LIÈGE université
Gembloux
Agro-Bio Tech

Analysis of Cassava Brown Streak Disease in Rwanda: Incidence, Dissemination, Genetic Diversity, and Innovative Mitigation Strategies.

Chantal NYIRAKANANI

FRENCH COMMUNITY OF BELGIUM
UNIVERSITY OF LIEGE _ GEMBLOUX AGRO-BIO TECH

Analysis of Cassava Brown Streak Disease in Rwanda: Incidence, Dissemination, Genetic Diversity, and Innovative Mitigation Strategies.

Chantal NYIRAKANANI

Original dissertation presented in the fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Agronomical Sciences and Bioengineering

Promoter: Professor Hervé VANDERSCHUREN
Co-promoter: Dr. Vincent de Paul BIGIRIMANA

Civil Year: 2023

© Chantal NYIRAKANANI, 2023

Chantal NYIRAKANANI (2023). Analysis of Cassava Brown Streak Disease in Rwanda: Incidence, Dissemination, Genetic Diversity, and Innovative Mitigation Strategies. (Ph.D. thesis). Gembloux Agro-Bio Tech, Liege University, Belgium. 27 Figure, 8 Tables.

Abstract

Cassava brown streak disease (CBSD), caused by Cassava brown streak ipomoviruses (CBSIs), namely *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), poses a significant threat to global food security. It particularly jeopardizes the food security of tropical Africa, where approximately 450 million people rely heavily on cassava as a staple food and vital income source. In Rwanda, CBSD has rapidly spread since its first report in 2009, with the incidence rising from 18.5% in 2012 to 69% in 2014. This widespread outbreak has resulted in severe consequences, including a shortage of planting materials and a 73% decline in cassava yields. CBSD is primarily transmitted to a longer distance through infected cuttings and, to a shorter distance, by the white fly vector. Infected plants exhibit symptoms on the leaves, stem, and, very importantly, on the storage roots, rendering them unsuitable for consumption.

In response to the CBSD outbreak, the government of Rwanda and researchers have joined forces to combat the burden of CBSD. Notably, they started to import and distribute tolerant cassava planting materials to farmers in 2015. Given that infected cuttings are a significant transmission mode for CBSD, it is imperative to break the cycle of disease transmission and minimize the risk of CBSD spread and its consequences by ensuring the availability of healthy planting materials. In this context, *in vitro* virus cleaning approaches were applied to combat the virus build-up effect over multiple cycles due to cassava's vegetative nature. However, these approaches take time and are often too costly for subsistence crops. Furthermore, studies have been conducted to assess the prevalence and diversity of CBSD causative agents in Rwanda, albeit with a focus limited to partial coat protein, and different breeding research projects have been initiated. These initiatives were reflecting important investments to mitigate the impact of CBSD while acknowledging the need for further extensive research to tackle the disease comprehensively.

In this regard, the first goal of this thesis was to conduct a countrywide cassava seed system survey to determine CBSD status following interventions and the risk factors that may contribute to its continued spread in Rwanda. To achieve this, 130 cassava farmers were interviewed across 13 major cassava-growing districts and their fields were visited to evaluate disease incidence. Leaf samples were collected and analyzed using RT-PCR (reverse transcription polymerase chain reaction) to confirm CBSIs infection. The findings revealed that CBSD has spread in all surveyed districts, and the overall incidence was 35.3%, with UCBSV being the most common, accounting for 61% of the infections. Several key risk factors that could contribute to the spread of the disease in Rwanda were also identified, including the source of planting materials, geographical location, knowledge of disease transmission, and disease management practices. These findings highlight the need to develop a robust seed

system and train farmers to increase awareness and skills to mitigate the spread and impact of CBSD in cassava farming communities.

Recognizing the pivotal role of robust diagnostic tools in fortifying the seed system, the second goal of this thesis was to investigate the genetic diversity of CBSD-causing agents in Rwanda by analyzing whole genomes with innovative methods to provide valuable insights into the evolutionary patterns of CBSIs in Rwanda. High-throughput sequencing (HTS) technologies were applied on 13 pooled samples (corresponding to 13 surveyed districts), enabling us to obtain comprehensive genomic data. Through HTS data analysis, 12 nearly complete consensus genomes of UCBSV were successfully reconstructed. Phylogenetic analysis of these genomes revealed a remarkable reduction in genetic diversity, with a maximum of 0.8% nucleotide divergence between the genomes. Further investigation beyond the consensus sequences utilizing the combination of fixation index (F_{ST}) calculation and Principal Component Analysis (PCA) based on SNPs patterns unveiled three distinct UCBSV haplotypes exhibiting geographic clustering. Interestingly, the distribution of haplotype two (H_2) was found to be associated with one of the CBSD tolerant cultivars widely distributed to farmers, "NAROCAS1". In addition, HTS allowed the assembly of the partial genome of *Manihot esculenta-associated virus 1*(MEaV-1) for the first time in Rwanda. Identifying distinct UCBSV haplotypes and their geographic distribution represents the first study in Rwanda, marking a significant advancement into the local patterns of UCBSV evolution, facilitating a better understanding of the disease's spread, and developing targeted control strategies.

Considering that the current main CBSD management relies on the distribution of tolerant cultivars susceptible to viral buildup effect, the third objective of this thesis was to transform existing *in vitro* virus cleaning methods into practical farmer-friendly approaches at the greenhouse and field levels toward CBSD mitigation. The present study assessed the effectiveness of combining greenhouse thermotherapy with chemotherapy and field chemotherapy, employing salicylic acid (SA) and Benzothiadiazole (BTH) on CBSIs-infected cuttings. The results revealed a remarkable reduction in viral loads, especially when combining thermotherapy with SA at 50 mg/L and thermotherapy with BTH at 50 mg/L, which exhibited the most substantial reduction compared to other treatments. Additionally, a significant decrease in the severity of CBSD root symptoms through field chemotherapy was observed among treated plants. These findings highlight the potential effectiveness of these combined approaches in mitigating the impact of CBSD and offer promising avenues for disease management in cassava. Furthermore, RNA sequencing on uninfected cassava plants exogenously treated with SA and BTH was conducted to investigate their impact on the cassava transcriptome. It was revealed that SA and BTH deregulate numerous cassava genes, including genes with potential involvement in plant defense, such as transcription factors (e.g., WRKY), Leucine Rich Repeat (LRR) Protein, Heat shock Protein (HSP), Mitogen-Activated Protein Kinase (MAPK), Cytochrome P450, and ethylene-responsive genes. The gene ontology (GO) enrichment analysis revealed that hormone signaling, defense response, response to stress, and regulation of transcription were among the enriched GO, suggesting their potential role in viral host response.

Overall, this thesis contributed significantly to understanding and managing CBSD, providing valuable knowledge for sustainable cassava farming in Rwanda. The countrywide farmers and cassava fields survey provided crucial findings on CBSD status and risk factors, emphasizing the urgency of a robust seed system and farmers' training. The high-resolution investigation into UCBSV genetic diversity through an innovative approach shed light on its evolutionary patterns and geographic distribution, offering insights for targeted control measures. Finally, transforming *in vitro* virus cleaning methods into greenhouse and field approaches showcased promising results in reducing CBSD severity and viral loads, supported by identifying potential defense-related cassava genes.

The present thesis can also serve as the basis for future research. Notably, the innovative approach to characterizing genetic diversity could be applied to study other crucial plant viruses. It is also worth investigating CBSV diversity and the factors driving the evolution of CBSIs in Rwanda. Furthermore, future research is needed to optimize the efficacy of greenhouse and field CBSD mitigation approaches and delve into the specific functions of genes regulated by SA and BTH. Moreover, assessing the impact of chemotherapy on the environment and microbiome and analyzing large-scale cost-benefit viability would provide invaluable insights. These avenues of exploration will undoubtedly contribute to a deeper understanding of virus management strategies and bolster efforts to safeguard plant health and agricultural productivity.

Keywords: Cassava, CBSD, UCBSV, CBSV, Ampelovirus, genetic diversity, SNP, thermotherapy, Chemotherapy, Rwanda

Chantal NYIRAKANANI (2023). Analysis of Cassava Brown Streak Disease in Rwanda: Incidence, Dissemination, Genetic Diversity, and Innovative Mitigation Strategies. (Ph.D. thesis). Gembloux Agro-Bio Tech, Liege University, Belgium. 27 Figure, 8 Tables.

Résumé

La maladie de la striure brune du manioc (CBSD) est causée par les ipomovirus de la striure brune du manioc (CBSIs), à savoir : le virus de la striure brune du manioc (CBSV) et le virus de la striure brune du manioc ougandais (UCBSV). Elle constitue une menace significative pour la sécurité alimentaire mondiale. Elle compromet particulièrement la sécurité alimentaire de l'Afrique tropicale, où environ 450 millions de personnes dépendent largement du manioc non seulement comme aliment de base mais aussi comme source de revenus vitale. Au Rwanda, la CBSD s'est rapidement propagée depuis son premier signalement en 2009, avec une incidence passant de 18,5% en 2012 à 69% en 2014. Cette épidémie généralisée a entraîné des conséquences graves, notamment une pénurie de matériels de plantation et une baisse considérable de 73% des rendements du manioc. La CBSD est principalement transmise sur de plus longues distances par des boutures infectées et sur de plus courtes distances par le vecteur appelé mouche blanche. Les plantes infectées présentent des symptômes sur les feuilles, les tiges et surtout sur les racines de stockage, ce qui les rend impropres à la consommation.

En réponse à l'épidémie de CBSD, le gouvernement Rwandais et les chercheurs ont décidé de travailler en synergie pour lutter contre le fardeau de la CBSD. De ce fait, en 2015, ils ont commencé à importer et à distribuer les matériels de plantation de manioc tolérants à la CBSD aux agriculteurs. Étant donné que les boutures infectées sont un mode de transmission significatif de la CBSD, il est impératif de rompre le cycle de transmission de la maladie et de minimiser le risque de propagation de la CBSD de même que ses conséquences en assurant la disponibilité du matériel de plantation sain. Dans ce contexte, des approches de nettoyage du virus *in vitro* ont été appliquées pour lutter contre l'effet d'accumulation du virus qui se produit sur plusieurs cycles en raison de la nature végétative du manioc. Cependant, ces différentes approches nécessitent du temps et des efforts. Par ailleurs, des études ont été menées pour évaluer la prévalence et la diversité des agents causatifs de la CBSD au Rwanda, bien que l'accent ait été mis sur la protéine partielle de la capside. De plus, différentes recherches en matière de sélection ont été lancées. Ces initiatives reflètent des investissements importants pour atténuer l'impact de la CBSD tout en reconnaissant la nécessité de mener des recherches supplémentaires plus approfondies afin d'aborder la maladie de manière globale.

À cet égard, le premier objectif de cette thèse était de réaliser une enquête nationale sur le système de semences de manioc pour déterminer l'état de la CBSD et les facteurs de risque qui peuvent contribuer à sa propagation continue au Rwanda. Pour y parvenir, 130 producteurs de manioc ont été interrogés dans 13 grands districts producteurs de manioc. De même, leurs champs ont été visités à l'effet d'évaluer l'incidence de la maladie. Enfin, des échantillons de feuilles ont été prélevés et analysés à l'aide de la RT-PCR pour confirmer l'infection par les CBSIs.

Les résultats ont révélé que la CBSD s'est répandue dans tous les districts enquêtés, avec une incidence globale de 35,3 %, l'UCBSV étant le plus commun, représentant 61 % des infections. Plusieurs facteurs de risque clés pouvant contribuer à la propagation de la maladie au Rwanda ont également été identifiés, en l'occurrence la source du matériel de plantation, l'emplacement géographique, la connaissance de la transmission de la maladie et les pratiques de gestion de la maladie. Ces résultats soulignent la nécessité de développer un solide système de semences et de former les agriculteurs pour accroître leur sensibilisation et leur compétence afin de lutter contre la propagation et l'impact de la CBSD dans les communautés de producteurs de manioc.

Reconnaissant le rôle essentiel des outils de diagnostic robustes dans le renforcement du système de semences, le deuxième objectif de cette thèse était d'étudier la diversité génétique des agents responsables de la CBSD au Rwanda en analysant l'ensemble des génomes grâce à des méthodes innovantes pour fournir des informations précieuses sur les modèles évolutifs des CBSI au Rwanda. Le séquençage à haut débit (HTS) a été utilisé sur 13 échantillons regroupés (correspondant aux 13 grands districts producteurs de manioc enquêtés), nous permettant d'obtenir des données génomiques complètes. Grâce à l'analyse des données HTS, 12 génomes presque complets de l'UCBSV ont été reconstruits avec succès. L'analyse phylogénétique de ces génomes a révélé une réduction remarquable de la diversité génétique, avec un maximum de 0,8 % de divergence nucléotidique entre les génomes de l'UCBSV. Des analyses supplémentaires au-delà des séquences de consensus, en utilisant la combinaison du calcul de l'indice de fixation (F_{ST}) et de l'analyse en composantes principales (ACP) basées sur les motifs de SNP ont révélé trois haplotypes distincts d'UCBSV présentant un regroupement géographique. Fait intéressant, la distribution de l'haplotype deux (H_2) a été associée à l'une des variétés de manioc tolérantes à la CBSD largement distribuées aux agriculteurs, "NAROCAS1". De plus, le HTS a permis pour la première fois l'assemblage du génome partiel d'un ampélovirus au Rwanda. L'identification des haplotypes distincts d'UCBSV et leur répartition géographique constituent la première étude au Rwanda marquant une avancée significative dans le modèle local d'évolution d'UCBSV, ce qui facilite une meilleure compréhension de la propagation de la maladie ainsi que l'élaboration de stratégies de contrôle ciblées.

Étant donné que la gestion principale actuelle de la CBSD repose sur la distribution de cultivars tolérants à la CBSD et susceptibles à l'effet d'accumulation virale, le troisième objectif de cette thèse était de transformer les méthodes existantes de nettoyage du virus *in vitro* en approches pratiques adaptées aux agriculteurs au niveau de la serre et du champ pour atténuer le CBSD. La présente étude a permis d'évaluer l'efficacité de la combinaison de la thermothérapie en serre avec la chimiothérapie et la chimiothérapie sur le terrain, en utilisant l'acide salicylique et le benzothiadiazole sur les boutures infectées par les CBSIs. Les résultats auxquels nous sommes parvenus, ont révélé une réduction remarquable de la charge virale, en particulier lors de la combinaison de la thermothérapie avec l'acide salicylique à 50 mg/L et le benzothiadiazole à 50 mg/L, qui a montré la réduction la plus substantielle par rapport aux autres traitements. En outre, nous avons observé une diminution significative de la gravité des symptômes racinaires du CBSD grâce à la chimiothérapie sur le terrain.

Ces résultats mettent en évidence l'efficacité potentielle de ces approches combinées pour atténuer l'impact du CBSD et offrent des perspectives prometteuses pour la gestion de la maladie du manioc. De même, nous avons effectué un séquençage d'ARN sur des plantes de manioc non infectés traités exogènement avec du SA et du BTH pour étudier leur impact sur le transcriptome du manioc. Il a été révélé que le SA et le BTH dérèglent de nombreux gènes du manioc, y compris des gènes potentiellement impliqués dans la défense des plantes, tels que des facteurs de transcription (par exemple, WRKY), des protéines riches en leucine répétées (LRR), des protéines de choc thermique (HSP), des protéines kinases activées par les mitogènes (MAPK), des cytochromes P450 et des gènes réactifs à l'éthylène. Bien plus, l'analyse de l'enrichissement de l'ontologie des gènes (GO) a révélé que la signalisation hormonale, la réponse immunitaire, la réponse au stress et la régulation de la transcription figuraient parmi les GO enrichis, suggérant leur rôle potentiel dans la réponse de l'hôte au virus.

Dans l'ensemble, cette thèse a apporté une contribution significative à la compréhension et à la gestion de la CBSD, fournissant ainsi des connaissances précieuses pour une agriculture durable du manioc au Rwanda. L'enquête nationale menée auprès des agriculteurs et dans les champs de manioc a fourni des résultats cruciaux sur l'état de la CBSD et les facteurs de risque, soulignant l'urgence d'un solide système de semences et de la formation des agriculteurs. Des analyses plus approfondies au-delà du consensus de la diversité génétique d'UCBSV grâce à une approche novatrice a également apporté des éclaircissements sur les modèles évolutifs et la répartition géographique, offrant des informations pour des mesures de contrôle ciblées. Enfin, la transformation des méthodes de nettoyage du virus *in vitro* en approches en serre sur le terrain a dévoilé des résultats prometteurs en ceci qu'elle contribue à la réduction des charges virales de la CBSD, étayée par l'identification de gènes potentiels du manioc liés à la défense.

La présente thèse pourra également servir de base pour des recherches futures. Ainsi, l'approche novatrice visant à caractériser la diversité génétique pourra être appliquée pour étudier d'autres virus végétaux cruciaux. Subséquemment, l'investigation des facteurs qui influencent l'évolution et la diversité des CBSIs au Rwanda sera significative. Bien plus, des recherches supplémentaires seront nécessaires pour optimiser l'efficacité des approches de lutte contre le CBSD en serre sur le terrain et approfondir les fonctions spécifiques des gènes régulés par l'acide salicylique et le benzothiadiazole. Enfin, réaliser une évaluation de l'impact de la chimiothérapie sur l'environnement et le microbiome, tout en analysant sa faisabilité en termes de coûts et d'avantages à grande échelle, fourniront des informations inestimables. Ces voies d'exploration contribueront sans aucun doute à une meilleure compréhension des stratégies de gestion des virus et renforceront les efforts visant à protéger la santé des plantes et la productivité agricole.

Mots-clés : Manioc, CBSD, UCBSV, CBSV, Ampelovirus, diversité génétique, SNP, thermothérapie, chimiothérapie, Rwanda.

Acknowledgments

First and foremost, I glorify my Lord, for He walked with me on this Ph.D. journey. Thank you, God, for this blessing you have bestowed on my life!

My sincere gratitude to the ARES (Académie de Recherche et d'Enseignement Supérieur [Academy of Research and Higher Education]) as well as to the University of Liege for funding my research work through the iCARE project. I also appreciate the University of Rwanda for recommending me to undertake this scholarship.

I am indebted to my promoter **Prof. Hervé VANDERSCHUREN,** and **Prof. Sébastien MASSART**; ***this journey could have been in vain without you***. Words cannot express the depth of my gratitude for guiding, training, and transforming me. ***Your tireless academic nurturing has made a profound impact on me***.

Prof. Vanderschuren, you were the backbone of my work and your guidance has been a driving force behind my growth and achievements. As the project coordinator, I'm grateful for your responsiveness to my multiple requests which resulted in the successful culmination of my research work. I also appreciate your efforts in organizing refreshing events that brought our team together. I can't forget the best moments we had while sharing Xmas joy as the Plant Genetic Team.

Prof. Massart, I'm forever grateful for opening my eyes to the new world of bioinformatics; thank you for believing in me even when I was doubting myself; you taught me perseverance, and for sure I deleted giving up in my dictionary. I cannot forget to appreciate you for the opportunity to participate in the conference at Aussois, France; in addition to building a scientific career, I also enjoyed doing my first skiing in the snowy Alps!

To all members of my thesis committee and Jury: **Prof. Patrick du Jardin, Prof. Claude Bragard, Prof. Ludivine Lassois, Prof. Gengler Nicolas, Dr. Athanase Nduwumuremyi,** and **Dr. Vincent de Paul Bigirimana,** thank you for your collaboration and acceptance to examine this thesis.

I sincerely thank **Prof. Charles BUCAGU**, who informed me about this Ph.D. opportunity when we met at the SASA-UR conference in 2017. I remember that I was about to give up before even starting, but your advice pushed me to embark on this adventure, and here I am, finished with pride. Thank you so much.

I wholeheartedly express my deepest gratitude to **Prof. Jean Claude TWIZERE** for his unwavering kindness and support throughout my Ph.D. journey and career development. Thank you immensely for everything.

A special thanks go to the Team of Plant Genetics, particularly **Sara, Ariadna, Shan, Leonard, Emanoella, Yvan,** and **Kuma**r. Besides being a united professional team, you also played the role of mentoring. Thank you so much for your patience

and everyday guidance. Your contribution to my career development will be forever remembered. I can't forget you **Soklay**; you are gifted; you play a mother role in our Plant Genetic lab, ensuring we had everything we needed; thank you so much for all the assistance. Many thanks also to the team phytopathology lab, especially **Lucie Tamiser, Johan,** and **Yves**; your collaboration was invaluable.

My gratitude also goes to my classmate: **Jean Pierre BIZIMANA**, your assistance in everything motivated me to keep going. Without you, I doubt I could have made it. I would also like to express my gratitude to the staff of RAB, particularly **Severin,** and **Ghislain**, for their cooperation during my journey.

This acknowledgment would be incomplete without appreciating the Catholic University of Rwanda. To my BLS Team particularly **Dr. Protais Muhayimana, Godefroid,** and **Japhet**; please accept my appreciation for your support was priceless.

Monica UWONKUNDA, I am deeply grateful for your presence and support as my mother in Belgium. **Claudine Namahoro**, please accept my heartfelt gratitude for being my caring elder sister. And to you, **Carine Ouedraogo**, a special appreciation for filling the role of my little sister. Your love and affection are engraved in my heart.

Finally, and more importantly, I wish to express ***my deepest gratitude to my family***, led by our father, **Protais HABAKURAMA**, and our mother, **Floride MUKANDABUKA**, who initiated us into the academic world. **My sisters**: *Claudine, Yvonne, Liliane & Brigitte,* and **brothers**: *Méthode, Procar & Protogene,* with this Ph.D. thesis, ***I honor your immeasurable sacrifices***. I take immense pride not only in accomplishing this splendid feat but also in your unwavering trust in me. To **my nephews**: *Melvin, Hervé, Shema, Tony, Kelly, Liam, Rwema & Logan*, and **my nieces**: *Gabriella, Briella, Brinna & Sheja*, thanks for spicing my Ph.D. journey in your unique ways, and as you grow up keep in mind that "***the only dreams impossible to reach are the ones you never pursue*** ".

Professor Morris Komakech, please find in this sentence my earnest gratitude for standing tall as my rock and a constant source of motivation.

This page isn't enough to list all of you who supported me during this journey, but in a few words full of gratitude, "**Thank You**".

Chantal NYIRAKANANI

Gembloux, August 1st 2023

Table of contents

Abstract i
Résumé iv
Acknowledgments vii
Table of contents ix
List of figures xiii
List of tables xv
List of abbeviations xvii
Chapter 1. General Introduction 3
General Introduction to Chapter 1 3
1.1. Background 4
1.1. Cassava Production and its Constraints 5
1.2. Epidemiology, Diversity, and Burden of major cassava viral diseases in sub-Saharan Africa 7
1.3.1. Cassava Mosaic Disease (CMD) 7
1.3.2. Cassava Brown Streak Disease (CBSD) 9
1.4. Status of CBSD in Rwanda 11
1.5. Strategies for cassava viral disease management 13
1.5.1. Knowledge of the pathogen 14
1.5.2. Diagnosis 14
1.5.3. Use of virus or whitefly-resistant cassava 16
1.5.4. Phytosanitary measures 17
1.5.5. Cassava virus cleaning 18
1.5.6. Biotechnology approach for managing CBSD 21
Chapter 2: Problem Statement, Objectives, and Thesis Structure 25
General Introduction to Chapter 2 25
2.1. Problem Statement and Justification 26
2.2. Research Questions and Thesis Objectives 27
2.2.1. Research Questions 27
2.2.2. Thesis Objectives 27
2.3. Thesis Structure 28
Chapter 3. Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated with Cassava Brown Streak Disease Incidence. ...31
General Introduction to Chapter 3 31
3.1. Abstract 33
3.2. Introduction 34
3.3. Materials and Methods 35

3.3.1. Study area ... 35
3.3.2. Farmers and fields selection ... 36
3.3.3. Farmers' interview ... 37
3.3.4. Disease severity assessment ... 37
3.3.5. Sampling test materials for RT-PCR ... 37
3.3.6. Molecular analysis ... 37
3.3.7. Data analysis ... 38
3.4. Results ... 38
3.4.1. Source of cassava cuttings, trust and disease management ... 38
3.4.2. CBSD incidence and severity score of observed symptoms. ... 39
3.4.3. Bivariate analysis of factors associated with CBSD incidence ... 41
3.4.4. Multivariate analysis of risk factors associated with CBSD ... 46
3.5. Discussion ... 48
3.6. Supplementary materials. ... 49
Chapter 4. Going beyond consensus genome sequences: an innovative SNP-based methodology reconstructs different UCBSV haplotypes at a nationwide scale in Rwanda. ... 53
General Introduction to Chapter 4. ... 53
4.1. Abstract ... 54
4.2. Introduction ... 56
4.3. Materials and Methods ... 58
4.3.1. Study Area and field sample collection ... 58
4.3.2. Total RNA extraction ... 59
4.3.3. RNA-Seq library preparation and sequencing ... 59
4.4. Bioinformatic analysis ... 59
4.4.1. Reads processing ... 59
4.4.2. De novo assembly, Mapping, and Phylogenetic analysis ... 59
4.4.3. Phylogenetic analysis ... 60
4.4.4. Nucleotide diversity ... 60
4.4.5. Genetic diversity analysis of UCBSV ... 60
4.4.6. SNP and haplotype validation ... 61
4.5. Results ... 61
4.5.1. UCBSV is detected in nearly all generated datasets at a high abundance 61
4.5.2. Phylogenetic analysis of consensus sequences of UCBSV revealed a high homogeneity throughout the sampled districts. ... 61
4.5.3. Validating SNPs by Sanger sequencing ... 63
4.5.4. Differences in nucleotide diversity are observed between districts ... 63

4.5.5. F_{ST} methodology revealed UCBSV genetic diversity beyond the consensus sequences....63
4.5.6. Principal Component Analysis of SNPs confirmed the clustering by *the* F_{ST} approach....64
4.5.7. The identified UCBSV haplotypes are geographically clustered....66
4.5.8. Haplotype 2 distribution appears to be associated with the presence of the CBSD-tolerant cultivar NAROCAS1....68
4.6. Discussion....68
4.7. Data availability....70
4.8. Supplementary materials....71
Chapter 5: Greenhouse thermotherapy combined with chemotherapy and field chemotherapy reduced viral load and severity of CBSD. 101
General introduction to chapter 5. 101
5.1. Abstract....102
5.2. Introduction....103
5.3. Materials and Methods....105
5.3.1. Field identification of symptomatic experimental plants....105
5.3.2. Initial virus infection confirmation....105
5.3.3. Greenhouse experiments: thermotherapy, thermotherapy combined with chemotherapy, and chemotherapy. 105
5.3.4. Field chemotherapy. 108
5.3.5. Selection of housekeeping genes....108
5.3.6. Quantification of UCBSV in treated plants....109
5.3.7. Statistical analysis....109
5.3.8. Transcriptomic experimental design....109
5.3.9. RNAseq and bioinformatic data analysis....110
5.4. Results....111
5.4.1. Initial CBSIs confirmation in experimental plant....111
5.4.2. Greenhouse thermotherapy combined with chemotherapy reduced the viral load. 111
5.4.3. Field chemotherapy reduced the severity of CBSD root symptoms....113
5.4.4. Overview of cassava transcriptome 24 hour after SA or BTH spraying. 115
5.4.5. Comparative analysis of DEGs reveals overlaps between SA, BTH treatments, and CBSV infection in cassava cultivars. 116
5.4.6. Important GO terms in cassava defense were enriched by SA and BTH treatments. 116
5.5. Discussion....118
5.6. Supplementary materials....122
Chapter 6. General Discussion, Conclusions, and Recommendations. 131

6.1. General Discussion 131
6.2. General Conclusions 136
6.3. Recommendations 138
Bibliography 143
Appendices. Peer review publications and Scientific communication 173
1. Peer review publications 173
1.1. Publications and manuscripts included in the thesis 173
1.2. Other peer review publications 173
2. Scientific Communication 174

List of figures

Figure 1. Symptoms of CBSD and CMD 6
Figure 2. Potential key factors enhancing the dissemination of CBSD. 7
Figure 3. The genome organization of cassava Geminiviruses. 8
Figure 4. Map of Africa illustrating where CBSD has been confirmed. 10
Figure 5. Organization of cassava brown streak ipomoviruses genome. 11
Figure 6. Key events in CBSD outbreak and research in Rwanda (2007–2023)... 12
Figure 7. Interaction of three main factors that are essential for CBSD spread. ... 14
Figure 8. Laboratory techniques for plant viral diagnosis. 16
Figure 9. General workflow of cassava seed system- case of Rwanda. 18
Figure 10. Schematic illustration outlining the structure of the thesis. 28
Figure 11. Map of Rwanda showing surveyed districts. 36
Figure 12. Sources of cassava planting materials, trust, and disease management among farmers in Rwanda during 2019. 39
Figure 13. Incidence of CBSD based on RT-PCR in different districts of Rwanda, 2019. 41
Figure 14. Phylogenetic analysis. 62
Figure 15. Analysis of UCBSV diversity. 64
Figure 16. Principal component analysis (PCA) of the virus populations. 65
Figure 17. SNPs contributing to the differentiation between UCBSV populations. 66
Figure 18. Country-wide distribution of UCBSV Haplotypes. 67
Figure 19. Greenhouse Experimental Design. 107
Figure 20. Field chemotherapy experiment. 108
Figure 21. Transcriptomic experimental design. 110
Figure 22. Weekly mean variation in temperature inside the thermotherapy box. 112
Figure 23. UCBSV quantification at 6- and 12-weeks post treatments. 113
Figure 24. Categories of CBSD root necrosis expressed at 12 MAP following field chemotherapy with SA or BTH. 114
Figure 25. Overview of the RNAseq data obtained post-SA and BTH treatment. 115
Figure 26. Some over-represented GO terms of the category "biological process" among DEGs induced by SA and BTH in 60444 cultivars at 24-hour post SA and BTH treatment. 117
Figure 27. Over-represented GO terms in the category "biological process" for overlapping DEGs between genes up-regulated by SA and BTH and potential CBSIs responsive genes from the literature. 118

Figure S 1. Detection of CBSV and UCBSV by reverse transcription-PCR in samples from surveyed distrcits. 49
Figure S 2. Map of Rwanda showing surveyed districts and their agro-ecological zones. 71
Figure S 3. Mapping of *Cassava brown streak virus* reads to the closest reference. 71
Figure S 4. Mapping of all reads to the Congolese genome sequence of MEaV-1 Ampelovirus (MT773588). 72
Figure S 5. Confirmation of novel Ampelovirus. 72
Figure S 6. Phylogenetic analysis. 73
Figure S 7. Pairwise identity matrices of the UCBSV haplotypes. 74
Figure S 8. Traditional thermotherapy box. 124
Figure S 9. Determination of the optimal number of reference gene in cassava plant treated with thermotherapy combined with **(A)** BTH or **(B)** SA. 125
Figure S 10. A gel picture of CBSIs initial confirmation in experimental plants. 126
Figure S 11. Variation in temperature outside the greenhouse. 126
Figure S 12. Weekly mean variation of relative humidity: **A)** inside the thermotherapy box and **B)** outside. 127
Figure S 13. Principal Component Analysis of the plants sprayed with SA, BTH and controls. 128
Figure S 14. Overlap between DEGs in 60444 plants treated with SA or BTH and CBSIs infected cultivars from published papers. 128

List of tables

Table 1. Summary of approaches used for virus cleaning or for inducing virus resistance. 20
Table 2. Field incidence, severity score and frequency of cassava plants showing aerial CBSD symptoms in Rwanda, 2019. 40
Table 3. Socio-demographic factors stratified by cassava infection....42
Table 4. Impact of accessibility of planting materials on cassava infection....43
Table 5. Impact of agronomic variables on cassava infection. 44
Table 6. Disease awareness variables stratified by cassava infection. 45
Table 7. Risk factors significantly associated with CBSD. 47
Table 8. The severity of CBSD root symptoms at 12 MAP following field chemotherapy. 114

Table S 1. GPS coordinates of sampled cassava fields....74
Table S 2. Characteristics of Agro-ecological Zones (AEZs) of the districts surveyed as well as common cassava cultivars and intensity of cassava cultivation per districts 75
Table S 3. RNA concentration, 260/280 Ratio and RIN of the used samples. 76
Table S 4. RT-PCR primer pairs used for the confirmation of the essential SNPs. 76
Table S 5. High Throughput Sequencing Data of UCBSV in Rwanda. Nearly complete UCBSV was recovered in 12 per 13 districts. 77
Table S 6. Analysis of CBSV analysis among 13 districts. 77
Table S 7. Percentages of identity at nucleotide level between UCBSV consensus sequences from 12 districts of Rwanda. 78
Table S 8. Percentages of identity between the amino acid sequence of the UCBSV consensus polyprotein from 12 districts of Rwanda. 78
Table S 9. A complete list of mutations contributed to the discrimination of UCBSV populations. Most mutations were silent, but few resulted in amino acid change. ...79
Table S 10. A list of mutations confirmed by Sanger sequencing. 90
Table S 11. A complete list of all 564 SNPs used to create five dimensions. 91
Table S 12. Confirmation of UCBSV haplotype 2 based on Sanger sequencing of individual samples. 97
Table S 13. Primer sequences used in chapter 4. 122
Table S 14. Selection of stably expressed genes under thermotherapy and chemotherapy 122
Table S 15. Independent *t* test for comparison of control and various treatments in terms of t shoot heights and number of leaves for Cyizere and Mushedule cultivars. 123
Table S 16. Overview of the RNA sequencing data and Mapping. 123
Table S 17. DESEq output for all Genes. 124
Table S 18. GO Enrichment Analysis within DEGs deregulated by BTH and SA. 124

Table S 19. Dataset for Potential CBSIs responsive genes created based on the Literature and GO analysis within the DEGs overlapping between the dataset and DEGs deregulated by BTH and SA. 124

List of abbeviations

aa	:	amino acid
ACMV	:	African Cassava Mosaic Virus
AGO	:	Argonaute
AOS	:	active oxygen species
BLAST	:	Basic Local Alignment Search Tool
BTH	:	Benzothiadizole
CBSD	:	Cassava Brown Streak Disease
CBSI	:	Cassava Brown Streak Ipomovirus
CBSV	:	Cassava Brown Streak Virus
CDS	:	Cocing Sequence
CMD	:	Cassava Mosaic Disease
CP	:	Coat Protein
CRISPR	:	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	:	Cetyltrimethylammonium Bromide
CT		Cycle Threshold
DCL	:	Dicer-like
DEG	:	Differentially Expressed Gene
DNA	:	Deoxyribonucleic acid
DRC	:	Democratic Republic of Congo
EACMV	:	East African Cassava Mosaic Virus
eIF	:	eukaryotic Initiation Factor
ETI	:	Effector Triggered Immunity
F_{ST}	:	Fixation index
GO	:	Gene Ontology
H	:	Haplotype
HR	:	Hypersensitive Response
HSP	:	Heat Shock Protein
HTS	:	High Throughput Sequencing
IITA	:	International Institute of Tropical Agriculture
LRR	:	Leucine Rich Repeat
MAP	:	Month After Planting
MAPK	:	Mitogen-Activated Protein Kinase
MEaV	:	Manihot Esculenta associated Virus
NCBI	:	National Center for Biotechnology Information
nt	:	nucleotide
PAL	:	Phenylalanine Ammonia-Lyase
PCA	:	Principal Component Analysis
PepGMV	:	Pepper golden mosaic virus

PTGS	:	Post-Transcriptional Gene Silencing
PTI	:	Pattern-Triggered Immunity
QDS	:	Quality Declared Seed
RAB	:	Rwanda Agriculture and Animal Ressources Development Board
RDRP	:	RNA Dependent RNA Polymerases
RIN	:	RNA Integrity Number
RISC	:	RNA-Induced Silencing Complex
RNA	:	Ribonucleic acid
RNAi	:	RNA interference
ROS	:	Reactive Oxygen Species
RT-PCR	:	Reverse Transcription-Polymerase Chain Reaction
SA	:	Salicylic acid
SAH	:	Semi-Autotrophic Hydroponics
SAR	:	Systemic Acquired Resistance
SNP	:	Single-Nucleotide Polymorphism
UCBSV	:	Ugandan Cassava Brown Streak Virus
UTR	:	Untranslated Region
WAP	:	Week After Planting

Chapter 1

General Introduction

Chapter 1. General Introduction

General Introduction to Chapter 1

In the inaugural chapter of this thesis, a systematic and thorough literature review was undertaken to unearth the wealth of knowledge, research, and insights accumulated over the years regarding cassava viruses and their intertwined dynamics in Sub-Saharan Africa, with a specific lens on the situation in Rwanda. The overarching objective here was to delve into the multifaceted aspects of the cassava brown streak disease (CBSD) outbreak, aiming to understand its prevalence, epidemiology, economic implications, existing management strategies to mitigate CBSD's detrimental effects, and the intricate web of challenges surrounding it. Various strategies and interventions were critically assessed, highlighting their successes and shortcomings. This chapter was essential in discerning the current state of CBSD management and identifying potential areas for improvement.

Chapter 1. General Introduction

1.1. Background

Plant viruses cause enormous yield losses and spoil the quality of the produce, which can significantly impact local and national income in countries where agriculture is the backbone of the national economy (1,2). The losses are exacerbated when crops enhancing food security are affected, as it can lead to food insecurity and famine (2). Globally, the estimated losses caused by plant viruses were over US dollars 30 billion in 2014 (3). The combination of viral pandemics, which represent almost half of the re-emerging and emerging plant diseases, and increasing food demand with the rapidly growing world population makes resistance to plant viruses a cornerstone strategy to increase the resilience of agrosystems (2,4,5). Vegetatively propagated crops such as cassava, potato, banana, and others are more vulnerable to viral infections than other crops because viruses accumulate during cycles of propagation and are disseminated from one cycle to the next (6).

Cassava (*Manihot esculenta* Crantz; family: *Euphorbiaceae*) originated from Latin America and was introduced to Africa by Portuguese seafarers in the sixteenth century (7). It is an important staple food for food security and income generation for approximately 800 million people across the globe in several developing and emerging countries (8). Regarding global annual production, cassava ranks among the ten most important crops, with 302,6 million tons (9). In Africa, the crop supports the livelihoods of over 300 million consumers and its importance peaks in Central-East Africa, where around 200 million people use cassava as a daily staple (10,11).

Cassava is often called a drought war famine crop because it is a resilient crop propagated vegetatively with a notable tolerance to dire environmental conditions and abiotic stresses. The crop can produce reasonable yields on depleted and marginal land, including acidic soil, where other crops would fail without using fertilizers (8,12,13). It is also a significant source of high-purity starch exploited in the industry and ranks second to maize (Zea mays L) for starch production globally (14).

In Rwanda, where over 80% of the population lives in rural areas and relies on agriculture, cassava is the third staple food after banana and sweet potato (15). However, it ranks second in terms of area underproduction after bananas. Cassava is particularly important in the East and South provinces, grown on about 21.5 % (197,400ha) of the cultivation area (National Institute of Statistics of Rwanda, 2020).

Around 70% of the world's cassava is consumed by humans, whereas the remaining 30% is used as animal feed and industrial raw materials (12,17). Several parts of cassava can be processed for consumption. First, the sweet cassava varieties (without cyanide) can be eaten raw or boiled. In contrast, the bitter varieties (with cyanide) must be processed into flour to make bread, acheke, beer brewing, etc. Second, cassava leaves are vegetables rich in protein, vitamins A, C, iron, and calcium and are prepared as vegetables in fresh or dried form to supplement the root, which is poor in proteins (18). Third, the peels from the roots are used for animal feed and industries, and stems, apart from being planting materials, can also be used as firewood or as a substrate for growing mushrooms (10,17,19,20).

Furthermore, cassava is a bioenergy crop that can revolutionize developing countries' economies due to its potential industrial application. Its starch could be used as raw materials for pharmaceuticals, textiles, butanol and ethanol production (21). Indeed, bioethanol is mainly produced from sugar crops such as sugarcane or starch-rich crops such as corn, wheat and cassava. Even though cassava has a high starch content that can be produced at a competitive price and efficiently transformed into ethanol, corn and wheat remain the most commonly used raw materials for bioethanol production, yielding 2050 L/T and 1560 L/T, respectively, while cassava could reach 6000 L/T (21). This is because significant cassava producers use its starchy root mainly for food, and its production is not at optimum, resulting in neglected bioenergy function (22).

1.1. Cassava Production and its Constraints

In 2019, the world's cassava production was 311.5 million tons, with over 60% of it being produced in Africa, making it the continent with the largest cassava production (23). Nigeria is the largest cassava producer with over 59 million tons, followed by DRC, Thailand and Ghana with 40, 31, and 22 million tons, respectively (24), and its production is mainly consumed domestically. Cassava has long been identified as a critical factor in revolutionizing Nigeria's economy through exports that would result in up to $2.98b in income yearly (25). Tanzania ranks first producer in East Africa, with around 4.5 million tons (8,26). Under optimum conditions, cassava yield would reach up to 90 tons of fresh root per hectare for improved varieties (30 tonnes of dry matter /ha)(12). Nevertheless, cassava's average yield remains low worldwide, in Asia, South America and Africa, with 12.8, 22, 14, and 9 t / ha, respectively (27).

Differences in use of improved genotypes, disease pressure, compliance with good agriculture practices (GAP), soil fertility management, weed management, late or early planting, fertilizer application are some of the factors that could explain the difference in yield without forgetting that high cassava production strongly depends on quality of stem cuttings grown which is also an element worth improvement for increasing cassava yield (28–32).

Many abiotic and biotic factors constrain the production of cassava. The rapid post-harvest physiological deterioration, estimated to cause up to 12 % total production loss, is one of the most critical abiotic constraints worldwide (18). One way of mitigating post-harvest loss is to rely on the possibility of staggering the cassava harvest; however, it can affect starch and root quality (Codycovefarm, 2022). Other abiotic challenges include a shortage of improved genotypes, drought, low fertile soil, inadequate fertilizer application, and small arable land that hamper cassava production (18,34). Besides the abiotic challenges, cassava production is also affected by biotic constraints, particularly viral diseases. The two most common viral diseases for cassava are cassava mosaic disease (CMD) which is present in all regions growing cassava in Africa, and cassava brown streak disease (CBSD), which has so far only been reported in Eastern and Central parts of Africa and is the most devastating threat to cassava (35,36). Viruses which cause both CBSD and CMD are propagated through the exchange of infected planting materials and by the whitefly vector *Bemisia tabaci* (Gennadius) (family Aleyrodidae) (37–40).

Unlike CMD, CBSD is transmitted semi-persistently by *Bemisia tabaci*, and its transmission capacity is very inefficient under laboratory conditions (35,41). Fortunately, grafting methods have been developed to offer alternative means of transmission for CBSD studies (42–45). On the other hand, CBSD spread is enhanced by the exchange of infected cassava cuttings among farmers. This is because some genotypes display very mild or no symptoms on the leaves, making it difficult to recognize even for trained people, which is not the case for CMD. Indeed, CBSD-infected cassava plant could display symptoms on leaves, stem and roots (**Figure 1**). However, sometimes the plants's leaves and stem may look asymptomatic when the roots are heavily affected; thus, there is a likelihood that stems of infected plants will be used by farmers to plant the crop for the subsequent season (35).

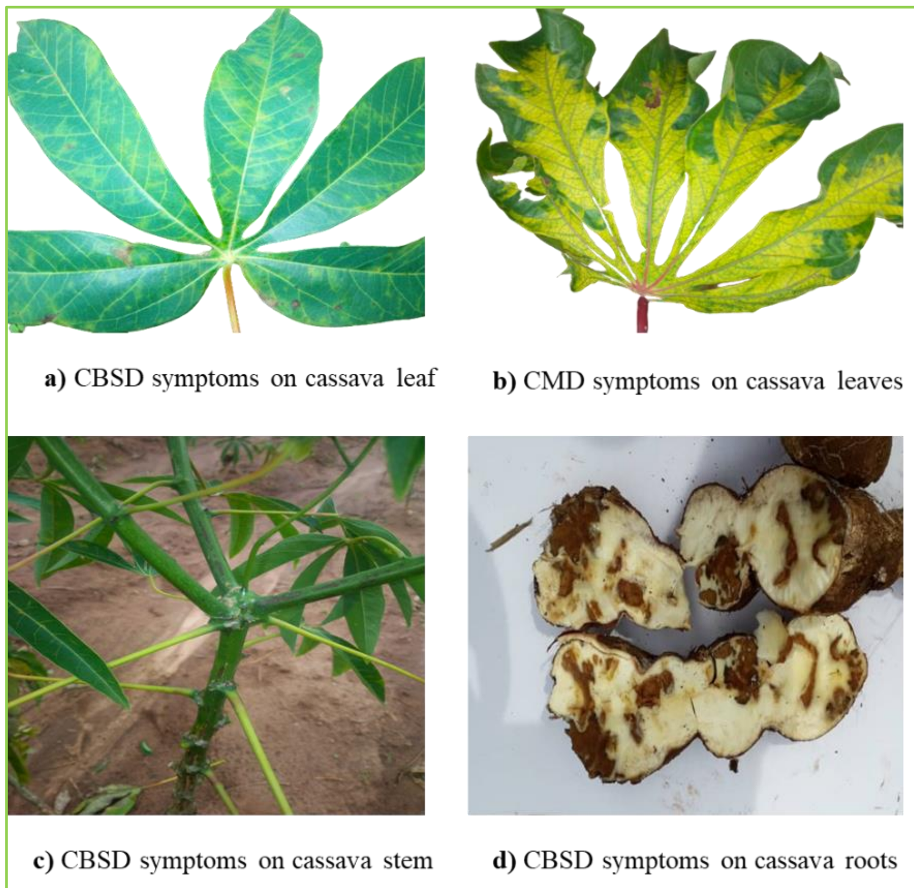


Figure 1. Symptoms of CBSD and CMD

a) CBSD leaf chlorosis running along the veins, **b)** chlorotic and distortion of leaf lamina on CMD infected leaf, **c)** CBSD dark streaks on the stem, **d)** Dark brown necrotic area within CBSD infected cassava roots.

Several studies have pointed out various other factors contributing to the dissemination of CBSD (35,46), which are summarized in **Figure 2**. Indeed, the fact that CBSD is mainly disseminated through planting materials highlights that identifying the pathways followed by cassava seeds from one actor to another in the cassava seed system, coupled with increasing farmers' awareness of viral disease management, would significantly contribute to reducing CBSD incidence (46). In addition, encouraging private sector investment in sustainable quality cassava seed systems would greatly improve the availability and accessibility of cassava quality seed in Rwanda which is an essential element for generating high income from cassava. Studies on commercial cassava seed systems have shown that cassava seed entrepreneurs could obtain a profit ranging from US$ 551 to 988/ha in Nigeria and US$ 1,000 to 1,500/ha in Tanzania (47).

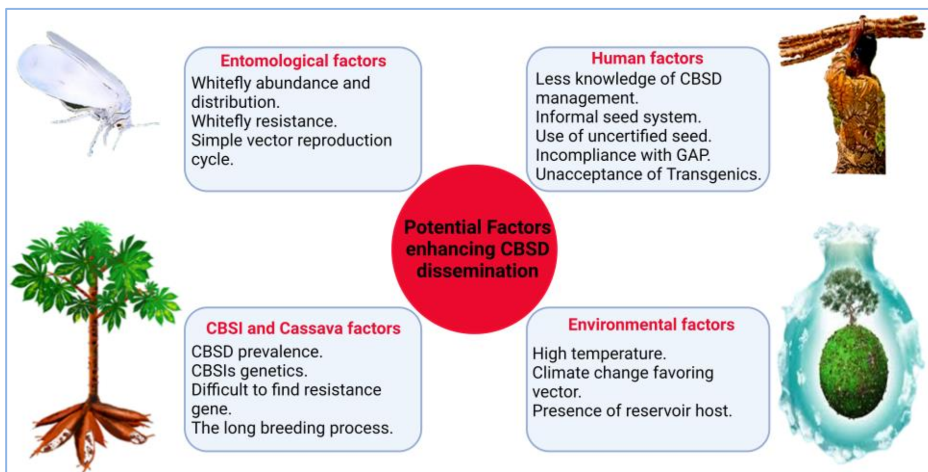


Figure 2. Potential key factors enhancing the dissemination of CBSD.

1.2. Epidemiology, Diversity, and Burden of major cassava viral diseases in sub-Saharan Africa

CBSD is currently the greatest enemy to cassava production in Central East African countries, and CMD remains the major constraint in countries where CBSD has not yet been reported. These diseases are two of the greatest threats to food security in sub-Saharan Africa and cause losses equivalent to 1 billion US$ annually (Legg *et al*., 2006; IITA, 2014).

1.3.1. Cassava Mosaic Disease (CMD)

CMD has been known for many years, and by the 1940s, it was already widely spread out in most cassava-growing countries in sub-Saharan Africa (4,49). It is caused by numerous circular single-stranded DNA with twin particles designated as DNA-A and DNA-B, each with a genome from 2.7–2.8 kb, in the family *Geminiviridae*, genus *Begomovirus*, collectively known as Cassava Mosaic Geminiviruses (CMGs).

At least nine species are associated with the CMD, and seven of them have been reported in sub-Saharan Africa (50). In East Africa, *African cassava mosaic virus* (ACMV) and *East African cassava mosaic virus* (EACMV) are the two most important causative agents of CMD and their synergistic effect results in severe loss in dual infections (36,51). **Figure 3** illustrates the genome structures of CMGs and the viral proteins they encode with their functions.

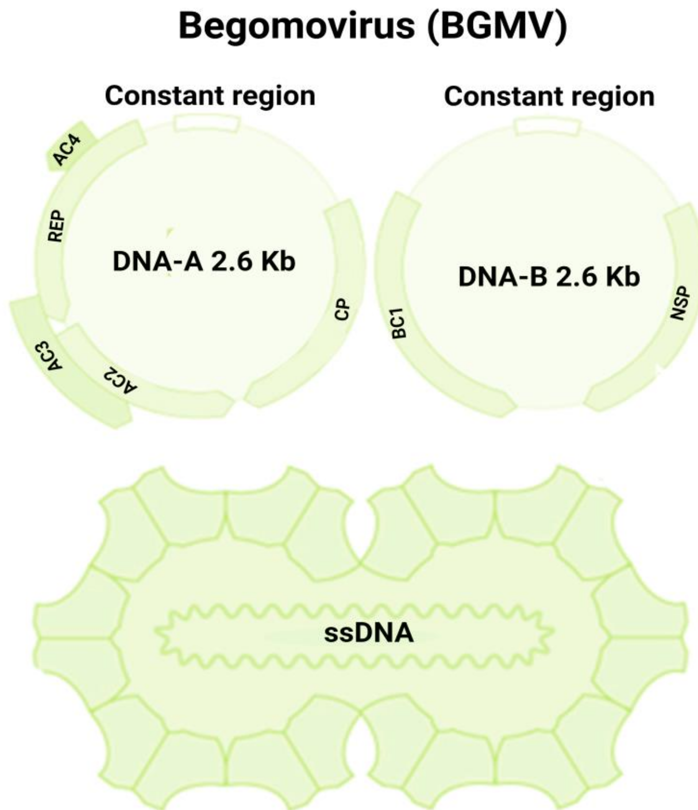


Figure 3. The genome organization of cassava Geminiviruses.

The genome of cassava Geminiviruses is Mono-, or bipartite, circular, positive sense single strand DNA genome (ssDNA+) of about 2.5-3.0 (monopartite) or 4.8-5.6 kb (bipartite). 3' terminus has no poly(A) tract. DNA-A encodes six open reading frames (ORF): AV1 and AV2 translated in the virion sense (which encodes coat protein and putative protein kinase respectively); Rep (encodes replication initiation protein), AC2 (encodes transcription activator protein, AC3 (encodes replication enhancer protein), and AC4 (encodes silencing suppressor protein) which are translated in the antisense. DNA-B has two ORF; The antisense *BC1 (*encodes movement protein) and the sense *NSP (*encodes nuclear shuttle protein), The common region contains an origine of replication that enable the virus to replicate through rolling circle amplification (52,53).

The CMD has caused yield losses of about 47% in East and Central Africa during the outbreak of an unusually severe form of the disease that occurred between the 1990s and mid-2000s and has resulted in a global economic loss of 1.2 to 2.3 billion USD, equivalent to 12-23 million tons of cassava roots (54–56). Planting infected materials leads to the highest yield loss, followed by early infection in the first 5 months after planting, whereas infection that starts from 6 months causes less yield loss (57). The production of cassava in Rwanda was heavily affected by the cassava mosaic pandemic, with a decrease evaluated at around 25% of the production at the country level (Legg *et al.*, 2001). The breeding for CMD resistance started after the discovery of a polygenic CMD resistance gene (referred to as CMD1) in *Manihot glaziovii* that was introgressed into cassava cultivar. Furthermore, the monogenic CMD resistance locus referred to as *CMD2* was identified in West African landraces (TME 3), and both CMD resistance has been combined through the crossing, giving rise to *CMD3* (58–60). Subsequent CMD management was primarily based on disseminating CMD-resistant cultivars (61); unfortunately, they all became susceptible to the CBSD.

1.3.2. Cassava Brown Streak Disease (CBSD)

The first report of CBSD occurred in Usambara Mountain in Tanganyika, now Tanzania, in 1936, but the source of the virus was not well known (62). Initially, the disease remained endemic in lowlands (i.e. below 500 m above sea level (masl)) in East Africa (63,64). In 2004, CBSD emerged in mid-altitude areas of Uganda (i.e. above 1200 masl) in which it had not previously been reported. It subsequently spread to affect over ten countries in East and Central Africa (35,36,65–68) (**Figure 4**). Although both species are widely distributed, epidemiological studies in different countries such as Mayotte, Malawi, Zambia, and Rwanda reported UCBSV to occur at a higher prevalence than CBSV (69–71).

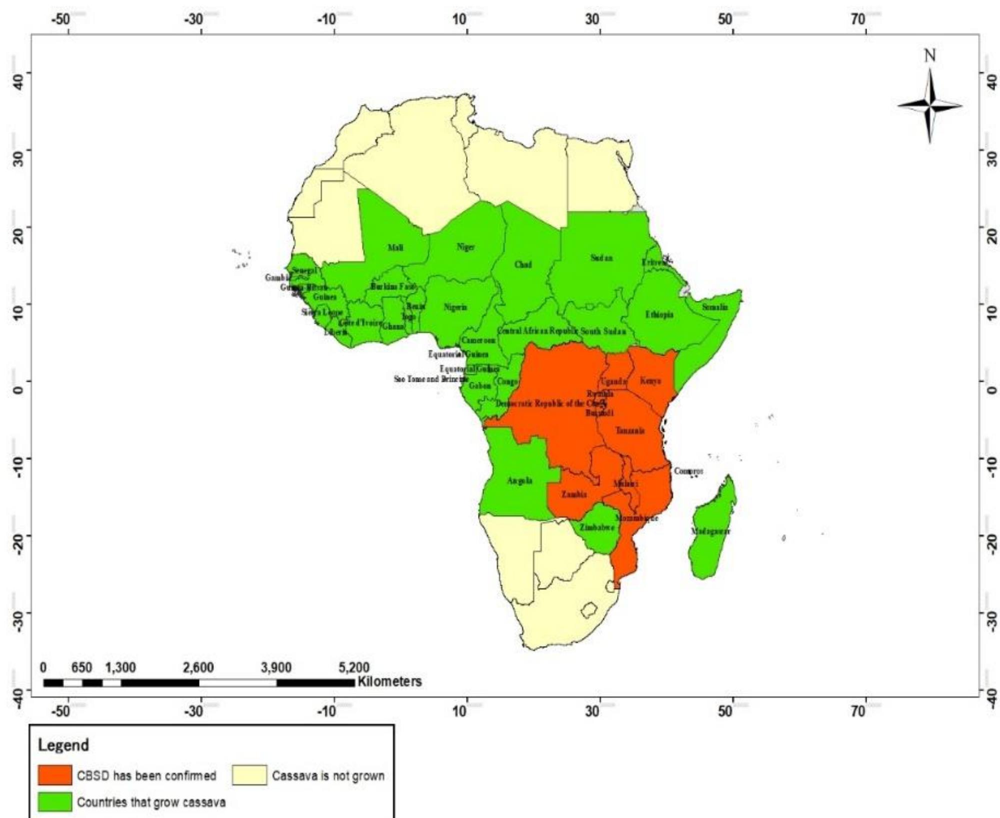


Figure 4. Map of Africa illustrating where CBSD has been confirmed.

CBSD is caused by two cassava brown streak Ipomovirus (CBSI) species from the Potyviridae family (72,73). *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), collectively designed cassava brown streak ipomoviruses (CBSIs), are both single-stranded RNA viruses with 73.6-74.4% and 69-70.3% amino acid and nucleotide sequence identity, respectively. The nucleotide identity between UCBSV species ranges from 87 -99%, whereas for CBSV species, it is 79-95% (74,75). The sequences analysis also revealed that UCBSV is 9070 nt with 5' 134 nt and 3' 227 nt untranslated region (UTR) while CBSV is 9069 nt with 5' 226 nt and 3' 131 nt UTR. UCBSV and CBSV have a single ORF, which encodes for a polyprotein of 2,902 amino acids. The polyprotein is cleaved into 10 mature viral proteins with different functions (**Figure 5**) (36,76,77).

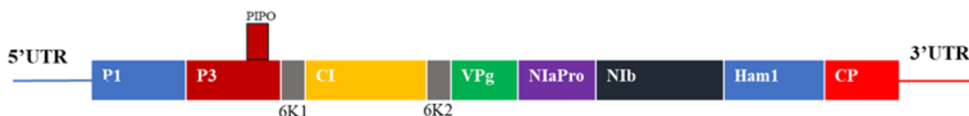


Figure 5. Organization of cassava brown streak ipomoviruses genome.

The genome of cassava brown streak ipomoviruses starts with a 5' UTR, **P1**encodes serine proteinase protein, ***P3*** encodes the third protein as well as P3N-PIPO (pretty interesting *Potyviridae* ORF), which is generated by a +2 frameshift, ***6K1*** encodes a 6-kDa protein, ***CI*** encodes a cylindrical inclusion protein, ***6K2*** encodes a 6-kDa protein, ***VPg*** encodes a viral genome-linked protein, ***NIaPro*** encodes a nuclear inclusion proteinase while ***NIb*** encodes nuclear inclusion polymerase, ***HAM1h*** encodes a reduction of mutation rate protein, ***CP*** encodes the coat protein. The sequence is completed by a 3' UTR and polyA tail. CBSIs have unusual features such as the presence of a single P1 serine protein that suppresses RNA silencing, the absence of a helper component proteinase protein (HC-Pro) and the presence of a novel Ham1 protein (77,78).

High Throughput Sequencing (HTS) has enabled the in-depth characterization of cassava ipomovirus sequences. The analysis revealed that CBSV evolves five times faster than UCBSV and Nla, followed by 6K2, Nlb, and P1 display high rates of evolution. Previous studies have also suggested that CBSV is more virulent and has more strategies to evade cassava immune response than UCBSV (79).

CBSD makes cassava roots unsuitable for use, which worsens with the plant's age and causes a substantial loss of yield of up to 70%, affecting income and food security (80). At the regional level, around 1.6 million tons of fresh roots are lost every year due to CBSD in only 8 countries of East and Central Africa, which is equivalent to over US$75 million (48,55,66). This highlights the huge food insecurity that can occur when the disease spreads as a pandemic and emphasizes the importance of mitigation and control.

1.4. Status of CBSD in Rwanda

In Rwanda, cassava has been selected as a priority crop by the government to ensure food security and increase small-scale farmers' income. Since 2007, the national food crop intensification programme has distributed to farmers 140 million stem cuttings of improved, CMD-resistant varieties such as TMS 30337, TMS 30572, TMS 30395, TMS 30001, and TMS 60142 delivered from International Institute of Tropical Agriculture (IITA), Nigeria (81), and provided them with fertilizer and extension advice (82). However, a CBSD outbreak reached the country in 2009, affecting the CMD-resistant cultivars distributed to farmers (83). The CBSD incidence has rapidly spread in Rwanda as it increased from 18.5% in 2012 to 69% in 2014, leading to significant crop loss and a shortage of planting materials (84). Consequently, a significant decline in cassava production from 3.3 million tonnes in 2011 to 900,000 tonnes in 2014 led to the price of cassava flour doubling by April 2015 compared to the average price in the previous five years (85). Cassava is a food security and a cash crop in Rwanda; the yield losses caused an essential impact on farmers who complained about losing money, reduced employment and food insecurity (84).

