**BALTAZAR DO AZARENTO ISABEL CHIPIRINGO**

# BEGOMOVIRUSES ASSOCIATED WITH WEEDS AND LEGUME CROPS IN MOZAMBIQUE: EMERGENCE, VARIABILITY AND RECOMBINATION-DRIVEN EVOLUTION

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de *Doctor Scientiae*.

Orientador: Francisco Murilo Zerbini Júnior
Coorientadora: Angélica Maria Nogueira

**VIÇOSA - MINAS GERAIS**
**2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa**

T
C541b
2021

Chipiringo, Baltazar do Azarento Isabel, 1981-
Begomoviruses associated with weeds and legume crops in Mozambique : emergence, variability and recombination-driven evolution / Baltazar do Azarento Isabel Chipiringo. – Viçosa, MG, 2021.
71 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Inclui anexos.
Orientador: Francisco Murilo Zerbini Júnior.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Begomovirus. 2. Recombinação (Genética). 3. Leguminosa - Doenças e pragas. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22. ed. 632.8

Bibliotecário(a) responsável: Alice Regina Pinto Pires CRB6 2523

**BALTAZAR DO AZARENTO ISABEL CHIPIRINGO**

**BEGOMOVIRUSES ASSOCIATED WITH WEEDS AND LEGUME CROPS IN MOZAMBIQUE: EMERGENCE, VARIABILITY AND RECOMBINATION-DRIVEN EVOLUTION**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de *Doctor Scientiae*.

APROVADA: 29 de junho de 2021.

Assentimento:

Baltazar do Azarento Isabel Chipiringo

Baltazar do Azarento Isabel Chipiringo
Autor

Francisco Murilo Zerbini Júnior
Orientador

A minha mãe Isabel Basílio (em memória)
A minha esposa Henriqueta da Teresa Saide Chipiringo,
Aos meus filhos: Isabel da Teresa Baltazar Chipiringo,
Baltazar Júnior Abasse Chipiringo,
Rachel da Ediluvia Baltazar Chipiringo e
Josué Saide Baltazar Chipiringo.

Pela paciência, fé, esperança, companheirismo e amor incondicional em todos os momentos da minha luta, dedico.

## AGRADECIMENTOS

A Deus toda honra e toda glória.

Aos meus pais (padrinhos) Américo Hilário Cantelo e Paulina Casimiro Mpagua por me aceitar como filho, pelo amor sincero, carinho e segurança.

Aos meus sogros Saide Abasse Saide e Teresa Abílio Mussa pelo carinho e força.

Aos meus irmãos Wiler Basílio Santana e Nilza Helena da Conceição Basílio por serem a minha plateia torcendo sempre pelo sucesso.

Ao professor Francisco Murilo Zerbini pelo profissionalismo, pela orientação, ensinamento e por acreditar em mim. Tenho na sua pessoa um espelho para a vida profissional e um exemplo na vida social e familiar.

A professora Poliane pelo zelo, atenção e por liberar o seu laboratório quando nos foi preciso.

Ao professor Pedro Chume, pelo incentivo, profissionalismo e confiança.

A Patrícia pelo profissionalismo, amizade incondicional, você é uma verdadeira irmã.

Aos colegas do laboratório Ayane, Ana Reis, Anelise Orílio, César, João Paulo, Lucas, Nathália, Osvaldo, Rena, Roberta, Ruiter e Tarsiane pelo verdadeiro espirito de guerreiros académicos.

Alberto Macamo pela força e amizade.

A todo corpo docente do Departamento de Fitopatologia pelos ensinamentos, profissionalismo e por serem provedores da ciência internacional.

A Universidade Federal de Viçosa (UFV) e do Departamento de Fitopatologia pelos ensinamentos e acolhimento.

Ao Ministério da Ciência e Tecnologia, Ensino Superior e Técnico Profissional de Moçambique, pela atribuição da bolsa de estudos.

A Universidade Zambeze (Unizambeze) pela dispensa nesse período de formação.

O presenta trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Código de Financiamento 001.

**O meu Muito Obrigado, a todos!**

**ABSTRACT**

CHIPIRINGO, Baltazar do Azarento Isabel, D.Sc., Universidade Federal de Viçosa, June, 2021. **Begomoviruses associated with weeds and legume crops in Mozambique: Emergence, variability and recombination-driven evolution.** Advisor: Francisco Murilo Zerbini Júnior. Co-advisor: Angélica Maria Nogueira

Agriculture in Mozambique is moving towards "modernization", with increased areas of monoculture, extended growing seasons, irrigation and mechanization replacing subsistence farming. However, such changes in agricultural practices can promote pathogen emergence. Understanding the emergence and evolution of plant viruses provides beneficial information on management practices. This work aimed at the detection, identification and molecular characterisation of begomoviruses associated with legumes and weeds in Mozambique. In the first study, a non-cultivated plant identified as *Pyrenecantha* sp. (Icacinaceae) with yellow mosaic symptoms was collected in the district of Malema, Nampula province. The viral genome was amplified using rolling-circle amplification, cloned and sequenced. The DNA-A has the highest identity (78%) with tomato leaf curl Namakele virus, while the DNA-B sequence has the highest identity (70%) with Deinbolia mosaic virus. The two components have a genomic organization typical of Old World, bipartite begomoviruses. Alignment of their common regions (CR) indicated a 35-nt insertion in the DNA-A CR. The identity between the CRs is 83%, increasing to 96% when the 35-nt insertion is removed from the alignment. The CRs have identical iterons, and the amino acid sequence of the Rep protein contains the iteron-related domain predicted to recognize these iterons. These results indicate that the cloned DNA-A and DNA-B are cognate components of the same virus, and the 35-nt insertion suggests a recombination event in the DNA-A CR. The name Pyrenecantha yellow mosaic virus is proposed for the new virus. For the second study, symptomatic samples of bean, cowpea and soybean plants were collected in the central and northern regions of Mozambique. Seven DNA-A and 11 DNA-B components were cloned and sequenced. The DNA-A sequences are 96% identical, indicating that they represent the same virus. The highest identity with other begomoviruses is 85% with cowpea golden mosaic virus. The DNA-B sequences are 93% identical amongst each other and have a maximum of 72% identity with cotton yellow mosaic virus. Alignment of the CRs indicates that the DNA-A and DNA-B components are cognate components of the same virus. The name Mozambique legume mosaic virus is proposed. Three recombination events were detected in the DNA-B. The DNA-B is more variable than the DNA-A, and the DNA-B nucleotide diversity is higher in the *MP* gene and in the short intergenic region. MLMV is an emerging begomovirus in legume crops in Mozambique, and its adaptation to new hosts appears to be driven by recombination in the DNA-B.

**Keywords:** Pyrenacanthae. PyYMV. 35-nt. MLMV. Mozlegume. Bean. Soybean. Cowpea.

# RESUMO

CHIPIRINGO, Baltazar do Azarento Isabel, D.Sc., Universidade Federal de Viçosa, junho de 2021. **Begomovírus associados a plantas daninhas e cultivos de leguminosas em Moçambique: Emergência, variabilidade e evolução por meio de recombinação**. Orientador: Francisco Murilo Zerbini Júnior. Coorientadora: Angélica Maria Nogueira

A agricultura em Moçambique caminha para a "modernização", com aumento de áreas de monocultura, períodos de cultivo prolongados, irrigação e mecanização substituindo a agricultura de subsistência. No entanto, tais mudanças nas práticas agrícolas podem promover a emergência de patógenos. Compreender a emergência e evolução dos vírus de plantas proporciona informações importantes sobre o manejo dessas doenças. Este trabalho objetivou a detecção, identificação e caracterização molecular de begomovírus associados a leguminosas e plantas não-cultivadas em Moçambique. No primeiro estudo, uma amostra de *Pyrenacantha* sp. (Icacinaceae) com sintomas de mosaico amarelo foi coletada no distrito de Malema, província de Nampula. O genoma viral foi amplificado, clonado e sequenciado. A sequência de nucleotídeos do DNA-A tem maior identidade (78%) com o tomato leaf curl Namakele virus. O DNA-B tem maior identidade (70%) com Deinbolia mosaic virus. Os dois componentes têm organização genômica típica dos begomovírus bissegmentados do Velho Mundo. O alinhamento das regiões comuns (CR) mostrou uma inserção de 35 nt na CR do DNA-A. A identidade entre as CRs é de 83%, aumentando para 96% quando a inserção de 35 nt é desconsiderada. As CRs têm iterons idênticos e a sequência de aminoácidos da proteína Rep contém o IRD que reconhece esses iterons. Esses resultados indicam que o DNA-A e o DNA-B são componentes cognatos do mesmo vírus, e a inserção de 35 nt sugere um evento de recombinação na CR do DNA-A. O nome Pyrenecantha yellow mosaic virus é proposto para o novo vírus. Para o segundo estudo, amostras sintomáticas de plantas de feijão, soja e caupi foram coletadas nas regiões Centro e Norte de Moçambique. Sete DNAs-A e 11 DNAs-B foram clonados e sequenciados. As sequências de DNA-A são 96% idênticas entre si, indicando que representam o mesmo vírus. A maior identidade com outros begomovírus é de 85% com o cowpea golden mosaic virus. Os DNAs-B possuem 93% de identidade entre si e máximo de 72% com cotton yellow mosaic virus. O alinhamento das CRs indica que os DNAs-A e DNAs-B são componentes cognatos. O nome Mozambique legume mosaic virus é proposto. Três eventos de recombinação foram detectados no DNA-B. O DNA-B é mais variável do que o DNA-A, e a diversidade de nucleotídeos do DNA-B é maior no gene *MP* e na região intergênica curta. O MLMV é um begomovírus emergente em leguminosas em Moçambique, e sua adaptação aos novos hospedeiros parece ser impulsionada por recombinação no DNA-B.

**Palavras-chave:** Pyrenacanthae. PyYMV. 35-nt. MLMV. Mozlegume. Feijão. Soja. Caupi.

## SUMMARY

# GENERAL INTRODUCTION

Mozambique is a country located on the east coast of southern Africa, covering an area of 801,590 km$^2$. With a population of 27,909,798 inhabitants, the backbone of agriculture is subsistence farming (Plumb-Dhindsa & Mondjane, 1984; Maunze *et al.*, 2019). The country has favourable agroclimatic conditions for agriculture and has food and cash crop production as priorities.

Agriculture is predominantly rain fed and constitutes the economic activity that occupies most of the country's population. Agricultural growth is fundamental and can be an essential source to promote economic growth. However, policies need to be adjusted to allow farmers to reach higher levels of production and trade surpluses efficiently. The fact of having a long coastline on the Indian Ocean can provide added value for exporting agricultural products as well as natural resources.

Increased production is necessary to meet food security and demand, as only a few regions of the country have favourable agro-ecological conditions for production. The factors detrimental to productivity, which include plant diseases, need to be increasingly studied and controlled to avoid losses (Giglioti *et al.*, 2015). Virus diseases are one of the main biotic causes of crop losses in tropical regions worldwide, and Mozambique is no exception.

The geminiviruses (family *Geminiviridae*) are plant viruses with circular, single-stranded DNA genomes encapsidated by a single structural protein in twinned icosahedral particles. They constitute a group of plant viruses of great relevance to agriculture due to the severity of diseases caused by them worldwide (Shepherd *et al.*, 2010; Rocha *et al.*, 2013; Sattar *et al.*, 2013; Mabvakure *et al.*, 2016; Rey & Vanderschuren, 2017). The many factors contributing to the emergence of plant virus diseases includes genetic changes in the viral genome and changes in agricultural practices which cause an increase in vector populations (Navas-Castillo *et al.*, 2011; Rojas *et al.*, 2018).

Geminiviruses exhibit considerable diversity in terms of genomic structure, host range and insect vector. Based on their genomic arrangement, type of insect vector, phylogenetic relationships and host range, geminiviruses are currently classified into 14 genera: *Begomovirus*, *Becurtovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Mastrevirus*, *Maldovirus*, *Mulcrilevirus*, *Opunvirus*, *Topocuvirus*, *Turncurtovirus* and *Topilevirus* (Zerbini *et al.*, 2017). Of all genera, *Begomovirus* constitutes the largest in number of species (currently, 445 species).

The begomoviruses can be divided into two major groups based on phylogenetic relationships and genome characteristics: Old World (OW; Europe, Africa and Asia) and New World (NW; the Americas) (Rybicki, 1994). Old World begomoviruses can be mono- or bipartite, and are often associated with two types of satellite DNAs, alpha- and betasatellites (Zhou, 2013; Brown *et al.*, 2015). In contrast, NW begomoviruses have two genomic components (bipartite) with only a few exceptions, and can, very rarely, be associated with alpha- and deltasatellites (Brown *et al.*, 2015; Ferro *et al.*, 2017; Ferro *et al.*, 2021).

Begomoviruses are transmitted in nature by whiteflies of the *Bemisia tabaci* cryptic species complex (Hemiptera: Aleyrodidae) to dicot plants (Brown *et al.*, 2015). These viruses are widespread in all tropical and subtropical regions of the world, and cause severe diseases in a number of economically relevant crops (Navas-Castillo *et al.*, 2011; Rojas *et al.*, 2018).

Begomovirus populations have a high degree of genetic variability. The frequent occurrence of recombination and pseudo-recombination (in viruses with bipartite genomes) and the high rate of nucleotide substitution of the viral genome are the major factors contributing to their high degree of genetic variability (Pita *et al.*, 2001; Duffy & Holmes, 2009; Lima *et al.*, 2013; Silva *et al.*, 2014; Lima *et al.*, 2017). Variation provides the basis for trait evolution through natural selection and results in plant viruses adapting to new hosts, new vectors and overcoming host resistance (Stobbe & Roossinck, 2016).

Mutation is the main mechanism for generating variability in begomoviruses (Lima *et al.*, 2017). Geminiviruses exhibit high levels of molecular variability within the host (Ge *et al.*, 2007; Van Der Walt *et al.*, 2008; Pinto *et al.*, 2021) and evolve at rates on the order of $10^{-3}$-$10^{-4}$ substitutions-site-year (Duffy & Holmes, 2009). Recombination events have been directly associated with the emergence of new begomoviruses and epidemics in crop plants. The most prominent examples include the devastating epidemics of cassava mosaic disease caused by a recombinant EACMV in Uganda and neighbouring countries (Zhou *et al.*, 1997; Pita *et al.*, 2001), the TYLCV epidemics caused by different recombinants in the western Mediterranean basin (Monci *et al.*, 2002; García-Andrés *et al.*, 2007; Belabess *et al.*, 2015), and the cotton leaf curl virus (CLCuV) epidemics in Pakistan, caused by a complex which includes a recombinant begomovirus associated with multiple betasatellites (Zhou *et al.*, 1997; Idris & Brown, 2002).

Agriculture in Mozambique tends to move towards modernization, with increased areas of monoculture, extended growing seasons, irrigation and mechanization replacing subsistence farming. However, such changes in agricultural practices can also promote pathogen introduction or emergence (Stobbe & Roossinck, 2016).

Understanding the emergence and evolution of plant viruses can provide beneficial information on management practices of these diseases in crops (Roossinck, 2015). Proper disease management requires information on types, distribution, incidence and genetic variation of the causal viruses. Such information is still limited in developing countries like Mozambique. Africa in general and Mozambique in particular is largely under-represented in plant virus identification (Mollel *et al.*, 2020).

Viruses have been co-evolving with indigenous plants growing in undisturbed plant communities in different regions of the world long before plants were domesticated (Vincent *et* *al.*, 2014). Wild plants may be the original hosts from which viruses spillover into crops as a first step towards emergence (García-Arenal & Zerbini, 2019). The simplification of ecosystems due to rapidly increasing human activity and the intensification of practices to increase food production to combat food insecurity create new scenarios that favour encounters between viruses and crop plants, which may lead to devastating epidemics (Roossinck & Garcia-Arenal, 2015; García-Arenal & Zerbini, 2019). In this context, cataloguing viruses infecting wild hosts is of obvious importance.

Legumes have great social and economic importance in Mozambique. Common bean (*Phaseolus vulgaris*) and cowpea (*Vigna unguiculata*) constitute important sources of protein in the diet of the population, and of family income for smallholder farmers. Both crops are produced mainly in intercropping with maize, sorghum and other dryland graminaceous crops, or in monocropping in small areas.

Common bean in the Americas is host to a large number of begomoviruses, including bean calico mosaic virus (BCMoV), bean dwarf mosaic virus (BDMV), bean golden mosaic virus (BGMV), bean golden yellow mosaic virus (BGYMV), macroptilium yellow spot virus (MaYSV), squash leaf curl virus (SLCV) and sida micrantha mosaic virus (SiGMV) (Lazarowitz, 1992; Frischmuth *et al.*, 1997; Garrido-Ramirez *et al.*, 2000; Brown *et al.*, 2015; Rojas *et al.*, 2018). Cowpea crops are affected by cowpea golden mosaic virus in Nigeria and by a complex of begomoviruses in South Asia (Qazi *et al.*, 2007).

Soybean (*Glycine max*) has enormous economic importance in generating income for farmers and raw material for feed production in the livestock industries in Mozambique. The soybean crop in Mozambique is relatively new and has experienced rapid growth, with much of the production dominated by small farmers. Besides being an excellent host for whiteflies, soybean is permissive to infection by several begomoviruses such as BGMV (Gilbertson *et al.*, 1991), mungbean yellow mosaic virus (MYMV), mungbean yellow mosaic India virus

(MYMIV), soybean blistering mosaic vrius (SbBMV), soybean mild mottle virus (SbMMV) and soybean chlorotic blotch virus (SbCBV) (Qazi *et al.*, 2007; Alabi *et al.*, 2010; Rodríguez-Pardina *et al.*, 2011).

Previous work carried out in our laboratory revealed the existence of three cryptic species of *B. tabaci* in Mozambique (Sande, 2020), highlighting the potential for crop losses due to the emergence of begomoviruses. The extent of productivity losses by begomoviruses will depend on the population level of the insect vector, the susceptibility of cultivars, climatic conditions and agronomic management of the crop (Garrido-Ramirez *et al.*, 2000; Rojas *et al.*, 2018). Research activities that provide a comprehensive understanding of the threat posed by viruses are necessary and of interest to the country.

The objectives of this work were the detection, identification and molecular characterization of begomoviruses infecting non-cultivated and cultivated plants in Mozambique. The work is presented in two chapters: (*i*) the molecular characterization of a new bipartite, recombinant begomovirus infecting the wild plant *Pyrenacantha* sp., and (*ii)* the emergence of a bipartite begomovirus and its recombination-driven adaptation to legume crops in Mozambique.

## Literature Cited

Alabi, O.J.; Kumar, P.L.; Mgbechi-Ezeri, J.U.; Naidu, R.A. Two new 'legumoviruses' (genus *Begomovirus*) naturally infecting soybean in Nigeria. **Archives of Virology,** v. 155, p. 643-656, 2010.

Belabess, Z.; Dallot, S.; El-Montaser, S.; Granier, M.; Majde, M.; Tahiri, A.; Blenzar, A.; Urbino, C.; Peterschmitt, M. Monitoring the dynamics of emergence of a non-canonical recombinant of tomato yellow leaf curl virus and displacement of its parental viruses in tomato. **Virology,** v. 486, p. 291-306, 2015.

Brown, J.K.; Zerbini, F.M.; Navas-Castillo, J.; Moriones, E.; Ramos-Sobrinho, R.; Silva, J.C.; Fiallo-Olive, E.; Briddon, R.W.; Hernandez-Zepeda, C.; Idris, A.; Malathi, V.G.; Martin, D.P.; Rivera-Bustamante, R.; Ueda, S.; Varsani, A. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. **Archives of Virology,** v. 160, p. 1593-619, 2015.

