Maf/ham1-like pyrophosphatases of non-canonical nucleotides are host-specific partners of viral RNA-dependent RNA polymerases

Adrian A. Valli[1, *], Rafael García López[1], María Ribaya[1, ¶], Francisco Javier Martínez[1], Diego García Gómez[2], Beatriz García[1], Irene Gonzalo[1], Alfonso Gonzalez de Prádena[1], Fabio Pasin[1, §], Inmaculada Montanuy[3], Encarnación Rodríguez-Gonzalo[2], Juan Antonio García[1].

[1]Centro Nacional de Biotecnología (CNB-CSIC), Madrid, 28049, Spain

[2]Departamento de Química Analítica, Nutrición y Bromatología, Universidad de Salamanca, Salamanca, 37008, Spain

[3]Facultad de Ciencias Experimentales, Universidad Francisco de Vitoria, Madrid, 28223, Spain

[*]Correspondence: Adrian A. Valli, avalli@cnb.csic.es

Current addresses

[¶]Centre for Research in Agricultural Genomics (CRAG-CSIC-IRTA-UAB-UB), Cerdanyola del Vallès, Barcelona, 08193, Spain.

[§]Instituto de Biología Molecular y Celular de Plantas (IBMCP, UPV-CSIC), Valencia, 46022, Spain

Abstract

Cassava brown streak disease (CBSD), dubbed the "Ebola of plants", is a serious threat to food security in Africa caused by two viruses of the family *Potyviridae*: cassava brown streak virus (CBSV) and Ugandan (U)CBSV. Intriguingly, U/CBSV, along with another member of this family and one secoviridae, are the only known RNA viruses encoding a protein of the Maf/ham1-like family, a group of widespread pyrophosphatase of non-canonical nucleotides (ITPase) expressed by all living organisms. Despite the socio-economic impact of CDSD, the relevance and role of this atypical viral factor has not been yet established. Here, using an infectious cDNA clone and reverse genetics, we demonstrate that UCBSV requires the ITPase activity for infectivity in cassava, but not in the model plant *Nicotiana benthamiana*. HPLC-MS/MS experiments showed that, quite likely, this host-specific constraint is due to an unexpected high concentration of non-canonical nucleotides in cassava. Finally, protein analyses and experimental evolution of mutant viruses indicated that keeping a fraction of the yielded UCBSV ITPase covalently bound to the viral RNA-dependent RNA polymerase (RdRP) optimizes viral fitness, and this seems to be a feature shared by the other members of the *Potyviridae* family expressing Maf/ham1-like proteins. All in all, our work (i) reveals that the over-accumulation of non-canonical nucleotides in the host might have a key role in antiviral defense, and (ii) provides the first example of an RdRP-ITPase partnership, reinforcing the idea that RNA viruses are incredibly versatile at adaptation to different host setups.

Keywords

RNA virus; virus/host coevolution; RNA-dependent RNA polymerase; RdRP; plant defence; ITP; XTP; Euphorbiaceae; *Potyviridae*; Ipomovirus.

Introduction

The family *Potyviridae* is the largest and most socio-economically relevant group of plant-infecting RNA viruses. With more than 200 assigned members sorted in 12 different genera, these viruses represent a major threat for basically every important crop on earth. Potyvirids (members of the family *Potyviridae*) share common features, such as (i) monopartite (except for a few bipartite viruses) and positive sense single-stranded RNA (+ssRNA) genome, (ii) transmission mediated by vectors, and (iii) picorna-like gene expression strategy based on large polyproteins further processed by

viral-encoded proteinases (Revers and García, 2015, Yang et al., 2021, Valli et al., 2021). Potyvirids, in most cases, produce 10 mature proteins: P1, HCPro, P3, P3N-PIPO, 6K1, CI, 6K2, NIa (VPg/NIaPro), NIb and CP. Of relevance to this study, NIaPro is a *cis*- and *trans*-acting proteinase that releases most of the mature factors from the polyprotein (Carrington and Dougherty, 1987a, Carrington and Dougherty, 1987b), and NIb is a RNA-dependent RNA polymerase (RdRP) that replicates the viral genome (Allison et al., 1986, Hong and Hunt, 1996).

With seven members described so far, the *Ipomovirus* genus is the most versatile group of potyvirids in term of genome organization, since only two of them follow the most common arrangement mentioned above (Dombrovsky et al., 2014). The remaining five ipomoviruses lack the HCPro coding region and express either one P1 proteinase or two P1s in tandem (Valli et al., 2006, Valli et al., 2007, Mbanzibwa et al., 2009). Two viruses infecting *Manihot esculenta* (cassava) in nature are classified in this genus: cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV), which cause the devastating cassava brown streak disease (CBSD), also dubbed the "Ebola of plants" (Patil et al., 2015, Tomlinson et al., 2018). Indeed, CBSD is considered among the seven most detrimental plant diseases in the world for its impact on the economy and food security in Africa, where it causes about 750 million US$ annual losses just in Tanzania, Uganda, Kenya and Malawi (Pennisi, 2010, Hillocks and Maruthi, 2015).

Even though CBSV and UCBSV are two distinct viral species, their genomes share around 72% nucleotide sequence identity, just below the species demarcation criteria in potyvirids (76%) (Winter et al., 2010). Moreover, these two viruses (i) encode a single P1 leader proteinase, (ii) lack HCPro and, as the most striking feature, (iii) present an extra cistron between NIb and CP that encodes a *bona fide* Maf/Ham1-like protein (Mbanzibwa et al., 2009). This protein (referred as HAM1 in this study) belongs to the inosine triphosphate (ITP) pyrophosphatase (ITPase) family, which hydrolyzes the pyrophosphate bonds in triphosphate substrates (ITP/XTP) to release the corresponding monophosphate (IMP/XMP) and a pyrophosphate molecule (Hwang et al., 1999, Lin et al., 2001, Chung et al., 2001, Chung et al., 2002). The presence of putative cleavage sites for the NIaPro proteinase at the N- and C-termini of HAM1 suggested that this protein accumulates into infected cells as a free product (Mbanzibwa et al., 2009).

HAM1-like enzymes are present in prokaryotes and eukaryotes, across all life kingdoms, where they are proposed to prevent (i) incorporation of non-canonical

nucleotides into nascent DNA and RNA molecules, (ii) RNA mistranslation, and (iii) inhibition of ATP-dependent enzymes (Simone et al., 2013). Although they are widespread in nature, HAM1-like proteins are not usually encoded in viral genomes; in fact, their presence has been reported in only four RNA viruses so far. Intriguingly, all these HAM1-expressing RNA viruses infect plants from the *Euphorbiaceae* family: three potyvirids [CBSV, UCSBV and euphorbia ringspot virus (EuRV, *Potyvirus* genus)] (Mbanzibwa et al., 2009, Knierim et al., 2017), and one virus from the *Secoviridae* family [cassava torrado-like virus (CsTLV)] (Jiménez Polo et al., 2018). Even though a recent study has shown that CBSV and UCBS HAM1s are genuine pyrophosphatases in *in vitro* experiments, and that they determine necrotic symptoms in the model plant *Nicotiana benthamiana* (Tomlinson et al., 2019), relevance and defined role of viral-derived HAM1 proteins are still unknown.

In this study, among other approaches, we used reverse genetics to manipulate an infectious cDNA clone of UCBSV in order to gain insight about the role of RNA virus-derived HAM1 proteins. Briefly, our experiments revealed that: (i) HAM1 is required for the virus to infect cassava, but not to produce a successful infection in the model plant *N. benthamiana*, and (ii) it works in partnership with the viral RdRP. The extremely high levels of non-canonical nucleotides that we have found in cassava, and likely present in other *Euphorbiaceae* plants, should have worked as a strong selection pressure to promote the acquisition of an ITP/XTP pyrophosphatase activity into virus RdRP in order to support successful replication and infection.

Material and Methods

Plants. Cassava plants were grown in a chamber with 16h/8h light/dark cycles at 28°C. *N. benthamiana* plants were grown in a greenhouse with 16h/8h light/dark cycles at 20-to-24°C with supplementary light. For viral infection, *N. benthamiana* plants were moved just after inoculation to the cassava-growing chamber.