According to the Ministry of Agriculture, 19 % of the country's households have been categorized as food insecure, and increasing average yields and overall cassava production could help increase food security for over 700,000 families (86). In response to that crisis, the Government of Rwanda, through the Ministry of Agriculture and the Rwanda Agriculture and Animal Ressources Development Board (RAB), imported cuttings of CMD and CBSD tolerant varieties from Uganda, including NASE14 and NAROCAS1. Imported cuttings were multiplied and disseminated at a large scale to farmers from 2015 onwards to mitigate the outbreak (84). Since cuttings did not reach every farmer directly, the '*kwitura'* concept was established whereby cassava farmers gave back to RAB the same numbers of cuttings they received so that they could be distributed to other farmers (87). Thanks to disease management and good agriculture practices, average cassava yields rose from 12.3 tonnes in 2011 to 14.2 tonnes in 2021 (88–90). **Figure 6** summarizes the critical cassava viral disease events in Rwanda.

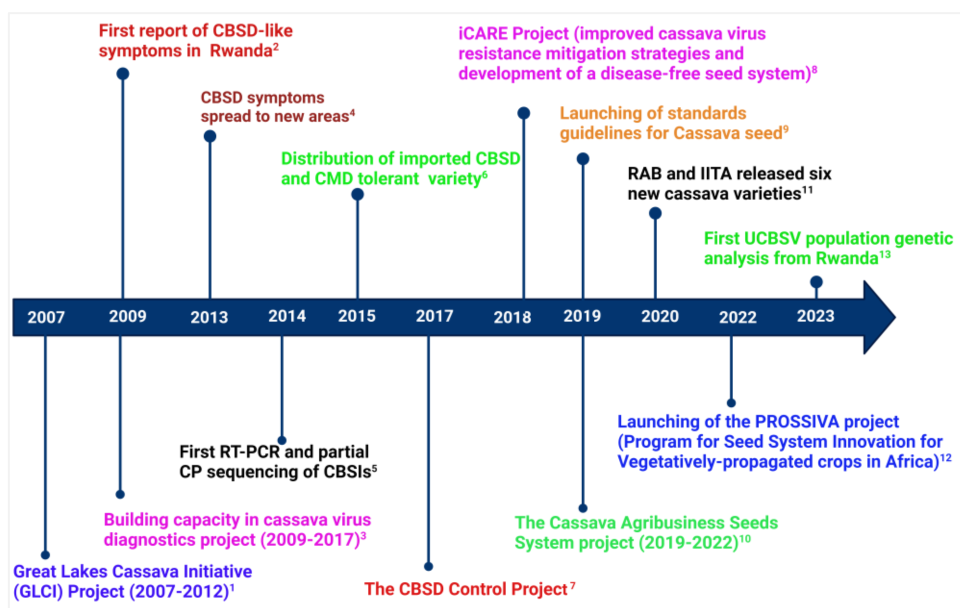


Figure 6. Key events in CBSD outbreak and research in Rwanda (2007–2023).

[1] The Great Lakes Cassava Initiative (GLCI) Project (2007-2012) enhanced surveillance, participatory varietal selection; germplasm multiplication; awareness raising; CBSD focus; seed systems; virus diagnostics (91). CMD-resistant varieties were released in 2008 (87). The government of Rwanda initiated crop intensification, with cassava among the prioritized crops.

[2]First CBSD-like symptoms were recorded in the Muhanga district in the south, Bugesera and Nyagatare districts in the East (92).

[3] The project of Building capacity in cassava virus diagnostics strengthened the capacity of CMD and CBSD virus diagnostics; surveillance; and sustainable virus management (91).

[4] CBSD symptoms spread to new areas in Gisagara, Nyanza, and Ruhango districts in the south and Kirehe districts in the East. There was a severe CBSD outbreak (RAB, unpublished data).
[5] First confirmation of CBSIs based on RT-PCR and partial coat protein (CP) sequencing with high CBSD incidence (69%)(93). There was a shortage of cuttings (87).
[6] CBSD and CMD tolerant cultivar (NASE14 and NAROCAS1) were imported from Uganda and distributed to cassava farmers(94).
[7] The CBSD Control Project introduced 17 elite cassava varieties, installed a new screen house for seed multiplication, trained tissue culture technicians, and initiated Semi-Autotrophic Hydroponics (SAH) (87,95)
[8] iCARE Project equipped biotechnology laboratory for virus indexing, Initiated breeding activities for CMD and CBSD resistance, CBSIs genetic diversity, seed system survey, and Capacity building.
[9] RSB and IITA Launched Standards guidelines for cassava seeds and offered training to carry out inspection and certification (87).
[10] The Cassava Agribusiness Seeds System (CASS) project was initiated with the aim to enable agribusiness development for scaling quality cassava seed systems for control of CBSD and CMD in Rwanda and Burundi (96).
[11] Six new cassava cultivars (Biseruka, Tegereza, Gikungu, Buryohe, Nsizebashonje, and Tebuka) with improved yield, CBSD and CMD resistance were released (97).
[12] IITA started a five years project to boost vegetatively-propagated crop seed systems in Africa (98).
[13] First UCBSV population genetic analysis from Rwanda was done (Nyirakanani *et al.*, 2023).

1.5. Strategies for cassava viral disease management

CBSI spread is the result of the continuous interaction of three factors: source of infection, mode of transmission and availability of susceptible or alternative hosts in nearby the fields. The management of CBSD requires multicomponent approaches that break their interaction, from knowledge of the pathogen, identifying and tracking it to preventing the spread and controlling damage (**Figure 7**).

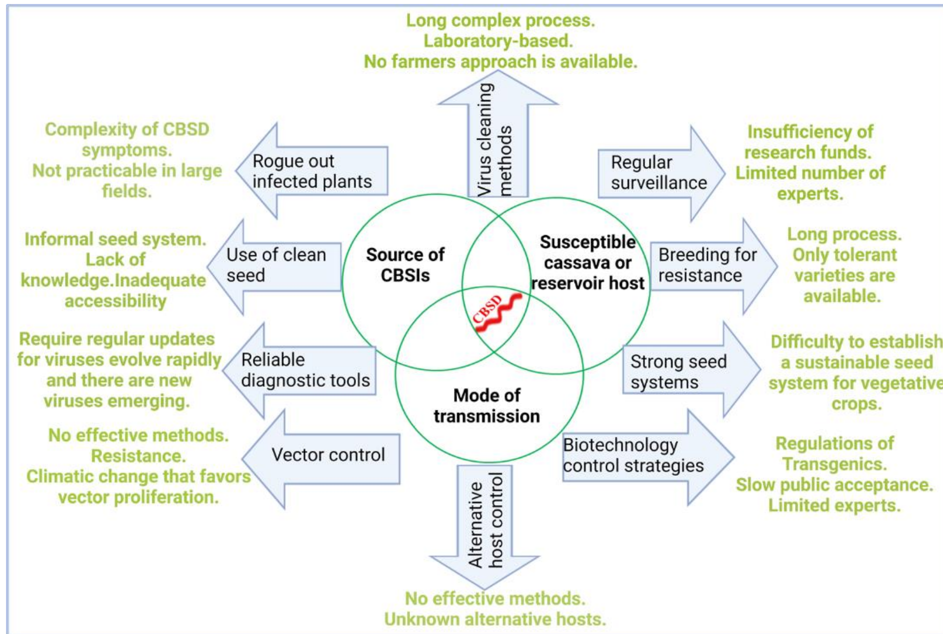


Figure 7. Interaction of three main factors that are essential for CBSD spread.
Blue arrows: management approaches; **Text in green**: limitations of the management approaches.

1.5.1. Knowledge of the pathogen

Knowledge and identification of the disease-causing agent is always the starting point of its management. CBSD produce symptoms which are not always easily recognized by farmers. Therefore, increasing awareness of the disease and its transmission among cassava farmers is an essential element in the management of the disease as it was realized that farmers with knowledge about the disease comply better with the management strategies (28,99,100). Furthermore, regular surveillance is required to monitor the geographical incidence of CBSD. Surveillance require accurate laboratory diagnosis using appropriate molecular diagnostic tools as CBSD visual inspection is often inadequate (65).

1.5.2. Diagnosis

Accurate and fast identification of the causal agent is a pivotal element in disease management. Diagnostic techniques must be sensitive enough to eliminate false negatives and specific to eliminate false positives and false negative (101). Several factors should be considered in order to get accurate and reliable results, including types of tissue, sampling method, period of sampling, ensuring quality assurance (QA), and quality control (QC) at pre-analytical, analytical, and post-analytical stages (101). Besides the direct virus visualization on electron microscopy, the diagnostic approaches commonly used are mainly serological methods like enzyme linked immunoassay (ELISA), lateral flow assay (LFA), dot immunobinding assay and tissue

blot immune assay (102–104), which applies the antigen-antibody recognition principle and molecular methods which apply principles of nucleic acid amplification, such as polymerase chain reaction (PCR) and its numerous variants (RT-PCR, nested PCR, multiplex PCR, etc.), real-time or quantitative PCR (qPCR) (105,106), LAMP (loop-mediated isothermal amplification) (107), Recombinase polymerase amplification (RPA) (108), Immunocapture-PCR (109) and hybridization like Western blot, Southern blot and Northern blot (110–112). Virus populations, particularly RNA viruses, display a high mutation rate, and many emerging and re-emerging plant viruses merit attention. Unfortunately, the mentioned diagnosis techniques are limited only to known viruses. High throughput sequencing (HTS) has revolutionized plant virus diagnosis as it enables the detection of both known and unknown viruses present in a sample allowing the discovery of novel viruses (113–115). For example, HTS has recently facilitated the discovery and comprehensive characterization of a previously unknown ampelovirus, identified as *Manihot esculenta-associated viruses 1* and *2* (MEaV-1 and 2). This ampelovirus has been found in cassava plants in the Democratic Republic of Congo (DRC), Mayotte, Madagascar and La Réunion (116). Ampeloviruses belong to the group of single-stranded RNA positive-sense viruses and possess a linear genome ranging in size from 13.7 to 18.5 kilobases (kb) (116). They are naturally transmitted semi-persistently by mealybugs and can also be disseminated over longer distances via infected stem materials. However, a comprehensive understanding of the symptoms induced by MEaV-1 and 2 and their broader impact on cassava production should be further investigated. The detailed characterization of virus genetic variability provides valuable information on the virus epidemiology and evolution, and it is essential for developing reliable diagnostic tools that contribute to disease management programs. The information on UCBSV and CBSV, whole genome sequences from Rwanda, is scarce and could provide a more detailed understanding of virus evolution across the country and help optimize the existing molecular diagnostic tools. **Figure 8** below summarizes the diagnostic techniques for detection of plant viruses.

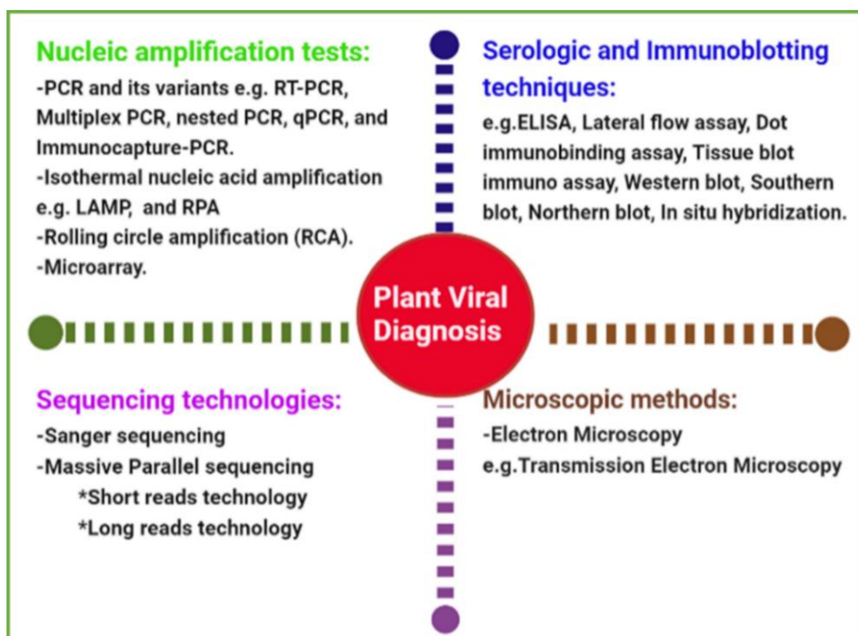


Figure 8. Laboratory techniques for plant viral diagnosis.

1.5.3. Use of virus or whitefly-resistant cassava

Mitigating vector transmission of viruses in open cassava fields remains challenging. Thus, the use of cassava varieties that are resistant to either the virus or the whitefly vector represents a robust strategy to address the issue of viral diseases in the field. Unfortunately, unlike for CMD, where *CMD2* confers geminivirus resistance, robust resistance against CBSD under field conditions remains to be identified and confirmed. Thanks to the extensive breeding studies that have been carried out, CBSD tolerant Namikonga cultivar was identified and incorporated in various crosses resulting in tolerant cultivars like NASE1 and NASE14 that were distributed to farmers (117–119). Interestingly, a recent screen with graft inoculation of CBSIs under greenhouse and field conditions has identified three cassava genotypes with high resistance levels against CBSIs. It includes two genotypes originating from Colombia (COL 40; COL 2182) and one from Peru (PER 556), which did not support the replication of CBSIs (120). Furthermore, promising results regarding breeding for dual resistance against CMD and CBSD, which represents the optimal solution for cassava viral management, were achieved in greenhouse experiments. However, additional field evaluations are still necessary to validate these findings (120). High whitefly populations have been linked to cassava virus epidemics in East Africa, and their control would significantly reduce the spread of current epidemics as well as the risk of the emergence of new virus variants (35).

Indeed, the distributed CMD-tolerant genotypes appeared highly susceptible to whitefly infestation (121). Therefore, whitefly resistance should be combined with *CMD2*-based tolerance and CBSD resistance to limit the virus's circulation.

Breeders have identified cassava genotypes with resistance to whitefly, including Ugandan landraces like Nabwire 1 and Ofumba Cai, as well as the South American genotype MEcu 72 (121,122), that could serve in breeding research. Whitefly control in other crop systems primarily relies on chemical pesticides (123). However, the overuse of pesticides in agriculture has led to resistance development and a negative impact on the environment, beneficial insects (pollinators), and humans (124). Fortunately, specific insect growth regulators (IGRs), sometimes referred to as insect birth control, represent an alternative to classical insecticides due to their ability to target specific insect stages, which makes them suitable for integrated pest management (IPM) due to their effectiveness, their environmental friendly nature (biodegradable, nonpolluting, and nonpersistent) and less toxicity to humans (125). To overcome rapid insecticide resistance, strategies based on offering a refuge host (not insecticide treated) to promote the susceptible whitefly population's survival while reducing the resistant population's fitness and using IGRs could be effective. A study on the sweet potato whitefly (*Bemisia tabaci*) from cotton fields proved that cotton refuges delayed its resistance to pyriproxyfen (126).

Transgenic approaches for controlling whiteflies have also been developed targeting the silencing of at least one of its vital genes, including acetylcholinesterase (AChE), *v-ATPase,* sex lethal (Sxl) protein, orcokinin (Orc), and ecdysone receptor (EcR) genes (127). A study done on lettuce crops genetically engineered to express dsRNAs targeting the whitefly novel v-ATPase gene showed a mortality rate of up to 98% within 5 days of feeding on transgenic lettuce and 95-fold less in whitefly eggs (128). Likewise, *Nicotiana tabacum* expressing dsRNA homologous to *Bemisia tabaci* AChE and EcR recorded whitefly mortality of over 90% within 3 days of feeding (129).

1.5.4. Phytosanitary measures

Phytosanitary measures such as roguing and selection of healthy stems for planting are essential for disease management. However, these methods are more practicable for CMD as infected plants present easily recognizable symptoms. By contrast, CBSD symptoms are much less conspicuous, which makes phytosanitary control difficult. For instance, a CBSD-infected plant may appear healthy on the aerial parts while presenting CBSD root symptoms (63). With plant disease outbreaks, it is essential to make healthy seeds accessible to farmers. To boost this aspect of phytosanitation in Rwanda, the Ministry of Agriculture, in collaboration with research institutions, has established a system to enable access to disease-free planting materials. The process involves a chain of actors, starting with the RAB, where breeders produce pre-basic clean seeds. Techniques like the high-ratio propagation technology called Semi-Autotrophic Hydroponic (SAH) have been adopted for quick multiplication (130).The SAH technique is more efficient than the conventional tissue culture technique as it offers a cheaper and quicker (2 weeks over 2 months for classical tissue culture) high clean seed multiplication ratio. Moreover, SAH allows almost a 100% success rate of weaning (transferring tissue culture plantlets into a field) (131). Next, clean seeds are then distributed to professional cassava seed multipliers in all major cassava growing areas for further multiplication. Professional seed multipliers are vital actors that help to scale up new disease-resistant cassava varieties to many farmers.

Standards for cassava quality declared seed (QDS) were approved by the Rwanda Standards Board (RSB) in 2018, making formal seed available for Rwandan farmers. To support this process, the agency of certification is involved in the inspection of cassava multiplication fields before their dissemination to farmers to ensure the quality of the planting materials (87,132–134) (**Figure 9**).

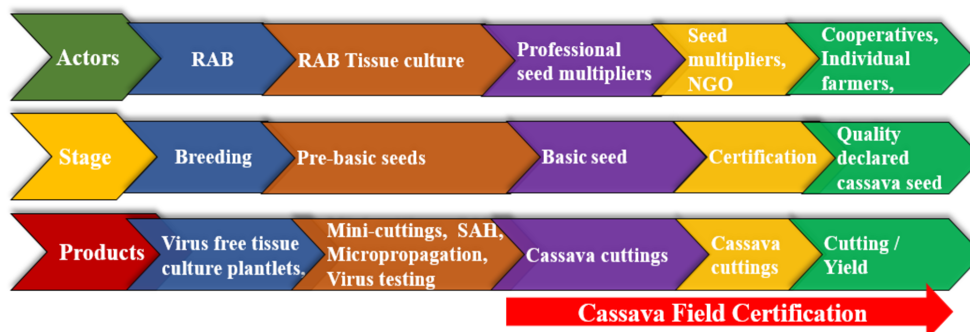


Figure 9. General workflow of cassava seed system- case of Rwanda.

Implementation of phytosanitary measures coupled with introduction of improved varieties has been shown to produce a significant reduction in CBSD incidence from > 90% to 27% at the community-level (135). However, the sustainability of a formal seed system needs to be improved by many factors, including the maintenance of an informal seed system by farmers, which often leads to the continuous supply of diseased planting material to cassava-growing communities.

1.5.5. Cassava virus cleaning

Supplying clean planting materials to farmers should be sustainable as available cassava cultivars are not immune to CBSD, and planting material becomes increasingly affected by virus disease over repeated cropping cycles (6). Therefore, to avoid virus build-up, laboratory techniques have been developed for virus cleaning to provide virus-free plants at the top level of the seed system. These techniques include meristem tip culture, thermotherapy, and chemotherapy (136–138).

i) Meristem tip culture

The principle of meristem tip culture relies on the non-uniform distribution of viruses within the plant and the rapid cell division characteristic of the meristem, which inhibits viral replication, resulting in a decreased virus concentration gradient towards the plant tip (139,140). This technique has been applied in cleaning viruses from different plants including cassava. Notably, the survival rate of these plantlets is inversely related to the size of the excised meristem tip, a factor critical to both the regeneration rate and the establishment of virus-free plants (141). Various studies indicated that a meristem tip measuring 0.5 mm in length enhances the production of virus-free plantlets, achieving success rates of up to 80%. Consequently, a meristem tips ranging from 0.1 to 0.5 mm in length is the optimal choice for the efficient elimination of pathogenic viruses, including CBSIs and CMGs (136).

ii) Thermotherapy

The effectiveness of elevated temperatures in eliminating viruses from plants was observed around a semi-century ago when researchers found that alfalfa and cucumber viruses could be successfully eradicated from *Nicotiana rustica* plants after subjecting them to a temperature of 32°C for a continuous period of 30 days (142). This intriguing phenomenon involves a mechanism primarily centered around the induction of the natural plant defense called RNA silencing, leading to the restriction of virus replication within the plant, ultimately resulting in reduced virus titers and, consequently, the complete elimination of the virus (143,144). This approach has also been successfully applied in managing cassava viruses, where a combination of thermotherapy followed by meristem culture completely cleared *cassava mosaic virus* (CMV) from cassava (145). Moreover, cassava plantlets grown at 38°C for 21 days, followed by meristem tip (0.5 mm) culture, produced over 80% of CBSV clean plantlets (136).

iii) Chemotherapy

The efficacy of chemotherapy to combat viral infections in plants is underpinned by the capacity of specific chemical inducers, which have been demonstrated to impede virus proliferation through distinct mechanisms including the induction of mutations in the virus, as exemplified by the action of ribavirin (146,147). Specific chemical inducers can also activate signaling pathways associated with disease resistance and RNA interference (RNAi), a vital defense mechanism against viruses (148–151). Numerous studies have been conducted to assess various chemicals' ability to induce resistance, cleanse, or reduce viral presence. For instance, salicylic acid has been found to induce resistance against the *tomato yellow leaf curl virus* (152). Notably, applying salicylic acid at a concentration of 30 mg/L eliminated cassava brown streak ipomovirus (CBSIs), achieving a 100% success rate (136). Ribavirin is another chemical agent that has been shown to effectively reduce the presence of *cassava mosaic virus* (CMV), achieving 80% reduction (138). Furthermore, benzothiadiazole has demonstrated the capability to induce resistance against the *pepper golden mosaic virus* (153).

Table 1 provides a comprehensive summary of the strategies employed to induce virus resistance or achieve virus cleaning, along with the corresponding research findings.

Table 1. Summary of approaches used for virus cleaning or for inducing virus resistance.

Approach / Method	Results	References
In vitro chemotherapy: Ribavirin 20mg/L	85.0% of *East African cassava mosaic virus* free plantlets.	(138)
In vitro chemotherapy: Salycilic acid (SA) 30mg/L	88.9% *East African Cassava mosaic virus*-free plantlets	(138)
In vitro chemotherapy: Ribavirin 0.10mM	Up to 40% CBSIs -free plants	(137)
In vitro chemotherapy: Ribavirin 30mg/L	88% plantlets free from CBSV	(136)
In vitro chemotherapy: Salycilic acid 30mg/L	100% CBSV-free plants	(136)
Exogenous application of 2 mM (SA)	Induced resistance to the *tomato yellow leaf curl virus*	(152)
Application of Acibenzolar-S-methyl (ASM) at 50 or 100 µg/ml in greenhouse	Induced resistance against *Colletotrichum lagenarium* fungi and *Cucumber mosaic virus* in cantaloupe	(154).
Spraying 300mg/L of benzothiadiazole (BTH)on pepper plants followed by *Pepper golden mosaic virus* (PepGMV) inoculation	Induced resistance to *pepper golden mosaic virus*	(153)
Foliar application of SA- 0.5 mM on tomato seedlings	SA collaborates with gene silencing in tomato defense against *tomato yellow leaf curl virus* (TYLCV).	(155)
Exogenous treatment with SA (1mM)	Increased RdRP expression in tobacco	(156)
Thermotherapy at 38 °C for 21 days followed by meristem tip culture (1mm)	68% of regenerated plants with 84% being CBSV-free	(136,157)
Combination of tissue culture, chemotherapy (ribavirin 25mg/L) and thermotherapy	In the first cycle, 27 of the 31 varieties (87%) were successfully cleaned of CMD and CBSD	(158)
Stem hot water therapy (55-60 °C) and two rounds of meristem tip culture	100% elimination of cassava mosaic begomoviruses (CMBs) and a significant reduction in the viral load of CBSV	(159)

1.5.6. Biotechnology approach for managing CBSD

i. Transgenic approaches

Genetic engineering has revolutionized the field of crop improvement. Unlike traditionally laborious conventional breeding, genetic engineering allows the insertion, deletion or modification of the target gene with no or limited alteration of the genome (160). For example, pathogen-derived resistance approach for managing plant viruses based on the expression of viral sequences engineered in the host plant can trigger resistance against viruses sharing similar sequences (161). The expression of such sequences triggers RNA silencing against viruses, which can block their replication and infection cycle (162–164). RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) was later fully implemented as a biotechnology approach for controlling plant viral diseases by expressing double-stranded RNA (dsRNA) or hairpin structured RNA (hpRNA) (160,165). The dsRNA is then cleaved by Dicer-like (DCL) proteins into small dsRNA of 20 to 24 nucleotides which are further unwounded, and single strands are loaded into argonaute proteins forming the RNA-induced silencing complex (RISC) where the ssRNA guide the RISC to its complementary target sequences leading to the degradation of the target mRNA (166–168). Indeed, producing dsRNA or hpRNA that contains exon or spacer sequences in transgenic plants will lead to the degradation of homologous viral RNA sequences. Whereas for DNA viruses, which replicate in the nucleus, dsRNA or hpRNA can also trigger RNA-directed DNA methylation of the promoter sequences preventing essential transcription factors from binding and thus inducing transcriptional gene silencing (165,169,170). It is very critical to select a suitable viral gene target for developing the hp-RNAi construct that would lead to adequate viral disease protection, and chimeric transgenes targeting either numerous viral proteins on the same genome or different virus strains have also been developed to achieve durable protection (171–173). This RNAi mechanism has been successfully used to engineer virus resistance in several important crops, including potatoes, beans, and tomatoes (160,165,174–177).

Studies have also demonstrated that RNAi can be applicable in controlling cassava viruses (178,179). Transgenic cassava lines expressing small interfering RNA against a near complete UCBSV coat protein gene were fully resistant to UCBSV (180,181). Vanderschuren *et al.* (2012) also proved that a hairpin construct targeting the CBSV CP sequence led to resistance against mixed CBSV and UCBSV infection (182). Following this successful record, in 2021, Kenya became the first country in East Africa to officially approve transgenic cassava resistant against CBSIs using RNAi technology (183). Unfortunately, like for other crops, deploying genetically modified (GM) CBSD-resistant varieties faces various challenges, including poor communication and lack of appropriate GM crop regulations, particularly in African countries. Furthermore, the misunderstanding of GM crops by the public, accompanied by anti-GM crops lobbyists who disseminate rumours about the negative impact of agro-biotech on biodiversity, indigenous crop, human health, and the environment, result in public distrust of agro-biotech products (177,184,185).

The potato story is a typical example where the successful development of a transgenic virus-resistant potato with high-quality yield was eventually banned from the market following an anti-biotechnology campaign (186).

ii. Gene Editing technologies and CRISPR Cas 9

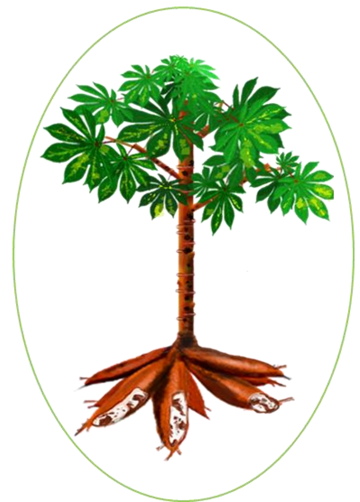
Advances in gene editing technologies have opened up new ways of achieving plant protection against various biotic stress. New tools have been developed based on a system developed around prokaryotic clustered regularly interspaced short palindromic repeats associated proteins (CRISPR / Cas). The system comprises a Cas protein with nuclease activity and a single guider RNA. Since its discovery, CRISPR/Cas has been shown to provide superior performance compared with other genome editing technologies due to its high rate of success, lower cost, and simplicity to design and use (187). The application of genome editing approaches requires knowledge of the target gene and the modifications to confer a desired trait. The CRISPR/Cas9 technology has been used in plant protection against many other plant viruses, such as *Tomato yellow leaf curl virus*, *Bean yellow dwarf virus*, *Cucumber vein yellowing virus*, *Zucchini yellow mosaic virus*, *Papaya ring spot mosaic virus*, and *Rice tungro spherical virus* (187–189).

In cassava, knockout of host susceptibility genes such as the cap-binding complex (eukaryotic translation initiation factors (eIF4F)) involved in the initiation of mRNA translation was identified as a potential target that can confer resistance against potyviruses. Because viruses depend on the host translation machinery, blocking viral RNA translation without affecting host translation could result in virus resistance. This aspect has been exploited in cassava, which has 5 genes coding for eIF4F. The resulting cassava with mutated novel cap-binding proteins 1 and 2 (NCBP1&2) reduced the incidence and severity of CBSD root and aerial symptoms (190).

Some drawbacks of this technology include the potential for non-specific Cas nuclease cleavage, the possibility of introducing off-target mutations, and the loss of resistance as target viruses evolve to overcome the single gene-based edited resistance (191). Furthermore, classifying edited plants as transgenic and subjecting them to GMO regulations slow their acceptance in different countries (192).

Chapter 2

Problem Statement, Objectives and Thesis structure



Chapter 2: Problem Statement, Objectives, and Thesis Structure

General Introduction to Chapter 2

This chapter is the keystone of the entire thesis, encompassing crucial elements vital for the research's success and comprehension. It starts by meticulously crafting a clear problem statement, highlighting the gaps in the CBSV research, the relevance and pressing need of the present studies. It then constructs pertinent research questions and the thesis's objectives, guiding the research towards defined milestones. By incorporating these elements, this chapter acts as a blueprint, providing clarity to the thesis.

Chapter 2: Problem Statement, Objectives, and Thesis Structure.

2.1. Problem Statement and Justification

The production of cassava, which is essential for food security in Sub-Saharan Africa (SSA) (8), remains constrained by CBSD, although different measures for its management have so far been implemented (48,157). CBSD arises from a complex interplay of triangular interactions involving the sources of CBSIs, transmission modes, and the presence of susceptible or reservoir hosts. Studies have been conducted in Rwanda to determine the prevalence of CBSD and identify its causative agents and reservoir hosts. However, the risk factors contributing to the ongoing spread of the disease in the country have yet to be thoroughly assessed. With the fact that the disease spreads mainly through the exchange of infected planting materials between farmers as they operate through vegetative propagation (79), it is imperative to ensure the availability of a sustainable disease-free cassava seed system and to find factors contributing to CBSD dissemination in the country. Moreover, it is essential to promote the adoption of good disease management practices by cassava seed multipliers and cassava farmers.

To effectively manage CBSD, it is crucial to comprehend the genetic diversity and evolution of the causative agents, enabling the development of accurate detection methods and continuous monitoring of virus changes. Advanced high-throughput sequencing technologies offer the means for in-depth analysis, providing millions of accurate short sequences from the viruses. Such thorough analysis beyond the consensus level is essential to revealing a precise picture of the virus evolution process, as even low-frequency point mutations can significantly impact the virus's biological characteristics (193–195). Nevertheless, in many genetic diversity studies of CBSIs infecting cassava in various African countries, the focus has primarily been on the analysis of consensus sequences of the virus population, disregarding potentially significant low-variant virus populations in the dataset (66,71,196–198).

In Rwanda, the viruses associated with CBSD were determined for the first time in 2014 based on partial coat protein sequence analysis, which revealed one type of CBSV and indications of diverse UCBSV. However, further studies based on complete genome sequencing were required for an extensive understanding (93). Analyzing the complete CBSI genome sequences from Rwanda beyond the consensus level would provide a comprehensive understanding of CBSIs' genetic diversity and evolution across the country, leading to appropriate approaches for sustainable disease management.

Currently, the management of CBSD relies on disseminating CBSI-tolerant cassava. However, tolerant varieties are not immune to CBSD; thus, their usefulness could be limited due to virus accumulation with every cycle of propagation (6). Therefore, establishing a clean cassava system with robust virus-cleaning techniques and distributing clean cassava seeds to farmers is one of the best ways of combating the yield loss associated with the viral build-up effect over time.

Efforts were made to clean the cassava virus using *in vitro* techniques like meristem tip culture, thermotherapy, and chemotherapy (136,137). Nevertheless, these techniques are complex, time-consuming, and cannot be applied in a greenhouse or field environment. A field or greenhouse virus cleaning approach applicable to farmers could be of paramount importance, as it could be instrumental in increasing the accessibility of clean cassava planting materials and minimizing the impact of CBSD on yield in Rwanda.

2.2. Research Questions and Thesis Objectives

2.2.1. Research Questions

The research questions of this thesis were:

1. What is the prevalence and distribution of CBSD in Rwanda, and what are the key factors influencing its spread and transmission within the country?

2. What is the genetic diversity of the CBSIs infecting cassava in Rwanda?

3. Can salicylic acid and benzothiadiazole boost the natural resistance mechanisms against CBSIs in cassava?

To address these research questions, the following thesis objectives were defined:

2.2.2. Thesis Objectives

1. To assess farmers' practices and knowledge of the biotic constraints and determine the status of CBSD as well as the critical factors associated with its spread across the country.

2. To determine the geographical distribution and genetic diversity of the cassava brown streak ipomovirus population infecting cassava in Rwanda

3. To study the effects of greenhouse thermotherapy and chemotherapy, along with field chemotherapy, on cassava viruses and to assess the impact of exogenous Salicylic acid (SA) and benzothiadiazole (BTH) application on the cassava transcriptome.

2.3. Thesis Structure

This thesis is organized into six (6) chapters. Below is the schematic illustration of the thesis structure (**Figure 10**).

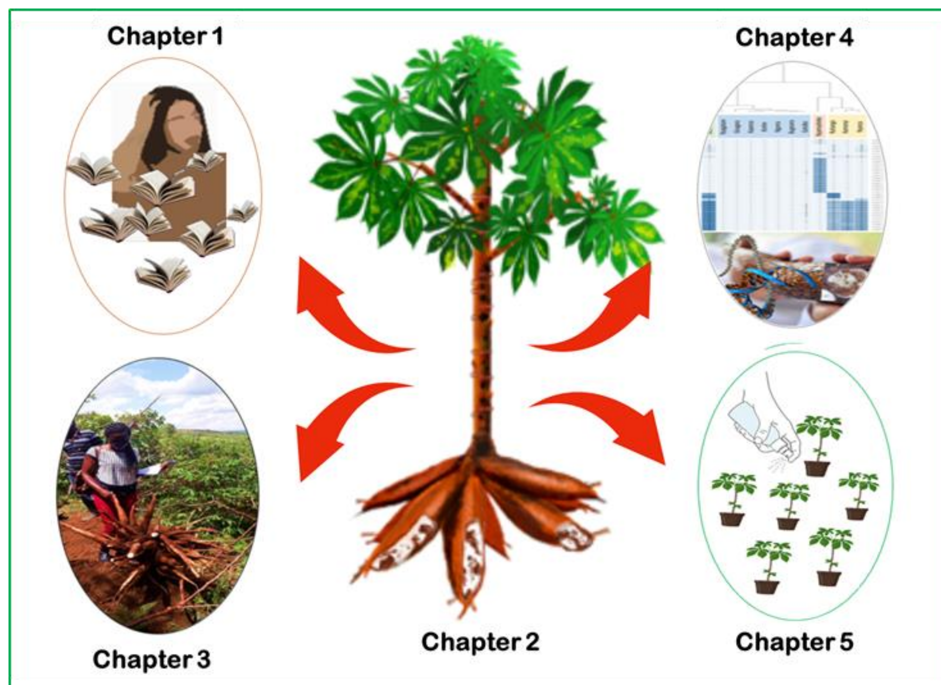


Figure 10. Schematic illustration outlining the structure of the thesis.

Chapter 1 provides a general introduction to cassava and cassava brown streak disease. It consists of a review of the CBSD and its management approaches.

Chapter 2 includes a problem statement and justification of the study, research questions, thesis objectives, and structure.

Chapter 3 addresses the first research question by providing the current status of the CBSD and critical factors associated with its spread in Rwanda.

Chapter 4 answers the second research question by analyzing the genetic diversity of CBSD causative agents and their geographical distribution in Rwanda.

Chapter 5 assesses the effect of greenhouse thermotherapy, chemotherapy, and field chemotherapy on cassava viruses as well as the impact of salicylic acid and benzothiadiazole on cassava transcriptome.

Chapter 6 presents a general discussion of the findings obtained, the general conclusion, and perspective for enhancing the management of cassava brown streak disease.

The list of scientific articles published, as well as scientific communications carried out during this PhD research, are in the appendices.

Chapter 3

Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated with Cassava Brown Streak Disease Incidence.

Chapter 3. Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated with Cassava Brown Streak Disease Incidence.

General Introduction to Chapter 3

In the opening chapters, the focus was directed towards understanding the substantial impact posed by cassava viral diseases, with a particular emphasis on cassava brown streak disease (CBSD). Despite considerable efforts to mitigate its impact, a crucial gap remained in understanding the factors influencing the spread of CBSD within the country and assessing its current status following various intervention measures. Indeed, effectively managing CBSD necessitates ongoing surveillance and active collaboration among all stakeholders within the cassava seed system. Thus, this chapter embarked on a comprehensive nationwide survey to determine the present CBSD situation, identify seed pathways, and examine farmers' practices that may contribute to disease transmission in Rwanda.

Chapter 3. Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated With Cassava Brown Streak Disease Incidence.

This chapter 3 is based on the original research article published in Frontiers in Sustainable food systems Volume 5; December, 3rd 2021 https://doi.org/10.3389/fsufs.2021.699655. Just text formatting was used to adapt the published version.

Chantal NYIRAKANANI [1,2] *, Jean Pierre BIZIMANA[1,3], Yves KWIBUKA[4,5], Athanase NDUWUMUREMYI[3], Vincent de Paul BIGIRIMANA[2], Charles BUCAGU[3], Ludivine LASSOIS[1], Eléonore MALICE[6], Nicolas GENGLER[7], Sébastien MASSART[4], Claude BRAGARD[8], Michael HABTU[9], Yves BROSTAUX [10], Cécile THONAR [1,11] and Hervé VANDERSCHUREN[1,12] *

[1]Plant Genetics and Rhizosphere Processes Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium.

[2]Department of Crop Science, School of Agriculture, University of Rwanda, Musanze, Rwanda.

[3]Department of Research, Rwanda Agriculture and Animal Resources Development Board, Huye, Rwanda.

[4]Integrated and Urban Plant Pathology Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium.

[5]Departement des Sciences Agronomiques, Université Catholique de Bukavu, Bukavu, République Démocratique du Congo.

[6]Unité Santé des Plantes et Forêts, Centre wallon de Recherches agronomiques, Gembloux, Belgium.

[7]Numerical Genetics, Genomics and Modeling Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium.

[8]Phytopathology Laboratory, University of Louvain, Louvain-la-Neuve, Belgium.

[9]Department of Public Health, School of Health Sciences, Mount Kenya University, Kigali, Rwanda.

[10]Modelisation and Development, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium.

[11]Agroecology Lab, Université Libre de Bruxelles (ULB), Brussels, Belgium.

[12]Tropical Crop Improvement Laboratory, Department of Biosystems, Katholieke Universiteit Leuven, Heverlee, Belgium.

3.1. Abstract

Cassava (*Manihot esculenta* Crantz) is a vital crop in Rwanda, where it ranks as the 3rd most consumed staple food. However, cassava productivity remains below its yield potential due to several constraints, including important viral diseases such as cassava brown streak disease (CBSD). The contribution of the present study is embedded in assessing current cassava seed system, farmers' practices and their knowledge of the biotic constraints to cassava production, determining countrywide CBSD status and critical factors associated with its spread in the country. A cross-sectional study was carried out from May to September 2019 in 13 districts of Rwanda. One hundred thirty farmers and cassava fields were visited, and the incidence and severity of CBSD were evaluated. CBSD was detected in all cassava-producing districts. The highest field incidence of CBSD was recorded in Nyanza district (62%; 95% CI = 56% - 67%), followed by Bugesera district (60%; 95% CI = 54% - 65%) which recorded the highest severity score of 3.0 ± 0,6. RT-PCR revealed the presence of CBSD at the rate of 35.3%. *Ugandan cassava brown streak virus* was predominant (21.5%), while *cassava brown streak virus* was 4% and mixed infection was 10%. The informal cassava seed system was dominant among individual farmers, whereas most cooperatives used quality seeds.

Disease management measures were practised by only half of the participants. Factors significantly associated with CBSD infection ($p<0.05$) were the source of cuttings, proximity to borders, age of cassava, and knowledge of CBSD transmission and management.

Key words: Cassava, Seed system, CBSD, Field survey, Rwanda.

3.2. Introduction

Cassava (*Manihot esculenta* Cranzt) ranks as the 6th most important food crop worldwide and the 4th after rice, maize, and wheat among developing and emerging countries (199,200). In Rwanda, cassava is the 3rd most important crop after banana and sweet potato (15). Because of its importance in several tropical regions and its relatively good performance on marginal lands under suboptimal climatic conditions (8), cassava is recognized as a critical crop to overcome food insecurity for the fast-growing population in areas prone to significant climatic changes (12,201,202).

The yield potential of cassava under optimum conditions is about 90 tons of fresh roots per hectare, equivalent to 30 tons of cassava dry matter per hectare (12). More than half (61%) of cassava production occurs in Africa. However, cassava yield in tropical countries is still far below its production potential. Indeed, in 2017, the world cassava yield was about 11.08 tons of fresh roots per hectare. The top cassava producer (Nigeria) had an average yield of 8.75 tons per hectare, followed by the Democratic Republic of Congo (DRC) with 8.14 tons per hectare (200,203). Cassava production in Rwanda varied between 3000 Mt – 3701 Mt of fresh roots per year from 2015 to 2018, with a reported average yield of about 14.5 tons per hectare (88)(RAB, 2020: unpublished data). Despite its resilience under adverse environmental conditions, cassava production remains constrained by several abiotic and biotic factors. The former includes post-harvest deterioration, infertile soils, planting unimproved traditional varieties, and inadequate farming practices, whereas the latter includes green mites, mealy bugs, cassava bacterial blight, and viral diseases (34,204).

Due to viral diseases and the lack of resistant varieties, cassava yields have drastically decreased in many countries. Cassava brown streak disease (CBSD) and Cassava mosaic disease (CMD) are the most economically essential cassava diseases, causing yield losses of over US$1 billion a year globally (36,48,61,205).

CBSD has so far only been reported in Africa. CBSD is devastating because it negatively impacts cassava tuberous roots quantitatively and qualitatively, causing substantial economic losses to African farmers (196). For decades, CMD has been managed through the dissemination of resistant varieties, but unfortunately, the distributed CMD-resistant varieties were found to be sensitive to CBSD in Rwanda and many other African countries (206–209). CBSD is caused by two species of single-stranded RNA (ssRNA) viruses of the family *Potyviridae,* Genus *Ipomovirus*: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) (77). In Rwanda, CBSD was first reported in 2009 in the Southern province (Muhanga, Bugesera, and Nyagatare districts). CBSD has since spread to reach most cassava-producing regions in the South (Kamonyi, Ruhango, Nyanza, and Gisagara districts) and East of the country (Kirehe and Gatsibo district) (93,208). A study conducted in Rwanda in 2014 reported a distribution of CBSD incidence: 74.2% UCBSV infection, 15.3% CBSV infection, and 10.5% mixed infection (93).

The CBSIs are transmitted by either the whitefly *Bemisia tabaci* and the exchange of infected planting materials between farmers. Plant pathologists and extension services have recognized the importance of establishing a disease-free seed system to mitigate the spread of CBSD (48,210).

In Rwanda, the formal distribution of clean planting seeds usually involves the whole production chain from RAB, where researchers produce basic clean seeds. Basics seeds are then distributed to seed multipliers across different regions for further multiplication, and before their dissemination to farmers, a quality seed certification agency is involved to ensure the quality of the planting materials (132,133).Quality seed refers to the seed preferred by farmers and consumers with good health (virus-free), genetic purity, appropriate physiological age and physical quality (133).

The presence of an informal seed system involves farmers producing and sharing seeds without following standardized quality seed certification protocols. This informal practice can facilitate the rapid dissemination and persistence of diseases within cassava agrosystems (211). Due to its widespread adoption in many cassava agrosystems globally, the informal seed system plays a crucial role in the accelerated spread of diseases (46,133).

Various management measures have been applied to reduce CBSD impact, including breeding of CBSD tolerant varieties (44,212–214), and the dissemination of disease-free planting material to farmers (79). Conversely, farmers who reuse cuttings from their fields will not escape the disease as this tends to maintain 30 to 50 % of infection, especially in CBSD hotspots (48,215). Unfortunately, the sustainability of the seed system remains fragile and needs to be sthrenghened in order to provide healthy planting cassava material to all cassava farmers. Despite the emergence of CBSD in cassava fields in Rwanda, there has been limited information about its dynamics of spreads and factors associated with its dissemination in the country.

The present study aimed to assess the current cassava seed system, farmers' practices, and their knowledge of the biotic constraints to cassava production. It also aimed to determine the status of CBSD and the critical factors associated with its spread through the seed system channels.

3.3. Materials and Methods

3.3.1. Study area

The study was conducted in 13 cassava-growing districts of Rwanda in 2019. Districts in the Southern and Eastern provinces are considered major cassava-producing areas. In the South, 5 districts were surveyed, namely Gisagara, Nyanza, Ruhango, Muhanga and Kamonyi, whereas, in the East, 6 districts were surveyed, namely Bugesera, Nyagatare, Kayonza, Gatsibo, Kirehe and Ngoma. In addition, two districts from Western and Northern provinces, Nyamasheke and Gakenke, respectively, were included in the study (**Figure 11**).

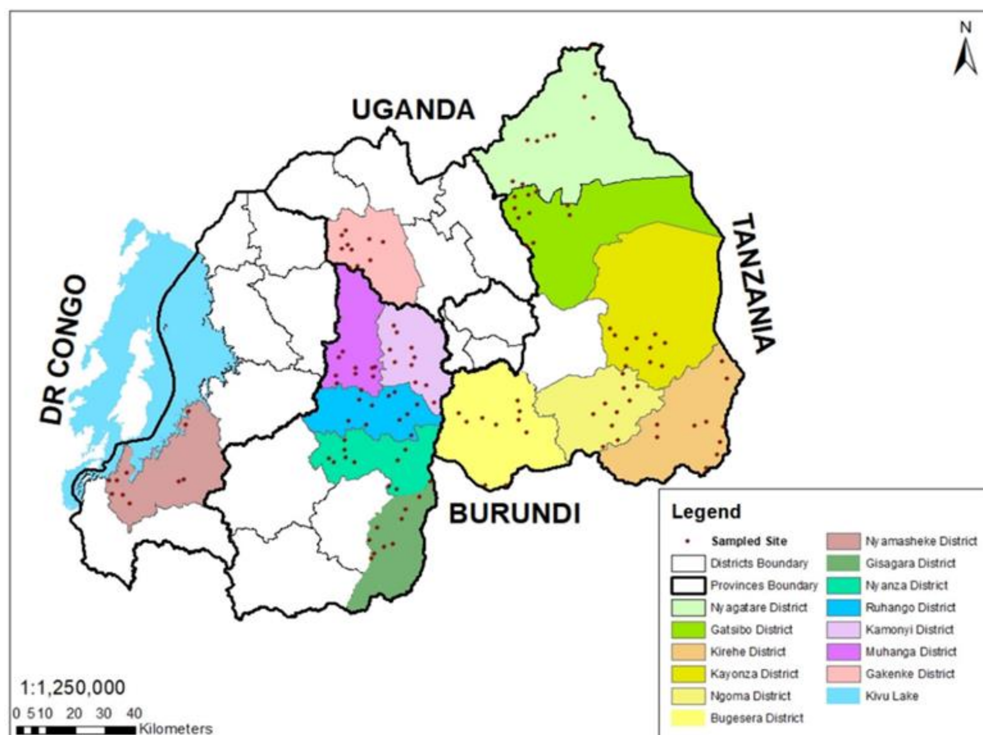


Figure 11. Map of Rwanda showing surveyed districts.
The red dots represent the location of the cassava fields assessed in the study.

3.3.2. Farmers and fields selection

A multistage sampling method was applied to select cassava farmers and fields. In the first stage, 13 districts representing major and minor cassava growing areas were selected. In the second stage, five sectors were purposively selected within each district based on their relative importance in cassava production according to information provided by district agronomists and RAB. In the third stage, 2 farmers per sector (one individual farmer and one farmer belonging to a cooperative) were selected from a sampling frame provided by sector agronomists using a simple random sampling, making a total of 10 farmers per district and 130 interviewed farmers for the 13 districts surveyed.

Furthermore, a field with cassava plants older than 6 months was also visited for disease evaluation for each participant. Within the selected fields, 30 plants were selected for leaf and stem CBSD symptoms examination, including 5 plants of the two diagonals and 5 of the 4 sides (Rwegasira *et al.,* 2011). The field incidence per district was recorded as the percentage of symptomatic plants out of the total examined. The ten plants examined at the 2 diagonals were further pooled and used for CBSIs indexing by RT-PCR.

3.3.3. Farmers' interview

Primary data used in the study were collected through a structured questionnaire, semi-structured interviews and observations on five key subject areas (socio-demography, agronomy, seed accessibility and availability factors and disease aspect) relevant to cassava disease spread. District and sector agriculture extension officers liaised with local community leaders and were involved in the mobilization of farmers. Permission to conduct research in the area was sought from the administration of the study area (district and sector agronomists) through official communication by RAB authorities. Participants were told the purpose of the research, and that participation was voluntary. Oral consent was given before starting the interview and field visit. All records were identified by study identification number to keep participant privacy and confidentiality.

3.3.4. Disease severity assessment

A 1-5 CBSD symptom scale was used to measure the degree of severity of CBSD aerial symptoms in the fields. The scale used was 1 = no apparent symptoms; 2 = slight leaf feathery chlorosis with no stem lesions; 3 = pronounced leaf feathery chlorosis, mild stem lesions and no dieback; 4 = severe leaf feathery chlorosis, severe stem lesions and no dieback; and 5 = defoliation, severe stem lesions and dieback (99). The average degree of severity was calculated by omitting the score of 1, representing asymptomatic plants, to provide an accurate picture of the severity in the fields assessed (216). An average disease severity per district was calculated based on observing 30 x 10 = 300 plants.

3.3.5. Sampling test materials for RT-PCR

Samples were collected from May to September 2019. In each cassava field assessed, 10 cuttings from 10 plants examined along the 2 diagonals were collected per field and established in the greenhouse. In total, 1300 plants were grown in the greenhouse from 130 fields. Five-month-old plants from collected cuttings were used for molecular analysis. In the laboratory, 20 leaf samples from the growth of cuttings collected on the same field were pooled (lower and middle leaves were used per plant). Thus, a total of 130 samples were tested for CBSIs using RT-PCR.

3.3.6. Molecular analysis

RNA Extraction, cDNA synthesis and RT-PCR

Total RNA was extracted from ~0.2 g cassava leaf using the cetyltrimethylammonium bromide protocol previously described (Abarshi et al., 2010). CDNA was synthesized using the ProtoScript II Reverse Transcriptase kit (BioLabs, UK) following the manufacturer's instructions. Briefly, a Master Mix containing 1 µL d(T)23 (50mM), 2 µL of buffer, 1 µL of 0.1mM DTT, 0.5 µL Protoscript II RT, 0.5 dNTP Mix, 3 µL of nuclease-free water was prepared. A 2 µL RNA template was added, making 10 µL per reaction. The reaction mixture was incubated in PCR thermocycler at 42°C for 1h for primer annealing and cDNA synthesis, followed by 20 min at 65°C for the ProtoScript II Reverse Transcriptase inactivation.

The synthesized cDNA was subjected to a polymerase chain reaction using a Taq G2 Hot Start Master Mix from Promega. The primer pair F:5'-CCTCCATCWCATGCTATAGACA-3' and R:5'-GGATATGGAGAAAGRKCTCC-3' that amplifies ~703bp of CBSV and ~800 bp of UCBSV isolates was used (Elegba, 2018). The 10 µL PCR reaction contained 5 µL, G2 Mix, 0.4 µL each primer (0.4 µM final concentration), and 1µL cDNA and the volume was brought to 10 µL with nuclease-free water. PCR conditions were as follows: Predenaturation at 95°C for two minutes was followed by 30 cycles of denaturation at 94°C for thirty seconds, annealing at 56°C for thirty seconds, elongation at 72°C for fifty seconds, and final elongation at 72°C for five minutes. An internal control gene from cassava called *Manihot esculenta* Protein Phosphatase 2 A (*MePP2A*) was detected in parallel using a pair of primers F: 5′-TGCAAGGCTCACACTTTCATC-3' and R: 5′-CTGAGCGTAAAGCAGGGAAG-3′ that amplifies 150bp of *Me*PP2A to ensure the accuracy of the PCR results by ruling out any false negative results (Moreno, Gruissem and Vanderschuren, 2011).

PCR amplification was checked by loading 10 µL of PCR products in 1% (w/v) agarose gel stained with Gel red in 1X Tris-acetate-EDTA (TAE) buffer for 1 hour at 200 V to allow the separation of amplicons from the two isolates. A UV gel documentation system was used for PCR product visualization and photography.

3.3.7. Data analysis

Raw data were transferred into Statistical Package for Social Sciences (SPSS) version 22 for analysis. Frequency and proportions for categorical variables were computed to describe the primary attributes of the respondents (farmers) and the occurrence of cassava infection (defined as the presence of CBSV or UCBSV after PCR) in sampled fields.

Bivariate analysis with chi-square tests was used to determine factors associated with cassava virus infections (categorical variable). Then multivariate logistic regression analyses were performed by considering all significant factors during bivariate analysis. The Hosmer-Lemeshow test was used to evaluate the goodness of fit. In all statistical tests, differences were considered statistically significant at $p < 0.05$.

3.4. Results

3.4.1. Source of cassava cuttings, trust and disease management

Among the 130 farmers interviewed, the majority (62.1%) reported that they obtained planting material from their fields and used the same materials over many seasons. A minority of farmers (25.1%) acquired planting materials from seed multipliers every season. Among those, a large proportion (86%) were cooperative members (**Figure 12A, B**). Although most farmers use seeds from their fields, about 45% questioned their quality and feared that their cassava fields might succumb to diseases.

A gap in cassava viral disease management was noted throughout the study as 50.3% of interviewed farmers took no action to control viral disease, and only 19.5% declared roguing out infected plants from their field (**Figure 12C, D**).

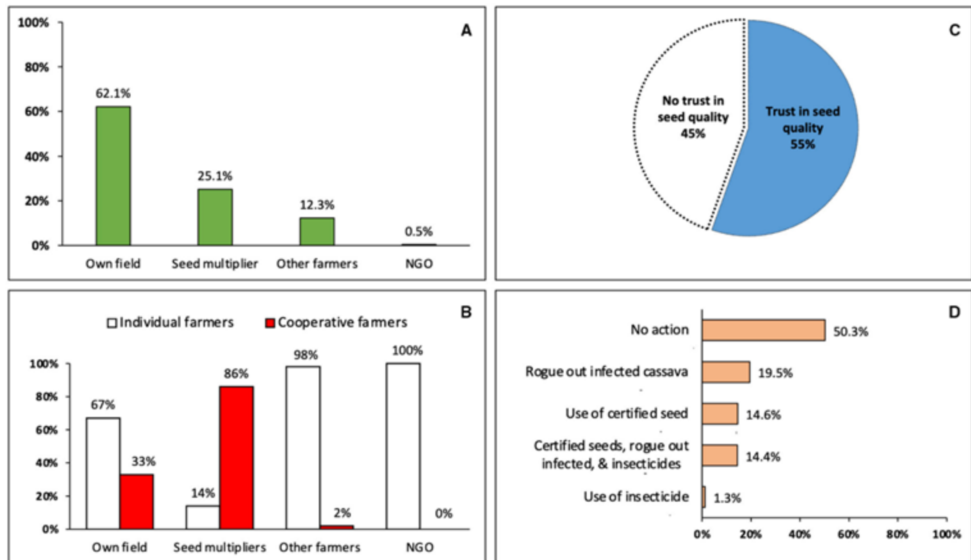


Figure 12. Sources of cassava planting materials, trust, and disease management among farmers in Rwanda during 2019.

(**A**) Sources of planting materials, (**B**) categories of farmers vs. sources of seeds, (**C**) level of trust in seed quality among participants, (**D**) disease management methods applied by participants (NGO, non-government organization).

3.4.2. CBSD incidence and severity score of observed symptoms.

Field incidence of CBSD and severity of aerial symptoms were evaluated in the 13 districts. The highest incidence (62%; 95%CI = 56% – 67%) was recorded in Nyanza district, while the lowest (12%; 95%CI = 8% - 16%) was observed in Gakenke district. Disease mean severity scores varied from district to district, ranging from 3.0 ± 0,6 in Bugesera to 2 ± 0,2 recorded in Nyamasheke (**Table 2**).

Table 2. Field incidence, severity score and frequency of cassava plants showing aerial CBSD symptoms in Rwanda, 2019.

Districts	Field incidence % (95% CI)	Mean severity score	Frequency of severity score (%)				
			1	2	3	4	5
South							
Gisagara	58.3 (52 - 63)	2.7 ± 0.8	42	29	19	8	2
Nyanza	62 (56 - 67)	2.9 ± 0.8	38	18	33	7	4
Ruhango	35.6 (30 - 41)	2.1 ± 0.3	61	35	4	0	0
Muhanga	30 (24 - 35)	2.2 ± 0.4	70	24	6	0	0
Kamonyi	25.3 (20 - 30)	2.2 ± 0.6	72	23	2	3	0
East							
Bugesera	60 (54 - 65)	3 ± 0.6	37	11	44	5	3
Kayonza	32.3 (27 - 37)	2.1 ± 0.3	68	28	4	0	0
Gatsibo	18 (13 - 22)	2.4 ± 0.5	82	10	8	0	0
Nyagatare	30 (24 - 35)	2.5 ± 0.6	70	16	12	2	0
Ngoma	20.6 (16 - 25)	2.2 ± 0.5	80	16	3	1	0
Kirehe	50 (44 - 55)	2.6 ± 0.5	50	22	26	2	0
North							
Gakenke	12 (8 - 16)	2.1 ± 0.3	87	11	2	0	0
West							
Nyamasheke	14.3 (10 - 18)	2.0 ± 0.2	86	13	1	0	0

Three hundred plants per district were examined. CI: Confidence interval

Molecular diagnostics were performed on 130 samples collected from 13 districts. RT-PCR analyzed samples for the detection of CBSV and UCBSV. The overall incidence of CBSIs was 35.3%. Among the positive samples, 61% (28/46) were UCBSV, whereas 11% (5/46) were CBSV, and 28% (13/46) had mixed infection of both CBSV and UCBSV (**Figure 13**) **Figure S1.** shows RT-PCR detection of CBSIs in field samples.

All the 13 districts surveyed were found to be affected by CBSIs based on RT-PCR results, and the highest incidence (60%) was recorded in the South, Nyanza district, followed by Gisagara district and Bugesera districts, both displaying an incidence of 50%. A single infection of UCBSV was found in all districts except Muhanga, whereas a single infection of CBSV occurred in Muhanga, Bugesera, Gatsibo and Gakenke. Mixed infections were recorded in most districts except Ngoma, Gatsibo, Gakenke and Nyamasheke (**Figure 13**).

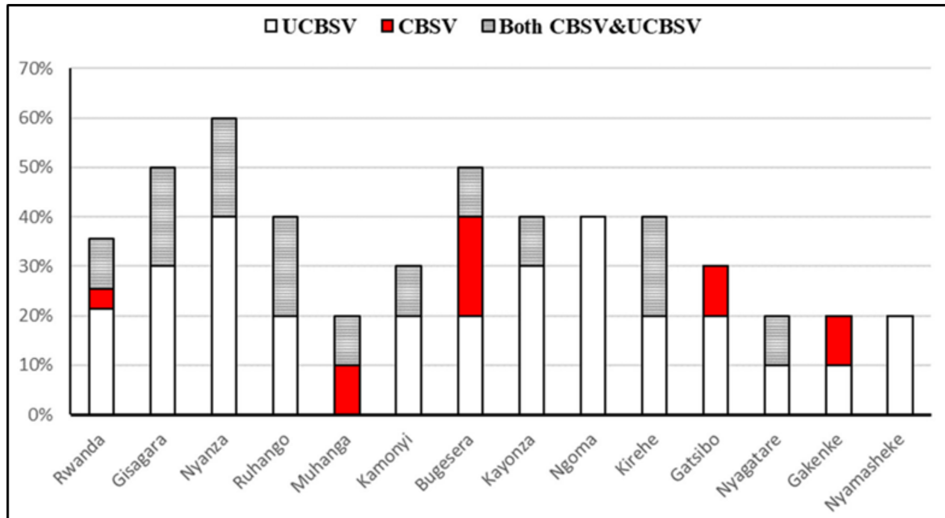


Figure 13. Incidence of CBSD based on RT-PCR in different districts of Rwanda, 2019. The highest incidence was recorded from Nyanza, followed by Gisagara and Bugesera districts. UCBSV: *Ugandan cassava brown streak virus*; CBSV: *Cassava brown streak virus*.

3.4.3. Bivariate analysis of factors associated with CBSD incidence

Fields from farmers working in cooperatives display lower CBSD infection rate

Utilizing the data gathered from the survey, an analysis was conducted to determine whether socio-demographic characteristics influence the level of cassava infection. Bivariate analysis (using the Chi-square test) of socio-demographic factors of 130 farmers (for whom cassava fields were visited) stratified by CBSV infection revealed that there is a significant association between the category of respondents and cassava infection, where individual farmers had more infected fields than farmers in cooperatives (p-value=0.023). Farmers' age also significantly influenced the level of cassava infection in their field (p-value =0.043). All the other socio-demographic factors did not influence levels of cassava infection (**Table 3**).

