390/v10070369.

Koonin, Eugene V, Valerian V Dolja, and T Jack Morris. 1993. “Evolution and Taxonomy of Positive-Strand RNA Viruses: Implications of Comparative Analysis of Amino Acid Sequences.” *Critical Reviews in Biochemistry and Molecular Biology* 28 (5): 375–430. https://doi.org/10.3109/10409239309078440.

Kreuze, Jan F., Ana Perez, Milton Untiveros, Dora Quispe, Segundo Fuentes, Ian Barker, and Reinhard Simon. 2009. “Complete Viral Genome Sequence and Discovery of Novel Viruses by Deep Sequencing of Small RNAs: A Generic Method for Diagnosis, Discovery and Sequencing of Viruses.” *Virology* 388 (1): 1–7. https://doi.org/10.1016/j.virol.2009.03.024.

Kroschel, J, M Sporleder, and P Carhuapoma. 2016. “Pest Distribution and Risk Atlas for Africa.” In *Potential Global and Regional Distribution and Abundance of Agricultural and Horticultural Pests and Associated Biocontrol Agents under Current and Future Climates*, edited by J Kroschel, N Mujica, P Carhuapoma, and M Sporleder, 7–23. International Potato Center.

Kufferath, H, and J Ghesquière. 1932. “La Mosaïque Du Manioc.” *C. R. Soc. Biol. Belge* 109: 1146 – 1148. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0006826098&partnerID=40&md5=4c59ebebb4750206685296fc58e4c3fd.

Kumar, P. L. and Legg, J. P. 2009. “Laboratory Manual for the Diagnosis of Cassava Virus Diseases.” *Tropical Agriculture*, 111. http://biblio.iita.org/documents/U09ManKumarLaboratoryNothomNodev.pdf-fbeb6a750d31604ab1e6f09e3856f081.pdf.

Kumar, P. Lava, Maritza Cuervo, J. F. Kreuze, Giovanna Muller, Gururaj Kulkarni, Safaa G. Kumari, Sebastien Massart, et al. 2021. “Phytosanitary Interventions for Safe Global Germplasm Exchange and the Prevention of Transboundary Pest Spread: The Role of Cgiar Germplasm Health Units.” *Plants* 10 (2): 1–29. https://doi.org/10.3390/plants10020328.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.” *Molecular Biology and Evolution* 35 (6): 1547–49. https://doi.org/10.1093/molbev/msy096.

Kwibuka, Yves, Espoir Bisimwa, Arnaud G. Blouin, Claude Bragard, Thierry Candresse, Chantal Faure, Denis Filloux, et al. 2021. “Novel Ampeloviruses Infecting Cassava in Central Africa and the South-West Indian Ocean Islands.” *Viruses* 13 (6): 1–17. https://doi.org/10.3390/v13061030.

Kwibuka, Yves, Chantal Nyirakanani, Jean Pierre Bizimana, Espoir Bisimwa, Yves Brostaux, Ludivine Lassois, Herve Vanderschuren, and Sebastien Massart. 2022. "Risk Factors Associated with Cassava Brown Streak Disease Dissemination through Seed Pathways in Eastern D.R. Congo." *Frontiers in Plant Science* 13 (July): 1–18. https://doi.org/10.3389/fpls.2022.803980.

Kyte, J, and R F Doolittle. 1982. "A Simple Method for Displaying the Hydropathic Character of a Protein." *Journal of Molecular Biology* 157 (1): 105–32. https://doi.org/10.1016/0022-2836(82)90515-0.

Langlands, B W. 1966. "Cassava in Uganda, 1860-1920." *Uganda Journal* 30 (2): 211–18. https://www.africabib.org/rec.php?RID=191289981.

Lebart, L, A Morineau, and M Piron. 1995. "Statistique Exploratoire Multidimensionnelle." *Statistique Exploratoire Multidimensionnelle*, 439.

Lebas, B, S Massart, D Kutnjak, and N Mehle. 2020. *Guidelines for the Selection, Development, Validation and Routine Use of High-Throughput Sequencing Analysis in Plant Health Diagnostic Laboratories: Grant Agreement N. 773139 : Deliverable N° 2.2*. Valitest EU project consortium. https://books.google.be/books?id=PbKnzgEACAAJ.

Lebas, Benedicte, Ian Adams, Maher Al Rwahnih, Steve Baeyen, Guillaume J. Bilodeau, Arnaud G Blouin, Neil Boonham, et al. 2022. "Facilitating the Adoption of High-throughput Sequencing Technologies as a Plant Pest Diagnostic Test in Laboratories: A Step-by-step Description." *EPPO Bulletin* 52 (2): 394–418. https://doi.org/10.1111/epp.12863.

Lebot, V. 2009. *Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids*. Crop Production Science in Horticulture, 17. CABI. https://books.google.be/books?id=rFwyrKRSMUMC.

Lefeuvre, Pierre, Darren P. Martin, Santiago F. Elena, Dionne N. Shepherd, Philippe Roumagnac, and Arvind Varsani. 2019. "Evolution and Ecology of Plant Viruses." *Nature Reviews Microbiology* 17 (10): 632–44. https://doi.org/10.1038/s41579-019-0232-3.

Legg, Christopher. 2008. "The Use of Geographic Information Systems (GIS) in Crop Disease Management- The C3P Crop Crisis Control Experience. Brief Report." https://www.crs.org/sites/default/files/tools-research/c3p-use-of-gis-in-crop-disease-management.pdf.

Legg, J. P. 1999. "Emergence, Spread and Strategies for Controlling the Pandemic of Cassava Mosaic Virus Disease in East and Central Africa." *Crop Protection* 18 (10): 627–37. https://doi.org/10.1016/S0261-2194(99)00062-9.

Legg, J. P., and C. M. Fauquet. 2004. "Cassava Mosaic Geminiviruses in Africa." *Plant Molecular Biology* 56 (4): 585–99. https://doi.org/10.1007/s11103-004-1651-7.

Legg, J. P., S. C. Jeremiah, H. M. Obiero, M. N. Maruthi, I. Ndyetabula, G. Okao-Okuja, H. Bouwmeester, et al. 2011. "Comparing the Regional Epidemiology of the Cassava Mosaic and Cassava Brown Streak Virus Pandemics in Africa." *Virus Research* 159 (2): 161–70. https://doi.org/10.1016/j.virusres.2011.04.018.

Legg, J. P., and M. D. Raya. 1998. "Survey of Cassava Virus Diseases in Tanzania." *International Journal of Pest Management* 44 (1): 17–23. https://doi.org/10.1080/096708798228473.

Legg, J.P., B. Owor, P. Sseruwagi, and J. Ndunguru. 2006. "Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of a Regional Pandemic." *Advanced Virus Research*, no. 67: 355–418.

Legg, James, Lava Kumar, and Edward Eneah Kanju. 2015. "Cassava Brown Streak." In *Virus Diseases of Tropical and Subtropical Crops*, edited by P. Tennant and G. Fermin, 15. Wallingford, United Kingdom: CAB International.

Legg, James P., P. Lava Kumar, T. Makeshkumar, Leena Tripathi, Morag Ferguson, Edward Kanju, Pheneas Ntawuruhunga, and Wilmer Cuellar. 2015. "Cassava Virus Diseases: Biology, Epidemiology, and Management." In *Advances in Virus Research*, 91:85–142. https://doi.org/10.1016/bs.aivir.2014.10.001.

Legg, James P, P Lava Kumar, T Makeshkumar, Leena Tripathi, Morag Ferguson, Edward Kanju, Pheneas Ntawuruhunga, and Wilmer Cuellar. 2015. "Cassava Virus Diseases: Biology, Epidemiology, and Management." *Advances in Virus Research* 91. https://doi.org/10.1016/bs.aivir.2014.10.001.

Legg, James, Eklou Attiogbevi Somado, Ian Barker, Larry Beach, Hernan Ceballos, Willmer Cuellar, Warid Elkhoury, et al. 2014. "A Global Alliance Declaring War on Cassava Viruses in Africa." *Food Security* 6. https://doi.org/10.1007/s12571-014-0340-x.

Legg, and M.J. Thresh. 2003. "Cassava Virus Diseases in Africa, in Plant Virology in Sub-Saharan Africa." In *Plant Virology in Sub-Saharan Africa*, 517–52.

Leinonen, Rasko, Hideaki Sugawara, and Martin Shumway. 2011. "The Sequence Read Archive." *Nucleic Acids Research* 39 (SUPPL. 1): 2010–12. https://doi.org/10.1093/nar/gkq1019.

Leiva, Ana M., Jenyfer Jimenez, Hector Sandoval, Shirley Perez, and Wilmer J. Cuellar. 2022a. "Complete Genome Sequence of a Novel Secovirid Infecting Cassava in the Americas." *Archives of Virology*, no. 0123456789: 10–13. https://doi.org/10.1007/s00705-021-05325-2.

———. 2022b. "Complete Genome Sequence of a Novel Secovirid Infecting Cassava in the Americas." *Archives of Virology* 167 (2): 665–68. https://doi.org/10.1007/s00705-021-05325-2.

Lennon, A. M., M. M. Aiton, and B. D. Harrison. 1987. "Purification and Properties of Cassava Green Mottle, a Previously Undescribed Virus from the Solomon

Islands." *Annals of Applied Biology : An International Journal of the AAB.* 110 (3): 545–55.

Lenth, Russell V. 2016. "Least-Squares Means: The R Package Lsmeans." *Journal of Statistical Software* 69 (1). https://doi.org/10.18637/jss.v069.i01.

Levy, Shawn E., and Richard M. Myers. 2016. "Advancements in Next-Generation Sequencing." *Annual Review of Genomics and Human Genetics* 17: 95–115. https://doi.org/10.1146/annurev-genom-083115-022413.