Duffy, S.; Holmes, E.C. Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. **Journal of General Virology,** v. 90, p. 1539-1547, 2009.

Ferro, C.G.; Silva, J.P.; Xavier, C.a.D.; Godinho, M.T.; Lima, A.T.M.; Mar, T.B.; Lau, D.; Zerbini, F.M. The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from *Sida* spp. and *Leonurus sibiricus*, plus two New World alphasatellites. **Annals of Applied Biology,** v. 170, p. 204-218, 2017.

Ferro, C.G.; Zerbini, F.M.; Navas-Castillo, J.; Fiallo-Olive, E. Revealing the complexity of sweepovirus-deltasatellite-plant host interactions: expanded natural and experimental helper virus range and effect dependence on virus-host combination. **Microorganisms,** v. 9, 2021.

Frischmuth, T.; Engel, M.; Lauster, S.; Jeske, H. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, *Sida*-infecting bipartite geminiviruses in Central America. **Journal of General Virology,** v. 78, p. 2675-2682, 1997.

García-Andrés, S.; Accotto, G.P.; Navas-Castillo, J.; Moriones, E. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. **Virology,** v. 359, p. 302-312, 2007.

García-Arenal, F.; Zerbini, F.M. Life on the edge: geminiviruses at the interface between crops and wild plant hosts. **Annual Review of Virology,** v. 6, p. 411-433, 2019.

Garrido-Ramirez, E.R.; Sudarshana, M.; Gilbertson, R.L. Bean golden yellow mosaic viru*s* from Chiapas, Mexico: Characterization, pseudorecombination with other bean-infecting geminiviruses and germ plasm screening. **Phytopathology,** v. 90, p. 1224-1232, 2000.

Ge, L.M.; Zhang, J.T.; Zhou, X.P.; Li, H.Y. Genetic structure and population variability of tomato yellow leaf curl China virus. **Journal of Virology,** v. 81, p. 5902-5907, 2007.

Giglioti, E.A.; Sumida, C.H.; Canteri, M.G. Disease phenomics. In: Fritsche-Neto, R. e Borém, A. (Ed.). **Phenomics: How Next-Generation Phenotyping is Revolutionizing Plant Breeding**. Cham, Switzerland: Springer International Publishing, 2015. p.101-123. ISBN 978-3-319-13677-6.

Gilbertson, R.L.; Faria, J.C.; Hanson, S.F.; Morales, F.J.; Ahlquist, P.G.; Maxwell, D.P.; Russell, D.R. Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. **Phytopathology,** v. 81, p. 980-985, 1991.

Idris, A.M.; Brown, J.K. Molecular analysis of cotton leaf curl virus-Sudan reveals an evolutionary history of recombination. **Virus Genes,** v. 24, p. 249-256, 2002.

Lazarowitz, S.G. Geminiviruses: Genome structure and gene function. **Critical Reviews in Plant Sciences,** v. 11, p. 327-349, 1992.

Lima, A.T.M.; Silva, J.C.F.; Silva, F.N.; Castillo-Urquiza, G.P.; Silva, F.F.; Seah, Y.M.; Mizubuti, E.S.G.; Duffy, S.; Zerbini, F.M. The diversification of begomovirus populations is predominantly driven by mutational dynamics. **Virus Evolution,** v. 3, p. vex005, 2017.

Lima, A.T.M.; Sobrinho, R.R.; Gonzalez-Aguilera, J.; Rocha, C.S.; Silva, S.J.C.; Xavier, C.a.D.; Silva, F.N.; Duffy, S.; Zerbini, F.M. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. **Journal of General Virology,** v. 94, p. 418-431, 2013.

Mabvakure, B.; Martin, D.P.; Kraberger, S.; Cloete, L.; Van Brunschot, S.; Geering, A.D.; Thomas, J.E.; Bananej, K.; Lett, J.M.; Lefeuvre, P.; Varsani, A.; Harkins, G.W. Ongoing geographical spread of tomato yellow leaf curl virus. **Virology,** v. 498, p. 257-64, 2016.

Maunze, X.H.; Dade, A.; Zacarias, M.D.F.; Cubula, B.; Alfeu, M.; Mangue, J.; Mouzinho, R.; Cassimo, M.N.; Zunguze, C.; Zavale, O.; Muchanga, I.; Aramuge, Z.; Julaya, F.; Tomo, J.; Nguenha, F.; Mula, C. **IV Recenseamento Geral da População e Habitação 2017**. Instituto Nacional de Estatística. Maputo, Moçambique. 2019

Mollel, H.G.; Ndunguru, J.; Sseruwagi, P.; Alicai, T.; Colvin, J.; Navas-Castillo, J.; Fiallo-Olivé, E. African basil (*Ocimum gratissimum*) is a reservoir of divergent begomoviruses in Uganda. **Plant Disease,** v. 104, p. 853-859, 2020.

Monci, F.; Sanchez-Campos, S.; Navas-Castillo, J.; Moriones, E. A natural recombinant between the geminiviruses tomato yellow leaf curl Sardinia virus and tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. **Virology,** v. 303, p. 317-26, 2002.

Navas-Castillo, J.; Fiallo-Olivé, E.; Sánchez-Campos, S. Emerging virus diseases transmitted by whiteflies. **Annual Review of Phytopathology** v. 49, p. 219-248, 2011.

Pinto, V.B.; Quadros, A.F.F.; Godinho, M.T.; Silva, J.C.; Alfenas-Zerbini, P.; Zerbini, F.M. Intra-host evolution of the ssDNA virus tomato severe rugose virus (ToSRV). **Virus Research,** v. 292, p. 198234, 2021.

Pita, J.S.; Fondong, V.N.; Sangare, A.; Otim-Nape, G.W.; Ogwal, S.; Fauquet, C.M. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. **Journal of General Virology,** v. 82, p. 655-65, 2001.

Plumb-Dhindsa, P.; Mondjane, A.M. Index of plant diseases and associated organisms of Mozambique. **Tropical Pest Management,** v. 30, p. 407-429, 1984.

Qazi, J.; Ilyas, M.; Mansoor, S.; Briddon, R.W. Legume yellow mosaic viruses: genetically isolated begomoviruses. **Molecular Plant Pathology,** v. 8, p. 343-348, 2007.

Rey, C.; Vanderschuren, H. Cassava mosaic and brown streak diseases: current perspectives and beyond. **Annual Review of Virology,** v. 4, p. 429-452, 2017.

Rocha, C.S.; Castillo-Urquiza, G.P.; Lima, A.T.M.; Silva, F.N.; Xavier, C.a.D.; Hora-Junior, B.T.; Beserra-Junior, J.E.A.; Malta, A.W.O.; Martin, D.P.; Varsani, A.; Alfenas-Zerbini, P.; Mizubuti, E.S.G.; Zerbini, F.M. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. **Journal of Virology,** v. 87, p. 5784-5799, 2013.

Rodríguez-Pardina, P.E.; Hanada, K.; Laguna, I.G.; Zerbini, F.M.; Ducasse, D.A. Molecular characterisation and relative incidence of bean- and soybean-infecting begomoviruses in northwestern Argentina. **Annals of Applied Biology,** v. 158, p. 69-78, 2011.

Rojas, M.R.; Macedo, M.A.; Maliano, M.R.; Soto-Aguilar, M.; Souza, J.O.; Briddon, R.W.; Kenyon, L.A.; Rivera-Bustamante, R.F.; Zerbini, F.M.; Adkins, S.; Legg, J.P.; Kvarnheden, A.; Wintermantel, W.M.; Sudarshana, M.R.; Peterschmitt, M.; Lapidot, M.; Martin, D.P.; Moriones, E.; Inoue-Nagata, A.K.; Gilbertson, R.L. World management of geminiviruses. **Annual Review of Phytopathology,** v. 56, p. 637-677, 2018.

Roossinck, M.J. A new look at plant viruses and their potential beneficial roles in crops. **Molecular Plant Pathology,** v. 16, p. 331-3, 2015.

Roossinck, M.J.; Garcia-Arenal, F. Ecosystem simplification, biodiversity loss and plant virus emergence. **Current Opinion in Virology,** v. 10, p. 56-62, 2015.

Rybicki, E.P. A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. **Archives of Virology,** v. 139, p. 49-77, 1994.

Sande, O.F.L. **Begomoviruses and their vectors in Mozambique: diversity, evolution and population dynamics**. 2020. (D.Sc. Dissertation). Dep. of Plant Pathology, Universidade Federal de Viçosa, Viçosa, MG.

Sattar, M.N.; Kvarnheden, A.; Saeed, M.; Briddon, R.W. Cotton leaf curl disease - an emerging threat to cotton production worldwide. **Journal of General Virology,** v. 94, p. 695-710, 2013.

Shepherd, D.N.; Martin, D.P.; Van Der Walt, E.; Dent, K.; Varsani, A.; Rybicki, E.P. Maize streak virus: an old and complex 'emerging' pathogen. **Molecular Plant Pathology,** v. 11, p. 1-12, 2010.

Silva, F.N.; Lima, A.T.M.; Rocha, C.S.; Castillo-Urquiza, G.P.; Alves-Júnior, M.; Zerbini, F.M. Recombination and pseudorecombination driving the evolution of the begomoviruses tomato severe rugose virus (ToSRV) and tomato rugose mosaic virus (ToRMV): two recombinant DNA-A components sharing the same DNA-B. **Virology Journal,** v. 11, p. 66, 2014.

Stobbe, A.; Roossinck, M.J. Plant Virus Diversity and Evolution. In: Wang, A. e Zhou, X. (Ed.). **Current Research Topics in Plant Virology**. Cham: Springer International Publishing, 2016. p.197-215. ISBN 978-3-319-32919-2.

Van Der Walt, E.; Martin, D.P.; Varsani, A.; Polston, J.E.; Rybicki, E.P. Experimental observations of rapid maize streak virus evolution reveal a strand-specific nucleotide substitution bias. **Virology Journal,** v. 5, p. 104, 2008.

Vincent, S.J.; Coutts, B.A.; Jones, R.a.C. Effects of introduced and indigenous viruses on native plants: Exploring their disease-causing potential at the agro-ecological interface. **PLoS ONE,** v. 9, p. e91224, 2014.

Zerbini, F.M.; Briddon, R.W.; Idris, A.; Martin, D.P.; Moriones, E.; Navas-Castillo, J.; Rivera-Bustamante, R.; Varsani, A.; ICTV Consortium. ICTV Virus Taxonomy Profile: *Geminiviridae*. **Journal of General Virology,** v. 98, p. 131-133, 2017.

Zhou, X. Advances in understanding begomovirus satellites. **Annual Review of Phytopathology,** v. 51, p. 357-381, 2013.

Zhou, X.; Liu, Y.; Calvert, L.; Munoz, C.; Otim-Nape, G.W.; Robinson, D.J.; Harrison, B.D. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. **Journal of General Virology,** v. 78, p. 2101-2111, 1997.

# CHAPTER 1

## A new bipartite begomovirus naturally infecting *Pyrenacantha* sp. in Mozambique

Chipiringo, B.A.I., Silva, J.P., Cascardo, R.S., Sande, O.F.L., Zerbini, F.M. (2021) A new bipartite begomovirus naturally infecting *Pyrenacantha* sp. in Mozambique. **Archives of Virology**, *in preparation*.

# **A new bipartite begomovirus naturally infecting *Pyrenacantha* sp. in Mozambique**

Baltazar A. I. Chipiringo[1,2], João P. Silva[1], Renan S. Cascardo[1], Osvaldo F.L. Sande[1,3], F. Murilo Zerbini[1]*

[1]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil

[2]Universidade Zambeze (Unizambeze), Faculdade de Ciências Agrárias, P.O. Box 213, Ulóngué, Tete 2306, Mozambique

[3]Divisão de Agricultura, Instituto Superior Politécnico de Manica (DivAG-ISPM), Manica, Mozambique

*Corresponding author:
Phone: (+55-31) 3612-2423
E-mail: zerbini@ufv.br

**Abstract**

Non-cultivated plants can act as alternative hosts and reservoirs for viruses important to agriculture. A *Pyrenacantha* sp. plant with yellow mosaic symptoms typical of begomovirus infection was collected in a maize production field in Mozambique. Total DNA was extracted and the presence of a begomovirus was confirmed by polymerase chain reaction with degenerate primers. The viral genome was cloned and sequenced. The complete genomic components (DNA-A and DNA-B) were determined to be 2766 and 2726 nucleotides (nt) in length, respectively. The DNA-A showed maximum nt sequence identity (78.5%) with tomato leaf curl Namakele virus, while the DNA-B showed higher nt identity (70.2%) with Deinbolia mosaic virus. Based on the current ICTV species demarcation criteria for the genus *Begomovirus*, the viral isolate, named Pyrenacantha yellow mosaic virus (PyYMV), is a member of a new species, for which the name *Begomovirus pyrenacanthae* is proposed. The DNA-A and DNA-B have a genomic organization typical of Old World, bipartite begomoviruses. Alignment of their common regions (CR) indicated a 35-nt insertion in the DNA-A CR. The nt sequence identity between the CRs is only 83% but increases to 96% when the 35-nt insertion is removed from the alignment. The CRs have identical iterons, and the amino acid sequence of the Rep protein contains the IRDs predicted to recognize these iterons. Together, these results indicate that the cloned DNA-A and DNA-B are cognate components of the same virus, and the 35-nt insertion suggests a recombination event in the DNA-A CR. Phylogenetic analysis showed congruence in phylogenetic positioning between DNA-A and DNA-B. This is the first report of a begomovirus naturally infecting *Pyrenacantha sp*.

**Keywords:** Pyrenacanthae. PyYMV. 35-nt.

Before the onset of agriculture, viruses were evolving with indigenous plants in undisturbed plant communities [1]. In natural ecosystems viruses may be associated with a wide variety of non-cultivated plant species, which act as reservoir and alternative hosts for the survival and spread of the virus that can eventually spillover to crop plants, as a first step towards emergence [2]. Non-cultivated plants have been reported as hosts of begomoviruses in several African countries [3-7].

The genus *Begomovirus* (family *Geminiviridae*) comprises one of the most important groups of emerging plant viruses. Geminiviruses exhibit considerable diversity in terms of genomic structure, host range and type of insect vector [8]. Begomoviruses, transmitted by whiteflies of the *Bemisia tabaci* cryptic species complex to dicot plants [9], constitute the largest genera of all the virosphere, with many ecologically successful begomoviruses infecting non-cultivated and cultivated plants in all tropical and sub-tropical regions of the world [8, 10].

Begomoviruses found in the "Old World" (OW; Africa, Europe, Asia and Oceania) can have mono- or bipartite genomes and are often associated with satellite DNAs [9, 11]. The genomic DNA of monopartite begomoviruses and the DNA-A component of OW bipartite begomoviruses have analogous genomic organization [12, 13] with six genes involved in replication (*Rep* and *REn*), transcription (*TrAP*), suppression of host defenses (*TrAP* and *AC4*), movement (*V2*) and encapsidation of viral progeny (*CP*) [14, 15]. The DNA-B of bipartite begomoviruses contains two genes that encode proteins required for virus movement across the nuclear envelope (*NSP*) and cell-to-cell movement across the plasmodesmata (*MP*) [16]. Transcription of the viral genome in begomoviruses is bidirectional [17, 18] and the viral and complementary transcription units of the genome are separated by an approximately 200-nt intergenic region (IR), known as the common region (CR) in bipartite viruses.

Sequence identity between the cognate DNA-A and DNA-B CRs is usually very high (>94%). The CR contains the nonanucleotide 5'-TAATATTAC-3', conserved in all begomoviruses and which constitutes the origin of replication, and specific binding sites for the Rep protein known as iterons [9, 19, 20].

Agriculture in Mozambique is predominantly rain fed and is the economic activity that occupies much of the population in the country. Agricultural growth is critical and can be a key source to promote economic growth. However, the intensification of practices to increase agricultural production and combat food insecurity may impact the emergence of plant virus diseases, which constitute one of the main biotic causes of yield losses in agriculture [21].

Despite the importance and damage caused by begomoviruses worldwide, there is an almost complete absence of information on begomoviruses in Mozambique. A recent study analyzed a population of a begomovirus in tomato crops [22], but there are no studies on begomovirus diversity on non-cultivated plants. This lack of knowledge on the occurrence and dynamics of viruses in wild plant communities limits the understanding of the emergence of plant viruses in crops [23].

*Pyrenacantha* sp. (family Icacinaceae) is a dioecious climbing herb that originates from a subterranean perennial tuberous rootstock. Endemic to East Africa (Tanzania, Malawi, Zambia, Zimbabwe and Mozambique), the type specimen was found in Cabora Bassa, Mozambique [24].

In this study, we identified and molecularly characterised a new bipartite begomoviruses that infects *Pyrenacantha* sp., named Pyrenacantha yellow mosaic virus. The species name *Begomovirus pyrenachantae* is proposed. To the best of our knowledge, this is the first report of a begomovirus infecting *Pyrenacantha* sp. not only in Africa but in the world.

A *Pyrenacantha* sp. plant with symptoms of yellow mosaic (Figure 1A) was collected in a maize field adjacent to a soybean field in the district of Malema (14°57.690'S, 37°23.428'E 628 masl), Nampula Province, Mozambique, in March 2019. Total DNA was extracted from dried leaf discs following the protocol described by Doyle and Doyle [25] and preserved at -20°C. To confirm the presence of a begomovirus in the sample, PCR was performed using universal primers for the genus (PAL1v1978/PAR1c496) [26], designed to amplify a 1.4 kbp fragment from the DNA-A component. Next, the DNA was subjected to rolling circle amplification (RCA) [27]. Concatamers were cleaved with the enzyme MspI and the band pattern observed in a 1.2% agarose gel. After confirmation of a cleavage pattern suggestive of a single bipartite begomovirus (sum of fragment sizes of approximately 5.4 kbp), the RCA product was cleaved with HindIII or SacI to obtain monomeric genomic fragments for cloning. Fragments of about 2.7 kbp, corresponding to one genomic copy of each linearized component, were ligated to the pBluescriptKS+ vector (Stratagene) previously cleaved with the same enzymes and dephosphorylated. The recombinant plasmids were used for transformation in *Escherichia coli* DH5α [28] and selected clones were sequenced commercially (Macrogen, South Korea). The full-length genome was assembled using SeqAssm v.1.0 (www.sequentix.de), starting from the nonanucleotide cleavage site (5'-TAATATT//AC-3').

The sequence was initially analysed with the BLAST*n* algorithm [29] to determine the viral species with which it shared the highest identity. Closely related begomovirus sequences were used to determine the taxonomical position of the new isolate using Species Demarcation Tool v. 1.2 [30], according to the criteria for species demarcation established by the ICTV *Geminiviridae* and *Tolecusatellitidae* study group (<91% nt sequence identity for the DNA-A) [9]. The DNA-B was identified as a cognate component based on sequence identity in the common region. The search for ORFs was performed with the Geminivirus Data Warehouse platform [31] and the ORF finder program (www.ncbi.nlm.nih.gov/projects/gorf/). A multiple sequence alignment was obtained using the MUSCLE algorithm implemented in the MEGA X program [32], and phylogenetic trees were built through Bayesian inference with the program MrBayes v. 3.2.7a [33] using the General Time Reversible nucleotide substitution model considering the gamma distribution with invariant sites (GTR+G+I), selected with the

MrModeltest v. 2.3 program [34] based on the Akaike Information Criterion. The trees were visualized and edited with the FigTree program (tree.bio.ed.ac.uk/software/figgtree/).