Table 3. Socio-demographic factors stratified by cassava infection.

Variables	Total, n(%)	CBSIs Positive, n(%)	Chi square value	df	p-value
Category of respondents					
Individual Farmers	76 (58.5)	32 (69.6)	5.168	1	**0.023**
Cooperative	54 (41.5)	14 (30.4)			
Province					
South	50 (38.5)	20 (43.4)	2.579	3	0.461
East	60 (46.2)	22 (48.0)			
West	10 (7.7)	2 (4.3)			
North	10 (7.7)	2 (4.3)			
Farmers' age in years					
25-35	42 (32.3)	15 (32.6)	3.946	2	**0.043**
35-45	41 (31.5)	10 (21.7)			
45 and above	47 (36.2)	21 (45.7)			
Gender					
Male	84 (64.6)	29 (63.0)	0.077	1	0.781
Female	46 (35.4)	17 (37.0)			
Marital status					
Married	87 (66.9)	32 (69.6)	0.534	2	0.766
Single	25 (19.2)	9 (19.6)			
Widow/er	17 (13.5)	5 (10.9)			
Education level					
Illiterate	23 (17.7)	8 (17.4)	1.006	2	0.605
Primary	94 (72.3)	35 (76.1)			
Secondary	13 (10.0)	3 (6.5)			

[*] Significant at $p<0.05$ bolded; df: degree of freedom

Fields established with planting material from seed multipliers have a lower probability to be CBSD infected

Furthermore, the link between cassava seed accessibility and CBSV infection was investigated. Our analysis showed that the source of cassava cuttings and proximity to the border significantly impacted cassava infection with p-values of 0.001 and 0.021, respectively (**Table 4**). Farmers who used seeds from their fields were more likely to have infected fields (60.9%) than those who got seeds from seed multipliers (10.9%). It was also noted that farmers near the country's border had more infected fields (54.3%) (**Table 4**).

Table 4. Impact of accessibility of planting materials on cassava infection

Variables	Total, n(%)	CBSIs Positive, n(%)	Chi square value	df	p-value
Source of cassava cuttings					
Seed multiplier	30(23.0)	5(10.9)	14.733	2	**0.001**
Own field	79(60.7)	28(60.9)			
Other farmers	21(16.2)	13(28.2)			
Distance to the source (Km)					
<1	70(53.8)	26(56.5)	4.032	3	0.258
1 to 4	28(21.5)	7(15.2)			
4 to 8	24(18.5)	8(17.4)			
>8	8(6.2)	5(10.9)			
Proximity to the tarmac road (Km)					
<1	14(10.8)	4(8.7)	1.732	3	0.630
1 to 4	24(18.5)	11(23.9)			
4 to 8	27(20.8)	10(21.7)			
>8	65(50.0)	21(45.7)			
Proximity to RAB (Km)					
<10	42(32.3)	23(50.0)	13.445	3	0.070
10 to 20	21(16.2)	3(6.5)			
20 to 30	21(16.2)	4(8.7)			
>30	46(35.4)	16(34.8)			
Proximity to the border (Km)					
<10	53(40.8)	25(54.3)	9.7	3	**0.021**
10 to 20	23(17.7)	10(21.7)			
21 to 50	10(7.7)	3(6.5)			
51 and above	45(34.6)	8(17.4)			

* Significant at p<0.05 bolded; df: degree of freedom

Most cassava fields surveyed were above 8 months old (56.2%). Most participants grew improved varieties (65.4%) and had access to extension services (83.1%). The analysis revealed that there was a significant association between the age of the plant and cassava infection (p-value <0.001), where the plants aged less than 8 months (57.1%) were significantly more likely to be healthy than to be infected (19.5%). Furthermore, a significant association was noted between the use of fertilizers and cassava infection (p-value =0.002), where the farmers using fertilizers (68%) were significantly more likely to have healthy than infected fields (39%) (**Table 5**)

Table 5. Impact of agronomic variables on cassava infection.

Variables	Total, n(%)	CBSIs Positive, n(%)	Chi square value	df	p-value
Age of plants in months					
<8	57(43.8)	9(19.5)	19.65	2	**<0.001**
8 to 10	56(43.1)	24(52.2)			
>10	17(13.1)	13(28.2)			
Type of cassava varieties grown					
Improved	85(65.4)	29(63.0)	7.2	2	0.067
Local	9(6.9)	0(0.0)			
Both improved & Local	36(27.7)	17(37.0)			
Farming system used					
Monoculture	60(46.2)	16(34.8)	3.7	1	0.054
Polyculture	70(53.8)	30(65.2)			
Using fertilizers to grow cassava					
Yes	75(57.7)	18(39.0)	10.05	1	**0.002**
No	55(42.3)	28(61.0)			
Access to extension services					
Yes	108(83.1)	37(80.4)	0.35	1	0.552
No	22(16.9)	9(19.6)			
Extension services benefited by farmers					
Visit of cassava field	42(32.3)	17(37.0)	2.06	4	0.725
Advice on diseases management	7(5.4)	3(6.5)			
Advice on farming practices	31(23.5)	9(19.5)			
None	22(16.9)	9(19.5)			
Field visit and advices on GAP	28(21.5)	8(17.4)			

* Significant at p<0.05 bolded; df: degree of freedom

Farmers' awareness of cassava viral diseases is associated with lower CBSD incidence

Although all farmers were aware of at least one cassava viral disease's existence, 34.6% of them did not know the symptoms of cassava viral diseases. Symptoms of CMD were quickly recognized by 31.5%, followed by 20.8% who recognized both CMD and CBSD. Farmers who were not aware that the viruses could be transmitted had more infected fields (65.2%) (p-value < 0.001), and likewise, those who did not know disease management had more infections in their fields (67.4%) compared to those who knew the management techniques (p-value < 0.001) (**Table 6**).

Table 6. Disease awareness variables stratified by cassava infection.

Variables	Total, n(%)	CBSIs Positive, n(%)	Chi square value	df	p-value*
Knowledge of cassava viral diseases					
CMD	69 (53.1)	26 (56.5)	0.79	2	0.673
CBSD	15 (11.5)	6 (13.1)			
Both CMD & CBSD	46 (35.4)	14 (30.4)			
Knowledge of cassava viral diseases transmission					
Yes	70 (53.8)	16 (34.8)	19.62	1	**<0.001**
No	60 (46.2)	30 (65.2)			
Knowledge of mode of transmission					
Infected cutting	53 (40.8)	13 (28.2)	5.24	2	0.073
Both whitefly & infected cutting	17 (13.0)	3 (6.5)			
None	60 (46.2)	30 (65.2)			
Type of known symptoms					
CMD Leaf symptoms	41 (31.5)	14 (30.4)	1.41	4	0.843
CBSD leaf symptoms	4 (3.1)	2 (4.3)			
CBSD roots symptom	13 (10.0)	3 (6.5)			
CMD & CBSD roots symptoms	27 (20.8)	10 (21.7)			
None	45 (34.6)	17 (37.0)			
Observed viral diseases and pests in the field					
CMD	23 (17.7)	11 (24.0)	5	5	0.417
CBSD	21 (16.2)	9 (19.6)			
Green mites	1 (0.8)	0 (0.0)			
White fly	12 (9.2)	2 (4.3)			
Both CBSD & CMD	7 (5.4)	3 (6.5)			
None	66 (50.8)	21 (45.6)			
Knowledge of cassava viral diseases management					
Yes	65 (50.0)	15 (31.3)	19.38	1	**<0.001**
No	65 (50.0)	31 (67.4)			

* Significant at $p<0.05$ bolded; df: degree of freedom

3.4.4. Multivariate analysis of risk factors associated with CBSD

Eight (8) factors that showed significant association ($p < 0.05$) during bivariate analysis (including the source of cassava cuttings, proximity to the border, age of the plants, use of fertilizers, category of respondents, age of farmers, knowledge of cassava viral diseases transmission and knowledge of cassava disease management) were considered together in a multivariable analysis to identify the variables associated with cassava infections. Upon fitting the factors using multiple logistic regression and specifying *the 'backward conditional'* method with removal at $p<0.05$, five factors remained in the final analysis, as shown in Table 6. After testing the goodness of fit using the Hosmer-Lemeshow test, the Chi-square value was 4.80 with a degree of freedom of 6, and the p-value was 0.570, which indicates that the fitted model was adequate.

Farmers who use cuttings from their own fields or other fields were at over seven-fold higher risk than those who use cuttings from seed multiplier ($p<0.05$). Respondents near the border had a 4-time higher risk of having CBSIs than those far away ($p<0.05$). Cassava plants under 8 months had less risk of infection than the older ones ($p<0.05$). Similarly, those who were not aware of the disease transmission and management had nearly a 3-time higher risk of having the infected plants ($p<0.05$) (**Table 7**).

Table 7. Risk factors significantly associated with CBSD.

Variables	AOR	95% CI		p-value*
		Lower	Upper	
Category of respondents				
Individual Farmers	0.43	0.12	1.52	0.191
Cooperative	Ref			
Farmers' age in years				
25-35	0.74	0.22	2.43	0.617
35-45	0.26	0.08	1.22	0.080
45 and above	Ref			
Source of cassava cuttings				
Seed multiplier	Ref			
Own field	7.31	1.52	35.06	**0.013**
Other farmers	10.1	1.73	58.81	**0.010**
Proximity to the border (Km)				
<10	4	1.33	12.05	**0.014**
10 to 20	4.24	1	17.97	**0.050**
21 to 50	1.7	0.27	10.65	0.571
51 and above	Ref			
Age of plants in months				
<8	Ref			
8 to 10	4.76	1.69	13.39	**0.003**
>10	18.47	3.93	86.78	**<0.001**
Using fertilizers to grow cassava				
Yes	Ref			
No	2.44	0.93	10.25	0.127
Knowledge of cassava viral diseases transmission				
Yes	Ref			
No	3.97	1.46	10.83	**0.007**
Knowledge of cassava viral diseases management				
Yes	Ref			
No	2.94	1.08	7.96	**0.034**

Ref= reference, * Significant at p<0.05 bolded; AOR= Adjusted Odds ratio; φ 95% CI: Confidence Interval

3.5. Discussion

The present study used a comprehensive cross-country survey to assess the current cassava seed system, farmers' practices and their knowledge of the biotic constraints, the status of CBSD, and critical factors associated with its spread throughout the cassava seed system in Rwanda.

The current findings confirmed the occurrence of CBSD (both CBSV and UCBSV) in Rwanda. The disease was found in all thirteen districts surveyed, indicating that it has spread out in all significant cassava-growing regions, including Kirehe and Nyagatare, where CBSIs were not detected in previous studies (93). Our survey found that districts near the border displayed a higher rate of CBSD incidence. The highest field incidences and severities were recorded in the 3 districts, namely Nyanza, Bugesera and Gisagara, bordering Burundi. In an earlier study by Munganyinka *et al.* in 2014, Nyanza and Gisagara districts were also displaying the highest CBSD incidence, confirming them as hotspots for CBSD (93). This observation might correspond to the informal movement of cassava cuttings across countries that leads to the importation of the infected cuttings or the use of genetic material that is more susceptible to CBSIs. Furthermore, the high CBSD incidence in those districts could be due to the fact that since its first report in 2009, the virus could have flourished in those areas season after season due to the relatively warm environments that favour the proliferation of whitefly vectors (217).

A CBSD survey performed in Burundi previously reported an average incidence and severity of 15.3 % and 2.3 respectively (67). A decade ago, UCBSV was the only viral species associated with the disease in Burundi, while it was already present in Tanzania (218). Based on RT-PCR diagnostics, the overall CBSIs incidence was found to be less than CBSIs incidences reported elsewhere in East-Central Africa (Kenya, Tanzania, Malawi, Zambia) (71,197,218,219). This difference might be due to the later introduction of CBSIs in Rwanda. UCBSV was prevalent nationwide, indicating that it is the commonest cause of CBSD. Similar findings were reported in the previous survey in Rwanda (93) and in DRC (220), Zambia (71).

Even though farmers are aware that quality seed is the cornerstone that impacts the output, it was observed that most farmers have difficulties identifying quality seed, as was reported earlier (221). Notably, most farmers (76.9%) use informal ways to get cuttings for free from their fields or neighbours. Because the supplied planting material often lacks quality control, farmers are more likely to plant virus-infected cuttings. The lack of knowledge on cassava viral diseases identified in the survey might further maintain the informal seed system, highlighting the need to increase farmers' awareness of using quality seeds and mobilize the private sector to invest in commercial cassava seed businesses. Previous studies have already highlighted the need to promote farmers' awareness (18,100,222) as farmers using cassava planting materials from appropriate sources (research institutions, NGOs, etc.) appear to have fields with reduced CBSD infection (99). Despite differences in CBSIs incidence between districts, farmers who used quality seeds had a lower CBSIs infection than those re-using seeds from their own field or from other farmers.

This observation also highlights the importance of the human factor (transport and exchange of unhealthy cuttings) contributing to the propagation and dissemination of CBSD (46,48).Following Government efforts to combat cassava viral diseases since its emergence in 2009, their effect on cassava production has decreased despite past and ongoing efforts to breed for virus resistant varieties, distribute clean planting material and promote GAP (223).

Our study revealed that viral diseases remain a constraint to cassava productivity, with a disease prevalence that has increased to 35.3% in Rwanda. Therefore, there is a need to continue efforts to introgress virus resistance traits into farmer-preferred varieties and establish a cassava seed system enabling a sustainable and affordable supply of clean planting material to farmers. Strengthening the cassava seed system will also require the development of essential capacities for virus diagnostics (224,225). There is also a need to increase farmers' awareness of cassava diseases and the immediate benefit of using quality seeds.

3.6. Supplementary materials.

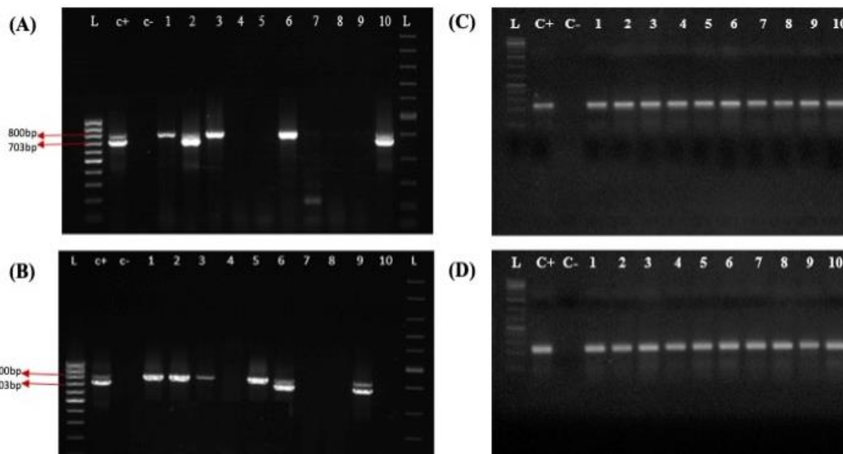
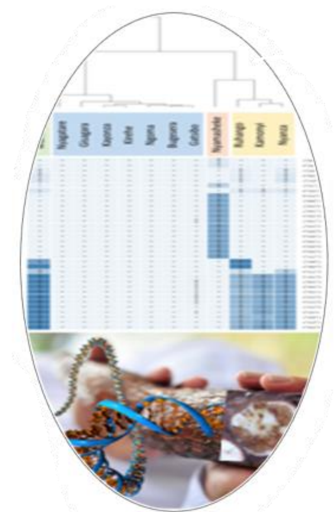


Figure S 1. Detection of CBSV and UCBSV by reverse transcription-PCR in samples from surveyed distrcits.

(A) Nyanza and (B) Bugesera districts. Expected PCR product sizes were CBSV (703 bp) and UCBSV (800 bp).From left to right, 100 bp DNA ladder; C+ correspond to positive control for both CBSV and UCBSV, whereas C– corresponds to negative control; lanes 1–10 represent pool of samples from fields 1–10; faster DNA ladder. (C) Detection of the internal control MePP2A for samples in (A); (D) detection of the internal control MePP2A for samples in (B); expected PCR product sizes was 150 bp From left to right, 100 bp DNA ladder; Lanes 1–10:samples; internal control was detected in all tested samples.

Chapter 4

Going beyond consensus genome sequences: an innovative SNP-based methodology reconstructs different Uganda cassava brown streak virus haplotypes at a nationwide scale in Rwanda.



Chapter 4. Going beyond consensus genome sequences: an innovative SNP-based methodology reconstructs different UCBSV haplotypes at a nationwide scale in Rwanda.

General Introduction to Chapter 4.

In the preceding chapter, which focused on a comprehensive field survey to assess the status of CBSD and identify the risk factors contributing to its spread in the country, it was revealed that the overall prevalence of CBSD was 35.3%. UCBSV was the most prevalent among the infections at 61%, followed by 28% of mixed infections and 11% of CBSV infections. In Rwanda, the genetic diversity of the CBSD causative agents has been predominantly limited to sanger sequencing of partial coat proteins. Therefore, the third chapter of this thesis presents a groundbreaking analysis of the UCBSV population. This analysis goes beyond the consensus level, providing crucial insights into the UCBSV evolution for the first time.

Chapter 4. Going beyond consensus genome sequences: an innovative SNP-based methodology reconstructs different Uganda cassava brown streak virus haplotypes at a nationwide scale in Rwanda.

This chapter 4 is based on the original research article published in Virus Evolution, August 24th 2023 https://doi.org/10.1093/ve/vead053. Just text formatting was used to adapt the accepted version.

Chantal Nyirakanani[1,3], Lucie Tamisier[2], Jean Pierre Bizimana[1,4], Johan Rollin[2,6], Athanase Nduwumuremyi[4], Vincent de Paul Bigirimana[3], Ilhem Selmi[2], Ludivine Lasois[1], Hervé Vanderschuren[1,5], and Sébastien Massart[2*]

[1]Plant Genetics and Rhizosphere Processes Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium

[2]Integrated and Urban Plant Pathology Laboratory, TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech, Gembloux, Belgium

[3]Department of Crop Sciences, School of Agriculture and Food Sciences, College of Agriculture, Animal Sciences and Veterinary Medicine, University of Rwanda, Musanze, Rwanda

[4]Rwanda Agriculture and Animal Resources Development Board, Huye, Rwanda

[5]Tropical Crop Improvement Laboratory, Department of Biosystems, KU Leuven, Heverlee, Belgium

[6]DNAVision, 6041 Gosselies, Belgium

4.1. Abstract

Cassava Brown Streak Disease (CBSD), which is caused by *cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), represents one of the most devastating threats to cassava production in Africa, including in Rwanda, where a dramatic epidemic in 2014 dropped cassava yield from 3.3 million to 900,000 tonnes (85). Studying viral genetic diversity at the genome level is essential in disease management, as it can provide valuable information on the origin and dynamics of epidemic events. To address the existing gap in genome-based studies on CBSIs diversity in Rwanda, a comprehensive nationwide survey of cassava ipomovirus genomic sequences was undertaken using high-throughput sequencing (HTS). It involved analyzing pooled plant samples collected from 130 cassava fields located in 13 districts known for cassava production. These districts span across seven distinct agro-ecological zones, each characterized by varying climatic conditions and diverse cassava cultivars. HTS allowed the assembly of a nearly complete consensus genome of UCBSV in 12 districts. The phylogenetic analysis revealed high homology between UCBSV genome sequences, with a maximum of 0.8 % divergence between genomes at the nucleotide level. An in-depth investigation based on Single Nucleotide Polymorphisms (SNP) was conducted to explore the genome diversity beyond the consensus sequences. First, to ensure the validity of the result, a panel of SNPs was confirmed by independent RT-PCR and Sanger sequencing.

Furthermore, the combination of fixation index (F_{ST}) calculation and Principal Component Analysis (PCA) based on SNPs patterns identified three different UCBSV haplotypes geographically clustered. The haplotype 2 (H_2) was restricted to the central regions, where the NAROCAS1 cultivar is predominantly farmed. RT-PCR and Sanger sequencing of individual NAROCAS1 plants confirmed their association with H_2. Haplotype 1 was widely spread, with a 100% occurrence in the Eastern region, while Haplotype 3 was only found in the Western region. These haplotypes' associations with specific cultivars or regions would need further confirmation. Our results prove that a much more complex picture of genetic diversity can be deciphered beyond the consensus sequences, with practical implications on virus epidemiology, evolution, and disease management. Our methodology proposes a high-resolution analysis of genome diversity beyond the consensus between and within samples. It can be used at various scales, from individual plants to pooled samples of virus-infected plants. Our findings also showed how subtle genetic differences could be informative on the potential impact of agricultural practices, as the presence and frequency of a virus haplotype could be correlated with the dissemination and adoption of improved cultivars.

Keywords: Cassava, High throughput sequencing, UCBSV, Ampelovirus, SNP, Rwanda.

4.2. Introduction

In many African regions, cassava (*Manihot esculenta* Crantz) is considered a key food security crop because of its capacity to cope with suboptimal climatic conditions and to grow on marginal land (226). The crop ranks as the sixth most important food crop in the world (200) and the third most important in Rwanda, with an average yield of 14 tons per hectare in 2021 (89). However, cassava production is still below its yield potential due to various constraints, including viral diseases (227). Cassava Brown Streak Disease (CBSD) is one of sub-Saharan Africa's most devastating threats to cassava. CBSD was first found in Tanzania in 1936 (62) and has now spread to ten East and Central African countries where cassava is a vital crop (35). The disease is caused by *Ugandan cassava brown streak virus* (UCBSV) and *cassava brown streak virus* (CBSV), which are both positive-sense, single-stranded RNA (+)ssRNA) virus species belonging to the genus *Ipomovirus* and the family *Potyviridae* (74). Both species are widely spread in Central and Eastern African countries, although UCBSV is often more prevalent than CBSV particularly in highland (28,66,71,228).

The disease is mainly vertically transmitted through planting material, in addition to the semipersistent transmission by the whitefly vector (46). After a dramatic outbreak of CBSD in Rwanda in 2014, the import and dissemination of CBSD-tolerant cassava cultivars (87) were instrumental in mitigating yield losses. However, the incidence of UCBSV and CBSV remained relatively high (28). RNA viruses often exist as a population of closely related mutants due to the low fidelity of RNA polymerases and, therefore, exhibit a fast yet constrained evolutionary rate enabling modification in virulence and transmissibility as well as a continuous virus adaptation to the changing environment (229). Therefore, studying the evolution of viral populations at different scales is of prime importance in managing a viral disease. For UCBSV, the analyses were mainly carried out on a few partial coat protein sequences. They reported a country-wide nucleotide divergence from under 1% in Mayotte (n=8 sequences), and Kenya (n=9) (69,230), to 12% in Rwanda (n=24) (93).

The first complete genome sequences of viruses were obtained by Sanger sequencing of PCR amplicons (231). They corresponded to the most frequent nucleotide at each position of the genome. However, this technique is not well adapted to detect minor alleles or Single Nucleotide Polymorphisms (SNPs - under 25% frequency) in the population. Unless coupled with cloning and sequencing of several clones per sample, amplicon sequencing generally fails to detect low-frequency polymorphisms, even though improvements have been made to detect minor alleles in tumours (232).

High throughput sequencing (HTS) has become the standard technology for studying virus population diversity, epidemiology, and evolution (193). For example, HTS-based profiling of plant virus genomes has allowed the reconstruction of the history of *potato virus V* evolution and dissemination (233) as well as deciphering the spread of the *turnip mosaic virus* along the silk road (234). For UCBSV, 63 genome sequences have been generated from Uganda, Tanzania, Kenya, DRC, Zambia, Malawi, Rwanda, and Burundi (66–68,71,74,79,235).

Although HTS technologies have facilitated the profiling of virus genomes, most publications only report consensus genome sequences for the detected and identified viruses, even from pooled samples, as done for the devastating viruses causing maize lethal necrosis (236). This consensus genome corresponds to the most frequent nucleotide at each position, providing an information level similar to Sanger sequencing technology and underexploiting the extent of sequencing data generated. Because HTS technologies generate millions of reads, tens to thousands of sequencing reads can cover each base of a viral genome. This sequencing coverage theoretically allows the identification of minor SNPs (frequency under 50%) even at a shallow frequency (below 1%) (66). The integration of minor SNPs in the analysis of viral genomes should become the rule rather than the exception, as they can drive evolutionary processes and the biological properties of viruses within their hosts (193,194,237,238). For example, minor variants in the *Coxsackie virus* have been shown to contribute to virus adaptation (195). Evolutionary studies on *barley yellow dwarf virus* (BYDV) showed that several virus populations might share the same consensus sequence while having different low-frequency SNPs patterns, highlighting the importance of characterizing virus genetic diversity beyond the consensus level (239). Therefore, it is critical to profile and reports the presence of individual low-frequency SNP present to improve the resolution of viral population analyses and to characterize the contribution of variants to virus evolution and adaptation (changes in viral load, virulence, transmission, host range, etc.).

When analyzing the genetic diversity of a virus species in the HTS dataset at the SNP level, generating the proper reference consensus sequence(s) is essential. Indeed, if the sample was infected by several divergent isolates, e.g., at least 5-10% of divergence, their consensus genome sequence reconstruction is possible using classical *de novo* assembly (75,240,241), and several consensus sequences can be obtained. The presence and frequency of SNPs can be further studied for each genome sequence. On the other hand, if the identity between isolates is higher, it becomes difficult to differentiate their genome sequences, and a unique consensus sequence is often generated. The comparison of viral populations using SNP frequencies can be performed with the fixation index (F_{ST}) (242).

F_{ST} is a measure of population differentiation usually applied to study the population genetics of pooled samples of vertebrates or plants. It has been recently applied to plant virus populations and has enabled an in-depth analysis of the virus population beyond consensus sequences reconstructed from HTS datasets (239,243). SNP frequencies can also be used as features for a principal component analysis (PCA) to reduce their complexity and increase their interpretability. SNP-based PCA has so far only been applied to study the evolution of a DNA virus infecting Drosophila (244).

In addition, the association between SNPs can be studied, e.g., if they belong to the same viral molecule. For eukaryotes, a haplotype is a set of genomic polymorphisms that tend to be inherited together. By extension, we can call haplotype a series of mutations present on the same viral molecule compared to the virus's consensus (or reference) sequence. Therefore, identifying haplotypes can improve the characterization of the viral population at the molecular level. SNPs can be associated

with haplotypes if they are located close to each other in the genome so that they can be observed on the same sequencing (paired) reads.

However, more distant SNPs are more difficult to associate with the short-read (i.e. 50-300 nt) sequencing technologies predominantly used in virus diagnostics and metagenomics studies. The emergence of new sequencing technologies generating long reads can solve this issue. Those technologies have already generated complete genomes of the viruses at high accuracy for small ssDNA circular genomes (245), but it can remain a challenge for longer viral genomes. To solve this challenge and exploit the massive amount of data generated by short-read technologies, an innovative methodology is suggested to reconstruct haplotypes of distant SNPs from short sequencing reads based on their relative frequencies within and between datasets.

In the present study, the goal was to decipher the presence and the genetic structure of UCBSV populations at the SNP and haplotype levels through a nationwide sampling in the major cassava production areas under diverse pedo-climatic zones and cultivar compositions. This study used high throughput sequencing (HTS) technologies combined with an innovative bioinformatics methodology based on the SNPs identified from the consensus sequences and their frequencies to reconstruct virus haplotypes.

4.3. Materials and Methods

4.3.1. Study Area and field sample collection

Cassava is grown on a large scale in the Central and Eastern parts of Rwanda (246). Thus, fields with cassava plants at least 6 months old from 13 cassava-producing districts were inspected for CBSD symptoms and sampled. They included five districts (Ruhango, Nyanza, Muhanga, Kamonyi, and Gisagara) from the Central regions, six districts (Bugesera, Kayonza, Kirehe, Nyagatare, Gatsibo, and Ngoma) from the Eastern regions, one district from Northern province (Gakenke) and one district from the Western province (Nyamasheke). In the latter two provinces, cassava is grown on a small scale. The sampled fields' locations are shown in **Figure S2**, and their GPS coordinates are available in **Table S1**.

The districts' climatic conditions differ in temperature, rainfall, and altitude. For example, the Eastern province is drier flatlands, the Northern province is cool mountains, whereas the low-lying valleys of southwestern provinces are warmer (247). These conditions divide the country into different agro-ecological zones, and detailed agro-ecological characteristics of each district surveyed are found in supplementary table 2 (**Figure S2**; **Table S2**).

From each district, ten fields were visited (**Figure S2**). They were selected from 5 main cassava producer sectors (selected purposely), and 2 cassava fields separated by 10 Km were selected in each sector, in order to provide a reliable overview of the ipomovirus diversity in those districts. From each field, leaf samples (symptomatic and asymptomatic) were collected from 10 plants selected randomly using the two diagonals approach across the field (248). Samples from the same field were pooled (one field was considered one sample for RNA extraction), totalling 130 pooled samples corresponding to 130 fields visited from 13 districts.

4.3.2. Total RNA extraction

Total RNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (249). Considering the cost of high throughput sequencing (HTS), RNA samples extracted from the same districts were pooled using equimolar concentration. Consequently, 13 samples of 100 plants (10 fields per district; 10 plants per field) were prepared for HTS. Bioanalyzer (Agilent) and Nanodrop (Thermofisher Scientific) determined the RNA integrity and concentration. **Table S3** shows the RNA concentration with the 260/280 ratio and the RNA integrity number of the used RNA (**Table S3**).

4.3.3. RNA-Seq library preparation and sequencing

Samples were processed at the GIGA facilities of Liège University (Liège, Belgium). Ribosomal depletion was performed by Ribo-Zero® rRNA Removal Kit (Illumina kit) following the manufacturer's guidelines. RNA Library was prepared following TruSeq Stranded Total RNA Sample Prep LS Protocol (Illumina kit) according to the manufacturer's instructions. The libraries were prepared with UDI (unique dual indexes), and a Free Adapter Removal (Illumina kit) was done on the pooled libraries. The HTS was done on the ILLUMINA NovaSeq 6000 with an S4 flowcell for 2*150 nt. Adapter removal was done with the bcltofastq v2.20 program of Illumina.

4.4. Bioinformatic analysis

4.4.1. Reads processing

First, the obtained raw reads were paired and trimmed using Geneious Prime 2023 (version 10.1.5, Biomatters) software (https://www.geneious.com). The low-quality nucleotides showing quality scores below 20 and reads with lengths lower than 35bp were trimmed using BBDuk V38.37 (250). Then, reads were merged, and duplicated reads were removed using the Dedupe V38.37 (251) plugin implemented in Geneious.

4.4.2. De novo assembly, Mapping, and Phylogenetic analysis

De novo assembly was further performed using the RNA-Spades (252) V3.13.0 assembler implemented in Geneious. The obtained contigs were subjected to a BLAST search (253) (blastn and blastx) against the viral RefSeq database (retrieved in September 2020 - release 201) to check which contigs matched viral sequences in GenBank. All contigs matching a viral reference sequence in GenBank were extracted. All reads were further mapped (Geneious mapper V10.1.5) on the viral contigs of interest with the parameter "Low sensitivity/ Fastest" used (10% mismatch). Furthermore, reads were mapped on the closest reference sequence in the database using the same parameters. The obtained mapped reads were used for SNP calling and in-depth analysis. SNP calling (Geneious V10.1.5) was performed using the following criteria: (i) a minimum coverage of 100x, (ii) a Minimum Strand-Bias > 65% (p-value ≤ 0.0005), and (iii) a minimum variant frequency of 1%.

To rule out the presence of CBSV in the analyzed samples, all samples were processed with Kraken2 (standard database from 06/2020, Version 2.1.1) on Galaxy (254,255), in addition to the BLAST mentioned above approach, and for positive samples (showing CBSV reads after Kraken results), simultaneous mapping of all reads on the closest CBSV (HG965221) and UCBSV (KX753356.1) reference genomes was performed at maximum 10 % mismatch to avoid non-specific Mapping of UCBSV reads on the CBSV sequence. In addition, to screen our samples for novel Ampelovirus recently reported to infect cassava in Central Africa and the South-West Indian Ocean Islands, all reads were mapped to the Congolese genome sequence of provisioary referred to as *Manihot esculenta-associated virus 1* (MEaV-1) (MT773588) (116). MEaV-1 was also tested by RT-PCR and sequencing of the PCR product as described previously (116).

4.4.3. Phylogenetic analysis

The percentage of identity between the consensus sequences of UCBSV and their polyproteins from 12 districts of Rwanda was conducted in MEGA X (V10.2.6) with the Poisson correction model (256).

The newly generated UCBSV whole genome sequences were aligned with the 23 UCBSV genomes from the NCBI GenBank nucleotide database using Clustal Omega V1.2.2 (257)(release 241, November 2020). Phylogenetic analysis and evolutionary divergence between nucleotides sequences were performed using the new genomes identified in the present study as well as complete or partial coat protein sequences from GenBank isolates of UCBSV using Molecular Evolutionary Genetics Analysis (MEGA-X V10.2.6) under the Neighbor-Joining method (258). The evolutionary distances were computed using the p-distance method (259). All ambiguous positions were removed for each sequence (pairwise deletion option).

4.4.4. Nucleotide diversity

SNPGenie (V1, May 2022) (260) was applied to the SNP tables to calculate the nucleotide diversity (π) for each district (sample). It represented a mean number of pairwise differences per nucleotide position in a population of sequences.

4.4.5. Genetic diversity analysis of UCBSV

A recently established methodology was applied for comparing UCBSV populations at both SNP and haplotype levels (239). Two distance measures were calculated using the consensus sequences or the fixation index (F_{ST}). F_{ST} methodology considers all the SNP detected in the UCBSV populations and their relative frequency (if above 1%) to compare samples. The consensus sequence considers only the dominant base (>50% frequency) at every position. For both methods, an index ranges between 0 and 1 was calculated, where differences between populations increase as the index changes from 0 to 1. The F_{ST} measures were obtained using Popoolation2 version 1.201 (261). First, the reads of each sample were mapped to the closest reference genome (accession number KX753357.1) as described above. Then, the analysis was performed with a single-window size defined by the size of the reference

genome (9.097 bp), a step size of 1, a minimum covered fraction of 0.1, minimum coverage of 50, and maximum coverage of 200,000.

The principal component analyses used ade4 and factoextra libraries from the R software version 4.0.3 (http://www.r-project.org/). The dendrograms were calculated using the hclust function implemented in the package stats (version 4.0.2).

4.4.6. SNP and haplotype validation

Sanger sequencing of the RT-PCR products from the 12 pooled samples and 3 individual plants of a tolerant cultivar (NAROCAS1) confirmed the presence of selected SNPs and one haplotype. PCR products were purified using a PCR product purification kit (Qiagen, German). A list of specific primer pairs sequences used for PCR amplification and sequencing is provided in supplementary table 4 (**Table S4**). In addition, the association between one of the UCBSV haplotypes and a CBSD tolerant cultivar was assessed by checking 12 SNP positions (spanning the whole genome) identified as specific for the selected haplotype.

4.5. Results

4.5.1. UCBSV is detected in nearly all generated datasets at a high abundance

The number of high-quality reads generated per sample ranged from 16,565,194 to 27,211,334, averaging 21,869,253 reads. The number of contigs generated by de novo assembly ranged from 24,503 to 49,784 per sequenced sample. In total, 12 genome sequences of UCBSV (ranging from 8,743 to 9,082 nt) with complete CDS (coding sequence) and nearly complete UTR (untranslated regions) regions were generated from 12 samples with a near-complete coverage of the reference (**KX753357.1** - 9,097 nt**)**. No UCBSV was found for one sample. The average genome fold coverage ranged between 143x to 963x, and the genome coverage was always >99% **(Table S5)**.

Complementary analyses using Kraken2 (255) identified specific CBSV reads. However, the number of reads did not exceed 25 reads per sample and did not allow the reconstruction of a complete CBSV genome sequence (**Figure S3; Table S6**). In the district of Ruhango, a recently discovered ampelovirus was detected in the dataset with 386 reads (**Figure S4)**, from which two partial ampelovirus contigs of 12,711 bp and 3,390 bp (OL579727; OL579728) were constructed. They shared 98% of their identity with the MEaV-1 Congolese isolate (MT773588). The presence of the ampelovirus was confirmed by RT-PCR (**Figure S5**) and subsequent Sanger sequencing of the amplicons (NCBI reference numbers: OL579729; and OL57973).

4.5.2. Phylogenetic analysis of consensus sequences of UCBSV revealed a high homogeneity throughout the sampled districts.

The bioinformatics analysis generated a consensus sequence of the nearly complete UCBSV genome for each of the 12 samples (districts). The phylogenetic analysis of the sequenced samples at nucleotide (nt) and amino acid (aa) levels and publicly

available sequences clustered the 12 UCBSV sequences in a single group reduction (**Figure S6**).

The 12 genomes showed very high homology between each other with a maximum of 0.8 % and 0.6 % of divergence at nucleotide (nt) and amino acid (aa) levels, respectively (**Table S7; Table S8**).

All the genomes had a very high identity (97 %) with a reference sequence of a UCBSV isolate from Tanzania (accession no. KX753357.1) (**Figure S6).** As UCBSV diversity was analyzed in 2014 based on the amplification and sequencing of 24 partial CP sequences (210 nt) (93), the partial CP sequences (210 nt) were selected from our 12 genomes. The phylogenetic analysis of the 36 sequences revealed that all the CP sequences from the present study clustered in a single group, with only two of the 24 CP sequences obtained previously. These sequences originated from Bugesera (KX168498) and Nyanza districts (KX168488) **(Figure 14)**.

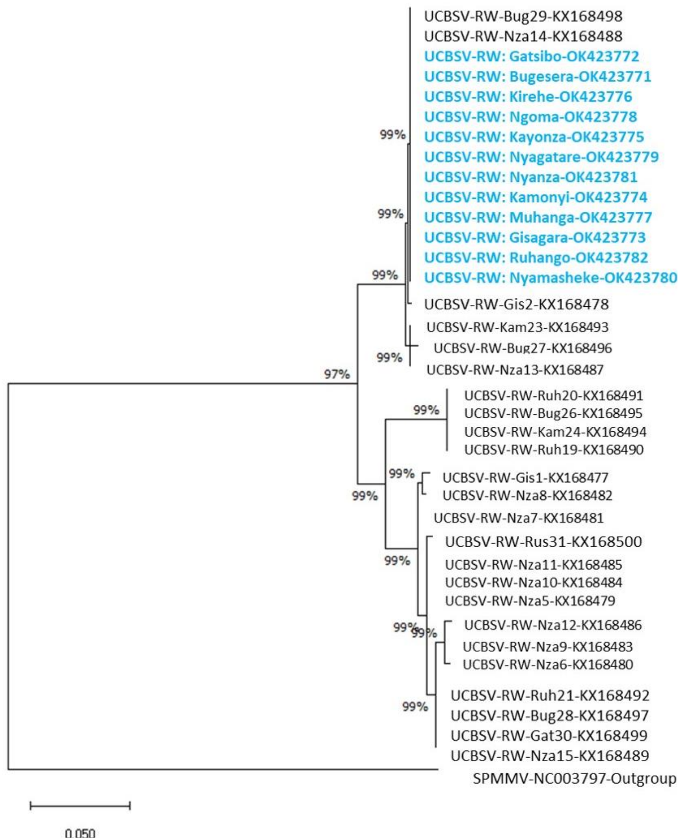


Figure 14. Phylogenetic analysis.

Phylogenetic tree of the 12 *Ugandan Cassava Brown Streak Virus* partial coat proteins (210 nt) from the present study (blue) in comparison with 24 partial coat proteins previously reported from Rwanda in 2014 (black) (93).

4.5.3. Validating SNPs by Sanger sequencing

SNPs were identified on 486 positions using all the UCBSV reads (12 samples) aligned on the closest UCBSV reference genome from Tanzania (KX753356.1). Among them, 192 corresponded to non-synonymous mutations and 294 to synonymous or silent mutations. The region with the highest NS mutations was CP, followed by Nib protein with 47/192 and 39/192, respectively (**Table S9**). The number of polymorphic sites per district ranged from 122 in Kirehe to 225 in Nyanza.

Before the in-depth analysis of SNPs, the robustness of the SNP identification was assessed by RT-PCR carried out on the twelve RNA extracts with seven primer pairs. The 84 amplification products were subsequently sequenced to check the nucleotides on 35 mutated positions scattered on the UCBSV genome. In total, 420 SNP positions (among which 91 SNPs corresponded to a mix of two alleles) contributing to the differentiation between UCBSV haplotypes were verified. High-quality sequencing was not achieved for 21 positions (5%). Ninety-seven per cent of the other SNPs (387/399) were confirmed, including mixed alleles for 80 SNPs. For 11 SNPs, only one of the two alleles was observed, mostly the allele with a higher frequency on the genome alignment (**Table S10**). This high rate of independent validation by Sanger sequencing confirmed the reliability of the obtained SNPs.

4.5.4. Differences in nucleotide diversity are observed between districts

Despite a high identity of consensus sequences of isolates between districts, the nucleotide diversity (π) appeared much more variable. The first cluster of districts (Gisagara, Kayonza, Kirehe, Ngoma, Bugesera, and Gatsibo) presented a π from 4.4 10^{-4} to 6.5 10^{-4} and corresponded to the district located in the Eastern part of Rwanda. Another cluster of districts, including Nyanza, Ruhango, Kamonyi, and Nyamasheke, presented a π nucleotide diversity that was ten times higher (4 4.6 10^{-3}). Three of these districts were contiguous and located in the central part of the country, while the fourth district (Nyamasheke) was the only western district surveyed. In addition, two districts (Nyagatare and Muhanga) presented an intermediate π nucleotide diversity with 1.3 10^{-3} and 8.1 10^{-4}, respectively (**Table S9**).

The variation in π nucleotide diversity across the districts prompted us to conduct an in-depth analysis of SNPs in each sample using an innovative methodology.

4.5.5. F_{ST} methodology revealed UCBSV genetic diversity beyond the consensus sequences.

To analyze the UCBSV genetic diversity in the sampled districts, the classical measure of distance between samples, based on the consensus sequence generated from each sample, was compared to an innovative approach based on the fixation index (F_{ST}). The F_{ST} calculation considers all the SNPs (>1% frequency) detected in the UCBSV populations from each sample. Figure 15A and B shows the dendrograms obtained by both methods, while supplementary figure 7A and 6B show the matrices (**Figure S7**). Both methods clustered seven districts together, but the F_{ST} method provided better discrimination between identical samples based on their consensus sequences.

Nyagatare district was the more distant, although it was identical to Kayonza at the consensus level. This difference aligns with the π 2.5X higher in the Nyagatare district compared to Kayonza. Nyamasheke district was also distinct from all other districts independently of the method used. Both methods clustered together Ruhango, Kamonyi, and Nyanza districts also presented similar π. Muhanga district clustered with these three districts based on the consensus approach but corresponded to a specific cluster by F_{ST} analysis (**Figure 15 A and B**), reflecting its lower π ($8.14\ 10^{-4}$) compared to the three districts. Overall, the discrimination between samples was improved by considering all the SNPs with a frequency above 1% for at least one sample. So, the next step of our analysis aimed to identify the SNPs discriminating the samples from each other.

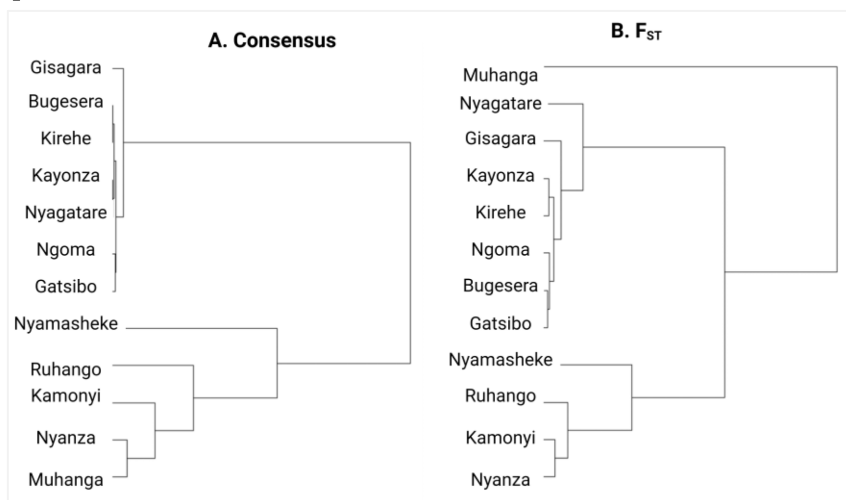


Figure 15. Analysis of UCBSV diversity.

(**A**) **Consensus approach**: Dendrogram built from pairwise distance matrices obtained after multiple alignments of the virus consensus sequences. (**B**) ***F*$_{ST}$ approach:** Dendrogram built from pairwise F_{ST} matrices obtained using the entire set of SNPs detected in the virus populations.

4.5.6. Principal Component Analysis of SNPs confirmed the clustering by *the F_{ST}* approach

The frequencies of the 486 SNPs detected in the samples from the 12 districts were used as variables to perform a Principal Component Analysis (PCA). The SNPs used to create each dimension and their frequencies are shown in supplementary table 11 (**Table S11**). The first, second, and third dimensions explained 22.8%, 15.5%, and 12.7% of the total variation, respectively (**Figure 16 A and B**). Most of the samples were clustered into two groups by dimension 1: (i) Nyanza, Ruhango, Kamonyi, and Muhanga, and (ii) Bugesera, Kirehe, Gatsibo, Kayonza, Nyamasheke, Gisagara, and Ngoma districts. Dimensions 1 and 2 discriminated Nyagatare district from the other districts. Dimension 3 separated the Nyamasheke district from the others.

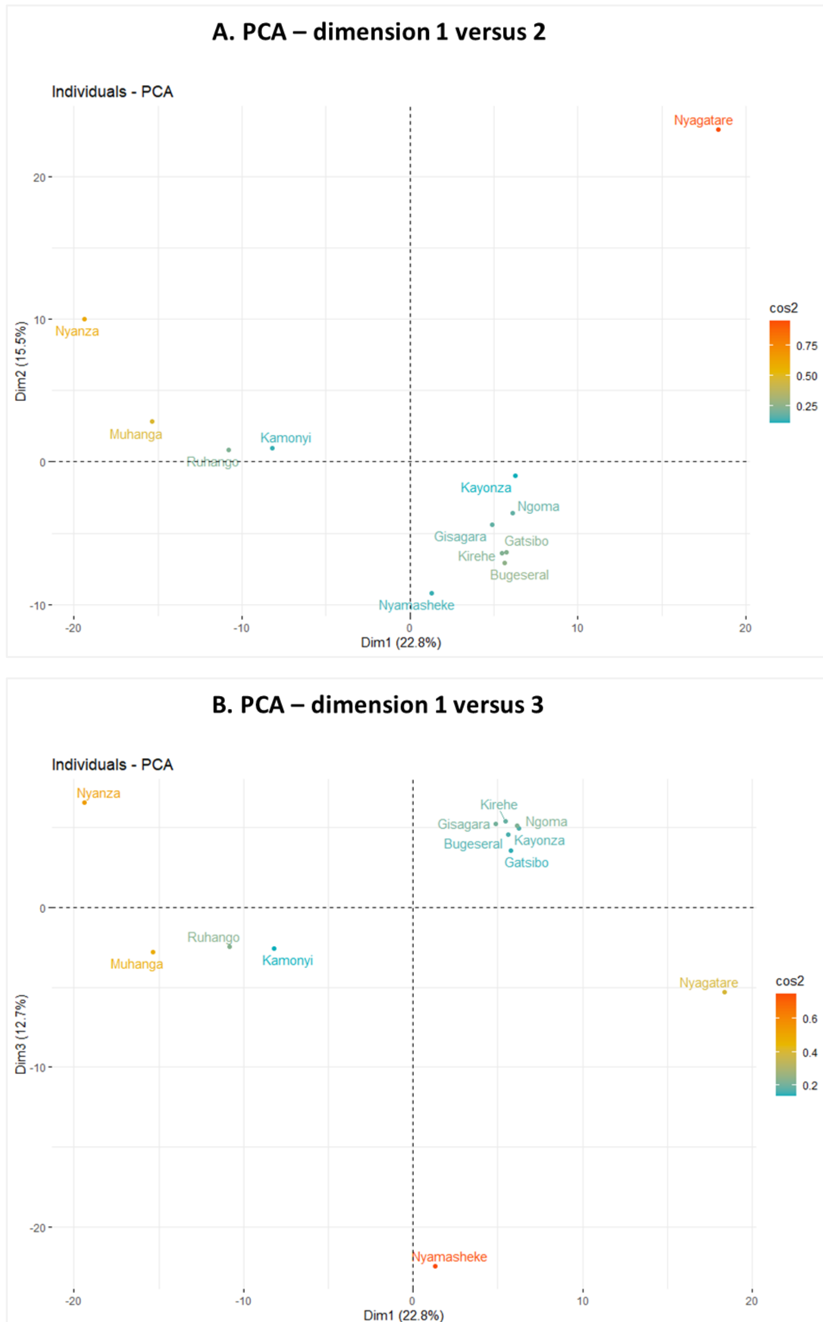


Figure 16. Principal component analysis (PCA) of the virus populations.
The PCA shows the first, second, and third dimensions obtained using all the detected SNPs as variables for *Ugandan cassava brown streak virus* sequences from the 12 districts.
(A) PCA-Dimension 1 versus 2. (B) PCA-Dimension 1 versus 3.

4.5.7. The identified UCBSV haplotypes are geographically clustered

To obtain further insight into the SNPs distribution, the frequencies of the 92 SNPs contributing the most to the first, second, or third dimension in each sample were extracted and compared to the previously obtained FST dendrograms (Figure 17).

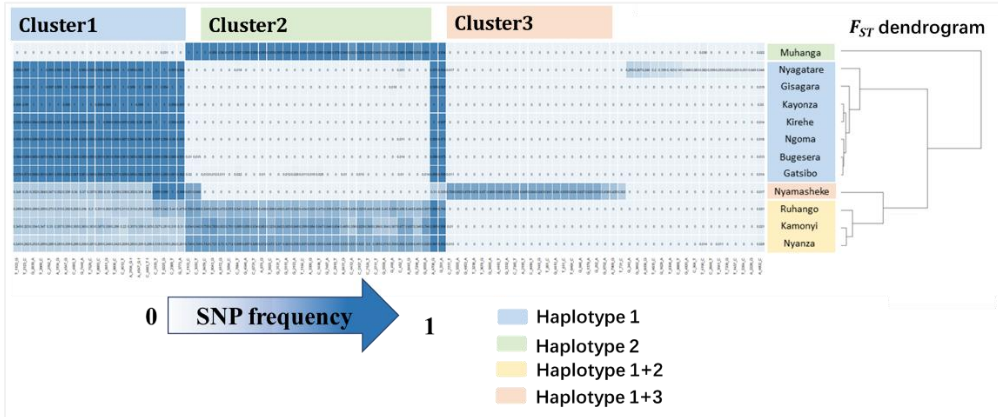


Figure 17. SNPs contributing to the differentiation between UCBSV populations. Dendrograms have been constructed from pairwise F_{ST} matrices obtained using the entire set of SNPs detected in the UCBSV populations. The frequencies of the 92 SNPs contributing to each sample's first, second, and third dimensions were extracted and compared to the F_{ST} dendrograms. The darker the blue, the higher the SNP frequency.

The SNP frequencies showed a pattern explaining the dendrogram. Several SNPs were always present at a very similar frequency for each sample and formed a cluster. Three major clusters were identified according to the following criteria: presenting at least 10 SNPs with a frequency higher than 10% in at least one sample and with a frequency varying similarly between the samples. Importantly, each cluster included SNPs present along the genome, sometimes at distant locations. For example, the first cluster included the SNPs T-1103-G and T-7529-C (SNPs being named according to their position on the genome and reference and alternative alleles) located 6,000 nt apart. Despite these distances, the frequency of the SNPs varied homogenously between the samples of a cluster, suggesting that they are linked and could constitute a haplotype. Our dataset highlighted 3 major haplotypes of UCBSV across Rwanda. Haplotype 1 (H_1) was defined by 30 SNPs at high frequency in clusters 1 and absent or low frequency for clusters 2 and 3. Among these SNPs, eight were non-synonymous and located mainly in the sequences of the P1 protein (3 SNPs) and the coat protein (2 SNPs) (**Table S9**). The frequency of this haplotype was close to 100% in Nyagatare, Gisagara, Kayonza, Kirehe, Ngoma, Bugesera, and Gatsibo districts (frequency of specific SNPs ranging between 86% and 100%). Its frequency was close to 30% in Ruhango, Kamonyi, Nyanza, and Nyamasheke districts, while it was absent in the Muhanga district. The presence of H_1 in seven samples was consistent with the results of PCA, where dimension one divided samples into two groups, one

of the groups being composed of districts showing only H_1 (Bugesera, Kirehe, Gatsibo, Kayonza, Gisagara, and Ngoma). Nyagatare district was slightly different from the other districts, with some minor SNPs in this virus population. This explains why PCA dimension two isolated the Nyagatare district from the others (**Figure 16 A**). The 28 SNPs of cluster 2 defined haplotype 2 (H_2), and their frequency was 100% in the Muhanga district. Among these SNPs, six were non-synonymous, and two were in the coat protein (2) (**Table S9**). It did not have any of the 30 SNPs similar to H_1, which could explain why this sample did not belong to any group in the F_{ST} dendrogram. H_1 and H_2 are mixed with frequencies around 30 and 70 %, respectively, in Ruhango, Kamonyi and Nyanza districts. Haplotype 3 (H_3), characterized by 28 SNPs, was only present in the Nyamasheke district, mixed with H_1. H_3 seemed to share the SNPs C-3155-T, G-6355-A, T-8225-G, G-8636-A, with H_1 as their frequency was 100%. H_3 explained the PCA results where dimension 3 separated Nyamasheke district from the rest of the districts (**Figure 16 B**). Compared to the two other haplotypes, H_3 presented seven non-synonymous SNPs located in the P1 protein (2) and the coat protein (2) (**Table S9**). Importantly, haplotype 1 was widely spread in the Eastern, Central-Southern and Western regions. In contrast, the haplotype 2 distribution was limited to the Central regions and haplotype 3 in the West of the country (**Figure 18**). When the 3 haplotypes were compared, no shared non-synonymous (NS) mutations existed. H_1, H_2 and H_3 had 8, 6 and 7 unique NS mutations, respectively. For synonymous (S) mutations, H_1 and H_2 had 22 unique S mutations each, while they had only 3 S mutations in common. H_3 had 21 unique S mutations. There were 4 S mutations common between H_1 and H_3 and 2 S mutations common between H_2 and H_3 (**Table S9**).

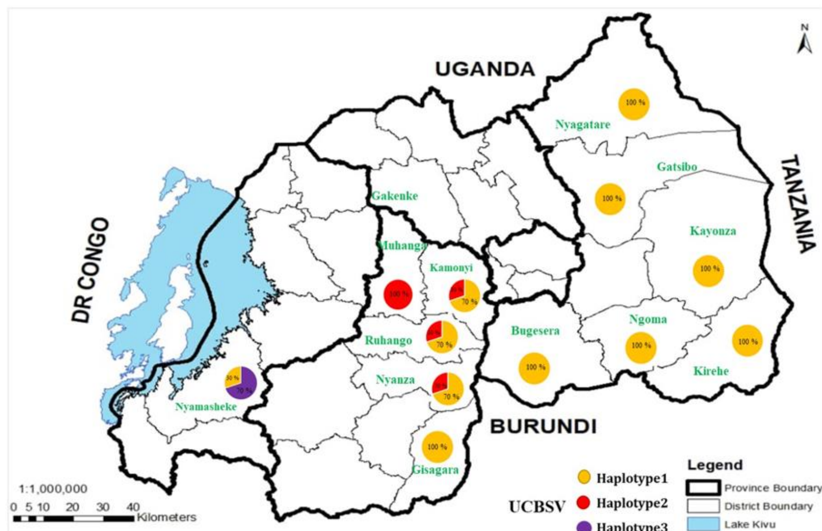


Figure 18. Country-wide distribution of UCBSV Haplotypes.
distribution of UCBSV Haplotypes identified was presented on the Rwandan map. A pie chart was used; single colour represents the presence of a single haplotype (100%). Two colours represent the presence of two different haplotypes.

4.5.8. Haplotype 2 distribution appears to be associated with the presence of the CBSD-tolerant cultivar NAROCAS1

An exciting association was observed between the haplotype H_2 and the NAROCAS1 cultivar. The NAROCAS1 cultivar is only present in the central regions (Kamonyi, Muhanga, Ruhango, and Nyanza districts at a frequency of around 25%). The same regions are also the only regions where H_2 was found (**Table S2**). The association between H_2 and NAROCAS1 was further verified by sequencing the RT-PCR amplicons from 3 individual NAROCAS1 samples using four primer pairs matching the UCBSV genome sequence. In total, 12 SNPs position spanning the whole genome that discriminates UCBSV haplotype H_2 from others were verified and confirmed the presence of H_2 in the three NAROCAS1 samples (**Table S12).**

4.6. Discussion

The present study computed SNP frequencies by combining F_{ST} analysis and PCA to study the UCBSV genome diversity and reconstruct haplotypes. It allowed the discrimination of very close virus isolates (>99%) and the characterization of 3 haplotypes whose distribution was clustered: H_2 was associated with the presence of tolerant cultivar NAROCAS1 in the central region, H_3 with the Western region (higher altitude and a specific cultivar, Mushedule was particularly abundant) while H_1 was found all over the country but at different frequencies. Two additional viruses were also detected in the 130 samples, namely CBSV and MEaV-1.

This methodology relies on high throughput sequencing of pooled samples and a combination of viral sequences analysis tools. First, the pooled plant samples were sampled using a balanced and systematic sampling scheme. The balanced and constant sampling used in the present study was instrumental in comparing the virus diversity between districts. Recent studies have proved that pooling is a cost-effective approach as a pooling of up to 50 samples has enabled comprehensive virome analysis of several virus species on a larger geographical scale in potatoes (262), *Poaceae* (263), flies or bees (264,265). Moreover, a recent study on pea viruses reported that while pooling 120 leaves into a single bulk field sample (BFS), viruses present at a low incidence were still detectable by HTS. Three of the BFS were re-tested in-depth by HTS, and no additional plant viruses were identified (266). Many studies have proven that analyzing minor variants at low frequency is essential to understand the virus diversity, its evolution and its interactions with the host (267,268) with examples of individual or pooled samples (269), with the *zucchini yellow mosaic virus* (238), or the Chlorovirus (270). Some studies above selected SNPs with a relative frequency above 1%, as in the present study (194,238,267). In addition, the obtained genomes were very well covered as they ranged from 142.9 to 963.5 for the analyzed samples.

In several studies, the conclusions remained constrained by the lack of independent validation (271,272). In the present study, the detection of a selected set of notable SNPs across the genome was independently validated, enhancing the credibility of the generated data. The validation process yielded a confirmation rate exceeding 97%.

The unconfirmed mutations occurred mainly in position with a mix of two bases, one minor at low frequency (<30%) that was not observed, probably due to the confirmed lower sensitivity of Sanger sequencing for detecting SNPs at low frequency (273). Our methodology was applied to 13 pools of plants representing each district and spanning different agro-ecological zones in Rwanda (S1 Fig). Overall, it revealed a low divergence of viral consensus sequences between districts. In addition, partial CP sequences (210 nt) extracted from the 12 consensus sequences presented 100% identity with two sequences obtained in 2014, suggesting a slow evolution of this short region from the UCBSV genome.

A reduction in UCSBV diversity was observed in the current dataset compared to the reported UCBSV diversity in 2014 and should be discussed technically and scientifically. First, the present study used HTS technologies that have improved inclusivity compared to targeted RT-PCR as demonstrated by many publications (241,274,275), so they could theoretically detect a broader range of isolates. Regarding the representativeness of the obtained sequences, the partial CP sequencing was carried out in 2014 using a limited number of plants per district, from one in the Gatsibo district to thirteen in the Nyanza district, thus most probably capturing the most abundant isolates. In contrast, the HTS protocol in the current study was applied to pools of 100 plants per district. Even if rare isolate present in one or a few plants might have been missed out by the sample pooling, Field-Based Sequencing (FBS) applied on pools of 120 plants following a similar protocol demonstrated the ability of HTS technologies to detect the viruses present in such pools reliably (266). Our bioinformatic methodology included SNPs with a frequency higher than 1%. Therefore, it theoretically identified other major isolates as the ones detected in 2014. One hypothesis on the observed difference in UCBSV diversity could originate from the strong shift in cassava cultivars experienced in Rwanda following the severe CBSD crisis. The present study used a majority (approximately 60%) of plant samples from the two CBSD tolerant cultivars introduced in 2015 (NASE14 and NAROCAS1), while, in 2014, the survey was carried out on local susceptible cultivars. The deployment of these imported cultivars has relied on official distribution to farmers and informal exchanges between farmers within and between districts (87).