Li, Shi-Yuan, Qiu-Xiang Cheng, Jing-Man Wang, Xiao-Yan Li, Zi-Long Zhang, Song Gao, Rui-Bing Cao, Guo-Ping Zhao, and Jin Wang. 2018. "CRISPR-Cas12a-Assisted Nucleic Acid Detection." *Cell Discovery* 4 (1): 20. https://doi.org/10.1038/s41421-018-0028-z.

Li, X H, L A Heaton, T J Morris, and A E Simon. 1989. "Turnip Crinkle Virus Defective Interfering RNAs Intensify Viral Symptoms and Are Generated de Novo." *Proceedings of the National Academy of Sciences* 86 (23): 9173–77. https://doi.org/10.1073/pnas.86.23.9173.

Li, Xiang-Hui, Xiao-Ling Zhang, Juan Wu, Ni Lin, Wei-Ming Sun, Min Chen, Qi-Shui Ou, and Zhen-Yu Lin. 2019. "Hyperbranched Rolling Circle Amplification (HRCA)-Based Fluorescence Biosensor for Ultrasensitive and Specific Detection of Single-Nucleotide Polymorphism Genotyping Associated with the Therapy of Chronic Hepatitis B Virus Infection." *Talanta* 191 (January): 277–82. https://doi.org/10.1016/j.talanta.2018.08.064.

Lin, Shengrong, Alexander G. McLennan, Kang Ying, Zhao Wang, Shaohua Gu, Hua Jin, Chaoqun Wu, et al. 2001. "Cloning, Expression, and Characterization of a Human Inosine Triphosphate Pyrophosphatase Encoded by the ITPA Gene." *Journal of Biological Chemistry* 276 (22): 18695–701. https://doi.org/10.1074/jbc.M011084200.

Liu, Fu Guo, Guo Fu Chen, Chun Yun Zhang, Yuan Yuan Wang, and Jin Zhou. 2019. "Exponential Rolling Circle Amplification Coupled with Lateral Flow Dipstick Strips as a Rapid and Sensitive Method for the Field Detection of Karlodinium Veneficum." *Journal of Applied Phycology* 31 (4): 2423–36. https://doi.org/10.1007/s10811-019-01762-4.

Lizak, Christian, Yao Yun Fan, Thomas Christian Weber, and Markus Aebi. 2011. "N-Linked Glycosylation of Antibody Fragments in Escherichia Coli." *Bioconjugate Chemistry* 22 (3): 488–96. https://doi.org/10.1021/bc100511k.

Lobato, Ivan Magriñá, and Ciara K. O'Sullivan. 2018. "Recombinase Polymerase Amplification: Basics, Applications and Recent Advances." *TrAC - Trends in Analytical Chemistry* 98: 19–35. https://doi.org/10.1016/j.trac.2017.10.015.

Lobell, David B, Marshall B Burke, Claudia Tebaldi, Michael D Mastrandrea, Walter P Falcon, and Rosamond L Naylor. 2008. "Prioritizing Climate Change

Adaptation Needs for Food Security in 2030." *Science* 319 (5863): 607–10. https://doi.org/10.1126/science.1152339.

Lopez, Raphael-Garcia, Irene Gonzalo, Juan-Antonio Garcia, and Adrian A Valli. 2022. "Cassava and Ugandan Cassava Brown Streak Virus, an Attractive Case of Plant-Pathogen Coevolution." In *III Reunion de La Red Nacional de Virologia de Plantas*, edited by Pedro-Gomez Lopez, 78. Murcia, Spain: RENAVIPLANT.

Louwaars, Neils, Philippe Le Coent, and Tom. Osborn. 2010. "Seed Systems and Plant Genetic Resources for Food and Agriculture." *Organization*, 1–24.

Lozano, J.C., and R.H. Booth. 1974. "Cassava Diseases and Pests." *PANS (Pest .Arctic. News Summ.)* 20: 40–45.

Ma, Hui, and Richard O'Kennedy. 2015. "The Structure of Natural and Recombinant Antibodies." *Methods in Molecular Biology (Clifton, N.J.)* 1348: 7–11. https://doi.org/10.1007/978-1-4939-2999-3_2.

Maclot, François. 2021. "Exploring Virus Ecology and Diversity in Cultivated and Wild Poacea Communities in Belgium." University of Liège.

Madeira, Fábio, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, et al. 2019. "The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019." *Nucleic Acids Research* 47 (W1): W636—W641. https://doi.org/10.1093/nar/gkz268.

Mahungu, N M, K W Tata Hangy, S M Bidiaka, and Antoine Frangoie. 2014. "Multiplication de Materiel de Plantation de Manioc et Gestion Des Maladies et Ravageurs." Kinshasa/R.D. Congo.

Mak, Wing Cheung, Valerio Beni, and Anthony P.F. Turner. 2016. "Lateral-Flow Technology: From Visual to Instrumental." *TrAC - Trends in Analytical Chemistry* 79: 297–305. https://doi.org/10.1016/j.trac.2015.10.017.

Manani, Duke, Elijah Ateka, Steven Nyanjom, and Laura Boykin. 2017. "Phylogenetic Relationships among Whiteflies in the Bemisia Tabaci (Gennadius) Species Complex from Major Cassava Growing Areas in Kenya." *Insects* 8 (1): 25. https://doi.org/10.3390/insects8010025.

Mann, Krin S., Melanie Walker, and Hélène Sanfaçon. 2017. "Identification of Cleavage Sites Recognized by the 3C-like Cysteine Protease within the Two Polyproteins of Strawberry Mottle Virus." *Frontiers in Microbiology* 8 (APR): 1–13. https://doi.org/10.3389/fmicb.2017.00745.

Manze, Francis, Patrick Rubaihayo, Alfred Ozimati, Paul Gibson, Williams Esuma, Anton Bua, Titus Alicai, Chris Omongo, and Robert S. Kawuki. 2021. "Genetic Gains for Yield and Virus Disease Resistance of Cassava Varieties Developed Over the Last Eight Decades in Uganda." *Frontiers in Plant Science* 12 (June): 1–11. https://doi.org/10.3389/fpls.2021.651992.

Marais, A, C Faure, B Bergey, and Thierry Candresse. 2018. "Viral Double-Stranded RNAs (DsRNAs) from Plants: Alternative Nucleic Acid Substrates for High-Throughput Sequencing." In *Viral Metagenomics: Methods and Protocols, Methods in Molecular Biology*, edited by Vitantonio Pantaleo and Michela Chiumenti, 1746:45–53. New York, NY: Springer Science+Business Media. https://doi.org/10.1007/978-1-4939-7683-6.

Margis, R, and L Pinck. 1992. "Effects of Site-Directed Mutagenesis on the Presumed Catalytic Triad and Substrate-Binding Pocket of Grapevine Fanleaf Nepovirus 24-KDa Proteinase." *Virology* 190 (2): 884–88. https://doi.org/10.1016/0042-6822(92)90931-e.

Martelli, G P, A A Agranovsky, Moshe Bar-joseph, D Boscia, T Candresse, R H A Coutts, V V Dolja, John S Hu, W Jelkmann, and Alexander V Karasev. 2012. "Family Closteroviridae." In *Ninth Report of the International Committee on Taxonomy of Viruses*, edited by A.M.Q. King, M.J. Adams, E.B. Carstens, and E.J. Lefkowitz, Elsevier A, 987–1001. New York, NY, USA. https://doi.org/10.1016/B978-0-12-384684-6.00085-9.

Martelli, Giovanni P., Michael J. Adams, Jan F. Kreuze, and Valerian V. Dolja. 2008. "Family Flexiviridae: A Case Study in Virion and Genome Plasticity." *Annual Review of Phytopathology* 45: 73–100. https://doi.org/10.1146/annurev.phyto.45.062806.094401.

Martin, Darren P., Ben Murrell, Michael Golden, Arjun Khoosal, and Brejnev Muhire. 2015. "RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes." *Virus Evolution* 1 (1): 1–5. https://doi.org/10.1093/ve/vev003.

Martin, E.F. 1928. "Report of the Mycologist. Annual Report of the Department of Agriculture." Uganda.

Maruthi, M. N., R. J. Hillocks, K. Mtunda, M. D. Raya, M. Muhanna, H. Kiozia, A. R. Rekha, J. Colvin, and J. M. Thresh. 2005. "Transmission of Cassava Brown Streak Virus by Bemisia Tabaci (Gennadius)." *Journal of Phytopathology* 153 (5): 307–12. https://doi.org/10.1111/j.1439-0434.2005.00974.x.

Maruthi, M.N. 2015. "Limiting the Impact of Cassava Brown Streak Disease on Smalholders, Womens and the Cassava Value Chain, Interim Project Narrative Report Submitted to the African Union Commission." Addis Abeba, Ethiopia.

Maruthi, M.N., Simon C. Jeremiah, Ibrahim U. Mohammed, and James P. Legg. 2017. "The Role of the Whitefly, Bemisia Tabaci (Gennadius), and Farmer Practices in the Spread of Cassava Brown Streak Ipomoviruses." *Journal of Phytopathology* 165 (11–12): 707–17. https://doi.org/10.1111/jph.12609.

Masoudi-Nejad, Ali, Zahra Narimani, and Nazanin Hosseinkhan. 2013. *Next Generation Sequencing and Sequence Assembly*. *Media*. Vol. 4. SpringerBriefs in Systems Biology. New York, NY: Springer New York.

https://doi.org/10.1007/978-1-4614-7726-6.

Massart, Sebastien, Ian Adams, Maher Al Rwahnih, Steve Baeyen, Guillaume J. Bilodeau, Arnaud G Blouin, Neil Boonham, et al. 2022. “Guidelines for the Reliable Use of High Throughput Sequencing Technologies to Detect Plant Pathogens and Pests.” *Peer Community Journal* 2 (October): e62. https://doi.org/10.24072/pcjournal.181.