We cloned and sequenced three begomovirus components (two DNA-As and one DNA-B). Comparison of the sequences obtained in this study with sequences deposited in GenBank using the BLAST*n* algorithm and the SDT program revealed a maximum nucleotide (nt) sequence identity of 78.5% with the DNA-A of tomato leaf curl Namakele virus (TLCNaV, GenBank accession number AM701764) for clones MZ-Mal170.2-19 and MZ-Mal170.11-19, and 70.2% nt sequence identity with the DNA-B of Deinbolia mosaic virus (DMV, KT878825) for clone MZ-Mal170.13-19. Based on current ICTV criteria for begomovirus species demarcation, the begomovirus cloned from *Pyrenacantha* sp., named Pyrenacantha yellow mosaic virus (PyYMV), represents a new species, for which the name *Begomovirus pyrenacanthae* is proposed. The two DNA-A components (both 2,766 nt long) have a genomic organization typical of Old World begomoviruses with six genes, two in the viral sense (*CP* and *AV2*) and four in the complementary sense (*Rep*, *TrAP*, *REn* and *AC4*). The DNA-B component (2,726 nt) has one gene in the viral sense (*NSP*) and one in the complementary sense (*MP*).

Interestingly, examination of the DNA-A and DNA-B CRs indicated a 35-nt insertion in the DNA-A CR, starting 38 nt downstream from the second direct repeat and ending 14 nt upstream from the nonanucleotide (Figure 2A). As a result of this insertion, the DNA-A CR is 243 nt long while the DNA-B CR is 208 nt long and they have only 82.7% nt sequence identity. Nevertheless, the two CRs have identical iterons, with the same inverted and direct repeats (TACCCC-GGTGTA-GGGGTA). Moreover, the iteron related domain of the Rep protein (MPPSRFKVN; Figure 2B), is predicted to recognise the GGTG iteron core sequence [35]. The sequence identity between the CRs when the 35-nt insertion is removed from the alignment is 96%. Together with the MspI restriction pattern (indicating the presence of only two begomovirus components in the sample), the presence of identical iterons and the high nt identity between the CRs when the 35-nt insertion is removed from the alignment indicate that the DNA-A and DNA-B components cloned from the *Pyrenacantha* sample are indeed cognate components of the same bipartite begomovirus. The 35-nt insertion likely originated from a recombination event. However, recombination analisys with RDP v. 4 [36] did not indicate a recombination event in the DNA-A, and BLAST analysis of the 35-nt insertion indicated no significant similarity with sequences in the databases.

Phylogenetic analysis indicates that the PyYMV DNA-A is located in a separate branch in a cluster with DMV and Asystasia mosaic Madagascar virus, two begomoviruses isolated from weeds (Figure 1B). The DNA-B also clusters with DMV (Figure 1C). In both cases, genetic distances between the PyYMV sequences and those of its closest relatives are large, indicating a high degree of divergence between PyYMV and other begomoviruses.

We have identified and molecularly characterized a new Old World, bipartite begomovirus infecting the weed plant *Pyrenacantha* sp. in Mozambique. Weeds can contribute to the prevalence and distribution of viruses in crops and act as a reservoir of begomovirus diversity [37-39]. Emerging viruses usually spread to a new host population from the wild host population [40]. Our surveys of common bean, soybean and cowpea fields near the area where PyYMV was found did not detect this virus in these crops (B.A.I. Chipiringo and F.M. Zerbini, unpublished results), and a recent study in tomato fields in Mozambique found a single begomovirus, tomato chlorotic stunt virus (ToCSV) [22]. Thus, PyYMV may be restricted to *Pyrenacantha* sp., similar to other begomoviruses from non-cultivated hosts which are also restricted to a single host species ("sealed containers") [2]. Whether it remains restricted to this host will depend on the virus- and ecosystem-specific factors that favors spill over events, such as host range and vector transmission properties (which are yet to be determined for PyYMV), or changes in agricultural practices that may lead to ecosystem simplification and increases in vector populations [21]. *Pyrenacantha* is a perennial shrub distributed across East Africa, and it is likely that it has been coexisting with PyYMV for a long time. Considering the rapid changes in agricultural practices currently taking place in Mozambique, the potential of PyYMV to emerge in crops plants should not be underestimated.

The identification and molecular characterization of PyYMV opens avenues for studying its diversity and distribution as well as to understand the viral and ecological factors that may drive its dissemination and eventual emergence in crops.

**Acknowledgements**

This work was funded by CAPES (Financial code 01), CNPq (409599/2016-6) and Fapemig (APQ-03276-18) grants to FMZ. BAIC was supported by a doctoral fellowship from Instituto de Bolsas de Estudos de Moçambique. OFLS was supported by a doctoral fellowship from Ministério da Ciência e Tecnologia, Ensino Superior e Técnico Profissional de Moçambique. The authors gratefully acknowledge the technical staff at the Fac. de Ciências Agrárias, Unizambeze, for their assistance in sample collection and processing.

**Data availability statement**

The sequences described in this study were deposited in GenBank under accession numbers MZ390982-MZ390894.

**References**

1. Vincent SJ, Coutts BA, Jones RAC (2014) Effects of introduced and indigenous viruses on native plants: Exploring their disease-causing potential at the agro-ecological interface. PLoS ONE 9:e91224
2. García-Arenal F, Zerbini FM (2019) Life on the edge: geminiviruses at the interface between crops and wild plant hosts. Annu Rev Virol 6:411-433
3. Mollel HG, Ndunguru J, Sseruwagi P, Alicai T, Colvin J, Navas-Castillo J, Fiallo-Olive E (2016) A novel East African monopartite begomovirus-betasatellite complex that infects Vernonia amygdalina. Arch Virol 162:1079-1082

4. Mollel HG, Ndunguru J, Sseruwagi P, Alicai T, Colvin J, Navas-Castillo J, Fiallo-Olivé E (2020) African basil (*Ocimum gratissimum*) is a reservoir of divergent begomoviruses in Uganda. Plant Dis 104:853-859
5. Mohammed HS, El Siddig MA, El Hussein AA, Navas-Castillo J, Fiallo-Olive E (2018) Complete genome sequence of datura leaf curl virus, a novel begomovirus infecting *Datura innoxia* in Sudan, related to begomoviruses causing tomato yellow leaf curl disease. Arch Virol 163:273-275
6. Kyallo M, Sseruwagi P, Skilton RA, Ochwo-Ssemakula M, Wasswa P, Ndunguru J (2017) Deinbollia mosaic virus: a novel begomovirus infecting the sapindaceous weed Deinbollia borbonica in Kenya and Tanzania. Arch Virol 162:1393-1396
7. Mollel HG, Sseruwagi P, Ndunguru J, Alicai T, Colvin J, Navas-Castillo J, Fiallo-Olivé E (2017) Desmodium mottle virus, the first legumovirus (genus *Begomovirus*) from East Africa. Arch Virol 162:1799-1803
8. Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, Navas-Castillo J, Rivera-Bustamante R, Varsani A, ICTV Consortium (2017) ICTV Virus Taxonomy Profile: *Geminiviridae*. J Gen Virol 98:131-133
9. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olive E, Briddon RW, Hernandez-Zepeda C, Idris A, Malathi VG, Martin DP, Rivera-Bustamante R, Ueda S, Varsani A (2015) Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Arch Virol 160:1593-1619
10. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Dempsey DM, Dutilh BE, Harrach B, Harrison RL, Hendrickson RC, Junglen S, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Nibert M, Orton RJ, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Varsani A, Zerbini FM, Davison AJ (2020) Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). Arch Virol 165:2737-2748
11. Zhou X (2013) Advances in understanding begomovirus satellites. Annu Rev Phytopathol 51:357-381
12. Padidam M, Beachy RN, Fauquet CM (1996) The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. Virology 224:390-404
13. Mansoor S, Briddon RW, Zafar Y, Stanley J (2003) Geminivirus disease complexes: An emerging threat. Trends Plant Sci 8:128-134
14. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL (2005) Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. Annu Rev Phytopathol 43:361-394
15. Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S (2013) Geminiviruses: Masters at redirecting and reprogramming plant processes. Nat Rev Microbiol 11:777-788
16. Noueiry AO, Lucas WJ, Gilbertson RL (1994) Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell 76:925-932
17. Lazarowitz SG, Beachy RN (1999) Viral movement proteins as probes for intracellular and intercellular trafficking in plants. Plant Cell 11:535-548
18. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D (1999) Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Plant Sci 18:71-106
19. Arguello-Astorga G, Herrera-Estrella L, Rivera-Bustamante R (1994) Experimental and theoretical definition of geminivirus origin of replication. Plant Mol Biol 26:553-556
20. Arguello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF (1994) Geminivirus replication origins have a group-specific organization of interative elements: a model for replication. Virology 203:90-100
21. Roossinck MJ, Garcia-Arenal F (2015) Ecosystem simplification, biodiversity loss and plant virus emergence. Curr Opin Virol 10:56-62
22. Sande OFL, Orílio AF, Chipiringo BAI, Xavier CAD, Zerbini FM (2021) Speciation driven by recombination in the evolution of tomato curly stunt virus in Mozambique. Plant Pathol 70:994-1002
23. Cooper I, Jones RA (2006) Wild plants and viruses: under-investigated ecosystems. Adv Virus Res 67:1-47

24. Hyde MA, Wursten BT, Ballings P, Coates Palgrave M (2021) Flora of Mozambique: Species information: *Pyrenacantha kaurabassana*. Available at: https://www.mozambiqueflora.com/speciesdata/species.php?species_id=137280. Accessed on January 3, 2021
25. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull 19:11-15
26. Rojas MR, Gilbertson RL, Russell DR, Maxwell DP (1993) Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. Plant Dis 77:340-347
27. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met 116:209-211
28. Sambrook J, Russel D (2001) Molecular Cloning - A Laboratory Manual (3ª ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410
30. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277
31. Silva JCF, Carvalho TFM, Basso MF, Deguchi M, Pereira WA, Sobrinho RR, Vidigal PMP, Brustolini OJB, Silva FF, Dal-Bianco M, Fontes RLF, Santos AA, Zerbini FM, Cerqueira FR, Fontes EPB (2017) Geminivirus data warehouse: a database enriched with machine learning approaches. BMC Bioinformatics 18:240
32. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547-1549
33. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574
34. Nylander JAA (2004) MrModeltest v2. Program distributed by the author Evolutionary Biology Centre, Uppsala University
35. Arguello-Astorga GR, Ruiz-Medrano R (2001) An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Arch Virol 146:1465-1485
36. Martin D, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003
37. García-Andrés S, Monci F, Navas-Castillo J, Moriones E (2006) Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: Evidence for the presence of a new virus species of recombinant nature. Virology 350:433-442
38. Silva SJC, Castillo-Urquiza GP, Hora-Júnior BT, Assunção IP, Lima GSA, Pio-Ribeiro G, Mizubuti ESG, Zerbini FM (2011) High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in northeastern Brazil. Arch Virol 156:2205-2213
39. Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, Lau D, Zerbini FM (2017) The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from *Sida* spp. and *Leonurus sibiricus*, plus two New World alphasatellites. Ann Appl Biol 170:204-218
40. Haydon DT, Cleaveland S, Taylor LH, Laurenson MK (2002) Identifying reservoirs of infection: A conceptual and practical challenge. Emerg Inf Dis 8:1468-1473

## Figure legends

**Figure 1. A.** Symptoms of yellow mosaic in the *Pyrenacantha* sp. sample. **B, C.** Bayesian phylogenetic trees based on the full-length nucleotide sequences of DNA-A (**B**) and DNA-B (**C**) of PyYMV (marked in red) and related begomoviruses. Labels indicate their acronyms and GenBank access numbers. Nodes with posterior probability values between 0.5 and 0.79 are indicated by gray circles, and nodes with values equal to or greater than 0.8 are indicated by black circles. The scale bar represents the number of nucleotide substitutions per site.

**Figure 2. A.** Nucleotide sequence alignment of the common regions (CRs) of PyYMV DNA-A and DNA-B. The iterons (two imperfect direct repeats and one inverted repeat) are underlined, the nonanucleotide is in bold italics, and the 35-nucleotide insertion in the DNA-A CR is highlighted in grey. **B.** Amino acid sequence alignment of the Rep protein N-terminal region of PyYMV and the most closely related begomoviruses. The iteron-related domain (IRD) is indicated in bold. Motif 1, involved in rolling-circle replication, is underlined.

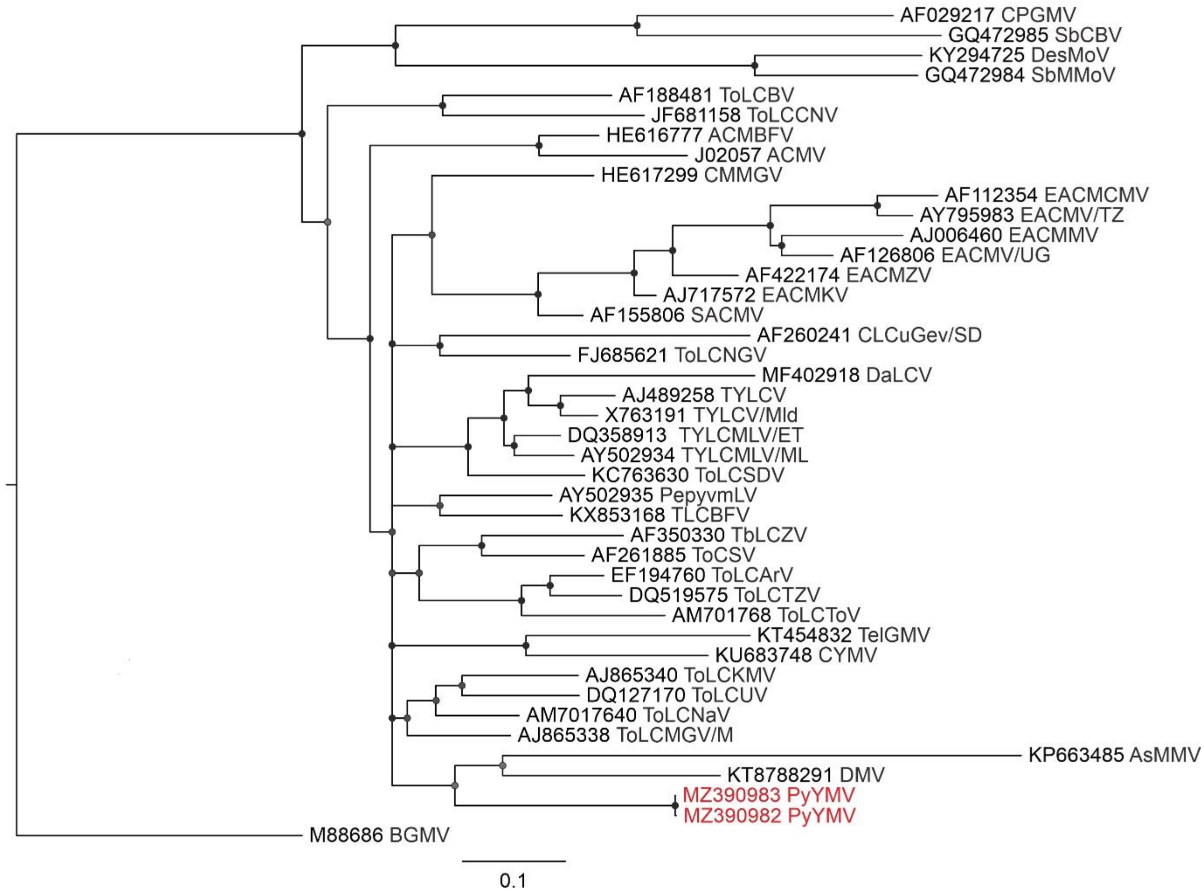
**Supplementary Figure S1**. Pairwise sequence comparisons of (**A**) DNA-A and (**B**) DNA-B of PyYMV and related begomoviruses. Identities were calculated using SDT v. 1.2 from a multiple sequence alignment prepared using the MUSCLE algorithm implemented in MEGA X.

**Figure 1**

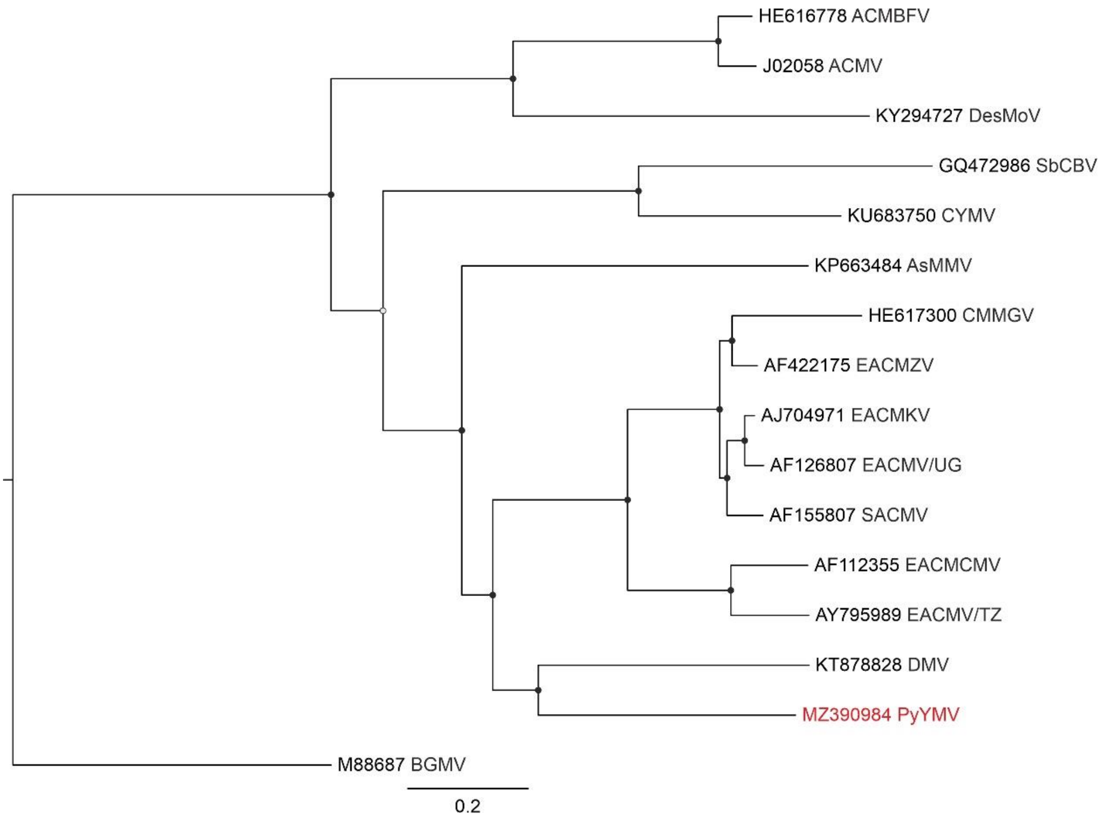
**A**


PyYMV

**B**


AF029217 CPGMV
GQ472985 SbCBV
KY294725 DesMoV
GQ472984 SbMMoV
AF188481 ToLCBV
JF681158 ToLCCNV
HE616777 ACMBFV
J02057 ACMV
HE617299 CMMGV
AF112354 EACMCMV
AY795983 EACMV/TZ
AJ006460 EACMMV
AF126806 EACMV/UG
AF422174 EACMZV
AJ717572 EACMKV
AF155806 SACMV
AF260241 CLCuGev/SD
FJ685621 ToLCNGV
MF402918 DaLCV
AJ489258 TYLCV
X763191 TYLCV/Mld
DQ358913 TYLCMLV/ET
AY502934 TYLCMLV/ML
KC763630 ToLCSDV
AY502935 PepyvmLV
KX853168 TLCBFV
AF350330 TbLCZV
AF261885 ToCSV
EF194760 ToLCArV
DQ519575 ToLCTZV
AM701768 ToLCToV
KT454832 TelGMV
KU683748 CYMV
AJ865340 ToLCKMV
DQ127170 ToLCUV
AM7017640 ToLCNaV
AJ865338 ToLCMGV/M
KP663485 AsMMV
KT8788291 DMV
MZ390983 PyYMV
MZ390982 PyYMV
M88686 BGMV
0.1