Plasmids. Oligonucleotides used for this study are listed in Supplementary table S1. UCBSV full-length clones derive from pLX-UCBSVi, a version of pLX-UCBSV (GenBank KY825157.1) (Pasin et al., 2017) that carries the second intron of *Solanum tuberosum* ST-LS1 gene to interrupt the UCBSV P3 cistron. To generate pLX-UCBSVi, the mentioned intron was first amplified by PCR from pIC-PPV (Lopez-Moya and Garcia, 2000) with primers #3257/#3258. The 3'-half part of the UCBSV P3 cistron

was amplified by PCR with primers #3259/#3260. An overlapping PCR with primers #3257/#3260 was used to join these two PCR products [intron-P3(3´half)]. UCBSV P1 and the 5'-half part of the UCBSV P3 cistron [P1-P3(5'-half)] were amplified with primers #3255/#3256. Finally, a DNA fragment that carries P1-P3(5'-half)-intron-P3(3´half) was produced by overlapping PCR with primers #3255/#3260, using P1-P3(5'-half) and intron-P3(3´half) as templates. This PCR product was digested with Bsu36I and NheI and introduced by ligation in pLX-UCBSV, which had been digested with the same enzymes, to replace the equivalent intron-less DNA segment.

To generate pLX-UCBSVi-eGFP (a GFP-tagged version of UCBSV), pLX-UCBSVi was used as backbone to introduce the GFP coding sequence between the HAM1 and CP cistrons. To allow the release of GFP during the infection, its coding sequence was flanked at both sides by synonymous sequences encoding the NIaPro cleavage site located between HAM1 and CP (LTIDVQ/A). First, eGFP (F64L, S65A, V163A) coding sequence was amplified by PCR with primers #3360/#3361, adding the coding sequence of NIaPro cleavage site in the reverse primer, by using P1P1b clone (Carbonell et al., 2012) as template. Then, the N-terminus of NIb and the whole HAM1 coding sequences were amplified by PCR with primers #3160/#3358, adding the coding sequence of the NIaPro cleavage site in the reverse primer, by using pLX-UCBSVi as template. A subsequent overlapping PCR with primers #3160//#3361 was used to join the two above-mentioned PCR products into one single DNA segment. Finally, a BstBI/StuI fragment (the last 25 nt from NIb, the whole HAM1 and 1 nt from CP) from pLX-UCBSVi was replaced by the larger PCR product digested with BstBI.

To generate a 2xMyc-tagged version of HAM1 in UCSBV, pLX-UCBSVi was used as backbone to introduce the 2xMYC (GLING<u>EQKLISEEDL</u>NG<u>EQKLISEEDL</u>) coding sequence just upstream the coding sequence that corresponds to the NIaPro cleavage site located between HAM1 and CP. First, the N-terminus of NIb and most of HAM1 coding sequences were PCR amplified with primers #3160/#3162, adding the coding sequence of 1xMyc in the reverse primer, by using pLX-UCBSVi as template. Then, a second PCR with primers #3160/#3163, adding the coding sequence of another 1xMyc and the NIaPro cleavage site (LTIDVQ/) in the reverse primer, was carried out by using the first PCR product as template. Finally, a BstBI/StuI fragment (the last 25 nt from NIb, the whole HAM1 and 1 nt from CP) from pLX-UCBSVi was replaced by the second PCR product digested with BstBI/StuI to generate pLX-UCBSVi-2xMyc.

Mutagenesis of HAM1 in both pLX-UCBSVi and pLX-UCBSVi-HAM1-2xMyc backbones was done by using a previously described method (Ho et al., 1989). In brief, two PCR products having overlapping ends, which carry the desired mutation, were used as template of a subsequent PCR to join both PCR products in a larger DNA fragment. Then, a BstBI/StuI fragment (the last 25 nt from NIb, the whole HAM1 and 1 nt from CP) in the corresponding backbone was replaced by the indicated PCR products digested with the same enzymes. A list of pLX-UCBSVi- and pLX-UCBSVi-HAM1-2xMyc-derivatives, as well as the name of primers used for the amplification of different inserts, are listed in Supplementary table S2.

The plasmid that expresses UCBSV-HAM1$_{T1A/D3G}$-2xMyc, a double mutant that carries T1A and D3G mutations in HAM1, was generated by replacing the BstBI/StuI fragment in pLX-UCBSVi-HAM1-2xMyc with the RT-PCR product amplified with primers #3160/#3130 from RNA of a cassava plant originally infected with UCBSV-HAM1$_{T1A}$-2xMyc after its digestion with the same restriction enzymes.

The plasmid pLX-UCBSVi-ΔHAM1, which has a full deletion of HAM1 cistron, was built by replacing the above-mentioned BstBI/StuI fragment in pLX-UCBSVi with a compatible end, short, double-stranded DNA fragment created by the annealing of oligonucleotides #3312/#3313.

Plasmids for transient expression of viral proteins in *N. benthamiana* leaves were built by the Gateway technology (Invitrogen) using pENTR1A as entry vector, and either pGWB702Ω (35S promoter, TMV 5'UTR, no tag, NOS terminator) or pGWB718 (35S promoter, 4xMyc tag for N-terminal fusion, NOS terminator) (Tanaka et al., 2011) as expression vectors. Briefly, cDNA fragments encoding NIa and NIb$_C$-HAM1-CP$_N$ from UCBSV, CBSV and EuRV were amplified by PCR and directly introduced into pENTR1A previously digested with XmnI/EcoRV (name of primers and templates used for each PCR are indicated in Supplementary table S3). The correctness of pENTR1A derivatives was confirmed by digestion with restriction enzymes and Sanger sequencing by Macrogen Europe. Then, those cDNAs were moved from pENTR1A derivatives to either pGWB702Ω (NIa) or pGWB718 (NIb$_C$-HAM1-CP$_N$) by LR recombination.

Alignment of primary amino acid sequences and 3D protein modeling. The primary amino acid sequences of the following HAM1 proteins were downloaded from NCBI: human ITPA (NP_258412.1), *E. coli* RdgB (NP_417429.1), yeast HAM1 (NP_012603.1), arabidopsis HAM1-like protein (NP_567410.1), and viral HAM1-like

proteins from CBSV (ACS71538.1), UCBSV (ASG92166.1) and EuRV (YP_009310049.1). Protein sequences were aligned with Clustal Omega from EMBL-EBI (Madeira et al., 2019) with default parameters, and results were visualized/colored with Jalview version 2.11.1.4 (Waterhouse et al., 2009). The tridimensional structure of UCBSV HAM1 bound to ITP was modeled by homology using the SWISS-MODEL server (Waterhouse et al., 2018).

Virus inoculation. Inoculation of UCBSV full-length clones (wild type and derivatives) was carried out by biolistic with the Helios Gene Gun System (Bio-Rad) by following a previously described protocol (Salvador et al., 2008). Helium pressures of 7 and 13 bar were used to inoculate *N. benthamiana* and cassava, respectively. Serial passages were done by manual inoculation of plants with sap extracts from infected plants as viral source. To do that, infected leaves were ground in a buffer containing 150 mM NaCl, 2.5 mM DTT and 50 mM Tris-HCl pH 7.5 (2ml/mg) with an ice-cold mortar and pestle, and the sap was finger-rubbed onto two leaves of plants that had previously been dusted with Carborundum.

Fluorescence imaging. GFP fluorescence was observed with an epifluorescence stereomicroscope using excitation and barrier filters at 470/40 nm and 525/50 nm, respectively, and photographed with an Olympus DP70 digital camera.

Transient expression by agro-infiltration. Two young leaves of 1-month-old *N. benthamiana* plants were infiltrated with *Agrobacterium tumefaciens* strain C58C1 carrying the indicated plasmids, as previously described (Valli et al., 2006). To boost protein expression, the potent silencing suppressor P14 from photos latent virus (Merai et al., 2005) was co-express along with the proteins of interest.