Beyond the very high genetic homogeneity of consensus sequences observed throughout the country, the genetic diversity within and between samples was analyzed through a combination of bioinformatic tools, including the nucleotide diversity (π), the calculation of the F_{ST}, and principal component analysis on SNPs to highlight differences between closely related genomes and to identify haplotypes representing molecules actually infecting the plants. This methodology of haplotype reconstruction on SNPs spanning the entire genome is complementary to the current methodologies based on SNPs shared on sequencing reads that often reconstruct haplotypes spanning partial genome sequences, as recently observed for the Lolium latent virus detected in pooled samples (263). The SNPs discriminating the three haplotypes were mainly synonymous, although each haplotype contained respectively 8, 6, and 7 unique and non-synonymous SNPs. Such observation deserves further investigation on its consequences as they are located in P1 and CP genes important in

plant-virus interactions (276,277). The H_3 was found only in the Western region of the country with higher altitude, lower temperature, higher rainfall and the high frequency of the Mushedule cultivar. On the other hand, the H_1 was widely spread with a 100% frequency in the Eastern region, which borders Uganda and Tanzania and is characterized by a lower altitude, less rain, a higher temperature and a high frequency of NASE14 cultivar. In contrast, the H_2 was restricted to the central regions and was found to be associated with the NAROCAS1 cultivar. However, this confirmatory work could also be carried out for H_3 and H_1 to clarify whether the haplotypes are linked with cassava cultivars or geographical occurrences. The plants analyzed in this study were sampled in the frame of a broad survey including 130 fields around the country. The disease incidence varied strongly while the disease severity on symptomatic plants was more constant across regions. For example, the CBSD mean severity scores were 3+/-0.6, 2.9+/-0.8 and 2+/-0.2 in Eastern, Central, and the Western part respectively (28). Noteworthy, in our context of field-based sampling, determining the association of disease severity with specific cultivars and haplotypes will require a new cultivar-based survey combined with greenhouse inoculation assays as recommended for evaluating robustly causal association (278).

In the future, the reported SNPs could serve as markers to investigate and decipher the factors impacting UCBSV genetic evolution in Rwanda and the geographical distribution of the 3 haplotypes. Incorporating testing the presence of these haplotypes on planting material in the regular viral testing should be a priority to ensure the distribution of healthy planting materials, which is an essential control measure in CBSD management. Furthermore, future research activities should investigate the impact of cassava varieties distribution on ipomoviruses diversity and distribution in Africa as well as the association between the identified UCBSD haplotypes and the CBSD severity symptoms on individual cassava genotypes, including the currently widely distributed cultivars NASE14 and NAROCAS1.

Overall, our results provided evidence that a much more complex picture of genetic diversity can be deciphered beyond the consensus sequences with practical implications on virus evolution and its management. Our methodology proposed a high-resolution analysis of genome diversity between and within samples. It can be used at various scales, from individual plants to plants pooled by geographical origin (from field to region) or any other factor (cultivar, phenotype).

4.7. Data availability

The datasets of genome sequences generated and analyzed during this study are available in the GenBank repository under the following accession numbers: OK423771; OK423772; OK423773; OK423774; OK423775; OK423776; OK423777; OK423778; OK423779; OK423780; OK423781; OK423782; OL579727; OL579728; OL579729; and OL57973. In addition, raw data were deposited in SRA (PRJNA768633) and can be found at https://www.ncbi.nlm.nih.gov/sra/PRJNA768633.

4.8. Supplementary materials

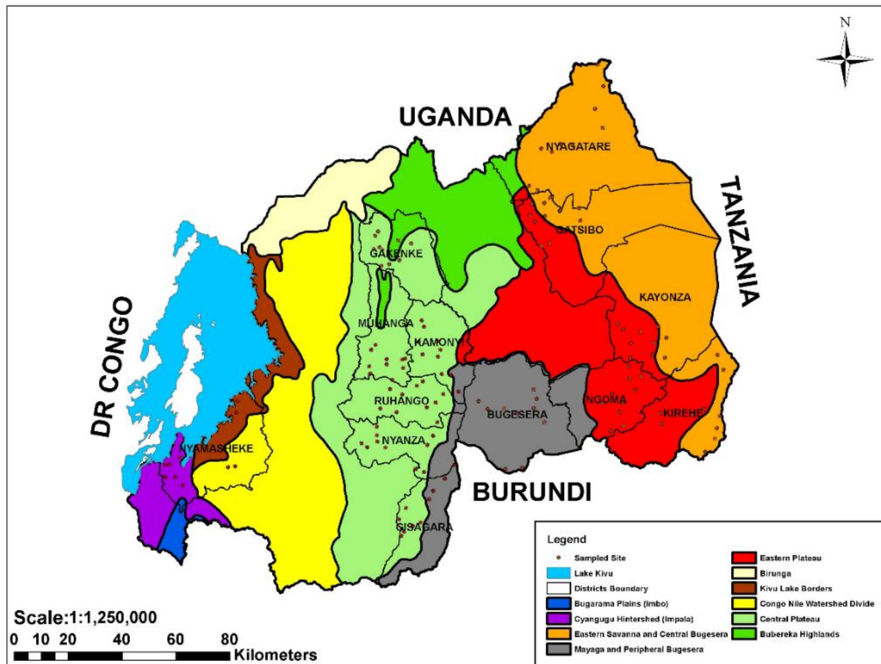


Figure S 2. Map of Rwanda showing surveyed districts and their agro-ecological zones. The dots represent the location of assessed cassava fields in 2019.

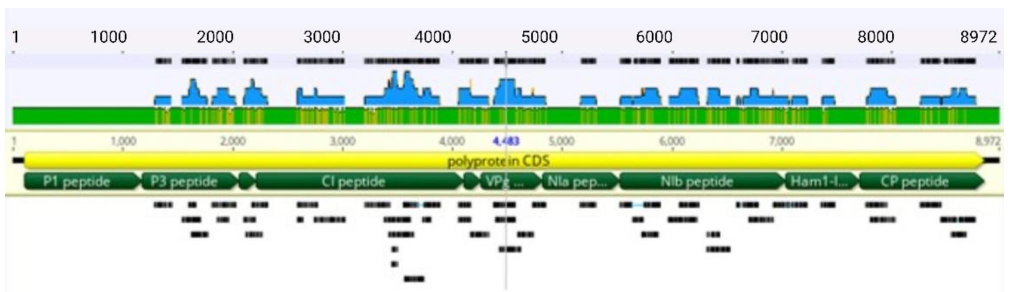


Figure S 3. Mapping of *Cassava brown streak virus* reads to the closest reference.

The CBSV reads from the 4 districts were mixed and mapped to the closest CBSV (HG965221).

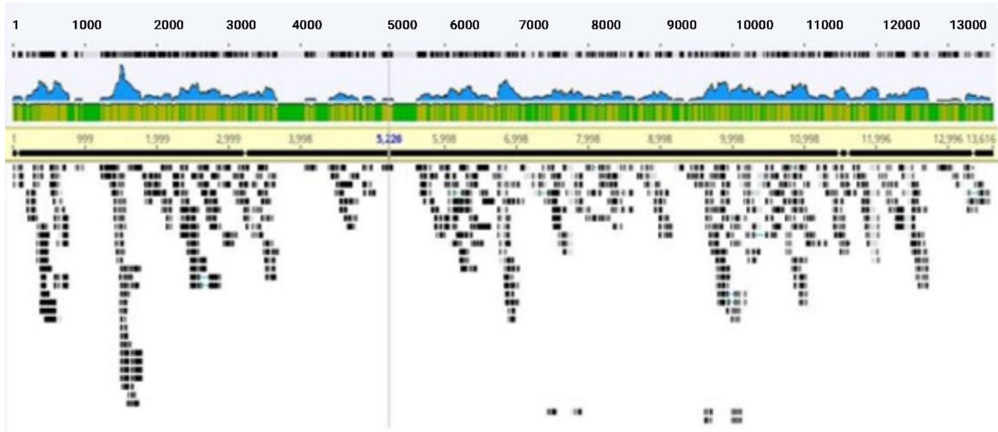


Figure S 4. Mapping of all reads to the Congolese genome sequence of MEaV-1 Ampelovirus (MT773588).

MEaV-1 Ampelovirus was present in one district of Rwanda called Ruhango with 98% pairwise identity to MT773588, with 386 reads mapping 87% of the genome.

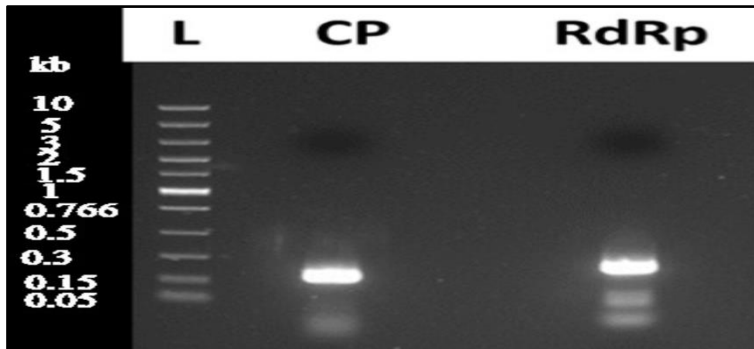


Figure S 5. Confirmation of novel Ampelovirus.

The gel image depicts RT-PCR of the Ampelovirus confirmation using primer pairs signed on coat protein that yield amplicons of 203 bp and RNA-dependent RNA polymerase (RDRP), which yield amplicons of 261 bp. L: Fast DNA Ladder.

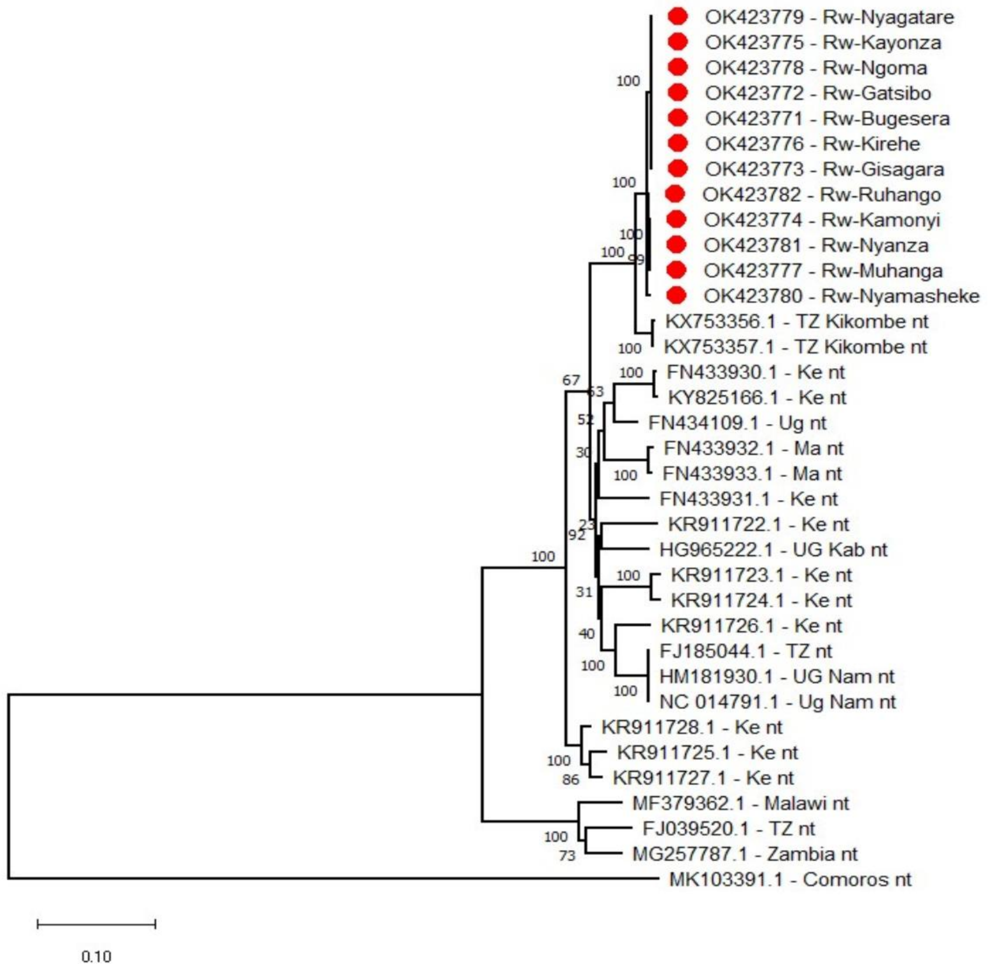


Figure S 6. Phylogenetic analysis.

Maximum likelihood phylogenetic tree (1000 bootstraps) generated from 12 *Ugandan cassava brown streak virus* (UCBSV) full or nearly complete sequences produced in this study compared with 23 UCBSV from NCBI. The analysis grouped the 12 sequences into 1 group. All the genomes showed a pairwise identity of 97% to the Tanzania isolate.

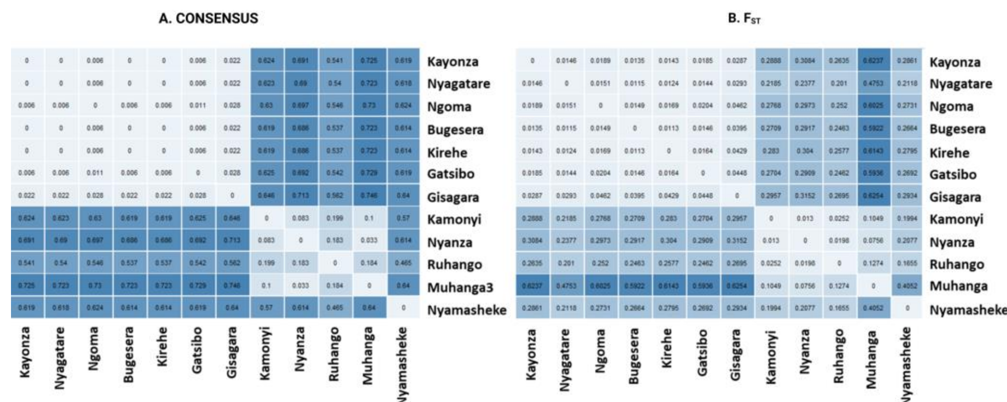


Figure S 7. Pairwise identity matrices of the UCBSV haplotypes.

Pairwise identity matrices were obtained from multiple alignments of the UCBSV consensus sequences and from pairwise F_{ST} matrices obtained using all the SNPs detected among the UCBSV populations. The values obtained with both methods ranged between 0 and 1; 0 means that two populations are genetically identical, and 1 means that two populations are completely divergent.

Table S 1. GPS coordinates of sampled cassava fields.

Districts	Cassava field	1	2	3	4	5	6	7	8	9	10
Nyagatare	Longitude	30.430	30.449	30.422	30.447	30.344	30.306	30.276	30.242	30.230	30.201
	Latitude	-1.064	-1.136	-1.212	-1.275	-1.333	-1.330	-1.357	-1.345	-1.484	-1.471
Gatsibo	Longitude	30.367	30.372	30.271	30.252	30.254	30.268	30.244	30.218	30.205	30.211
	Latitude	-1.546	-1.587	-1.512	-1.528	-1.572	-1.664	-1.674	-1.591	-1.567	-1.533
Ngoma	Longitude	30.517	30.472	30.502	30.449	30.479	30.514	30.556	30.537	30.576	30.533
	Latitude	-2.275	-2.295	-2.231	-2.201	-2.169	-2.199	-2.156	-2.119	-2.110	-2.077
Kirehe	Longitude	30.830	30.850	30.838	30.828	30.821	30.789	30.785	30.750	30.643	30.640
	Latitude	-2.038	-2.089	-2.244	-2.285	-2.319	-2.364	-2.220	-2.240	-2.270	-2.234
Kayonza	Longitude	30.540	30.542	30.516	30.497	30.573	30.619	30.613	30.654	30.660	30.633
	Latitude	-2.048	-2.000	-1.953	-1.935	-1.962	-1.997	-2.037	-2.049	-1.981	-1.944
Bugesera	Longitude	30.217	30.226	30.253	30.217	30.161	30.118	30.063	30.033	30.179	30.122
	Latitude	-2.154	-2.186	-2.265	-2.217	-2.235	-2.219	-2.221	-2.194	-2.418	-2.423
Kamonyi	Longitude	29.966	29.940	29.910	29.897	29.906	29.898	29.850	29.843	29.833	29.846
	Latitude	-2.163	-2.107	-2.103	-2.045	-2.022	-1.993	-1.944	-1.923	-1.995	-2.037
Muhanga	Longitude	29.778	29.781	29.745	29.672	29.668	29.675	29.689	29.788	29.726	29.729
	Latitude	-2.054	-2.083	-2.123	-2.102	-2.077	-2.022	-2.005	-2.051	-2.053	-2.071
Ruhango	Longitude	29.896	29.860	29.889	29.915	29.778	29.759	29.705	29.737	29.847	29.827
	Latitude	-2.264	-2.215	-2.199	-2.170	-2.172	-2.231	-2.219	-2.154	-2.128	-2.142
Nyanza	Longitude	29.851	29.823	29.862	29.881	29.694	29.722	29.695	29.661	29.642	29.694
	Latitude	-2.432	-2.422	-2.341	-2.311	-2.309	-2.350	-2.330	-2.349	-2.336	-2.280
Gisagara	Longitude	29.920	29.881	29.868	29.839	29.813	29.784	29.774	29.768	29.791	29.952
	Latitude	-2.454	-2.496	-2.524	-2.602	-2.613	-2.632	-2.647	-2.591	-2.553	-2.408
Gakenke	Longitude	29.808	29.768	29.735	29.709	29.711	29.703	29.683	29.688	29.697	29.765
	Latitude	-1.665	-1.721	-1.734	-1.741	-1.686	-1.676	-1.684	-1.639	-1.624	-1.654
Nyamasheke	Longitude	29.228	29.224	29.220	29.197	29.036	29.003	28.988	28.988	29.021	29.045
	Latitude	-2.193	-2.232	-2.413	-2.416	-2.383	-2.406	-2.408	-2.448	-2.449	-2.477

Table S 2. Characteristics of Agro-ecological Zones (AEZs) of the districts surveyed as well as common cassava cultivars and intensity of cassava cultivation per districts

Districts (Sample N°)	AEZs	Rainfall (av.mm)	Temper at (Av. °C)	Altitude (Av.m)	Soil type	Dry period (days/year)	% of commonly grown cultivars										Ranking of cassava cultivation; A scale from 1 to 13; 1= the highest cassava cultivation
							Nase14	Narocas1	Makadamia	Kwatamumpare	Gumino	Kicaro	Rutanihisha	Mushedule	Gacyacyari	Others	
Muhanga	Central plateau	1298	19	1749	Humiferous	59	50	25	0	0	0	0	0	0	0	25	8
Nyanza	Central plateau	1298	19	1749	Humiferous	59	60	20	0	0	0	0	0	0	0	20	2
	Mayaga	1101	21	1403	Clayey, schists	109											
Ruhango	Central plateau	1298	19	1749	Humiferous	59	60	25	10	0	0	0	0	0	0	5	1
	Mayaga	1101	21	1403	Clayey, schists	109											
Kamonyi	Central plateau	1298	19	1749	Humiferous	59	45	20	10	0	0	0	0	0	0	25	3
	Mayaga	1101	21	1403	Clayey, schists	109											
Gisagara	Central plateau	1298	19	1749	Humiferous	59	46	20	0	0	0	0	0	0	0	34	10
	Mayaga	1101	21	1403	Clayey, schists	109											
Bugesera	Mayaga	1101	21	1403	Clayey, schists	109	45	25	0	0	0	0	0	0	0	30	5
	Eastern savanna	902	21	1386	Strongly weathered	126											
Nyagatare	Eastern savanna	902	21	1386	Strongly weathered	126	40	0	0	0	0	25	0	0	0	35	7
Kayonza	Eastern savanna	902	21	1386	Strongly weathered	126	40	0	0	30	0	0	0	0	0	30	6
	Eastern plateau	1038	20	1578	Laterite soil	86											
Gatsibo	Eastern savanna	902	21	1386	Strongly weathered	126	45	0	0	20	0	0	0	0	0	35	12
	Eastern plateau	1038	20	1578	Laterite soil	86											
Kirehe	Eastern savanna	902	21	1386	Strongly weathered	126	40	0	0	0	25	0	0	0	0	35	4
	Eastern plateau	1038	20	1578	Laterite soil	86											
Ngoma	Eastern savanna	902	21	1386	Strongly weathered	126	40	0	0	25	0	0	0	0	0	35	9
	Eastern plateau	1038	20	1578	Laterite soil	86											
Nyamash	Cyangugu hinterl	1710	19	1666	Vey fine, red < basalt	56	0	0	0	0	0	0	0	46	20	34	11
	Lake Kivu border	1225	20	1638	Shallow, clay loam	66											
	Congo-Nile wate	1542	17	2058	Humiferous, acid	27											
Gakenke	Buberuka highlan	1267	17	1957	Laterite soil	41	0	0	0	0	0	0	28	0	25	47	13

Table S 3. RNA concentration, 260/280 Ratio and RIN of the used samples.

Samples/Districts	Concentration (ng/µl)	A260 / A 280	RIN
Nyanza	1145	2.008	6.1
Ruhango	960.4	2.018	5.7
Bugesera	1050	2.005	5.1
Kirehe	1098	1.959	5.1
Nyamasheke	908	1.974	5.4
Gatsibo	1209	1.956	4.9
Gakenke	806.4	1.953	4.8
Kamonyi	1561	1.943	6.1
Muhanga	1081	1.9	6.6
Kayonza	920.4	1.942	6
Nyagatare	1351	1.907	5.4
Gisagara	848.4	1.935	5.8
Ngoma	1774	1.972	5.2

Table S 4. RT-PCR primer pairs used for the confirmation of the essential SNPs.

Primer pairs	Sequences (5' => 3')	Tm
898=>1398	Fwd: GGAAAGCTATACCAATGATC	53.2
	Rev: CAATTATCGGACTACCCT	51.41
2640=>2814	Fwd: GGTCGAGTCAGAACTGGTTC	59.35
	Rev: GCTCAACTCACGCAACAATG	57.3
2894=>3307	Fwd: ACACACGCTCGAGATAATGG	57.3
	Rev: TCYCGACTCTCCCAAATC	54.83
3166=>4035	Fwd: GTGTYACCCTTGAYGTTG	53.69
	Rev: GAGACGCTGAAYCCGATTG	57.75
4350 => 4833	Fwd: GAAGTGCTGCAGTTGAAC	53.69
	Rev: GCCTCCATCTACCCTCTTC	58.82
5555 => 6053	Fwd: RGCRAATGAAGGTGAAGAAT	53.2
	Rev: CCGCTCCAAATACCAACATC	57.3
7200=>7564	Fwd: GGCATTRCAAGTGATGTTA	51.27
	Rev: CCCAGCCAAATCCATTGTTA	55.25

Table S 5. High Throughput Sequencing Data of UCBSV in Rwanda. Nearly complete UCBSV was recovered in 12 per 13 districts.

Districts	N° of reads obtained	N° of reads after trimming	N° of contigs produced	N° of UCBSV contigs	Accession number (GenBank)	Accession of NCBI Ref seq	Length of consensus sequence	N° of reads mapped to NCBI ref	Average coverage	Query coverage %	Pairwise Identity%
Nyanza	17,955,754	16,565,194	31312	1	OK423781	KX753357.1	8950	15291	260.1	100	97.44
Ruhango	19,536,212	17,995,068	33169	1	OK423782	KX753357.1	9047	52541	897	100	97.61
Bugesera	29,249,980	27,211,334	29781	1	OK423771	KX753357.1	9054	49028	833.4	100	97.36
Kirehe	21,239,054	19,855,700	24503	1	OK423776	KX753357.1	9044	37741	644.9	100	97.37
Nyamasheke	20,308,612	19,375,730	43545	1	OK423780	KX753357.1	9082	56545	963.5	100	97.51
Gatsibo	21,725,568	20,320,336	49784	1	OK423772	KX753357.1	9061	50128	855.7	100	97.37
Gakenke	24,641,094	23,493,718	41336	None							
Kamonyi	20,735,394	19,741,696	37840	1	OK423774	KX753357.1	9075	25297	434.2	99	97.5
Muhanga	19,840,361	18,368,666	34704	1	OK423777	KX753357.1	9057	38933	193.4	100	97.43
Kayonza	20,030,480	18,572,746	36285	1	OK423775	KX753357.1	9047	16549	142.9	100	97.36
Nyagatare	19,199,016	17,981,060	45103	1	OK423779	KX753357.1	9059	32202	288.35	100	97.36
Gisagara	20,215,250	18,822,054	35257	1	OK423773	KX753357.1	8743	47008	407.65	100	97.42
Ngoma	25,708,336	24,127,742	38821	1	OK423778	KX753357.1	9053	33376	284.3	100	97.36

Table S 6. Analysis of CBSV analysis among 13 districts.
The CBSV analysis by three approaches (Blast, Mapping, and Kraken) revealed that CBSV was detected in 4 districts: Ruhango, Muhanga, Bugesera, and Gatsibo.

Districts	Mapping	Blast	Kraken	percentage of ipomovirus sequence reads; a number of CBSV reads	CBSV presence
Gisagara	_	-	_	0.004%; 1 read	_
Nyanza	_	-	_	0.002%; 1 read	_
Ruhango	+	-	+	0.02; 9 reads	+
Muhanga	+	-	+	0.07%; 13 reads	+
Kamonyi	_	-	_	No read of CBSV	_
Bugesera	+	-	+	0.2%; 25 reads	+
Kayonza	_	-	_	No read of CBSV	_
Ngoma	_	-	_	No read of CBSV	_
Kirehe	_	-	_	No read of CBSV	_
Gatsibo	+	-	+	0.1%; 14 reads	+
Nyagatare	_	-	_	No read of CBSV	_
Gakenke	_	-	_	No read of CBSV	_
Nyamasheke	-	-	-	No read of CBSV	-

Table S 7. Percentages of identity at nucleotide level between UCBSV consensus sequences from 12 districts of Rwanda.

	Kirehe	Bugesera	Gatsibo	Ngoma	Nyagatare	Kayonza	Gisagara	Nyamasheke	Ruhango	Kamonyi	Nyanza	Muhanga
Kirehe												
Bugesera	100											
Gatsibo	100	100										
Ngoma	100	100	100									
Nyagatare	100	100	100	100								
Kayonza	100	100	100	100	100							
Gisagara	100	100	100	100	100	100						
Nyamasheke	99.4	99.4	99.4	99.4	99.4	99.4	99.4					
Ruhango	99.5	99.5	99.5	99.5	99.5	99.5	99.4	99.5				
Kamonyi	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.4	99.9			
Nyanza	99.3	99.3	99.3	99.3	99.3	99.3	99.3	99.4	99.8	99.9		
Muhanga	99.3	99.3	99.3	99.3	99.3	99.3	99.2	99.4	99.8	99.9	100	

Table S 8. Percentages of identity between the amino acid sequence of the UCBSV consensus polyprotein from 12 districts of Rwanda.

Kirehe											
Bugesera	100										
Gatsibo	100	100									
Ngoma	100	100	100								
Nyagatare	100	100	100	100							
Kayonza	100	100	100	100	100						
Gisagara	99.9	99.9	99.9	99.9	99.9	99.9					
Nyamasheke	99.5	99.5	99.5	99.5	99.5	99.5	99.5				
Ruhango	99.7	99.7	99.7	99.7	99.7	99.7	99.6	99.8			
Nyanza	99.6	99.6	99.6	99.6	99.6	99.6	99.5	99.6	99.9		
Kamonyi	99.6	99.6	99.6	99.6	99.6	99.6	99.5	99.6	99.9	100	
Muhanga	99.5	99.5	99.5	99.5	99.5	99.5	99.4	99.6	99.8	100	100

Table S 9. A complete list of mutations contributed to the discrimination of UCBSV populations. Most mutations were silent, but few resulted in amino acid change.

Table S9 is also available at https://doi.org/10.5281/zenodo.8222816.

				H2	H1	H1	H1	H1	H1	H1	H1	H1 + H3	H1 + H2	H1 + H2	H1 + H2	
SNP	aa Change	CDS	Polymorp-hism Type	Muhanga	Nyagatare	Gisagara	Kayonza	Kirehe	Ngoma	Bugesera	Gatsibo	Nyamasheke	Ruhango	Kamonyi	Nyanza	Intrpretation
T_992_C		P1 protein	SNP	0.999	0.998	0.999	0.998	0.872	0.974	0.998	1	0.999	1	1	1	SNP present in all haplotypes and almost all districts
T_6657_C		NIb protein	SNP	0.996	0.996	0.998	1	0.996	1	0.996	1	0.998	0.996	0.8	1	
GG_294_AA	G -> K	P1 protein	SNP	0.993	0	0.995	0.989	0.992	0.991	0.996	0	0.999	0.998	0.988	0.993	
G_8636_A		Coat protein	SNP	0.997	1	0.998	1	1	0.997	0.995	0.998	0.998	0.812	0.996	0.97	
G_6355_A	R -> K	NIb protein	SNP	1	0.991	0.998	1	1	0.991	0.999	0.999	0.999	0.737	1	0.944	
C_4385_T		Vpg protein	SNP	0.726	0.995	1	1	1	0.999	0.999	0.999	1	0.885	1	0.858	
A_4706_G		Vpg protein	SNP	1	0.975	0.999	1	1	0.999	0.996	0.999	0.31	0.999	0.995	0.997	SNP absent from H3 but present in all others haplotypes
G_551_A		P1 protein	SNP	0.994	0.994	0.997	1	0.997	0.973	0.975	0.999	0.329	0.766	0.998	0.982	
C_3668_T		CI protein	SNP	1	0.995	1	1	0.994	0.976	0.992	0.997	0.313	0.772	1	0.965	
T_1103_C		P1 protein	SNP	1	0	0	0	0	0	0.01	0.02	0.651	0.709	0.654	0.749	SNP present in H2 and H3
C_3050_T		CI protein	SNP	1	0	0	0	0	0	0.015	0	0.646	0.666	0.642	0.749	
T_8459_C		Coat protein	SNP	1	0	0	0	0	0	0	0.012	0	0.548	0.735	0.742	SNP present in H2
T_8043_G	S -> A	Coat protein	SNP	0.995	0	0	0	0	0	0	0.012	0	0.539	0.785	0.723	
A_8172_G	I -> V	Coat protein	SNP	0.96	0	0	0	0	0	0	0.011	0	0.472	0.602	0.72	
T_5894_C		NIb protein	SNP	0.979	0	0	0	0	0	0	0	0	0.572	0.701	0.712	
C_7664_T		HAM1 protein	SNP	0.997	0.018	0	0	0	0	0	0.022	0	0.541	0.689	0.698	
G_4046_A		CI protein	SNP	0.999	0	0	0	0	0	0	0	0	0.544	0.698	0.697	
C_6731_T		NIb protein	SNP	0.998	0	0	0	0	0	0	0	0	0.519	0.634	0.689	
A_670_G	N -> S	P1 protein	SNP	0.999	0	0	0	0	0	0	0.01	0	0.458	0.663	0.672	
T_5825_C		NIb protein	SNP	0.959	0	0	0	0	0	0	0	0	0.508	0.525	0.672	
G_3131_T		CI protein	SNP	0.997	0	0	0	0	0	0	0	0	0.492	0.642	0.67	
G_5115_A	V -> I	NIa pro	SNP	0.995	0	0	0	0	0	0	0.012	0	0.491	0.605	0.663	
C_452_T		P1 protein	SNP	0.989	0.051	0	0	0	0.011	0.014	0.01	0	0.496	0.677	0.662	
G_2153_A		6K1 protein	SNP	0.981	0	0	0	0	0	0	0.026	0	0.522	0.648	0.658	
T_1184_C		P1 protein	SNP	0.99	0	0	0	0	0	0	0.011	0	0.434	0.599	0.657	
A_3380_G		CI protein	SNP	0.975	0	0	0	0	0	0	0.018	0	0.525	0.684	0.655	
C_1436_T		P3 protein	SNP	0.991	0	0	0	0	0	0	0.028	0	0.473	0.647	0.653	
G_7487_A		HAM1 protein	SNP	0.986	0	0	0	0	0	0	0	0	0.539	0.646	0.646	
C_2918_T		CI protein	SNP	0.976	0	0	0	0	0	0	0	0	0.527	0.654	0.643	
A_6445_G	Y -> C	NIb protein	SNP	0.998	0	0	0	0	0	0	0	0	0.445	0.657	0.636	
A_6431_G		NIb protein	SNP	0.949	0	0	0	0	0	0	0.01	0	0.439	0.469	0.633	
C_4412_A		Vpg protein	SNP	0.822	0	0	0	0	0	0	0	0	0.456	0.356	0.631	
G_8555_A		Coat protein	SNP	0.991	0	0	0	0	0	0	0.014	0	0.557	0.744	0.626	
C_5157_T		NIa pro	SNP	0.993	0	0	0	0	0	0	0.013	0	0.525	0.629	0.625	
C_7124_T		HAM1 protein	SNP	0.953	0	0	0	0	0	0	0	0	0.442	0.568	0.6	
C_2711_T		CI protein	SNP	0.849	0	0	0	0	0	0	0	0	0.445	0.366	0.585	
G_5558_A		NIb protein	SNP	0.835	0	0	0	0	0	0	0	0	0.45	0.364	0.574	
G_458_A		P1 protein	SNP	0.858	0	0.018	0	0	0	0	0	0	0.559	0.367	0.558	
G_7206_A	A -> T	HAM1 protein	SNP	0.835	0	0	0	0	0	0	0	0	0.404	0.348	0.491	
A_1775_G		P3 protein	SNP	0.369	0	0	0	0	0	0	0	0	0.177	0	0.282	

T_8306_C		Coat protein	SNP	0	0.987	1	1	0.996	0.997	1	0.991	0.308	0.559	0.331	0.43	
T_7799_A	D -> E	Coat protein	SNP	0	0.994	1	1	0.994	0.993	0.985	0.981	0.316	0.246	0.262	0.392	
C_7772_T		Coat protein	SNP	0	1	0.997	1	0.997	0.997	0.984	0.981	0.289	0.22	0.274	0.364	
G_7757_T	E -> D	Coat protein	SNP	0	0.988	0.997	1	0.994	0.974	0.978	0.976	0.28	0.211	0.263	0.338	
G_5204_A		NIa pro	SNP	0	0.964	0.995	0.969	0.992	0.975	0.977	0.967	0.293	0.252	0.293	0.314	
A_7228_G	N -> S	HAM1 protein	SNP	0	0.996	0.999	0.996	0.998	0.992	0.986	0.992	0.355	0.296	0.408	0.309	
G_5948_A		NIb protein	SNP	0.056	0.915	0.957	0.896	0.89	0.932	0.943	0.863	0.303	0.293	0.349	0.306	
T_4982_C		NIa pro	SNP	0	0.981	0.999	1	0.995	0.983	0.937	0.99	0.34	0.288	0.341	0.299	
G_5838_A	E -> K	NIb protein	SNP	0	0.995	0.996	1	0.994	0.977	0.986	0.984	0.333	0.265	0.325	0.299	
C_5819_T		NIb protein	SNP	0	1	1	1	0.998	0.986	0.99	0.982	0.333	0.274	0.335	0.296	
T_2669_C		CI protein	SNP	0	1	1	1	0.996	0.991	0.987	0.991	0.364	0.295	0.347	0.289	
C_2762_T		CI protein	SNP	0	0.995	0.997	1	0.995	0.974	0.984	0.991	0.347	0.273	0.33	0.288	
T_1158_C	Y -> H	P1 protein	SNP	0	1	0.998	0.996	0.994	0.989	0.989	0.983	0.336	0.31	0.415	0.286	
C_4883_T		NIa pro	SNP	0	1	0.999	1	0.995	0.99	0.985	0.987	0.34	0.292	0.305	0.286	
T_5693_G		NIb protein	SNP	0	0.996	0.995	0.994	0.995	0.967	0.986	0.992	0.308	0.318	0.377	0.286	SNP present in H1
A_515_G		P1 protein	SNP	0	0.991	0.999	0.995	0.996	0.972	0.979	0.989	0.33	0.284	0.337	0.284	
A_1164_G	I -> V	P1 protein	SNP	0	0.991	0.997	0.996	0.994	0.988	0.986	0.987	0.337	0.309	0.403	0.284	
A_3104_G		CI protein	SNP	0	0.995	0.999	1	0.992	0.994	0.985	0.995	0.352	0.313	0.357	0.283	
T_7529_C		HAM1 protein	SNP	0	0.995	0.997	1	0.993	0.993	0.995	0.988	0.357	0.3	0.303	0.281	
A_5453_G		NIa pro	SNP	0	0.983	0.998	0.996	0.996	0.983	0.984	0.992	0.321	0.293	0.345	0.275	
T_6683_C		NIb protein	SNP	0	0.996	1	0.989	1	0.993	0.991	0.989	0.393	0.251	0.387	0.265	
T_2123_C		6K1 protein	SNP	0	0.997	0.998	0.99	0.994	0.978	0.988	0.973	0.35	0.293	0.323	0.262	
A_4547_G		Vpg protein	SNP	0	0.993	1	0.995	0.99	0.984	0.982	0.982	0.338	0.292	0.359	0.259	
C_4807_T	A -> V	Vpg protein	SNP	0	0.992	0.986	0.987	0.988	0.986	0.988	0.983	0.295	0.276	0.34	0.259	
G_7448_A		HAM1 protein	SNP	0	0.982	0.991	1	0.998	0.992	0.989	0.981	0.37	0.256	0.365	0.258	
G_2639_A		CI protein	SNP	0	1	1	1	0.994	0.989	0.983	0.988	0.352	0.289	0.334	0.253	
A_3911_G		CI protein	SNP	0	0.994	0.997	0.995	0.998	0.974	0.991	0.978	0.35	0.262	0.338	0.248	
T_1103_G	D -> E	P1 protein	SNP	0	0.992	0.998	0.998	0.992	0.986	0.984	0.976	0.348	0.289	0.345	0.245	
T_8606_C		Coat protein	SNP	0	0.989	1	1	0.995	0.986	0.996	0.977	0.416	0.327	0.286	0.242	
C_8012_T		Coat protein	SNP	0	1	1	1	0.997	0.987	0.981	0.985	0.336	0.311	0.22	0.209	
C_3155_T		CI protein	SNP	0	1	1	1	1	0.997	0.999	0.991	0.999	0.507	0.327	0.318	
T_8225_G		Coat protein	SNP	0.031	1	0.994	1	1	0.998	0.996	0.988	0.998	0.542	0.281	0.318	SNP present in H1 and H3
C_2369_T		CI protein	SNP	0	0.997	1	0.998	0.992	0.99	0.985	0.991	1	0.44	0.315	0.262	
G_3773_A		CI protein	SNP	0	0.994	1	0.997	0.993	0.985	0.991	0.982	0.998	0.473	0.31	0.261	
C_1799_T		P3 protein	SNP	0	0	0.184	0	0	0	0	0	0	0.123	0	0.179	
T_6311_C		NIb protein	SNP	0.101	0	0	0	0	0	0	0	0	0.059	0	0.123	
T_989_C		P1 protein	SNP	0.169	0	0	0	0	0	0	0	0	0.084	0.328	0.112	
A_5324_G		NIa pro	SNP	0.11	0	0	0	0	0	0	0	0	0.088	0.099	0.109	
A_7296_G	K -> E	HAM1 protein	SNP	0.108	0	0	0	0	0	0	0	0	0.041	0	0.094	
C_194_T		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.194	0	0.083	
C_1195_T	T -> I	P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.262	0	0.074	
T_3203_C		CI protein	SNP	0.054	0	0	0	0	0	0	0	0	0.036	0	0.062	
GA_6355_AG	R -> K	NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.256	0	0.056	
A_6925_G	E -> G	NIb protein	SNP	0.031	0	0	0.011	0	0	0	0	0	0	0	0.055	
T_1349_C		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0.238	0	0.054	

C_785_T		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.169	0	0.048
A_6572_G		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.193	0	0.046
G_2987_A		CI protein	SNP	0	0	0.044	0	0	0	0	0	0	0.042	0	0.045
C_1401_T		P3 protein	SNP	0.036	0	0	0	0	0	0	0	0	0.013	0	0.044
_1624_A		P3 protein	on (tandem r	0.046	0.073	0.062	0.078	0.051	0.039	0.061	0.05	0.061	0.057	0.055	0.043
A_8551_C	Y -> S	Coat protein	SNP	0.012	0.034	0	0.022	0.026	0.015	0.029	0.019	0.015	0.011	0.02	0.042
C_3089_T		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.201	0	0.041
G_6971_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.177	0	0.039
C_3761_T		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.188	0	0.038
C_6998_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.161	0	0.038
AG_7065_CAT	K -> H	NIb protein	SNP	0.012	0	0.027	0.011	0	0.029	0.018	0.035	0.025	0	0	0.038
T_692_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.194	0	0.036
C_1886_T		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0.208	0	0.036
T_6323_C		NIb protein	SNP	0	0	0.059	0	0	0	0	0	0	0	0	0.035
T_7396_C	L -> S	HAM1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.035
G_125_A		5' UTR	SNP	0	0	0	0	0	0	0	0	0	0	0	0.034
A_2630_G		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.193	0	0.034
C_6461_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.227	0	0.034
A_272_C	Q -> H	P1 protein	SNP	0	0.018	0	0.027	0.023	0.013	0.01	0.013	0.017	0.012	0	0.033
A_7101_C	I -> L	HAM1 protein	SNP	0.021	0.029	0.011	0.02	0.014	0.017	0	0.013	0.014	0	0	0.033
T_326_C		P1 protein	SNP	0	0	0	0	0	0.013	0	0	0	0	0	0.032
_5955_A		NIb protein	Insertion	0.043	0.074	0.029	0.01	0.044	0.049	0.046	0.058	0.054	0.064	0.062	0.032
A_461_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0.012	0	0	0.031
A_1730_G		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0.201	0	0.031
G_3134_A		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.21	0	0.031
G_3416_A		CI protein	SNP	0.031	0	0	0	0	0	0	0	0	0.016	0	0.031
A_7713_C	K -> Q	HAM1 protein	SNP	0	0.011	0	0	0	0	0	0	0	0	0.013	0.031
G_554_A		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.228	0	0.03
A_7868_G		Coat protein	SNP	0.012	0	0	0	0	0	0	0	0	0	0	0.029
G_7984_A	R -> K	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0.183	0	0.029
T_509_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.028
C_1889_T		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.028
A_4602_C	M -> L	Vpg protein	SNP	0.022	0.046	0.019	0.03	0.014	0.016	0.013	0.014	0.017	0.021	0.021	0.028
G_7367_A		HAM1 protein	SNP	0	0.014	0	0	0	0	0	0	0	0.012	0	0.028
T_3648_C		CI protein	SNP	0.01	0.013	0	0	0.018	0.017	0	0.01	0.016	0	0	0.027
T_4241_C		6K2 protein	SNP	0	0	0	0	0	0	0	0	0	0.01	0	0.027
A_8513_G		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.027
A_8514_G	I -> V	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.027
T_659_C		P1 protein	SNP	0.011	0.017	0	0.014	0	0	0	0.011	0	0	0	0.026
T_3139_G	I -> S	CI protein	SNP	0.018	0.027	0.012	0.012	0.013	0	0.013	0	0.016	0.013	0.013	0.026
A_4071_C	I -> L	CI protein	SNP	0.033	0.043	0.026	0.034	0.028	0.021	0.026	0.018	0.037	0.022	0.016	0.026
T_6674_C		NIb protein	SNP	0.027	0	0	0.01	0.013	0.019	0.02	0.03	0.019	0.02	0	0.026
A_8723_G		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.026
G_1970_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.025
A_3309_C	M -> L	CI protein	SNP	0.016	0.032	0.018	0.02	0.01	0	0.013	0.014	0.016	0.011	0.015	0.025
A_5094_G	I -> V	NIa pro	SNP	0.029	0	0	0	0	0	0	0	0	0.016	0	0.025

A_5425_C	N -> T	NIa pro	SNP	0.015	0.026	0.021	0	0.021	0.012	0.018	0.018	0.015	0.021	0	0.025	
T_8534_C		Coat protein	SNP	0	0	0	0	0	0.012	0	0	0	0	0	0.025	
T_188_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.024	
A_5570_C	E -> D	NIb protein	SNP	0	0.018	0	0	0	0	0.011	0	0	0	0.014	0.024	
G_6384_A	A -> T	NIb protein	SNP	0	0	0	0.022	0	0	0	0	0	0	0	0.024	
A_4684_G	K -> R	Vpg protein	SNP	0	0	0	0	0	0	0	0	0	0.212	0	0.023	
A_7973_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.023	
T_8167_G	I -> S	Coat protein	SNP	0.015	0.033	0	0.011	0.024	0.018	0	0	0	0.016	0	0.023	
C_8612_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.023	
T_176_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.022	
A_455_G		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.185	0	0.022	
A_2019_C	K -> Q	P3 protein	SNP	0.018	0.029	0	0.024	0	0.021	0	0	0	0.015	0.019	0.022	
T_5706_C		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.022	
A_5755_C	K -> T	NIb protein	SNP	0	0.017	0.014	0	0	0	0	0	0	0	0.011	0.022	
C_7976_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0.214	0.022	
A_7978_G	N -> S	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.022	
G_8054_A		Coat protein	SNP	0	0	0	0	0	0	0.019	0	0	0	0	0.022	
T_8453_C		Coat protein	SNP	0.019	0	0	0	0	0	0	0	0	0.02	0	0.022	
C_185_T		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
T_1172_C		P1 protein	SNP	0.019	0	0	0	0	0	0	0	0	0	0	0.021	
A_1901_G		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
G_1907_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
T_1919_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
T_5714_C		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
G_6438_T	D -> Y	NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
A_6786_C	I -> L	NIb protein	SNP	0.012	0.038	0.011	0.014	0.013	0	0	0	0.012	0	0.015	0.021	
A_7020_C	I -> L	NIb protein	SNP	0.015	0.011	0.013	0.013	0.014	0	0.013	0.012	0.014	0.021	0.027	0.021	
G_7991_A		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
G_8011_A	S -> N	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
G_8019_A	D -> N	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
G_8042_A		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
T_8043_C	S -> P	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
A_8440_G	Q -> R	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.021	
A_8663_G		Coat protein	SNP	0	0	0.012	0	0	0	0	0	0	0	0	0.021	
G_1598_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0.191	0	0.02	
G_1976_T		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.02	
C_4316_T		Vpg protein	SNP	0	0	0.019	0	0	0.01	0	0	0	0	0	0.02	
A_4364_G		Vpg protein	SNP	0	0	0	0	0	0	0	0.176	0	0	0	0.02	
G_5201_A		NIa pro	SNP	0	0.013	0	0	0	0	0	0	0	0	0	0.02	
A_5342_T		NIa pro	SNP	0	0	0	0	0	0	0	0	0	0	0	0.02	
G_5730_A	V -> I	NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.02	
A_6863_C	E -> D	NIb protein	SNP	0	0.012	0	0	0	0	0	0	0	0	0	0.02	
A_7885_C	K -> T	Coat protein	SNP	0	0	0	0.02	0	0	0.01	0	0	0	0	0.02	
G_8001_A	V -> I	Coat protein	SNP	0	0	0	0	0	0	0	0.247	0	0	0	0.02	
A_8022_C	K -> Q	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.02	
C_8430_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0	0.02	

A_1546_C	N -> T	P3 protein	SNP	0.024	0.033	0.012	0.017	0.02	0.015	0.025	0.018	0.016	0.026	0.028	0.019
T_7070_G	D -> E	NIb protein	SNP	0	0.033	0	0.011	0	0	0.016	0.012	0.014	0	0.02	0.019
G_8138_A		Coat protein	SNP	0.132	0	0	0	0	0	0	0	0	0.051	0	0.018
A_8512_C	K -> T	Coat protein	SNP	0	0.011	0	0.04	0	0.011	0	0.017	0	0.012	0	0.018
A_162_C	I -> L	P1 protein	SNP	0.025	0.094	0.028	0.035	0.024	0.043	0.022	0.032	0.01	0.023	0.032	0.017
A_2438_C	E -> D	CI protein	SNP	0.023	0.032	0.015	0.024	0.018	0.025	0.017	0.015	0.025	0.027	0.032	0.017
A_2933_C	E -> D	CI protein	SNP	0.025	0.038	0.019	0.026	0.028	0.021	0.02	0.02	0.016	0.024	0.014	0.017
A_4205_C		6K2 protein	SNP	0.015	0.038	0.018	0.019	0	0.034	0	0.015	0.02	0.017	0.015	0.017
A_4493_C		Vpg protein	SNP	0.013	0.021	0.013	0	0.011	0.018	0	0.018	0	0.015	0.018	0.017
A_5661_T	T -> S	NIb protein	SNP	0.031	0.047	0.015	0	0.025	0.02	0	0.019	0.019	0.019	0	0.017
A_5661_G	T -> A	NIb protein	SNP	0	0	0	0.036	0	0	0	0	0	0	0	0.017
A_6668_C	E -> D	NIb protein	SNP	0.011	0.032	0.011	0.024	0.014	0	0	0.012	0.012	0.017	0.02	0.017
A_8505_C	M -> L	Coat protein	SNP	0.016	0.017	0.012	0.016	0	0.021	0.01	0	0.012	0.018	0.012	0.017
A_8600_G		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0.175	0	0.017
C_425_T		P1 protein	SNP	0	0	0	0	0	0	0.048	0.05	0	0	0	0.016
A_470_C		P1 protein	SNP	0.024	0.036	0.024	0.01	0.067	0.024	0.028	0.026	0.023	0.013	0.024	0.016
A_4714_G	N -> S	Vpg protein	SNP	0	0	0.027	0	0	0	0	0	0	0	0	0.016
T_6986_C		NIb protein	SNP	0	0.013	0	0.022	0	0	0	0	0	0	0	0.016
A_7946_C	E -> D	Coat protein	SNP	0	0.018	0	0.02	0	0	0	0	0	0	0.015	0.016
A_8557_C	N -> T	Coat protein	SNP	0	0.059	0.017	0.035	0.028	0.034	0.022	0.029	0.029	0	0	0.016
T_566_G	D -> E	P1 protein	SNP	0.011	0.019	0.021	0.019	0	0.013	0.014	0	0.023	0.012	0.011	0.015
A_648_C	I -> L	P1 protein	SNP	0	0.017	0	0.012	0.011	0.011	0.011	0.012	0.021	0.012	0	0.015
C_823_T	A -> V	P1 protein	SNP	0	0	0.778	0.375	0	0	0	0	0	0.013	0	0.015
A_1818_C	I -> L	P3 protein	SNP	0.045	0.064	0.044	0.042	0.031	0.036	0.026	0.032	0.043	0.038	0.031	0.015
A_5800_C	N -> T	NIb protein	SNP	0	0.021	0	0.02	0	0.02	0	0	0	0	0	0.015
C_7777_T	A -> V	Coat protein	SNP	0	0.017	0	0	0	0	0	0.011	0.709	0	0.01	0.015
A_906_C	I -> L	P1 protein	SNP	0.022	0.028	0.012	0	0	0.011	0.012	0	0	0.016	0.018	0.014
T_4190_C		6K2 protein	SNP	0.036	0.062	0	0	0	0	0	0	0	0	0	0.014
T_5585_C		NIb protein	SNP	0.011	0.022	0	0	0	0	0	0.014	0.021	0	0.011	0.014
A_8250_C	I -> L	Coat protein	SNP	0.014	0.024	0.016	0.017	0	0.011	0.013	0.014	0.013	0.012	0	0.014
A_4023_G	I -> V	CI protein	SNP	0	0	0	0	0	0	0	0	0	0.199	0	0.013
A_5610_C	N -> H	NIb protein	SNP	0.017	0.024	0.024	0	0.021	0.018	0.016	0.017	0.012	0.014	0.032	0.013
T_6674_G		NIb protein	SNP	0	0	0	0.01	0	0.012	0	0	0	0	0.029	0.013
A_1722_C	I -> L	P3 protein	SNP	0.018	0.026	0.01	0	0.023	0.011	0.012	0.018	0.013	0.011	0.01	0.012
G_2278_A	R -> K	CI protein	SNP	0	0	0	0	0	0	0	0.462	0	0	0	0.012
T_3197_G	D -> E	CI protein	SNP	0	0.02	0	0	0	0	0.012	0	0	0	0	0.012
A_4936_C	N -> T	NIa pro	SNP	0	0.023	0.011	0.012	0	0	0	0	0.014	0.01	0	0.012
G_5054_A		NIa pro	SNP	0	0	0.027	0	0	0	0	0	0	0.012	0	0.012
A_5678_C	Q -> H	NIb protein	SNP	0	0.037	0	0.012	0	0	0.012	0	0.012	0	0	0.012
A_5686_C	N -> T	NIb protein	SNP	0.021	0	0	0.011	0	0	0	0	0	0	0	0.012
A_6274_C	N -> T	NIb protein	SNP	0.019	0.046	0.02	0.019	0.024	0.042	0	0.021	0.015	0.02	0.022	0.012
T_3041_C		CI protein	SNP	0	0.053	0	0	0	0	0	0	0	0	0	0.011
A_3072_C		CI protein	SNP	0.013	0.023	0	0.019	0	0.023	0	0.014	0.019	0.011	0.014	0.011
A_1065_C	I -> L	P1 protein	SNP	0	0.022	0	0	0	0.01	0	0	0	0	0	0.01
A_3946_C	N -> T	CI protein	SNP	0.027	0.043	0.018	0.032	0	0.02	0.017	0.019	0.026	0.019	0.012	0.01
A_5322_C	T -> P	NIa pro	SNP	0	0.029	0.012	0.021	0	0	0.01	0.016	0.012	0	0	0.01

A_6795_C	I -> L	NIb protein	SNP	0.013	0.038	0.013	0.014	0	0.011	0	0	0.015	0	0	0.01
A_6904_C	Y -> S	NIb protein	SNP	0	0.027	0	0	0	0	0	0.012	0.011	0	0	0.01
G_39_A		5' UTR	SNP	0	0	0	0	1	0	1	0.981	0	0.207	0	0
T_40_C		5' UTR	SNP	0	0	0	0	1	0	1	1	0	1	0	0
A_47_T		5' UTR	SNP	0	0	0.016	0	0	0	0	0.027	0	0	0	0
T_50_C		5' UTR	SNP	0	0	1	0	1	0	1	0.987	0.411	0.214	0	0
G_66_T		5' UTR	SNP	0	0.033	0.011	0	0	0.024	0	0	0	0	0	0
G_68_T		5' UTR	SNP	0	0.023	0.015	0	0	0	0	0	0	0	0	0
A_77_G		5' UTR	SNP	0	0	0	0	0	0	0	0	0	0.04	0	0
G_81_A		5' UTR	SNP	0	0	0	0	0	0	0	0	0.082	0	0	0
T_82_C		5' UTR	SNP	0	0	0	0	0.138	0	0	0	0	0	0	0
G_86_A		5' UTR	SNP	0	0.011	0	0	0	0	0	0	0.048	0	0	0
A_96_G		5' UTR	SNP	0.996	0.986	0.997	1	1	1	0.992	1	1	1	1	0
A_109_C		5' UTR	SNP	0	0	0.011	0	0	0	0	0	0	0.021	0.016	0
A_120_T		5' UTR	SNP	0	0.025	0	0	0	0	0.015	0.011	0	0	0	0
A_152_C		P1 protein	SNP	0	0.027	0	0.01	0.012	0	0	0	0	0	0.032	0
G_186_A	V -> I	P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.085	0
G_202_A	G -> E	P1 protein	SNP	0	0	0	0	0	0	0	0	0.612	0	0	0
G_240_A	A -> T	P1 protein	SNP	0	0.293	0	0	0	0	0	0	0	0	0	0
A_282_G	K -> E	P1 protein	SNP	0	0	0	0	0	0	0	0	0.118	0	0	0
A_284_G		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0.049	0	0
G_306_A	E -> K	P1 protein	SNP	0	0	0	0	0	0	0.105	0	0	0	0	0
G_316_A	R -> Q	P1 protein	SNP	0	0	0	0	0	0	0	0	0.032	0	0	0
T_322_C	V -> A	P1 protein	SNP	0	0.024	0	0	0	0	0	0	0	0	0	0
G_450_A	A -> T	P1 protein	SNP	0	0.025	0	0	0	0	0	0	0	0	0	0
A_465_G	R -> G	P1 protein	SNP	0.015	0.012	0	0	0	0.025	0	0	0.01	0	0	0
G_543_A	A -> T	P1 protein	SNP	0	0	0	0	0	0.023	0	0	0	0	0	0
T_557_C		P1 protein	SNP	0	0.043	0	0	0	0	0	0	0	0	0	0
G_590_T	K -> N	P1 protein	SNP	0	0.026	0	0.01	0	0	0	0	0	0	0.011	0
T_611_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0.635	0	0	0
A_669_G	N -> D	P1 protein	SNP	0	0	0	0	0	0	0	0.028	0	0	0	0
G_713_A		P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.025	0
T_731_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0.302	0	0	0
G_800_A		P1 protein	SNP	0	0	0	0	0.115	0.031	0.015	0	0	0	0	0
C_802_T	T -> I	P1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.079	0
T_818_C		P1 protein	SNP	0	0.041	0	0	0	0	0	0	0	0	0	0
G_846_A	E -> K	P1 protein	SNP	0	0	0	0	0	0	0	0	0.627	0	0	0
A_867_C	I -> L	P1 protein	SNP	0.014	0.012	0	0	0.011	0	0.013	0.012	0.022	0	0	0
T_881_C		P1 protein	SNP	0	0	0	0	0	0	0	0	0.641	0	0	0
C_904_T	A -> V	P1 protein	SNP	0	0.063	0	0	0	0	0	0	0	0	0	0
C_1028_A		P1 protein	SNP	0	0	0	0	0	0	0	0.059	0	0	0	0
T_1032_C	Y -> H	P1 protein	SNP	0	0	0	0	0	0	0	0	0.036	0	0	0
T_1058_C		P1 protein	SNP	0.013	0	0.014	0	0.024	0.011	0.027	0.022	0.018	0.018	0.014	0
G_1182_A	V -> I	P1 protein	SNP	0	0.023	0	0	0	0	0	0	0	0	0	0
G_1202_A	M -> I	P1 protein	SNP	0	0	0	0	0	0	0	0.029	0	0	0	0
A_1254_G	K -> E	P3 protein	SNP	0	0	0	0.026	0	0.013	0	0	0	0	0	0
A_1339_C	K -> T	P3 protein	SNP	0.013	0.02	0.015	0.014	0.019	0.017	0	0	0	0	0	0

C_1343_T		P3 protein	SNP	0	0	0	0	0.025	0	0	0	0	0	0	0	
G_1379_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0.625	0	0	0	
T_1382_C		P3 protein	SNP	0.019	0.02	0.033	0.017	0.017	0.022	0.015	0.016	0.014	0.02	0.012	0	
G_1388_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0.076	0	0	0	
T_1397_C		P3 protein	SNP	0	0	0	0	0	0	0	0	0	0.022	0	0	
G_1423_A	G -> D	P3 protein	SNP	0	0.02	0	0	0	0	0	0	0	0	0	0	
T_1457_C		P3 protein	SNP	0	0.051	0	0	0	0	0	0	0	0	0	0	
G_1628_A		P3 protein	SNP	0	0.199	0	0	0	0	0	0	0	0	0	0	
G_1673_A		P3 protein	SNP	0	0	0	0	0	0	0	0	0.036	0	0	0	
A_1722_T	I -> L	P3 protein	SNP	0	0	0	0	0	0	0.031	0	0	0	0	0	
G_1776_A	E -> K	P3 protein	SNP	0	0	0	0	0	0	0	0	0.038	0	0	0	
T_2027_C		P3 protein	SNP	0	0.026	0	0	0	0	0	0	0	0	0	0	
T_2036_C		P3 protein	SNP	0	0	0	0	0	0	0.026	0	0	0	0	0	
T_2120_C		6K1 protein	SNP	0	0	0	0	0	0.025	0	0	0	0	0	0	
C_2143_T	A -> V	6K1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.028	0	
A_2173_C	Y -> S	6K1 protein	SNP	0	0.015	0	0.019	0	0.023	0	0.01	0.015	0	0	0	
A_2195_T		6K1 protein	SNP	0	0	0	0	0	0	0	0	0.649	0	0	0	
A_2226_G	S -> G	6K1 protein	SNP	0	0.049	0	0	0	0	0	0	0	0	0	0	
T_2261_C		CI protein	SNP	0	0.015	0.014	0	0	0	0.02	0	0	0.011	0.012	0	
T_2267_C		CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.045	0	
C_2274_T	L -> F	CI protein	SNP	0	0	0	0	0	0	0	0.044	0	0	0	0	
C_2282_T		CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.041	0	
T_2306_C		CI protein	SNP	0	0	0.02	0.011	0	0.01	0	0	0	0	0	0	
C_2312_T		CI protein	SNP	0	0	0.764	0	0	0	0	0	0	0	0	0	
A_2330_G		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.033	0	0	
T_2504_C		CI protein	SNP	0.015	0	0	0	0.023	0	0.015	0.013	0.014	0	0.019	0	
A_2505_C	K -> Q	CI protein	SNP	0	0.02	0	0	0.01	0.011	0	0	0	0	0.025	0	
A_2585_G		CI protein	SNP	0.021	0	0	0	0.021	0	0	0	0	0	0	0	
C_2648_T		CI protein	SNP	0	0.059	0	0	0	0	0	0	0	0	0	0	
A_2689_C	Y -> S	CI protein	SNP	0	0.024	0	0.011	0.011	0.015	0	0.012	0.012	0.011	0.013	0	
A_2713_C	N -> T	CI protein	SNP	0	0.025	0	0	0	0.016	0	0	0	0	0	0	
A_2742_G	I -> V	CI protein	SNP	0	0.027	0	0	0	0	0	0	0	0	0	0	
G_2913_A	D -> N	CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.036	0	
T_3170_C		CI protein	SNP	0	0	0	0.072	0	0	0	0	0	0	0	0	
C_3179_T		CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.101	0	
T_3185_G	D -> E	CI protein	SNP	0	0.025	0	0.011	0	0	0	0	0	0.011	0	0	
T_3236_A	D -> E	CI protein	SNP	0	0	0	0	0	0	0	0	0.672	0	0	0	
A_3271_T	Q -> L	CI protein	SNP	0	0.02	0	0	0	0	0	0	0	0	0	0	
A_3272_G		CI protein	SNP	0	0.035	0	0	0	0	0	0	0	0	0	0	
G_3305_A		CI protein	SNP	0	0	0	0	0	0	0	0	0.084	0	0	0	
C_3320_T		CI protein	SNP	0	0	0	0	0	0	0.106	0	0	0	0	0	
G_3332_A		CI protein	SNP	0	0	0	0	0	0	0	0	0.652	0	0	0	
G_3374_A		CI protein	SNP	0	0.026	0	0	0	0	0	0	0	0	0	0	
G_3446_A		CI protein	SNP	0	0	0	0	0	0	0	0	0.047	0	0	0	
G_3498_T	A -> S	CI protein	SNP	0	0	0	0.02	0	0	0	0	0	0	0	0	
G_3563_A		CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.061	0	