Massart, Sebastien, Thierry Candresse, José Gil, Christophe Lacomme, Lukas Predajna, Maja Ravnikar, Jean Sébastien Reynard, et al. 2017. “A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies.” *Frontiers in Microbiology* 8 (JAN). https://doi.org/10.3389/fmicb.2017.00045.

Massart, Sebastien, Michela Chiumenti, Kris De Jonghe, Rachel Glover, Annelies Haegeman, Igor Koloniuk, Petr Komínek, et al. 2019. “Virus Detection by High-Throughput Sequencing of Small RNAs: Large-Scale Performance Testing of Sequence Analysis Strategies.” *Phytopathology* 109 (3): 488–97. https://doi.org/10.1094/PHYTO-02-18-0067-R.

Massart, Sebastien, Antonio Olmos, Haissam Jijakli, and Thierry Candresse. 2014. “Current Impact and Future Directions of High Throughput Sequencing in Plant Virus Diagnostics.” *Virus Research* 188: 90–96. https://doi.org/10.1016/j.virusres.2014.03.029.

Maxam, A M, and W Gilbert. 1977. “A New Method for Sequencing DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 74 (2): 560–64. https://doi.org/10.1073/pnas.74.2.560.

Mbanzibwa, D R, Y P Tian, A K Tugume, S B Mukasa, F Tairo, S Kyamanywa, A Kullaya, and J P T Valkonen. 2011. “Simultaneous Virus-Specific Detection of the Two Cassava Brown Streak-Associated Viruses by RT-PCR Reveals Wide Distribution in East Africa, Mixed Infections, and Infections in Manihot Glaziovii.” *Journal of Virological Methods* 171 (2): 394–400. https://doi.org/10.1016/j.jviromet.2010.09.024.

Mbanzibwa, Y. P. Tian, a. K. Tugume, S. B. Mukasa, F. Tairo, S. Kyamanywa, A. Kullaya, and Jari P T Valkonen. 2009. “Genetically Distinct Strains of Cassava Brown Streak Virus in the Lake Victoria Basin and the Indian Ocean Coastal Area of East Africa.” *Archives of Virology* 154 (2): 353–59. https://doi.org/10.1007/s00705-008-0301-9.

Mbanzibwa, Yanping Tian, Settumba B Mukasa, and Jari P T Valkonen. 2009. “Cassava Brown Streak Virus (Potyviridae) Encodes a Putative Maf/HAM1 Pyrophosphatase Implicated in Reduction of Mutations and a P1 Proteinase That Suppresses RNA Silencing but Contains No HC-Pro.” *Journal of Virology* 83 (13): 6934–40. https://doi.org/10.1128/JVI.00537-09.

Mbewe, Willard, Fred Tairo, Peter Sseruwagi, Joseph Ndunguru, Siobain Duffy, Ssetumba Mukasa, Ibrahim Benesi, Samar Sheat, Marianne Koerbler, and Stephan Winter. 2017. "Variability in P1 Gene Redefines Phylogenetic Relationships among Cassava Brown Streak Viruses." *Virology Journal* 14 (1): 1–7. https://doi.org/10.1186/s12985-017-0790-9.

McEwan, Margaret A, Conny JM Almekinders, Jorge JL Andrade-Piedra, Erik Delaquis, Karen A Garrett, Lava Kumar, Sarah Mayanja, Bonaventure A Omondi, Srinivasulu Rajendran, and Graham Thiele. 2021. "'Breaking through the 40% Adoption Ceiling: Mind the Seed System Gaps.' A Perspective on Seed Systems Research for Development in One CGIAR." *Outlook on Agriculture*, 003072702198934. https://doi.org/10.1177/0030727021989346.

McGuire, Shawn, and Louise Sperling. 2016. "Seed Systems Smallholder Farmers Use." *Food Security* 8 (1): 179–95. https://doi.org/10.1007/s12571-015-0528-8.

McQuaid, C. F., P. Sseruwagi, A. Pariyo, and F. van den Bosch. 2016. "Cassava Brown Streak Disease and the Sustainability of a Clean Seed System." *Plant Pathology* 65 (2): 299–309. https://doi.org/10.1111/ppa.12453.

Melzer, M. J., D. M. Sether, A. V. Karasev, W. Borth, and J. S. Hu. 2008. "Complete Nucleotide Sequence and Genome OrgangRNA Ization of Pineapple Mealybug Wilt-Associated Virus-1." *Archives of Virology* 153 (4): 707–14. https://doi.org/10.1007/s00705-008-0051-8.

Mendiburu, Felipe de, and Mohammad Yassen. 2020. "Agricolae: Statistical Procedures for Agricultural Research.R Package Version 1.4.0." https://myaseen208.github.io/agricolae/.

Menzel, Wulf, Reinhard Goetz, Dietrich E. Lesemann, and H. Josef Vetten. 2009. "Molecular Characterization of a Closterovirus from Carrot and Its Identification as a German Isolate of Carrot Yellow Leaf Virus." *Archives of Virology* 154 (8): 1343–47. https://doi.org/10.1007/s00705-009-0428-3.

Mezger, Anja, Christina Öhrmalm, David Herthnek, Jonas Blomberg, and Mats Nilsson. 2014. "Detection of Rotavirus Using Padlock Probes and Rolling Circle Amplification." *PLOS ONE* 9 (11): 1–6. https://doi.org/10.1371/journal.pone.0111874.

Mickey, J., and S. Greenland. 1989. "A Study of the Impact of Confounder Selection Criteria on Effect Estimation." *American Journal of Epidemiology* 129: 125–37.

Mohammed, I U, M M Abarshi, B Muli, R J Hillocks, and M N Maruthi. 2012. "The Symptom and Genetic Diversity of Cassava Brown Streak Viruses Infecting Cassava in East Africa." *Advances in Virology* 2012 (795697): 1–10. https://doi.org/10.1155/2012/795697.

Mohsen, Michael G., and Eric T. Kool. 2016. "The Discovery of Rolling Circle Amplification and Rolling Circle Transcription." *Accounts of Chemical*

Research 49 (11): 2540–50. https://doi.org/10.1021/acs.accounts.6b00417.

Monde, G., J. Walangululu, S. Winter, and C. Bragard. 2010. "Dual Infection by Cassava Begomoviruses in Two Leguminous Species (Fabaceae) in Yangambi, Northeastern Democratic Republic of Congo." *Archives of Virology* 155 (11): 1865–69. https://doi.org/10.1007/s00705-010-0772-3.

Monger, W. A., S. Seal, S. Cotton, and G. D. Foster. 2001. "Identification of Different Isolates of Cassava Brown Streak Virus and Development of a Diagnostic Test." *Plant Pathology* 50 (6): 768–75. https://doi.org/10.1046/j.1365-3059.2001.00647.x.

Monger, W. A., S. Seal, A. M. Isaac, and G. D. Foster. 2001. "Molecular Characterization of the Cassava Brown Streak Virus Coat Protein." *Plant Pathology* 50 (4): 527–34. https://doi.org/10.1046/j.1365-3059.2001.00589.x.

Monger, W a, S Seal, a M Isaac, and G D Foster. 2001. "Molecular Characterization of the Cassava Brown Streak Virus Coat Protein." *Plant Pathology* 50 (4): 527–34. https://doi.org/Doi 10.1046/J.1365-3059.2001.00589.X.

Mongkolsiriwattana, Chawin, Angel Y.S. Chen, and James C.K. Ng. 2011. "Replication of Lettuce Chlorosis Virus (LCV), a Crinivirus in the Family Closteroviridae, Is Accompanied by the Production of LCV RNA 1-Derived Novel RNAs." *Virology* 420 (2): 89–97. https://doi.org/10.1016/j.virol.2011.08.017.

Moreno, Isabel, Wilhelm Gruissem, and Hervé Vanderschuren. 2011. "Reference Genes for Reliable Potyvirus Quantitation in Cassava and Analysis of Cassava Brown Streak Virus Load in Host Varieties." *Journal of Virological Methods* 177 (1): 49–54. https://doi.org/10.1016/j.jviromet.2011.06.013.

Muhindo, Honoré, Sabrine Yasenge, Clérisse Casinga, Médard Songbo, Benoît Dhed'a, Titus Alicai, Justin Pita, and Godefroid Monde. 2020. "Incidence, Severity and Distribution of Cassava Brown Streak Disease in Northeastern Democratic Republic of Congo." *Cogent Food and Agriculture* 6 (1): 1–13. https://doi.org/10.1080/23311932.2020.1789422.

Mulimbi, W, X Phemba, and B Assumani. 2012. "First Report of Ugandan Cassava Brown Streak Virus on Cassava in Democratic Republic of Congo." *New Disease ...*, 5197. https://doi.org/10.5197/j.2044-0588.2012.026.011.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. "Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction." *Cold Spring Harbor Symposia on Quantitative Biology* 51 (1): 263–73. https://doi.org/10.1101/SQB.1986.051.01.032.

Munganyinka, E., E. M. Ateka, A. W. Kihurani, M. C. Kanyange, F. Tairo, P. Sseruwagi, and J. Ndunguru. 2018. "Cassava Brown Streak Disease in Rwanda, the Associated Viruses and Disease Phenotypes." *Plant Pathology* 67 (2): 377–

87. https://doi.org/10.1111/ppa.12789.

Munthali, D. 1992. "Effect of Cassava Variety on the Biology of Bemisia Afer (Priesner & Hosny) (Hemiptera: Aleyrodidae)." *Insect Science and Its Application* 13 (3): 459–65.

Munyahali, Wivine, Pieter Pypers, Rony Swennen, Jean Walangululu, Bernard Vanlauwe, and Roel Merckx. 2017. "Responses of Cassava Growth and Yield to Leaf Harvesting Frequency and NPK Fertilizer in South Kivu, Democratic Republic of Congo." *Field Crops Research* 214 (September): 194–201. https://doi.org/10.1016/j.fcr.2017.09.018.