Immunodetection of proteins by western blot. The preparation of protein samples under denaturing conditions, the separation on SDS-PAGE and the electroblotting to nitrocellulose membranes was previously described (Gallo et al., 2018). UCBSV was detected using anti-CP (Ref. AS-1153, DSMZ) as primary antibody and horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (Ref. 111-035-003, Jackson ImmunoResearch) as the secondary reagent. GFP and Myc-tag were detected using anti-GFP (Ref. 11814460001, Roche) and anti-Myc (either Ref. M20002, AbMART; or Ref.

05-724, Millipore) as primary antibodies, respectively, and HRP-conjugated sheep anti-mouse IgG (Ref. NA931, Amersham) as the secondary reagent. Immunostained proteins were visualized by enhanced chemiluminescence detection with Clarity ECL Western blotting substrate (Bio-Rad) in a ChemiDoc apparatus (Bio-Rad). Ponceau red was used to verify equivalent loading of total proteins in each sample.

Reverse transcription followed by PCR. Firstly, total RNA was isolated from *N. benthamiana* and cassava leaves by using the FavorPrep Plant Total RNA Purification Mini Kit (Ref. FAPRK 001, Favorgen Biotech) and Spectrum Plant Total RNA Kit (Ref. STRN50, Sigma), respectively. The RNA integrity was verified by electrophoresis in agarose gel. Secondly, cDNA was synthesized from 500 ng of total RNA with retrotranscriptase from Moloney murine leukemia virus (Ref. M0253, New England BioLabs) and random hexanucleotides as primers by following the manufacturer's instructions. Then, the cDNAs were used as template to amplify the region that encodes NIb_C-HAM1-CP_N with primers #3160/#3130 or the one that encodes a short fragment of CP with primers #3547/#3130. In the particular case of samples from cassava, which are prone to be contaminated with RT-PCR inhibitors such as polyphenols, RNA quality was checked by RT-PCR amplification of the UBQ10 housekeeping gene (Moreno et al., 2011) in order to validate the lack of amplification of UCBSV-derived fragments in samples from non-infected plants. Finally, when required, PCR products were Sanger sequenced by Macrogen Europe.

Measurement of NTPs in plant leaves. Free NTPs were extracted from young leaves of *N. benthamiana* and cassava by using a previously described method (Riondet et al., 2005). Extracts were immediately injected into a Vanquish UHPLC system equipped with a Q Exactive Focus Orbitrap spectrometry detector (Thermo Fisher Scientific). NTPs were separated by means of a Primesep SB column (3 µm, 4.6 x 150 mm) (SIELC Technologies) with a mobile phase formed by a mixture of (A) acetonitrile and water (5/95 v/v) with 30 mM of ammonium acetate (pH 4.5) and (B) acetonitrile and water (10/90 v/v) with 200 mM of ammonium acetate (pH 4.5) flowing at 1.0 ml/min with a gradient from 50-to-100% of A in 15 minutes. Injection was set to 5 µl and the column temperature to 25°C. Electrospray ionization was done at 4000V, setting the capillary temperature to 400°C. Desolvation was carried out with nitrogen with sheath gas and auxiliary gas flow rates of 70 and 20 (500°C), respectively. NTPs were detected

in MS/MS experiments (scan range from 50 to 550) based on the transition from the molecular protonated cation ([M+H]+) to the breakdown product consisting of the corresponding protonated nucleobase ([Nb+H]+) at collision energy of 25 eV.

Results

UCBSV does not require HAM1 to infect *N. benthamiana*.

With the aim of tracking the UCBSV infection *in planta*, an UCBSV full-length cDNA clone was manipulated to introduce the GFP coding sequence between NIb and CP cistrons (Figure 1A). The infection efficiency of UCBSV-eGFP was compared with that of the wild-type UCBSV in the model plant *N. benthamiana*. Plants inoculated with the wild-type virus (n=3) started to display clear symptoms of viral infection in upper non-inoculated leaves at 10 dpi, whereas those inoculated with UCBSV-eGFP (n=3) had a delay in symptom appearance of 2-to-3 days. At 15 dpi all inoculated plants showed equivalent symptoms in apical leaves, including strong leaf curling and vein clearing (Figure 1B); however, in line with the delay in symptom appearance, the height of plants inoculated with UCBSV-eGFP was in the middle of the untreated plants (tallest) and those infected with the wild-type UCBSV (shortest) (Figure 1B). As expected, upper non-inoculated leaves of plants infected with UCBSV-eGFP displayed GFP-derived fluorescence when observed under UV light (Figure 1B). In accordance with the other infection parameters, viral load in upper non-inoculated leaves, inferred from UCBSV CP immunodetection, was slightly higher in plants infected with the wild-type virus (Figure 1C). The immunodetection analysis also showed that GFP produced by UCBSV-eGFP was properly released from the viral polyprotein during the infection (Figure 1C).

After a plant-to-plant passage, unlike in plants initially inoculated with cDNA clones, we observed no differences among plants infected with wild-type and GFP-tagged viruses regarding the time of appearance and intensity of systemic symptoms. When upper non-inoculated leaves of plants infected with UCBSV-eGFP were observed under UV light at 20 dpi, curiously, fluorescence was not detected, suggesting that the GFP cistron had been deleted from the viral genome (data not shown). Indeed, a deeper analysis of viral populations from these plants confirmed this assumption, as DNA products amplified by RT-PCR with primers flanking the HAM1-GFP coding region were much smaller than those produced from plants originally infected by shooting (Figure 1D). Remarkably, direct Sanger sequencing of these products showed that not

only GFP-, but also HAM1-coding sequences, had been either totally or partially deleted from UCBSV-eGFP after the first passage (Figure 1D). This result support the idea that HAM1 is not required for the virus to infect *N. benthamiana*.

To confirm that HAM1 is unnecessary for UCBSV to infect the experimental host *N. benthamiana*, and rule out the possibility that a small fraction of the wild-type virus was complementing the deletion mutant, we built an infectious cDNA clone that carries a complete deletion of HAM1 cistron (Figure 2A). *N. benthamiana* plants inoculated with plasmids expressing either UCBSV or UCBSV-ΔHAM1 (n=3 per construct) started to display clear infection symptoms in upper non-inoculated leaves at the same time, and these symptoms were similar in intensity and type (Figure 2B). Estimation of viral load in these plants was carried out in samples from systemically infected leaves by RT-qPCR to detect small differences, if any. As observed in Figure 2C, accumulation of viral RNA did not show significant differences between both viruses. Together, experiments shown in Figure 1 and 2 demonstrate that HAM1 is not required to produce an UCBSV wild-type-like infection in *N. benthamiana*.

UCBSV requires HAM1 pyrophosphatase activity to infect its natural host.

Based on the above results, we hypothesized that the presence of HAM1 in UCBSV is a specific requirement for the virus to infect its natural host. To test this guess, we inoculated cassava and *N. benthamiana* plants in parallel with UCBSV and UCBSV-ΔHAM1 (n=3 per virus and plant species). As in the previous experiment, *N. benthamiana* plants displayed clear symptoms of UCBSV infection at 9-to-10 dpi in upper non-inoculated leaves independently of the presence/absence of HAM1 cistron in the viral genome (data not shown). Cassavas, in turn, developed typical UCBSV symptoms (yellow mottling along the major veins in leaves and dark brown streaks in stems) by 45-to-60 dpi in plants inoculated with the wild-type virus (Figure 2D). In contrast, plants inoculated with UCBSV-ΔHAM1, as those untreated, had normal leaf coloring and lacked streaks in stems (Figure 2D), even after 180 dpi (data not shown). The presence of UCBSV in these plants was tested by RT-PCR in samples collected at 60 dpi from upper non-inoculated leaves. The result confirmed our visual observation: only plants inoculated with wild-type UCBSV accumulated viral RNA in upper non-inoculated tissues (Figure 2E).