A_3574_C	N -> T	CI protein	SNP	0	0.022	0	0.014	0	0	0	0	0	0	0.011	0	
G_3602_A		CI protein	SNP	0	0.267	0	0	0	0	0	0	0	0	0	0	
G_3669_A	D -> N	CI protein	SNP	0	0	0	0	0	0	0	0	0.031	0	0	0	
T_3678_G	S -> A	CI protein	SNP	0	0	0	0	0	0	0	0	0.67	0	0	0	
G_3696_A	V -> I	CI protein	SNP	0.045	0	0	0	0	0	0	0	0	0	0	0	
T_3734_C		CI protein	SNP	0	0.024	0	0	0	0	0	0	0	0	0	0	
GT_3741_ATC	V -> I	CI protein	SNP	0	0	0	0	0	0.053	0.034	0	0	0	0	0	
C_3761_A		CI protein	SNP	0	0	0	0	0	0	0	0	0.025	0	0	0	
G_3782_A		CI protein	SNP	0	0	0.07	0	0	0	0	0	0	0	0	0	
T_3788_C		CI protein	SNP	0.036	0	0	0	0	0	0	0	0	0	0	0	
A_3905_C	Q -> H	CI protein	SNP	0	0.034	0	0	0	0.011	0.013	0	0.015	0.012	0.011	0	
T_3929_C		CI protein	SNP	0	0	0	0	0	0	0	0	0	0.206	0	0	
T_3962_C		CI protein	SNP	0	0	0	0	0.029	0	0	0	0	0	0	0	
A_4007_C	E -> D	CI protein	SNP	0	0.021	0	0.018	0	0.012	0	0	0	0.011	0	0	
C_4016_T		CI protein	SNP	0	0	0	0	0	0	0	0	0	0	0.111	0	
C_4049_T		CI protein	SNP	0	0	0	0	0	0.375	0	0	0	0	0	0	
G_4166_T	E -> D	6K2 protein	SNP	0.034	0.011	0	0	0	0	0	0	0	0	0	0	
T_4196_C		6K2 protein	SNP	0	0	0	0	0.066	0	0	0	0	0	0	0	
A_4202_G		6K2 protein	SNP	0	0	0	0.108	0	0	0	0	0	0	0	0	
A_4215_C	M -> L	6K2 protein	SNP	0.017	0.03	0.013	0.019	0.019	0.023	0.017	0.026	0.016	0.013	0.035	0	
A_4328_G		Vpg protein	SNP	0	0	0	0	0.246	0	0	0	0	0	0	0	
G_4410_A	A -> T	Vpg protein	SNP	0	0	0	0	0	0	0	0	0.64	0	0	0	
A_4436_G		Vpg protein	SNP	0	0	0	0	0	0	0	0.075	0	0	0	0	
A_4485_C	M -> L	Vpg protein	SNP	0.012	0.022	0.01	0.02	0	0.016	0	0.015	0.015	0.012	0.013	0	
A_4493_T		Vpg protein	SNP	0	0	0	0	0	0	0	0	0.666	0	0	0	
C_4514_T		Vpg protein	SNP	0	0.025	0	0	0	0	0	0	0	0	0	0	
C_4631_T		Vpg protein	SNP	0	0	0	0	0	0	0	0	0.055	0	0	0	
T_4635_C		Vpg protein	SNP	0	0.2	0	0	0	0	0	0	0	0	0	0	
G_4639_A	R -> Q	Vpg protein	SNP	0	0	0	0	0	0	0	0	0	0	0.061	0	
G_4784_A		Vpg protein	SNP	0	0.023	0	0	0	0	0	0	0	0	0	0	
A_4800_T	T -> S	Vpg protein	SNP	0	0.023	0	0	0	0	0	0	0	0	0.017	0	
G_4812_A	V -> M	Vpg protein	SNP	0	0	0	0	0	0.012	0.032	0	0	0	0	0	
G_4820_A		Vpg protein	SNP	0	0.066	0	0	0	0	0	0	0	0	0	0	
A_4835_G		Vpg protein	SNP	0	0	0	0	0	0	0	0	0.026	0	0	0	
G_4853_A		Vpg protein	SNP	0	0	0	0	0	0	0	0	0.678	0	0	0	
T_4856_C		NIa pro	SNP	0.04	0	0	0	0	0	0	0	0	0	0	0	
C_4870_T	P -> L	NIa pro	SNP	0	0	0	0	0	0	0	0	0.022	0	0	0	
T_4877_C		NIa pro	SNP	0	0	0	0	0	0	0.047	0	0	0	0	0	
A_4931_C		NIa pro	SNP	0	0.026	0.015	0	0.014	0.015	0.013	0	0.012	0.011	0	0	
G_4934_A		NIa pro	SNP	0	0	0	0	0	0	0	0	0	0	0.021	0	
G_4992_A	D -> N	NIa pro	SNP	0	0	0	0	0	0	0	0.036	0	0	0	0	
C_5066_T		NIa pro	SNP	0	0	0	0	0	0.022	0.012	0	0	0	0	0	
G_5088_A	V -> I	NIa pro	SNP	0	0	0	0	0.021	0	0	0	0	0	0	0	
A_5182_C	Y -> S	NIa pro	SNP	0	0.038	0.023	0.027	0.012	0	0	0.015	0.01	0	0	0	
G_5210_A		NIa pro	SNP	0	0	0	0.113	0	0	0	0	0	0	0	0	
C_5231_T		NIa pro	SNP	0	0.032	0	0	0	0	0	0	0	0	0.02	0	

A_5345_G		NIa pro	SNP	0	0.016	0	0.033	0	0	0	0	0	0	0.011	0
T_5354_C		NIa pro	SNP	0	0.051	0	0	0	0	0	0	0	0	0	0
A_5384_G		NIa pro	SNP	0	0.064	0	0	0	0	0	0	0	0	0.026	0
T_5393_C		NIa pro	SNP	0	0	0	0	0	0.091	0	0	0	0	0	0
T_5402_C		NIa pro	SNP	0	0.016	0	0	0	0.02	0.017	0	0	0	0	0
A_5410_C	N -> T	NIa pro	SNP	0.012	0.019	0.022	0.019	0.027	0.021	0.024	0.016	0.03	0.019	0.028	0
A_5504_G		NIa pro	SNP	0	0	0	0	0	0	0	0	0.026	0	0	0
G_5555_A		NIa pro	SNP	0	0	0	0	0	0	0	0	0.692	0	0	0
A_5560_C	N -> T	NIb protein	SNP	0	0.026	0	0	0.013	0.011	0	0	0.015	0	0	0
A_5586_C	N -> H	NIb protein	SNP	0	0.032	0	0	0	0	0	0	0	0.011	0	0
G_5648_A		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0.053	0
A_5705_C	E -> D	NIb protein	SNP	0.013	0.011	0.011	0	0	0.011	0	0.014	0	0.013	0.021	0
C_5723_T		NIb protein	SNP	0	0	0	0	0	0	0	0.038	0	0	0	0
T_5748_G	F -> V	NIb protein	SNP	0.013	0.013	0.014	0	0	0	0.01	0.014	0.018	0.011	0.02	0
C_5810_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0	0.021	0
G_5855_A		NIb protein	SNP	0	0	0	0	0	0	0	0	0.669	0	0	0
A_5887_G	D -> G	NIb protein	SNP	0	0	0.016	0	0.02	0	0	0	0	0	0	0
A_5892_C	I -> L	NIb protein	SNP	0	0.037	0.016	0.051	0.015	0.023	0.011	0	0.02	0.01	0	0
C_5910_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0	0.029	0	0
A_5949_C	K -> Q	NIb protein	SNP	0	0.016	0	0.021	0	0	0	0	0	0	0	0
T_5987_A	D -> E	NIb protein	SNP	0	0	0	0	0.077	0	0	0	0	0	0	0
A_6042_C	I -> L	NIb protein	SNP	0	0.024	0.013	0	0	0	0	0	0.012	0	0	0
A_6056_G		NIb protein	SNP	0	0.246	0	0	0	0	0	0	0	0	0	0
A_6146_G		NIb protein	SNP	0	0.018	0	0	0	0	0	0	0	0.021	0	0
T_6152_C		NIb protein	SNP	0	0.039	0	0	0	0	0	0	0	0	0	0
A_6216_C	I -> L	NIb protein	SNP	0	0.023	0.017	0	0.018	0	0.015	0	0	0.017	0	0
A_6245_G		NIb protein	SNP	0	0	0	0	0	0	0	0	0.065	0	0	0
T_6308_A		NIb protein	SNP	0	0.163	0	0	0	0	0	0	0	0	0	0
T_6317_C		NIb protein	SNP	0	0.021	0	0	0	0	0	0	0.013	0	0.017	0
T_6464_C		NIb protein	SNP	0	0	0	0	0	0	0	0	0.634	0	0	0
T_6590_C		NIb protein	SNP	0.011	0	0	0.014	0	0	0	0	0.011	0	0.024	0
T_6745_C	L -> P	NIb protein	SNP	0	0	0	0	0	0.433	0	0	0	0	0	0
CG_6771_TTA		NIb protein	SNP	0	0.993	1	1	0.989	0.998	0.997	1	1	0	1	0
A_6874_C	Y -> S	NIb protein	SNP	0.013	0.022	0.016	0.012	0.015	0.021	0	0	0	0.012	0.016	0
T_6890_C		NIb protein	SNP	0	0	0	0	0	0	0	0.033	0	0	0	0
A_6957_C	K -> Q	NIb protein	SNP	0.012	0.074	0.023	0.011	0.047	0.027	0.022	0.034	0.021	0.026	0.018	0
A_6992_T		NIb protein	SNP	0	0	0	0	0	0	0	0	0.083	0	0	0
G_7067_C	K -> N	NIb protein	SNP	0	0.024	0	0	0	0	0	0	0	0	0	0
A_7080_G	R -> G	HAM1 protein	SNP	0	0	0	0	0.033	0	0	0	0	0	0	0
A_7174_C	K -> T	HAM1 protein	SNP	0	0.021	0	0	0	0	0.011	0	0.01	0	0.017	0
T_7195_G	I -> S	HAM1 protein	SNP	0	0	0	0	0	0	0	0	0	0.027	0	0
A_7286_C	Q -> H	HAM1 protein	SNP	0.011	0.014	0.013	0.02	0	0	0	0	0.013	0	0.019	0
T_7338_G	C -> G	HAM1 protein	SNP	0	0.052	0	0	0	0	0	0	0	0	0	0
C_7352_T		HAM1 protein	SNP	0	0	0	0	0	0	0	0.015	0.638	0	0	0
C_7385_T		HAM1 protein	SNP	0	0	0	0	0	0	0	0	0.65	0	0	0
T_7386_C		HAM1 protein	SNP	0	0	0	0	0	0	0	0	0.096	0	0	0
T_7395_C		HAM1 protein	SNP	0.013	0	0	0	0	0	0.249	0	0	0	0	0

T_7400_A		HAM1 protein	SNP	0.04	0	0	0	0	0	0	0	0	0	0	0	
T_7400_C		HAM1 protein	SNP	0.015	0	0	0	0.013	0	0	0.012	0	0	0.022	0	
A_7445_G		HAM1 protein	SNP	0	0	0	0	0	0	0	0	0.643	0	0	0	
A_7457_C		HAM1 protein	SNP	0.02	0.011	0	0.024	0	0	0	0	0.012	0	0	0	
T_7460_C		HAM1 protein	SNP	0	0	0	0	0.013	0.06	0.021	0	0	0	0	0	
T_7465_C	F -> S	HAM1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.021	0	
A_7519_C	N -> T	HAM1 protein	SNP	0.013	0.022	0	0	0.015	0.019	0	0	0	0	0	0	
T_7544_A		HAM1 protein	SNP	0	0	0	0	0	0	0	0	0	0	0.156	0	
C_7637_T		HAM1 protein	SNP	0.02	0.124	0	0	0	0	0	0	0.011	0	0	0	
A_7686_G	M -> V	HAM1 protein	SNP	0	0	0	0	0	0.08	0	0	0	0	0	0	
G_7700_T	E -> D	HAM1 protein	SNP	0	0.021	0	0	0	0	0	0	0	0	0	0	
G_7715_A		HAM1 protein	SNP	0	0	0	0	0	0	0.01	0	0	0	0.03	0	
T_7721_C		HAM1 protein	SNP	0.011	0.022	0	0	0.01	0	0	0	0	0	0	0	
A_7790_C		Coat protein	SNP	0	0.022	0.014	0.03	0.012	0.01	0.014	0.016	0.022	0	0	0	
G_7866_T	V -> L	Coat protein	SNP	0	0.022	0	0	0	0	0	0	0	0	0	0	
C_7887_T		Coat protein	SNP	0	0.035	0	0	0	0	0	0	0	0	0	0	
G_7968_A	D -> N	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0.061	0	
A_7969_G	D -> G	Coat protein	SNP	0	0	0	0	0	0	0	0	0.429	0	0	0	
G_7982_A		Coat protein	SNP	0	0	0	0	0.231	0	0	0	0	0	0	0	
A_8026_G	Q -> R	Coat protein	SNP	0	0	0	0.022	0	0	0	0	0	0	0	0	
A_8058_G	R -> G	Coat protein	SNP	0	0	0	0.021	0	0	0	0	0	0	0	0	
T_8087_C		Coat protein	SNP	0.022	0	0.012	0	0.013	0.015	0.018	0.019	0.018	0.015	0	0	
C_8090_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0.644	0	0	0	
C_8152_T	S -> L	Coat protein	SNP	0	0	0	0	0	0	0	0.034	0	0	0	0	
T_8192_C		Coat protein	SNP	0	0	0.017	0.02	0	0	0	0	0	0	0	0	
C_8239_T	P -> L	Coat protein	SNP	0	0	0	0	0	0	0	0	0.047	0	0	0	
T_8243_C		Coat protein	SNP	0	0	0	0	0.011	0	0.015	0	0	0	0.258	0	
A_8252_C		Coat protein	SNP	0	0	0	0	0	0.021	0.014	0	0	0	0	0	
C_8255_T		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0.022	0	
A_8274_C	I -> L	Coat protein	SNP	0.016	0.028	0.017	0.014	0.023	0.024	0.02	0.037	0.02	0.024	0.037	0	
A_8287_C	N -> T	Coat protein	SNP	0.025	0	0	0.014	0	0.011	0	0	0.02	0.011	0.012	0	
A_8295_C	I -> L	Coat protein	SNP	0	0.018	0	0	0	0	0	0.016	0.011	0.013	0.024	0	
A_8323_G	N -> S	Coat protein	SNP	0.029	0	0	0	0	0	0	0	0	0	0	0	
T_8348_C		Coat protein	SNP	0	0	0	0	0	0	0.281	0	0	0	0	0	
A_8401_C	K -> T	Coat protein	SNP	0	0	0	0.024	0	0	0	0	0	0	0	0	
G_8451_		Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0.03	0	
G_8465_A		Coat protein	SNP	0	0	0	0.038	0	0	0	0	0.012	0	0	0	
C_8469_T	P -> S	Coat protein	SNP	0	0	0	0	0	0.026	0	0	0	0	0	0	
G_8477_A		Coat protein	SNP	0	0	0.013	0	0	0	0	0	0	0	0.031	0	
A_8485_T	Y -> F	Coat protein	SNP	0	0	0	0	0.011	0.01	0.036	0	0	0	0	0	
A_8541_C	I -> L	Coat protein	SNP	0.029	0.049	0.019	0.043	0.023	0.033	0.017	0.02	0.013	0.021	0.012	0	
C_8543_T		Coat protein	SNP	0	0	0	0.022	0	0	0	0	0	0	0	0	
A_8574_C	K -> Q	Coat protein	SNP	0.015	0.033	0	0.021	0	0.021	0	0.017	0.013	0	0.016	0	
C_8669_T		Coat protein	SNP	0	0.141	0	0	0	0	0	0	0	0	0	0	
T_8681_C		Coat protein	SNP	0	0.032	0	0.04	0	0	0	0	0	0	0	0	
G_8702_A		Coat protein	SNP	0	0	0	0	0	0	0	0	0.438	0	0	0	

C_8722_T	T -> I	Coat protein	SNP	0	0	0	0.02	0	0	0	0	0	0	0	0
A_8763_G	N -> D	Coat protein	SNP	0	0	0	0	0.024	0	0	0	0	0	0	0
G_8772_C	D -> H	Coat protein	SNP	0	0.024	0	0	0	0	0	0	0	0	0	0
A_8779_C	E -> A	Coat protein	SNP	0	0	0.022	0	0	0	0	0	0	0	0	0
T_8801_C		Coat protein	SNP	0.994	0.986	1	0	0.983	0.984	0.977	1	1	1	1	0
A_8813_G		Coat protein	SNP	0.994	0.973	0.983	0	1	1	0.991	1	1	0.989	1	0
A_8815_G	H -> R	Coat protein	SNP	0.02	0	0	0	0.013	0	0	0	0.014	0.013	0	0
A_8815_C	H -> P	Coat protein	SNP	0	0.022	0.02	0	0	0.018	0	0	0	0	0	0
G_8818_T	S -> I	Coat protein	SNP	0	0.031	0	0	0	0	0	0	0	0	0	0
A_8821_T	Y -> F	Coat protein	SNP	0	0	0	0	0	0	0	0	0	0	0.218	0
C_8830_T	A -> V	Coat protein	SNP	0	0.031	0	0	0	0	0	0	0	0	0	0
A_8843_G		3' UTR	SNP	0	0	0	0	0	0	0	0	0	0	0.028	0
A_8847_G		3' UTR	SNP	1	0.984	1	0	1	0	0.994	1	1	1	0	0
A_8864_C		3' UTR	SNP	0	0.984	1	0	1	0	1	1	1	1	1	0
A_8878_G		3' UTR	SNP	0	0	0	0	0	0	0.024	0	0	0	0	0
A_8880_G		3' UTR	SNP	0	0.017	0.039	0	0	0	0.036	0.013	0	0	0	0
T_8881_C		3' UTR	SNP	0	0	0	0	0	0	0	0.026	0	0	0	0
A_8894_G		3' UTR	SNP	0	1	0	0	0	0	1	1	1	1	0	0
A_8899_		3' UTR	SNP	0	0	0	0	0	0	0	0	0.021	0	0	0
C_8913_		3' UTR	SNP	0	1	0	0	0	0	0.987	1	0.972	0.976	0	0
C_8913_T		3' UTR	SNP	0	0	0	0	0	0	0	0	0	0.024	0	0
C_8913_A		3' UTR	SNP	0	0	0	0	0	0	0	0	0.021	0	0	0
T_8919_C		3' UTR	SNP	0	1	0	0	0	0	1	1	1	1	0	0
T_8929_C		3' UTR	SNP	0	0.991	0	0	0	0	1	1	1	1	0	0
A_8938_G		3' UTR	SNP	0	1	0	0	0	0	1	1	0.993	1	0	0
G_8956_A		3' UTR	SNP	0	1	0	0	0	0	1	1	1	1	0	0
T_8957_C		3' UTR	SNP	0	1	0	0	0	0	1	1	0.993	1	0	0
T_8962_C		3' UTR	SNP	0	0.029	0	0	0	0	0	0	0	0.018	0	0
A_8965_G		3' UTR	SNP	0	0	0	0	0	0	0.028	0	0	0	0	0
A_8979_G		3' UTR	SNP	0	0	0	0	0	0	0.023	0	0	0	0	0
A_8996_G		3' UTR	SNP	0	0	0	0	0	0	1	0	0	0	0	0
		SNPGenie morphic		133	209	127	133	122	134	138	149	179	198	172	225
		Nucleotide diver		8.1 10-4	1.25 10-3	4.4 10-4	5.4 10-4	5.1 10-4	6.2 10-4	5.5 10-4	6.5 10-4	3.4 10-3	4.6 10-3	3.8 10-3	3.9 10-3

H2	H3
0	0
6	0
0	7

Shared S mutations

	H1	H2	H3
H1	22	3	4
H2	3	22	2
H3	4	2	21

No Nonsynonymous mutation was shared between different Haplotypes

	H1	**H2**	**H3**
H1		3 S	4 S
H2	0 NS		2 S
H3	0 NS	0 NS	

Repartition of non-synonymous on the proteins

	CP	P1 protein	NIa	CI protein	Vpg	Nib	HAM1
H1	2	3	0	0	1	1	1
H2	2	1	1	0	0	1	1
H3	2	2	0	2	1	0	0

Table S 10. A list of mutations confirmed by Sanger sequencing.
A representative number of mutations that contributed to the discrimination of the UCBSV population were validated by Sanger sequencing.

Primers	Position	Reference	Kayonza	Kirehe	Nyagatare	Gatsibo	Ngoma	Bugesera	Gisagara	Muhanga	Kamonyi	Nyanza	Ruhango	Nyamasheke	SNP validation
Primers 1															
4350 => 4833	4385	C	T	T	T	T	T	T	T	TC	T	T	T	T	OK
4350 => 4833	4410	G	G	G	G	G	G	G	G	G	G	G	G	AG	OK
4350 => 4833	4493	A	A	A	A	A	A	A	A	A	A	A	A	TA	OK
4350 => 4833	4547	A	G	G	G	G	G	G	G	A	AG	G	AG	AG	OK
4350 => 4833	4706	A	G	G	G	G	G	G	G	G	G	G	G	AG	OK
Primers 2															
2894=>3307	2918	C	C	C	C	C	C	C	C	T	T	TC	CT	C	OK (TC for Kamonyi was expected)
2894=>3307	3050	C	C	C	C	C	C	C	C	T	T	TC	TC	TC	OK (TC for Ruhango was expected)
2894=>3307	3104	A	G	G	G	G	G	G	G	A	A	GA	GA	GA	OK (AG was expected for kamony)
2894=>3307	3131	G	G	G	G	G	G	G	G	G	T	TG	TG	G	OK (TG was expected for Kamonyi)
2894=>3307	3155	C	T	T	T	T	T	T	T	C	C	CT	CT	T	OK (CT was expected for Kamonyi)
2894=>3307	3236	T	T	T	T	T	T	T	T	T	T	T	T	AT	OK
Primers 3															
7200=>7564	7228	A	G	G	G	G	G	G	G	A	AG	AG	AG	A	OK (AG was expected for Nyamasheke)
7200=>7564	7352	C	C	C	C	C	C	C	C	C	C	C	C	TC	OK
7200=>7564	7385	C	C	C	C	C	C	C	C	C	C	C	C	CT	OK
7200=>7564	7445	A	A	A	A	A	A	A	A	A	A	A	A	GA	OK
7200+>7564	7448	G	A	A	A	A	A	A	A	G	GA	GA	GA	GA	OK
7200=>7564	7487	G	G	G	G	G	G	G	G	A	AG	AG	AG	G	OK
7200=>7564	7529	T	C	C	C	C	C	C	C	T	TC	TC	TC	TC	OK
Primers 4															
898=>1396	992	T	C	C	C	C	C	C	C	C	C	C	C	C	OK
898=>1396	1103	T	G	G	G	G	G	G	G	C	C	C	C	G	OK
898=>1396	1158	T	C	C	C	C	C	C	C	T	TC	TC	TC	C	OK (Nyamasheke was expected to have TC)
898=>1396	1164	A	G	G	G	G	G	G	G	A	A	AG	AG	G	OK (Kamonyi& Nyamasheke were expected to have AG)
898=>1396	1184	T	T	T	T	T	T	T	T	C	C	CT	TC	T	OK (Kamonyi was expected to have CT)
Primers5															
5555 => 6053	5693	T	G	G	G	TG	G	G	G	T	TG	TG	TG	TG	OK
5555 => 6053	5819	C	T	T	T	T	T	T	T	C	TC	CT	CT	CT	OK
5555 => 6053	5825	T	T	T	T	T	T	T	T	C	CT	CT	CT	T	OK
5555 => 6053	5838	G	A	A	A	GA	A	A	A	G	GA	GA	GA	GA	OK
5555 => 6053	5894	T	T	T	T	T	T	T	T	C	CT	CT	CT	T	OK
5555 => 6053	5948	G	A	A	A	AG	A	A	A	G	GA	GA	GA	GA	OK
Primers 6															
2640=>2814	2669	T	C	C	C			C	C	T	TC	T	TC		OK (Nyanza was expected to have TC)
2640=>2814	2762	C	T	T	T	T		T	C	C	CT	CT	CT	C	OK
Primers 7															
3166=>4035	3380	A	A	A	A			A	A	G		GA	GA		OK
3166=>4035	3668	C	T	T	T			T	T	T		T	T		OK
3166=>4035	3773	G	A	A	A			A	A	G		GA	GA		OK
3166=>4035	3911	A	G	G	G			G	G	A		AG	AG		OK

Haplotype 1
Haplotype 2
Haplotype 1 + 2
Haplotype 3

Table S 11. A complete list of all 564 SNPs used to create five dimensions.

SNP /position	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
A_5324_G	0.756785205	0.131302358	0.000252495	0.112280201	0.000376571
C_4412_A	0.740126975	0.13182193	5.65E-05	0.028639029	0.005770939
A_8172_G	0.731503661	0.124627829	0.000541331	0.091306536	0.010861813
C_2369_T	0.731128378	0.117706469	0.001023999	0.130086572	0.004771567
T_5825_C	0.72874403	0.12078114	0.001148477	0.08636833	7.34E-05
T_5894_C	0.727130888	0.116211149	0.002128876	0.151011283	0.004091601
G_3773_A	0.72621383	0.119490713	0.000934235	0.127792392	0.011045417
T_8459_C	0.724875239	0.11718445	0.00155761	0.147274153	0.01235156
G_5558_A	0.722654896	0.121151238	0.000766785	0.052018712	0.005956784
C_2711_T	0.72143515	0.121991349	0.00064574	0.048583304	0.004236189
C_6731_T	0.720318202	0.116366062	0.001909603	0.127737732	0.005133815
G_4046_A	0.718027709	0.113997399	0.002416521	0.155704716	0.007575259
G_458_A	0.715251713	0.109662862	0.001820192	0.073483396	0.04133063
G_2153_A	0.714047396	0.109529068	0.002368272	0.153555892	0.004763978
G_5115_A	0.712873565	0.113394919	0.001926392	0.125851889	0.005785086
A_6431_G	0.712056442	0.118116521	0.000929012	0.073384125	0.001207874
C_7664_T	0.71194976	0.122193789	0.002369091	0.153085589	0.006630498
T_8043_G	0.711610348	0.11249519	0.002223333	0.180318841	0.021065314
G_7487_A	0.711277104	0.109692191	0.003257228	0.16105017	0.002588279
G_3131_T	0.711091059	0.113975738	0.002217867	0.138745484	0.010092627
C_3155_T	0.710898873	0.112870004	0.002097513	0.146719399	0.014325981
C_2918_T	0.710062067	0.109783464	0.003222102	0.165268121	0.004544536
A_3380_G	0.710016638	0.108543096	0.002846691	0.173289335	0.007840651
T_8606_C	0.709039887	0.034841618	0.264622651	0.028926996	0.002992424
T_8225_G	0.704245002	0.111838536	0.001848704	0.163207104	0.033514805
C_1436_T	0.704181763	0.109158834	0.002138283	0.14987085	0.01382287
C_5157_T	0.703891307	0.105922218	0.00349179	0.164812308	0.002412028
A_670_G	0.703709272	0.112881327	0.001956444	0.143353327	0.021673371
T_1184_C	0.70347726	0.114300202	0.001564467	0.118749484	0.017063995
T_6683_C	0.701966887	0.029148646	0.288322418	0.014992535	0.00538765
C_7124_T	0.698506929	0.110207775	0.002701894	0.136829004	0.008752009
C_8012_T	0.696358035	0.02600469	0.30472784	0.026001626	0.008319277
A_3911_G	0.695456819	0.025534471	0.313150319	0.015952061	0.000729088
G_2639_A	0.694915673	0.024378338	0.314463894	0.016342674	7.98E-06
T_1103_G	0.694408017	0.025677798	0.315932612	0.013150852	7.16E-05
G_7448_A	0.693680511	0.029835792	0.307246572	0.015701399	0.003175616
T_1103_C	0.693571451	0.024836463	0.318222471	0.013692703	5.85E-05
G_8555_A	0.692500118	0.100326845	0.004932557	0.228807783	0.007736489
C_3050_T	0.691998686	0.026179455	0.316603165	0.010420975	0.000657867
T_2123_C	0.691866356	0.023364733	0.318503654	0.019217565	6.35E-05
A_6445_G	0.691669108	0.109080417	0.002877382	0.160188772	0.023118838
G_7206_A	0.687556249	0.107989372	0.002389778	0.079218854	0.003231606
A_4547_G	0.687229356	0.023495731	0.329614215	0.012019069	0.000201567
T_2669_C	0.686816542	0.023180721	0.321629551	0.021344769	0.000160089
T_7529_C	0.684363969	0.024348841	0.324339697	0.027632262	0.00042122
C_452_T	0.683876759	0.137118092	0.003259484	0.166753491	0.012333299
A_3104_G	0.682459323	0.023630854	0.331289383	0.016364197	3.72E-08
C_2762_T	0.681828442	0.022001993	0.33255835	0.023903369	0.000416221
C_4883_T	0.680176365	0.020838494	0.33740076	0.026450679	0.000154073
T_1158_C	0.677605447	0.020406614	0.343920914	0.007182444	0.001237711
A_515_G	0.676417245	0.021371742	0.346082651	0.0191595	0.000127417
A_1164_G	0.675957868	0.022558662	0.344961803	0.008583147	0.000856141
C_5819_T	0.67575454	0.019575651	0.347696106	0.022641491	0.000503003
A_7228_G	0.675107792	0.021708787	0.34006616	0.013147246	0.002595945
C_4807_T	0.674716196	0.019756274	0.361539901	0.011631509	9.13E-05
G_5838_A	0.67428428	0.019711322	0.347976247	0.025998866	0.000661939
A_5453_G	0.672281576	0.023650026	0.354467622	0.015056278	2.99E-05
T_4982_C	0.671947079	0.021188477	0.346493868	0.022534331	0.000491856
T_5693_G	0.668131958	0.018579618	0.365764822	0.009515832	6.04E-07
G_5948_A	0.656042392	0.017353704	0.385771403	0.015578377	0.000172504
T_3203_C	0.648598333	0.154906794	0.010313387	0.096765458	0.103417589
G_5204_A	0.647842918	0.01996168	0.392628944	0.031484971	0.000193546
G_7757_T	0.641565486	0.013497844	0.402525322	0.046102872	0.001652603
T_6311_C	0.641307484	0.162982761	0.014627458	0.129914829	0.056526801
C_7772_T	0.637536484	0.011641019	0.406300017	0.050499435	0.001920118
A_8557_C	0.634745364	0.119086928	0.002720168	0.27862096	0.019627708
G_3416_A	0.634643062	0.14890773	0.008114171	0.081250288	0.059701695
T_8453_C	0.634035014	0.133114576	0.004350033	0.046405114	0.331587574
T_7799_A	0.632109626	0.012566081	0.399405057	0.066476597	0.00050741

A_5094_G	0.628513496	0.133740192	0.003335899	0.041560102	0.101802235
CG_6771_TTA	0.62319746	0.124182399	0.001926055	0.024539108	0.364944211
A_1775_G	0.60840142	0.124767031	0.001811003	0.027881967	0.081160041
C_1401_T	0.605548795	0.167779023	0.020323892	0.173689813	0.003228684
A_7296_G	0.605453455	0.139642394	0.006024104	0.065564519	0.024493587
A_7790_C	0.585859146	0.011326545	0.044114997	0.214142786	0.046372239
T_8306_C	0.574690282	0.011516065	0.44722383	0.031016863	0.055653198
C_4385_T	0.527527215	0.099749338	0.000392566	0.000233341	0.095678895
T_1172_C	0.495738697	0.152941659	0.023861229	0.201164589	0.056162139
A_6957_C	0.484983346	0.192025562	0.025168984	0.095284041	0.16021944
A_5182_C	0.475015691	0.192946096	0.003269976	0.033508924	0.009623307
A_7868_G	0.447694881	0.182827174	0.064370251	0.471471238	0.070858165
A_6925_G	0.443537857	0.174271232	0.072035423	0.4298162	0.106611051
T_989_C	0.410586906	0.050641592	0.011418017	0.63283542	0.306753634
A_5892_C	0.394751056	0.050608995	0.005315706	0.081176093	0.005003656
A_4931_C	0.392606733	0.073657414	0.05465114	0.002271975	0.299322202
T_4241_C	0.378857368	0.160011035	0.072142975	0.49275177	0.040118651
A_8541_C	0.366020285	0.178714781	0.002036194	0.07391962	0.02427539
A_7020_C	0.359162788	0.029418395	0.033494931	0.153224715	0.017212875
G_8138_A	0.342333869	0.030706227	0.018057515	0.080294255	0.14724492
A_162_C	0.335503173	0.723443791	0.002117121	0.087858208	9.79E-06
A_120_T	0.331096787	0.32254708	0.001754481	0.001607645	0.056372256
A_5560_C	0.326192157	0.208690132	0.316610664	0.029818342	1.04E-05
A_5322_C	0.319780425	0.255842709	0.016511183	0.276370717	0.021485193
G_66_T	0.314863034	0.421668147	6.28E-05	0.001004349	0.002710218
A_5686_C	0.294668878	0.077007788	0.022737344	0.066981044	0.110154123
A_2689_C	0.290932671	0.15838398	0.171894719	0.115626773	0.001629571
A_2173_C	0.288818538	0.003748084	0.07032638	0.123482517	0.063960181
A_2713_C	0.283747297	0.54670687	0.007462045	0.002027927	0.003453171
A_8815_C	0.279614698	0.209046491	0.015858908	0.000114742	0.001586294
G_125_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_176_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
C_185_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_188_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_509_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
C_1889_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_1901_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_1907_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_1919_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_1970_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_1976_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_5342_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_5706_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_5714_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_5730_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_6438_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_7396_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_7973_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_7978_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_7991_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_8011_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_8019_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_8022_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
G_8042_A	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
T_8043_C	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
C_8430_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_8440_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_8513_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_8514_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
C_8612_T	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_8723_G	0.278416274	0.160798339	0.103704336	0.703106355	0.065876756
A_96_G	0.272508109	0.170364074	0.103183371	0.703793322	0.064821083
A_6042_C	0.266185715	0.327135422	0.345142888	0.02984898	0.00066761
G_68_T	0.262977036	0.512185529	0.006555111	0.001442654	0.009938606
X_1624_A	0.258731675	0.057473049	0.080689067	0.00088332	4.17E-05
A_1818_C	0.258093659	0.217346716	0.311159801	0.09918971	0.045155383
T_5402_C	0.249815773	0.066915163	0.028584933	2.13E-07	0.013481797
G_240_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_904_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_1457_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_1628_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
A_2226_G	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_2648_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_3602_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_4635_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487

G_4820_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_5354_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
A_6056_G	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_6308_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_7338_G	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_8669_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_322_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_450_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_557_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_818_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_1182_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_1423_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_2027_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
A_2742_G	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
A_3271_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
A_3272_G	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_3374_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_3734_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_4514_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_4784_A	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
T_6152_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_7067_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_7700_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_7866_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_7887_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_8772_C	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
G_8818_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_8830_T	0.249350754	0.871547581	0.067582463	0.004379156	0.01254487
C_194_T	0.244673028	0.038388069	0.000306015	0.002748701	1.181102404
A_461_C	0.244598858	0.061796815	0.009998412	0.978798735	0.095924366
GG_294_AA	0.236954592	0.256706606	0.003266197	0.002857923	0.030692289
A_5410_C	0.234868859	0.312931603	0.26505166	0.365304251	0.013550416
A_4215_C	0.23415263	0.010824069	0.042551131	0.965750943	0.183720663
A_1339_C	0.23221612	0.101143122	0.10140621	0.005101482	0.016199945
A_2933_C	0.228749627	0.463493033	0.007385945	0.007483052	0.245777847
T_1382_C	0.218355512	0.044215102	0.004636587	0.216115761	0.212953649
T_8681_C	0.217042672	0.336509664	0.000827673	0.016697974	0.046748478
A_5949_C	0.210179921	0.314674193	0.001408977	0.01786356	0.050029852
A_4007_C	0.195601774	0.398604548	0.000523326	0.000550025	0.10255029
G_2987_A	0.195219685	0.029221013	0.092091037	0.150063616	0.361956609
T_326_C	0.194005936	0.107278732	0.165293284	0.659037958	0.07461673
G_590_T	0.192210596	0.761471945	0.041796522	0.122120586	0.115781391
C_7637_T	0.192077376	0.856775712	0.146140704	0.004888794	0.007944974
C_785_T	0.191697966	0.021376546	0.000757586	0.00416537	1.351883788
C_1195_T	0.191091484	0.021206728	0.00078593	0.004338321	1.353614027
A_8574_C	0.189722233	0.273624194	0.12254421	0.066863012	0.236383603
A_3905_C	0.184794935	0.372020365	0.395672994	0.050097431	0.061556267
A_5678_C	0.184736973	0.658688518	0.12161928	0.184124066	0.006936065
A_6274_C	0.184725143	0.325921284	0.001970741	0.094627386	2.79E-07
A_3574_C	0.176841138	0.621432986	0.020220853	0.11084956	0.191516621
T_6674_C	0.176078323	0.113819549	0.00246091	0.281405197	0.202353376
A_6216_C	0.175831572	0.09667757	0.019032865	0.03926291	0.686589411
T_8534_C	0.175449101	0.096107179	0.171932973	0.631647569	0.073848316
A_8880_G	0.175433116	0.00809351	0.0989975	0.000821787	0.037574627
G_3696_A	0.174442231	0.012703132	0.018591343	0.083647468	0.002829375
T_3788_C	0.174442231	0.012703132	0.018591343	0.083647468	0.002829375
T_4856_C	0.174442231	0.012703132	0.018591343	0.083647468	0.002829375
T_7400_A	0.174442231	0.012703132	0.018591343	0.083647468	0.002829375
A_8323_G	0.174442231	0.012703132	0.018591343	0.083647468	0.002829375
A_6572_G	0.174087174	0.016674303	0.001793317	0.010625992	1.400061247
A_7713_C	0.17395882	0.494757439	0.027757545	0.162063002	0.271415968
C_6998_T	0.173191036	0.016447794	0.001857886	0.011034992	1.402397623
T_1349_C	0.16965801	0.015566972	0.00212378	0.012724106	1.411498812
C_1799_T	0.168342725	0.028195041	0.121558546	0.212210668	0.17640184
G_6971_T	0.167124073	0.014947271	0.002325668	0.014011215	1.417917737
A_8663_G	0.16676373	0.073529263	0.179691246	0.602749271	0.0531631
GA_6355_AG	0.166509643	0.014798528	0.002376036	0.014332881	1.419460469
T_3185_G	0.165901716	0.728920866	0.035432391	0.009599301	0.215912036
C_3089_T	0.160802765	0.013445592	0.002870444	0.01750031	1.433531502
C_3761_T	0.160087783	0.013279754	0.00293579	0.017920166	1.435261334
T_692_C	0.153707531	0.011836461	0.003552917	0.021896775	1.450367706
A_5384_G	0.153003795	0.826743207	0.088469035	0.177785514	0.042476691
A_8274_C	0.152629604	0.021250688	0.052929394	0.97981779	0.002694801
T_3041_C	0.151166226	1.0283762	0.037100898	0.011571147	0.003431343
A_5586_C	0.150196761	0.8466522	0.08582658	0.027552654	0.278790301

A_2630_G	0.150097542	0.011049338	0.003929507	0.024332188	1.458648878
C_1886_T	0.148913771	0.010795934	0.004057369	0.025160369	1.46132196
A_6795_C	0.148630711	0.656813494	0.162420517	0.117538685	0.035612956
G_8636_A	0.148029875	0.007019885	0.005021948	0.035977539	1.469733935
G_6355_A	0.147133633	0.020382823	0.002709629	0.015616528	1.454745255
G_7984_A	0.143334251	0.009633275	0.00468951	0.029263266	1.473634947
A_1730_G	0.141720638	0.009306906	0.004881504	0.030511905	1.477106817
A_3946_C	0.14069477	0.326439543	0.258372976	0.002853515	0.007537921
A_4485_C	0.14064557	0.109382795	0.239337049	0.099911635	0.007618022
C_6461_T	0.140031318	0.008970036	0.005086977	0.031849367	1.48069818
G_3134_A	0.139212242	0.008808486	0.005188259	0.03250905	1.482423386
A_8815_G	0.137411271	0.048451719	0.380459339	0.018052387	0.257560969
A_8287_C	0.135268836	0.035024178	0.50255874	0.077050193	0.041283317
G_7367_A	0.133204565	0.635349749	0.016190868	0.372002549	0.092128146
G_554_A	0.133161562	0.007651684	0.00597046	0.037612293	1.494837694
A_470_C	0.132177696	0.006736936	0.014194798	0.036372575	0.002878608
A_7519_C	0.130330304	0.258450919	0.001849771	0.033199391	0.001069526
A_455_G	0.128427397	0.006792477	0.006625122	0.041893642	1.504136865
A_6904_C	0.126776492	0.565214926	0.219662583	0.165513944	6.40E-05
A_8864_C	0.124645092	0.041334438	0.158342543	0.211324464	0.185207273
A_4684_G	0.124558873	0.006121148	0.007188675	0.045585546	1.511458117
G_1598_A	0.123164671	0.005886129	0.007398213	0.046959625	1.514034196
A_8600_G	0.120384473	0.005428624	0.007826418	0.049769773	1.519070616
T_50_C	0.111141118	0.468494369	0.074432137	0.009039822	0.111025985
A_4023_G	0.108903232	0.00370264	0.009746555	0.062401364	1.538395796
A_7174_C	0.105716738	0.251590952	0.360299882	0.192628223	0.173358932
A_4071_C	0.103942091	0.254859915	0.330453511	0.25975395	0.000753756
T_7721_C	0.103228444	0.614504097	0.034543244	0.043930858	0.008440393
C_5231_T	0.101339111	0.719507786	0.089474192	0.322743171	0.120041181
A_109_C	0.100607481	0.000501467	0.00414056	0.629449049	0.264594784
C_8913_	0.099254028	0.00182343	0.354832136	0.006880728	0.699030653
G_8054_A	0.098378753	0.015186525	0.167434779	0.474080979	0.007854571
A_5345_G	0.097640764	0.157034847	0.004876425	0.014085406	0.276375631
A_8938_G	0.095478115	0.001314181	0.362124948	0.007042749	0.709936808
T_8957_C	0.095478115	0.001314181	0.362124948	0.007042749	0.709936808
A_8894_G	0.095298653	0.00120786	0.36632469	0.007348621	0.70684938
T_8919_C	0.095298653	0.00120786	0.36632469	0.007348621	0.70684938
G_8956_A	0.095298653	0.00120786	0.36632469	0.007348621	0.70684938
T_8929_C	0.094085766	0.000905878	0.366047101	0.007432544	0.708431783
T_6674_G	0.0933702	0.018197068	0.036503441	0.181073796	0.970618643
A_1065_C	0.09276839	0.910853757	1.35E-06	0.087586943	0.001115717
A_77_G	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
A_284_G	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
T_1397_C	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
A_2330_G	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
T_3929_C	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
C_5910_T	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
T_7195_G	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
C_8913_T	0.086599026	0.001184845	0.014252901	0.092189754	1.56840419
A_8250_C	0.085036828	0.31407315	0.016722827	0.351164466	0.136989644
A_648_C	0.084668141	0.022417178	0.214679855	0.769768193	0.127250318
G_6384_A	0.082579258	0.080240434	0.177419216	0.636181152	0.201764808
A_6874_C	0.082316196	0.196482955	0.034535536	0.510363072	0.00173471
A_5800_C	0.08048834	0.44929567	0.080312071	0.234580661	0.092875435
A_4800_T	0.079123089	0.65695155	0.087891602	0.395346468	0.16555111
T_8962_C	0.078824148	0.716672004	0.087708989	0.050957394	0.621586587
C_7976_T	0.077709612	0.006644846	0.00849254	0.703080706	0.697945097
T_2306_C	0.073204691	0.055663629	0.203618862	0.020154473	0.039271181
T_7400_C	0.070920782	0.018798569	0.000788879	0.833884256	0.351723415
T_6317_C	0.068736349	0.322434125	0.643588079	0.153581393	0.245992255
A_8813_G	0.066823707	0.082213931	0.176345721	0.607041827	0.19575474
T_6590_C	0.066227317	0.007298323	0.223573451	0.369194348	0.917034501
T_8801_C	0.066207909	0.073899679	0.184203293	0.611997157	0.204979712
G_39_A	0.066111554	0.259356701	0.173016858	0.002682819	0.151652415
A_4205_C	0.06516225	0.435860673	0.09725061	0.00905302	0.003030832
G_4166_T	0.062716071	0.164472388	0.046470631	0.09226005	0.00027488
A_2505_C	0.061815295	0.24992694	0.008604573	0.628213187	0.347820042
A_152_C	0.060914551	0.335399922	0.018354329	0.559806418	0.422297205
T_7460_C	0.060360752	0.085452401	0.143693441	0.000567042	1.34E-05
T_2261_C	0.059234832	0.062809117	0.002627758	0.2833866	0.118301373
A_8485_T	0.057101946	0.154768727	0.134693439	0.0006055	0.024292853
A_8847_G	0.05691155	0.02550088	0.164635048	0.007845656	0.761169218
A_867_C	0.056662876	0.032572424	0.619630938	0.046099355	0.001714367
C_4316_T	0.056096389	0.014268644	0.265858793	0.43290612	0.046370659
A_1254_G	0.055311098	0.010461546	0.117788403	0.042751618	0.130701218

T_7070_G	0.054690685	0.492312038	0.104178621	0.018330877	0.267922559
A_8252_C	0.054439452	0.084563827	0.12228846	0.001603311	0.001289225
A_8505_C	0.054296674	0.29658397	0.022494625	0.011550436	0.008419866
GT_3741_ATC	0.054044719	0.082756659	0.121482589	0.001578078	0.001061844
A_4936_C	0.052571173	0.494687122	0.228949051	0.317201547	0.040421654
T_3197_G	0.052523353	0.701077504	0.003205609	0.145620588	0.002399609
C_5066_T	0.052083372	0.07511955	0.11739363	0.001467604	0.000366568
A_1722_C	0.05161471	0.175236543	0.042913621	0.018437808	0.08736103
T_8192_C	0.05117802	0.022580056	0.135740441	0.036911567	0.078742872
T_3139_G	0.050329387	0.667051167	0.112680917	0.111463942	0.011489121
G_186_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_713_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_802_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_2143_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
T_2267_C	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_2282_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_2913_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_3179_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_3563_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_4016_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_4639_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_4934_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_5648_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_5810_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
T_7465_C	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
T_7544_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_7968_A	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
C_8255_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
G_8451_	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
A_8821_T	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
A_8843_G	0.049901023	0.001596819	0.015596507	0.848773407	0.649063838
T_6657_C	0.049386081	0.002096133	0.018287837	0.88235161	0.616460699
G_800_A	0.045028688	0.109217411	0.130863955	0.001533577	0.001598524
G_5054_A	0.044989045	0.000226835	0.100412416	0.062062667	0.16575765
A_47_T	0.044663201	0.102938608	0.086421039	0.000325862	0.015204439
T_8243_C	0.04392161	0.000158525	0.010237268	0.860750593	0.639587801
A_5887_G	0.043531111	0.105825522	0.148834455	0.000329539	0.001056175
G_4812_A	0.043317034	0.107018626	0.094948286	0.001707032	0.018634986
A_7885_C	0.041858064	0.025913349	0.2432172	0.602900918	0.147047252
A_6863_C	0.041572092	0.738490318	0.022042533	0.510106	0.028695915
A_2585_G	0.039842305	0.011303079	0.009284136	0.063328102	0.000138405
T_992_C	0.037721972	0.078849002	0.099130027	0.001676492	0.000409355
C_823_T	0.035249167	0.030758413	0.125898642	0.019197129	0.020242541
A_906_C	0.034363239	0.753778306	0.001773487	0.265351196	0.025148768
T_659_C	0.03396126	0.60940629	0.063005868	0.433838766	0.080700493
A_5661_T	0.033530531	0.432597299	0.104093985	0.013222169	0.216685574
G_5201_A	0.031617306	0.778714674	0.017959861	0.485173828	0.025917979
G_8465_A	0.031545242	0.022785157	0.010440442	0.128106586	0.158967658
T_3170_C	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
G_3498_T	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
A_4202_G	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
G_5210_A	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
A_8026_G	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
A_8058_G	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
A_8401_C	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
C_8543_T	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
C_8722_T	0.028921605	0.001419695	0.059118988	0.044484909	0.125265131
A_5570_C	0.028482755	0.619038148	0.015445673	0.044674008	0.200453617
G_7715_A	0.028268176	0.002864563	0.002389949	0.788199897	0.535026346
G_543_A	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
T_2120_C	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
C_4049_T	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
T_5393_C	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
T_6745_C	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
A_7686_G	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
C_8469_T	0.02764715	0.020989489	0.063859583	0.00054668	0.004982926
A_4602_C	0.026932248	1.128514299	0.019044114	0.01357403	0.013619288
T_6323_C	0.025992892	0.002975532	0.161141648	0.22777283	0.017907195
A_4714_G	0.02591035	0.002952457	0.161077774	0.227437816	0.017877612
G_8477_A	0.025524257	0.001068752	0.000270197	0.747559787	0.588422733
A_669_G	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
C_1028_A	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
G_1202_A	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
C_2274_T	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
A_4436_G	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448

G_4992_A	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
C_5723_T	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
T_6890_C	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
C_8152_T	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
T_8881_C	0.024564791	0.063935271	0.030866328	3.01E-07	0.018419448
G_306_A	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_1722_T	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
T_2036_C	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
C_3320_T	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
T_4877_C	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
T_8348_C	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_8878_G	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_8965_G	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_8979_G	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_8996_G	0.023464567	0.080911923	0.050327838	0.00115727	0.028170055
A_6786_C	0.022467504	0.866242037	0.07019414	0.027618814	0.281682771
G_86_A	0.022289664	0.023689987	1.330858322	0.222338291	0.020106527
T_82_C	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
C_1343_T	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
T_3962_C	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
T_4196_C	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
A_4328_G	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
G_5088_A	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
T_5987_A	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
A_7080_G	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
G_7982_A	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
A_8763_G	0.022056702	0.065569055	0.070901885	0.00251072	0.00139342
A_465_G	0.02130877	0.036952092	0.114959763	0.000140688	0.009970641
A_2019_C	0.021098122	0.798509935	0.010643364	0.017264781	0.062396055
G_2278_A	0.02053968	0.059015436	0.034013046	0.000452603	0.016722018
T_3648_C	0.017973369	0.071948913	0.007682346	0.669498559	0.060359797
C_2312_T	0.017739612	0.03134233	0.065895675	0.001217529	9.01E-06
G_3782_A	0.017739612	0.03134233	0.065895675	0.001217529	9.01E-06
A_8779_C	0.017739612	0.03134233	0.065895675	0.001217529	9.01E-06
A_7286_C	0.017384675	0.064502892	0.229348079	0.138480081	0.68613079
T_7395_C	0.017377187	0.078127603	0.047505985	0.000360363	0.027431179
A_3072_C	0.017049129	0.219720863	0.27378035	0.002768085	0.076625091
A_5755_C	0.016870543	0.62834408	0.029222219	0.072155236	0.193233928
X_5955_A	0.016306399	0.1080801	0.336236743	0.374167616	0.232995434
T_4190_C	0.015569345	0.974924363	0.058291501	0.001645811	0.000440568
A_4493_C	0.01399206	0.395314591	0.093242962	0.191321748	0.014292433
A_5705_C	0.0133948	0.053757328	0.000842649	1.122928687	0.001709496
G_8001_A	0.013104549	0.048969365	0.041043696	0.004573051	0.013318974
AG_7065_CAT	0.012761948	0.102640442	0.041872126	0.81832467	0.064576463
A_6146_G	0.011328313	0.441203386	0.074174413	0.082244421	1.151523261
C_425_T	0.010492036	0.091097199	0.140679725	0.05201096	0.026564762
A_4364_G	0.009437783	0.04330318	0.045415804	0.009045275	0.011441943
T_40_C	0.009262526	0.198000413	0.102821277	0.03525817	0.87258587
A_6668_C	0.008483421	0.661332124	0.058800078	0.016564783	0.020871442
A_8512_C	0.008053896	0.062824888	0.164587909	0.214566143	0.000553913
A_8295_C	0.007553753	0.105778571	0.355030174	0.495325061	0.00053327
T_566_G	0.007175237	0.077876998	0.275319747	0.193375218	0.042823517
C_3668_T	0.0055888	0.114009601	1.204566984	0.181736176	0.05415645
A_5661_G	0.005502995	0.020232455	0.137357846	0.323886562	0.198524653
T_2504_C	0.005265398	0.269338273	0.066857041	0.252521114	0.183095703
G_551_A	0.004388471	0.122778431	1.212204646	0.157320626	0.0675905
T_6986_C	0.004245311	0.415950668	0.067168545	0.391567748	0.144244789
A_272_C	0.00420108	0.063108314	0.0293158	0.971414275	0.000154322
A_5610_C	0.003493122	0.045236282	0.000179575	0.697487703	0.070823796
A_3309_C	0.003234285	0.707781381	0.043594033	0.124552467	0.070129541
A_7457_C	0.002875838	0.054736119	0.177455543	0.020662909	0.113157434
A_4706_G	0.002623328	0.113488542	1.237507929	0.236184051	0.02857862
A_2438_C	0.002543988	0.314975165	0.40616696	0.447158688	0.017929054
T_1058_C	0.002540163	0.720453933	0.007434547	0.158270622	0.221519308
A_5425_C	0.002172085	0.131576141	0.002428406	0.311651228	0.717958659
T_8167_G	0.001991532	0.67571554	0.044404892	0.054994192	0.100467121
T_8087_C	0.001648392	0.512001636	0.054078437	0.002205611	0.440194863
C_7352_T	0.001554642	0.140876188	1.206064595	0.240241307	0.027340423
C_7777_T	0.001295071	0.117879095	1.222539988	0.245522675	0.032895751
G_202_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_611_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_731_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_846_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_881_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_1379_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096

A_2195_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_3236_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_3332_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_3678_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_4410_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_4493_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_4853_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_5555_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_5855_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_6464_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_7385_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_7445_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_7969_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_8090_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_8702_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_81_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_282_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_316_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_1032_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_1388_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_1673_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_1776_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_3305_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_3446_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
G_3669_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_3761_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_4631_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_4835_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_4870_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_5504_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_6245_G	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_6992_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
T_7386_C	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_8239_T	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_8899_	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
C_8913_A	0.001272391	0.13593289	1.210648366	0.239359751	0.028302096
A_7946_C	0.000403364	0.603405982	0.013598655	0.007809147	0.512317284
T_5585_C	0.00028401	0.299500819	0.668836635	0.057229555	0.163010529
A_7101_C	0.000264439	0.465997368	0.005689312	0.68196507	0.154159774
T_5748_G	0.000252265	0.000777852	0.628825671	0.421399873	0.001441321
A_1546_C	0.000200845	0.532449327	0.051699496	0.488004789	0.059746932
A_8551_C	1.52E-05	0.343714413	0.049074566	0.255918208	0.109343321

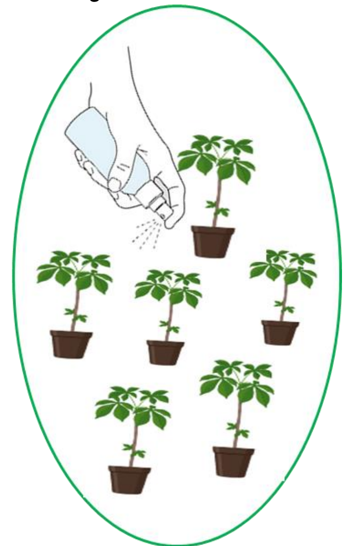
Table S11: also available at**:** https://doi.org/10.5281/zenodo.8222816.

Table S 12. Confirmation of UCBSV haplotype 2 based on Sanger sequencing of individual samples.

			Samples from Narocas1				Interpretation		
Primers	**Position**	**Reference**	**1**	**2**	**3**		**1**	**2**	**3**
Primers 1									
898=>1396	1103	T	C	C	C		H2	H2	H2
898=>1396	1158	T	T	T	T		H2	H2	H2
898=>1396	1184	T	C	C	C		H2	H2	H2
Primers 2									
4350 => 4833	4412	C	A	A	A		H2	H2	H2
Primer 3									
5555 => 6053	5693	T	T	T	T		H2	H2	H2
5555 => 6053	5819	C	C	C	C		H2	H2	H2
5555 => 6053	5825	T	C	C	C		H2	H2	H2
5555 => 6053	5838	G	G	G	G		H2	H2	H2
5555 => 6053	5894	T	C	C	C		H2	H2	H2
5555 => 6053	5948	G	G	G	G		H2	H2	H2
Primers 4									
7200=>7564	7228	A	NA	NA	NA				
7200=>7564	7487	G	A	A	A		H2	H2	H2

Chapter 5

Greenhouse thermotherapy combined with chemotherapy and field chemotherapy reduced viral load and severity of CBSD.



Chapter 5: Greenhouse thermotherapy combined with chemotherapy and field chemotherapy reduced viral load and severity of CBSD.

General introduction to chapter 5.

The third and fourth studies unequivocally confirmed that in Rwanda, CBSD is caused by CBSV and UCBSV, with UCBSV being the most prevalent. The investigation into CBSIs' genetic diversity revealed the presence of three distinct UCBSV haplotypes in cassava fields across the country. As disease-free planting materials are crucial for managing seed-borne viral diseases, this chapter evaluated the efficacy of greenhouse thermotherapy when combined with Salicylic acid (SA) or Benzothiadiazole (BTH) for combating CBSD. Furthermore, the effect of SA and BTH on CBSD root symptoms was evaluated through field chemotherapy, and an investigation into their impact on the cassava transcriptome was conducted.

Chapter 5. Greenhouse thermotherapy combined with chemotherapy and field chemotherapy reduced viral load and severity of CBSD.

Chantal NYIRAKANANI, Ariadna PICART PICOLO, Jean Pierre BIZIMANA, and Herve VANDERSCHUREN

5.1. Abstract

Cassava, which is a vital food crop in tropical regions, is undergoing a remarkable transformation from a mere staple crop to an industrial powerhouse that has the potential to revolutionize livelihoods. However, its productivity is being severely hampered by devastating viral diseases particularly cassava brown streak disease (CBSD). Exploring more straightforward and affordable methods for cleaning cassava planting material is crucial to mitigate yield loss and preserve the most productive cultivars.