Myhrvold, Cameron, Catherine A Freije, Jonathan S Gootenberg, Omar O Abudayyeh, Hayden C Metsky, Ann F Durbin, Max J Kellner, et al. 2018. "Field-Deployable Viral Diagnostics Using CRISPR-Cas13." *Science* 360 (6387): 444–48. https://doi.org/10.1126/science.aas8836.

Naidu, R. A., H. Bottenberg, P. Subrahmanyam, F. M. Kimmins, D. J. Robinson, and J. M. Thresh. 1998. "Epidemiology of Groundnut Rosette Virus Disease: Current Status and Future Research Needs." *Annals of Applied Biology* 132 (3): 525–48. https://doi.org/10.1111/j.1744-7348.1998.tb05227.x.

Ndunguru, Joseph, Peter Sseruwagi, Fred Tairo, Francesca Stomeo, Solomon Maina, Appolinaire Djinkeng, Monica Kehoe, Laura M. Boykin, and Ulrich Melcher. 2015. "Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation." *PLoS ONE* 10 (10): 1–18. https://doi.org/10.1371/journal.pone.0139321.

Nduwumuremyi, Athanase, Rob Melis, Paul Shanahan, and Theodore Asiimwe. 2016. "Participatory Appraisal of Preferred Traits, Production Constraints and Postharvest Challenges for Cassava Farmers in Rwanda." *Food Security* 8 (2): 375–88. https://doi.org/10.1007/s12571-016-0556-z.

Neri, Uri, Yuri I. Wolf, Simon Roux, Antonio Pedro Camargo, Benjamin Lee, Darius Kazlauskas, I. Min Chen, et al. 2022. "Expansion of the Global RNA Virome Reveals Diverse Clades of Bacteriophages." *Cell* 185 (21): 4023-4037.e18. https://doi.org/10.1016/j.cell.2022.08.023.

Newbery, Fay, Aiming Qi, and Bruce DL Fitt. 2016. "Modelling Impacts of Climate Change on Arable Crop Diseases: Progress, Challenges and Applications." *Current Opinion in Plant Biology* 32: 101–9. https://doi.org/10.1016/j.pbi.2016.07.002.

Nichols, R F W. 1950. "The Brown Streak Disease of Cassava: Distribution, Climatic Effects and Diagnostic Symptoms." *The East African Agricultural Journal* 15 (3): 154–60.

Niekerk, Jaci van, and Rachel Wynberg. 2017. "Traditional Seed and Exchange

Systems Cement Social Relations and Provide a Safety Net: A Case Study from KwaZulu-Natal, South Africa." *Agroecology and Sustainable Food Systems* 41 (9–10): 1099–1123. https://doi.org/10.1080/21683565.2017.1359738.

Noskov, V. N., K. Staak, P. V. Shcherbakova, S. G. Kozmin, K. Negishi, B. C. Ono, H. Hayatsu, and Y. I. Pavlov. 1996. "HAM1, the Gene Controlling 6-N-Hydroxylaminopurine Sensitivity and Mutagenesis in the Yeast Saccharomyces Cerevisiae." *Yeast* 12 (1): 17–29. https://doi.org/10.1002/(SICI)1097-0061(199601)12:1<17::AID-YEA875>3.0.CO;2-I.

Notomi, Tsugunori, Hiroto Okayama, Harumi Masubuchi, Toshihiro Yonekawa, Keiko Watanabe, Nobuyuki Amino, and Tetsu Hase. 2000. "Loop-Mediated Isothermal Amplification of DNA." *Nucleic Acids Res* 28 (12): E63. https://doi.org/10.1093/nar/28.12.e63.

Nurk, Sergey, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, et al. 2022. "The Complete Sequence of a Human Genome." *Science* 376 (6588): 44–53. https://doi.org/10.1126/science.abj6987.

Nweke, Felix I, D.S.C. Spencer, and John Lynam. 2002. *The Cassava Transformation: Africa's Best-Kept Secret.* Internatio. East Lansing, United States: Michigan State University press. https://hdl.handle.net/10568/92722.

Ogbe, F. O., Dixon AGO, J. d'A. Hughes, O. J. Alabi, R. Okechukwu, a. G. O. Dixon, J. d'A. Hughes, O. J. Alabi, and R. Okechukwu. 2006. "Status of Cassava Begomoviruses and Their New Natural Hosts in Nigeria." *Plant Disease* 90 (5): 548–53. https://doi.org/10.1094/PD-90-0548.

Okereke, N., J. Tomlinson, I. Adams, L. Laurenson, N. Fairless, N. Boonham, A.F. Okunade, J Onyeka, J Smith, and 1National. 2013. "Protocol for Cost-Effective Detection of Cassava Mosaic Virus" 16 (22): 119–28.

Okigbo, Bede N. 1980. "Nutritional Implications of Projects Giving High Priority to the Production of Staples of Low Nutritive Quality: The Case for Cassava (Manihot Esculenta, Crantz) in the Humid Tropics of West Africa." *Food and Nutrition Bulletin* 2 (4): 1–12. https://doi.org/10.1177/156482658000200404.

Okogbenin, E., C. N. Egesi, B. Olasanmi, O. Ogundapo, S. Kahya, P. Hurtado, J. Marin, et al. 2012. "Molecular Marker Analysis and Validation of Resistance to Cassava Mosaic Disease in Elite Cassava Genotypes in Nigeria." *Crop Science* 52 (6): 2576–86. https://doi.org/10.2135/cropsci2011.11.0586.

Olmos, A., N. Boonham, T. Candresse, P. Gentit, B. Giovani, D. Kutnjak, L. Liefting, et al. 2018. "High-Throughput Sequencing Technologies for Plant Pest Diagnosis: Challenges and Opportunities." *EPPO Bulletin* 48 (2): 219–24. https://doi.org/10.1111/epp.12472.

Olsen, Kenneth M. 2004. "SNPs, SSRs and Inferences on Cassava's Origin." *Plant*

Molecular Biology 56 (4): 517–26. https://doi.org/10.1007/s11103-004-5043-9.

Orek, C. O. 2018. "An Optimised Cetyltrimethylammonium Bromide (CTAB)-Based Protocol for Extracting RNA from Young and Old Cassava Leaves." *Journal of Advances in Biology & Biotechnology* 19 (4): 1–7. https://doi.org/10.9734/JABB/2018/45370.

Otim-Nape, G. W., T. Alicai, and J. M. Thresh. 2001. "Changes in the Incidence and Severity of Cassava Mosaic Virus Disease, Varietal Diversity and Cassava Production in Uganda." *Annals of Applied Biology* 138 (3): 313–27. https://doi.org/10.1111/j.1744-7348.2001.tb00116.x.

Otim-Nape, G. W., J. M. Thresh, and M. W. Shaw. 1997. "The Effects of Cassava Mosaic Virus Disease on Yield and Compensation in Mixed Stands of Healthy and Infected Cassava." *Annals of Applied Biology* 130 (3): 503–21. https://doi.org/10.1111/j.1744-7348.1997.tb07678.x.

Otim-Nape, G.W., and ~et al. 2000. "The Current Pandemic of Casava Mosaic Virus Disease in East Africa and Its Control." Chatham, UK: NARO/NRI/DFID publication.

Otoo, John A. 1994. "Rapid Multiplication of Cassava. Research Guide 51." Ibadan, Nigeria: Internation Institute of tropical agriculture.

Otun, Sarah, Ainoa Escrich, Ikechukwu Achilonu, Molemi Rauwane, Jordy Alexis Lerma-Escalera, José Rubén Morones-Ramírez, and Leonardo Rios-Solis. 2022. "The Future of Cassava in the Era of Biotechnology in Southern Africa." *Critical Reviews in Biotechnology* 0 (0): 1–19. https://doi.org/10.1080/07388551.2022.2048791.

Ou, Tey Putita, Chanvannak Yun, Heidi Auerswald, Saraden In, Rithea Leang, Rekol Huy, Rithy Choeung, Philippe Dussart, and Veasna Duong. 2020. "Improved Detection of Dengue and Zika Viruses Using Multiplex RT-QPCR Assays." *Journal of Virological Methods* 282 (May): 113862. https://doi.org/10.1016/j.jviromet.2020.113862.

Owor, B., J. P. Legg, G. Okao-Okuja, R. Obonyo, and M. W. Ogenga-Latigo. 2004. "The Effect of Cassava Mosaic Geminiviruses on Symptom Severity, Growth and Root Yield of a Cassava Mosaic Virus Disease-Susceptible Cultivar in Uganda." *Annals of Applied Biology* 145 (3): 331–37. https://doi.org/10.1111/j.1744-7348.2004.tb00390.x.

Palanga, Essow, Denis Filloux, Darren P Martin, Emmanuel Fernandez, Zakaria Bouda, Daniel Gargani, Romain Ferdinand, et al. 2016. "Metagenomic-Based Screening and Molecular Characterization of Cowpea- Infecting Viruses in Burkina Faso." *PLoS ONE* 11 (10): 1–21. https://doi.org/10.1371/journal.pone.0165188.

Palanga, Essowè, Denis Filloux, Darren P. Martin, Emmanuel Fernandez, Daniel

Gargani, Romain Ferdinand, Jean Zabré, et al. 2016. "Metagenomic-Based Screening and Molecular Characterization of Cowpea-Infecting Viruses in Burkina Faso." *PLoS ONE* 11 (10): 1–21. https://doi.org/10.1371/journal.pone.0165188.

Papadakis, G., A. K. Pantazis, N. Fikas, S. Chatziioannidou, V. Tsiakalou, K. Michaelidou, V. Pogka, et al. 2022. "Portable Real-Time Colorimetric LAMP-Device for Rapid Quantitative Detection of Nucleic Acids in Crude Samples." *Scientific Reports* 12 (1): 1–15. https://doi.org/10.1038/s41598-022-06632-7.