Our results strongly suggest that UCBSV requires a pyrophosphatase activity to infect cassava. In order to test this hypothesis, and to rule out the possibility that the lack of infectivity of UCBSV-ΔHAM1 in cassava was rather due to an undesired side effect caused by the deletion of the whole HAM1 cistron from the viral genome, we aimed to introduce just a single point mutation in HAM1 to specifically disrupt its pyrophosphatase activity. Based on previous reports on the crystal structure of the human HAM1 (named ITPA) bound to ITP (Stenmark et al., 2007), we modeled with high confidence (QMEAN = -0.66) the tridimensional conformation of a UCBSV HAM1 dimer bound to this non-canonical nucleotide (Figure 3A). The K amino acid at position 38, which is located in the second α-helix, is among the fully conserved amino acids in HAM1-like proteins from potyvirids and from organisms as diverse as *Escherichia coli,* baker's yeast, *Arabidopsis* and human (Supplementary Figure S1). In ITPA, this particular K (K19) is proposed to be part of the protein catalytic centre, as its side chain directly interacts with the triphosphate group of ITP (Stenmark et al., 2007) (Figure 3A). Moreover, in line with such relevance, a mutation of this K (K13) in RdgB, the HAM1-like protein from *E. coli*, abolishes its capacity to hydrolyze ITP *in vitro* (Savchenko et al., 2007). Based on these data, we build an UCBSV cDNA clone that carries the mutation K38A in HAM1 (Figure 3B). The wild-type and mutant versions of UCBSV were inoculated in *N. benthamiana* and cassava in parallel (n=3 per virus and plant species). As expected, there were no differences in *N. benthamiana* plants inoculated with each of these viruses in infectivity, time of appearance and intensity of symptoms in upper non-inoculated leaves (Figure 3C), as well as in viral accumulation measured by RT-qPCR in samples from these tissues (Figure 3D). Conversely, only the three cassava plants inoculated with the wild-type virus displayed symptoms of viral infection in upper non-inoculated leaves at 60 dpi (Figure 3E). Further analysis by RT-PCR confirmed that the wild-type UCBSV, but not the mutant variant that carries the K38A mutation in HAM1, was able to infect cassavas (Figure 3F).

Together, these results indicate that an active pyrophosphatase contributes to UCBSV infection, and the requirement of this activity depends on the particular host.

Differential accumulation of NTPs in *M. esculenta* versus *N. benthamiana*.

Our observation that pyrophosphatase activity is only required for UCBSV infection in cassava prompted us to investigate the accumulation of canonical and non-canonical

nucleotides in *M. esculenta* and *N. benthamiana* plants. To do that, NTPs were extracted from equivalent amount of tissue powder from both UCBSV hosts (n=12 per plant species) and the relative concentrations of ATP, CTP, GTP, UTP, ITP and XTP were estimated by high performance liquid chromatography coupled with tandem mass spectrometry. Whereas the concentration (measured as the area under the curve) corresponding to CTP and ATP where equivalent in both plants, showing no significant differences, that of XTP, ITP, GTP and UTP were significantly higher in cassava relative to *N. benthamiana* (Figure 4A). This difference was particularly relevant in the case of the non-canonical nucleotides XTP (4.5 folds) and ITP (3.6 folds) (Figure 4B). Importantly, an independent repetition of this experiment showed equivalent differences when comparing the population of NTPs in leaves of these two plant species. Therefore, we can conclude that *M. esculenta*, the natural host of UCBSV, accumulates much higher levels of XTP and ITP in leaves than the *N. benthamiana* counterpart.

Suboptimal cleavage at NIb/HAM1 junction during UCBSV infection.

When the presence of HAM1 cistron in the genome of UCBSV (named CBSV at that time) was reported for the first time, authors proposed that NIb and HAM1 might accumulate as two independent mature factors in infected cells due to the presence of a putative target for the viral-derived protease NIaPro (Mbanzibwa et al., 2009). A canonical NIaPro cleavage site is formed by 9 moderately conserved amino acids, and cleavage occurs between residues 6 and 7 (P1 and P1', Figure 5A). Amino acid primary sequence analysis shows that positions P4, P1 and P1' have high degree of conservation. For P1', for instance, either A, S or G residues were observed in 92% of the cases (n = 343 from 49 viral genomes × 7 cleavage sites (Adams et al., 2005). Curiously, a T residue occupies this position in the cleavage site located at the NIb/HAM1 junction of UCBSV (Figure 5A), which is not a common amino acid at P1' with a representation of 2%. In fact, a seminal study about the NIaPro-mediated cleavage at the NIb/CP junction of tobacco each virus, another potyvirus, showed that S x T mutation at P1' strongly reduced cleavage efficiency in an *in vitro* system (see Figure 4D in (Dougherty et al., 1988).

The above-mentioned antecedents prompted us to investigate whether the proposed cleavage site located between NIb and HAM1 is efficiently processed during UCBSV infection. To do that we built an infectious cDNA clone in which HAM1 was tagged with two copies of the Myc epitope (UCBSV-HAM1-2xMyc, Figure 5B) for the easy

detection of HAM1 in extracts of infected tissues. This clone, and the clone that expresses the wild-type UCBSV as control, were inoculated in *N. benthamiana* plants (n=3 per virus). No differences among inoculated plants were observed in term of viral symptoms (Figure 5C) and accumulation as estimated by western blot against UCBSV CP (Figure 5D). This result indicated that the tag does not have a noticeable negative impact on viral fitness in *N. benthamiana*. Immunodetection with Myc antibody revealed the presence of two defined protein species in samples infected with UCBSV-2xMYC. The one with less electrophoretic mobility had the expected size for the Myc-tagged NIb-HAM1 fusion product (86.3 kDa), whereas the smaller species had the expected size for the sole Myc-tagged HAM1 (28.2 kDa) (Figure 5D). The ratio between larger and smaller species was estimated in 1.5 based on the densitometric analysis of chemiluminescence signals.

Our results, along with previous antecedents (see above), suggested that T at position P1' causes an inefficient NIaPro-mediated processing at the cleavage site located in the NIb/HAM1 junction. To test this idea, we introduced mutations in the UCBSV cDNA clone to express two types of P1' mutants: (i) T1A and T1S, as A and S are among the most frequent residues at this position, and (ii) TxP, as P is not present at the P1' position in any NIaPro cleavage sites (Adams et al., 2005). When mutated and wild-type versions of UCBSV-2xMYC were inoculated in *N. benthamiana* (n=2 per virus), all of them produced indistinguishable infections, with comparable symptoms (Supplementary Figure 2) and virus accumulation in upper non-inoculated leaves as observed by immunodetection of UCBSV CP (Figure 5E). As anticipated from conservation of amino acids present at the P1' position, the T1A and T1S mutants accumulated only the protein species that corresponds to free HAM1, while T1P mutant only produced the NIb-HAM1 complex. Altogether, these results indicate that the NIaPro-mediated separation of NIb and HAM1 is inherently inefficient in UCBSV, which is due to the presence of a T residue at the P1' position of the cleavage site.

Relevance of the inefficient cleavage at NIb/HAM1 junction in cassava.

To estimate the relevance of the poor separation of NIb from HAM1 in the UCBSV natural host, we inoculated cassava plants with the Myc-tagged wild-type virus as well as the T1A and T1P variants (n=3 per virus). Clear symptoms of infection appeared at 60 dpi in the upper leaves of all inoculated plants independently of the infecting virus (Figure 6A). At that time, RT-PCR confirmed that upper non-inoculated leaves from all

inoculated plants were successfully infected with the Myc-tagged viruses (Figure 6B). Moreover, Sanger sequencing analysis of these RT-PCR products indicated that the introduced mutations were maintained after two months (Figure 6C). At 120 dpi, we divided plants infected with each virus in two groups, such as one plant was kept growing (plant 1), whereas the remaining two plants were propagated through stem cuttings (plant 2 and plant 3). At 180 dpi, samples were taken from the upper leaves of all plants and the identity of infecting viruses was determined. Remarkably, whereas the NIb-HAM1 junction from both wild-type and the T1P variants remained unchanged in all the analyzed plants (data not shown), that of the T1A variant evolved to introduce mutations (Figure 6C and Supplementary Figure S3). In the case of plant 1, a second mutation appeared at the P3' position, so that the original D amino acid was replaced by G (D3G) to give rise to a T1A/D3G double mutant (Figure 6C). Cuttings made from plant 2 also accumulated a viral variant carrying the second mutation G3D, reinforcing the idea that the T1A single mutant evolves to T1A/D3G when adapting to cassava (Supplementary Figure S3). Finally, cuttings made from plant 3 showed accumulation of a variant that encodes the wild-type cleavage site, so that the T1A mutation reverted to T (Supplementary Figure S3).