The present study explored novel approach for mitigating CBSD by treating cassava cuttings infected with the virus. The treatments included greenhouse thermotherapy, greenhouse thermotherapy combined with salicylic acid (SA) at 30 mg/L and 50 mg/L, greenhouse thermotherapy combined with benzothiadiazole (BTH) at 10 mg/L and 50 mg/L, chemotherapy with SA at 30 mg/L and 50 mg/L, and BTH at 10 mg/L and 50 mg/L. The impact of the treatments on CBSIs was assessed by analyzing quantitative RT-PCR data at 6 and 12 weeks after planting with the comparative CT method and *t-test*. The impact of chemotherapy on CBSD root symptoms was assessed through field chemotherapy using 1 mM SA and 1 mM BTH. The severity scoring was conducted at harvest, 12 months after planting, and analyzed by an independent *t-test*. In addition, alterations in the cassava transcriptome caused by SA and BTH treatments were examined by RNAseq of treated plants.

Our findings revealed a significant reduction in viral load across all the treatments. Notably, the combination of thermotherapy with SA at 50 mg/L and thermotherapy with BTH at 50 mg/L demonstrated the most substantial reduction in viral load compared to other treatments suggesting a potential synergy between the treatments. Furthermore, a noteworthy reduction in CBSD root symptoms severity was observed among the SA and BTH treated plants. Moreover, various plant defense genes were activated by SA and BTH, including but not limited to transcription factors (e.g., WRKY), Leucine Rich Repeat (LRR) proteins, Heat shock proteins (HSP), Mitogen-Activated Protein Kinase (MAPK), Cytochrome P450, ethylene-responsive genes, and Protein TIFY. These findings underscore the promising potential of greenhouse thermotherapy and chemotherapy as measures to mitigate the impact of CBSD. However, it is essential to note that despite the observed reductions in viral load and CBSD root symptoms, the virus persisted across all treatments. This indicates the need for further optimization to combat CBSD effectively. Additionally, it is crucial to delve deeper into understanding the specific functions of genes regulated by these treatments in the context of host-virus interactions.

Keywords: Cassava, CBSD, thermotherapy, chemotherapy, Rwanda.

5.2. Introduction

Cassava is a highly resilient crop to climatic changes and is used as a food and cash crop in sub-Saharan Africa (279). Despite its good performance under adverse conditions, cassava suffers severe yield loss caused by cassava brown streak disease (CBSD). The pandemics of CBSD lead to substantial economic losses every year and food insecurity (35,80,280). The disease is caused by the *cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), collectively designated as cassava brown streak ipomoviruses (CBSIs) (75). Although whitefly vectors can transmit CBSIs in a semi-persistent manner (46), CBSIs are mainly transmitted through stem cuttings as vegetative propagation remains the primary mode of seed multiplication for cassava field establishment (28,79).

Hitherto, no CBSD-resistant cultivar is available, but the effort in breeding for CBSD resistance has resulted in the development of tolerant cultivars. These cultivars are not immune to the disease but can remain asymptomatic or have a delayed onset of root symptoms (6,281,282). While the current CBSD management relies on the widespread adoption of tolerant cultivars, the long-term effectiveness of these cultivars is limited since virus accumulation occurs with each cycle of propagation (6). Hence, it is crucial to implement virus-cleaning methods to address the detrimental impact of viral buildup and the subsequent yield loss. These methods would enable the cleaning of susceptible and tolerant cultivars that continue to be used or have been adopted by farmers because of consumer preferences and pandemics of CBSIs.

Several cassava virus cleaning approaches have been established, including thermotherapy, chemotherapy, and meristematic culture (158,283,284). For instance, *in vitro* chemotherapy with salicylic acid (30mg/L) enabled 100% CBSV elimination from infected cassava plants (136,138). Unfortunately, most established cassava virus cleaning methods are based on expensive *in vitro* techniques that are difficult to implement under greenhouse or field conditions by seed multipliers and farmers.

High temperatures in thermotherapy impact the interaction between virus and plant by modulating RNA silencing (143). Studies have shown that high temperatures can induce mild symptoms and reduce virus load, a phenomenon known as heat masking (143,285,286). For instance, RNA silencing for geminivirus resulted in fewer symptoms, and this effect became even more pronounced with an increase in temperature (287). On the other hand, the chemical inducers used in chemotherapy act by inducing mutations in the virus (e.g., ribavirin) or by inducing signaling pathways involved in disease resistance or RNA interference (e.g., salicylic acid) (148–151).

Transcriptome sequencing is now widely used to improve our knowledge of plant-pathogen interactions, identify critical pathways, and discover genes associated with virus resistance in model and crop species, including *Arabidopsis thaliana* (288), wheat (289), potato (290), banana (291), cassava (213) and many more. Indeed, transcriptomics studies enable the quantitative analysis of plant gene expression changes occurring during various stresses (292). The genetic basis of CBSD resistance remains poorly understood as their infection sometimes results in the induction of none of the typical known resistance gene analogues (293).

Thanks to the release of the cassava reference genome in 2012 (294), numerous transcriptomic and proteomic studies were conducted and provided insights into CBSI-cassava interaction, identifying some potential genes involved in the response against CBSIs (293,295–297). However, cassava genes associated with natural resistance against CBSIs have remained elusive. A recent study found that the phenylpropanoid's PAL-1 (Phenylalanine Ammonia-Lyase) gene could induce CBSIs resistance (296). Another study reported that CBSIs infection induced the expression of genes belonging to secondary metabolites such as terpenoids, phenylpropanoids, and hormone pathways, both implicated in plant resistance (293).

Compounds that induce plant defense responses have been identified as potent resistance elicitors in various pathosystems (298–300). Applying elicitor compounds can activate biosynthetic pathways, producing active oxygen species (AOS). AOS is involved in the hypersensitivity response (HR) that limits pathogen growth and can activate systemic acquired resistance (SAR) in distant plant parts. Furthermore, other resistance mechanisms, including strengthening of the plant cell walls by callose and propanoic compounds, accumulation of pathogenesis-related proteins, and different defense enzymes, have also been reported upon the application of elicitor compounds on plant organs (301,302).

Both salicylic acid (SA) and benzothiadiazole (BTH) display elicitor activities when they are sprayed exogenously on leaves (148). Exogenous application of SA was demonstrated to reduce the severity of clubroot disease by increasing the activities of antioxidant enzymes, reactive oxygen species (ROS) scavenging, and osmotic regulation (303). Another study confirmed that SA is a resistance-inducing elicitor that can enhance tomato plant resistance against *Tomato yellow leaf curl virus* by altering the expression of ROS-scavenging enzymes and inducing the expression of pathogenesis-related genes to produce systemic acquired resistance (304). Moreover, SA was found to induce the expression of genes involved in RNA silencing, such as RdRp1(RNA-dependent RNA polymerase) and AGO2 (argonaute), in both *Arabidopsis thaliana* and *Nicotiana benthamiana* (156,305,306).

Benzothiadiazole (BTH), a synthetic analogue of SA, was also found to have prime resistance to the *turnip crinkle virus* in *Arabidopsis* (307). Likewise, another study showed that an exogenous spray of BTH on pepper plants induced resistance against *Pepper golden mosaic virus* (PepGMV) by activating the SA pathway. However, the effectiveness of the treatment decreased as the time between treatment and virus inoculation increased (153).

Considering the implication of engaging the community in disease management, implementing a straightforward, farmer-friendly approach for virus cleaning holds great promise in reducing yield losses caused by CBSD. The present study evaluated the effectiveness of greenhouse thermotherapy combined with chemotherapy and field chemotherapy for CBSD mitigation. Furthermore, the study explored potential plant defense genes activated in cassava in response to two specific compounds, SA and BTH. Understanding the activation of these genes can provide valuable insights into the mechanisms underlying cassava's defense against CBSD. By shedding light on these aspects, the research aims to offer practical and scientifically-backed solutions for mitigating CBSD.

5.3. Materials and Methods

5.3.1. Field identification of symptomatic experimental plants

Two CBSD-sensitive cassava cultivars, Cyizere and Mushedule, were selected for this experiment. Cassava plants showing clear signs of CBSD infections with a symptom score of 3 were selected from the Rwanda Agricultural and Animal Resources Development Board (RAB) experimental field at the Rubona research station in the South. For CBSIs infection confirmation, pieces of mature leaves were sampled from the mother plants and stored in tubes with 70% ethanol before RNA extraction.

5.3.2. Initial virus infection confirmation

Total RNA was extracted from 16 samples (8 from each of the 2 cultivars) following the protocol of CTAB (308). Extracted RNA were treated with DNase I enzyme to remove cassava genomic DNA according to the manufacturer's instructions (New England Biolabs, Leiden, the Netherlands). The quality, purity, quantity and integrity of RNA was assessed by a Quantus Fluorometer (Promega, Leiden, the Netherlands) and 1% gel electrophoresis. Subsequently, 0.5 µg was subjected to first strand cDNA synthesis by reverse transcription using Superscript III 1st Strand Superscript II Kit following manufacturer instructions (Thermo Scientific, United States). The generated cDNA was subjected to PCR reaction using a Platinum II Hot-Start green PCR Master Mix (2x) (Thermofisher, Lithuania). The primer pair F:5'CCTCCATCWCATGCTATAGACA-3' and R:5'GGATATGGAGAAAGRKCTCC-3' that amplifies ~703bp of CBSV and ~800 bp of UCBSV isolates was used. The PCR reaction contained 5 µL G2 Mix, 0.4 µL each primer, 1 µL cDNA and the volume was brought to 10 µl by adding nuclease free water. PCR conditions were: predenaturation at 95°C for 2min, followed by 30 cycles of denaturation at 94°C for 30", annealing at 56°C for 30" and elongation at 72°C for 50" and final elongation at 72°C for 5 min. To ensure the accuracy of the PCR results by ruling out any false negative results, internal control gene PP2A was detected in parallel using a pair of primers F: 5′-TGCAAGGCTCACACTTTCATC-3' & R: 5′-CTGAGCGTAAAGCAGGGAAG-3′. PCR amplification was checked by loading PCR products in 1% (w/v) agarose gel stained with Gel red in 1X Tris-acetate-EDTA (TAE)buffer for 1hour at 200 V for better separation of the 2 isolates. UV gel documentation system was used to visualize and photograph the PCR products.

5.3.3. Greenhouse experiments: thermotherapy, thermotherapy combined with chemotherapy, and chemotherapy.

In the greenhouse experiments conducted from mid-June to mid-September 2020, stem cuttings from CBSD-infected cassava plants were used, selecting eight cuttings (four from each of the two cultivars). These 2 nodes cuttings (with around 6 cm) were transported to the RAB Rubona Research Station greenhouse. Before planting, the cuttings underwent a disinfection process using 70% ethanol and the fungicide benlate (3.5 g/L) to ensure a clean starting point.

For the thermotherapy experiment (**Figure 19A**), the infected cuttings were placed inside the greenhouse thermotherapy box for up to 3 months. In the combined thermotherapy and chemotherapy approach (**Figure 19B**), infected cuttings were grown in the thermotherapy box and simultaneously sprayed with two potential resistance inducers, benzothiadiazole (BTH) at concentrations of 10 and 50 mg/L (153,309), and salicylic acid (SA) at concentrations of 30 and 50 mg/L (138,151,306). The plants remained in the thermotherapy box until they were 12 weeks old. Throughout this period, the temperature and relative humidity in the thermotherapy box was monitored at four-time points daily: 6:00 AM, 10:00 AM, 2:00 PM, and 6:00 PM. Each treatment group consisted of eight plants, with four biological replicates from each of the two cultivars. Thus, 40 cuttings (two chemicals, two concentrations, thermotherapy, and two cultivars with four biological replicates) were planted in a traditional thermotherapy box (**Figure S8**).

The chemotherapy-only group (**Figure 19C),** had infected cuttings grown outside in an open environment and subjected to the same spraying regimen with SA or BTH. Negative controls were also grown outside in an open environment and sprayed with water. The spraying of treatments began after sprouting (2 weeks after planting [WAP]) and continued thrice a week for 4 weeks. Thus, another set of 40 cuttings was grown outside for the chemotherapy experiment (two chemicals, two levels of concentration, two cultivars with four biological replicates, and controls). The plant's growth vigor was evaluated on a weekly basis, considering parameters like shoot height and leaf count. Samples were gathered from lower, middle, and upper leaves to assess the treatments' efficacy in countering CBSD. This was done through viral quantification at 6 and 12 weeks after planting (WAP) (**Figure 19**).

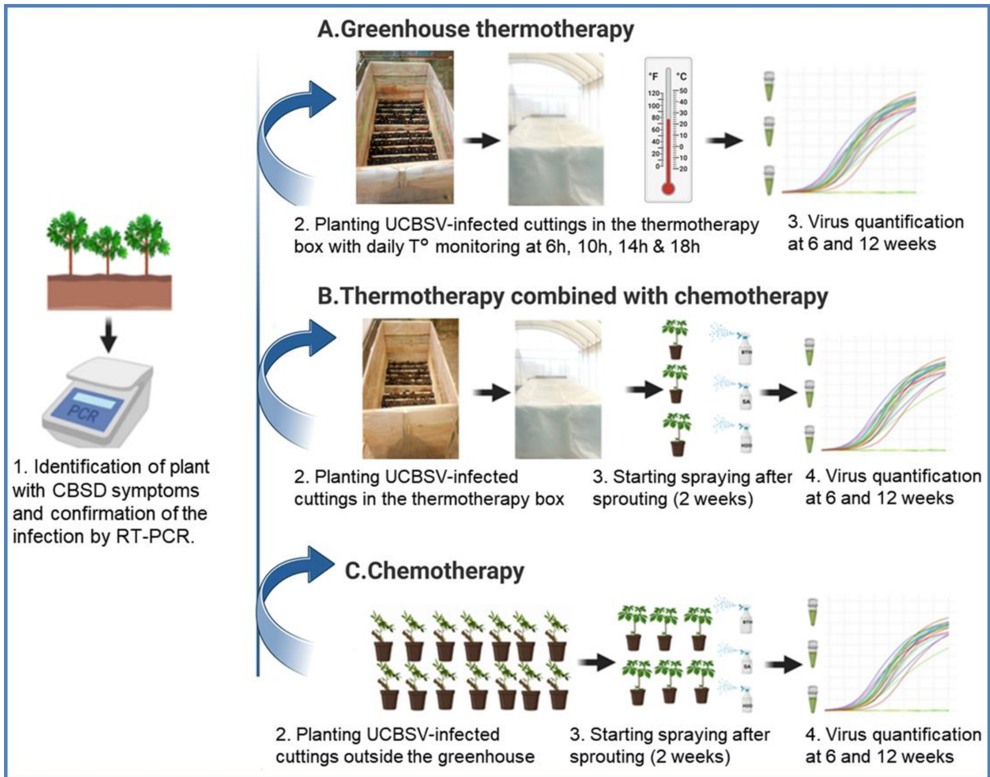


Figure 19. Greenhouse Experimental Design.

A. Thermotherapy Experiment: Cassava plants exhibiting CBSD symptoms were selected from Mushedule and Cyizere cultivars. Leaves samples were collected from these plants and subjected to RT-PCR to confirm the presence of CBSIs infection. The UCBSV infected cuttings were then planted in a specially designed wooden box covered with a plastic bag, referred to as the "thermotherapy box," for up to 3 months. Throughout this period, the temperature inside the thermotherapy box was diligently monitored daily at 6h, 10h, 14h, and 18h. **B. Combined Thermotherapy and Chemotherapy:** Infected cuttings were subjected to the abovementioned thermotherapy. After two weeks of planting, the treatments involving SA at 30 and 50 mg/L and BTH at 10 and 50 mg/L were applied through spraying. This spraying process was repeated thrice a week for 4 weeks. **C. Chemotherapy Only:** Infected cuttings were grown outside in an open environment. Two weeks after planting, the spraying of treatments, including SA at 30 and 50 mg/L and BTH at 10 and 50 mg/L, was initiated. Like the combined approach, the spraying occurred thrice a week for 4 weeks. Infected plants were also grown outside and sprayed with water as a control group. The load of CBSIs was quantified among the treated plants and compared to the controls at six and twelve weeks after planting using real-time RT-PCR to assess the impact of the treatments.

5.3.4. Field chemotherapy.

A field chemotherapy experiment was set where cassava stem cuttings from two CBSD-infected cassava cultivars (CBSD tolerant: NASE14 and CBSD susceptible: Cyizere) were grown in the experimental field of RAB at a spacing of 1 meter by 1 meter from a randomized complete block design from October 2021 till December 2022. Four biological replicates were used for each treatment. Following complete sprouting (3 WAP), plants were sprayed thrice a week with either 1mM of SA or 1mM of BTH for 3 months. At 12 months after planting, CBSD root symptoms were evaluated following the 1 to 5 scale where 1 means no visible root necrosis, 2 means less than 5% root necrosis, 3 means 5 to 20% root necrosis, 4 means 30 to 40% root necrosis and 5 means severe root necrosis greater or equal to 50% (99,310) (**Figure 20**).

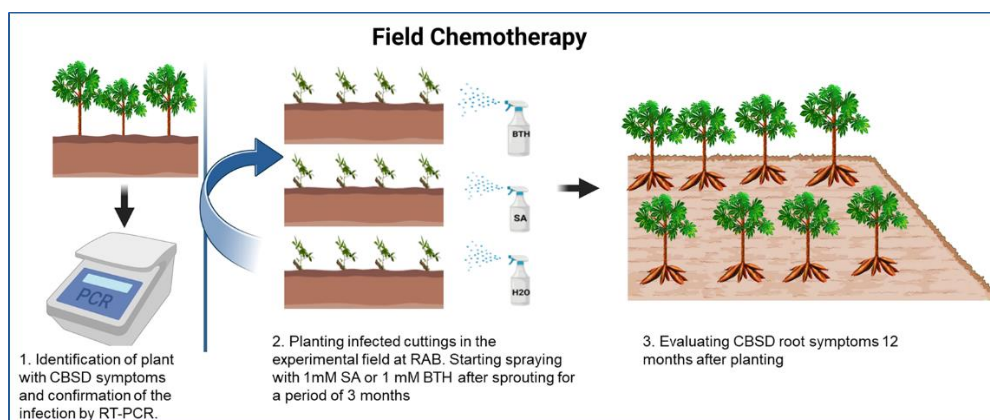


Figure 20. Field chemotherapy experiment.

UCBSV infected cutting of NASE14 and Cyizere were grown in open field following normal standards of cassava farming. After 3 weeks, they were sprayed three times per week for up to 3 months with 1mM of SA or 1mM of BTH and CBSD root symptoms were evaluated at 12 months.

5.3.5. Selection of housekeeping genes

A normalization experiment was conducted using reference genes that are well-documented in the cassava literature to identify stably expressed genes under our experimental conditions. These reference genes include Serine-threonine phosphatase (PP2A), Ubiquitin 100 (UBQ10), vacuolar ATP synthase (vATP), GTP binding (CTPb), and Clp protease ATP-binding subunit (ClpA). The primer sequences for these genes were obtained from Moreno et al.'s work (**Table S13**) (311). The real-time PCR reaction mix consisted of 6 µL of nuclease-free H2O, 10 µL of GoTaq qPCR Master Mix (Promega), 2 µL of primers (at a final concentration of 1 µM), and 0.2 µg of cDNA template, resulting in a total volume of 20 µL. Amplification commenced with an initial denaturation step at 95°C for 20 seconds, followed by 40 cycles

involving denaturation at 95°C for 3 seconds, annealing at 60°C for 15 seconds, and extension at 72°C for 30 seconds. Amplification and detection were performed using the StepOnePlus Real-Time PCR Detection System (Applied Biosystems, Belgium). Gene expression stability analysis was performed in geNorm, NormFinder and BestKeeper programs. The geNorm calculates the average pairwise expression ratio to evaluate expression stability. The lower the M value, the more stable the expression of the reference gene, and values of M that surpass the cutoff value of 1.5 are not considered stable across treatments (312).

NormFinder has a similar mathematical model to GeNorm, where genes with low stability values (SV) are considered more stable (313), and BestKeeper program evaluates the gene expression stabilities by the standard deviation (SD) of their Ct values. The lower the SD value, the higher the expression stability (314). The sum score of the 3 software was assigned where the most stable genes had the lowest score and the most unstable had the highest score (**Table S14**). The optimum number of genes for accurate normalization was selected according to the pairwise variation V from geNorm (**Figure S9 A and B**).

5.3.6. Quantification of UCBSV in treated plants

Leaf samples were collected, total RNA extracted by CTAB method (308), and cDNA synthesized as previously described (section 5.3.2). For quantification of UCBSV, a 20 µL qPCR reaction mixture contained 10 µL of GoTaq qPCR Master Mix (Promega), 2 µL of primers (with a final concentration of 1 µM), 6 µL of distilled water, and 0.2 µg of cDNA template. Amplification was detected using the StepOnePlus Real-Time PCR Detection System (Applied Biosystems, Belgium). The PCR thermal cycling procedure followed this sequence: An initial denaturation step lasting 20 seconds at 95°C, followed by 40 cycles of denaturation at 95°C for 3 seconds, annealing at 60°C for 15 seconds, and extension at 72°C for 30 seconds. The relative quantification method was applied whereby the generated Ct (cycle threshold) values were used to determine the fold change in the virus gene expression.

5.3.7. Statistical analysis

The relative virus titers were calculated using 2-ΔΔCt method for relative normalized expression analysis using appropriate reference genes. A *t-test* was used to compare treatments with the calibrator. To assess the effect of the treatment on the plant growth, independent student *t-test* was used to compare shoot height and number of leaves between control and treated plants.

5.3.8. Transcriptomic experimental design

The uninfected susceptible cultivar 60444 was multiplied *in vitro* at the Plant Genetics and Rhizosphere Processes laboratory (Gembloux Agro-Biotech, Belgium) using cassava basic media (CBM) (4.4 g/L of CBM media were mixed with sucrose 20g/L, 2mMCuSO4 1mL/L and the PH adjusted to 5.8 by NaOH or HCl). At 4 weeks old, the plantlets were transferred to a greenhouse under 27°C, 16h light, 75% humidity. After 4 weeks, the plantlets were sprayed with salicylic acid (1mM) or benzothiadiazole (1mM). Plants sprayed with distilled water were used as the negative

controls. After 24 hours, the leaves samples were collected from the four biological replicates and used for RNA extraction followed by RNASeq (**Figure 21**).

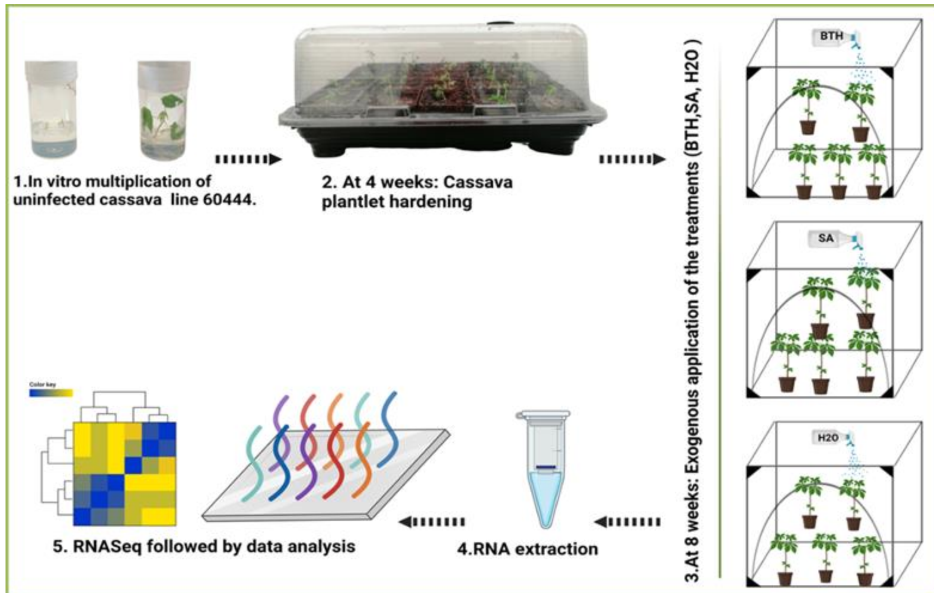


Figure 21. Transcriptomic experimental design.

Uninfected cassava susceptible cultivar 60444 were first multiplied invitro and transferred into soil at four weeks. At eight weeks old, plants were sprayed with either 1mM SA or 1mMBTH. Four biological replicates were used. Control plants were treated with water. Twenty-four hours post treatment, leaves samples were collected for RNA extraction and RNAsequencing.

5.3.9. RNAseq and bioinformatic data analysis

Total RNA was extracted as previously described (308), and only RNA samples with an RNA integrity number (RIN) > 6 were used to perform RNASeq at the GIGA sequencing facility (ULiege). According to the manufacturer's instructions, the Illumina stranded mRNA library preparation kit produced high-quality sequencing libraries (Illumina, California, USA). The sequencing was done using an S4 Flowcell on a Novaseq instrument (Illumine, California, USA), which generated paired-end reads of 150 nucleotides in length each (2*150bp).

The quality of the resulting sequences was assessed using fastqc (version 0.12.0) (315) and the sequences were filtered by quality, only reads with >90% of nucleotides with a score equal to or higher than 34 were retained. The adapters were trimmed using Trimmomatic (version 3.31, default parameters) (316). The paired-end reads from each sample were mapped to the latest version of the cassava reference genome from Phytozome (https://phytozome-next.jgi.doe.gov/info/Mesculenta_v8_1) (International Cassava Genetic Map Consortium, 2015) with HISAT2 (version 2.1.0, default parameters (317). Transcripts quantification was achieved using HTSeq-count

with default parameters (version 2.0.1) (318). The differential expression analysis was performed with the R (R version 3.2.2, (http://www.r-project.org/) package DESeq2 (319). Differentially Expressed Genes (DEGs) were then filtered based on a False Discovery Rate (FDR) -corrected P-value of <0.01 and an absolute fold–change of ≥2. Gene set enrichment analysis with the PlantGSEA Toolkit (http://bioinformatics.cau.edu.cn/PlantGSEA/analysis.php) (320), and PlantRegMap (http://plantregmap.gao-lab.org/go.php) (321), allowed the identification of enriched Gene Ontology (GO) categories. Overrepresented GO categories were identified among the upregulated and downregulated genes for SA and BTH separately.

5.4. Results

5.4.1. Initial CBSIs confirmation in experimental plant

Four plants confirmed to be infected with UCBSV were selected from the two cassava cultivars: Mushedule and Cyizere, for thermotherapy and chemotherapy experiments (**Figure S10**).

5.4.2. Greenhouse thermotherapy combined with chemotherapy reduced the viral load.

The weekly mean variation in temperature and relative humidity inside the thermotherapy box was calculated. The highest weekly mean temperature of 42.2°C (+/- 2.5°C) was recorded at 14:00, while the lowest mean temperature of 23.5°C (+/- 1.78°C) was observed at 6:00 (**Figure 22**). For the control group grown outside the greenhouse, the mean temperature ranged from 17.6°C (+/- 1.4°C) to 28°C (+/- 1.13°C) (**Figure S11**). Furthermore, the weekly mean variation in relative humidity inside the thermotherapy box varied between 74 and 85 %, while in the controls, it varied between 67 and 75% (**Figure S12**).

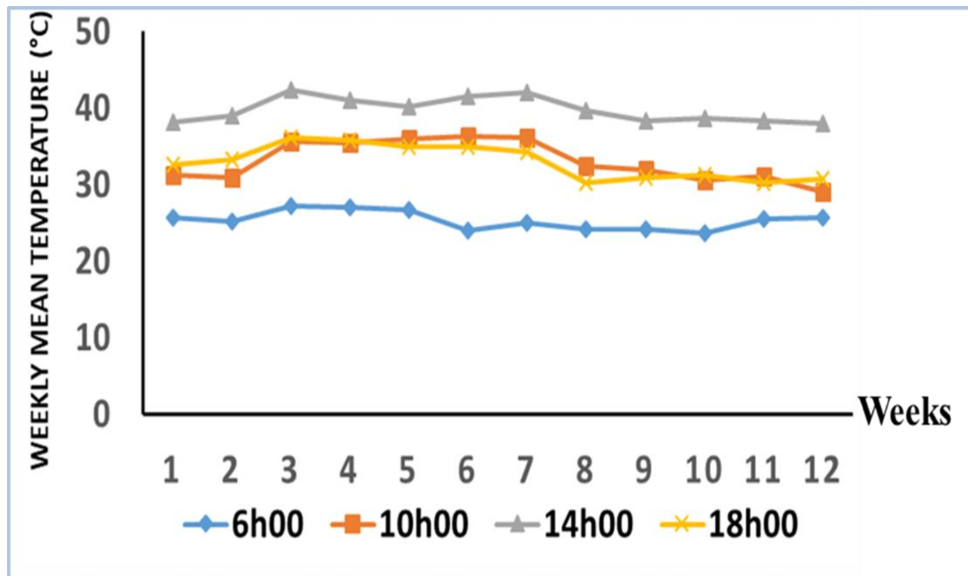


Figure 22. Weekly mean variation in temperature inside the thermotherapy box. The temperature inside the thermotherapy box was daily recorded at 6, 10, 14 and 18h00.

Evaluation of treatment impacts on UCBSV loads was performed by RT-qPCR at 6- and 12-weeks post treatments. Quantitative data analysis revealed a statistically significant decrease in UCBSV load in all the treatments. Noteworthy, the combination of thermotherapy with SA at 50 mg/L and thermotherapy with BTH at 50 mg/L demonstrated the most substantial reduction in viral load compared to other treatments. Furthermore, these combined treatments exhibited more significant reductions in viral load than using thermotherapy, SA, or BTH alone, suggesting a potential synergy between the treatments (**Figure 23**). Interestingly, the treatments used did not affect the plant growth as the independent *t-test* for shoot heights and number of leaves under various treatments has a p-value ranging from 0.16 to 0.93 (**Table S15**).

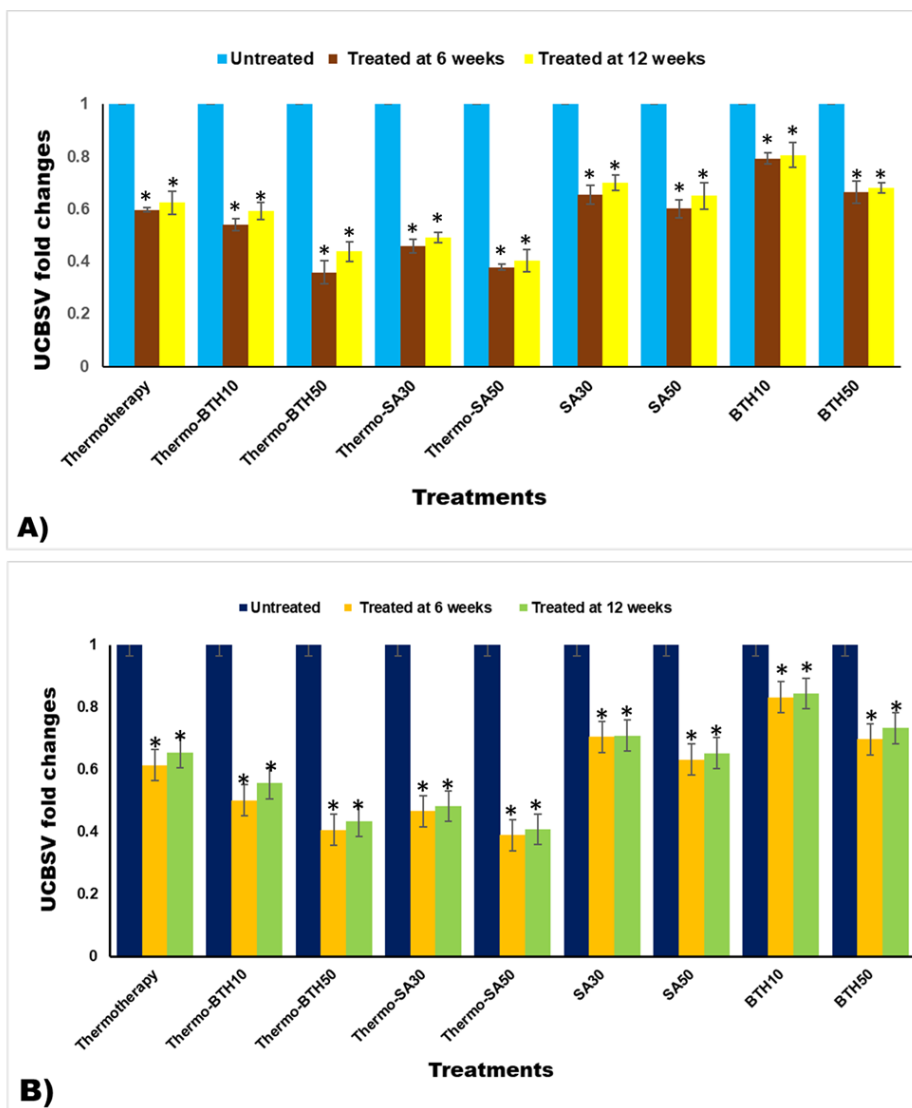


Figure 23. UCBSV quantification at 6- and 12-weeks post treatments. **A)** Mushedule and **B)** Cyizere varieties. Data are mean ± SD of 3 or 4 biological replicates. Statistical significance according to the *t-test* is represented by *, $P < 0.01$.

5.4.3. Field chemotherapy reduced the severity of CBSD root symptoms.

CBSD root necrosis was evaluated among the plant treated with either SA or BTH in comparison to the untreated (control). Interestingly, the sprayed plant developed less severe CBSD root necrosis than the control (**Table 8; Figure 24**).

Table 8. The severity of CBSD root symptoms at 12 MAP following field chemotherapy.

	Severity score for CBSD root symptoms _ Cyizere cultivar		
Independent *t-test*	**Control_Cyizere**	**Cyizere_SA**	**Cyizere_BTH**
Mean	3	2.06	2.31
Std deviation	0	0.11	0.1
t value		4.62	3.19
p-value		**<0.001***	**0.003***
	Severity score for CBSD root symptoms _ NASE14 cultivar		
Independent *t-test*	**Control -Nase14**	**Nase14_SA**	**Nase14_BTH**
Mean	2.75	1.75	1.81
Std deviation	0.17	0.18	0.27
t value		4.39	3.66
p-value		**<0.001***	**<0.001***

STD: standard deviation; * significant

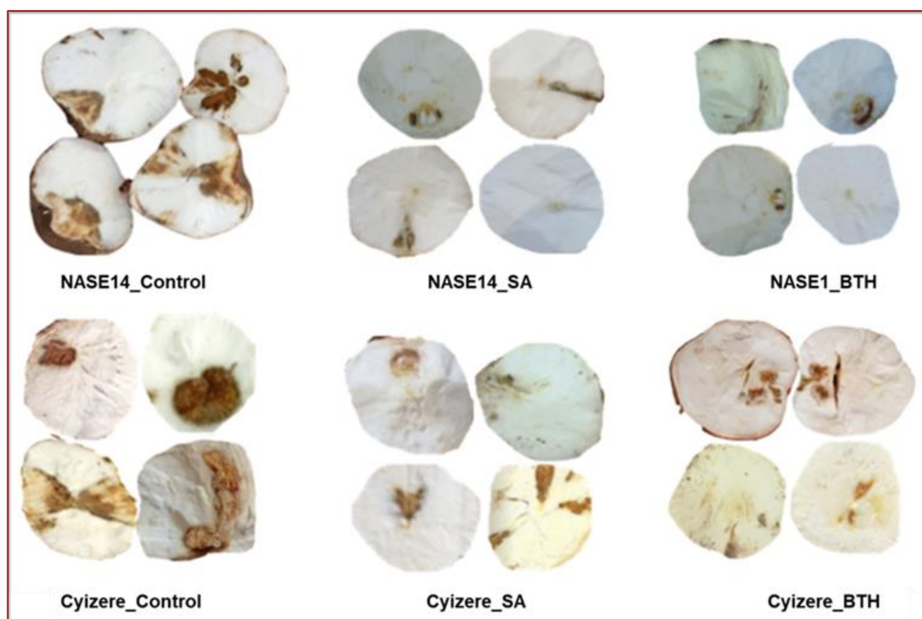


Figure 24. Categories of CBSD root necrosis expressed at 12 MAP following field chemotherapy with SA or BTH.

Two varieties NASE14 and Cyizere were sprayed with 1mM SA or 1mM BTH for 3 months while the controls were sprayed with water.

5.4.4. Overview of cassava transcriptome 24 hour after SA or BTH spraying.

A total of 17 to 23 million paired reads per sample were mapped against the cassava reference genome. The rate of successful mapping to reference (*M. esculenta*_671_v8.1.) was over 80% of the total reads, and genome coverage ranged between 8.4 and 10.9-fold (**Table S16**). The principal component analysis (PCA) was used to characterize the variation between samples, which revealed good clustering of the biological replicates used per treatment. The first 2 axes of PCA, accounting for 56% and 23% of the total variance, respectively, separated the datasets in terms of the treatment, indicating that the experimental conditions induced actual gene expression differences between groups (**Figure S13**). DESeq analysis revealed that many genes were deregulated following SA and BTH treatment (**Table S17**). The plant treated with SA had the largest number of deregulated genes (639) compared to the plant treated with BTH (370) (Log2FC $\geq$1; adj. p-value $\leq$ 0.01). SA downregulated more genes (437) compared to BTH (115), whereas BTH induced slightly more genes (255) than SA (202). Both treatments commonly upregulated 96 differentially expressed genes (DEGs), while both treatments commonly downregulated 58 DEGs. Noteworthy, ten DEGs downregulated by SA were induced by BHT (**Figure 25**).

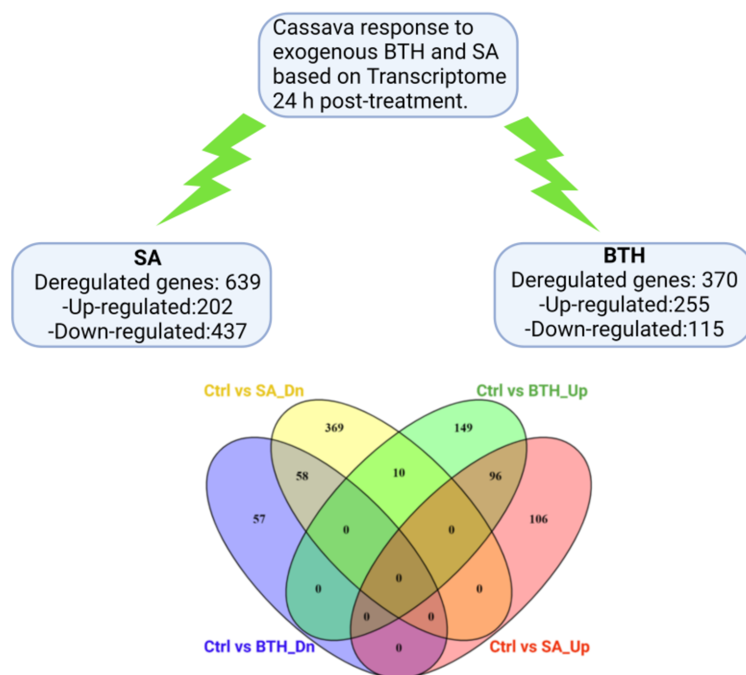


Figure 25. Overview of the RNAseq data obtained post-SA and BTH treatment.

from 60444 cassava line at 24 hours post exogenous treatment by 1mMSA or 1mM BTH. Differentially Expressed Genes among treatments and their commonalities at P-value $\leq$0.01; Log (FC) > |1|

5.4.5. Comparative analysis of DEGs reveals overlaps between SA, BTH treatments, and CBSV infection in cassava cultivars.

Among the DEGs induced by SA and BTH, 35% and 36% had *Arabidopsis thaliana* (AT) orthologs, respectively, while among the DEGs repressed by SA and BTH, 33% and 30% had AT orthologs, respectively. The AT orthologs from the SA and BTH DEGs were extracted and used for comparison with the DEGs deregulated by CBSIs infection in different cassava cultivars such as CBSIs-tolerant like Kaleso (293), KBH (295), and Namikonga (297); and CBSIs-sensitive like Albert (293) and 60444 (295).

Our findings revealed significant overlaps between DEGs deregulated by BTH and SA treatments and DEGs deregulated by CBSIs infection in various cassava cultivars. It was noted that 24% and 33% of the DEGs prompted by BTH and SA treatments, respectively, exhibited a similar response in the CBSIs-tolerant Namikonga cultivar two days after UCBSV infection via grafting (297) (**Figure S14A**). Furthermore, 76% and 14% of DEGs downregulated by SA and BTH were also repressed in infected Namikonga (297) (**Figure S14B).** In the Albert cultivar, only two genes were deregulated two days post-UCBSV infection by grafting; thus, no overlapping was found (**Figure S14A and B**).

For the CBSIs-tolerant Kaleso cultivar (293), it was observed that 8% and 6% of the DEGs induced by SA and BTH, respectively, overlapped with DEGs induced by CBSIs infection (**Figure 14C**). Additionally, 28% and 6% of the DEGs repressed by SA and BTH, respectively, overlapped with the CBSIs downregulated DEGs in Kaleso (**Figure S14D**).

Conversely, it was observed that 15% and 10% of the DEGs induced by SA and BTH overlapped with the upregulated DEGs in CBSIs infected 60444. In contrast, 7% and 5% of DEGs induced by BTH and SA treatments, respectively, overlapped with the DEGs induced by CBSIs infection in the resistant KBH cultivar (295) (**Figure S14E**). Nevertheless, 36% and 10% of the DEGs down-regulated by SA and BTH, respectively, exhibited an overlap with the DEGs repressed by CBSIs infection in the resistant KBH cultivar. In comparison, 18% of the DEGs repressed by SA exhibited an overlap with CBSIs-induced DEGs in infected 60444 (**Figure S14F**).

5.4.6. Important GO terms in cassava defense were enriched by SA and BTH treatments.

GO terms enrichment analysis revealed that biological processes potentially involved in plant pathogen defense were enriched within the upregulated genes following either SA or BTH treatment. The enriched GO biological process among the commonalities between genes upregulated by SA and BTH included hormone-mediated signaling pathway (GO:0009755), response to stimulus (GO:0009607), response to stress (GO:0006950), defense response (GO:0006952), response to biotic stimulus (GO:0009607), etc. (**Figure 26**). The detailed enriched GO terms within BTH and SA-deregulated genes are presented in **Table S18.**

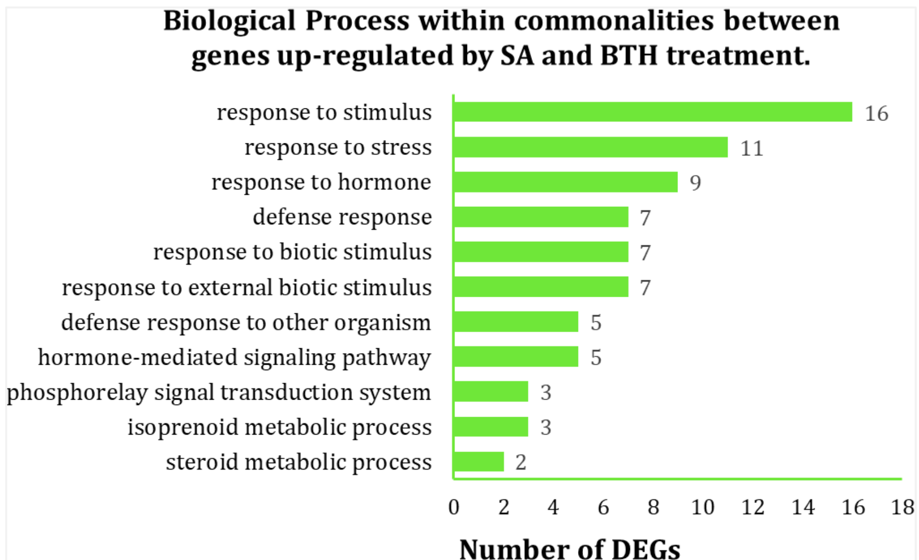


Figure 26. Some over-represented GO terms of the category "biological process" among DEGs induced by SA and BTH in 60444 cultivars at 24-hour post SA and BTH treatment.

Furthermore, a set of 132 genes potentially associated with plant defense mechanisms were manually selected from previously published papers to construct the dataset for potential CBSIs responsive genes (213,296,297,322–327) (**Table S19**). After creating the dataset, a comparison was made with the DEGs induced by SA and BTH treatments. Interestingly, 60 of 147 genes overlap between the dataset and the DEGs induced by these treatments.

Among the overlapping genes, 17 were common to BTH-induced and SA-induced upregulated genes. In contrast, 10 genes were shared between the dataset and the upregulated genes induced by SA, and 16 genes were identified as shared between the dataset and the upregulated genes induced by BTH treatment (**Table S19**). Notably, the commonly upregulated transcripts included genes already observed in previous research as potential players in plant defense mechanisms. These genes fall into various categories, such as transcription factors (e.g., WRKY), Leucine Rich Repeat (LRR) Proteins, Heat shock Proteins (HSPs), Mitogen-Activated Protein Kinases (MAPKs), Cytochrome P450 enzymes, and TIFY protein.

The GO analysis of the common upregulated DEGs, shared between the dataset of potential CBSIs responsive genes and those upregulated by BTH or SA treatments, provided valuable insights into the potential biological processes involved in plant defense. A summary of these processes is presented in **Figure 27**, while the complete list of enriched GO terms is available in **Table S19**. However, no significant intersection was observed between the genes down-regulated by SA and BHT treatments and the CBSIs dataset.

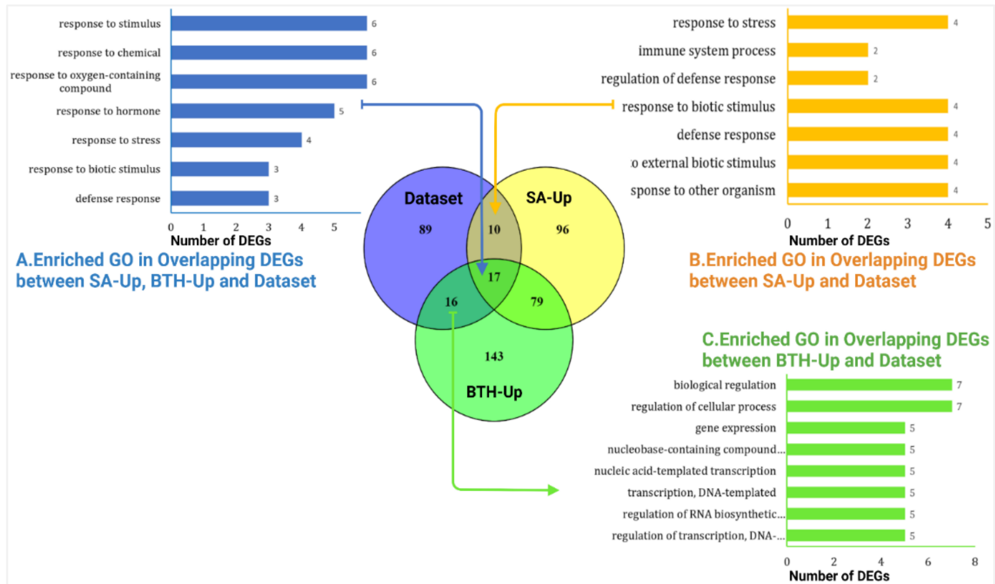


Figure 27. Over-represented GO terms in the category "biological process" for overlapping DEGs between genes up-regulated by SA and BTH and potential CBSIs responsive genes from the literature.

The dataset of potential CBSIs responsive genes created from the literature (213,296,297,322–327).

5.5. Discussion

The present study marks the pioneering effort in mitigating CBSD through a multi-pronged approach, integrating greenhouse thermotherapy with chemotherapy and field chemotherapy, employing SA and BTH. Remarkably, all treatments led to a significant reduction in UCBSV loads. Combining thermotherapy with SA and BTH at 50 mg/L showed better promising results. Moreover, field chemotherapy also contributed significantly to a notable decrease in CBSD root symptoms.

The impact of high temperature on plant virus interaction has long been reported when Walkey *et al*. noticed that cucumber and alfa alfa viruses were eliminated from *Nicotiana rustica* after 30 days at 32°C (142). Furthermore, it was proved in a study conducted by Szittya *et al*. (2003) that the plant's defense mechanism, known as RNA silencing, is influenced by temperature. Higher temperatures trigger its activation while lower temperatures hinder both virus-induced and transgene-induced RNA silencing, resulting in increased vulnerability to viruses and the loss of silencing-mediated transgenic phenotypes (143).

Previous studies on cassava virus cleaning using *in vitro* thermotherapy had reported 50% of virus-free plantlets at 40°C for three weeks (137) and 40% for 4 weeks at temperatures between 36 and 40°C (284). In the present study, UCBSV-infected cuttings were exposed to high temperatures in the box, which varied between 25 and

42°C for three months compared to control cuttings that were grown at ambient temperature in an open environment (18-29°C), indicating its contribution in the observed virus decrease. A field evaluation of hot water thermotherapy on CMD-infected cassava stem treated at 49°C hot water for 30 minutes revealed a significant symptom decrease and maintained the same yield production as a healthy plant (283).

Our data showed that foliar application of salicylic acid on cassava resulted in a statistically significant reduction of UCBSV load and CBSD root symptoms. This could be explained by the role played by the salicylic acid pathway in plant defense against biotic stress. Indeed, Salicylic is known to induce viral resistance by inhibiting viral replication and intercellular movement (328–330). A study conducted by Tian *et al.* (2015), confirmed that SA inhibits the replication of *Tomato bushy stunt virus* (TBSV) by directly inhibiting a host factor called glyceraldehyde 3-phosphate dehydrogenase (GAPDH) essential in TBSV replication (331). Nevertheless, this factor was not significantly induced by SA or BTH, which indicates that there might be a different mechanism of action for CBSIs.

Application of benzothiadiazole on infected cassava plants also resulted in significant UCBSV loads and root symptom decrease. BTH is known to have broad-spectrum activity against a wide range of diseases. For example, BTH application induces resistance to various viruses and pest including *pepper golden mosaic virus* (153), *Tobacco Mosaic Virus* (332), yellow strain of *Cucumber mosaic virus* (333), Citrus Exocortis Viroid (CEVd)(334), and *Bemisia tabaci* (335) without harming either the productivity or the quality of the fruit (153,332–335).

The chemotherapy's effectiveness in reducing viral load increased with higher concentrations, as demonstrated by the lower viral load observed at higher concentrations. This finding aligns with a study by Mwangagi *et al.* (2014). Nonetheless, it is essential to be cautious in order to prevent potential phytotoxicity that could arise from elevated concentrations, potentially harming the host plant. In the current study, an essential observation was made regarding the synergistic effect of thermotherapy when combined with either SA or BTH. This combination led to a significantly greater decrease in viral load than individual treatments. These findings are consistent with previous studies on other crops. For example, previous *in vitro* investigations have reported higher efficiency in eliminating *Potato Virus Y* through simultaneous thermotherapy and chemotherapy (336). A noteworthy distinction in the present study was the absence of phytotoxicity, which contrasts with the results reported by Mwangagi *et al.* (2014). In their study, *in vitro* plants treated with thermotherapy combined with chemotherapy experienced total mortality. The lack of phytotoxicity in our study could be attributed to the variation in temperature recorded within the thermotherapy box used, which provided a protective environment that prevented the plants from being exposed to excessively high temperatures. Furthermore, it is essential to consider that *in vitro* plantlets, as utilized in the study by Mwangagi *et al.* (2014), are more delicate and vulnerable than the cuttings employed in our current study. This disparity in plant material could contribute also to the differences in the observed outcomes.

RNASeq data analysis revealed that 33 and 27 DEGs induced by BTH and SA, respectively (including 17 common DEGs between SA and BTH treatments) were

already reported to be implicated in viral defense such as transcription factors, LRR-containing genes, Heat shock Protein, Mitogen-Activated Protein Kinase (MAPK), Cytochrome P450, Ethylene responsive transcription factor, Protein TIFY, and NSP Interacting Kinase. In numerous RNA sequencing and proteomics-based investigations, performing an independent validation of the generated data is customary. This validation is typically done through quantitative PCR or western blot analysis to ascertain the results' accuracy and robustness (337,338). However, it is noteworthy that numerous studies have consistently demonstrated a robust correlation between RNA-seq data and qPCR results (339,340). This correlation suggests that validating RNA-seq data may not be essential when the experimental design rigorously follows state-of-the-art protocols and includes a sufficient number of biological replicates (337). Thus, in the present study, differentially expressed genes were not subjected to independent validation based on an adequate number of biological replicates that was applied.

The analysis of DEGs induced by SA and BTH revealed enrichments in various Gene Ontology (GO) categories with potential roles in the viral host response, such as hormone signaling, defense response, stress response, and transcriptional regulation. These findings enhance our understanding of the genetic responses triggered by the BTH and SA treatments. They could guide further research into the specific molecular pathways and mechanisms underlying cassava defense against CBSIs, and such knowledge could aid in developing targeted strategies to combat CBSD.

It was observed from previous cassava transcriptomic study done on cassava cultivar challenged by UCBSV that the LRR domain, TFs (e.g. WRKY, NAC), and heat shock protein contributes to cassava resistance against UCBSV (297). Other studies have also highlighted that WRKY TFs are the major regulators of immune responses and signalling involving pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) (325). For instance, in cassava, WRKY TFs (e.g. MeWRKY81) were found to be involved in the regulation of CMD tolerance or susceptibility (341). In rice, the WRKY gene conferred resistance against fungal blast caused by *Magnaporthe grisea* (342). Another study has provided further evidence of the role of TIFY4B in the host's defense pathways against geminiviruses, with its increased expression leading to the inhibition of virus replication (324), and their involvement in RNA virus defense cannot be ruled out.

Furthermore, exogenous application of SA was reported to cause the accumulation of pathogenesis-related (PR) proteins and induce systemic acquired resistance against *Tobacco Mosaic Virus* (TMV) (148). Thulke and Conrath (1998) reported that pre-treatment with salicylic acid or functionally related inducers like benzothiadiazole can induce the generation of hydrogen peroxide and expression of defense-related genes such as PAL or 4 coumarate CO-A ligase (Thulke and Conrath, 1998). In our study, spraying of SA or BTH upregulated 4-coumarate-CoA ligase (4CL: Manes.02G082000) 2.4 and 3-fold for BTH and SA, respectively. The 4CL plays a role in plant secondary compounds biosynthesis at different points from general phenylpropanoid metabolism (343). Likewise, the activity of peroxidase enzymes implicated in SA-induced defense response was upregulated. Peroxidases are

involved in the metabolism of ROS and reactive nitrogen species (RNS), biosynthesis of phytoalexins and switching on the hypersensitivity response (322). Furthermore, the cytochromes P450, which are involved in plant protection against biotic stress by synthesis of plant defense compounds such as terpenoids, alkaloids, etc. and by regulating plant hormones (344), were induced by both SA and BTH treatments. On the other side, foliar applications of salicylic acid (0,5-1mM) enhanced the resistance to *Tomato yellow leaf curl virus* (TYLCV) in tomato plants and the expression of RNA-silencing-related genes, such as DCL1, DCL2, DCL4, RDR2, RdRp1, RDR3a, RDR6a, AGO1, AGO2 and AGO4, were significantly triggered by its exogenous application (155,156). However, in the current study, none of them was differentially induced. A similar observation was reported by Maruthi *et al*. (2014) after analyzing RNseq data from infected cassava cultivars 12 months post-grafting.

It could be explained by the multigenic nature of CBSD resistance, which is often controlled by recessive genes, as it was proved that inbreeding enhance resistance to CBSD (345). Moreover, it may indicate that other unidentified CBSD resistance genes are induced by SA and BTH treatments. In the present study, it was observed that treatment with SA downregulated ankyrin protein (ANK) (Manes.07G134100) by 5.2-fold, suggesting a potential delay in viral movement within the plant due to SA treatment. ANK protein was found to be involved in the movement and infection process of tobacco viruses, as its repression resulted in a notable reduction in viral infection. In contrast, its upregulation enhanced viral infection (346). Additionally, apart from the genes previously identified as being involved in CBSV defense, numerous other genes were observed to be activated by SA and BTH. However, their specific involvement in the antiviral response remains to be studied and confirmed.

The findings from the current study showed that greenhouse thermotherapy and chemotherapy using salicylic acid and benzothiadiazole could significantly benefit farmers as it resulted in a notable reduction in viral load and alleviation of CBSD root symptoms. Nevertheless, attaining the necessary high temperatures inside the greenhouse thermotherapy box might present challenges during colder seasons, as external weather conditions can influence the temperature. Hence, strategically implementing this approach during the summer months, in preparation for the September planting season, becomes crucial for achieving optimal results. In addition, it is imperative to exercise caution and precision in selecting elicitors, considering critical factors such as the timing, concentration, and frequency of application to the specific crop under treatment to achieve the best possible outcomes.

To further advance this promising approach, additional research efforts are necessary. Optimizing the protocol to achieve complete or maximum virus reduction is paramount. Additionally, investigating the specific functions of the genes deregulated by the treatments will provide valuable insights into the underlying mechanisms. Furthermore, a cost-benefit analysis to confirm the proposed approach's affordability would enable their adoption by farmers. These comprehensive research steps are essential before the approach could be confidently integrated into CBSD management strategies.

5.6. Supplementary materials

Table S 13. Primer sequences used in chapter 4.

Primer name	Sequence 5' => 3'	Analysis	Amplicon size (bp)	References
CBSV-UCBSV-F	CCTCCATCWCATGCTATAGACA	RT-PCR	CBSV-703	(Elegba, 2018)
CBSV-UCBSV-R	GGATATGGAGAAAGRKCTCC		UCBSV-800	
PP2A-F	TGCAAGGCTCACACTTTCATC	Normalization	187	(Moreno, Gruissem and Vanderschuren, 2011)
PP2A-R	CTGAGCGTAAAGCAGGGAAG			
UBQ10-F	TGCATCTCGTTCTCCGATTG		166	
UBQ10-R	GCGAAGATCAGTCGTTGTTGG			
vATPs-F	GTTGAATGGCTTTGTGCTCAG		128	
vATPs-R	CCATCTGCGTGAACAAAAGAA			
GTPb-F	GTTGCCTTCTTTTGCGTTTCT		114	
GTPb-R	GCAATTTGATCCGTTTTCCAT			
ClpA-F	GCCAAGATCCTATACGGCAAG		109	
ClpA-R	TTTCTGGCTCTGGTTCTCCAT			
UCBSV-F	CCATTTGAGGCTAGGAGATTGA	RT-qPCR	157	(Vanderschuren *et al.*, 2012)
UCBSV-R	ACTTCCCCATCATCTGGTTCTC			

Table S 14. Selection of stably expressed genes under thermotherapy and chemotherapy

The reference genes were chosen after ranking in the geNorm, NormFinder, and BestKeeper. The most stable genes were identified based on the total score obtained by all the program. The genes in yellow are the most stable genes (lowest score) and were chosen for normalization.

Thermotherapy combined with BTH					
Genes	geNorm	NorFinder	BestKeeper	Score	Ranking
Ubiquitin 10	5	5	5	15	5
PP2A	4	2	4	10	4
GTPb	3	1	3	7	2
ClpA	2	3	2	7	2
vATP	1	4	1	6	1
Thermotherapy combined with SA					
Genes	geNorm	NormFinder	BestKeeper	Score	Ranking
Ubiquitin 10	4	5	5	14	5
PP2A	5	1	2	8	2
GTPb	1	3	4	8	2
ClpA	3	4	1	8	2
vATP	2	2	3	7	1

Table S 15. Independent *t* test for comparison of control and various treatments in terms of t shoot heights and number of leaves for Cyizere and Mushedule cultivars.

Independent Samples Test	Control	Thermotherapy	Thermo + SA30	Thermo + SA50	Thermo + BTH10	Thermo + BTH50	SA50	SA30	BTH10	BTH50
Shoot heights (in cm) following various treatments for Cyizere variety										
Mean	47,11	50,11	54,91	54,32	53,05	55,48	43,16	49,07	43,11	46,36
Std deviation	20,54	18,94	22,81	22,55	22,29	19,24	17,17	19,33	17,94	20,53
t value		0,459	-0,842	-0,783	-0,649	-0,162	0,49	-0,23	1,459	0,086
p value		0,16	0,41	0,443	0,524	0,873	0,63	0,821	0,16	0,933
Shoot heights (in cm) following various treatment for Mushedule variety										
Mean	48,24	51,48	48,09	47,85	50,48	45,48	40,06	41,12	44,5	40,86
Std deviation	20,56	20,13	20,71	21,63	22,13	20,47	19,93	18,7	21,24	20,02
t value		0,183	0,472	0,599	0,193	0,888	1,18	0,85	1,093	0,853
p value		0,749	0,642	0,556	0,849	0,385	0,252	0,405	0,287	0,404
Leaf number following various treatment for Cyizere variety										
Mean	10,26	10,05	9,28	7,98	9,7	8,55	9,63	11,48	8,9	8,51
Std deviation	4,55	3,91	2,51	3,39	3,13	2,96	3,64	7,65	3,07	3,91
t value		0,213	0,121	1,333	0,338	1,043	0,362	-0,454	0,824	0,121
p value		0,801	0,905	0,198	0,739	0,309	0,721	0,655	0,42	0,905
Leaf number following various treatment for Mushedule variety										
Mean	7,94	8,62	9,75	8,05	8,16	7,62	8,79	9,01	8,15	9,46
Std deviation	2,37	2,73	3,51	3,68	3,32	2,83	2,97	6,11	3,21	4,74
t value		0,30	0,286	-0,089	-0,181	0,286	-0,745	-0,543	-0,181	-0,955
p value		0,40	0,778	0,93	0,858	0,778	0,465	0,593	0,858	0,351

Table S 16. Overview of the RNA sequencing data and Mapping.