Parida, Manmohan M., Santhosh Sannarangaiah, Paban Kumar Dash, P. V.L. Rao, and Kouichi Morita. 2008. "Loop Mediated Isothermal Amplification (LAMP): A New Generation of Innovative Gene Amplification Technique; Perspectives in Clinical Diagnosis of Infectious Diseases." *Reviews in Medical Virology* 18 (6): 407–21. https://doi.org/10.1002/rmv.593.

Parker, Jane E., Giuliana Hessler, and Haitao Cui. 2022. "A New Biochemistry Connecting Pathogen Detection to Induced Defense in Plants." *New Phytologist* 234 (3): 819–26. https://doi.org/10.1111/nph.17924.

Parmar, Aditya, Sascha M Kirchner, Henning Langguth, Thomas F Döring, and Oliver Hensel. 2017. "Boxwood Borer Heterobostrychus Brunneus (Coleoptera: Bostrichidae) Infesting Dried Cassava: A Current Record from Southern Ethiopia." *Journal of Insect Science (Online)* 17 (1). https://doi.org/10.1093/jisesa/iew106.

Parmar, Aditya, Barbara Sturm, and Oliver Hensel. 2017. "Crops That Feed the World: Production and Improvement of Cassava for Food, Feed, and Industrial Uses." *Food Security* 9 (5): 907–27. https://doi.org/10.1007/s12571-017-0717-8.

Patil, B. L., J. P. Legg, E. Kanju, and C. M. Fauquet. 2015. "Cassava Brown Streak Disease: A Threat to Food Security in Africa." *Journal of General Virology* 96 (Pt_5). https://doi.org/10.1099/vir.0.000014.

Patil, B L, S Rajasubramaniam, C Bagchi, and I Dasgupta. 2005. "Both Indian Cassava Mosaic Virus and Sri Lankan Cassava Mosaic Virus Are Found in India and Exhibit High Variability as Assessed by PCR-RFLP." *Archives of Virology* 150 (2): 389 – 397. https://doi.org/10.1007/s00705-004-0399-3.

Peremyslov, V. V., Y.-W. Pan, and V. V. Dolja. 2004. "Movement Protein of a Closterovirus Is a Type III Integral Transmembrane Protein Localized to the Endoplasmic Reticulum." *Journal of Virology* 78 (7): 3704–9. https://doi.org/10.1128/jvi.78.7.3704-3709.2004.

Petrzik, Karel, Igor Koloniuk, Jaroslava Přibylová, and Josef Špak. 2016. "Complete Genome Sequence of Currant Latent Virus (Genus Cheravirus, Family Secoviridae)." *Archives of Virology* 161 (2): 491–93.

https://doi.org/10.1007/s00705-015-2679-5.

Piepenburg, Olaf, Colin H. Williams, Derek L. Stemple, and Niall A. Armes. 2006. “DNA Detection Using Recombination Proteins.” *PLoS Biology* 4 (7): 1115–21. https://doi.org/10.1371/journal.pbio.0040204.

Pita, J. S., V. N. Fondong, A. Sangaré, G. W. Otim-Nape, S. Ogwal, and C. M. Fauquet. 2001. “Recombination, Pseudorecombination and Synergism of Geminiviruses Are Determinant Keys to the Epidemic of Severe Cassava Mosaic Disease in Uganda.” *Journal of General Virology* 82 (3): 655–65. https://doi.org/10.1099/0022-1317-82-3-655.

Powell, Michael L, Frank R Bowler, Aurore J Martinez, Catherine J Greenwood, Niall Armes, and Olaf Piepenburg. 2018. “New Fpg Probe Chemistry for Direct Detection of Recombinase Polymerase Amplification on Lateral Flow Strips.” *Analytical Biochemistry* 543 (February): 108–15. https://doi.org/10.1016/j.ab.2017.12.003.

Prochnik, Simon, Pradeep Reddy Marri, Brian Desany, Pablo D. Rabinowicz, Chinnappa Kodira, Mohammed Mohiuddin, Fausto Rodriguez, et al. 2012. “The Cassava Genome: Current Progress, Future Directions.” *Tropical Plant Biology* 5 (1): 88–94. https://doi.org/10.1007/s12042-011-9088-z.

QGIS Development Team. 2021. “QGIS Geographic Information System. Open Source Geospatial Foundation Project.” http://qgis.osgeo.org.

Quer, Josep, Sergi Colomer-Castell, Carolina Campos, Cristina Andrés, Maria Piñana, Maria Francesca Cortese, Alejandra González-Sánchez, et al. 2022. “Next-Generation Sequencing for Confronting Virus Pandemics.” *Viruses* 14 (3). https://doi.org/10.3390/v14030600.

Quyen, Than Linh, Tien Anh Ngo, Dang Duong Bang, Mogens Madsen, and Anders Wolff. 2019. “Classification of Multiple DNA Dyes Based on Inhibition Effects on Real-Time Loop-Mediated Isothermal Amplification (LAMP): Prospect for Point of Care Setting.” *Frontiers in Microbiology* 10 (October): 1–12. https://doi.org/10.3389/fmicb.2019.02234.

Rabbi, Ismail Y., Martha T. Hamblin, P. Lava Kumar, Melaku A. Gedil, Andrew S. Ikpan, Jean Luc Jannink, and Peter A. Kulakow. 2014. “High-Resolution Mapping of Resistance to Cassava Mosaic Geminiviruses in Cassava Using Genotyping-by-Sequencing and Its Implications for Breeding.” *Virus Research* 186: 87–96. https://doi.org/10.1016/j.virusres.2013.12.028.

Rabobank. 1994. “The World Seed Market.”

Reddy, P. Parvatha. 2015a. *Plant Protection in Tropical Root and Tuber Crops*. *Plant Protection in Tropical Root and Tuber Crops*. Springer I. Bangalore, Karnataka, India. https://doi.org/10.1007/978-81-322-2389-4_1.

———. 2015b. *Plant Protection in Tropical Root and Tuber Crops*. *Plant Protection in Tropical Root and Tuber Crops*. New Delhi: Springer India. https://doi.org/10.1007/978-81-322-2389-4.

Regenmortel, M.H.V. Van. 2016. "Classes, Taxa and Categories in Hierarchical Virus Classification: A Review of Current Debates on Definitions and Names of Virus Species." *Bionomina* 10 (1): 1–21. https://doi.org/10.11646/bionomina.10.1.1.

Riley, Melissa, Margaret Williamson, and Otis Maloy. 2002. "Plant Disease Diagnosis." *Plant Health Instructor*. https://doi.org/10.1094/PHI-I-2002-1021-01.

Rojas, J. Alejandro, Timothy D. Miles, Michael D. Coffey, Frank N. Martin, and Martin I. Chilvers. 2017. "Development and Application of QPCR and RPA Genusand Species-Specific Detection of Phytophthora Sojae and P. Sansomeana Root Rot Pathogens of Soybean." *Plant Disease* 101 (7): 1171–81. https://doi.org/10.1094/PDIS-09-16-1225-RE.

Roossinck, Marilyn J. 2015. "Plants, Viruses and the Environment: Ecology and Mutualism." *Virology* 479–480: 271–77. https://doi.org/10.1016/j.virol.2015.03.041.

Roossinck, Marilyn J., Darren P. Martin, and Philippe Roumagnac. 2015. "Plant Virus Metagenomics: Advances in Virus Discovery." *Phytopathology* 105 (6): 716–27. https://doi.org/10.1094/PHYTO-12-14-0356-RVW.

Roossink, M.J. 2014. "Metagenomics of Plant and Fungal Viruses Reveals an Abundance of Persistent Lifestyles." *Frontiers in Microbiology* 5: 767. https://doi.org/10.3732/ajb.1000298.

Roossink, Marilyn J. 2012. "Plant Virus Metagenomics: Biodiversity and Ecology." *Annual Review of Genetics* 46 (1): 359–69. https://doi.org/10.1146/annurev-genet-110711-155600.

Rosas-Cárdenas, Flor de Fátima, Noé Durán-Figueroa, Jean Philippe Vielle-Calzada, Andrés Cruz-Hernández, Nayelli Marsch-Martínez, and Stefan de Folter. 2011. "A Simple and Efficient Method for Isolating Small RNAs from Different Plant Species." *Plant Methods* 7 (1): 1–7. https://doi.org/10.1186/1746-4811-7-4.

Rose, D J W. 1978. "Epidemiology of Maize Streak Disease." *Annual Review of Entomology* 23 (1): 259–82. https://doi.org/10.1146/annurev.en.23.010178.001355.

Ross, H.B. 1975. "The Diffusion of the Manioc Plant from South America to Africa: An Essay in Ethnobotacnical Culture History." Columbia University. https://library.ciat.cgiar.org/cgi-bin/koha/opac-detail.pl?biblionumber=11387&shelfbrowse_itemnumber=13200#.

Rubio, Luis, Hsin Hung Yeh, Tongyan Tian, and Bryce W. Falk. 2000. "A Heterogeneous Population of Defective RNAs Is Associated with Lettuce

Infectious Yellows Virus." *Virology* 271 (1): 205–12. https://doi.org/10.1006/viro.2000.0318.

Rwegasira, G M, and M. E.C. Rey. 2012. "Response of Selected Cassava Varieties to the Incidence and Severity Ofle Cassava Brown Streak Disease in Tanzania," no. May. https://doi.org/10.5539/jas.v4n7p237.

Rwegasira, Gration, and Chrissie Rey. 2012. "Relationship between Symptoms Expression and Virus Detection in Cassava Brown Virus Streak-Infected Plants." *Journal of Agricultural Science* 4 (7): p246. https://doi.org/10.5539/jas.v4n7p246.