The observed reversion in cuttings from plant 3 strongly suggested that the virus requires the inefficient processing of the cleavage site located at the NIb-HAM1 junction for a successful infection. If that were the case, then one would expect that the T1A/D3G double mutant mimics this phenotype. To test this idea, we built the double mutant T1A/D3G by directed mutagenesis of the UCBSV-HAM1-2xMyc clone, and this plasmid was used to inoculate *N. benthamiana* plants (n=4) for easy detection of processing products by western blot. Both UCBSV-HAM1-2xMyc and UCBSV-HAM1$_{T1A}$-2xMyc variants were used as control (n=2 per variant). The T1A/D3G double mutant behaved as controls in term of infection timing and visible symptoms (data not shown), as well as viral accumulation in upper non-inoculated leaves as estimated by immunodetection of UCBSV CP (Figure 6D). Detection of Myc-tagged proteins in samples from systemically infected tissue showed that the T1A/D3G double mutant, as in the case of the wild-type virus, and unlike the T1A variant, accumulated two different protein species: NIb-HAM1 and free HAM1 (Figure 6D). Therefore, our results indicate that, at least for a relevant fraction of the total NIb and HAM1 produced during UCBSV infection, (i) these two proteins stay covalently bound, and (ii) the NIb-HAM1 partnership is indeed essential when UCBSV infects its natural host. In addition,

our viral evolution experiment highlights the importance of the usually underestimated amino acids located at P3' position of NIaPro cleavage sites for the actual NIaPro processing.

The expression of a joint NIb-HAM1 product is a common future of potyvirds encoding HAM1.

CBSV and EuRV are also potyvirids encoding HAM1 in their genomes, and this cistron is located, as in the case of UCBSV, just downstream of NIb. Importantly, the previously proposed NIaPro cleavage site at the NIb/HAM1 junction in both viruses (Mbanzibwa et al., 2009) does not fit the conventional conservation rules. CBSV has a V at P1', which is not a common residue at this position (Figure 5A) (Adams et al., 2005). In the case of EuRV, P1 is occupied by R, which is a strongly underrepresented amino acid at this position (Figure 5A) (Adams et al., 2005). Therefore, we hypothesized that HAM1 also remains bound to CBSV and EuRV NIbs. To test this idea, and due to the lack of infectious cDNA clones for these two viruses, we transiently expressed 4xMyc c-terminal tagged versions of NIb-HAM1-CP either with or without the presence of their cognate NIa (VPg-NIaPro) proteinases (Figure 7A). We did the same with the equivalent fragments of UCBSV as control for comparison. As expected, the expression of UCBSV fragments mimicked the results that we got with the full-length UCBSV-2xMYC virus, so that NIb-HAM1-CP was processed only in the presence of NIa, and it happened at the cleavage site located at the HAM1/CP junction and, with much lower efficiency, at the site placed at the NIb/HAM1 junction (Figure 7B). Remarkably, CBSV behaved just like UCBSV, as the main product, by far, was the one produced after cleavage at the HAM1/CP junction, with just a residual accumulation of the small fragment corresponding to the processing at the NIb/HAM1 junction (Figure 7B). Finally, for EuRV we only detected the product that corresponds to the NIaPro-mediated processing of the cleavage site located between HAM1 and CP (Figure 7B). Altogether, we conclude that most of the NIb and HAM1 might also be covalently bound during CBSV and EuRV infections.

Discussion

RNA viruses are widespread in nature, where they display a great diversity of particle structures, genome arrangements and expressed proteins (Dolja and Koonin, 2018). Despite these differences, they are all replicated by viral-encoded RdRPs sharing, at

least in all cases reported so far, a highly conserved core architecture folded into three subdomains (thumb, palm, and fingers) resembling a cupped right hand (Venkataraman et al., 2018). In some cases, other key protein domains implicated in viral replication, and/or transcription, are acquired by basic RdRP cores. The flaviviral replicase (NS5), for instance, possesses a capping enzyme domain required to synthetize the 5'-cap structure of genomic RNA (Lu and Gong, 2017, Brand et al., 2017). The potexviral replicase, in turn, not only has a capping enzyme domain, but also a helicase (Park et al., 2013). Remarkably, the covalent association between a viral RdRP and a HAM1-like protein had not been described so far. Data presented in this study indicate that (i) particular potyvirid replicases are covalently bound to, and work in association with, a HAM1-like pyrophosphatase, and (ii) the requirement of this partnership is host specific, which might be due to the peculiar accumulation of XTP/ITP in some hosts.

Regarding the precise role of viral HAM1 enzymes during the infection, the simple fact that a high fraction of this protein stays covalently attached to the viral replicase strongly suggests that HAM1 participates in replication. As UCBSV- and CBSV-derived HAM1s are pyrophosphatases with preference for non-canonical nucleotides (Tomlinson et al., 2019), it is logical to hypothesise that HAM1 hydrolyses ITP/XTP in order to prevents their incorporation into the viral genome, which would otherwise cause inhibition of RNA synthesis and/or further genome mutations. In other words, it seems quite likely that ITP/XTP behave as natural antiviral molecules, similarly to artificial nucleoside- and nucleotide-like analogues used against plus-stranded RNA viruses in animals (Deval et al., 2014). Intriguingly, our experiments with UCBSV (data not shown), as well as previous results with CBSV (Tomlinson et al., 2019), showed that the absence of HAM1 does not increase the complexity of UCBSV and CBSV mutant swarms in infected *N. benthamiana* plants. The incapacity of pyrophosphatase-defective UCBSV variants to infect cassava (Figure 2 and 3) precluded us to test whether the absence of this activity increases the variability of UCBSV genome sequence in its natural host, where HAM1 is strictly required.

Theoretically, the concentration of ITP/XTP in the pool of free nucleotides inside cells are tightly maintained at low levels by ITPases to avoid their deleterious effects over DNA and RNA molecules (Simone et al., 2013). Therefore, results showing that cassava, and probably other euphorbiaceous, accumulates high amounts of ITP/XTP (Figure 4) question this rule. To conciliate our result in cassava with that broadly accepted idea, we hypothesise that some plants accumulate unexpectedly high

concentration of ITP/XTP in certain subcellular compartments, whereas in those locations where they have damaging consequences, such as in the nucleus, ITP/XTP are kept at much lower concentration. The recent suggestion that euphorbiaceous HAM1-like proteins might harbour a nuclear localization signal (James et al., 2021) fits pretty well with this assumption. Therefore, it is possible that viruses infecting plants from the *Euphorbiacea* family (e.g. UCBSV, CBSV, EuRV and CsTLV) have to face high levels of ITP/XTP in the cytoplasm, where they replicate, thus explaining the incorporation of a HAM1 enzyme as an active module of the viral replicase. This possibility also fits well with the expression of some free HAM1 during the infection (Figure 5, 6 and 7), so that it might also help to get rid of ITP/XTP in all those cellular environments where the virus is replicating.

All in all, our findings inform about a novel and interesting case of virus/host coevolution, highlighting (i) the striking peculiarity of cassava plants, and presumably other euphorbiaceous, of accumulating high levels of ITP/XTP into cells, and (ii) the flexibility of RNA viruses to incorporate additional factors when required. Whether this peculiar feature of cassava regarding the high concentration of non-canonical nucleotides evolved as a *bona fide* strategy to prevent multiplication of pathogens, and how this plant copes with the harmful effect of ITP/XTP, are indeed exciting questions deserving special attention in future studies.

Acknowledgments.

We would like to thank Tsuyoshi Nakagawa, Daniel Silhavy and Gary Foster, for providing Gateway expression vectors, pBIN61-P14 and pYES2-CBSV-F2, respectively. We are also grateful to the Mass Spectrometry Facility (Nucleus, USAL) for its kind help. This work was funded by BIO2015-73900-JIN (AEI-FEDER), PID2019-110979RB-I00/ AEI / 10.13039/501100011033, RYC2018-025523-I and 202020E001 to A.A.V, BIO2016-80572-R (AEI-FEDER) and PID2019-109380RBI00/ AEI /301 10.13039/501100011033 to J.A.G, and Funding Program for Research Groups (M.C2 from USAL) to D.G.G.