ID	Sequencing	HISAT2			Genome coverage	HTSeq Count							
	N° of Paired reads	Mapped reads (%)	# mapped reads	# Unmapped		total_ features	% feature	no_feature	ambiguous	too_low_Q uality	not_aligned pairs	not_aligned single	alignment_n ot_unique
1	19580039	83.3	16310172	3269867	9.2	1.5E+07	89.8	1394176	1281520	544104	6497695	3248848	425203
2	17868993	85.2	15224382	2644611	8.4	1.4E+07	88.9	1304210	937.06	344121	6625317	3312659	380.24
3	18962351	84.1	15947337	3015014	8.9	1.4E+07	89.3	1426017	1159664	483581	6348863	3174432	404.13
4	23231501	85.4	19839702	3391799	10.9	1.8E+07	89.6	1709845	1395251	557868	7123938	3561969	501278
5	19055120	82.7	15758584	3296536	8.9	1.4E+07	89.3	1383522	1224852	526668	6643857	3321929	429228
6	18852031	84.1	15854558	2997473	8.8	1.4E+07	88.9	1426608	1130172	465575	6445899	3222950	420835
7	19310556	82	15834656	3475900	9.1	1.4E+07	88.6	1431232	1068589	458742	6884517	3442259	433705
8	20930467	84.9	17769966	3160501	9.8	1.6E+07	89.1	1567464	1227319	405976	6996126	3498063	456888
9	20549667	85	17467217	3082450	9.6	1.5E+07	88.6	1502338	869233	346934	7882244	3941122	451718
10	19424918	84	16316931	3107987	9.1	1.5E+07	88.9	1435388	1122370	452081	6823108	3411554	440417
11	21020809	85.4	17951771	3069038	9.9	1.6E+07	88.7	1557358	1040279	430248	7877643	3938822	456453
12	22607424	83.8	18945021	3662403	10.6	1.7E+07	88.9	1672236	1205291	487374	7749155	3874578	506448

Table S 17. DESEq output for all Genes.

Table S 18. GO Enrichment Analysis within DEGs deregulated by BTH and SA.

Table S 19. Dataset for Potential CBSIs responsive genes created based on the Literature and GO analysis within the DEGs overlapping between the dataset and DEGs deregulated by BTH and SA.

The above-mentioned supplementary tables due to their size are available at https://doi.org/10.5281/zenodo.8222816.

Figure S 8. Traditional thermotherapy box.

Plant were grown in this box covered with a plastic bag to create high temperature inside the box. The temperature and relative humidity inside the box were recorded daily at peak hours; 6h00, 10h00, 14h00, and 18h00.

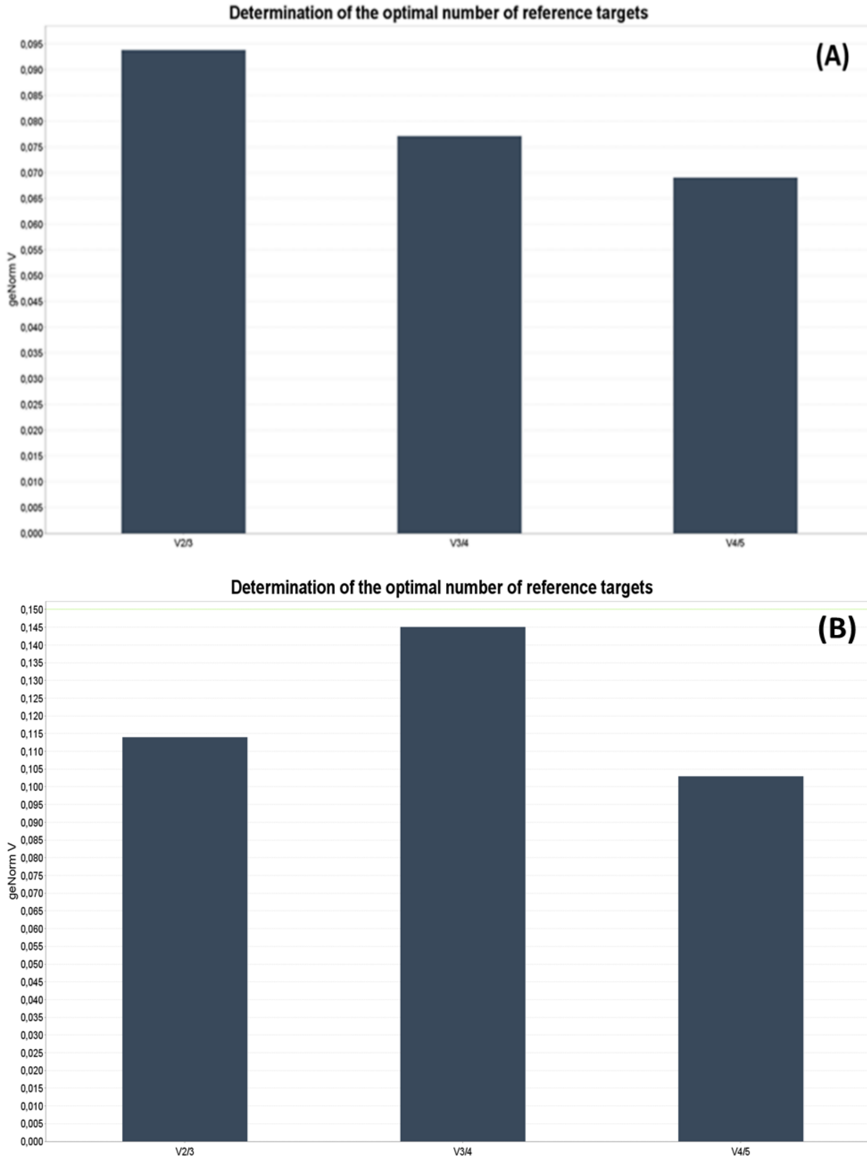


Figure S 9. Determination of the optimal number of reference gene in cassava plant treated with thermotherapy combined with **(A)** BTH or **(B)** SA.

Determination of the optimal number of reference genes in treated cassava plant, according to the pairwise variation V from geNorm (y-axis), for an accurate normalization. The optimum number of genes to be used for normalization equals to the number of reference genes bellow the cut-off pairwise variation value V < 0.15 (312). For both, thermotherapy combined with SA or BTH, the optimal number of reference genes was 2.

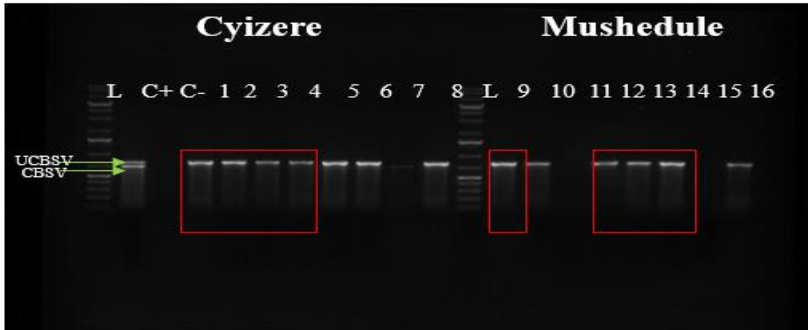


Figure S 10. A gel picture of CBSIs initial confirmation in experimental plants.

From left to right, DNA ladder; C+ correspond to positive control for both CBSV and UCBSV, whereas corresponds to negative control; lanes 1–8 represent 8 screened cassava plant of Cyizere cultivar, lanes 9-16 represent 8 cassava plant of Mushedule cultivar. The red box shows the selected plants for further experiment.

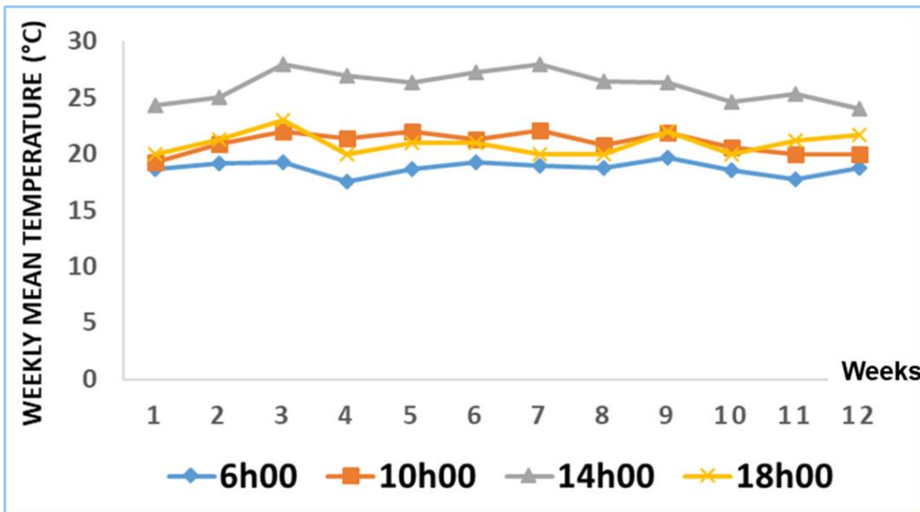


Figure S 11. Variation in temperature outside the greenhouse.

The temperature outside the greenhouse, where the control group was grown, was recorded at 6:00, 10:00, 14:00, and 18:00, and the weekly mean variation was subsequently calculated.

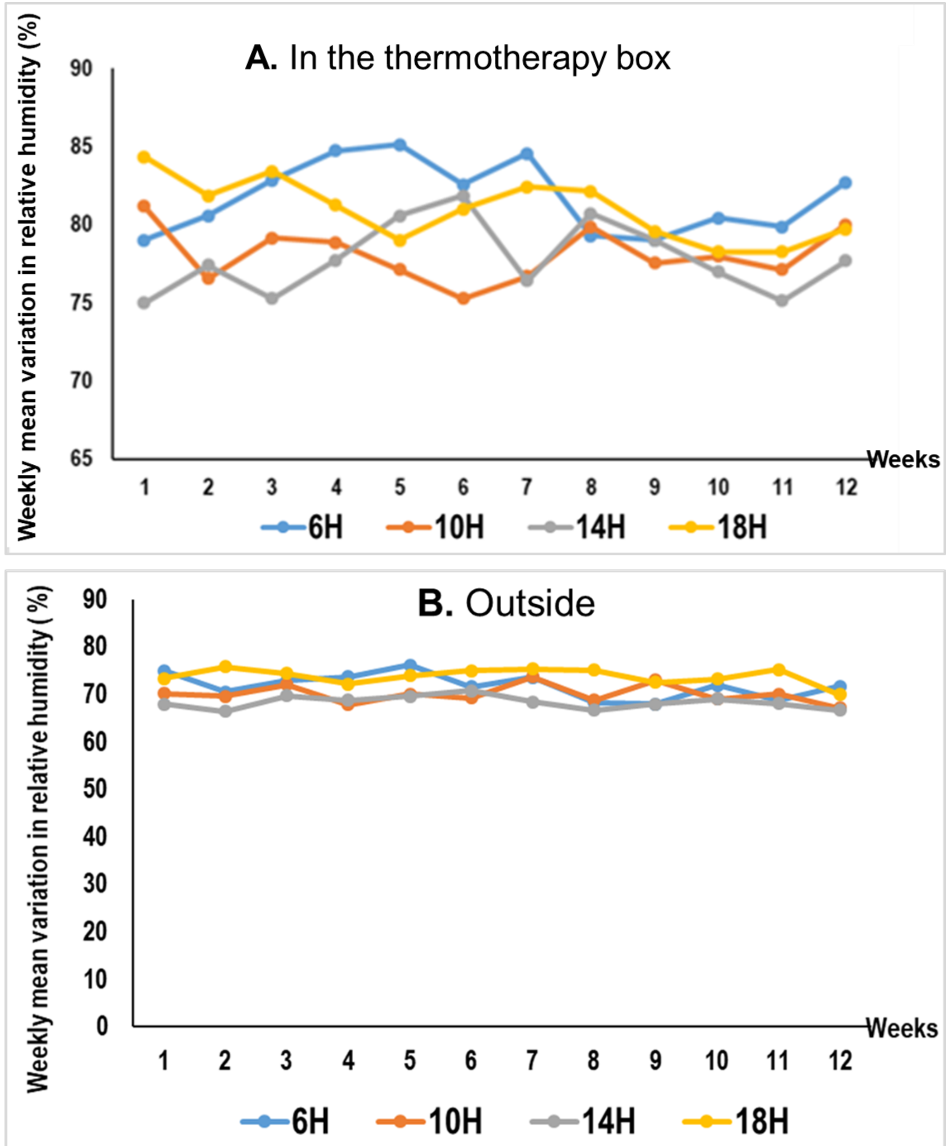


Figure S 12. Weekly mean variation of relative humidity: **A)** inside the thermotherapy box and **B)** outside.

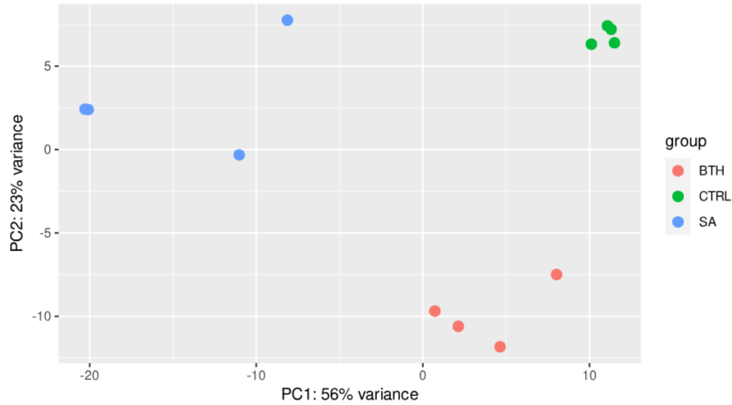


Figure S 13. Principal Component Analysis of the plants sprayed with SA, BTH and controls.

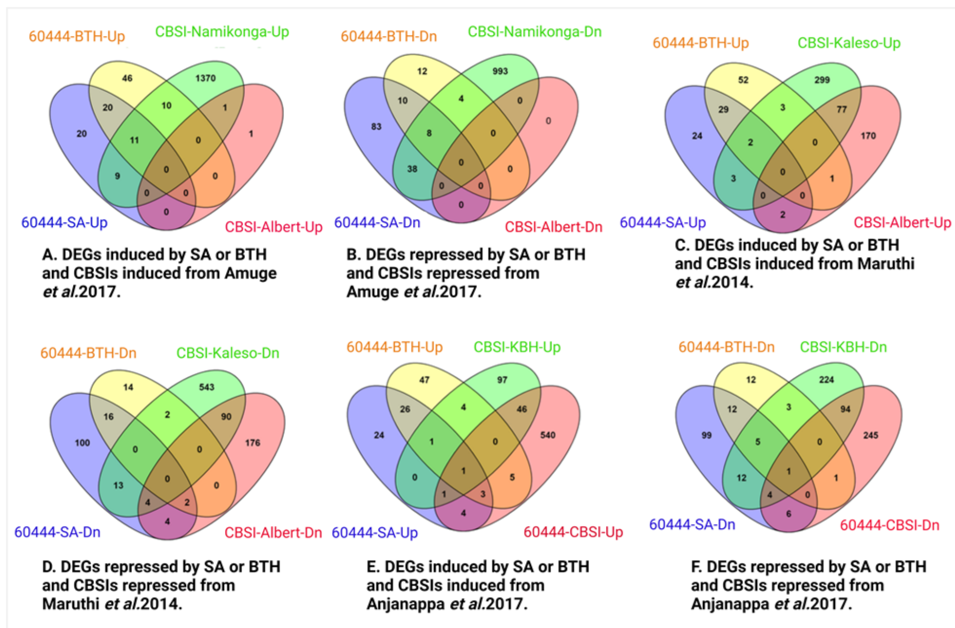


Figure S 14. Overlap between DEGs in 60444 plants treated with SA or BTH and CBSIs infected cultivars from published papers.

A and B represent the overlap of deregulated genes between SA and BTH treatments, and the DEGs in CBSIs-tolerant Namikonga and CBSIs susceptible Albert cultivars 2 days post infection by grafting (297). **C and D**. represent the overlap between SA and BTH deregulated genes with the CBSIs deregulated genes in CBSIs-tolerant Kaleso and CBSIs susceptible Albert cultivars (293); **E and F**. represent the overlap between SA and BTH deregulated genes with the CBSIs deregulated genes in CBSI-tolerant KBH and 60444 cultivar (295).

Chapter 6

General Discussion, Conclusions and Recommendations.

Chapter 6. General Discussion, Conclusions, and Recommendations.

6.1. General Discussion

Cassava possesses excellent potential to contribute significantly to meeting the ever-growing food demand of our increasing population. However, its production has been hindered by the resurgence of viral diseases over the past few decades (35). One of the most concerning diseases is CBSD, which threatens food security, particularly in Central-Eastern African countries, even after extensive research efforts. Thanks to several research outputs, different interventions for the mitigation of CBSD have been initiated. Indeed, epidemiologists and virologists have characterized the causative agents of CBSD, their genetic diversity, and their dissemination in many countries (35,66,75). The outcomes and tools generated from phytopathology and genetic studies set the foundation for initiating breeding research. So far, breeders' efforts have resulted in the obtention of tolerant cultivars that can delay or show no CBSD root symptoms (117–119).

Nevertheless, there is a need to minimize the risk of continuous spreading within the country and to other unaffected Western African countries, which can heavily affect people's food security. Thus, continuous efforts to monitor and manage the disease should be maintained. In this regard, the work presented in this thesis first contributed by exploring the epidemiology of CBSD and factors associated with its dissemination in Rwanda through nationwide cassava fields and farmers' surveys. It was conducted in 13 districts, including 11 major cassava-growing districts and 2 minor cassava-growing districts, to get an accurate picture of the CBSD spread. Munganyinka *et al.* (2014) was the first study to explore the molecular epidemiology of the CBSD causative agents before our study. Our survey study also responded to their recommendation of conducting regular CBSD surveillance. In the previous study, they surveyed only eight cassava districts. In the present study, an extensive analysis of the nationwide situation was conducted to gain a deeper understanding of the distribution of CBSD. In terms of causative agents, as previously reported in Rwanda (93), CBSD was found to be caused by CBSV and UCBSV, with a field incidence of 88% for UCBSV and 39% for CBSV, among which 28% were mixed infections.

The previously reported incidence was higher and could be explained by the current study conducted after the CBSD intervention of distributing cassava-tolerant cultivars. Unlike the study by Munganyinka *et al.*2017, which did not detect CBSIs in two districts (Nyagatare and Kirehe), It was revealed that CBSD has spread in every district surveyed, including five new districts where it was not confirmed before (Nyagatare, Kirehe, Ngoma, Gakenke, and Nyamasheke), with UCBSV being the most common causative species in Rwanda. Similar trends in higher UCBSV prevalence were reported in Rwanda (93), Zambia (71), Burundi (67), and the DRC (220,228).

The spread of plant viruses is a result of triangular factors, including the source of inoculum, mode(s) of transmission, and presence of susceptible host(s) or reservoir host(s) (347). The present study was the first to determine the risk factors associated with CBSD incidence in Rwanda. The factors found to be associated with CBSD were limited awareness of its transmission and management of cassava viral diseases, which were also reported in Uganda (348), Malawi (349), and Tanzania (135). The age of the plants was another risk factor because the severity of CBSD root symptoms increased as the plants got older. Farmers cope with this by early harvesting, which also causes another form of yield decrease from immature cassava roots (350). The use of infected cassava planting stems, as well as the distance to bordering countries with reported presence of viral pandemics, were other identified risk factors. Indeed, when it comes to vegetatively propagated crops, the informal seed system is commonly practiced by farmers, increasing the risk of using infected planting materials for repeated cycles (351,352). However, what motivates the selection of sources of planting materials among farmers still needs to be determined. These findings highlight the need to reinforce the cassava seed system to boost the quality of planting material and increase farmers' awareness and knowledge about cassava viral diseases.

A study conducted in DRC on the *banana bunchy top virus* revealed that multiple introductions of improved varieties, the uncontrolled exchange of non-certified planting materials, and the use of local varieties of uncertain origin have significantly contributed to the spread of the disease (353). Therefore, the community phytosanitary approach, also called area-wide disease management, which engages cassava farmers on a large scale, should be encouraged. Indeed, a study conducted in Tanzania proved that community phytosanitary practices reduced primary CBSD inoculum pressure, sustained low incidence in the grown tolerant cultivar, and improved yield (135). This approach has been successfully applied in managing other vector-borne diseases, including tomato-infecting begomovirus (354), Laurel wilts disease in avocado trees (355), and Huanglongbing disease in sweet orange trees (356). The success of this community phytosanitary approach depends on the availability of disease-free planting material and farmers' awareness about viral diseases and dissemination modes. Both factors are essential to promote compliance with good agricultural practices. Nevertheless, in the current study, half of the participants did not act to manage CBSD in their field and most relied on informal cassava seed systems, highlighting the need to address gaps in seed systems and enhance knowledge about cassava viral diseases.

Income-oriented cassava farming requires a disease-free quality seed, and the sustainability of improved seed systems critically depends on incorporating commercial systems. Interestingly, a model for commercial seed production in Tanzania and Nigeria for long-term supply and access to healthy and improved quality cassava seeds has proved to have promising results. Cassava seed entrepreneurs have reported that they make profits up to USD 551–988 /ha in Nigeria, and the benefit could reach up to USD 1,500 at the end of two seasons following the ratooning strategy in some areas. In Tanzania, the benefit reached USD 1000 – 1500, indicating a positive business case sustainability (47). The information and communications technology (ICT) app developed by IITA << Seed Tracker ™>> facilitates the data

collection from seed production to certification and helps buyers easily find where to buy quality seed (47). Furthermore, promising success has been reported in Ghana and Nigeria for commercial yam seed systems. Yam farmers have started buying seeds every season instead of producing their own to ensure seed quality (357).

Although the commercial seed system offers numerous advantages for both seller and buyer, including improving yield, speeding up dissemination of improved varieties, getting information on farmer preference traits important for further breeding research, generating income and collaboration between private seed companies and breeder, the journey to its success for vegetative crops is still long as only 1% of cassava planting materials was quality seed in Tanzania (47). The primary challenge vegetative crops face is the limited market demand for seeds. This concern can be effectively tackled by maintaining a steady commitment to nurturing the system through collaborative efforts between governmental bodies and research institutes. Indeed, the insights gained into the dynamics of CBSD spread and the crucial risk factors, particularly the origin of planting materials, underscore areas for improvement in sustainable cassava production. It is essential not only to educate farmers about cassava disease transmission and management but also to raise awareness about the pivotal significance of utilizing high-quality seeds and promoting seed replacement practices. Furthermore, a comprehensive grasp of the geographical factors influencing disease spread empowers targeted interventions, allocating resources and efforts to regions at high risk. These investigations can also be extended to other crops, ultimately advancing comprehension of risk factors associated with essential diseases enhancing their management in agricultural ecosystems for sustainable food production. Fortunately, there is cause for optimism as a program dedicated to seed system innovation for vegetatively propagated crops in Africa, known as PROSSIVA, has been successfully launched in five countries, including Rwanda. This initiative is poised to fortify collaboration with the private sector and establish efficient sustainable seed systems for vegetative crops (98).

The strength of a seed system lies in the possibility of implementing robust viral diagnostic tools for routine surveillance of both known and unknown virus epidemiology and evolution (87). Massive parallel sequencing technologies have revolutionized virology studies and allowed the characterization, genetic diversity, and evolution of unknown and known cassava viruses (66,75,116). However, before the current study, the efforts to study the genetic diversity of CBSD causative agents in Rwanda were limited to partial coat protein genes.

In the second study of this thesis, high throughput sequencing on pools of samples was applied to explore cassava virome in Rwanda. Thanks to that, twelve nearly complete genomes (complete CDS and partial UTR) of UCBSV were reconstructed with 97% pairwise identity to KX753357.1 from Tanzania. The evolution and diversity of the UCBSV population, which is the most familiar species in Rwanda (28), was further explored for the first time with a high-resolution methodology based on SNP analysis beyond the consensus coupled with the fixation index (F_{ST}) calculation and Principal Component Analysis (PCA).

Indeed, based on the classical consensus approach, the generated UCBSV genomes had high homogeneity. Fortunately, the F_{ST} approach allowed better discrimination

between identical sequences based on their consensus. As a result, three different UCBSV haplotypes that were geographically clustered were identified. There is a strong hypothesis that links the distribution of the UCBSV haplotype with the prevalence of CBSD-tolerant cultivars that were widely spread across the country in 2015 (94), as it appeared that the distribution of haplotype two was linked with that of NAROCAS1 cultivar. The observed substantial reduction of UCBSV diversity compared to the previous study performed five years before could be due to the difference in cultivars sampled, as the previous study was carried out before the distribution of the tolerant cultivars (93). Indeed, a study on rice stripe virus populations revealed that the host plant impacts the population structure and evolution of the virus (358). Studies have also reported that virus populations can take divergent evolutionary trajectories when passed on to several plants of the same host species; some viruses increase fitness and adapt to their host, others remain the same, whereas others go to extinction, which could be a sign of selection and genetic drift effects (359).

Studying viral evolution is crucial to its control not only because changing virus populations may render diagnostic tools ineffective but also to monitor the emergence of resistance breakdown. Breeders and viral epidemiologists must work more closely to manage viral diseases effectively. Plant breeders tend to focus on the genetic diversity of the host but frequently overlook the genetic diversity of the virus, which should be considered when testing novel resistance genes to reduce the likelihood of resistance breakdown (101).

In the hitherto study, the approach of HTS on pools of samples was cost-effective and generated novel, exciting results. However, some challenges were faced. For example, the association between the three UCBSV Haplotypes and the CBSD severity symptoms could not be determined because of pooling samples which, resulted in pooling different cassava genotypes. Moreover, the association of the haplotypes and cassava cultivar still need further confirmation. Furthermore, even though enough CBSV-specific reads were identified to confirm its presence, the whole genome of CBSV could not be reconstructed, which could have resulted from the dilution effect. Therefore, future research projects should consider providing insights into these challenges.

High-throughput sequencing stands out for its remarkable ability to uncover previously unknown viruses (360). The present research enabled the discovery of a novel *Manihot esculenta associated virus* in the genus Ampelovirus, family Closteroviridae, for the first time in Rwanda. Notably, this newly discovered virus exhibited a striking 98% genetic similarity with the MT773588 strain found in the Democratic Republic of Congo (116). However, our understanding of this virus remains incomplete, necessitating further investigation into its biological characteristics, epidemiological patterns, transmission dynamics, and potential impact on cassava crops. In addition, these findings underscore the importance of monitoring and managing emerging plant diseases, a vital contribution to the sustainable support of global agricultural systems.

The current CBSD management approach centers on distributing diverse CBSD-tolerant cultivars, which have been successfully developed through breeding

programs. However, it is imperative to complement this strategy with methods to clean the virus from infected plants to mitigates the detrimental consequences of viral accumulation on crop yields and ensuring the enduring effectiveness of these cultivars. While in vitro approaches have yielded promising results (136,138,158), it is worth exploring field or greenhouse-based methods, which are more practical and feasible for farmers.

Hence, the third part of this thesis assessed the impact of greenhouse thermotherapy, chemotherapy, and field-based chemotherapy on CBSD using SA and BTH treatments. Interestingly, the combined approach of greenhouse thermotherapy and chemotherapy resulted in lower viral loads, and field chemotherapy led to decreased CBSD root necrosis, which indicates their potential in CBSD management. The observed effect could be explained by the revealed ability of those treatments to upregulate genes potentially involved in CBSI defense, such as Leucine Rich Repeat (LRR) Proteins, transcription factors (e.g., WRKY), Heat shock Proteins (HSP), Mitogen-Activated Protein Kinase (MAPK), Peroxidases, Cytochromes P450, ethylene-responsive genes, and NSP-interacting kinase. Similar genes were induced in CBSD-resistant cultivars challenged by UCBSV through grafting, suggesting they may contribute to the host defense against CBSIs (297).

Similarly, researchers observed a significant upregulation of genes associated with stimulus-response and hormone pathways in a separate RNAseq study conducted on *Salvia miltiorrhiza* following SA induction. Notably, the WRKY transcription factor, Peroxidase, and Cytochrome P450 showed increased expression levels, indicating their potential involvement in the SA signaling pathway during the plant's defense response (361). Moreover, applying BTH triggered the induction of cabbage's LRR proteins, Peroxidase, and transcription regulatory factors (362). These findings further support the notion that these identified components could play crucial roles in SA-mediated signaling pathways involved in the defense response of plants. Our findings suggest that Mitogen-Activated Protein Kinases (MAPKs) are potentially involved in the signaling process of plant defense responses against pathogens activated in response to SA. It aligns with previous research indicating that MAPKs enhance antiviral defense and improve *tomato yellow leaf curl virus* tolerance through either the SA or JA defense signaling pathways (355). Interestingly, in the SA-treated plants, the expression of ankyrin repeat family proteins known to enable the movement of the virus via plasmodesmata by promoting callose degradation (213) was significantly repressed, which suggests that the treatment may have the ability to impede the movement of the virus, resulting in less severe root symptoms. Although the resistance inducers used in the present study thus far may not have a curative effect, the significant reduction in severity and viral load highlights their potential application as a valuable tool.

The current study has provided insights into the transcriptional response of cassava to exogenous salicylic acid and benzothiadiazole. Nevertheless, the specific roles of the deregulated genes in viral plant defense remain unclear. Therefore, further in-depth functional studies are imperative to identify critical genes that could be targeted for breeding or genome editing purposes, ultimately enhancing resistance against CBSD.

Application of resistance inducers like SA and BTH has demonstrated noteworthy success across various crops against viral, parasitic, and fungal diseases (300,307,364–366). Likewise, in the current study, exogenous application of SA and BTH to cassava has shown promise in reducing the impact of CBSD. However, additional optimization efforts are necessary to achieve the optimal effect. Indeed, the efficacy of induced resistance can hinge upon several factors, including the host genotype, concentration, timing, frequency of applications, the chosen method, and environmental conditions (367).

The foremost objective of agricultural development is to ensure a sustainable food supply capable of meeting the growing demands of our global population while preserving the delicate equilibrium of our ecosystem. Although the utilization of some agrochemicals has yielded negative effects on the environment, biodiversity, and human health, it's important to recognize that their absence would result in a substantial 78% reduction in fruit production, a 54% decrease in vegetable output, and a 32% decline in cereal yields (368,369). Thus, achieving a balance between food production and ecology is imperative by adopting eco-friendly and human health safe agrochemicals. Although the inducers used in the present study are typically not associated with toxicity (370,371), residues or runoff from chemicals used in agriculture can persist in the soil and contaminate water bodies, impacting the microbiome or non-target organisms within the environment and ultimately reducing biodiversity (372). Thus, future studies should also assess the balance between the economic benefits of SA and BTH on CBSD mitigation and their impact on the environment to ensure their practical viability and sustainability.

6.2. General Conclusions

Rwanda, a densely populated country in Central Africa with around 525 individuals per square kilometer, relies heavily on agriculture, contributing over a third of its gross domestic product (GDP) (373). In 2009, the country initiated the crop intensification program (CIP) to address agricultural challenges, targeting six priority crops, including cassava (374). While cassava is essential for food and income, it faces significant production constraints due to CBSD in central and eastern African countries. Many efforts have been put into its management to limit its spread to other cassava-growing countries and minimize its economic impact. However, it remains and may remain, a threat to cassava production for a long time. Basic research is always essential to understanding the virus epidemiology, the role of farmers in disease transmission, and genetic variability that would enable better approaches to managing the disease.

In this context, the first part of this thesis contributed to the knowledge of CBSD epidemiology and risk factors contributing to the spread of the disease through a nationwide farmer and cassava field survey. It was found that CBSD had expanded in all thirteen districts surveyed, including five districts where it had not been confirmed before, and its incidence had remained considerably high. The limits of the interventions implemented to reduce CBSD transmission are reflected in the recent observed CBSD expansion. It was revealed that factors such as the source of cuttings,

proximity to the border, knowledge of CBSD transmission, and management were associated with the CBSD incidence in the country. This insight underscores the solid need to develop a sustainable disease-free cassava seed system to minimize disease transmission, as cuttings play a central role in disseminating this disease. Furthermore, awareness among farmers about viral disease transmission and management should be increased, which will, in turn, strengthen the seed system as farmers will understand the importance of using quality planting materials. Nevertheless, understanding cost-benefit analysis to ensure that potential benefits outweigh the costs is essential to motivating farmers and the private sector to invest in cassava quality seeds.

In promotion of this investment, the government should provide subsidies to farmers to promote quality seed accessibility at a lower price, providing them with regular extension services, exploring cassava value-added opportunities by arranging for importers, and establishing policy linkages with agro-processing industries (e.g., promoting the use of cassava flour in bakery) to increase cassava market and improve farmers overall farming outcomes. Efforts are being made through collaboration between governments and different research institutes to overcome the bottlenecks that hinder the sustainability of seed systems for vegetative crops (98). Regular surveillance of viral genetic diversity and evolution is essential for sustainable disease management. In fact, if the genetic diversity and evolution of virus populations are not considered, it can easily defeat the efficacy and durability of some mitigation strategies.

In this regard, the second part of this thesis has made a groundbreaking contribution to knowledge by shedding light on the genetic diversity of cassava ipomovirus in Rwanda, particularly after the widespread distribution of tolerant cultivars. Twelve nearly complete genomes of UCBSV were successfully generated for the first time in over a decade following its first report in Rwanda.

This milestone allowed a more in-depth analysis that surpassed the traditional consensus level, identifying three distinct UCBSV haplotypes that exhibited geographic clustering within the sampled regions. Furthermore, the analysis of cassava virome diversity revealed a noteworthy decrease in UCBSV diversity following the widespread distribution of tolerant cultivars compared to previous records. Notably, Haplotype 2 is associated with the NAROCAS1 cultivar, shedding light on the possible cultivar impact on viral genetic diversity. In addition to these results, a new virus named MEaV-1 ampelovirus (116) was identified in the Ruhango district, further underscoring the importance of integrating High-Throughput Sequencing (HTS) technology into routine disease surveillance and seed certification protocols. This integration could enable the screening of both known and unknown viruses.

Moreover, embracing HTS technology provides a pragmatic remedy to counteract the pitfalls of misdiagnosis resulting from reliance solely on visual symptom assessments. This issue is especially evident in instances like CBSD, characterized by diverse symptom manifestations, varying recovery patterns, and uneven plant virus distribution. Embracing HTS empowers decision-makers to make well-informed choices and proactively implement measures to safeguard against emerging viral

threats. This approach simultaneously bolsters the prospects of sustainable agriculture in the face of viral adversity.

Fortunately, HTS is becoming more affordable, although there are still challenges to applying this technology in Africa, particularly due to the fewer people with skills in bioinformatics tools and data analysis. Different capacity-building initiatives have been launched, including the BecA-ILRI hub (Biosciences Eastern and Central Africa-International Livestock Research Institute) (375), the Next Einstein Initiative (376), and collaborations with developed countries (377). Nonetheless, significant efforts are still required to strengthen the capacity building of local scientists, enabling them to become self-reliant in utilizing advanced bioinformatic tools and conducting data analysis.

In the sustainable management of a disease transmitted via planting materials, the quality of planting materials is crucial. Currently, CBSD management mainly relies on tolerant cultivars that weaken as the virus accumulates over cycles. Nevertheless, it was realized that the long-term distribution of clean seeds was fragile, because when support for the program ended, the provision of improved seeds ceased (223). Therefore, it is essential to mitigate this effect by raising farmers' awareness of the importance of changing seeds and practicing other cost-effective measures to lower the impact of the disease, such as boosting cassava's immune defense.

In that aspect, the last goal of this thesis brought a valuable contribution to knowledge by transforming *in vitro* methods into more feasible approach for viral cleaning. The focus shifted to greenhouse and field experiments to assess the efficacy of thermotherapy, Salicylic Acid, and Benzothiadiazole in virus cleaning. The experimental results yielded valuable insights, demonstrating the potential of these interventions in reducing viral load and mitigating CBSD root symptoms, thus highlighting their possible application in combating the disease's impact. Moreover, transcriptomic data analysis following the exogenous spray of Salicylic Acid or Benzothiadiazole on uninfected cassava plants revealed the activation of several intriguing DEGs previously identified in other studies investigating the natural cassava response to CBSIs in resistant cultivars, indicating their potential roles in defense mechanisms. (292). Within this set of DEGs, notable examples include genes featuring a Leucine-rich repeat (LRR) domain, Heat Shock Proteins, Mitogen-Activated Protein Kinases, Cytochrome P450 enzymes, ethylene-responsive genes, and diverse transcription factors like WRKY.

The findings from the present study offer promising prospects for bolstering cassava's viral defense through thermotherapy, SA, and BTH treatments. However, it is crucial to investigate the specific functions of the DEGs deregulated by SA and BTH in cassava's defense mechanisms. This understanding lays the foundation for their future applications in crop protection, fostering sustainable cassava production while countering viral risks.

6.3. Recommendations

The present thesis presented essential results on the CBSIs epidemiology and risk factors associated with its spread in Rwanda. It also provided the first findings on the

whole UCBSV genomes population analysis, which shed light on UCBSV diversity and evolution in Rwanda. Moreover, a study on a farmer-friendly approach for virus cleaning was done, which provided the basis for optimizing an effective protocol for field CBSI cleaning to minimize its impact on yield.

Based on the findings from the current thesis, we articulate the following recommendations for farmers, policymakers, researchers, and governments to address the complex challenges posed by CBSD and promote sustainable cassava production:

i) Farmers and seed multipliers:

1. Quality seed and seed replacement: Farmers should prioritize using quality planting materials and regularly replace their cassava seeds to reduce the impact of virus accumulation over several cycles.

2. **Awareness and training:** Quality cassava starts with quality seed. Our findings highlight that farmers need training about the importance of using certified seeds, viral disease transmission, and management. Increased awareness of the impact of practicing an informal seed system would encourage them to use quality planting materials and other disease mitigation practices.

3. **Quality Seed Certification:** The seed certification agency should enhance the training of professional seed multipliers in efficient technologies and emphasize the significance of implementing rigorous quarantine measures in the nurseries where seed multiplication occurs. Moreover, conducing regular audits would help ensure compliance with certification requirements.

ii) Policy Makers and Government:

1. Support Sustainable Seed Systems: Establishing a sustainable, disease-free cassava seed system is imperative to ensure a reliable supply of healthy planting materials, sustaining cassava productivity, and supporting livelihoods in cassava-dependent communities. This success hinges on collaborative efforts among the government, research institutions, seed production stakeholders, private sectors, seed certification agencies, and cassava farmers. e.g., supporting a commercial seed system, and evaluating its success would significantly improve cassava production. The success of such a system in vegetative crops would be a groundbreaking achievement that would significantly improve cassava production.

2. Subsidies and Extension Services: The government should provide subsidies to farmers to promote the accessibility of high-quality seeds at lower prices while promoting private sector investment in vegetative quality seed production to enhance the sustainability of their seed system. Strengthening extension services would also assist farmers in adopting best practices.

3. Market Expansion: Encourage policies that explore cassava value-added opportunities, such as promoting the use of cassava flour in the bakery industry to increase the cassava market and improve farmers' overall economic outcomes while motivating them to use quality seed.

4. Support Research: Provide financial support for research initiatives that would enhance major disease management. Strengthen capacity-building initiatives for local scientists to enhance their skills in utilizing advanced technologies such as bioinformatic tools, and data analysis.

iii) Researchers:

1. Other agricultural researchers can build upon the current nationwide survey by conducting similar studies on vital crops, contributing to a broader understanding of disease spread and risk factors for better management.

2. Future study should further investigate the factors contributing to the evolution of CBSIs in Rwanda. Such an investigation holds the potential to provide valuable insights, aiding in the focused allocation of research endeavors and the formulation of disease management strategies grounded in evidence.

3. Embrace HTS technology for more accurate disease surveillance and seed certification protocols. Promote the integration of HTS into routine disease management practices for simultaneous detection of known and unknown viruses not only in cassava but also in other key crops; this would help detect new or emerging cassava viruses early, enabling prompt action to contain their spread and implement control measures to prevent potential crop losses.

4. The innovative approach for studying virus genetic diversity used in the current research revealed that a better diversity could be explored beyond the consensus and thus could also be applied to study the diversity of other important viruses toward better management. Furthermore, exploring the distribution of ampelovirus in the country, its biological properties, and its impact on cassava yield would contribute to understanding its biology and impact on cassava farming.

5. The assessment of greenhouse thermotherapy and chemotherapy approaches for CBSD mitigation provided valuable insights into disease management techniques and Researchers can expand upon this work, optimizing and adapting these approaches for different crops and diseases. Moreover, RNAi-mediated gene silencing should investigate specific functions of potential DEGs induced by SA and BTH.

6. Studies that could boost cassava growth or branching to increase the number of cuttings per plant would strengthen the profitability of the seed system.

Bibliography

Bibliography

1. Thresh JM. The Impact of Plant Virus Diseases in Developing Countries. In: Virus and virus-Like Diseases of Major Crops in Developing Countries. 2003. p. 1–2.
2. Jones RAC, Naidu RA. Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. Annu Rev Virol. 2019;6:387–409.
3. Sastry SK. Plant Virus and Viroid Diseases in the Tropics. Vol. 2 Epidemio. 2014. 361 p.
4. Fargette D, Konat G, Fauquet C, Muller E, Peterschmitt M, Thresh JM. Molecular Ecology and Emergence of Tropical Plant Viruses. Annu Rev Phytopathol. 2006;235–62.
5. Jones RAC. Global plant virus disease pandemics and epidemics. Plants. 2021;10(2):1–41.
6. Abaca A, Kawuki R, Tukamuhabwa P, Baguma Y, Pariyo A, Alicai T, et al. Genetic Relationships of Cassava Genotypes That are Susceptible or Tolerant to Cassava Brown Streak Disease in Uganda. J Agric Sci. 2013;5(7).
7. Charrier A, Lefevre F. La Diversite Genetique Du Manioc : Son Origine , Son Evaluation Et Son Utilisation. Evaluation. 1988;(Figure 1):71–81.
8. Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T. Cassava: the Drought, War and Famine Crop in a Changing World. Sustainability. 2010;(ISSN 2071-1050):3572–607.
9. FAOSTAT. Food and Agriculture Organization of the United Nations (FAO). Rome, Italy [Internet]. 2020 [cited 2022 Aug 11]. Available from: https://www.fao.org/faostat/en/#data/QCL
10. FAO. Save and Grow: Cassava A guide to sustainable producrion intensification. 2013.
11. Mwebaze P. Socio-economic factors influencing productivity of cassava farmers in E . Africa. Aust Agric Resour Econ Soc. 2016;(February):2–5.
12. El-sharkawy MA. Cassava biology and physiology. Plant Mol Biol. 2004;(June).
13. Legg J, Somado EA, Barker I, Beach L, Ceballos H, Cuellar W, et al. A global alliance declaring war on cassava viruses in Africa. 2014;231–48.
14. Reinhardt Howeler, NeBambi Lutaladio GT. SAVE AND GROW Cassava A GUIDE TO SUSTAINABLE PRIDUCTION INTENSIFICATION. Rome; 2013.
15. Night G, Asiimwe P, Gashaka G, Nkezabahizi D, Legg PJ, Okao-Okuja G, et al. Occurrence and distribution of cassava pests and diseases in Rwanda. "Agriculture, Ecosyst Environ [Internet]. 2011;(January 2018). Available from: http://dx.doi.org/10.1016/j.agee.2011.01.014
16. National institute of statistics of Rwanda. Seasonal Agricultural Survey. 2020.
17. Nuwamanya E, Baguma Y, Rey CME. An African Perspectives: Developing an African bioressources based industry - the case for cassava. In: Creating Sustainable

Bioeconomies. 2016.

18. Nduwumuremyi A, Melis R, Shanahan P, Asiimwe T. Participatory appraisal of preferred traits, production constraints and postharvest challenges for cassava farmers in Rwanda. Food Secur. 2016;8(2):375–88.

19. Balagopalan C. Cassava utilization in food, feed and industry. In: Cassava: biology, production and utilization. 2002. p. 301–18.

20. Guira F, Some K, Kabore D, Sawadogo-Lingani H, Traore Y, Savadogo A. Origins, production, and utilization of cassava in Burkina Faso, a contribution of a neglected crop to household food security. Food Sci Nutr. 2016;5(3):415–23.

21. Jansson C, Westerbergh A, Zhang J, Hu X, Sun C. Cassava, a potential biofuel crop in China. Appl Energy [Internet]. 2009;86(SUPPL. 1):S95–9. Available from: http://dx.doi.org/10.1016/j.apenergy.2009.05.011

22. Fathima AA, Sanitha M, Tripathi L, Muiruri S. Cassava (Manihot esculenta) dual use for food and bioenergy: A review. Food Energy Secur [Internet]. 2022;(March):1–26. Available from: https://doi.org/10.1002/fes3.380

23. Imarc. Cassava Processing Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2023-2028 [Internet]. Market Research Report. 2023 [cited 2023 Jul 12]. Available from: https://www.imarcgroup.com/cassava-processing-plant#:~:text=What is the size of,311.5 Million Tons in 2022.

24. Krungsri. Industry Outlook 2022-2024: Cassava Industry [Internet]. 2023 [cited 2023 May 29]. Available from: https://www.krungsri.com/en/research/industry/industry-outlook/agriculture/cassava/io/io-cassava-21

25. Akinfenwa G. Nigeria has potential of earning $2.98b from cassava export yearly. [Internet]. 2021. Available from: https://guardian.ng/features/nigeria-has-potential-of-earning-2-98b-from-cassava-export-yearly/

26. IITA. Transforming African Agriculture: Tanzania to improve cassava in Africa with NextGen Cassava project. 2016;6421.

27. FAOSTAT. Food and Agriculture Organisation of the United Nations, Rome, Italy. FAOSTAT statistics database. http://www.fao.org/faostat/en/#data/QC. Accessed on June 6, 2021. 2019.

28. Nyirakanani C, Bizimana JP, Kwibuka Y, Nduwumuremyi A, Bigirimana V de P, Bucagu C, et al. Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated With Cassava Brown Streak Disease Incidence and Cassava Productivity. Front Sustain Food Syst. 2021;5(December):1–14.

29. Fermont AM, van Asten PJA, Tittonell P, van Wijk MT, Giller KE. Closing the cassava yield gap: An analysis from smallholder farms in East Africa. F Crop Res. 2009;112(1):24–36.

30. Fermont AM, Tittonell ÆPA, Giller KE. Towards understanding factors that govern fertilizer response in cassava : lessons from East Africa. 2010;133–51.

31. Rosenthal DM, Slattery RA, Miller RE, Grennan AK, Cavagnaro TR, Fauquet CM, et al. Cassava about-FACE: Greater than expected yield stimulation of cassava (Manihot

esculenta) by future CO 2 levels. Glob Chang Biol. 2012;18(8):2661–75.

32. Kintché K, Hauser S, Mahungu NM, Ndonda A, Lukombo S, Nhamo N, et al. Cassava yield loss in farmer fields was mainly caused by low soil fertility and suboptimal management practices in two provinces of the Democratic Republic of Congo. Eur J Agron [Internet]. 2017;89(June):107–23. Available from: http://dx.doi.org/10.1016/j.eja.2017.06.011

33. codycovefarm. Plant Profile: Cassava (Manihot esculenta) [Internet]. 2022. Available from: https://codycovefarm.com/plant-profile-cassava-manihot-esculenta/

34. Kombate K, Dossou-Aminon I, Dansi A, Adjatin RA, Kpemoua K, Dassou GA, et al. Cassava (Manihot esculenta Crantz) Production Constraints, Farmers' Preference Criteria and Diversity Management in Togo. Int J Curr Microbiol Appl Sci. 2017;6(6):3328–3240.

35. Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, et al. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. Virus Res [Internet]. 2011;159(2):161–70. Available from: http://dx.doi.org/10.1016/j.virusres.2011.04.018

36. Rey C, Vanderschuren H. Cassava Mosaic and Brown Streak Diseases: Current Perspectives and beyond. Annu Rev Virol. 2017;4:429–52.

37. Antony B, Lisha VS, Palaniswami MS. Evidences for transmission of Indian cassava mosaic virus through Bemisia tabaci-cassava biotype. Arch Phytopathol Plant Prot. 2009;42(10):922–9.

38. Jose A, Makeshkumar T, Edison S. Survey of Cassava Mosaic Disease in Kerala. J Root Crop. 2011;(January 2011).

39. Duraisamy R, Natesan S, Raveendran M, Gandhi K, Lakshmanan P, Karuppusamy N, et al. Molecular studies on the transmission of indian cassava mosaic virus (ICMV) and Sri Lankan Cassava Mosaic Virus (SLCMV) in cassava by bemisia tabaci and cloning of ICMV and SLCMV replicase gene from cassava. Mol Biotechnol. 2013;53(2):150–8.

40. Legg JP, Shirima R, Tajebe LS, Guastella D, Boniface S, Jeremiah S, et al. Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. Pest Manag Sci. 2014;70(10):1446–53.

41. Maruthi MN, Hillocks RJ, Mtunda K, Raya MD, Muhanna M, Kiozia H, et al. Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). 2005;312:307–12.

42. Wagaba H, Beyene G, Trembley C, Alicai T, Fauquet CM, Taylor NJ. Efficient transmission of Cassava brown streak disease viral pathogens by chip bud grafting. BMC Res Notes. 2013;

43. Anjanappa RB, Mehta D, Maruthi MN, Kanju E, Gruissem W, Vanderschuren H. Characterization of Brown Streak Virus – Resistant Cassava. Am Phythological Soc. 2016;29(7):527–34.

44. Sheat S, Fuerholzner B, Stein B, Winter S. Resistance against cassava brown streak

viruses from africa in cassava germplasm from South America. Front Plant Sci. 2019;10(May).

45. Elegba W, Gruissem W, Vanderschuren H. Screening for resistance in farmer-preferred cassava cultivars from ghana to a mixed infection of cbsv and ucbsv. Plants. 2020;9(8):1–16.

46. Maruthi MN, Jeremiah SC, Mohammed IU, Legg JP. The role of the whitefly , Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. J Phytopathol. 2017;(July):707–17.

47. Legg PJ, Elohor D-O, David E, Michael F, Edward Kanju, Regina Kapinga PLK, Sanni L, et al. Commercially Sustainable Cassava Seed Systems in Africa. In: Root, Tuber and Banana Food System Innovations. 2022. p. 483–509.

48. Patil BL, Legg JP, Kanju E, Fauquet CM. Cassava brown streak disease: a threat to food security in Africa. J Gen Virol. 2015;96(5):956–68.

49. Hillocks RJ, Jennings DL. Cassava brown streak disease : A review of present knowledge and research needs. Internatinal J Pest Manag. 2010;0874.

50. J.M.Thresh RJH and. CASSAVA MOSAIC AND CASSAVA BROWN STREAK VIRUS DISEASES IN AFRICA : A comparative guide to symptoms and aetiologies. 2000;7(December):1–8.

51. Legg JP, Kumar PL, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, et al. Cassava Virus Diseases : Biology , Epidemiology , and Management. Adv Virus Res. 2014;

52. Jeske H. Germiniviruses. In: Villiers E-M de, Hausen H Zur, editors. TT Viruses_ The Still Elusive Human Pathogens. Berlin; 2009. p. 185–226.

53. Fondong VN. Geminivirus protein structure and function. Mol Plant Pathol. 2013;14:635–49.

54. Alabi OJ. Cassava Mosaic Disease : A Curse to Food Security in Sub- Saharan Africa Cassava and Its Importance in Sub-. APSnet Featur Stories [Internet]. 2011;(44):1–31. Available from: http://www.apsnet.org/publications/apsnetfeatures/Pages/cassava.aspx

55. Patil BL. Plant Viral Diseases: Economic Implications [Internet]. Elsevier. Elsevier Ltd.; 2020. 81–97 p. Available from: http://dx.doi.org/10.1016/B978-0-12-809633-8.21307-1

56. Sseruwagi P, Tairo F, Ndunguru J. The Cassava Diagnostics Project (CDP). 2018.

57. Thresh JM, Otirn-nape DFGW. Effects of African cassava mosaic geminivirus on the yield of cassava. Trop Sci. 1994;(September 1993):26–42.

58. Akano A., Dixon AG., Mba C, Barrera E, Fregene M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor Appl Genet. 2002;521–5.

59. Lokko Y, Dixon AGO, Offei SK, Danquah EY. Gene complementarity of resistance to the Cassava Mosaic Disease among the African cassava accessions. African Crop Sci J. 2006;

60. Fondong VN. The Search for Resistance to Cassava Mosaic Geminiviruses : How

Much We Have Accomplished , and What Lies Ahead. Front Microbiol. 2017;8(March):1–19.

61. Legg JP, Owor B, Sseruwagi P, Ndunguru J. Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of a Regional Pandemic. Adv Virus Res. 2006;67(06).

62. Storey HH, Nichols RFW. Virus Diseases of East African Plants. East African Agric J. 1938;3(6):446–9.

63. Nichols RFW. The Brown Streak Disease of Cassava. East African Agric J. 1950;15(3):154–60.

64. Hillocks RJ, Thresh JM, Tomas J, Botao M, Macia R, Zavier R. Cassava brown streak disease in northern Mozambique. Int J Pest Manag. 2002;48(3):178–81.

65. Legg JP, Lava Kumar P, Makeshkumar T, Tripathi L, Ferguson M, Kanju E, et al. Cassava virus diseases: Biology, epidemiology, and management. Adv Virus Res. 2015;91(1):85–142.

66. Ndunguru J, Sseruwagi P, Tairo F, Stomeo F, Maina S, Djinkeng A, et al. 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. 2015;10(10):1–18.

67. Bigirimana S, Barumbanze P, Ndayihanzamaso P, Shirima R, Legg JP. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. 2011;5197.

68. Mulimbi W, Phemba X, Assumani B, Kasereka P, Muyisa S, Ugentho H, et al. First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. 2012;5197.

69. Roux-Cuvelier M, Teyssedre D, Chesneau T, Jeffray C, Massé D, Jade K, et al. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Mayotte Island . New Dis Reports. 2014;30(1):28–28.

70. Mbewe W, Winter S, Mukasa S, Tairo F, Peter S, Joseph N, et al. Deep Sequencing Reveals a Divergent Ugandan cassava brown streak virus Isolate from Malawi. Am Soc Microbiol. 2017;5(33):8–9.

71. Mulenga RM, Makulu M, Laura M. Cassava Brown Streak Disease and Ugandan cassava brown streak virus Reported for the First Time in Zambia. Plant Dis. 2018;102(7):1410–8.

72. Monger WA, Seal S, Isaac AM, Foster GD. Molecular characterization of the Cassava brown streak virus coat protein. Plant Pathol. 2001;527–34.

73. Monger WA, Seal S, Cotton S, Foster GD. Identification of different isolates of Cassava brown streak virus and development of a diagnostic test. Plant Pathol. 2001;768–75.

74. Mbanzibwa DR, Tian YP, Tugume AK, Patil BL, Yadav JS, Bagewadi B, et al. Evolution of cassava brown streak disease-associated viruses. J Gen Virol. 2011;92(4):974–87.

75. Winter S, Koerbler M, Stein B, Pietruszka A, Paape M, Butgereitt A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol. 2010;91(2010):1365–72.

76. Hull R. Comparative Plant Virology. Angewandte Chemie International Edition, 6(11), 951–952. 2008. 5–24 p.

77. Mbanzibwa DR, Tian Y, Mukasa SB, Valkonen JPT. 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 . J Virol. 2009;83(13):6934–40.

78. Chung BYW, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the Potyviridae. Proc Natl Acad Sci U S A. 2008;105(15):5897–902.

79. Alicai T, Ndunguru J, Sseruwagi P, Tairo F, Okao-okuja G, Nanvubya R, et al. a rapidly evolving genome : implications for virus speciation , variability , diagnosis and host resistance. Nat Publ Gr. 2016;(October):1–14.

80. Hillocks RJ, RAya MD, Mtunda KM, Kiozia H. Effects of Brown Streak Virus Disease on Yield and Quality of Cassava in Tanzania. J Phytopathol. 2001;394.

81. Infonet biovision. African cassava mosaic virus (ACMV) [Internet]. 2019 [cited 2023 Mar 16]. Available from: https://infonet-biovision.org/PlantHealth/Pests/African-cassava-mosaic-virus-ACMV

82. Ndushabandi E, Rutaysire C, Mwangi L, Bizimana V. Crop Intensification Program (CIP) Citizen ' S Satisfaction Survey. Institute of Research and Dialogue for Peace. 2018.

83. Casinga CM, Shirima RR, Mahungu NM, Bashizi KB, Munyerenkana CM, Ughento H, et al. Expansion of the Cassava Brown Streak Disease Epidemic in Eastern Democratic Republic of Congo. Plant Dis. 2021;2011(August):2177–88.

84. FAO. Food and Agriculture Organization of the United Nations. 2019; Available from: https://www.fao.org/resilience/news-events/detail/en/c/1208321/

85. United Nations Rwanda. Restoring cassava farming in Rwanda [Internet]. 2019. Available from: https://rwanda.un.org/en/22965-restoring-cassava-farming-rwanda

86. Paridaens A-M, Jayasinghe S. Rwanda:Comprehensive Food Security & Vulnerability Analysis (CFSVA). 2018.

87. Douthwaite B. Development of a cassava seed certification system in Rwanda: Evaluation of CGIAR contributions to a policy outcome trajectory. 2020.

88. FAO. Food Outlook- Biannual report on global food markets - November 2018 [Internet]. Global information and early warning system on food and agriculture. 2018. 104 p. Available from: http://www.fao.org/docrep/013/al969e/al969e00.pdf

89. MINAGRI. Annual report 2020-2021: Ministry of agriculture and animal resources. 2021.

90. National Institute of Statistics of Rwanda. Seasonal Agricultural Survey. 2021.

91. Legg JP, Patil BL, Legg JP, Kanju E, Fauquet CM. Cassava brown streak disease : A

threat to food security in Africa Cassava brown streak disease : a threat to food security in Africa. 2014;(December).

92. Nyirahorana C. Drivers Behind Adoption Of Cassava Brown Streak Disease Control Measures In Rwanda. 2017;6(11):113–7.

93. Munganyinka E, Ateka EM, Kihurani AW, Kanyange MC, Tairo F. Cassava brown streak disease in Rwanda , the associated viruses and disease phenotypes. Plant Pathol. 2017;67:377–87.

94. Miklyaev M, Jenkins G, Shobowale D. Sustainability of agricultural crop policies in rwanda: An integrated cost–benefit analysis. Sustain. 2021;13(1):1–22.

95. IITA. New disease resistant cassava varieties introduced in Rwanda [Internet]. 2020. Available from: https://blogs.iita.org/index.php/new-disease-resistant-cassava-varieties-introduced-in-rwanda/

96. IITA. Cassava Agribusiness Seed Systems (CASS) [Internet]. 2019 [cited 2023 Aug 3]. Available from: https://www.iita.org/iita-project/cassava-agribusiness-seed-systems-cass/

97. IITA. Rwandan farmers name new cassava varieties during the cassava week [Internet]. 2020. Available from: https://www.iita.org/news-item/rwandan-farmers-name-new-cassava-varieties-during-the-cassava-week/

98. IITA. IITA embarks on project to boost vegetatively-propagated crop seed systems in Africa [Internet]. 2022 [cited 2023 Jul 24]. Available from: https://www.iita.org/news-item/iita-embarks-on-project-to-boost-vegetatively-propagated-crop-seed-systems-in-africa/

99. Gondwe F, N.M. MahunguI, R.J. HillocksP, M.D. RayaQ, C.C. MoyoR MMS, ChipunguS FP, Benesi TRM. Economic Losses Experienced by Small-scale Farmers in Malawi due to Cassava Brown Streak Virus Disease. Nat Ressources Int Ltd. 2003;(October).

100. Chipeta MM, Shanahan P, Melis R, Sibiya J, Ibrahim RM. Farmers ' knowledge of cassava brown streak disease and its management in Malawi. Internatinal J Pest Manag [Internet]. 2016;0874. Available from: https://doi.org/10.1080/09670874.2016.1167268

101. Rubio L, Galipienso L, Ferriol I. Detection of Plant Viruses and Disease Management: Relevance of Genetic Diversity and Evolution. Front Plant Sci. 2020;11(July):1–23.