**C**


HE616778 ACMBFV
J02058 ACMV
KY294727 DesMoV
GQ472986 SbCBV
KU683750 CYMV
KP663484 AsMMV
HE617300 CMMGV
AF422175 EACMZV
AJ704971 EACMKV
AF126807 EACMV/UG
AF155807 SACMV
AF112355 EACMCMV
AY795989 EACMV/TZ
KT878828 DMV
MZ390984 PyYMV
M88687 BGMV
0.2

# Figure 2

## A

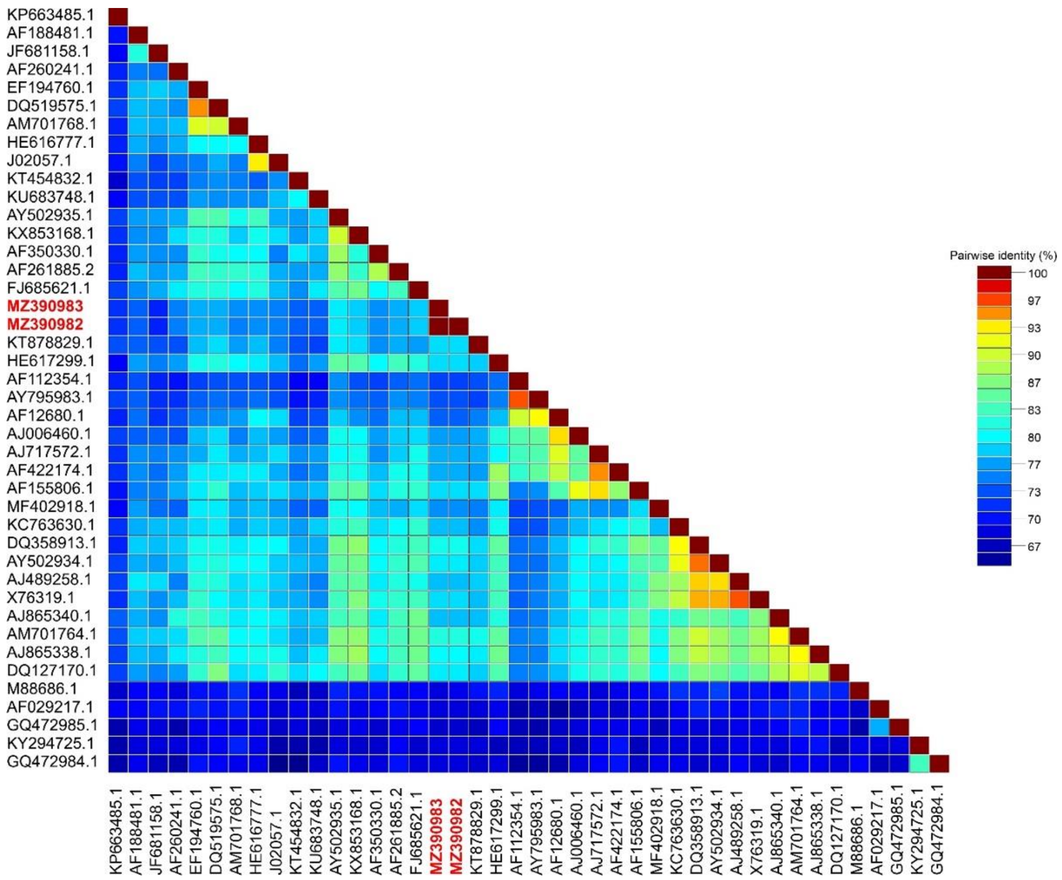
```
DNA-A (MZ390982)    TTTGGTCAATGTACCCCGATTGACCCCCTCTCTATCTATTCCCTGATATCGGTGTAACGG    60
DNA-A (MZ390983)    TTTGGTCAATGTACCCCGATTGACCCCCTCTCTATCTATTCCCTGATATCGGTGTAACGG    60
DNA-B (MZ390984)    TTTGGTCGTTGTACCCCGATTGACCCCCTCTCTATCTATTCCCTGATATCGGTGTAACGG    60

DNA-A (MZ390982)    GGTACAATATATACTCGTACCCCGAATGGCACGCTCGTAATTTCTGCTGAAATTTACCGC    120
DNA-A (MZ390983)    GGTACAATATATACTCGTACCCCGAATGGCACGCTCGTAATTTCTGCTGAAATTTACCGC    120
DNA-B (MZ390984)    GGTACAATATATACTCGTACCCCGAATGGCACGCTTGTAATT------------------    102

DNA-A (MZ390982)    TCAAATTTTGAATTTTTTCGCGGCCATCCTTTAATATTACCGGATGGCCGCGCCCCCGCC    180
DNA-A (MZ390983)    TCAAATTTTGAATTTTTTCGCGGCCATCCTTTAATATTACCGGATGGCCGCGCCCCCGCC    180
DNA-B (MZ390984)    -----------------TCGCGGCCATCCTTTAATATTACCGGATGGCCGCGCCCCCGCC    145

DNA-A (MZ390982)    CGACGTGGGCCCCACACGAAAGAGACAGGGCAACCAATCATATTTGGCGCTGAATGGTTA    240
DNA-A (MZ390983)    CGACGTGGGCCCCACACGAAAGAGACAGGGCAACCAATCATATTTGGCGCTGAATGGTTA    240
DNA-B (MZ390984)    CGACGTGGGCCCCACACGAAAGAGACAGGGCAACCAATCATATTGGCCGCTGAATGGTAA    205

DNA-A (MZ390982)    AAT   243
DNA-A (MZ390983)    AAT   243
DNA-B (MZ390984)    ATT   208
```

## B

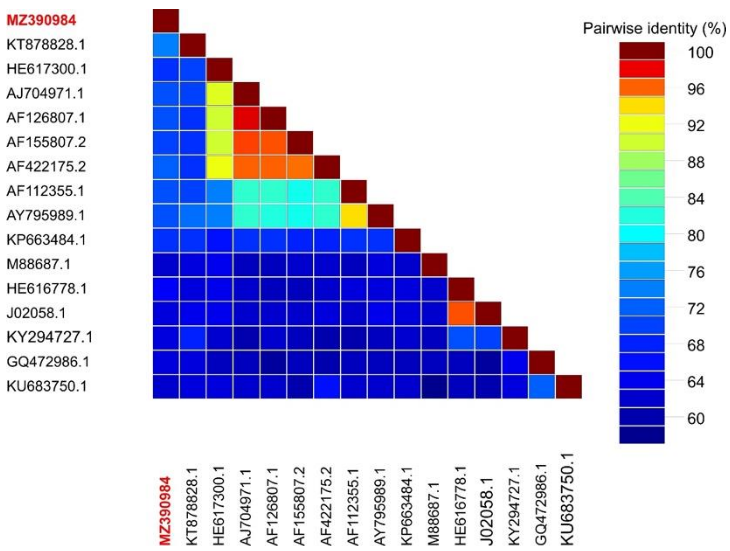
```
PYYMV  (MZ390982)   MAPPSRFKVNAKNYFLTYPRCSLSKEEALAQLQNLNTPTKKLFIKVCREFHENGEPHLHA   60
DMV    (KT878829)   MPRAGRFSINAKNYFLTYPNCPLDKNEALSQLQAKPTPVNKLFIKVCRELHESGEPHLHA   60
ToLCNaV(AM701764)   MAPPRRFKIFAKNFFLTYPKCSLTKEEALSQIQALQTPVNKLFIKICRELHENGEPHLHM   60
SACMV  (KJ887927)   MPRAGRFSIKAKNYFLTYPKCNLSKEAALDQLRQLQTPTNKLFIKICRELHDNGEPHLHA   60
                          IRD    Motif1
```

**Supplementary Figure S1**

**A**




**B**



**Supplementary Table S1**. Begomovirus sequences obtained from GenBank.

**DNA-A**

| **Species** | **Virus name** | **Isolate** | **Accession #** | **Acronym** |
|---|---|---|---|---|
| *African cassava mosaic Burkina Faso virus* | African cassava mosaic Burkina Faso virus | BF-Oua-127-08 | HE616777 | ACMBFV |
| *African cassava mosaic virus* | African cassava mosaic virus | Cameroon-1998 | J02057 | ACMV |
| *Bean golden mosaic virus* | Bean golden mosaic virus | Brazil-Campinas 1-1978 | M88686 | BGMV |
| *Cassava mosaic Madagascar virus* | Cassava mosaic Madagascar virus | Madagascar-Toliary-2006 | HE617299 | CMMGV |
| *Cotton leaf curl Gezira virus* | Cotton leaf curl Gezira virus | Sudan-Gezira-1996 | AF260241 | CLCuGev-SD |
| *Cowpea golden mosaic virus* | Cowpea golden mosaic virus | Nigeria-Nsukka-1990 | AF029217 | CPGMV |
| *Datura leaf curl virus* | Datura leaf curl virus | SD-Kha435-16 | MF402918 | DaLCV |
| *Deinbollia mosaic virus* | Deinbollia mosaic virus | Tanzania-DB_T1A-2015 | KT878829 | DMV |
| *Desmodium mottle virus* | Desmodium mottle virus | Uganda-Ki_UG5-2015 | KY294725 | DesMoV |
| *East African cassava mosaic Cameroon virus* | East African cassava mosaic Cameroon virus | Cameroon-1998 | AF112354 | EACMCMV |
| *East African cassava mosaic Kenya virus* | East African cassava mosaic Kenya virus | Kenya-Mitaboni-K298-2002 | AJ717572 | EACMKV |
| *East African cassava mosaic Malawi virus* | East African cassava mosaic Malawi virus | Malawi-K-1996 | AJ006460 | EACMMV |
| *East African cassava mosaic virus* | East African cassava mosaic virus | Uganda-Grave 2-1997 | AF126806 | EACMV-UG |
| *East African cassava mosaic virus* | East African cassava mosaic virus | Tanzania-1-2001 | AY795983 | EACMV-TZ |
| *East African cassava mosaic Zanzibar virus* | East African cassava mosaic Zanzibar virus | Tanzania-Uguja-1998 | AF422174 | EACMZV |
| *Pepper yellow vein Mali virus* | Pepper yellow vein Mali virus | Mali River | AY502935 | PepyvmLV |
| *South African cassava mosaic virus* | South African cassava mosaic virus | South Africa | AF155806 | SACMV |
| *Soybean chlorotic blotch virus* | Soybean chlorotic blotch virus | Nigeria-Soybean 19-2007 | GQ472985 | SbCBV |
| *Soybean mild mottle virus* | Soybean mild mottle virus | Nigeria-Soybean 17-2007 | GQ472984 | SbMMoV |
| *Telfairia golden mosaic virus* | Telfairia golden mosaic virus | Cameroon-BBT20-2014 | KT454832 | TelGMV |
| *Tobacco leaf curl Zimbabwe virus* | Tobacco leaf curl Zimbabwe virus | Zimbabwe | AF350330 | TbLCZV |
| *Tomato curly stunt virus* | Tomato curly stunt virus | South Africa-Onderberg-1998 | AF261885 | ToCSV |
| *Tomato leaf curl Arusha virus* | Tomato leaf curl Arusha virus | Tanzania-Kilimandjaro-2005 | EF194760 | ToLCArV |
| *Tomato leaf curl Burkina Faso virus* | Tomato leaf curl Burkina Faso virus | Burkina Faso-Loumbila-Tomate51B1-2013 | KX853168 | TLCBFV |
| *Tomato leaf curl Bangladesh virus* | Tomato leaf curl Bangladesh virus | Bangladesh-BD2 | AF188481 | ToLCBV |
| *Tomato leaf curl China virus* | Tomato leaf curl China virus | China-Guangxi-Tobacco-JX-2-2010 | JF681158 | ToLCCNV |
| *Tomato leaf curl Comoros virus* | Tomato leaf curl Comoros virus | Mayotte-Kahani-2003 | AJ865340 | ToLCKMV |
| *Tomato leaf curl Madagascar virus* | Tomato leaf curl Madagascar virus | Madagascar-Morondova-2001 | AJ865338 | ToLCMGV-M |
| *Tomato leaf curl Namakely virus* | Tomato leaf curl Namakely virus | Madagascar-Namakely-2001 | AM701764 | ToLCNaV |
| *Tomato leaf curl Nigeria virus* | Tomato leaf curl Nigeria virus | Nigeria-2006 | FJ685621 | ToLCNGV |
| *Tomato leaf curl Sudan virus* | Tomato leaf curl Sudan virus | Sudan-WM-2011 | KC763630 | ToLCSDV |
| *Tomato leaf curl Uganda virus* | Tomato leaf curl Uganda virus | Uganda-Iganga-2005 | DQ127170 | ToLCUV |
| *Tomato leaf curl Tanzania virus* | Tomato leaf curl Tanzania virus | TZ-Ten-05 | DQ519575 | ToLCTZV |
| *Tomato leaf curl Toliara virus* | Tomato leaf curl Toliara virus | Madagascar-Miandrivazo-2001 | AM701768 | ToLCToV |
| *Tomato yellow leaf curl Mali virus* | Tomato yellow leaf curl Mali virus | Ethiopia-Melkassa-2005 | DQ358913 | TYLCMLV-ET |
| *Tomato yellow leaf curl Mali virus* | Tomato yellow leaf curl Mali virus | Mali-2003-Mali | AY502934 | TYLCMLV-ML |
| *Tomato yellow leaf curl virus* | Tomato yellow leaf curl virus | Spain-Almeria-Pepper-1999 | AJ489258 | TYLCV |
| *Tomato yellow leaf curl virus* | Tomato yellow leaf curl virus | Israel-1993-Mild | X76319 | TYLCV-Mld |
| *Cotton yellow mosaic virus* | Cotton yellow mosaic virus | Benin-Gos_San2-2014 | KU683748 | CYMV |

| | | | | |
|---|---|---|---|---|
| *Asystasia mosaic Madagascar virus* | Asystasia mosaic Madagascar virus | Madagascar-MG493-2011 | KP663485 | AsMMV |

**DNA-B**

| **Species** | **Virus name** | **Isolate** | **Accession #** | **Acronym** |
|---|---|---|---|---|
| *African cassava mosaic Burkina Faso virus* | African cassava mosaic Burkina Faso virus | BF-Oua-127-08 | HE616778 | ACMBFV |
| *African cassava mosaic virus* | African cassava mosaic virus | Cameroon-1998 | J02058 | ACMV |
| *Bean golden mosaic virus* | Bean golden mosaic virus | Brazil-Campinas 1-1978 | M88687 | BGMV |
| *Cassava mosaic Madagascar virus* | Cassava mosaic Madagascar virus | Madagascar-Toliary-2006 | HE617300 | CMMGV |
| *Deinbollia mosaic virus* | Deinbollia mosaic virus | Tanzania-DB_T1A-2015 | KT878828 | DMV |
| *Desmodium mottle virus* | Desmodium mottle virus | Uganda-Ki_UG5-2015 | KY294727 | DesMoV |
| *East African cassava mosaic Cameroon virus* | East African cassava mosaic Cameroon virus | Cameroon-1998 | AF112355 | EACMCMV |
| *East African cassava mosaic Kenya virus* | East African cassava mosaic Kenya virus | Kenya-Mitaboni-K298-2002 | AJ704971 | EACMKV |
| *East African cassava mosaic virus* | East African cassava mosaic virus | Uganda-Grave 2-1997 | AF126807 | EACMV-UG |
| *East African cassava mosaic virus* | East African cassava mosaic virus | Tanzania-1-2001 | AY795989 | EACMV-TZ |
| *East African cassava mosaic Zanzibar virus* | East African cassava mosaic Zanzibar virus | Tanzania-Uguja-1998 | AF422175 | EACMZV |
| *South African cassava mosaic virus* | South African cassava mosaic virus | South Africa | AF155807 | SACMV |
| *Soybean chlorotic blotch virus* | Soybean chlorotic blotch virus | Nigeria-Soybean 19-2007 | GQ472986 | SbCBV |
| *Cotton yellow mosaic virus* | Cotton yellow mosaic virus | Benin-Gos_San2-2014 | KU683750 | CYMV |
| *Asystasia mosaic Madagascar virus* | Asystasia mosaic Madagascar virus | Madagascar-MG493-2011 | KP663484 | AsMMV |

# CHAPTER 2

**Emergence of a bipartite begomovirus in Mozambique and its recombination-driven adaptation to legume crops**

Chipiringo, B.A.I., Barbosa, T.M.C., Cascardo, R.S., Sande, O.F.L., Quadros, A.F.F., Zerbini, F.M. (2021) Emergence of a bipartite begomovirus in Mozambique and its recombination-driven adaptation to legume crops. **Plant Pathology**, *in preparation*.

# Emergence of a bipartite begomovirus in Mozambique and its recombination-driven adaptation to legume crops

Baltazar A. I. Chipiringo[1,2], Tarsiane M.C. Barbosa[1], Renan S. Cascardo[1], Osvaldo F.L. Sande[1,3], Ayane F.F. Quadros[1], F. Murilo Zerbini[1*]

[1]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil

[2]Universidade Zambeze (Unizambeze), Faculdade de Ciências Agrárias, P.O. Box 213, Ulónguè, Tete 2306, Mozambique

[3]Divisão de Agricultura, Instituto Superior Politécnico de Manica (DivAG-ISPM), Manica, Mozambique

*Corresponding author:

Phone: (+55-31) 3612-2423

E-mail: zerbini@ufv.br

**Abstract**

Begomoviruses cause serious constraints to crop production in many tropical and subtropical areas of the world. In March 2019, symptomatic samples of bean, soybean and cowpea plants were collected in central and northern Mozambique. Seven begomovirus DNA-A components were cloned and sequenced. The seven sequences are 96% identical, indicating that they represent the same virus. The highest identity with other begomoviruses is 85% with cowpea golden mosaic virus. Eleven DNA-B components were cloned and sequenced, with 93% identity amongst each other and a maximum of 72% sequence identity with cotton yellow mosaic virus. Alignment of the CRs and the identification of identical iterons indicated that the DNA-A and DNA-B components are cognate components of the same virus, for which the name Mozambique legume mosaic virus (MLMV) is proposed. Three recombination events (including both intra- and interspecific events) were detected in the DNA-B, affecting the *MP* gene and the short intergenic region (SIR). Analysis of nucleotide diversity indicates that the MLMV DNA-B is more variable than the DNA-A. MLMV is an emerging begomovirus in legume crops in Mozambique, and its adaptation to new hosts appears to be driven by recombination acting upon the DNA-B.