Author contribution

Conceptualization, A.A.V, J.A.G; Investigation, A.A.V, R.G.L., M.R., F.J.M, D.G.G., B.G.; I.G., A.G.P. and I.M.; Writing – Original Draft, A.A.V.; Writing – Review &

Editing, A.A.V, J.A.G.; Funding Acquisition, A.A.V, J.A.G., D.G.G.; Resources, F.P.; Supervision, A.A.V.

Declaration of interests

The authors declare no competing interests.

References

ADAMS, M. J., ANTONIW, J. F. & BEAUDOIN, F. 2005. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol Plant Pathol,* 6**,** 471-87.

ALLISON, R., JOHNSTON, R. E. & DOUGHERTY, W. G. 1986. The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology,* 154**,** 9-20.

BRAND, C., BISAILLON, M. & GEISS, B. J. 2017. Organization of the Flavivirus RNA replicase complex. *Wiley Interdiscip Rev RNA,* 8.

CARBONELL, A., DUJOVNY, G., GARCIA, J. A. & VALLI, A. 2012. The Cucumber vein yellowing virus silencing suppressor P1b can functionally replace HCPro in Plum pox virus infection in a host-specific manner. *Mol Plant Microbe Interact,* 25**,** 151-64.

CARRINGTON, J. C. & DOUGHERTY, W. G. 1987a. Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology,* 160**,** 355-62.

CARRINGTON, J. C. & DOUGHERTY, W. G. 1987b. Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *J Virol,* 61**,** 2540-8.

CHUNG, J. H., BACK, J. H., PARK, Y. I. & HAN, Y. S. 2001. Biochemical characterization of a novel hypoxanthine/xanthine dNTP pyrophosphatase from Methanococcus jannaschii. *Nucleic Acids Res,* 29**,** 3099-107.

CHUNG, J. H., PARK, H. Y., LEE, J. H. & JANG, Y. 2002. Identification of the dITP- and XTP-hydrolyzing protein from Escherichia coli. *J Biochem Mol Biol,* 35**,** 403-8.

DEVAL, J., SYMONS, J. A. & BEIGELMAN, L. 2014. Inhibition of viral RNA polymerases by nucleoside and nucleotide analogs: therapeutic applications against positive-strand RNA viruses beyond hepatitis C virus. *Curr Opin Virol,* 9**,** 1-7.

DOLJA, V. V. & KOONIN, E. V. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res,* 244**,** 36-52.

DOMBROVSKY, A., REINGOLD, V. & ANTIGNUS, Y. 2014. Ipomovirus--an atypical genus in the family Potyviridae transmitted by whiteflies. *Pest Manag Sci,* 70**,** 1553-67.

DOUGHERTY, W. G., CARRINGTON, J. C., CARY, S. M. & PARKS, T. D. 1988. Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J,* 7**,** 1281-7.

GALLO, A., VALLI, A., CALVO, M. & GARCIA, J. A. 2018. A Functional Link between RNA Replication and Virion Assembly in the Potyvirus Plum Pox Virus. *J Virol,* 92.

HILLOCKS, R. J. & MARUTHI, M. N. 2015. *Post-harvest impact of cassava brown streak disease in four countries in eastern Africa,* Practical Action Publishing.

HO, S. N., HUNT, H. D., HORTON, R. M., PULLEN, J. K. & PEASE, L. R. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene,* 77**,** 51-9.

HONG, Y. & HUNT, A. G. 1996. RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology,* 226**,** 146-51.

HWANG, K. Y., CHUNG, J. H., KIM, S. H., HAN, Y. S. & CHO, Y. 1999. Structure-based identification of a novel NTPase from Methanococcus jannaschii. *Nat Struct Biol,* 6**,** 691-6.

JAMES, A. M., SEAL, S. E., BAILEY, A. M. & FOSTER, G. D. 2021. Viral inosine triphosphatase: A mysterious enzyme with typical activity, but an atypical function. *Mol Plant Pathol,* 22**,** 382-389.

JIMÉNEZ POLO, J., LEIVA, A. M. & CUELLAR, W. J. 2018. Identification of a torradovirus-encoded protein that complements the systemic movement of a potexvirus lacking the TGB3 gene. *International Congress of Plant Pathology (ICPP) 2018: Plant Health in A Global Economy, Boston.*

KNIERIM, D., MENZEL, W. & WINTER, S. 2017. Analysis of the complete genome sequence of euphorbia ringspot virus, an atypical member of the genus Potyvirus. *Arch Virol,* 162**,** 291-293.

LIN, S., MCLENNAN, A. G., YING, K., WANG, Z., GU, S., JIN, H., WU, C., LIU, W., YUAN, Y., TANG, R., XIE, Y. & MAO, Y. 2001. Cloning, expression, and characterization of a human inosine triphosphate pyrophosphatase encoded by the itpa gene. *J Biol Chem,* 276**,** 18695-701.

LOPEZ-MOYA, J. J. & GARCIA, J. A. 2000. Construction of a stable and highly infectious intron-containing cDNA clone of plum pox potyvirus and its use to infect plants by particle bombardment. *Virus Res,* 68**,** 99-107.

LU, G. & GONG, P. 2017. A structural view of the RNA-dependent RNA polymerases from the Flavivirus genus. *Virus Res,* 234**,** 34-43.

MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R. N., POTTER, S. C., FINN, R. D. & LOPEZ, R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res,* 47**,** W636-W641.

MBANZIBWA, D. R., TIAN, Y., MUKASA, S. B. & VALKONEN, J. P. 2009. Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *J Virol,* 83**,** 6934-40.

MERAI, Z., KERENYI, Z., MOLNAR, A., BARTA, E., VALOCZI, A., BISZTRAY, G., HAVELDA, Z., BURGYAN, J. & SILHAVY, D. 2005. Aureusvirus P14 is an efficient RNA silencing suppressor that binds double-stranded RNAs without size specificity. *J Virol,* 79**,** 7217-26.

MORENO, I., GRUISSEM, W. & VANDERSCHUREN, H. 2011. Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *J Virol Methods,* 177**,** 49-54.

PARK, M. R., SEO, J. K. & KIM, K. H. 2013. Viral and nonviral elements in potexvirus replication and movement and in antiviral responses. *Adv Virus Res,* 87**,** 75-112.

PASIN, F., BEDOYA, L. C., BERNABE-ORTS, J. M., GALLO, A., SIMON-MATEO, C., ORZAEZ, D. & GARCIA, J. A. 2017. Multiple T-DNA Delivery to Plants

Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synthetic Biology,* 6**,** 1962-1968.

PATIL, B. L., LEGG, J. P., KANJU, E. & FAUQUET, C. M. 2015. Cassava brown streak disease: a threat to food security in Africa. *J Gen Virol,* 96**,** 956-68.

PENNISI, E. 2010. Armed and dangerous. *Science,* 327**,** 804-5.

REVERS, F. & GARCÍA, J. A. 2015. Molecular biology of potyviruses. *Advances in Virus Research,* 92**,** 101-99.

RIONDET, C., MOREL, S. & ALCARAZ, G. 2005. Determination of total ribonucleotide pool in plant materials by high-pH anion-exchange high-performance liquid chromatography following extraction with potassium hydroxide. *J Chromatogr A,* 1077**,** 120-7.

SALVADOR, B., DELGADILLO, M. O., SAENZ, P., GARCIA, J. A. & SIMON-MATEO, C. 2008. Identification of Plum pox virus pathogenicity determinants in herbaceous and woody hosts. *Mol Plant Microbe Interact,* 21**,** 20-9.

SAVCHENKO, A., PROUDFOOT, M., SKARINA, T., SINGER, A., LITVINOVA, O., SANISHVILI, R., BROWN, G., CHIRGADZE, N. & YAKUNIN, A. F. 2007. Molecular basis of the antimutagenic activity of the house-cleaning inosine triphosphate pyrophosphatase RdgB from Escherichia coli. *J Mol Biol,* 374**,** 1091-103.