102. Givord L, Fargette D, Kounounguissa D, Thouvenel J, Walter B, Regenmortel VM. Detection of geminiviruses from tropical countries by a double monoclonal antibody ELISA using antibodies to African cassava mosaic virus. Plant Pathol. 1994;(May).

103. Martinelli F, Scalenghe R, Davino S, Panno S, Scuderi G, Ruisi P, et al. Advanced methods of plant disease detection. A review. Agron Sustain Dev. 2015;35(1):1–25.

104. Abd El-Aziz M. Three modern serological methods to detect plant viruses. J Plant Sci Phytopathol. 2019;3(3):101–6.

105. Kumar PL, Legg J. Laboratory manual for the diagnosis of cassava virus diseases. [Internet]. Tropical Agriculture. 2009. Available from:

http://biblio.iita.org/documents/U09ManKumarLaboratoryNothomNodev.pdf-fbeb6a750d31604ab1e6f09e3856f081.pdf

106. Adams IP, Abidrabo P, Miano DW, Alicai T, Kinyua ZM, Clarke J. High throughput real-time RT-PCR assays for specific detection of cassava brown streak disease causal viruses , and their application to testing of planting material. Plant Pathol. 2013;233–42.

107. Tomlinson JA, Ostoja-starzewska S, Adams IP, Miano DW, Abidrabo P, Kinyua Z, et al. Loop-mediated isothermal amplification for rapid detection of the causal agents of cassava brown streak disease. J Virol Methods [Internet]. 2013;191(2):148–54. Available from: http://dx.doi.org/10.1016/j.jviromet.2012.07.015

108. Londoño MA, Harmon CL, Polston JE. Evaluation of recombinase polymerase amplification for detection of begomoviruses by plant diagnostic clinics. Virol J [Internet]. 2016;13(1):1–9. Available from: http://dx.doi.org/10.1186/s12985-016-0504-8

109. Le Provost G, Iskra-Caruana ML, Acina I, Teycheney PY. Improved detection of episomal Banana streak viruses by multiplex immunocapture PCR. J Virol Methods. 2006;137(1):7–13.

110. Rodoni BC, Ahlawat YS, Varma A, Dale JL, Harding RM. Identification and characterization of banana bract mosaic virus in India. Plant Dis. 1997;81(6):669–72.

111. Shcherbakova LA. Comprehensive and Molecular Phytopathology. In: Comprehensive and Molecular Phytopathology [Internet]. 2007. p. 75–116. Available from: http://www.sciencedirect.com/science/article/pii/B9780444521323500067

112. Mcclure L V, Lin Y, Sullivan CS. Detection of Viral microRNAs by Northern Blot Analysis. In: Methods in molecular biology [Internet]. 2011. p. 153–71. Available from: http://www.springerlink.com/index/10.1007/978-1-61779-037-9

113. Roossinck MJ. Deep sequencing for discovery and evolutionary analysis of plant viruses. Virus Res [Internet]. 2017;239:82–6. Available from: http://dx.doi.org/10.1016/j.virusres.2016.11.019

114. Adams IP, Fox A, Boonham N, Massart S. The impact of high throughput sequencing on plant health diagnostics. Eur J Plant Pathol. 2018;909–19.

115. Massart S, Chiumenti M, Jonghe K De, Glover R, Haegeman A, Koloniuk I, et al. Virus Detection by High-Throughput Sequencing of Small RNAs : Large-Scale Performance Testing of Sequence Analysis Strategies. Phytopathology. 2019;488–97.

116. Kwibuka Y, Bisimwa E, Blouin AG, Bragard C, Candresse T, Faure C, et al. Novel ampeloviruses infecting cassava in central africa and the south-west indian ocean islands. Viruses. 2021;13(6):1–17.

117. Kawuki RS, Kaweesi T, Esuma W, Pariyo A, Kayondo IS, Ozimati A, et al. Eleven years of breeding efforts to combat cassava brown streak disease. Breed Sci. 2016;66(4):560–71.

118. Mukiibi DR, Alicai T, Kawuki R, Okao-Okuja G, Tairo F, Sseruwagi P, et al. Resistance of advanced cassava breeding clones to infection by major viruses in

Uganda. Crop Prot. 2019;115(February 2018):104–12.

119. Masinde EA, Kimata B, Ogendo JO, Mulwa RMS, Mkamilo G, Maruthi MN. Developing dual-resistant cassava to the two major viral diseases. Crop Sci. 2021;61(3):1567–81.

120. Sheat S, Winter S. Developing broad-spectrum resistance in cassava against viruses causing the cassava mosaic and the cassava brown streak diseases. Front Plant Sci. 2023;14(January):1–10.

121. Omongo CA, Kawuki R, Bellotti AC, Alicai T, Baguma Y, Maruthi MN, et al. African Cassava Whitefly , Bemisia tabaci , Resistance in African and South American Cassava Genotypes. J Integr Agric [Internet]. 2012;11(2):327–36. Available from: http://dx.doi.org/10.1016/S2095-3119(12)60017-3

122. Parsa S, Medina C, Rodríguez V. Sources of pest resistance in cassava. Crop Prot [Internet]. 2015;68:79–84. Available from: http://dx.doi.org/10.1016/j.cropro.2014.11.007

123. Naveen NC, Rahul C, Kumar D, Reb KB, Rajagopal R. Insecticide resistance status in the whitefly , Bemisia tabaci genetic groups Asia-I , Asia-II-1 and Asia-II-7 on the Indian subcontinent. 2017.

124. Sani I, Ismail SI, Abdullah S, Jalinas J, Jamian S, Saad N. A Review of the Biology and Control of Whitefly, Bemisia tabaci (Hemiptera: Aleyrodidae), with Special Reference to Biological Control Using Entomopathogenic Fungi. Insects. 2020;

125. Iffcobazar. Insect Growth Regulators: Meaning and Benefits [Internet]. 2023. Available from: https://www.iffcobazar.in/en/blog/insect-growth-regulators-meaning-and-benefits

126. Carrière Y, Ellers-Kirk C, Hartfield K, Larocque G, Degain B, Dutilleul P, et al. Large-scale, spatially-explicit test of the refuge strategy for delaying insecticide resistance. Proc Natl Acad Sci U S A. 2012;109(3):775–80.

127. Zubair M, Zuhaib M, Rauf I, Raza A, Hussain A, Hassan I, et al. Artificial micro RNA (amiRNA) -mediated resistance against whitefly (Bemisia tabaci) targeting three genes. Crop Prot [Internet]. 2020;137(June):105308. Available from: https://doi.org/10.1016/j.cropro.2020.105308

128. Ibrahim AB, Monteiro TR, Cabral GB, Aragão FJL. RNAi-mediated resistance to whitefly (Bemisia tabaci) in genetically engineered lettuce (Lactuca sativa). Transgenic Res. 2017;26(5):613–24.

129. Malik HJ, Raza A, Amin I, Scheffler JA, Scheffler BE, Brown JK, et al. RNAi-mediated mortality of the whitefly through transgenic expression of double-stranded RNA homologous to acetylcholinesterase and ecdysone receptor in tobacco plants. Sci Rep. 2016;6(May):1–11.

130. Olugboyega P, Benjamin G, Adejumobi I, Olusola T, Odom-Kolombia O, Adeosun T, et al. Semi-Autotrophic Hydroponics: A potential seed system technology for reduced breeding cycle and rapid quality seed delivery. 2019.

131. Hortidaily. Semi Autotrophic Hydroponic (SAH) laboratory for cassava in Rwanda

[Internet]. 2022 [cited 2023 Mar 3]. Available from: https://www.hortidaily.com/article/9453676/semi-autotrophic-hydroponic-sah-laboratory-for-cassava-in-rwanda/

132. Broek J van den, Byakweli JM. Exploratory study on Rwanda's Seed Sector : Options for Dutch support. Centre for Development Innovtion, Wageningen UR (University & Research centre). CDI report CDI-14-002. Wageningen. 2014.

133. Andrade-Piedra J, Bentley J, Almekinders C, Walsh K, Thiele G. Case Studies of Roots, Tubers and Banana Seed Systems. CGIAR Res Progr Roots, Tubers Banan (RTB),Lima RTB Work Pap N° 2016-3ISSN 2309-6586244P [Internet]. 2016;244. Available from: https://www.researchgate.net/publication/317199951

134. Tumwegamire S. Government of Rwanda adopts standards for quality declared seed (QDS) of cassava ,. 2021.

135. Legg J, Ndalahwa M, Yabeja J, Ndyetabula I, Bouwmeester H, Shirima R, et al. Community phytosanitation to manage cassava brown streak disease. Virus Res [Internet]. 2017;241(May):236–53. Available from: http://dx.doi.org/10.1016/j.virusres.2017.04.020

136. Mwangangi M, Ateka E, Nyende A, Kagundu A. Elimination of Cassava Brown Streak Virus from Infected Cassava. J Biol Agric Healthc. 2014;4(13):34–41.

137. Mohammed IU, Ghosh S, Maruthi MN. Generating virus-free cassava plants by in vitro propagation with chemical and heat treatment. African J Biotechnol. 2017;16(27):1551–60.

138. Ndunguru JC, Kidulile CE, Ateka EM, Alakonya AE. Efficacy of chemotherapy and thermotherapy in elimination of East African cassava mosaic virus from Tanzanian cassava landrace. J Phytopathol. 2018;(March):739–45.

139. Kanichi M. Production of Virus-Free Plants by Means of Meristem Culture. JARQ. 1971;6(1).

140. Thresh JM. CONTROL OF TROPICAL PLANT VIRUS DISEASES. Adv Virol. 2006;67(06):245–95.

141. Cha-um S, Hien T, Kirdmanee C. Disease-free Production of Sugarcane Varieties (Saccharum officinarum L.) Using in vitro Meristem Culture. Biotechnology. 2006;5(1682).

142. WALKEY DGA. High temperature inactivation of cucumber and alfalfa mosaic viruses in Nicotiana rustica cultures. Ann Appl Biol. 1976;84(2):183–92.

143. Szittya G, Silhavy D, Molnár A, Havelda Z, Lovas Á, Lakatos L, et al. Low temperature inhibits RNA silencing-mediated defence by the control of siRNA generation. EMBO J. 2003;22(3):633–40.

144. Chellappan P, Vanitharani R, Ogbe F, Fauquet CM. Effect of Temperature on Geminivirus-Induced RNA Silencing in Plants 1. Plant phy. 2005;138(August):1828–41.

145. Okori, J. B. ; Nakabonge G. In-vitro techniques for elimination of viruses causing cassava mosaic disease and cassava brown streak disease. RUFORUM Work Doc Ser.

2016;2(14):567–73.

146. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe : Direct molecular test by using ribavirin. PNAS. 2001;98(12).

147. Dietz J, Schelhorn S, Fitting D, Mihm U, Susser S, Welker M, et al. Deep Sequencing Reveals Mutagenic Effects of Ribavirin during Monotherapy of Hepatitis C Virus Genotype 1-Infected Patients. J Virol. 2013;87(11):6172–81.

148. Bektas Y, Eulgem T. Synthetic plant defense elicitors. Front Plant Sci. 2015;5(January):1–17.

149. Chen H, Zhang Z, Teng K, Lai J, Zhang Y, Huang Y, et al. Up-regulation of LSB1 / GDU3 affects geminivirus infection by activating the salicylic acid pathway. Plant J. 2010;12–23.

150. Shang J, Fei DX, Wang XS, Cao S, Zhao MXP, Zhang SJZ, et al. A broad-spectrum , e Y cient and nontransgenic approach to control plant viruses by application of salicylic acid and jasmonic acid. Planta. 2011;299–308.

151. Zhu F, Xi D, Yuan S, Xu F, Zhang D, Lin H. Salicylic Acid and Jasmonic Acid Are Essential for Systemic Resistance Against Tobacco mosaic virus in Nicotiana benthamiana. Mol Plant-Microbe Interact. 2014;27(6):567–77.

152. Li T, Huang Y, Xu Z, Wang F, Xiong A. Salicylic acid-induced differential resistance to the Tomato yellow leaf curl virus among resistant and susceptible tomato cultivars. BMC Plant Biol. 2019;1–14.

153. Trejo-Saavedra DL, Rivera-Bustamante RF, Garcia-Neria MA. Benzothiadiazole (BTH) induces resistance to Pepper golden mosaic virus (PepGMV) in pepper (Capsicum annuum L .). Biol Res. 2013;46(2008):333–40.

154. Smith-becker J, Keen NT, Becker JO. Acibenzolar-S-methyl induces resistance to Colletotrichum lagenarium and cucumber mosaic virus in cantaloupe. Crop Prot. 2003;22:769–74.

155. Li Y, Muhammad T, Wang Y, Zhang D, Crabbe MJC, Liang Y. Salicylic acid collaborates with gene silencing to tomato defense against tomato yellow leaf curl virus (TYLCV). Pakistan J Bot. 2018;50(5):2041–54.

156. Xie Z, Fan B, Chen C, Chen Z. An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. PNAS. 2001;98(11).

157. Tumwegamire S, Kanju E, Legg J, Shirima R, Kombo S, Mkamilo G, et al. Exchanging and managing in-vitro elite germplasm to combat Cassava Brown Streak Disease (CBSD) and Cassava Mosaic Disease (CMD) in Eastern and Southern Africa. Food Secur. 2018;(Cmd).

158. Maruthi MN, Whitfield EC, Otti G, Tumwegamire S, Kanju E, Legg JP, et al. A method for generating virus-free cassava plants to combat viral disease epidemics in Africa. Physiol Mol Plant Pathol [Internet]. 2019;105(January 2018):77–87. Available from: https://doi.org/10.1016/j.pmpp.2018.09.002

159. Nakabonge G, Nangonzi R, Tumwebaze BS, Kazibwe A, Samukoya C, Baguma Y. Production of virus-free cassava through hot water therapy and two rounds of

meristem tip culture. Cogent Food Agric [Internet]. 2020;6(1). Available from: https://doi.org/10.1080/23311932.2020.1800923

160. Dong OX, Ronald PC. Genetic engineering for disease resistance in plants: Recent progress and future perspectives. Plant Physiol. 2019;180(1):26–38.

161. Wilson TMA. Review Strategies to protect crop plants against viruses : Pathogen-derived resistance blossoms. Proc Natl Acad Sc USA. 1993;90(April):3134–41.

162. Kaniewski W, Lawson C, Sammons B, Haley L, Hart J, Delannay X, et al. Field resistance of transgenic Russet Burbank Potato to effects of infection by Potato Virus X and Potato Virus Y. Nat Publ Gr. 1990;

163. Reimann-philipp U. Expression of Coat Protein. In: Mechanisms of Resistance. 1998. p. 521–32.

164. Lindbo JA, Falk BW. The Impact of " Coat Protein-Mediated Virus Resistance " in Applied Plant Pathology and Basic Research. Phytopathology. 2017;624–34.

165. Rosa C, Kuo Y, Wuriyanghan H, Falk BW. RNA Interference Mechanisms and Applications in Plant Pathology. Annu Rev Phytopathol. 2018;

166. Ruiz-ferrer V, Voinnet O. Roles of Plant Small RNAs in Biotic Stress Responses. Annu Rev Plant Biol. 2009;

167. Borges F, Martienssen RA. The expanding world of small RNAs in plants. Nat Publ Gr. 2015;16(December).

168. Pooggin MM. Small RNA-Omics for Plant Virus Identification , Virome Reconstruction , and Antiviral Defense Characterization. Front Microbiol. 2018;9(November):1–20.

169. Waterhouse PM, Wang M, Lough T. Gene silencing as an adaptive defence against viruses. Nature. 2001;

170. Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. Geminiviruses: Masters at redirecting and reprogramming plant processes. Nat Rev Microbiol. 2013;11(11):777–88.

171. Bucher E, Lohuis D, Poppel PMJA Van, Geerts-dimitriadou C, Goldbach R, Prins M. Communication Multiple virus resistance at a high frequency using a single transgene construct. J Gen Virol. 2006;3697–701.

172. Ribeiro SG, Lohuis H, Goldbach R, Prins M. Tomato Chlorotic Mottle Virus Is a Target of RNA Silencing but the Presence of Specific Short Interfering RNAs Does Not Guarantee Resistance in Transgenic Plants □. J Virol. 2007;81(4):1563–73.

173. Pooggin MM. RNAi-mediated resistance to viruses: A critical assessment of methodologies. Elsevier. 2017;1–14.

174. Klas FE, Fuchs M, Gonsalves D. Comparative spatial spread overtime of Zucchini Yellow Mosaic Virus (ZYMV) and Watermelon Mosaic Virus (WMV) in fields of transgenic squash expressing the coat protein genes of ZYMV and WMV , and in fields of nontransgenic squash. Transgenic Res. 2006;(November).

175. Krubphachaya P, Jurícek M, Kertbundit S. Induction of RNA-mediated Resistance to

Papaya Ringspot Virus Type W. J Biochem Mol Biol. 2007;40(3):404–11.

176. Kundu JK, Briard P, Hily JM, Ravelonandro M, Scorza R. Role of the 25 – 26 nt siRNA in the resistance of transgenic Prunus domestica graft inoculated with plum pox virus. Virus Genes. 2008;215–20.

177. Collinge DB, Sarrocco S. Transgenic approaches for plant disease control : Status and prospects 2021. Plant Pathol. 2021;(July):1–19.

178. Vanderschuren H, Akbergenov R, Mikhail MP, Thomas H, Wilhelm G, Peng Z. Transgenic cassava resistance to African cassava mosaic virus is enhanced by viral DNA-A bidirectional promoter-derived siRNAs. Plant Mob Biol. 2007;549–57.

179. Lin ZJD, Taylor NJ, Bart R. Engineering Disease-Resistant Cassava. Cold Spring Harb Perspect Biol. 2019;

180. Patil BL, Ogwok E, Wagaba H, Mohammed IU, Yadav JS, Bagewadi B, et al. RNAi-mediated resistance to diverse isolates belonging to two virus species involved in Cassava brown streak disease. Mol Plant Pathol. 2011;12:31–41.

181. Yadav JS, Ogwok E, Wagaba H, Patil BL, Bagewadi B, Alicai T, et al. RNAi-mediated resistance to Cassava brown streak Uganda virus in transgenic cassava. Mol Plant Pathol. 2011;12:677–87.

182. Vanderschuren H, Moreno I, Anjanappa RB, Zainuddin IM, Gruissem W. Exploiting the Combination of Natural and Genetically Engineered Resistance to Cassava Mosaic and Cassava Brown Streak Viruses Impacting Cassava Production in Africa. PLoS One. 2012;7(9).

183. National Biodiversity Authority. NBA Approves Environmental Release Application for GM cassava. Kenya; 2021.

184. Ezezika OC, Nations U, Woman E, Child E, Marketplace I, Daar A, et al. Factors influencing agbiotech adoption and development in sub-Saharan Africa. Nat Biotechnol. 2012;(August 2016).

185. Gbadegesin, Lanre Anthony, Ayodeji E, Kwesi C, Anthony V, Olayemi O, Kojo J, et al. GMOs in Africa : Status , adoption and public acceptance. Food Control [Internet]. 2022;141(November 2021):109193. Available from: https://doi.org/10.1016/j.foodcont.2022.109193

186. Kaniewski WK, Thomas PE. The Potato Story. AgBioForum. 2004;7:41–6.

187. Yin K, Qiu JL. Genome editing for plant disease resistance: Applications and perspectives. Philos Trans R Soc B Biol Sci. 2019;374(1767).

188. Borrelli VMG, Brambilla V, Rogowsky P, Marocco A, Lanubile A. The enhancement of plant disease resistance using crispr/cas9 technology. Front Plant Sci. 2018;9(August).

189. Wheatley MS, Yang Y. Versatile applications of the CRISPR/cas toolkit in plant pathology and disease management. Phytopathology. 2021;111(7):1080–90.

190. Gomez MA, Lin ZD, Moll T, Chauhan RD, Hayden L, Renninger K, et al. Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and

nCBP-2 reduces cassava brown streak disease symptom severity and incidence. Plant Biotechnol J. 2019;17(2):421–34.

191. Tyagi S, Kumar R, Kumar V, Won SY, Shukla P. Engineering disease resistant plants through CRISPR-Cas9 technology. GM Crop Food. 2021;12(1):125–44.

192. Callaway E. CRISPR plants now subject to tough GM laws in European Union. Nature. 2018;560(7716):16.

193. Massart S, Olmos A, Jijakli H, Candresse T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res [Internet]. 2014;188:90–6. Available from: http://dx.doi.org/10.1016/j.virusres.2014.03.029

194. da Silva W, Kutnjak D, Xu Y, Xu Y, Giovannoni J, Elena SF, et al. Transmission modes affect the population structure of potato virus Y in potato. PLoS Pathog [Internet]. 2020;16(6 June):1–23. Available from: http://dx.doi.org/10.1371/journal.ppat.1008608

195. Bordería A V., Isakov O, Moratorio G, Henningsson R, Agüera-González S, Organtini L, et al. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. PLoS Pathog [Internet]. 2015;11(5):1–20. Available from: http://dx.doi.org/10.1371/journal.ppat.1004838

196. Mohammed IU, Abarshi MM, Muli B, Hillocks RJ, Maruthi MN. The Symptom and Genetic Diversity of Cassava Brown Streak Viruses Infecting Cassava in East Africa. Adv Virol. 2012;2012.

197. Mbewe W, Kumar PL, Changadeya W, Ntawuruhunga P, Legg J. Diversity , Distribution and Effects on Cassava Cultivars of Cassava Brown Streak Viruses in Malawi Diversity , Distribution and Effects on Cassava Cultivars of Cassava Brown Streak Viruses in Malawi. J Phytopathol. 2014;(January 2018).

198. Amisse J, Tairo F, Boykin LM, Virus C, Project A, Kehoe MA. Analyses of seven new whole genome sequences of cassava brown streak viruses in Mozambique reveals two distinct clades: evidence for new species. Plant Pathol. 2019;(February).

199. Saranraj P, Behera SS, Ray RC. Traditional Foods From Tropical Root and Tuber Crops. 2019.

200. Otekunrin OA, Sawicka B. Cassava , a 21st Century Staple Crop : How can Nigeria Harness Its Enormous Trade Potentials ? Acta Sci Agric. 2019;3(8).

201. Chavez AL, Sanchez T, Jaramillo G, Bedoya JM, Echeverry J. Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica. 2005;125–33.

202. Lobell BD, Burke BM, Tebaldi C, Mastrandrea DM, Falcon PW, Naylor LR. Needs for Food Security in 2030 Region. Science (80-). 2008;319(February):607–10.

203. FAO. FAOSTAT Statistical Database, Statistical Division. Rome; 2019.

204. Bull SE, Ndunguru J, Gruissem W, Beeching JR, Vanderschuren H. Cassava: Constraints to production and the transfer of biotechnology to African laboratories. Plant Cell Rep. 2011;30(5):779–87.

205. Legg JP, Thresh JM. Cassava virus diseases in Africa. Plant Virol sub-Saharan Africa. 2003;(June 2003).

206. Legg JP, Okao-Okuja G, Mayala R, Muhinyuza J-B. Spread into Rwanda of the severe cassava mosaic virus disease pandemic and the associated Uganda variant of East African cassava mosaic virus (EACMV-Ug). Plant Pathol. 2001;(November 2017).

207. Bua B. Farmers ' Knowledge of Cassava Streak Virus Disease in Selected Districts of Central Uganda. Int J Environ Agric Biotechnol. 2017;2(1):230–4.

208. Nyirahorana C, Mburu DM, Mulyungi P, Ntaganira E, Nduguru J, Sseruwagi P, et al. Drivers Behind Adoption Of Cassava Brown Streak Disease Control Measures In Rwanda. Int J Sci Technol Res. 2017;6(11):113–7.

209. Thresh JM, Cooter RJ. Strategies for controlling cassava mosaic virus disease in Africa. Plant Pathol. 2005;54:587–614.

210. Mbanzibwa DR, Tian YP, Tugume AK, Mukasa SB, Tairo F, Kyamanywa S, et al. Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch Virol. 2009;1996:353–9.

211. Muthoni J, Nyamongo DO. Seed systems in Kenya and their relationship to on-farm conservation of food crops. J New Seeds. 2008;9(4):330–42.

212. Masumba EA, Kapinga F, Mkamilo G, Salum K, Kulembeka H, Rounsley S, et al. QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. Theor Appl Genet. 2017;130(10):2069–90.

213. Anjanappa RB, Mehta D, Okoniewski MJ, Szabelska-Beręsewicz A, Gruissem W, Vanderschuren H. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. Mol Plant Pathol. 2018;19(2):476–89.

214. Shirima RR, Legg JP, Maeda DG, Tumwegamire S, Mkamilo G, Mtunda K, et al. Genotype by environment cultivar evaluation for cassava brown streak disease resistance in Tanzania. Virus Res. 2020;286(January).

215. Rwegasira GM, Rey MEC HN. Approaches to diagnosis and detection of Cassava Brown Streak Virus(Potiviridae, Ipomovirus) in field grown cassava crop. African J Food,Agriculture, Nutr Dev. 2011;11(3):4739–56.

216. Sseruwagi P, Sserubombwe WS, Legg JP, Ndunguru J, Thresh JM. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: A review. Virus Res. 2004;100(1):129–42.

217. Campo BVH, Hyman G, Bellotti A. Threats to cassava production: Known and potential geographic distribution of four key biotic constraints. Food Secur. 2011;3(3):329–45.

218. Rwegasira GM, Momanyi G, Rey MEC, Kahwa G, Legg JP. Widespread Occurrence and Diversity of Cassava brown streak virus (Potyviridae : Ipomovirus) in Tanzania. Virology. 2011;101(10):1159–67.

219. Koima IN, Orek CO, Nguluu SN. Distribution of Cassava Mosaic and Cassava Brown

Streak Diseases in Agro-Ecological Zones of Lower Eastern Kenya. Int J Innov Sci Res Technol. 2018;(January).

220. Muhindo H, Yasenge S, Casinga C, Songbo M, Dhed'a B, Alicai T, et al. Incidence, severity and distribution of Cassava brown streak disease in northeastern Democratic Republic of Congo. Cogent Food Agric. 2020;6(1).

221. Minot N, Smale M, Kling J, Eicher C. Seed Development Programs in Sub-Saharan Africa : A Review of Experiences Seed development programs in sub-Saharan Africa : A review of experiences Prepared by : Nicholas Minot Melinda Smale Submitted to : Contact information : Nicholas Minot Markets , Tr. ResearchGate. 2007;(May 2014).

222. Bentley J, Olanrewaju A, Madu T, Olaosebikan O, Abdoulaye T, Assfaw Wossen T, et al. Cassava farmers' preferences for varieties and seed dissemination system in Nigeria: gender and regional perspectives. 2017.

223. Catholic Relief Sercives. Final Report on the Great Lakes Cassava Initiative. Vol. 1. 2012.

224. Mukasa S. Developing a community-based cassava seed system for increased productivity and market linkages in Uganda; 2015.

225. Wossen T, Spielman DJ, Abdoulaye T, Kumar PL. The cassava seed system in Nigeria : Opportunities and challenges for policy and regulatory reform. 2020.

226. Mbewe W, Hanley-Bowdoin L, Ndunguru J, Duffy S. CHAPTER 7: Cassava Viruses: Epidemiology, Evolution, and Management. In 2020. p. 133–57.

227. Tomlinson KR, Bailey AM, Alicai T, Seal S, Foster GD. Cassava brown streak disease: historical timeline, current knowledge and future prospects. Mol Plant Pathol. 2018;19(5):1282–94.

228. Kwibuka Y, Nyirakanani C, Bizimana JP, Bisimwa E, Brostaux Y, Massart S. Risk factors associated with cassava brown streak disease dissemination through seed pathways in Eastern. Front Plant Sci. 2022;(July):1–18.

229. Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol. 2018;16(8):1–6.

230. Kathurima TM, Ateka EM. Diversity and Phylogenetic Relationships of Full Genome Sequences of Cassava Brown Streak Viruses in Kenya. Biotechnol J Int. 2019;23(3):1–11.

231. Mbanzibwa DR, Tian YP, Tugume AK, Mukasa SB, Tairo F, Kyamanywa S, et al. 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. J Virol Methods [Internet]. 2011;171(2):394–400. Available from: http://dx.doi.org/10.1016/j.jviromet.2010.09.024

232. Lane AA, Odejide O, Kopp N, Kim S, Yoda A, Erlich R, et al. Low frequency clonal mutations recoverable by deep sequencing in patients with aplastic anemia. Leukemia. 2013;27(4):968–71.

233. Fuentes S, Gibbs AJ, Adams IP, Hajizadeh M, Kreuze J, Fox A, et al. Phylogenetics and Evolution of Potato Virus V: Another Potyvirus that Originated in the Andes. Plant

Dis. 2022;106(2):691–700.

234. Kawakubo S, Gao F, Li S, Tan Z, Huang YK, Adkar-Purushothama CR, et al. Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road. Proc Natl Acad Sci U S A. 2021;118(12):1–10.

235. Mbewe W, Winter S, Mukasa S, Tairo F. Deep Sequencing Reveals a Divergent Ugandan cassava brown streak virus isolate from Malawi. Viruses. 2017;43:8–9.

236. Adams IP, Harju V., Hodges T, Hany U, Skelton A, Rai S, et al. First report of maize lethal necrosis disease in Rwanda. New Dis Reports. 2014;29(1):22–22.

237. Cuevas JM, Willemsen A, Hillung J, Zwart MP, Elena SF. Temporal dynamics of intrahost molecular evolution for a plant RNA virus. Mol Biol Evol. 2015;32(5):1132–47.

238. Simmons HE, Dunham JP, Stack JC, Dickins BJA, Pagán I, Holmes EC, et al. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. J Gen Virol. 2012;93(8):1831–40.

239. Tamisier L, Colson C, Maclot F, Zhang P, Wang X. The tree that hides the forest : going beyond the consensus to analyse within-plant virus population diversity (in preparation). (4):1–23.

240. Kim NK, Lee HJ, Kim SM, Jeong RD. Identification of Viruses Infecting Oats in Korea by Metatranscriptomics. Plants. 2022;11(3).

241. Hanafi M, Rong W, Tamisier L, Berhal C, Roux N, Massart S. Detection of Banana Mild Mosaic Virus in Musa In Vitro Plants: High-Throughput Sequencing Presents Higher Diagnostic Sensitivity Than (IC)-RT-PCR and Identifies a New Betaflexiviridae Species. Plants. 2022;11(2).

242. Weir B, Clark Cockerham C. Estimating F-Statistics for the Analysis of Population Structure. Evolution (N Y). 1984;38(6):1358–70.

243. Wijayasekara D, Ali A. Evolutionary study of maize dwarf mosaic virus using nearly complete genome sequences acquired by next-generation sequencing. Sci Rep [Internet]. 2021;11(1):1–14. Available from: https://doi.org/10.1038/s41598-021-98299-9

244. Hill T, Unckless RL. Recurrent evolution of high virulence in isolated populations of a DNA virus. Elife. 2020;9:1–53.

245. Mehta D, Stürchler A, Anjanappa RB, Zaidi SSEA, Hirsch-Hoffmann M, Gruissem W, et al. Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. Genome Biol. 2019;20(1):1–10.

246. NISR. Seasonal Agricultural Survey (SAS) ANNUAL REPORT 2019. 2019.

247. Nduwumuremyi A, Kabirigi M. Yield Gap Analysis of Key Agricultural Commodities in Rwanda. 2018.

248. Rwegasira G, Rey M, Nawabu H. Approaches to Diagnosis and Detection of Cassava Brown Streak Virus (Potiviridae, Ipomovirus) In Field-Grown Cassava Crop. African

J Food,Agriculture, Nutr Dev. 2011;11(3):4739–56.

249. Ling Z, Zhike Z, Shunquan L, Tingting Z, Xianghui Y. Evaluation of Six Methods for Extraction of Total RNA from Loquat. Not Bot Horti Agrobot Cluj-Napoca. 2013;41(1):313–6.

250. Kechin A, Boyarskikh U, Kel A, Filipenko M. CutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017;24(11):1138–43.

251. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. PLoS One. 2017;12(10):1–15.

252. Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. Gigascience. 2019;8(9):1–13.

253. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

254. Pond SK, Wadhawan S, Chiaromonte F, Ananda G, Chung WY, Taylor J, et al. Windshield splatter analysis with the Galaxy metagenomic pipeline. Genome Res. 2009;19(11):2144–53.

255. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;

256. Zuckerkandl E, Pauling L. Molecules as documents of history. J Theor Biol. 1965;8(2):357–66.

257. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7(539).

258. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.

259. Nei M, Kumar S. Molecular Evolution and Phylogenetics. Vol. 31, Oxford University Press. New York; 2000. 1029–1029 p.

260. Nelson CW, Moncla LH, Hughes AL. SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinformatics. 2015;31(22):3709–11.

261. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27(24):3435–6.

262. Schumpp O, Dupuis B, Bréchon A, Wild W, Frei P, Pellet D, et al. Diagnostic moléculaire à haut débit pour détecter les viroses des plants de pomme de terre. Rech Agron Suisse. 2016;7(10):456–65.

263. Maclot F, Candresse T, Filloux D, Malmstrom CM, Roumagnac P, van der Vlugt R, et al. Illuminating an Ecological Blackbox: Using High Throughput Sequencing to Characterize the Plant Virome Across Scales. Front Microbiol. 2020;11(October):1–

16.

264. Wallace MA, Coffman KA, Gilbert C, Ravindran S, Albery GF, Abbott J, et al. The discovery, distribution, and diversity of DNA viruses associated with Drosophila melanogaster in Europe. Virus Evol. 2021;7(1):1–23.

265. Roberts JMK, Ireland KB, Tay WT, Paini D. Honey bee-assisted surveillance for early plant virus detection. Ann Appl Biol. 2018;173(3):285–93.

266. Fowkes AR, McGreig S, Pufal H, Duffy S, Howard B, Adams IP, et al. Integrating high throughput sequencing into survey design reveals turnip yellows virus and soybean dwarf virus in pea (Pisum sativum) in the united kingdom. Viruses. 2021;13(12):1–18.

267. Clerc S Le, Coulonges C, Delaneau O, Manen D Van, Herbeck T, Limou S, et al. SCREENING LOW FREQUENCY SNPS FROM GENOME WIDE ASSOCIATION STUDY REVEALS A NEW RISK ALLELE FOR. J Acquir Immune Defic Syndr. 2011;56(3):279–84.

268. Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, et al. Population genomics of intrapatient HIV-1 evolution. Elife. 2015;4(DECEMBER2015):1–26.

269. Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC. Validation of SNP Allele Frequencies Determined by Pooled Next-Generation Sequencing in Natural Populations of a Non-Model Plant Species. PLoS One. 2013;8(11).

270. Retel C, Kowallik V, Becks L, Feulner PGD. Strong selection and high mutation supply characterize experimental Chlorovirus evolution. Virus Evol. 2022;8(1):1–14.

271. McCann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, Rees-George J, et al. Genomic Analysis of the Kiwifruit Pathogen Pseudomonas syringae pv. actinidiae Provides Insight into the Origins of an Emergent Plant Disease. PLoS Pathog. 2013;9(7).

272. Aimone CD, Lavington E, Hoyer JS, Deppong DO, Mickelson-Young L, Jacobson A, et al. Population diversity of cassava mosaic begomoviruses increases over the course of serial vegetative propagation. J Gen Virol. 2021;102(7).

273. Huang T. Next generation sequencing to characterize mitochondrial genomic DNA heteroplasmy. Curr Protoc Hum Genet. 2011;(SUPPL. 71):1–12.

274. Vu A, Stankovi I. Detection of Four New Tomato Viruses in Serbia Using Post Hoc High-Throughput Sequencing Analysis of Samples From a Large-Scale Field Survey. Plant Dis. 2021;(September):2325–32.

275. Villamor DEV, Keller KE, Martin RR, Tzanetakis IE. Comparison of High Throughput Sequencing to Standard Protocols for Virus Detection in Berry Crops. Plant Dis. 2022;106(2):518–25.

276. Pablo-rodríguez JL, Bailey AM, Foster GD. Characterization of Cassava brown streak virus proteins draw their sides during mixed infections and reveal P1 as a silencing suppressor. Res Sq. 2022;1–22.

277. Ivanov KI, Mäkinen K. Coat proteins, host factors and plant viral replication. Curr Opin Virol. 2012;2(6):712–8.

278. Fox A. Reconsidering causal association in plant virology. Plant Pathol. 2020;69(6):956–61.

279. Jarvis A, Ramirez-villegas J, Vanessa B, Campo H, Navarro-racines C. Is Cassava the Answer to African Climate Change Adaptation ? Trop Plant Biol. 2012;9–29.

280. Hillocks RJ, Maruthi MN. Post-harvest impact of cassava brown streak disease in four countries in eastern Africa. Food Chain. 2015;5(1–2):116–22.

281. Masinde EA, Mkamillo G, Ogendo JO, Hillocks R, Mulwa RMS, Kimata B, et al. Genotype by environment interactions in identifying cassava (Manihot esculenta Crantz) resistant to cassava brown streak disease. F Crop Res. 2018;215(September 2017):39–48.

282. Kaweesi T, Kawuki R, Kyaligonza V, Baguma Y, Tusiime G, Ferguson ME. Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. Virol J. 2014;11(1).

283. Zinga I, Chiroleu F, Kamba E, Giraud-carrier C, Harimalala M, Komba EK, et al. Field Evaluation of the Effectiveness of Thermotherapy against Cassava Mosaic Disease in Central African Republic. Am J Exp Agric. 2014;4(July 2011):1232–41.

284. Wasswa P, Alicai T, Mukasa SB. Optimization of in vitro techniques for cassava brown streak virus elimination from infected cassava clones. African Crop Sci J. 2010;18(4):235–41.

285. Wang Q, J. CUELLAR W, RAJAMÄKI M-L, HIRATA Y, JARI P. T. V. Combined thermotherapy and cryotherapy for efficient virus eradication : relation of virus distribution , subcellular changes , cell survival and viral RNA degradation in shoot tips. Mol Plant Pathol. 2008;9:237–50.

286. Velazquez K, Renovell A, Comellas M, Serra P, Garcı ML, Navarro L, et al. Effect of temperature on RNA silencing of a negative- stranded RNA plant virus : Citrus psorosis virus. Plant Pathol. 2010;982–90.

287. Chellappan P, Vanitharani R, Ogbe F, Fauquet CM. Effect of Temperature on Geminivirus-Induced RNA Silencing in Plants 1. Plant Physiol. 2005;138(August):1828–41.

288. Tahmasebi A, Khahani B, Tavakol E, Afsharifar A, Shahid MS. Microarray analysis of Arabidopsis thaliana exposed to single and mixed infections with Cucumber mosaic virus and turnip viruses. Physiol Mol Biol Plants [Internet]. 2021;27(1):11–27. Available from: https://doi.org/10.1007/s12298-021-00925-3

289. Sharaf A, Nuc P, Ripl J, Alquicer G, Ibrahim E, Wang X, et al. Transcriptome Dynamics in Triticum aestivum Genotypes Associated with Resistance against the Wheat Dwarf Virus. Viruses. 2023;15(3):689.

290. Osmani Z, Sabet MS, Shams-Bakhsh M, Moieni A, Vahabi K. Virus-specific and common transcriptomic responses of potato (Solanum tuberosum) against PVY, PVA and PLRV using microarray meta-analysis. Plant Breed. 2019;138(2):216–28.

291. Cheng Z, Yu X, Li S, Wu Q. Genome-wide transcriptome analysis and identification of benzothiadiazole-induced genes and pathways potentially associated with defense

response in banana. BMC Genomics. 2018;19(1):1–19.

292. Wang X, Li N, Li W, Gao X, Cha M, Qin L, et al. Advances in Transcriptomics in the Response to Stress in Plants. Glob Med Genet. 2020;07(02):030–4.

293. Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One. 2014;9(5).

294. Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, et al. The Cassava Genome: Current Progress, Future Directions. Trop Plant Biol. 2012;5(1):88–94.

295. Anjanappa RB, Mehta D, Okoniewski MJ, Szabelska A, Gruissem W, Vanderschuren H. Early transcriptome response to brown streak virus infection in susceptible and resistant cassava varieties. bioRxiv. 2017;1–22.

296. Kavil S, Otti G, Bouvaine S, Armitage A, Maruthi MN. PAL1 gene of the phenylpropanoid pathway increases resistance to the Cassava brown streak virus in cassava. Virol J. 2021;18(1):1–10.

297. Amuge T, Berger DK, Katari MS, Myburg AA, Goldman SL, Ferguson ME. A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. Sci Rep [Internet]. 2017;7(1):1–21. Available from: http://dx.doi.org/10.1038/s41598-017-09617-z

298. Oostendorp M, Kunz W, Dietrich B, Staub T. Induced disease resistance in plants by chemicals. Eur J Plant Pathol. 2001;19–28.

299. GOMEZ-VASQUEZ R, DAY R, BUSCHMANN H, RANDLES S, BEECHING JR, COOPER RM. Phenylpropanoids , Phenylalanine Ammonia Lyase and Peroxidases in Elicitor-challenged Cassava (Manihot esculenta) Suspension Cells and Leaves. Ann Bot. 2004;

300. Thakur M, Sohal BS. Role of Elicitors in Inducing Resistance in Plants against Pathogen Infection : A Review. Hindawi Publ Corp ISRN Biochem. 2013;2013.

301. LOON LC VAN, STRIEN EA VAN. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol. 1999;85–97.

302. HEIL M, BOSTOCK RM. Induced Systemic Resistance (ISR) Against Pathogens in the Context of Induced Plant Defences. Ann Appl Biol. 2002;

303. Xi D, Li X, Gao L, Zhang Z, Zhu Y, Zhu H. Application of exogenous salicylic acid reduces disease severity of Plasmodiophora brassicae in pakchoi (Brassica campestris ssp. chinensis Makino). PLoS One [Internet]. 2021;16(6 June):1–11. Available from: http://dx.doi.org/10.1371/journal.pone.0248648

304. Li T, Huang Y, Xu Z, Wang F, Xiong A. Salicylic acid-induced differential resistance to the Tomato yellow leaf curl virus among resistant and susceptible tomato cultivars. 2019;1–14.

305. Hunter LJR, Westwood JH, Heath G, Macaulay K, Smith AG, Macfarlane SA, et al. Regulation of RNA-Dependent RNA Polymerase 1 and Isochorismate Synthase Gene

Expression in Arabidopsis. PLoS One. 2013;8(6).

306. Diao P, Zhang Q, Sun H, Ma W, Cao A, Yu R, et al. miR403a and SA Are Involved in NbAGO2 Mediated Antiviral Defenses Against TMV Infection in Nicotiana benthamiana. Genes (Basel). 2019;526(10).

307. Lawton KA, Friedrich L, Hunt M, Weymann K, Delaney T, Kessmann H, et al. Benzothiadiazole induces disease resistance in Arabidopsis by activation of the systemic acquired resistance signal transduction pathway. Plant J. 1996;10(1):71–82.

308. Abarshi MM, Mohammed IU, Wasswa P, Hillocks RJ, Holt J, Legg JP, et al. Optimization of diagnostic RT-PCR protocols and sampling procedures for the reliable and cost-effective detection of Cassava brown streak virus. J Virol Methods. 2010;163(2):353–9.

309. Obrepalska-Steplowska A, Pospieszny H, Smiglak M, Patryk F. Assessment of the Efficacy and Mode of Action of Benzo(1,2,3)-Thiadiazole-7-Carbothioic Acid S-Methyl Ester (BTH) and Its Derivatives in Plant Protection Against Viral Disease. Int J Mol Sci. 2019;

310. Zeyimo B, Zola M, Boykin L, Pita J, Amand M, Monde G, et al. Original Research Article Original Research Article Open Access First Report and Preliminary Evaluations of Cassava Brown Streak-Like. Int J Dev Res. 2018;08:22400–7.

311. Moreno I, Gruissem W, Vanderschuren H. Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. J Virol Methods [Internet]. 2011;177(1):49–54. Available from: http://dx.doi.org/10.1016/j.jviromet.2011.06.013

312. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):1–12.

313. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245–50.

314. Pfaffl MW, Tichopad A, Prgomet C, Neuvians T. Determination of most stable housekeeping genes, differentially regulated target genes and sample integrity : BestKeeper. Biotechnol Lett. 2004;26:509–15.

315. Andrews S. FastQC: a quality control tool for high throughput data [Internet]. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

316. Bolger AM, Lohse M, Usadel B. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

317. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods [Internet]. 2015;12(4):357–60. Available from: https://doi.org/10.1038/nmeth.3317

318. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

319. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.

320. Yi X, Du Z, Su Z. PlantGSEA: a gene set enrichment analysis toolkit for plant community. Nucleic Acids Res. 2013;41(Web Server issue):98–103.

321. Tian F, Yang DC, Meng YQ, Jin J, Gao G. PlantRegMap: Charting functional regulatory maps in plants. Nucleic Acids Res. 2020;48(D1):D1104–13.

322. Almagro L, Gómez Ros L V., Belchi-Navarro S, Bru R, Ros Barceló A, Pedreño MA. Class III peroxidases in plant defence reactions. J Exp Bot. 2009;60(2):377–90.

323. Santos AA, Lopes KVG, Apfata JAC, Fontes EPB. NSP-interacting kinase, NIK: A transducer of plant defence signalling. J Exp Bot. 2010;61(14):3839–45.

324. Chung H, Sunter G. Interaction between the transcription factor AtTIFY4B and begomovirus AL2 protein impacts pathogenicity. Plant Mol Biol. 2014;(October).

325. Amorim L, Santos R, Neto J, Guida-Santos M, Crovella S, Benko-Iseppon A. Transcription Factors Involved in Plant Resistance to Pathogens. Curr Protein Pept Sci. 2016;18(4):335–51.

326. Xu J, Wang XY, Guo WZ. The cytochrome P450 superfamily: Key players in plant development and defense. J Integr Agric [Internet]. 2015;14(9):1673–86. Available from: http://dx.doi.org/10.1016/S2095-3119(14)60980-1

327. Ogwok E, Ilyas M, Alicai T, Rey MEC, Taylor NJ. Comparative analysis of virus-derived small RNAs within cassava (Manihot esculenta Crantz) infected with cassava brown streak viruses. Virus Res [Internet]. 2016;215:1–11. Available from: http://dx.doi.org/10.1016/j.virusres.2016.01.015

328. Faoro F, Gozzo F. Is modulating virus virulence by induced systemic resistance realistic ? Plant Sci [Internet]. 2015;234:1–13. Available from: http://dx.doi.org/10.1016/j.plantsci.2015.01.011

329. Palukaitis P, Yoon J, Choi S, Carr JP. ScienceDirect Manipulation of induced resistance to viruses. Curr Opin Virol [Internet]. 2017;26:141–8. Available from: http://dx.doi.org/10.1016/j.coviro.2017.08.001

330. Dempsey DA, Klessig DF. How does the multifaceted plant hormone salicylic acid combat disease in plants and are similar mechanisms utilized in humans? BMC Biol. 2017;15(1):1–11.

331. Tian M, Sasvari Z, Gonzalez PA, Friso G, Rowland E, Liu X, et al. Salicylic Acid Inhibits the Replication of Tomato bushy stunt virus by Directly Targeting a Host Component in the Replication Complex. Am Phythological Soc. 2015;28(4):379–86.

332. Aleksandra O-S, Patryk F, Henryk P, Marcin S. Assessment of the Efficacy and Mode of Action of S-Methyl Ester (BTH) and Its Derivatives in Plant Protection Against Viral Disease. Int J Mol Sci. 2019;20.

333. Anfoka GH. systemic resistance in tomato (Lycopersicon esculentum . Mill cv . Vollendung) to Cucumber mosaic v irus. Crop Prot. 2000;19:401–5.

334. López-Gresa MP, Lisón P, Yenush L, Conejero V, Rodrigo I, Bellés JM. Salicylic acid

is involved in the basal resistance of tomato plants to citrus exocortis viroid and tomato spotted wilt virus. PLoS One. 2016;11(11):1–25.

335. Nombela G, Pascual S, Aviles M, Guillard E. Benzothiadiazole Induces Local Resistance to Bemisia tabaci (Hemiptera : Aleyrodidae) in Tomato Plants. J Econ Entomol. 2005;98(6):2266–71.

336. Nascimento LC, Pio-Ribeiro G, Willadino L, Andrade GP. Stock Indexing and Potato Virus Y Elimination from Potato Plants Cultivated in Vitro. Sci Agric. 2003;60(3):525–30.

337. Coenye T. Do results obtained with RNA-sequencing require independent verification? Biofilm [Internet]. 2021;3(January):100043. Available from: https://doi.org/10.1016/j.bioflm.2021.100043

338. Mehta D, Ahkami AH, Walley J, Xu S, Uhrig RG. The incongruity of validating quantitative proteomics using western blots. Nat plants. 2022;8(December):1320–1.

339. Everaert C, Luypaert M, Maag JL V, Cheng QX, Marcel E, Hellemans J, et al. Benchmarking of RNA-sequencing analysis workflows using whole- transcriptome RT-qPCR expression data. Sci Rep [Internet]. 2017;(July 2016):1–11. Available from: http://dx.doi.org/10.1038/s41598-017-01617-3

340. Shi Y, He M. Differential gene expression identi fi ed by RNA-Seq and qPCR in two sizes of pearl oyster (Pinctada fucata). Gene [Internet]. 2014;538(2):313–22. Available from: http://dx.doi.org/10.1016/j.gene.2014.01.031

341. Freeborough W, Gentle N, Rey MEC. WRKY Transcription Factors in Cassava Contribute to Regulation of Tolerance and Susceptibility to Cassava Mosaic Disease through Stress Responses. MDPI Viruses. 2021;

342. Zeng T, Hongbo L, Deyun Q, Yan Z, Xianghua L, Caiguo X, et al. A Pair of Allelic WRKY Genes Play Opposite Roles in. Plant Physiol. 2009;151(October):936–48.

343. Bjorn H, Hahlbrock K. The 4-coumarate : CoA ligase gene family in Arabidopsis thaliana comprises one rare , sinapate- activating and three commonly occurring isoenzymes. PNAS. 2004;101(7):1–6.

344. Pandian BA, Sathishraj R, Djanaguiraman M, Prasad PVV, Jugulam M. Role of cytochrome P450 enzymes in plant stress response. MDPI Antioxidants. 2020;9(5):1–15.

345. Kaweesi T, Kyaligonza V, Baguma Y, Kawuki R, Morag F. Inbreeding enhances field resistance to cassava brown streak viruses. J Plant Breed Crop Sci. 2016;8(8):138–49.

346. Ueki S, Spektor R, Natale DM, Citovsky V. ANK, a host cytoplasmic receptor for the Tobacco mosaic virus cell-to-cell movement protein, facilitates intercellular transport through plasmodesmata. PLoS Pathog. 2010;6(11).

347. Jones RAC. Plant virus ecology and epidemiology: Historical perspectives, recent progress and future prospects. Ann Appl Biol. 2014;164(3):320–47.

348. Kumakech A, Acipa A, Tumwine V, Maiteki GA. Knowledge on cassava disease management: The case of cassava brown streak disease awareness in Northern Uganda. African J Plant Sci. 2013;7(12):597–601.

349. Chipeta MM, Shanahan P, Melis R, Sibiya J, Benesi IRM. Farmers' knowledge of cassava brown streak disease and its management in Malawi. Int J Pest Manag [Internet]. 2016;62(3):175–84. Available from: https://doi.org/10.1080/09670874.2016.1167268

350. Kanju E, Uzokwe VNE, Ntawuruhunga P, Tumwegamire S, Yabeja J, Pariyo A, et al. Varietal response of cassava root yield components and root necrosis from cassava Brown streak disease to time of harvesting in Uganda. Crop Prot [Internet]. 2019;120:58–66. Available from: https://doi.org/10.1016/j.cropro.2019.02.013

351. Almekinders CJM, Walsh S, Jacobsen KS, Andrade-piedra JL, Mcewan MA, Haan S De, et al. Why interventions in the seed systems of roots , tubers and bananas crops do not reach their full potential. Food Secur. 2019;23–42.

352. McEwan MA, Almekinders CJM, Andrade-Piedra JJL, Delaquis E, Garrett KA, Kumar L, et al. "Breaking through the 40% adoption ceiling: Mind the seed system gaps." A perspective on seed systems research for development in One CGIAR. Outlook Agric. 2021;50(1):5–12.

353. Mukwa LFT, Muengula M, Zinga I, Kalonji A, Iskra-Caruana ML, Bragard C. Occurrence and Distribution of Banana bunchy top virus Related Agro-Ecosystem in South Western, Democratic Republic of Congo. Am J Plant Sci. 2014;05(05):647–58.

354. Filho AB, Inoue-Nagata AK, Bassanezi RB, Belasque J, Amorim L, Macedo MA, et al. The importance of primary inoculum and area-wide disease management to crop health and food security. Food Secur. 2016;8(1):221–38.

355. Evans EA, Crane JH, Ploetz RC, Ballen FH. Cost-benefit analysis of area-wide management of Laurel wilt disease in Florida commercial avocado production area. Mark Commer Stream. 2015;(February 2011):467–71.

356. Bassanezi RB, Montesino LH, Gimenes-Fernandes N, Yamamoto PT, Gottwald TR, Amorim L, et al. Efficacy of area-wide inoculum reduction and vector control on temporal progress of huanglongbing in young sweet orange plantings. Plant Dis. 2013;97(6):789–96.

357. Maroya N, Balogun M, Aighewi B, Mignouna DB, Kumar PL, Asiedu R. Transforming Yam Seed Systems in West Africa. In: Root, Tuber and Banana Food System Innovations. 2022.

358. Lu L, Wu S, Jiang J, Liang J, Zhou X, Wu J. Whole genome deep sequencing revealed host impact on population structure, variation and evolution of Rice stripe virus. Virology [Internet]. 2018;524(March):32–44. Available from: https://doi.org/10.1016/j.virol.2018.08.005

359. Kutnjak D, Elena FS, Ravnika M. Time-Sampled Population Sequencing Reveals the Interplay of Selection and Genetic Drift in Experimental Evolution of Potato Virus Y. J Virol. 2017;91(16):1–17.

360. Minicka J, Zarzyńska-Nowak A, Budzyńska D, Borodynko-Filas N, Hasiów-Jaroszewska B. High-throughput sequencing facilitates discovery of new plant viruses in Poland. Plants. 2020;9(7):1–20.

361. Zhang X, Dong J, Liu H, Wang J, Qi Y, Liang Z. Transcriptome Sequencing in

Response to Salicylic Acid in Salvia miltiorrhiza. PLoS One. 2016;1–27.

362. Ge Y, Wang Y, Han J, Lu Y, Yue X, Zhang X, et al. Scientia Horticulturae Transcriptome analysis reveals resistance induced by Benzothiadiazole against soft rot in Chinese cabbage. Sci Hortic (Amsterdam) [Internet]. 2023;315(September 2022):111978. Available from: https://doi.org/10.1016/j.scienta.2023.111978

363. Li Y, Qin L, Zhao J, Muhammad T, Cao H, Li H, et al. SlMAPK3 enhances tolerance to tomato yellow leaf curl virus (TYLCV) by regulating salicylic acid and jasmonic acid signaling in tomato (Solanum lycopersicum). PLoS One. 2017;12(2):1–21.

364. Neerja S, Sohal BS, Lore JS. Foliar Application of Benzothiadiazole and Salicylic Acid to Combat Sheath Blight Disease of Rice. Rice Sci. 2013;20(5):349–55.

365. Khan MIR, Fatma M, Per TS, Anjum NA, Khan NA. Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants. Front Plant Sci. 2015;6(JUNE):1–17.

366. Jarecka-Boncela A, Spychalski M, Ptaszek M, Włodarek A, Smiglak M, Kukawka R. The Effect of a New Derivative of Benzothiadiazole on the Reduction of Fusariosis and Increase in Growth and Development of Tulips. Agric. 2023;13(4).

367. Walters D, Walsh D, Newton A, Lyon G. Induced resistance for plant disease control: Maximizing the efficacy of resistance elicitors. Phytopathology. 2005;95(12):1368–73.

368. Mandal A, Sarkar B, Mandal S, Vithanage M. Impact of agrochemicals on soil health [Internet]. Agrochemicals Detection, Treatment and Remediation. LTD; 2020. 161–188 p. Available from: http://dx.doi.org/10.1016/B978-0-08-103017-2.00007-6

369. Tudi M, Ruan HD, Wang L, Lyu J, Sadler R, Connell D, et al. Agriculture development, pesticide application and its impact on the environment. Int J Environ Res Public Health. 2021;18(3):1–24.

370. Koo YM, Heo AY, Choi HW. Salicylic Acid as a Safe Plant Protector and Growth Regulator. Plant Pathol J. 2020;36(1):1–10.

371. Jiang Y, Li J, Bi Y, Ma T, Zhang B. Pre-Harvest Benzothiadiazole Spraying Promotes the Cumulation of Phenolic Compounds in Grapes. MDPI foods. 2022;31–3.

372. Sandroni M, Liljeroth E, Mulugeta T, Alexandersson E. Plant resistance inducers (PRIs): perspectives for future disease management in the field. CAB Rev. 2020;(001).

373. MINAGRI. Rwanda's Agriculture sector transformation journey over the last 27 years [Internet]. 2023 [cited 2023 Sep 18]. Available from: https://www.minagri.gov.rw/updates/news-details/rwandas-agriculture-sector-transformation-journey-over-the-last-27-years

374. MINAGRI. National Agriculture Policy. 2017.

375. Gisel A, Stavolone L, Olagunju T, Landi M, van Damme R, Niazi A, et al. EpiCass and CassavaNet4Dev Advanced Bioinformatics Workshop. EMBnet.journal, 29, e1045. 2023;

376. AIMS-NEI. Next Einstein Initiative (AIMS-NEI) [Internet]. 2022 [cited 2023 Jul 27].

Available from: https://www.wise-qatar.org/project/next-einstein-initiative-south-africa/

377. Njenga GL. Global One Health initiative recognizes ILRI as a long-standing partner [Internet]. 2021 [cited 2023 Jul 27]. Available from: https://www.ilri.org/news/global-one-health-initiative-recognizes-ilri-long-standing-partner

Appendices

Peer review publications and Scientific communication

Appendices. Peer review publications and Scientific communication

1. Peer review publications

1.1. Publications and manuscripts included in the thesis

This thesis is a sum of three manuscript articles including one published article (Article 1), one article accepted for publication (Article 2) and one manuscript under review (Article 3).

Article 1 : Nyirakanani C, Bizimana JP, Kwibuka Y, Nduwumuremyi A, Bigirimana V de P, Bucagu C, *et al*. Farmer and Field Survey in Cassava-Growing Districts of Rwanda Reveals Key Factors Associated with Cassava Brown Streak Disease Incidence and Cassava Productivity. Front Sustain Food Syst. 2021;5(December):1–14.

Article 2: Nyirakanani C, Tamisier L, Bizimana JP, Rollin J, Nduwumuremyi A, Bigirimana, V, Selmi I, *et al*. Going beyond consensus genome sequences: an innovative SNP-based methodology reconstructs different Uganda cassava brown streak virus haplotypes at the country-wide level. Virus Evol. 2023;14–27.

Article 3: Nyirakanani C, Picolo P A, Bizmana, JP, Vanderschuren H, Greenhouse thermotherapy combined with chemotherapy and field chemotherapy effectively reduces viral load and severity of Cassava Brown Streak Disease. ***Manuscript.***

1.2. Other peer review publications

1. Kwibuka Y, **Nyirakanani C**, Bizimana JP, Bisimwa E, Brostaux Y, Massart S. Risk factors associated with cassava brown streak disease dissemination through seed pathways in Eastern. Front Plant Sci. 2022;(July):1–18.

2. Shakir S, Zaidi SS-A, Hashemi FSG, **Nyirakanani C**, Vanderschuren H. Harnessing plant viruses in the metagenomics era: from the development of infectious clones to applications. Trends Plant Sci [Internet]. 2022;1–15. Available from: https://doi.org/10.1016/j.tplants.2022.10.005

2. Scientific Communication

1. **Chantal Nyirakanani** (2023). *A nationwide analysis of the Ugandan cassava brown streak virus genomes in Rwanda suggests an association of the virus population with the deployment of tolerant cassava cultivars.* Rencontres de Virologie Végétale (RVV2023) in Centre Paul Langevin, Aussois, France. January 15-19, 2023.

2. **Chantal Nyirakanani** (2021). "*Mitigation strategies for cassava brown streak disease (CBSD) in Rwanda.* virtual live Keystone eSymposia joint conference, Plant Genome Engineering: From Lab to Field. March 8-9, 2021.

3. **Chantal Nyirakanani** (2020). *Cassava seed system, from field to laboratory: farmers' practices, knowledge of biotic constraints and molecular characterization of ipomoviruses associated with cassava brown streak disease in Rwanda*. The 25th National Symposium for Applied Biological Sciences (NSABS) conference at Université de Liège, Gembloux Agro-Bio Tech, Espace Senghor, on 31st January 2020.