**Keywords:** MLMV. Mozlegume. Bean. Soybean. Cowpea.

## Introduction

Legume (pulse) crops are of extreme importance in developing countries as a vital source of dietary protein that substitutes for animal protein. In Mozambique, legume crops occupy a prominent socio-economic place due to their importance in food security and income generation for farmers. Smallholder farmers are mainly responsible for the production in an environment dominated by subsistence agriculture (INE, 2011, Haber *et al.*, 2015). The main legume crops in Mozambique are common bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*) and soybean (*Glycine max*). Common bean and cowpea are produced mainly as intercrops with maize or in monoculture in small areas. Soybean is a recently introduced cash crop produced in monoculture by small and medium farmers.

Increased agricultural production is an urgent necessity in Mozambique, given the growing demand for food. With the diversification and intensification of practices to increase food production and combat food insecurity, the emergence of plant diseases, and particularly of plant viral diseases, may be impacted (Roossinck and Garcia-Arenal, 2015). Diseases caused by viruses are one of the main biotic causes of production losses in legumes. Detrimental factors to productivity, which includes diseases, need to be increasingly understood and controlled to avoid losses (Giglioti *et al.*, 2015).

The geminiviruses (family *Geminiviridae*) constitute a group of plant viruses of great relevance to agriculture due to the severity of the diseases caused by them to vegetable and field crops worldwide, particularly in tropical and subtropical regions (Shepherd *et al.*, 2010, Rocha *et al.*, 2013, Sattar *et al.*, 2013, Jacobson *et al.*, 2018). Geminiviruses are classified into fourteen genera based on their genomic arrangement, type of insect vector, phylogenetic relationships and host range (Zerbini *et al.*, 2017). The genus *Begomovirus* is the largest in the family, with 445 currently recognized species (Zerbini *et al.*, 2017, Walker *et al.*, 2020).

Begomoviruses are transmitted in nature by whiteflies of the *Bemisia tabaci* cryptic species complex to dicotyledonous plants (Brown *et al.*, 2015). They can be divided into two major groups based on phylogenetic relationships and genome characteristics: Old World (OW; Europe, Africa and Asia) and New World (NW; the Americas) (Rybicki, 1994). Old World begomoviruses can have mono- or bipartite genomes and are often associated with satellite DNAs known as alpha- and betasatellites (Zhou, 2013, Brown *et al.*, 2015). Begomoviruses affecting legumes in the OW have mostly bipartite genomes and constitute a phylogenetically isolated group sometimes refered to as "legumovirses" (Qazi *et al.*, 2007).

The two genomic components of bipartite begomoviruses, termed DNA-A and DNA-B, have no significant sequence identity, except for an intergenic region (termed the common region, CR) of about 200 nucleotides. The CR is important for maintaining the integrity of the bipartite genome and allows both components to be replicated at similar rates (Lazarowitz, 1992). The CRs of cognate genomic components have nucleotide sequence identity >94% and contain a sequence capable of forming a stem-loop structure with the nonanucleotide 5'-TAATATTAC-3', conserved in all begomoviruses and which constitutes the functional origin of replication (Fontes *et al.*, 1994b, Laufs *et al.*, 1995).

The DNA-A of Old World begomoviruses contains six genes: *Rep*, responsible for encoding the only protein essential for replication, initiator of the rolling circle replication mechanism (Fontes *et al.*, 1994a, Orozco *et al.*, 1997); *TrAP*, which encodes a transcriptional factor of the *CP* and *NSP* genes and which also acts as a suppressor of plant defence responses (Voinnet *et al.*, 1999, Wang *et al.*, 2005); *REn*, responsible for encoding an accessory (non-essential) factor of viral replication (Sunter *et al.*, 1990, Pedersen and Hanley-Bowdoin, 1994); *AC4*, responsible for encoding a protein involved in the suppression of RNA-mediated gene silencing (Vanitharani *et al.*, 2004); *CP*, responsible for encoding the capsid protein and which is also essential for virus transmission by the insect vector (Briddon *et al.*, 1990, Hofer *et al.*, 1997); and *V2* ("pre-coat"), responsible for encoding a protein that acts in the movement of the virus in the plant (Padidam *et al.*, 1996). The DNA-B contains two genes: *NSP*, encoding a protein responsible for transporting the DNA across the nuclear envelope (nucleus-cytoplasm) (Noueiry *et al.*, 1994, Sanderfoot and Lazarowitz, 1995), and *MP*, encoding a protein involved in the cell-to-cell movement of the virus across the plasmodesmata (Noueiry *et al.*, 1994).

The frequency with which new begomoviruses have emerged in the past few decades suggests that these viruses have a strong capacity to adapt to new hosts in an agricultural context, and therefore pose a serious threat to agricultural sustainability. Although the factors that favour the emergence of viral diseases are not completely understood, the persistent infection of non-cultivated (wild) host species, the expansion of monocultures, and the dissemination of highly polyphagous and invasive vectors may contribute to the transfer of viruses, including begomoviruses, from non-cultivated hosts to crops (Péréfarres *et al.*, 2012, Rey *et al.*, 2012, García-Arenal and Zerbini, 2019, McLeish *et al.*, 2020).

Despite the importance of legume crops in Mozambique, there has never been an effort to screen for legume viruses. In this context, in March 2019, common bean, cowpea and soybean samples were collected in the central and northern regions of Mozambique in order to identify and characterize begomoviruses infecting these plants. Eighteen complete genomes (seven DNA-A and eleven DNA-B) were cloned and sequenced. Analyses indicated the presence of a new OW bipartite begomovirus, for which the name Mozambique legume mosaic virus (MLMV) is proposed. We describe here the molecular characterization of MLMV and the assessment of the genetic variability and evolutionary history of this emerging begomovirus.

## Materials and Methods

**Sample collection.** Plants of common bean, cowpea and soybean displaying symptoms of mosaic and leaf deformation were collected in fields in central and northern Mozambique in March 2019 (Figure 1A). The collected samples were pressed onto newspaper sheets in a wooden frame and transported to the Microbiology Laboratory of the Faculty of Health Sciences of Tete in Mozambique for further analysis. The geographical coordinates of each collection point were recorded and digital images of each collected sample were obtained.

**Extraction of viral DNA and cloning of complete genomes.** Total DNA was extracted from dried leaf discs following the protocol described by Doyle and Doyle (1987) and stored in a freezer (-20°C). To confirm the presence of begomoviruses in the samples, PCR was performed using universal oligonucleotides (PAL1v1978/PAR1c496), designed to amplify a 1.4 kbp fragment of the DNA-A component (Rojas *et al.*, 1993). Next, for the PCR-positive samples, the DNA was used as a template for rolling circle amplification (RCA) of the viral genomes (Inoue-Nagata *et al.*, 2004). Concatamers were initially cleaved with the 4-base cutter enzyme MspI and the band pattern in agarose gels (1.2%) was analyzed. After confirmation of a cleavage pattern indicative of infection by begomoviruses (*ie*, sum of the fragments being approximately 5.4 kbp), the concatamers were cleaved with restriction enzymes (BamHI, EcoRI, HindIII, KpnI and PstI) to obtain monomeric genomic fragments for cloning. Fragments of about 2,700 nucleotides, corresponding to one genomic copy of a linearized component, were ligated to the pBluescriptKS+ vector (Stratagene) previously cleaved with the same enzyme and dephosphorylated. The recombinant plasmids were used for transformation in *Escherichia coli* DH5α (Sambrook and Russel, 2001) and selected clones were sequenced commercially (Macrogen, South Korea).

**Sequence assembly, pairwise comparisons and species identification.** Full-length begomovirus genomes were assembled using SeqAssm v.1.0 (www.sequentix.de) and adjusted to start from the nonanucleotide cleavage site (5'-TAATATT//AC-3'). The sequences were initially analysed with the BLAST*n* algorithm (Altschul *et al.*, 1990) to determine the viral species with which they shared the highest identity. Sequences of the most closely related viruses (Suppl. Table S1) were then used to assign taxonomy to the new isolates using Species Demarcation Tool v. 1.2 (Muhire *et al.*, 2014). The taxonomic positioning of the viral isolate was defined according to the criteria for begomovirus species and strain demarcation established by the ICTV *Geminiviridae* and *Tolecusatellitidae* Study Group (<91% and <94% nucleotide sequence identity for the full-length DNA-A, respectively) (Brown *et al.*, 2015). Cognate DNA-A and DNA-B components were identified on the basis of sequence identity in their common regions. The search for ORFs was carried out in the Geminivirus Data Warehouse platform (Silva *et al.*, 2017) and the ORF Finder program (www.ncbi.nlm.nih.gov/projects/gorf/).

**Phylogenetic analysis**. The complete genomes were aligned using the MUSCLE algorithm implemented in the MEGA X program (Kumar *et al.*, 2018). Phylogenetic trees were built through Bayesian inference with the program MrBayes v. 3.2.7a (Ronquist *et al.*, 2012). The CIPRES Science Gateway was used to run phylogenetic analyses (Miller *et al.*, 2010). The General Time Reversible nucleotide substitution model considering the gamma distribution with invariant sites (GTR+G+I) was selected with the MrModeltest v. 2.3 program (Nylander, 2004) based on the Akaike Information Criterion (AIC). The estimation of divergence time of the isolates was calculated using the time-structured evolutionary model. The analysis was based on the Markov Chain Monte Carlo (MCMC) algorithm considering 10,000,0000 generations and the trees generated every 2,000 generations. The trees were visualized and edited with FigTree (tree.bio.ed.ac.uk/software figgtree/).

**Genetic variability.** The main descriptors of genetic variability were quantified using DnaSP v. 6 (Rozas *et al.*, 2017): mean number of nucleotide differences (K), nucleotide diversity ($\pi$), number of haplotypes (H), haplotype diversity (Hd). The hypothesis of selection acting on viral genes was tested by means of three methods implemented in the DnaSP program (Tajima D, Fu & LiD* and Fu & Li F*).

**Recombination analysis.** For analysis of the recombination events a data set was created with the complete genomes obtained plus additional sequences of related begomoviruses retrieved from GenBank. These genomes were aligned based on the MUSCLE algorithm implemented in MEGA X. The detection of putative recombination events was performed using the RDP, Bootscan, Maximum chi square, Chimaera, SisterScan, 3Seq and Geneconv methods implemented in RDP v. 5.3 (Martin *et al.*, 2020). Only recombination events detected by at least four methods were considered reliable.

## Results

A total of 52 bean, 41 cowpea and 77 soybean samples were collected. The symptoms ranged from vein clearing, yellowing and mosaic (Figure 1B), suggestive of infection by begomoviruses. Indeed, the presence of begomoviruses was confirmed by conventional PCR using universal primers in 38, 39 and 47 samples of bean, cowpea and soybean, respectively (data not shown). From the PCR-positive samples, a total of 18 clones were obtained (three, six and nine from bean, cowpea and soybean samples, respectively; Table 1) and completely sequenced.

BLAST*n* and SDT analyses showed that seven clones (MZ-Bar82-19, MZ-Cum149-19, MZ-Gur81.5-19, MZ-Lic152-19, MZ-Lic153-19, MZ-Mal164-19 and MZ-Met163-19), with 95.8-100% nucleotide (nt) sequence identity with each other, corresponded to a begomovirus DNA-A, and had the highest nt sequence identity (85%) with the DNA-A of cowpea golden mosaic virus from Nigeria (CGMV, GenBank accession AF029217) (Suppl. Figure S1). Eleven clones (MZ-Bar82.1-19, MZ-Bar84-19, MZ-Bar85-19, MZ-Bar85.9-19, MZ-Bar86-19, MZ-Bar86.9-19, MZ-Gur81.4-19, MZ-Lic154-19, MZ-Mal164.1-19, MZ-Sus89-19 and MZ-Sus90.6-19), with 93.3-100% nt sequence identity with each other, corresponded to a begomovirus DNA-B, and had the highest nt sequence identity (72%) with the DNA-B from cotton yellow mosaic virus from Benin (CYMV, KU683750) (Suppl. Figure S2).

All DNA-A components (2739-2740 nt) have a genomic organization typical of Old World begomoviruses, with six genes, two in the viral sense (*CP* and *AV2*) and four in the complementary sense (*Rep*, *TrAP*, *REn* and *AC4*). The DNA-B components (2669-2717 nt) have one gene in the viral sense (*NSP*) and one in the complementary sense (*MP*).

From three samples (cowpea sample NH81 and soybean samples S82 and S164), both a DNA-A and a DNA-B were cloned. The common regions of these three DNA-A and DNA-B pairs have >94% nt sequence identities (Figure 2A). These results indicate that the DNA-A and DNA-B components are the cognate components of the same begomovirus. The CRs contain the conserved nonanucleotide 5′-TAATATTAC-3′ and the iterative sequences (iterons) typical of begomoviruses, with two imperfect direct repeats (GGTGTA-GGGGTA) and one inverted repeat (TACACC) (Figure 2A). The Rep protein contains an iteron related domain (FRVN) which is predicted to recognize the core iteron sequence GGTG (Figure 2B) (Arguello-Astorga and Ruiz-Medrano, 2001).

Based on the current ICTV criteria for begomovirus species demarcation (<91% nt sequence identity for the DNA-A component), the begomovirus identified infecting legume crops in north-central Mozambique represents a new bipartite, Old World begomovirus, for which the name Mozambique legume mosaic virus (MLMV) is proposed. The species name *Begomovirus mozlegume* is proposed.

Analysis of Bayesian phylogenetic trees based on the complete DNA-A nt sequences showed that MLMV clusters with other begomoviruses infecting legumes in Africa in a monophyletic clade (Figure 3A). Phylogenetic analysis based on the DNA-B showed incongruence with the DNA-A tree. The MLMV isolates clustered with cotton yellow mosaic virus (CYMV) in a monophyletic clade adjacent to a clade comprised of soybean chlorotic blotch virus (SbCBV) isolates (Figure 3B). This suggested the involvement of recombination in the evolutionary history of the MLMV DNA-B. It should be noted, however, that there is no complete DNA-B sequence for CPGMV (the virus with the closest relationship with MLMV based on the DNA-A).

Analysis of putative recombination events for the MLMV intraspecific DNA-A data set indicated two events with recombination signals supported by four to seven methods. However, these events had high *p*-values and will not be detailed here. For the intraspecific DNA-B data set, three well-supported events were detected (Suppl. Table S2). The first event, detected in isolates MZ-Bar82.1-19, MZ-Bar85.9-19 and MZ-Bar86-19, has breakpoints at nucleotides 984 to 1391 (encompassing *NSP*, the SIR and the 3'-region of *MP*), with isolates MZ-Bar86.9-19 and MZ-Mal164.1-19 as major and minor parents, respectively. The second and third events, detected in isolates MZ-Lic154-19 and MZ-Mal164.1-19, respectively, have breakpoints encompassing *MP* and the LIR (Suppl. Table S2).

We determined the descriptors of genetic variability for MLMV and for other Old World, legume-infecting begomoviruses. The MLMV DNA-A showed higher nucleotide diversity ($\pi$=0.02686) than SbCBV ($\pi$=0.01762) and similar to MYMV ($\pi$=0.02956) (Table 2; Suppl. Table S3). The haplotype diversity indices calculated for all viruses were similar and close to 1, indicating that most isolates were unique within each data set. The nucleotide diversity of the DNA-B was higher than that of the DNA-A for all viruses analysed, including MLMV (Table 2; Suppl. Table S3). As observed for the DNA-A, the nucleotide diversity of the MLMV DNA-B is higher than that of SbCBV and similar to that of MYMV, and it is lower than that of ACMV and EACMV (Table 2; Suppl. Table S3).

We analysed the nucleotide diversity indices for each gene and intergenic region in the DNA-A and DNA-B datasets of MLMV and MYMV, as these two viruses showed similar π values for their complete genomes. Although some differences were observed (eg, the MLMV *MP* gene being more variable than the MYMV *MP* gene), overall, the values were similar (Table 2), indicating that these two viruses have, indeed, equivalent levels of genetic variability. The degree of genetic variability is an intrinsic property of each begomovirus, regardless of whether it infects a cultivated or non-cultivated host (Ramos-Sobrinho et al., 2014, Rodelo-Urrego et al., 2015). Studies indicate the existence of a correlation between the number of recombination events and genetic variability (Xavier et al., 2021).

Three neutrality tests were used to assess the evidence of selection acting on the MYMV population. Although negative values were obtained in all cases, they were not statistically significant (Table 3).

## Discussion

In recent years, the discovery of new plant viruses has been increasing worldwide. They pose a major threat to plants in cultivated agroecosystems. Cowpea is native to Africa and has probably been cultivated in Mozambique long before the introduction of the common bean crop. Soybean is a recently introduced crop, whose social and economic importance has been increasing over the last 20 years.

In our study, we identified and molecularly characterised a new bipartite, Old World begomovirus infecting these three legume crops in Mozambique, which we named Mozambique legume mosaic virus (MLMV). This report is significant for a number of reasons. Firstly, it is the first report of a begomovirus infecting common bean in Africa, and only the second and third begomovirus reported infecting cowpea and soybean, respectively, in this continent (Alabi *et al.*, 2010). Secondly, MLMV was detected in samples of these three plants collected over 600 km apart. And thirdly, it was the only virus detected in the samples. Although a larger number of clones should be sequenced, it seems that MLMV is the prevalent (if not the only) begomovirus infecting these three important legume crops in Mozambique. We did collect a large number of samples (170 in total) and a large proportion of those samples (73%) were PCR-positive with universal begomovirus primers. Unfortunately, the DNA extracted from the samples in Mozambique was transported to Brazil under unfavourable conditions, which is probably the reason why it was not possible to obtain a larger number of full-length genomic clones using RCA. We plan to return to the field in Mozambique to collect additional samples and assess the true prevalence of MLMV in legume crops in the country. Sampling in the neighboring countries of Malawi, South Africa, Tanzania, Zambia and Zimbabwe would be desirable as well.

Two hypotheses could explain the emergence of a begomovirus infecting these three legume crops, both assuming that MLMV is indigenous to Mozambique or to southeastern Africa. It could have co-evolved with wild relatives of cowpea (which is native to Africa), from which it spilled over to, and became established in, cowpea crops, and was then transferred to common bean and soybean (both exotic to Africa). Or it could be the result of horizontal transfer of an indigenous begomovirus infecting a more distantly related, indigenous wild host. In the first scenario, the virus would be better adapted to cowpea compared to common bean and soybean, while in the second scenario it would have an equivalent degree of adaptation to all three crops. Thus, these hypotheses can be tested with biological experiments to assess the fitness of MLMV to each one of its three hosts (for example, based on virus accumulation in each host and the efficiency of transmission by the insect vector from and to each host). Moreover, it is logical to assume that additional non-cultivated and cultivated species could be hosts, highlighting the importance of future surveys not only to estimate the prevalence of MLMV in legume crops (as mentioned above) but also to identify possible additional hosts.

Begomoviruses reach new hosts by means of their polyphagous vectors, whiteflies of the *B. tabaci* cryptic species complex. Because all begomoviruses are transmitted by the same vectors, mixed infections are common and favour the occurrence of recombination, which is recognized as a defining force driving the evolution of begomoviruses worldwide (Lefeuvre *et al.*, 2007, Lefeuvre and Moriones, 2015). Whiteflies are widespread in Mozambique, where they have been associated with the dissemination of cassava mosaic disease (Cossa, 2011, Jacobson *et al.*, 2018). A recent study from our laboratory indicated the presence of three whitefly species in Mozambique, all indigenous to the African continent (Sub-Saharan Africa 1, SSA5 and SSA14) (Sande, 2020). The fact that MLMV was the only begomovirus detected in our samples is therefore surprising, and suggests that other begomoviruses infecting non-cultivated hosts may be poorly adapted to legume crops. Curiously, we have also detected a single begomovirus, tomato curly stunt virus (ToCSV) in a recent study of tomato crops (Sande *et al.*, 2021). It could be that the low-input, smallholder agricultural practices which still predominate in Mozambique do not favour the emergence of viruses, with MLMV and ToCSV being particularly well adapted to their (mostly exotic) hosts. However, this is obviously speculative and would have to be demonstrated experimentally.