SCHECHTER, I. & BERGER, A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun,* 27**,** 157-62.

SIMONE, P. D., PAVLOV, Y. I. & BORGSTAHL, G. E. 2013. ITPA (inosine triphosphate pyrophosphatase): from surveillance of nucleotide pools to human disease and pharmacogenetics. *Mutat Res,* 753**,** 131-46.

STENMARK, P., KURSULA, P., FLODIN, S., GRASLUND, S., LANDRY, R., NORDLUND, P. & SCHULER, H. 2007. Crystal structure of human inosine triphosphatase. Substrate binding and implication of the inosine triphosphatase deficiency mutation P32T. *J Biol Chem,* 282**,** 3182-7.

TANAKA, Y., NAKAMURA, S., KAWAMUKAI, M., KOIZUMI, N. & NAKAGAWA, T. 2011. Development of a series of gateway binary vectors possessing a tunicamycin resistance gene as a marker for the transformation of Arabidopsis thaliana. *Biosci Biotechnol Biochem,* 75**,** 804-7.

TOMLINSON, K. R., BAILEY, A. M., ALICAI, T., SEAL, S. & FOSTER, G. D. 2018. Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Mol Plant Pathol,* 19**,** 1282-1294.

TOMLINSON, K. R., PABLO-RODRIGUEZ, J. L., BUNAWAN, H., NANYITI, S., GREEN, P., MILLER, J., ALICAI, T., SEAL, S. E., BAILEY, A. M. & FOSTER, G. D. 2019. Cassava brown streak virus Ham1 protein hydrolyses mutagenic nucleotides and is a necrosis determinant. *Mol Plant Pathol,* 20**,** 1080-1092.

VALLI, A., GARCÍA, J. A. & LÓPEZ-MOYA, J. J. 2021. Potyviruses (Potyviridae). *In:* BAMFORD, D. H. & ZUCKERMAN, M. (eds.) *Encyclopedia of Virology (Fourth Edition).* Oxford: Academic Press.

VALLI, A., LOPEZ-MOYA, J. J. & GARCIA, J. A. 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J Gen Virol,* 88**,** 1016-28.

VALLI, A., MARTIN-HERNANDEZ, A. M., LOPEZ-MOYA, J. J. & GARCIA, J. A. 2006. RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. *J Virol,* 80**,** 10055-63.

VENKATARAMAN, S., PRASAD, B. & SELVARAJAN, R. 2018. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*, 10.

WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46**,** W296-W303.

WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25**,** 1189-91.

WINTER, S., KOERBLER, M., STEIN, B., PIETRUSZKA, A., PAAPE, M. & BUTGEREITT, A. 2010. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *J Gen Virol*, 91**,** 1365-72.

YANG, X., LI, Y. & WANG, A. 2021. Research advances in potyviruses: From the laboratory bench to the field. *Annual Review of Phytopathology*, 59**,** *In Press*.

Figure legends

Figure 1. GFP-tagged UCBSV loses HAM1 and GFP coding sequences after one passage in *N. benthamiana*. (A) Schematic representation of viral constructs based on the pLX-UCBSV (Pasin et al., 2017) used in this experiment. Boxes represent mature viral factors as they are encoded in the viral genome. The presence of an intron in the P3 coding sequence is also indicated. p35S: 35S promoter from cauliflower mosaic virus; tNOS: terminator from the NOS gene of *Agrobacterium tumefaciens*. (B) Representative pictures taken at 15 days post-inoculation of infected and non-treated *N. benthamiana* plants under UV radiation and visible light (white bar = 1 cm; black bar = 4 cm). (C) Detection of GFP and UCBSV CP by immunoblot analysis in protein samples from upper non-inoculated leaves of *N. benthamiana* plants infected with the indicated viruses. Blot stained with *Ponceau* red showing the large subunit of the ribulose-1,5-bisphosphate carboxylase-oxygenase is included as a loading control (D) Agarose gel electrophoresis analysis of a viral genomic fragment amplified by RT-PCR from plants infected with UCBSV-GFP after one passage. The upper part shows a schematic representation of the amplified fragment. Black arrows represent primers used for amplification. Sizes of expected PCR products are indicated. Amino acids around the NIaPro cleavage sites are depicted at the bottom.

Figure 2. Virus-derived HAM1 is required for the successful infection of UCBSV in cassava plants, but not in *N. benthamiana*. (A) Schematic representation of the

NIb-to-CP genomic segment of viruses used in these experiments. Amino acids around the NIaPro cleavage sites are depicted. (B) Representative pictures of infected and non-treated *N. benthamiana plants* taken at 12 days post-inoculation. White bar = 4 cm. (C) RT-qPCR measuring the accumulation of viral RNA in upper non-inoculated leaves of *N. benthamiana plants* infected with the indicated viruses. Each bar represents the average of three plants (error bar = 1 standard deviation). For normalization, the average of wild type UCBSV is equal to 1. (D) Representative pictures of upper non-inoculated leaves, at 60 days post-inoculation, of cassava plants inoculated with the indicated viruses. White bar = 4 cm. Black arrows indicates the presence of brown streaks in the stem of an infected plant. (E) Analysis by agarose gel electrophoresis of a fragment of the UCBSV genome (V) and of a plant housekeeping gene (H) amplified by RT-PCR. RNA samples from upper non-inoculated leaves of 3 independent cassava plants inoculated with the indicated viruses were used as template.

Figure 3. The pyrophosphatase activity of UCBSV HAM1 is required for the successful infection of cassava plants. (A) Model of the ITP-bound UCBSV HAM1 tridimensional structure. Interaction between K38 and ITP is highlighted at the left. (B) Schematic representation of the NIb-to-CP genomic segment of viruses used in these experiments. Amino acids around the NIaPro cleavage sites are depicted. The presence of the K38A mutation is indicated with a red line. (C) Representative pictures of infected and non-treated *N. benthamiana plants* taken at 11 days post-inoculation. White bar = 4 cm. (D) RT-qPCR measuring the accumulation of viral RNA in upper non-inoculated leaves of *N. benthamiana plants* infected with the indicated viruses. Each bar represents the average of three plants (error bar = 1 standard deviation). For normalization, the average of wild type UCBSV is equal to 1. (E) Representative pictures of upper non-inoculated leaves, taken at 60 days post-inoculation, of cassava plants inoculated with the indicated viruses. White bar = 4 cm. (F) Analysis by agarose gel electrophoresis of a fragment of the UCBSV genome (V) and of a plant housekeeping gene (H) amplified by RT-PCR. RNA samples from upper non-inoculated leaves of 3 independent cassava plants inoculated with the indicated viruses were used as template.

Figure 4. High accumulation of non-canonical nucleotides in cassava. Base peak chromatogram in arbitrary units (AU) for representative samples of total NTPs