Saah, Alfred J., and Donald R. Hoover. 1997. "'Sensitivity' and 'Specificity' Reconsidered: The Meaning of These Terms in Analytical and Diagnostic Settings." *Annals of Internal Medicine* 126 (1): 91–94. https://doi.org/10.7326/0003-4819-126-1-199701010-00026.

Sakamoto, Y, M Ishiguro, and G Kitagawa. 1986. *Akaike Information Criterion Statistics*. Mathematics and Its Applications. D. Reidel Publishing Company. https://books.google.be/books?id=w6wQnQEACAAJ.

Sanfaçon, Hélène. 2013. "Investigating the Role of Viral Integral Membrane Proteins in Promoting the Assembly of Nepovirus and Comovirus Replication Factories." *Frontiers in Plant Science* 3 (JAN): 1–7. https://doi.org/10.3389/fpls.2012.00313.

Sanfaçon, Hélène, Indranil Dasgupta, Marc Fuchs, Alexander V. Karasev, Karel Petrzik, Jeremy R. Thompson, Ioannis Tzanetakis, René van der Vlugt, Thierry Wetzel, and Nobuyuki Yoshikawa. 2020. "Proposed Revision of the Family Secoviridae Taxonomy to Create Three Subgenera, 'Satsumavirus', 'Stramovirus' and 'Cholivirus', in the Genus Sadwavirus." *Archives of Virology* 165 (2): 527–33. https://doi.org/10.1007/s00705-019-04468-7.

Sanfaçon, Hélène, Joan Wellink, Olivier Le Gall, Alexander Karasev, René Van Der Vlugt, and Thierry Wetzel. 2009. "Secoviridae: A Proposed Family of Plant Viruses within the Order Picornavirales That Combines the Families Sequiviridae and Comoviridae, the Unassigned Genera Cheravirus and Sadwavirus, and the Proposed Genus Torradovirus." *Archives of Virology* 154 (5): 899–907. https://doi.org/10.1007/s00705-009-0367-z.

Sanger, F, G M Air, B G Barrell, N L Brown, A R Coulson, J C Fiddes, C A Hutchison, P M Slocombe, and M Smith. 1977. "Nucleotide Sequence of Bacteriophage ΦX174 DNA." *Nature* 265 (5596): 687–95. https://doi.org/10.1038/265687a0.

Saraswathy, Nachimuthu, and Ponnusamy Ramalingam. 2011. "Genome Mapping." *Concepts and Techniques in Genomics and Proteomics*, 77–93. https://doi.org/10.1533/9781908818058.77.

Saunders, Keith, Ian D. Bedford, Tetsukazu Yahara, and John Stanley. 2003. "The

Earliest Recorded Plant Virus Disease." *Nature* 422 (6934): 831. https://doi.org/10.1038/422831a.

Saunders, Keith, Nazeera Salim, Vasant R Mali, Varagur G Malathi, Rob Briddon, Peter G Markham, and John Stanley. 2002. "Characterisation of Sri Lankan Cassava Mosaic Virus and Indian Cassava Mosaic Virus: Evidence for Acquisition of a DNA B Component by a Monopartite Begomovirus." *Virology* 293 (1): 63–74. https://doi.org/10.1006/viro.2001.1251.

Sayers, Eric W, Evan E Bolton, J Rodney Brister, Kathi Canese, Jessica Chan, Donald C Comeau, Ryan Connor, et al. 2022. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 50 (D1): D20–26. https://doi.org/10.1093/nar/gkab1112.

Scussel, Sarah, Thierry Candresse, Armelle Marais, Sohini Claverie, Murielle Hoareau, Hamza A Azali, Eric Verdin, et al. 2019. "High-Throughput Sequencing of Complete Genomes of Ipomoviruses Associated with an Epidemic of Cassava Brown Streak Disease in the Comoros Archipelago." *Archives of Virology* 164 (8): 2193–96. https://doi.org/10.1007/s00705-019-04228-7.

Seguin, Jonathan, Rajendran Rajeswaran, Nachelli Malpica-López, Robert R. Martin, Kristin Kasschau, Valerian V. Dolja, Patricia Otten, Laurent Farinelli, and Mikhail M. Pooggin. 2014. "De Novo Reconstruction of Consensus Master Genomes of Plant RNA and DNA Viruses from SiRNAs." *PLoS ONE* 9 (2). https://doi.org/10.1371/journal.pone.0088513.

Seif, A. A. 1982. "Effect of Cassava Mosaic Virus on Yield of Cassava." *Plant Disease* 66 (1): 661. https://doi.org/10.1094/pd-66-661.

SENASEM. 2019. "Catalogue National Variétal Des Cultures Vivrières: Répertoire Des Variétés Homologuées Des Plantes à Racines, Tubercules et Bananier." Kinshasa/R.D. Congo. https://minagri.gouv.cd/wp-content/uploads/2020/11/Catalogue_National_Cultures_Racine_Tubercule_Bananier_RDC_2019.pdf.

Sether, Diane M., Michael J. Melzer, Wayne B. Borth, and John S. Hu. 2009. "Genome Organization and Phylogenetic Relationship of Pineapple Mealybug Wilt Associated Virus-3 with Family Closteroviridae Members." *Virus Genes* 38 (3): 414–20. https://doi.org/10.1007/s11262-009-0334-5.

Shankar, K, L Chavan, S Shinde, and B. Patil. 2010. "An Improved DNA Extraction Protocol from Four in Vitro Banana Cultivars." *Asian Journal of Biotechnology* 3 (1): 84–90. https://doi.org/10.3923/ajbkr.2011.84.90.

Shaw, M. W., and M. Pautasso. 2014. "Networks and Plant Disease Management: Concepts and Applications." *Annual Review of Phytopathology* 52 (June): 477–93. https://doi.org/10.1146/annurev-phyto-102313-050229.

Sheat, Samar, Bettina Fuerholzner, Beate Stein, and Stephan Winter. 2019. “Resistance against Cassava Brown Streak Viruses from Africa in Cassava Germplasm from South America.” *Frontiers in Plant Science* 10 (May). https://doi.org/10.3389/fpls.2019.00567.

Sheat, Samar, and Stephan Winter. 2022. “HAM1 in Cassava Brown Streak Viruses Determines Symptom Expression, Pathogenicity and Resistance Responses of Cassava.” In *International Advances in Plant Virology*, edited by Association of Applied Biologists, 37–38. Ljubljana, Slovenia.

———. 2023. “Developing Broad-Spectrum Resistance in Cassava against Viruses Causing the Cassava Mosaic and the Cassava Brown Streak Diseases.” *Frontiers in Plant Science* 14 (January): 1–10. https://doi.org/10.3389/fpls.2023.1042701.

Sheat, Samar, Xiaofei Zhang, and Stephan Winter. 2022. “High-Throughput Virus Screening in Crosses of South American and African Cassava Germplasm Reveals Broad-Spectrum Resistance against Viruses Causing Cassava Brown Streak Disease and Cassava Mosaic Virus Disease.” *Agronomy* 12 (5): 13. https://doi.org/10.3390/agronomy12051055.

Shirima, R.R., Everlyne Nafula Wosula, Abdou Azali Hamza, Nobataine Ali Mohammed, Hadji Mouigni, Salima Nouhou, Naima Mmadi Mchinda, et al. 2022. “Epidemiological Analysis of Cassava Mosaic and Brown Streak Diseases, and Bemisia Tabaci in the Comoros Islands.” *Viruses* 14 (10): 1–18. https://doi.org/10.3390/v14102165.

Shirima, Rudolph R., James P. Legg, Daniel G. Maeda, Silver Tumwegamire, Geoffrey Mkamilo, Kiddo Mtunda, Heneriko Kulembeka, et al. 2020. “Genotype by Environment Cultivar Evaluation for Cassava Brown Streak Disease Resistance in Tanzania.” *Virus Research* 286 (January). https://doi.org/10.1016/j.virusres.2020.198017.

Shirima, Rudolph R, Daniel G Maeda, Edward Kanju, Gloria Ceasar, Flora I Tibazarwa, and James P Legg. 2017. “Absolute Quantification of Cassava Brown Streak Virus MRNA by Real-Time QPCR.” *Journal of Virological Methods* 245: 5–13. https://doi.org/10.1016/j.jviromet.2017.03.003.

Simmonds, Peter, Mike J. Adams, Mária Benk, Mya Breitbart, J. Rodney Brister, Eric B. Carstens, Andrew J. Davison, et al. 2017. “Consensus Statement: Virus Taxonomy in the Age of Metagenomics.” *Nature Reviews Microbiology* 15 (3): 161–68. https://doi.org/10.1038/nrmicro.2016.177.

Southern, E.M. 1975. “Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis.” *Journal of Molecular Biology* 98 (3): 503–17. https://doi.org/10.1016/S0022-2836(75)80083-0.

Spadiut, Oliver, Simona Capone, Florian Krainer, Anton Glieder, and Christoph Herwig. 2014. “Microbials for the Production of Monoclonal Antibodies and

Antibody Fragments." *Trends in Biotechnology* 32 (1): 54–60. https://doi.org/10.1016/j.tibtech.2013.10.002.

Sperling, Louise, and David H. Cooper. 2003. "Understanding Seed Systems and Strengthening Seed Security: A Background Paper." Workshop on Effective and Sustainable Seed Relief Activities, 26-28 May 2003. Rome.

Stace-Smith, R. 1966. "Purification and Properties of Tomato Ringspot Virus and an RNA-Deficient Component." *Virology* 29 (2): 240–47. https://doi.org/https://doi.org/10.1016/0042-6822(66)90030-4.

Stansly, Philip A., Steven E. Naranjo, Judith K. Brown, A. Rami Horowitz, James P. Legg, Jane E. Polston, Dan Gerling, and Moshe Lapidot. 2010. "Bemisia: Bionomics and Management of a Global Pest." *Bemisia: Bionomics and Management of a Global Pest*, no. June: 1–540. https://doi.org/10.1007/978-90-481-2460-2.