Phylogenetic analyses indicated that the MLMV DNA-A clusters with begomoviruses that infect leguminous plants in the Old World. This is unusual for begomoviruses, which usually cluster based on geographical origin rather than host, but is consistent with the assertion that OW legume-infecting begomoviruses are genetically isolated (Qazi *et al.*, 2007). The incongruence between the observed DNA-A and DNA-B phylogenies could be the result of recombination acting on the DNA-B, although the absence of a complete DNA-B sequence for CPGMV is a confounding factor. However, even if the MLMV DNA-B is closely related to CPGMV, the placement of CYMV between MLMV and SbCBV (the other legume-infecting begomovirus from Africa) would probably not be affected.

The three recombination events of DNA-B were confirmed by RDP recombination analyses. The recombinant region comprises mainly the *MP* gene and the SIR. The inability of a virus to infect plants may result from defects in replication, movement or assembly of the virus (Noris *et al.*, 1998). The DNA-B of bipartite begomoviruses is required for systemic infection. Recombination provides viruses with a great potential for variation and therefore greater opportunity to evolve and rapidly adapt to changing environments (Seal *et al.*, 2006). Our results are consistent with a scenario in which the virus is gradually overcoming barriers in the infection process by recombination in the DNA-B, that is, this component is evolving in response to the pressure to adapt to new hosts.

Despite the small sample size, analysis of the nucleotide diversity index indicated that the MLMV DNA-B is more variable than the DNA-A, something that has been consistently observed for both OW and NW begomoviruses (Briddon *et al.*, 2010, Xavier *et al.*, 2021). Selection analysis may have been impaired by the small sample size. Nevertheless, these analyses did not indicate any significant differences in the variability of MLMV compared to other begomoviruses.

MLMV is an emerging begomovirus in legume crops in Mozambique, and its adaptation to new hosts appears to be driven by recombination acting on the DNA-B. Additional studies are necessary to better understand its variability and epidemiology, and to set up disease management strategies in case it becomes more widespread and/or severe in bean, cowpea or soybean crops in the future.

**Acknowledgements**

This work was funded by CAPES (Financial code 01), CNPq (409599-2016-6) and Fapemig (APQ-03276-18) grants to FMZ. BAIC was supported by a doctoral fellowship from Instituto de Bolsas de Estudos de Moçambique. OFLS was supported by a doctoral fellowship from Ministério da Ciência e Tecnologia, Ensino Superior e Técnico Profissional de Moçambique. The authors gratefully acknowledge the technical staff at the Fac. de Ciências Agrárias, Unizambeze, for their assistance in sample collection and processing.

**Data availability statement**

The sequences described in this study were deposited in GenBank under accession numbers MZ485479-MZ485496.

**References**

Alabi OJ, Kumar PL, Mgbechi-Ezeri JU, Naidu RA (2010). Two new 'legumoviruses' (genus *Begomovirus*) naturally infecting soybean in Nigeria. *Archives of Virology* 155, 643-656.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-410.

Arguello-Astorga GR, Ruiz-Medrano R (2001). An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Archives of Virology* 146, 1465-1485.

Briddon RW, Patil BL, Bagewadi B, Nawaz-Ul-Rehman MS, Fauquet CM (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology* 10, 97.

Briddon RW, Pinner MS, Stanley J, Markham PG (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology* 177, 85-94.

Brown JK, Zerbini FM, Navas-Castillo J, *et al.* (2015). Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. *Archives of Virology* 160, 1593-1619.

Cossa NS (2011). *Epidemiology of cassava mosaic disease in Mozambique*. Johannesburg, South Africa: University of the Witwatersrand, Ph.D. Thesis.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Fontes EPB, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L (1994a). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *Journal of Biological Chemistry* 269, 8459-8465.

Fontes EPB, Gladfelter HJ, Schaffer RL, Petty ITD, Hanley-Bowdoin L (1994b). Geminivirus replication origins have a modular organization. *Plant Cell* 6, 405-416.

García-Arenal F, Zerbini FM (2019). Life on the edge: geminiviruses at the interface between crops and wild plant hosts. *Annual Review of Virology* 6, 411-433.

Giglioti EA, Sumida CH, Canteri MG (2015). Disease phenomics. In: Fritsche-Neto R, Borém A, eds. *Phenomics: How Next-Generation Phenotyping is Revolutionizing Plant Breeding*. Cham, Switzerland: Springer International Publishing, 101-123.

Haber LL, Carvalho CE, Bowen W, Resende FV (2015). *Horticultura em Moçambique: caracteristicas, tecnologias de produção e de pós-colheita*. Brasília, DF: Embrapa, 2015.

Hofer P, Bedford ID, Markham PG, Jeske H, Frischmuth T (1997). Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. *Virology* 236, 288-295.

Ine (2011). *Censo Agropecuário 2009-2010: Resultados definitivos, Moçambique*

Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. *Journal of Virological Methods* 116, 209-211.

Jacobson AL, Duffy S, Sseruwagi P (2018). Whitefly-transmitted viruses threatening cassava production in Africa. *Current Opinion in Virology* 33, 167-176.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35, 1547-1549.

Laufs J, Traut W, Heyraud F*, et al.* (1995). *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences, USA* 92, 3879-3883.

Lazarowitz SG (1992). Geminiviruses: Genome structure and gene function. *Critical Reviews in Plant Sciences* 11, 327-349.

Lefeuvre P, Martin DP, Hoareau M*, et al.* (2007). Begomovirus 'melting pot' in the south-west Indian Ocean islands: Molecular diversity and evolution through recombination. *Journal of General Virology* 88, 3458-3468.

Lefeuvre P, Moriones E (2015). Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. *Current Opinion in Virology* 10, 14-19.

Martin D, Murrell B, Golden M, Khoosal A, Muhire B (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1, vev003.

Mcleish MJ, Fraile A, García-Arenal F (2020). Population genomics of plant viruses: the ecology and evolution of virus emergence. *Phytopathology* 111, 32-39.

Miller MA, Pfeiffer W, Schwartz T (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1-8.

Muhire BM, Varsani A, Martin DP (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9, e108277.

Noris E, Vaira AM, Caciagli P, Masenga V, Gronenborn B, Accotto GP (1998). Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *Journal of Virology* 72, 10050-10057.

Noueiry AO, Lucas WJ, Gilbertson RL (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell* 76, 925-932.

Nylander JaA (2004). *MrModeltest v2.* Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Orozco BM, Miller AB, Settlage SB, Hanley-Bowdoin L (1997). Functional domains of a geminivirus replication protein. *Journal of Biological Chemistry* 272, 9840-9846.

Padidam M, Beachy RN, Fauquet CM (1996). The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology* 224, 390-404.

Pedersen TJ, Hanley-Bowdoin (1994). Molecular characterization of the AL3 protein encoded by a bipartite geminivirus. *Virology* 202, 1070-1075.

Péréfarres F, Thierry M, Becker N, *et al.* (2012). Biological invasions of geminiviruses: case study of TYLCV and Bemisia tabaci in Reunion Island. *Viruses* 4, 3665-3688.

Qazi J, Ilyas M, Mansoor S, Briddon RW (2007). Legume yellow mosaic viruses: Genetically isolated begomoviruses. *Molecular Plant Pathology* 8, 343-348.

Ramos-Sobrinho R, Xavier CaD, Pereira HMB, *et al.* (2014). Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. *Journal of General Virology* 95, 2540-2552.

Rey MEC, Ndunguru J, Berrie LC, *et al.* (2012). Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the South-West Indian Ocean islands. *Viruses* 4, 1753-1791.

Rocha CS, Castillo-Urquiza GP, Lima ATM, *et al.* (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *Journal of Virology* 87, 5784-5799.

Rodelo-Urrego M, García-Arenal F, Pagán I (2015). The effect of ecosystem biodiversity on virus genetic diversity depends on virus species: A study of chiltepin-infecting begomoviruses in Mexico. *Virus Evolution* 1, vev004.

Rojas MR, Gilbertson RL, Russell DR, Maxwell DP (1993). Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* 77, 340-347.

Ronquist F, Teslenko M, Van Der Mark P, *et al.* (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61, 539-542.

Roossinck MJ, Garcia-Arenal F (2015). Ecosystem simplification, biodiversity loss and plant virus emergence. *Current Opinion in Virology* 10, 56-62.

Rozas J, Ferrer-Mata A, Sánchez-Delbarrio JC, *et al.* (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology* 34, 3299-3302.

Rybicki EP (1994). A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. *Archives of Virology* 139, 49-77.

Sande OFL (2020). *Begomoviruses and their vectors in Mozambique: diversity, evolution and population dynamics*. Viçosa, MG: Universidade Federal de Viçosa, D.Sc. Dissertation.

Sande OFL, Orílio AF, Chipiringo BaI, Xavier CaD, Zerbini FM (2021). Speciation driven by recombination in the evolution of tomato curly stunt virus in Mozambique. *Plant Pathology* 70, 994-1002.

Sanderfoot AA, Lazarowitz SG (1995). Cooperation in viral movement: The geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. *Plant Cell* 7, 1185-1194.

Sattar MN, Kvarnheden A, Saeed M, Briddon RW (2013). Cotton leaf curl disease - an emerging threat to cotton production worldwide. *Journal of General Virology* 94, 695-710.

Schenk S, Laddaga RA (1992). Improved method for electroporation of *Staphylococcus aureus*. *FEMS Microbiology Letters* 94, 133-138.

Seal SE, Van Den Bosch F, Jeger MJ (2006). Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. *Critical Reviews in Plant Sciences* 25, 23-46.

Shepherd DN, Martin DP, Van Der Walt E, Dent K, Varsani A, Rybicki EP (2010). Maize streak virus: an old and complex 'emerging' pathogen. *Molecular Plant Pathology* 11, 1-12.

Silva JCF, Carvalho TFM, Basso MF, *et al.* (2017). Geminivirus data warehouse: a database enriched with machine learning approaches. *BMC Bioinformatics* 18, 240.

Sunter G, Hartitz MD, Hormuzdi SG, Brough CL, Bisaro DM (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for eficient DNA replication. *Virology* 179, 69-77.

Vanitharani R, Chellappan P, Pita JS, Fauquet CM (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *Journal of Virology* 78, 9487-9498.

Voinnet O, Pinto YM, Baulcombe DC (1999). Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences, USA* 96, 14147-14152.

Walker PJ, Siddell SG, Lefkowitz EJ, *et al.* (2020). Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). *Archives of Virology* 165, 2737-2748.

Wang H, Buckley KJ, Yang X, Buchmann RC, Bisaro DM (2005). Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. *Journal of Virology* 79, 7410-7418.

Xavier CAD, Godinho MT, Mar TB, *et al.* (2021). Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis. *Molecular Ecology*, doi 10.1111-mec.15997.

Zerbini FM, Briddon RW, Idris A, *et al.* (2017). ICTV Virus Taxonomy Profile: *Geminiviridae*. *Journal of General Virology* 98, 131-133.

Zhou X (2013). Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* 51, 357-381.

**Table 1.** Begomovirus sequences reported in this study.

| Sample code | Isolate name | Sampling date | Location (District, Province) | Geographical coordinates | Host | Enzyme | | Clone | GenBank accession # |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | DNA-A | DNA-B | | |
| F152 | MZ-Lic152-19 | 27-Feb-19 | Lichinga, Niassa | 13°19.892'S, 35°15.211'E | Bean | BamHI | | MZ-Lic152-19 | MZ485484 |
| F153 | MZ-Lic153-19 | 27-Feb-19 | Lichinga, Niassa | 13°20.021'S, 35°15.086'E | Bean | BamHI | | MZ-Lic153-19 | MZ485485 |
| F154 | MZ-Lic154-19 | 27-Feb-19 | Lichinga, Niassa | 13°20.025'S, 35°15.091'E | Bean | | EcoRI | MZ-Lic154-19 | MZ485496 |
| NH81 | MZ-Gur81-19 | 22-Feb-19 | Guro, Manica | 17°26.833'S, 33°20.596'E | Cowpea | KpnI | | MZ-Gur81.5-19 | MZ485479 |
| | | | | | | | KpnI | MZ-Gur81.4-19 | MZ485486 |
| NH89 | MZ-Sus89-19 | 23-Feb-19 | Sussundenga, Manica | 19°19.697'S, 33°14,091'E | Cowpea | | KpnI | MZ-Sus89-19 | MZ485487 |
| NH90 | MZ-Sus90-19 | 23-Feb-19 | Sussundenga, Manica | 19°19.697'S, 33°14,088'E | Cowpea | | KpnI | MZ-Sus90.6-19 | MZ485488 |
| NH149 | MZ-Cum149-19 | 26-Feb-19 | Cumba, Niassa | 14°45.622'S, 36°32.453'E | Cowpea | PstI | | MZ-Cum149-19 | MZ485480 |
| NH163 | MZ-Met163-19 | 28-Feb-19 | Metarica, Niassa | 13°11.004'S, 36°56.491'E | Cowpea | HindIII | | MZ-Met163-19 | MZ485481 |
| S82 | MZ-Bar82-19 | 22-Feb-19 | Barue, Manica | 17°31.977'S, 33°16.998'E | Soybean | PstI | | MZ-Bar82-19 | MZ485482 |
| | | | | | | | HindIII | MZ-Bar82.1-19 | MZ485489 |
| S84 | MZ-Bar84-19 | 22-Feb-19 | Barue, Manica | 17°31.991'S, 33°17.010'E | Soybean | | KpnI | MZ-Bar84-19 | MZ485490 |
| S85 | MZ-Bar85-19 | 22-Feb-19 | Barue, Manica | 17°32.822'S, 33°16.479'E | Soybean | | EcoRI | MZ-Bar85-19 | MZ485491 |
| | | | | | | | KpnI | MZ-Bar85.9-19 | MZ485492 |
| S86 | MZ-Bar86-19 | 22-Feb-19 | Barue, Manica | 17°32.822'S, 33°14.126'E | Soybean | | EcorI | MZ-Bar86-19 | MZ485493 |
| | | | | | | | KpnI | MZ-Bar86.9-19 | MZ485494 |
| S164 | MZ-Mal164-19 | 01-Mar-19 | Malema, Nampula | 14°57.718'S, 37°23.429'E | Soybean | HindIII | | MZ-Mal164-19 | MZ485483 |
| | | | | | | | EcoRI | MZ-Mal164.1-19 | MZ485495 |

**Table 2.** Descriptors of the genetic variability of the population of Mozambique legume mosaic virus (MLMV). For comparison, the same descriptors are also presented for a population of the legume-infecting OW begomovirus mungbean yellow mosaic virus (MYMV).

| | L[#] | | N | | Hd | | π | |
|---|---|---|---|---|---|---|---|---|
| **Region[*]** | **MLMV** | **MYMV** | **MLMV** | **MYMV** | **MLMV** | **MYMV** | **MLMV** | **MYMV** |
| *CP* | 762 | 774 | 7 | 11 | 0.905 | 0.964 | 0.03143 | 0.02669 |
| *V2* | 339 | 300 | 7 | 11 | 0.905 | 0.964 | 0.00618 | 0.01897 |
| *Rep* | 1089 | 1089 | 7 | 11 | 0.905 | 0.982 | 0.01850 | 0.02975 |
| *TrAP* | 408 | 408 | 7 | 11 | 0.905 | 0.945 | 0.04272 | 0.02010 |
| *REn* | 405 | 405 | 7 | 11 | 0.905 | 1 | 0.04127 | 0.03012 |
| *AC4* | 300 | 264 | 7 | 11 | 0.810 | 0.891 | 0.00349 | 0.02355 |
| IR-A | 276 | 252 | 7 | 11 | 0.905 | 0.873 | 0.02174 | 0.04293 |
| *NSP* | 786 | 769 | 11 | 17 | 0.982 | 0.978 | 0.02188 | 0.02787 |
| *MP* | 792 | 799 | 11 | 17 | 0.964 | 0.993 | 0.04116 | 0.02413 |
| LIR-B | 890 | 903 | 11 | 17 | 0.982 | 1 | 0.02981 | 0.06320 |
| SIR-B | 39 | 31 | 11 | 17 | 0.855 | 0.838 | 0.06620 | 0.06262 |
| DNA-A | 2739 | 2724 | 7 | 11 | 0.905 | 1 | 0.02686 | 0.02956 |
| DNA-B | 2651 | 2588 | 11 | 15 | 1 | 1 | 0.03145 | 0.03526 |

[*] *CP*, coat protein; *Rep*, replication associated protein; *TrAP*, transactivating protein; *REn*: replication enhancer protein; IR-A, intergenic region of the DNA-A; *NSP*: nuclear shuttle protein; *MP*, movement protein; LIR-B, large intergenic region of the DNA-B; SIR, short intergenic region of the DNA-B

[#] L: median length in nucleotides; N: number of sequences analyzed; Hd: haplotype diversity; π: nucleotide diversity

**Table 3.** Results of neutrality tests for the *MP* and *NSP* genes of viral isolates comprising populations of the legume infecting, OW begomoviruses Mozambique legume mosaic virus (MLMV) and mungbean yellow mosaic virus (MYMV).

| Virus | Gene | Tajima's D | Fu and Li's D | Fu and Li's F | dN/dS |
|---|---|---|---|---|---|
| MLMV | *NSP* | -0.23399 ns* | -0.33632 ns | -0.35195 ns | 0.02188 |
| | *MP* | -0.59494 ns | -0.85579 ns | -0.89579 ns | 0.04116 |
| MYMV | *NSP* | -1.37484 ns | -1,67311 ns | -1.83964 ns | 0.02787 |
| | *MP* | -1.10027 ns | -1.39619 ns | -1.51920 ns | 0.02413 |

*n.s, non-significant

**Figure legends**

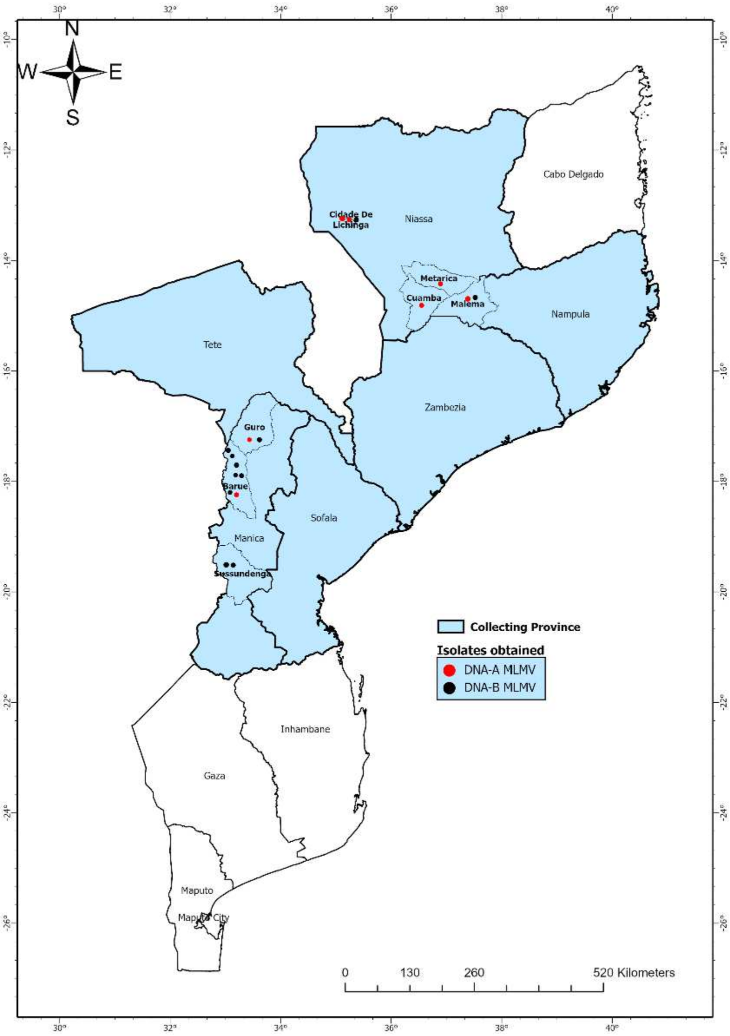
**Figure 1. A.** Map of Mozambique showing the provinces (highlighted in blue) where symptomatic common bean, cowpea and soybean samples were collected. The dots correspond to the locations where MLMV isolates were obtained. **B.** Symptoms in common bean, cowpea and soybean plants infected with MLMV. Left image, common bean sample F153, from which isolate MZ-Lic153-19 was obtained; center, cowpea sample NH90, from which isolate MZ-Sus90-19 was obtained; right, soybean sample S85, from which isolate MZ-Bar85-19 was obtained.

**Figure 2. A.** Nucleotide sequence alignment of the common regions (CRs) of three pairs of MLMV DNA-A and DNA-B components, each pair cloned from the same sample. The iterons (two imperfect direct repeats and one inverted repeat) are highlighted in red and underlined, the nonanucleotide is in bold, and the TATA box and start codon of the *Rep* gene are underlined in italics. **B.** Amino acid sequence alignment of the Rep protein N-terminal region of MLMV and the most closely related begomoviruses. The iteron-related domain (IRD) is indicated. Motifs I, II, III, GRS and α-helix2, required for specific dsDNA binding, protein conformation, DNA cleavage and initiation of rolling circle replication, are also indicated.

**Figure 3. A.** Midpoint-rooted Bayesian phylogenetic tree based on the full-length genomes (or DNA-A components) of Old World begomoviruses. Nodes with posterior probability values between 0.6 and 0.79 are indicated by grey circles, and nodes with values equal to or greater than 0.8 are indicated by black circles. The scale bar represents the number of nucleotide substitutions per site. **B.** Bayesian phylogenetic trees based on the full-length nucleotide sequences of DNA-B components of Old World begomoviruses. Nodes with posterior

probability values between 0.7 and 0.79 are indicated by grey circles, and nodes with values equal to or greater than 0.8 are indicated by black circles. The scale bar represents the number of nucleotide substitutions per site. See Suppl. Table S1 for full virus names.

**Figure 1A**


N
W
E
S
Cabo Delgado
Niassa
Cidade De Lichinga
Metarica
Cuamba
Malema
Nampula
Tete
Zambezia
Guro
Barue
Sofala
Manica
Sussundenga
Collecting Province
Isolates obtained
DNA-A MLMV
DNA-B MLMV
Inhambane
Gaza
Maputo
Maputo City
0
130
260
520 Kilometers

**B**

# Figure 2

## A

```
94%
DNA-A(MZ-Gur81.5-19)    GTGGCATTTTTGCAATTCTATCGTTTACACCGATAGCTCTTTCGCTTACACCGATTCATC   60
DNA-B(MZ-Gur81.4-19)    GTGGCATTTTTGCAATTCTGGCGTTTACACCGATTGCTCTCTCGCTTACACCGATCCATC   60

DNA-A(MZ-Gur81.5-19)    TATCGTATTATATCGGTGTAACGGGGTACCTAATATACTACTTACCCCAATTGATATCAG  120
DNA-B(MZ-Gur81.4-19)    TATCGTATTATATCGGTGTAACGGGGTACCTAATATACCTCTTACCCCAATTGATATCAG  120

DNA-A(MZ-Gur81.5-19)    CTTGACACGTGTCCCTAGCCCCACACCTAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180
DNA-B(MZ-Gur81.4-19)    CTTGACACCTGTCCCTAGCCCCACACCTAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180

DNA-A(MZ-Gur81.5-19)    CCGCGCTGTATCGGTGTACC  200
DNA-B(MZ-Gur81.4-19)    CCGCGCGCAATTGGTGTACC  200

96%
DNA-A(MZ-Bar82-19)      GTGGCATTTTTGCAATTCTATCGTTTACACCGATTGCTCTCTCGCTTACACCGATTCATC   60
DNA-B(MZ-Bar82.1-19)    GTGGCATTTTTGCAATTCTAGCGTTTACACTGATTGCTCTCTCGCTTACACCGATCCATC   60

DNA-A(MZ-Bar82-19)      TATCGTATTATATCGGTGTAACGGGGTACCTAATATACTACTTACCCCAATTGATATCAG  120
DNA-B(MZ-Bar82.1-19)    TATCGTATTATATCGGTGTAACGGGGTACCTAATATACCTCTTACCCCAATTGATATCAG  120

DNA-A(MZ-Bar82-19)      CTTGACACGTGTCCCTAGCCCCACTCCAAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180
DNA-B(MZ-Bar82.1-19)    CTTGACACCTGTCCCTAGCCCCACACCTAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180

DNA-A(MZ-Bar82-19)      CCGCGCGCAATCGGTGTACC  200
DNA-B(MZ-Bar82.1-19)    CCGCGCGCAATTGGTGTACC  200

97%
DNA-A(MZ-Mal164-19)     GTGGCATTTTTGCAATTCTATCGTTTACACCGATAGCTCTCTCGCTTACACCGATTCATC   60
DNA-B(MZ-Mal164.1-19)   GTGGCATTTTTGCAATTCTAGCGTTTACACCGATTGCTCTCTCGCTTACACCGATCCATC   60

DNA-A(MZ-Mal164-19)     TATCGTATTATATCGGTGTAACGGGGTACCTAATATACCTCTTACCCCAATTGATATCAG  120
DNA-B(MZ-Mal164.1-19)   TATCGTATTATATCGGTGTAACGGGGTACCTAATATACCTCTTACCCCAATTGATATCAG  120

DNA-A(MZ-Mal164-19)     CTTGACACCTGTCCCTAGCCCCACACCTAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180
DNA-B(MZ-Mal164.1-19)   CTTGACACCTGTCCCTAGCCCCACACCTAAAGCGGCCCTCAGTATAATATTACCTGAGGG  180

DNA-A(MZ-Mal164-19)     CCGCGCTGTATTGGTGTACC  200
DNA-B(MZ-Mal164.1-19)   CCGCGCGCAATTGGTGTACC  200
```

## B

```
                          IRD    MotifI                         α-helix2 MotifII
MYMV MN698275     MPRLGRFAINAKNYFLTYPRCPLTKEDVLEQLLALSTPVNKKFIRVCRELHDDGEPHLHV   60
ACMV J02058       M-RTPRFRIQAKNVFLTYPKCSIPKEHLLSFIQTLSLQSNPKFIKICRELHQNGEPHLHA   59
SbCBV KC508642    MPRSGAFRVHAKNIFLTYPRCSLSKDEALELLLGVTTPVNKKFIKVARELHEDGQPHLHV   60
CPGMV AF029217    MPRSGAFRVNAKNIFATYPRCSLPKEEALELLRQIPTAVNKKFIKVARELHEDGEPHLHV   60
MLMV MZ485481     MPRSGAFRVNAKNIFATYPRCTIPKEEALELLRQIPTAVNKKFIKVAQELHENGEPHLHV   60
MLMV MZ485483     MPRSGAFRVNAKNIFATYPRCTIPKEEALELLRQIPTAVNKKFIKVAQELHENGEPHLHV   60
MLMV MZ485484     MPRSGAFRVNAKNIFATYPRCTIPKEEALDLLRQIPTAVNKKFIKVAKELHENGEPHLHV   60
MLMV MZ485485     MPRSGAFRVNAKNIFATYPRCTIPKEEALDLLRQIPTAVNKKFIKVAKELHENGEPHLHV   60
MLMV MZ485482     MPRSGAFRVNAKNIFATYPRCTIPKEEALELLRQIPTAVNKKFIKVAKELHENGEPHLHV   60
MLMV MZ485479     MPRSGAFRVNAKNIFATYPRCTIPKEEALELLRQIPTAVNKKFIKVAKELHENGEPHLHV   60
MLMV MZ485480     MPRSGAFRVNAKNIFATYPRCTIPKEETLELLRQIPTAVNKKFIKVAKELHENGEPHLHV   60

                                      GRS                 MotifIII
MYMV MN698275     LLQFEGKLQTKNERFFDLVSPTRSAHYHPNVQAAKSASDVKSYMDKDGDVLDHGSFQVDG  120
ACMV J02058       LIQFEGKITITNNRLFDCVHPSCSTSFHPNIQGAKSSSDVKSYLDKDGDTVEWGQFQIDG  119
SbCBV KC508642    LLQFEGKLQITNPRLFDLVSRTSAHVFHPNVQGAKSSSDVKSYIDKDGDTVSWGEFQIDA  120
CPGMV AF029217    LLQFEGKLQITNPRLFDLVSGNRSAVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485481     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485483     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485484     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485485     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485482     LLQFEGKLQITNPRLFDLVYGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485479     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
MLMV MZ485480     LLQFEGKLQITNPRLFDLVSGCRAQVFHPNIQGAKSSSDVKSYIEKDGDTISWGEFQIDG  120
```

**Figure 3A**

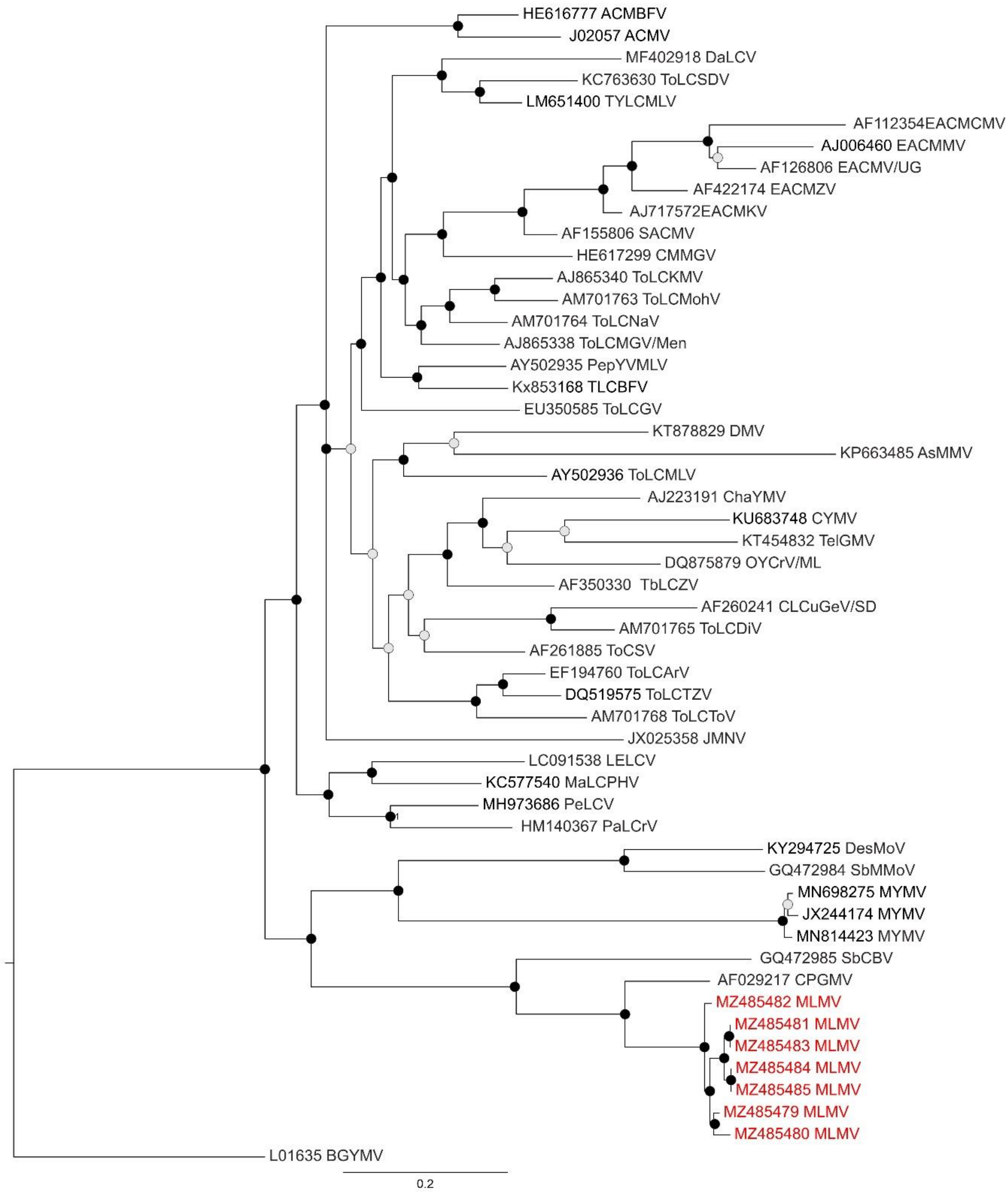
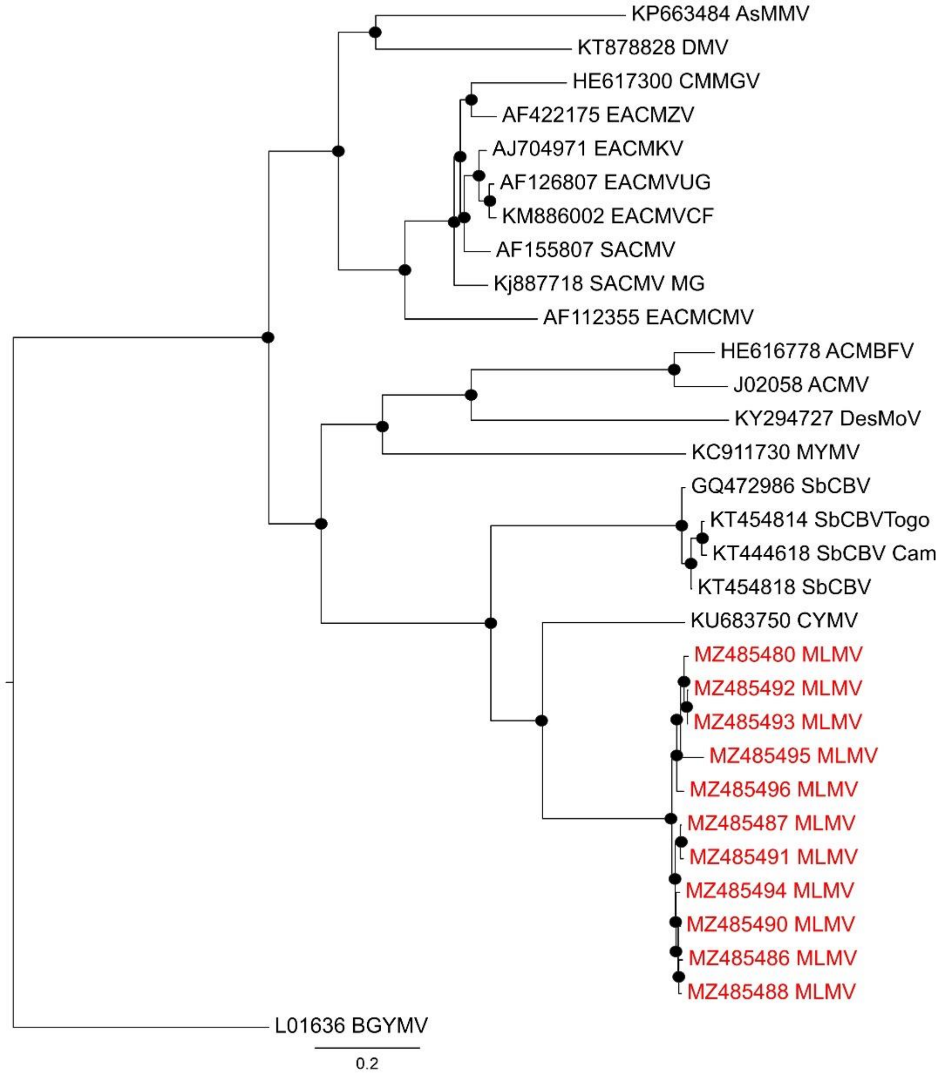

HE616777 ACMBFV
J02057 ACMV
MF402918 DaLCV
KC763630 ToLCSDV
LM651400 TYLCMLV
AF112354EACMCMV
AJ006460 EACMMV
AF126806 EACMV/UG
AF422174 EACMZV
AJ717572EACMKV
AF155806 SACMV
HE617299 CMMGV
AJ865340 ToLCKMV
AM701763 ToLCMohV
AM701764 ToLCNaV
AJ865338 ToLCMGV/Men
AY502935 PepYVMLV
Kx853168 TLCBFV
EU350585 ToLCGV
KT878829 DMV
KP663485 AsMMV
AY502936 ToLCMLV
AJ223191 ChaYMV
KU683748 CYMV
KT454832 TelGMV
DQ875879 OYCrV/ML
AF350330 TbLCZV
AF260241 CLCuGeV/SD
AM701765 ToLCDiV
AF261885 ToCSV
EF194760 ToLCArV
DQ519575 ToLCTZV
AM701768 ToLCToV
JX025358 JMNV
LC091538 LELCV
KC577540 MaLCPHV
MH973686 PeLCV
HM140367 PaLCrV
KY294725 DesMoV
GQ472984 SbMMoV
MN698275 MYMV
JX244174 MYMV
MN814423 MYMV
GQ472985 SbCBV
AF029217 CPGMV
MZ485482 MLMV
MZ485481 MLMV
MZ485483 MLMV
MZ485484 MLMV
MZ485485 MLMV
MZ485479 MLMV
MZ485480 MLMV
L01635 BGYMV
0.2

**Figure 3B**


KP663484 AsMMV
KT878828 DMV
HE617300 CMMGV
AF422175 EACMZV
AJ704971 EACMKV
AF126807 EACMVUG
KM886002 EACMVCF
AF155807 SACMV
Kj887718 SACMV MG
AF112355 EACMCMV
HE616778 ACMBFV
J02058 ACMV
KY294727 DesMoV
KC911730 MYMV
GQ472986 SbCBV
KT454814 SbCBVTogo
KT444618 SbCBV Cam
KT454818 SbCBV
KU683750 CYMV
MZ485480 MLMV
MZ485492 MLMV
MZ485493 MLMV
MZ485495 MLMV
MZ485496 MLMV
MZ485487 MLMV
MZ485491 MLMV
MZ485494 MLMV
MZ485490 MLMV
MZ485486 MLMV
MZ485488 MLMV
L01636 BGYMV
0.2