Storey, H H. 1935. "Report of the Plant Pathologist." *Report of the Plant Pathologist.*

———. 1936. "Virus Diseases of East African Plants. VI-A Progress Report on Studies of the Disease of Cassava." *East African Agricultural Journal* 2: 34–39.

———. 1938. "Virus Diseases of East African Plants." *The East African Agricultural Journal* 3 (6): 446–49. https://doi.org/10.1080/03670074.1938.11663806.

Storey, H H, and R F W Nichols. 1938. "Studies of the Mosaic Disease of Cassava." *Annals of Applied Biology* 25 (4): 790–806. https://doi.org/10.1111/j.1744-7348.1938.tb02354.x.

Takami, N. 1901. "On Dwarf Disease of Rice Plant and 'Tsumaguro-Yokabai.'" *J. Jpn. Agric. Soc* 241: 22–30.

TANAGRA. 2009. "Understanding the 'Test Value' Criterion. Tutorial." http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/en_Tanagra_Comprendre_La_Valeur_Test.pdf.

Tangpukdee, Noppadon, Chatnapa Duangdee, Polrat Wilairatana, and Srivicha Krudsood. 2009. "Malaria Diagnosis: A Brief Review." *Korean Journal of Parasitology* 47 (2): 93–102. https://doi.org/10.3347/kjp.2009.47.2.93.

Tanner, Nathan A., Yinhua Zhang, and Thomas C. Evans. 2012. "Simultaneous Multiple Target Detection in Real-Time Loop-Mediated Isothermal Amplification." *BioTechniques* 53 (2): 81–89. https://doi.org/10.2144/0000113902.

Tatineni, Satyanarayana, and Gary L. Hein. 2023. "Plant Viruses of Agricultural Importance: Current and Future Perspectives of Virus Disease Management Strategies." *Phytopathology®* 113 (2): 117–41. https://doi.org/10.1094/phyto-05-22-0167-rvw.

Teo, Jeanette W.P., Donald Chiang, Roland Jureen, and Raymond T.P. Lin. 2015.

"Clinical Evaluation of a Helicase-Dependant Amplification (HDA)-Based Commercial Assay for the Simultaneous Detection of HSV-1 and HSV-2." *Diagnostic Microbiology and Infectious Disease* 83 (3): 261–62. https://doi.org/10.1016/j.diagmicrobio.2015.07.018.

The UniProt Consortium. 2021. "The Universal Protein Knowledgebase in 2021." *Nucleic Acids Research* 49 (D1): D480–89. https://doi.org/10.1093/nar/gkaa1100.

Thompson, Jeremy R., Indranil Dasgupta, Marc Fuchs, Toru Iwanami, Alexander V. Karasev, Karel Petrzik, Hélène Sanfaçon, et al. 2017. "ICTV Virus Taxonomy Profile: Secoviridae." *Journal of General Virology* 98 (4): 529–31. https://doi.org/10.1099/jgv.0.000779.

Thresh, J. M. 1991. "The Ecology of Tropical Plant Viruses." *Plant Pathology* 40 (3): 324–39. https://doi.org/10.1111/j.1365-3059.1991.tb02386.x.

Thresh, J. M., and R. J. Cooter. 2005. "Strategies for Controlling Cassava Mosaic Virus Disease in Africa." *Plant Pathology* 54 (5): 587–614. https://doi.org/10.1111/j.1365-3059.2005.01282.x.

Thresh, J.M. 1988. "Cocoa Swollen Shoot: An Archetypal Crowd Disease." *Journal of Plant Diseases and Protection* 95 (4): 428–46.

———. 2003. "Control of Plant Virus Diseases Un Sub Saharan Africa: The Possibility and Feasability of an Integrated Approach." *African Crop Science Journal* 11 (3): 199–223.

Thresh, J.M., and M.W. Mbwana. 1998. "Cassava Mosaic and Cassava Brown Streak Virus Diseases in Zanzibar." *Roots* 5: 6–8.

Thresh, J.M., G.W. Otim-Nape, J.P. Legg, and D. Fargette. 1997. "African Cassava Mosaic Virus Disease: The Magnitude of the Problem." *African Journal of Root Tuber Crop* 2: 13–19.

Thresh, J M. 1982. "Cropping Practices and Virus Spread." *Annual Review of Phytopathology* 20 (1): 193–216. https://doi.org/10.1146/annurev.py.20.090182.001205.

Thresh, J M, and D Fargette. 2001. "The Epidemiology of African Plant Viruses: Basic Principles and Concepts." In *Virus and Virus-like Diseases of Major Crops in Developing Countries.*, edited by Gad Loebenstein and George Thottappilly, Kluwer Aca, 61–111. Dordrecht, Netherland: Kluwer Academic Publisher.

Thresh, J M, Denis Fargette, and G W Otim-Nape. 1994. "Effects of African Cassava Mosaic Geminivirus on the Yield of Cassava." *Tropical Science* 34: 26–42. https://www.documentation.ird.fr/hor/fdi:010022175.

Thresh, J M, and G W Otim-Nape. 1994. "Strategies for Controlling African Cassava

Mosaic Geminivirus." In *Advances in Disease Vector Research*, edited by Kerry F Harris, 215–36. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4612-2590-4_8.

Tiendrébéogo, Fidèle, Pierre Lefeuvre, Murielle Hoareau, Mireille A Harimalala, Alexandre De Bruyn, Julie Villemot, Valentin S E Traoré, et al. 2012. "Evolution of African Cassava Mosaic Virus by Recombination between Bipartite and Monopartite Begomoviruses." *Virology Journal* 9 (1): 67. https://doi.org/10.1186/1743-422X-9-67.

Tomlinson, J. a., S. Ostoja-Starzewska, I. P. Adams, D. W. Miano, P. Abidrabo, Z. Kinyua, T. Alicai, et al. 2013. "Loop-Mediated Isothermal Amplification for Rapid Detection of the Causal Agents of Cassava Brown Streak Disease." *Journal of Virological Methods* 191 (2). https://doi.org/10.1016/j.jviromet.2012.07.015.

Tomlinson, J a, M J Dickinson, and N Boonham. 2010. "Rapid Detection of Phytophthora Ramorum and P. Kernoviae by Two-Minute DNA Extraction Followed by Isothermal Amplification and Amplicon Detection by Generic Lateral Flow Device." *Phytopathology* 100 (2): 143–49. https://doi.org/10.1094/PHYTO-100-2-0143.

Tomlinson, Katie R., Andy M. Bailey, Titus Alicai, Sue Seal, and Gary D. Foster. 2018a. "Cassava Brown Streak Disease: Historical Timeline, Current Knowledge and Future Prospects." *Molecular Plant Pathology* 19 (5): 1282–94. https://doi.org/10.1111/mpp.12613.

———. 2018b. "Cassava Brown Streak Disease: Historical Timeline, Current Knowledge and Future Prospects." *Molecular Plant Pathology* 19 (5): 1282–94. https://doi.org/10.1111/mpp.12613.

Tomlinson, Katie R., José Luis Pablo-Rodriguez, Hamidun Bunawan, Sarah Nanyiti, Patrick Green, Josie Miller, Titus Alicai, Susan E. Seal, Andy M. Bailey, and Gary D. Foster. 2019. "Cassava Brown Streak Virus Ham1 Protein Hydrolyses Mutagenic Nucleotides and Is a Necrosis Determinant." *Molecular Plant Pathology* 20 (8): 1080–92. https://doi.org/10.1111/mpp.12813.

Tsongalis, Gregory J., and Lawrence M. Silverman. 2006. "Molecular Diagnostics: A Historical Perspective." *Clinica Chimica Acta* 369 (2): 188–92. https://doi.org/10.1016/j.cca.2006.02.044.

Uddin, Shah Mukim, Abkar Sayad, Jianxiong Chan, Duc Hau Huynh, Efstratios Skafidas, and Patrick Kwan. 2021. "Heater Integrated Lab-on-a-chip Device for Rapid Hla Alleles Amplification towards Prevention of Drug Hypersensitivity." *Sensors* 21 (10): 1–14. https://doi.org/10.3390/s21103413.

Uke, Ayaka, Sophary Khin, Kohei Kobayashi, Takuma Satou, Ok Kyung Kim, Trinh Xuan Hoat, Keiko T. Natsuaki, and Masashi Ugaki. 2022. "Detection of Sri

Lankan Cassava Mosaic Virus by Loop-Mediated Isothermal Amplification Using Dried Reagents.” *Journal of Virological Methods* 299 (January 2020): 114336. https://doi.org/10.1016/j.jviromet.2021.114336.

Valli, Adrian A., Rafael Garcia Lo pez, Maria Ribaya, Francisco Javier Martinez, Diego Garcia Gomez, Beatriz Garcia, Irene Gonzalo, et al. 2022. “Maf/Ham1-like Pyrophosphatases of Noncanonical Nucleotides Are Host-Specific Partners of Viral RNA-Dependent RNA Polymerases.” *PLoS Pathogens* 18 (2): 1–23. https://doi.org/10.1371/journal.ppat.1010332.

Vanderschuren, Hervé, and Chrissie Rey. 2017. “Cassava Mosaic and Brown Streak Diseases: Current Perspectives and Beyond.” *Annual Review of Virology* 4 (1): annurev-virology-101416-041913. https://doi.org/10.1146/annurev-virology-101416-041913.

Varshney, Rajeev K., Jean Marcel Ribaut, Edward S. Buckler, Roberto Tuberosa, J. Antoni Rafalski, and Peter Langridge. 2012. “Can Genomics Boost Productivity of Orphan Crops?” *Nature Biotechnology* 30 (12): 1172–76. https://doi.org/10.1038/nbt.2440.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Springer (. New York, NY, USA.