**Supplementary Table S1.** Begomovirus sequences retrieved from GenBank.

| DNA-A<br>Species | Virus name | GenBank Accession # | Acronym |
|---|---|---|---|
| *African cassava mosaic Burkina Faso virus* | African cassava mosaic Burkina Faso virus | HE616777 | ACMBFV |
| *African cassava mosaic virus* | African cassava mosaic virus | J02057 | ACMV |
| *Asystasia mosaic Madagascar virus* | Asystasia mosaic Madagascar virus | KP663485 | AsMMV |
| *Bean golden yellow mosaic virus* | Bean golden yellow mosaic virus | L01635 | BGYMV |
| *Chayote yellow mosaic virus* | Chayote yellow mosaic virus | AJ223191 | ChaYMV |
| *Cassava mosaic Madagascar virus* | Cassava mosaic Madagascar virus | HE617299 | CMMGV |
| *Cotton leaf curl Gezira virus* | Cotton leaf curl Gezira virus | AF260241 | CLCuGeV |
| *Cotton yellow mosaic virus* | Cotton yellow mosaic virus | KU683748 | CYMV |
| *Cowpea golden mosaic virus* | Cowpea golden mosaic virus | AF029217 | CPGMV |
| *Datura leaf curl virus* | Datura leaf curl virus | MF402918 | DaLCV |
| *Desmodium mottle virus* | Desmodium mottle virus | KY294725 | DesMoV |
| *Deinbollia mosaic virus* | Deinbollia mosaic virus | KT878829 | DMV |
| *East African cassava mosaic Cameroon virus* | East African cassava mosaic Cameroon virus | AF112354 | EACMCMV |
| *East African cassava mosaic virus* | East African cassava mosaic virus | AF126806 | EACMV-[UG] |
| *East African cassava mosaic Malawi virus* | East African cassava mosaic Malawi virus | AJ006460 | EACMMV |
| *East African cassava mosaic Zanzibar virus* | East African cassava mosaic Zanzibar virus | AF422174 | EACMZV |
| *East African cassava mosaic Kenya virus* | East African cassava mosaic Kenya virus | AJ717572 | EACMKV |
| *Lisianthus enation leaf curl virus* | Lisianthus enation leaf curl virus | LC091538 | LELCV |
| *Jatropha mosaic Nigeria virus* | Jatropha mosaic Nigeria virus | JX025358 | JMNV |
| *Malvastrum leaf curl Philippines virus* | Malvastrum leaf curl Philippines virus | KC577540 | MaLCPHV |
| *Mungbean yellow mosaic virus* | Mungbean yellow mosaic virus | MN698275 | MYMV |
| *Mungbean yellow mosaic virus* | Mungbean yellow mosaic virus | MN814423 | MYMV |
| *Mungbean yellow mosaic virus* | Mungbean yellow mosaic virus | JX244174 | MYMV |
| *Okra yellow crinkle virus* | Okra yellow crinkle virus | DQ875879 | OYCrV |
| *Papaya leaf crumple virus* | Papaya leaf crumple virus | HM140367 | PaLCrV |
| *Pedilanthus leaf curl virus* | Pedilenthus leaf curl virus | MH973686 | PeLCV |
| *Pepper yellow vein Mali virus* | Pepper yellow vein Mali virus | AY502935 | PepYVMLV |
| *South African cassava mosaic virus* | South African cassava mosaic virus | AF155806 | SACMV |
| *Soybean chlorotic blotch virus* | Soybean chlorotic blotch virus | GQ472985 | SbCBV |
| *Soybean mild mottle virus* | Soybean mild mottle virus | GQ472984 | SbMMoV |
| *Telfairia golden mosaic virus* | Telfairia golden mosaic virus | KT454832 | TelGMV |
| *Tobacco leaf curl Zimbabwe virus* | Tobacco leaf curl Zimbabwe virus | AF350330 | TbLCZV |
| *Tomato curly stunt virus* | Tomato curly stunt virus | AF261885 | ToCSV |
| *Tomato leaf curl Arusha virus* | Tomato leaf curl Arusha virus | EF194760 | ToLCArV |
| *Tomato leaf curl Diana virus* | Tomato leaf curl Diana virus | AM701765 | ToLCDiV |
| *Tomato leaf curl Ghana virus* | Tomato leaf curl Ghana virus | EU350585 | ToLCGV |
| *Tomato leaf curl Comoros virus* | Tomato leaf curl Comoros virus | AJ865340 | ToLCKMV |
| *Tomato leaf curl Madagascar virus* | Tomato leaf curl Madagascar virus | AJ865338 | ToLCMGV |

| | | | |
|---|---|---|---|
| *Tomato leaf curl Namakely virus* | Tomato leaf curl Namakely virus | AM701764 | ToLCNaV |
| *Tomato leaf curl Sudan virus* | Tomato leaf curl Sudan virus | KC763630 | ToLCSDV |
| *Tomato leaf curl Tanzania virus* | Tomato leaf curl Tanzania virus | DQ519575 | ToLCTZV |
| *Tomato leaf curl Toliara virus* | Tomato leaf curl Toliara virus | AM701768 | ToLCToV |
| *Tomato leaf curl Moheli virus* | Tomato leaf curl Moheli virus | AM701763 | ToLCMohV |
| *Tomato yellow leaf curl Mali virus* | Tomato yellow leaf curl Mali virus | LM651400 | TYLCMLV |
| *Tomato leaf curl Burkina Faso virus* | Tomato leaf curl Burkina Faso virus | KX853168 | TLCBFV |
| *Tomato leaf curl Mali virus* | Tomato leaf curl Mali virus | AY502936 | ToLCMLV |

**DNA-B**

| **Species** | **Virus name** | **GenBank Accession #** | **Acronym** |
|---|---|---|---|
| *African cassava mosaic Burkina Faso virus* | African cassava mosaic Burkina Faso virus | HE616778 | ACMBFV |
| *African cassava mosaic virus* | African cassava mosaic virus | J02058 | ACMV |
| *Asystasia mosaic Madagascar virus* | Asystasia mosaic Madagascar virus | KP663484 | AsMMV |
| *Bean golden yellow mosaic virus* | Bean golden yellow mosaic virus | L01636 | BGYMV |
| *Cassava mosaic Madagascar virus* | Cassava mosaic Madagascar virus | HE617300 | CMMGV |
| *Cotton yellow mosaic virus* | Cotton yellow mosaic virus | KU683750 | CYMV |
| *Desmodium mottle virus* | Desmodium mottle virus | KY294727 | DesMoV |
| *Deinbollia mosaic virus* | Deinbollia mosaic virus | KT878828 | DMV |
| *East African cassava mosaic Cameroon virus* | East African cassava mosaic Cameroon virus | AF112355 | EACMCMV |
| *East African cassava mosaic virus* | East African cassava mosaic virus | AF126807 | EACMV-[UG] |
| *East African cassava mosaic Zanzibar virus* | East African cassava mosaic Zanzibar virus | AF422175 | EACMZV |
| *East African cassava mosaic Kenya virus* | East African cassava mosaic Kenya virus | AJ704971 | EACMKV |
| *Soybean chlorotic blotch virus* | Soybean chlorotic blotch virus | GQ472986 | SbCBV |
| *South African cassava mosaic virus* | South African cassava mosaic virus | AF155807 | SACMV |
| *Soybean chlorotic blotch virus* | Soybean chlorotic blotch virus | KT454818 | SbCBV |
| *Mungbean yellow mosaic virus* | Mungbean yellow mosaic virus | AJ132574 | MYMV |
| *Mungbean yellow mosaic viru* | Mungbean yellow mosaic viru | KC911730 | MYMV |
| *East African cassava mosaic virus* | East African cassava mosaic virus | KM886002 | EACMV |
| *South African cassava mosaic virus* | South African cassava mosaic virus | KJ887718 | SACMV |

**Supplementary Table S2.** Recombination events detected in the genomes of begomovirus isolates from Mozambique.

| Event | Recombinant sequence (DNA-B) | Breakpoint positions Begin | End | Minor parent | Major parent | Detection method* | *p*-value# |
|---|---|---|---|---|---|---|---|
| 1 | MZ-Bar85.9-19<br>MZ-Bar82.1-19<br>MZ-Bar86-19<br>MZ-Bar85.9-19 | 984 | 1391 | MZ-Mal164.1-19 | MZ-Bar86.9-19 | RGBMCS**<u>3</u>** | $1.269x10^{-24}$ |
| 2 | MZ-Lic154-19 | 1266 | 2258 | MZ-Sus89-19<br>MZ-Gur81.4-19<br>MZ-Sus90.6-19<br>MZ-Bar84-19<br>MZ-Bar85-19<br>MZ-Bar86.9-19 | MZ-Mal164.1-19 | RGBM**<u>S</u>**3 | $1.269x\ 10^{-24}$ |
| 3 | MZ-Mal164.1-19 | 1388 | 2364 | Unknown (MZ-Bar85-19) | MZ-Bar85.9-19 | RGBMCS**<u>3</u>** | $1.007x\ 10^{-22}$ |

* R, RDP; G, GENECONV; B, BootScan; M, MaxChi; C, Chimaera; S, Siscan; 3, 3Seq.

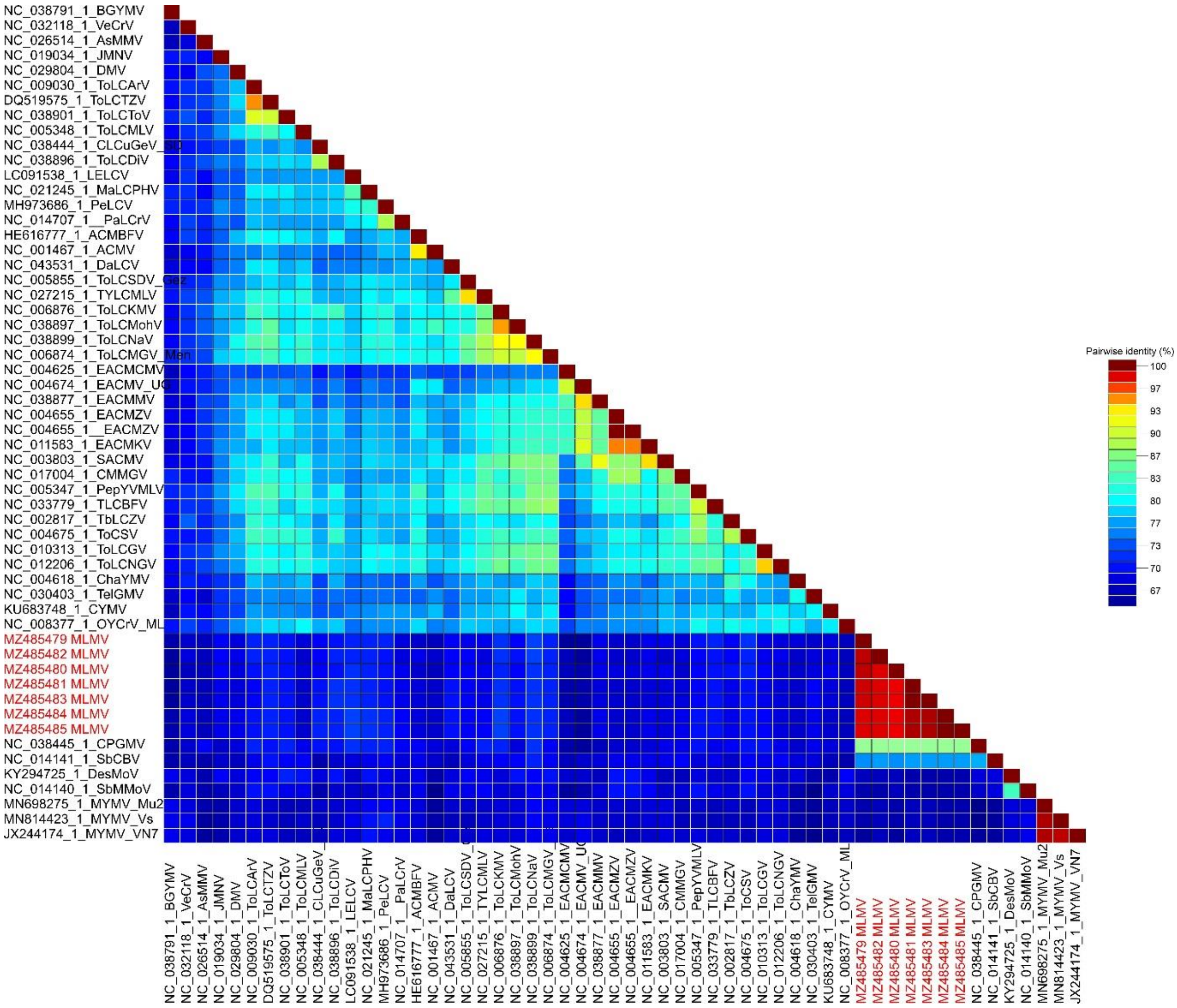
# The *p*-value informed refers to the bold, underlined detection method and is the lowest calculated *p*-value for the region in question.

**Suplementary Table S3.** Descriptors of genetic variability for legume-infecting begomoviruses from the Old World.

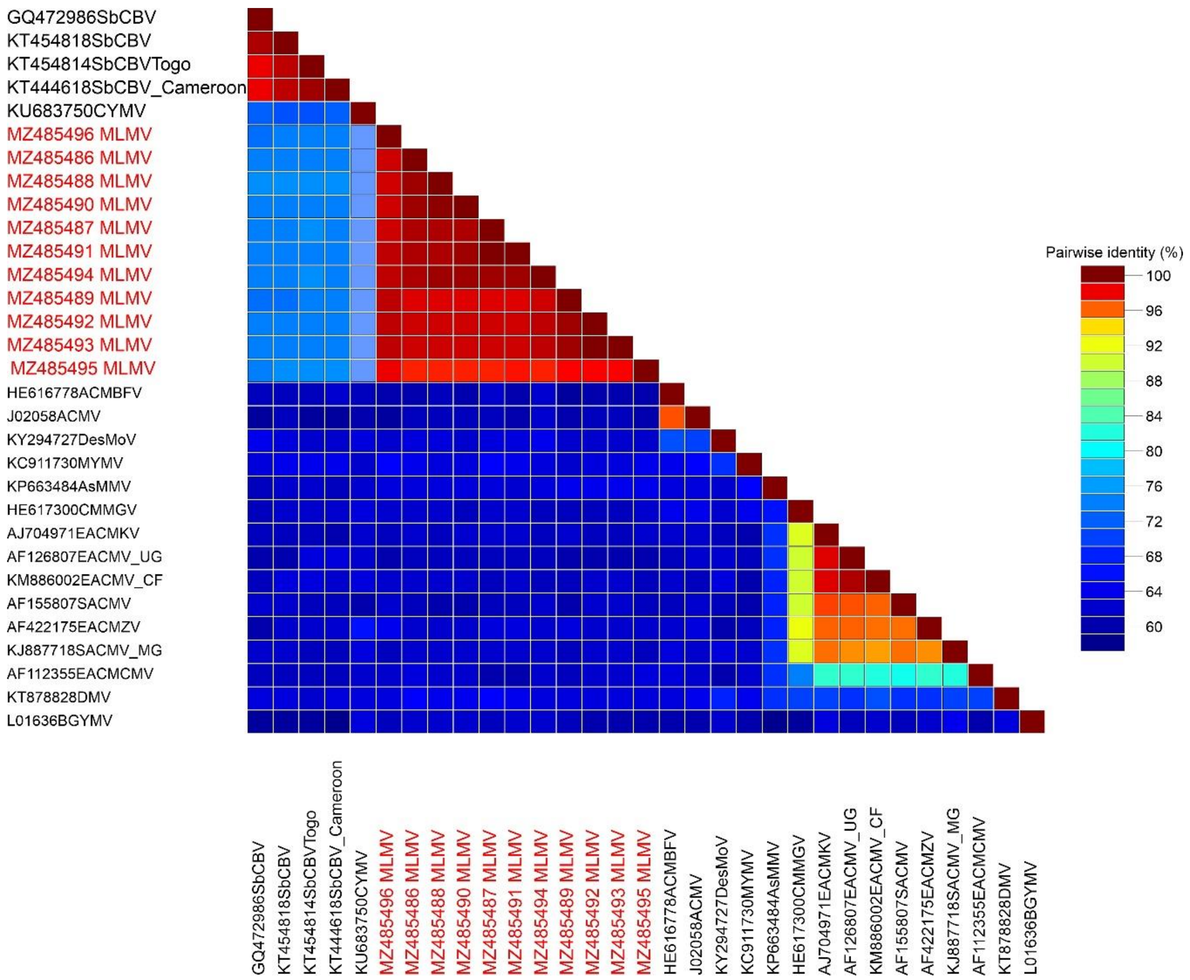
| DNA-A | L | N | H | Hd | π |
|---|---|---|---|---|---|
| MLMV | 2739 | 7 | 5 | 0.905 | 0.02686 |
| SbCBV | 2658 | 11 | 11 | 1 | 0.01762 |
| MYMV | 2724 | 11 | 11 | 1 | 0.02956 |
| **DNA-B** | | | | | |
| MLMV | 2651 | 11 | 11 | 1 | 0.03145 |
| SbCBV | 2275 | 17 | 17 | 1 | 0.02209 |
| MYMV | 2588 | 15 | 15 | 1 | 0.03526 |
| EACMV | 2719 | 31 | 31 | 1 | 0.04115 |
| ACMV | 2693 | 19 | 19 | 1 | 0.06151 |

N: number of sequences; L: median genome length in nucleotides; H: number of haplotypes; Hd: haplotype diversity; π: nucleotide diversity

**Supplementary Figure S1.** Pairwise nucleotide sequence identity matrix of the complete genome or the DNA-A component of Old World begomoviruses, calculated using SDT v. 1.2 from a multiple sequence alignment prepared using the MUSCLE algorithm implemented in MEGA .



**Supplementary Figure S2.** Pairwise nucleotide sequence identity matrix of the DNA-B component of Old World begomoviruses, calculated using SDT v. 1.2 from a multiple sequence alignment prepared using the MUSCLE algorithm implemented in MEGA X.



# CONCLUSIONS

We identified and characterized the complete genome of two new bipartite begomoviruses that infect weeds and legume crops, and which so far have only been found in Mozambique.

Pyrenacantha yellow mosaic virus (PyYMV; *Begomovirus pyrenacanthae*) is a new Old World, bipartite begomovirus that naturally infects *Pyrenacantha* sp. in Mozambique. The presence of a 35-nt insertion in the intergenic (common) region of its DNA-A indicates the occurrence of an ancient recombination event, as recombination analysis did not detect a recombination signal and the 35-nt sequence has no significant identity to any other sequence in the databases. *Pyrenacantha* sp. is identified for the first time as a host of a begomoviruse, worldwide.

Mozambique legume mosaic virus (MLMV; *Begomovirus mozlegume*) is an Old World, bipartite begomovirus emerging in legume crops (common bean, cowpea and soybean) in Mozambique. Its adaptation to new hosts seems to be driven by recombination acting on the DNA-B. The increase in nucleotide diversity in some genes may be evidence of evolution by adaptation to the environment (new hosts). MLMV is the first begomovirus described in common bean in Africa, and only the second and third described in cowpea and soybean, respectively.

Overall, the results of this work highlight the need for a survey of viral diversity in cultivated and non-cultivated plants in Mozambique.

# PERSPECTIVES

The development of disease control and management strategies depends primarily on prior knowledge of the pathogen and its molecular and biological characteristics. However, because agriculture in Mozambique is mostly rainfed, with a long period between crops, crop rotation and intercropping, damage caused by plant pathogens can go unnoticed and be underestimated. Our results provide evidence of the need for extensive, intensive and continuous research.

The identification of two new species of begomoviruses in weeds and legumes opens up a range of questions that need to be answered. We intend to address these questions soon, and suggest the use of HTS in future research to allow detection of both known and unknown viruses.

Studies on the genetic variability of the newly identified begomoviruses, the epidemiological situation and implications for agricultural productivity and economic losses must be carried out. We propose to the Ministry of Agriculture and Rural Development, together with the Ministry of Science and Technology of Mozambique, to redouble efforts in this research area, taking into account the relevance of diseases for agriculture.