Verdaguer, Bertrand, Alexandre De Kochko, Roger N Beachy, and Claude Fauquet. 1996. “Isolation and Expression in Transgenic Tobacco and Rice Plants, of the Cassava Vein Mosaic Virus (CVMV) Promoter.” *Plant Molecular Biology* 31 (6): 1129 – 1139. https://doi.org/10.1007/BF00040830.

Verdin, Eric, Catherine Wipf-Scheibel, Patrick Gognalons, François Aller, Mireille Jacquemond, and Mark Tepfer. 2017. “Sequencing Viral SiRNAs to Identify Previously Undescribed Viruses and Viroids in a Panel of Ornamental Plant Samples Structured as a Matrix of Pools.” *Virus Research* 241 (September): 19–28. https://doi.org/10.1016/j.virusres.2017.05.019.

Villamor, D. E.V., T. Ho, M. Al Rwahnih, R. R. Martin, and I. E. Tzanetakis. 2019. “High Throughput Sequencing for Plant Virus Detection and Discovery.” *Phytopathology* 109 (5): 716–725. https://doi.org/10.1094/PHYTO-07-18-0257-RVW.

Vu, Nam Tuan, Juan Manuel Pardo, Elizabeth Alvarez, Ham Huy Le, Kris Wyckhuys, Kim Lien Nguyen, and Dung Tien Le. 2016. “Establishment of a Loop-Mediated Isothermal Amplification (LAMP) Assay for the Detection of Phytoplasma-Associated Cassava Witches’ Broom Disease.” *Applied Biological Chemistry* 59 (2): 151–56. https://doi.org/10.1007/s13765-015-0134-7.

Walker, Peter J., Stuart G. Siddell, Elliot J. Lefkowitz, Arcady R. Mushegian, Evelien M. Adriaenssens, Donald M. Dempsey, Bas E. Dutilh, et al. 2020. “Changes to Virus Taxonomy and the Statutes Ratified by the International Committee on

Taxonomy of Viruses (2020).” *Archives of Virology* 165 (11): 2737–48. https://doi.org/10.1007/s00705-020-04752-x.

Wang, A., and H. Sanfacon. 2000. “Proteolytic Processing at a Novel Cleavage Site in the N-Terminal Region of the Tomato Ringspot Nepovirus RNA-1-Encoded Polyprotein in Vitro.” *Journal of General Virology* 81 (11): 2771–81. https://doi.org/10.1099/0022-1317-81-11-2771.

Wang, Aiming, Sumin Han, and Hélène Sanfaçon. 2004. “Topogenesis in Membranes of the NTB – VPg Protein of Tomato Ringspot Nepovirus : Definition of the C-Terminal Transmembrane Domain.” *Journal of General Virology* 85: 535–45. https://doi.org/10.1099/vir.0.19612-0.

Watson, I D. 2000. “Clinical Diagnosis: Chromatography.” *Encyclopedia of Separation Science*, 2484–90. https://doi.org/10.1016/B0-12-226770-2/01061-9.

WATSON, J D, and F H C CRICK. 1953. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” *Nature* 171 (4356): 737–38. https://doi.org/10.1038/171737a0.

Watson, Mick, and Amanda Warr. 2019. “Errors in Long-Read Assemblies Can Critically Affect Protein Prediction.” *Nature Biotechnology* 37 (2): 124–26. https://doi.org/10.1038/s41587-018-0004-z.

Weemen, B.K. Van, and A.H.W.M. Schuurs. 1971. “Immunoassay Using Antigen-Enzyme Conjugates.” *FEBS Letters* 15 (3): 232–36. https://reader.elsevier.com/reader/sd/pii/0014579371803198?token=0F7ADA583F861A7C24CCB4A839ABEEB3D3D4252B01262377C4B74D3E1B871099F7078DE883EE7EB545F95412D212B57D&originRegion=eu-west-1&originCreation=20220812132638.

Wetzel, Thierry, Joan Chisholm, Alexandra Bassler, and Hélène Sanfaçon. 2008. “Characterization of Proteinase Cleavage Sites in the N-Terminal Region of the RNA1-Encoded Polyprotein from Arabis Mosaic Virus (Subgroup A Nepovirus).” *Virology* 375 (1): 159–69. https://doi.org/10.1016/j.virol.2008.01.046.

White, K.A., and T.J. Morris. 1995. “RNA Determinants of Junction Site Selection in RNA Virus Recombinants and Defective Interfering RNAs.” *Rna* 1 (10): 1029–40.

Wilson, Edward O. 1998. *CONCILIENCE, the Unity of Knowledge*. Edited by A. Knopf. Edward O. Vintage Books? Random house.

Winter, Stephan, Marianne Koerbler, Beate Stein, Agnes Pietruszka, Martina Paape, and Anja Butgereitt. 2010a. “Analysis of Cassava Brown Streak Viruses Reveals the Presence of Distinct Virus Species Causing Cassava Brown Streak Disease in East Africa.” *Journal of General Virology* 91 (5): 1365–72.

https://doi.org/10.1099/vir.0.014688-0.

———. 2010b. “Analysis of Cassava Brown Streak Viruses Reveals the Presence of Distinct Virus Species Causing Cassava Brown Streak Disease in East Africa.” *Journal of General Virology* 91 (5): 1365–72. https://doi.org/10.1099/vir.0.014688-0.

Wylie, Stephen J., Hua Li, Krishnapillai Sivasithamparam, and Michael G.K. Jones. 2014. “Complete Genome Analysis of Three Isolates of Narcissus Late Season Yellows Virus and Two of Narcissus Yellow Stripe Virus: Three Species or One?” *Archives of Virology* 159 (6): 1521–25. https://doi.org/10.1007/s00705-013-1969-z.

Xie, Xiulan, Zhouqing He, Nifan Chen, Zizhong Tang, Qiang Wang, and Yi Cai. 2019. “The Roles of Environmental Factors in Regulation of Oxidative Stress in Plant.” *BioMed Research International* 2019: 21–27. https://doi.org/10.1155/2019/9732325.

Xing, Yanru, John Fredy, Hernandez Nopsa, Kelsey Andersen, and Fenton Douglas Beed. 2020. “Global Cropland Connectivity : A Risk Factor for Invasion and Saturation by Emerging Pathogens and Pests,” no. July. https://doi.org/10.1093/biosci/biaa067.

Yu, Naitong, Zhiwen Luo, Hongyan Fan, Zhili Zhang, Xianghong Li, Jianhua Wang, Zhixin Liu, and Fan He. 2015. “Complete Genomic Sequence of a Pineapple Mealybug Wilt-Associated Virus-1 from Hainan Island, China.” *European Journal of Plant Pathology* 141 (3): 611–15. https://doi.org/10.1007/s10658-014-0545-z.

Zamzami, Mazin A. 2022. “Inosine Triphosphate Pyrophosphatase (ITPase): Functions, Mutations, Polymorphisms and Its Impact on Cancer Therapies.” *Cells* 11 (3). https://doi.org/10.3390/cells11030384.

Zell, R., E. Delwart, A. E. Gorbalenya, T. Hovi, A. M.Q. King, N. J. Knowles, A. M. Lindberg, et al. 2017. “ICTV Virus Taxonomy Profile: Picornaviridae.” *Journal of General Virology* 98 (10): 2421–22. https://doi.org/10.1099/jgv.0.000911.

Zerbini, F. Murilo, Rob W. Briddon, Ali Idris, Darren P. Martin, Enrique Moriones, Jesús Navas-Castillo, Rafael Rivera-Bustamante, Philippe Roumagnac, and Arvind Varsani. 2017. “ICTV Virus Taxonomy Profile: Geminiviridae.” *Journal of General Virology* 98 (2): 131–33. https://doi.org/10.1099/jgv.0.000738.

Zeyimo, Bakelana, Justin Pita, Monde Godefroid, Mahungu Nzola, Lema Munseki, Tshilenge Kanana, and Kalonji Mbuyi. 2020. “Cassava Root Necrosis Disease (CRND): A New Crop Disease Spreading in Western Democratic Republic of Congo and in Some Central African Countries.” *Journal of Agricultural Science* 12 (3): 105. https://doi.org/10.5539/jas.v12n3p105.

Zhang, Guangzhi, and Hélène Sanfaçon. 2006. "Characterization of Membrane Association Domains within the Tomato Ringspot Nepovirus X2 Protein , an Endoplasmic Reticulum-Targeted Polytopic Membrane Protein □ †" 80 (21): 10847–57. https://doi.org/10.1128/JVI.00789-06.

Zhang, J, and J.McD. Stewart. 2016. "Economical and Rapid Method for Extracting Cotton Genomic DNA." *The Journal of Cotton Science* 4 (March): 193–201. https://www.cotton.org/journal/2000-04/3/upload/jcs04-193.pdf.

Zhang, Liding, Xuewei Du, Congjie Chen, Zhixin Chen, Li Zhang, Qinqin Han, Xueshan Xia, Yuzhu Song, and Jinyang Zhang. 2018. "Development and Characterization of Double-Antibody Sandwich ELISA for Detection of Zika Virus Infection." *Viruses* 10 (11): 634. https://doi.org/10.3390/v10110634.

Zhang, Xuzhi, Stuart B. Lowe, and John Justin Gooding. 2014. "Brief Review of Monitoring Methods for Loop-Mediated Isothermal Amplification (LAMP)." *Biosensors and Bioelectronics* 61: 491–99. https://doi.org/10.1016/j.bios.2014.05.039.

Zhou, X., Y. Liu, L. Calvert, C. Munoz, G. W. Otim-Nape, D. J. Robinson, and B. D. Harrison. 1997. "Evidence That DNA-A of a Geminivirus Associated with Severe Cassava Mosaic Disease in Uganda Has Arisen by Interspecific Recombination." *Journal of General Virology* 78 (8): 2101–11. https://doi.org/10.1099/0022-1317-78-8-2101.