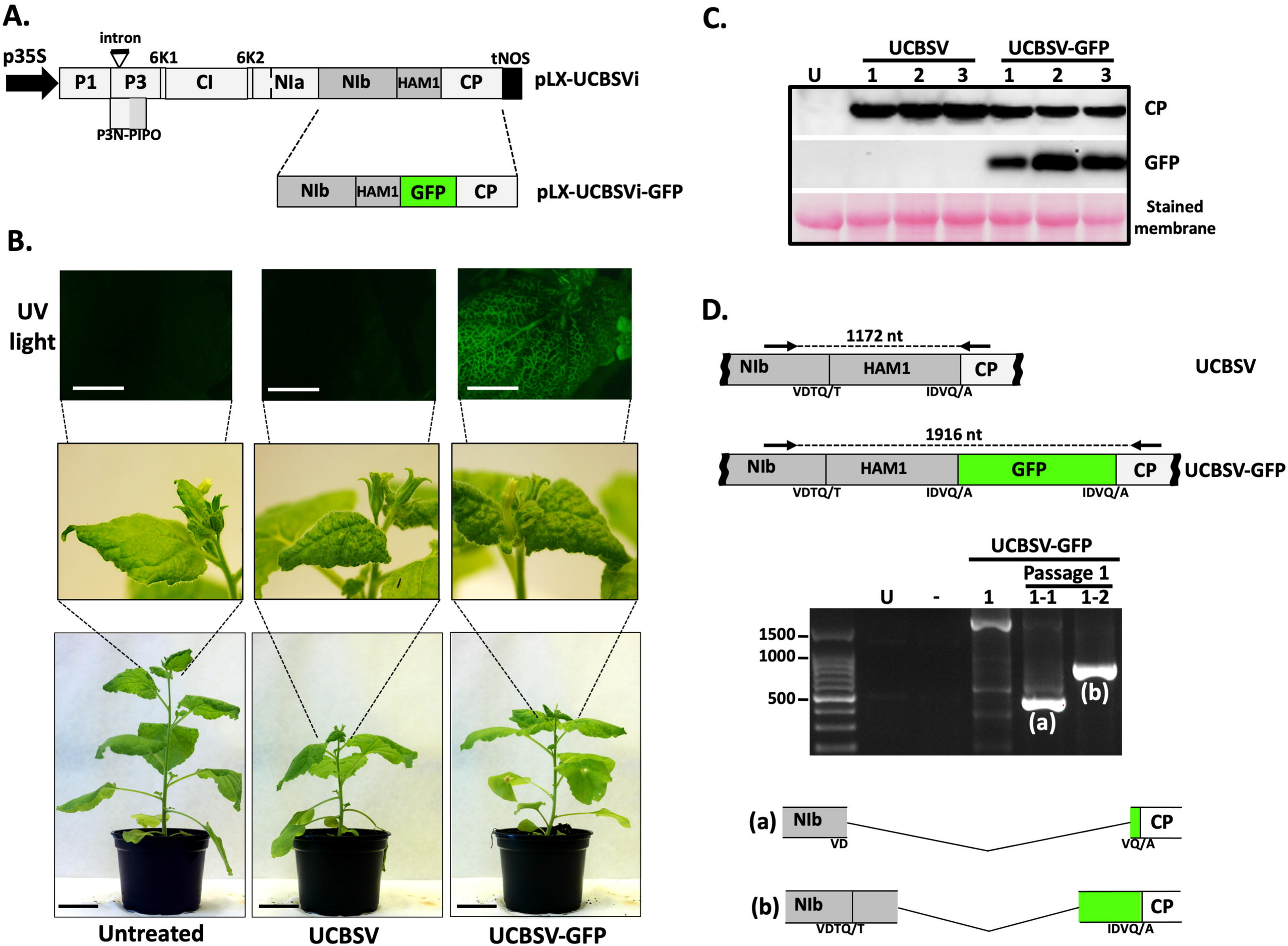
from *Manihot esculenta* (blue) and *Nicotiana benthamiana* (orange). The *M. esculenta/N. benthamiana* ratios for the average concentration (n=12 per plant species) of each NTP are indicated in parentheses. Non-canonical nucleotides are highlighted in red.

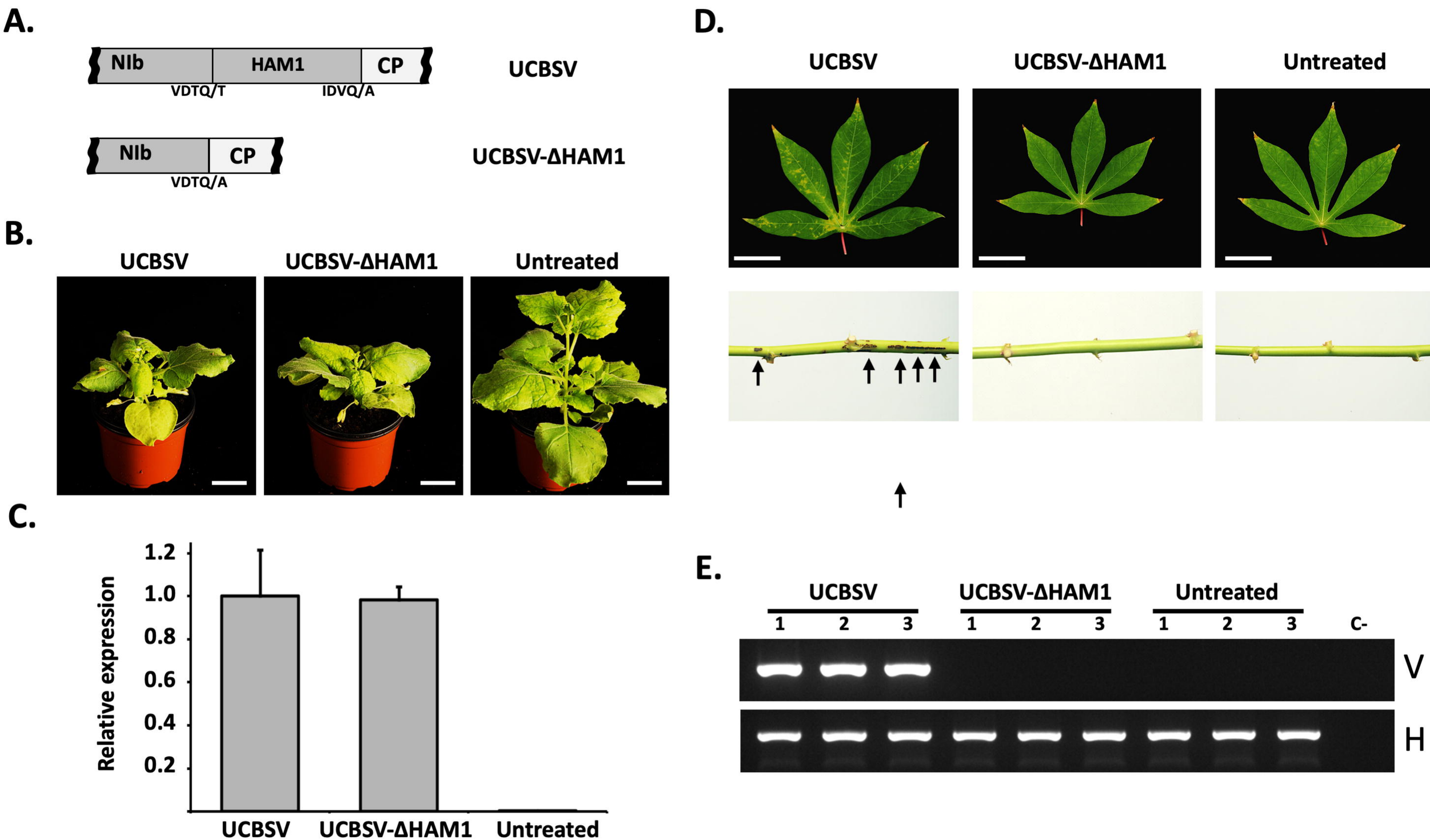
Figure 5. Suboptimal separation of NIb-HAM1 during UCBSV infection. (A) Schematic representation of a NIaPro cleavage site. Substrate residues at both sides of the scissile bond are labeled by following a previously proposed nomenclature (Schechter and Berger, 1967). The consensus sequence of NIaPro substrates, as well as those residues present at the NIb-HAM1 junction in UCBSV, CBSV and EuRV, are indicated. The non-conserved residue at the NIb-HAM1 cleavage site of each virus is surrounded by a blue circle. (B) Schematic representation of the NIb-to-CP genomic segment of viruses used in these experiments. Amino acids around the NIaPro cleavage sites are depicted. (C) Representative pictures of infected and non-treated *N. benthamiana plants* taken at 13 days post-inoculation. White bar = 4 cm. (D and E) Detection of Myc-tagged HAM1 and CP by immunoblot analysis in samples from upper non-inoculated leaves of *N. benthamiana* plants infected with the indicated viruses. The positions of prestained molecular mass markers (in kilodaltons) run in the same gels are indicated to the right. The black asterisk indicates the presence of a cross-reacting band in all the samples, including the untreated control. Blots stained with *Ponceau* red showing the large subunit of the ribulose-1,5-bisphosphate carboxylase-oxygenase are included as a loading control.

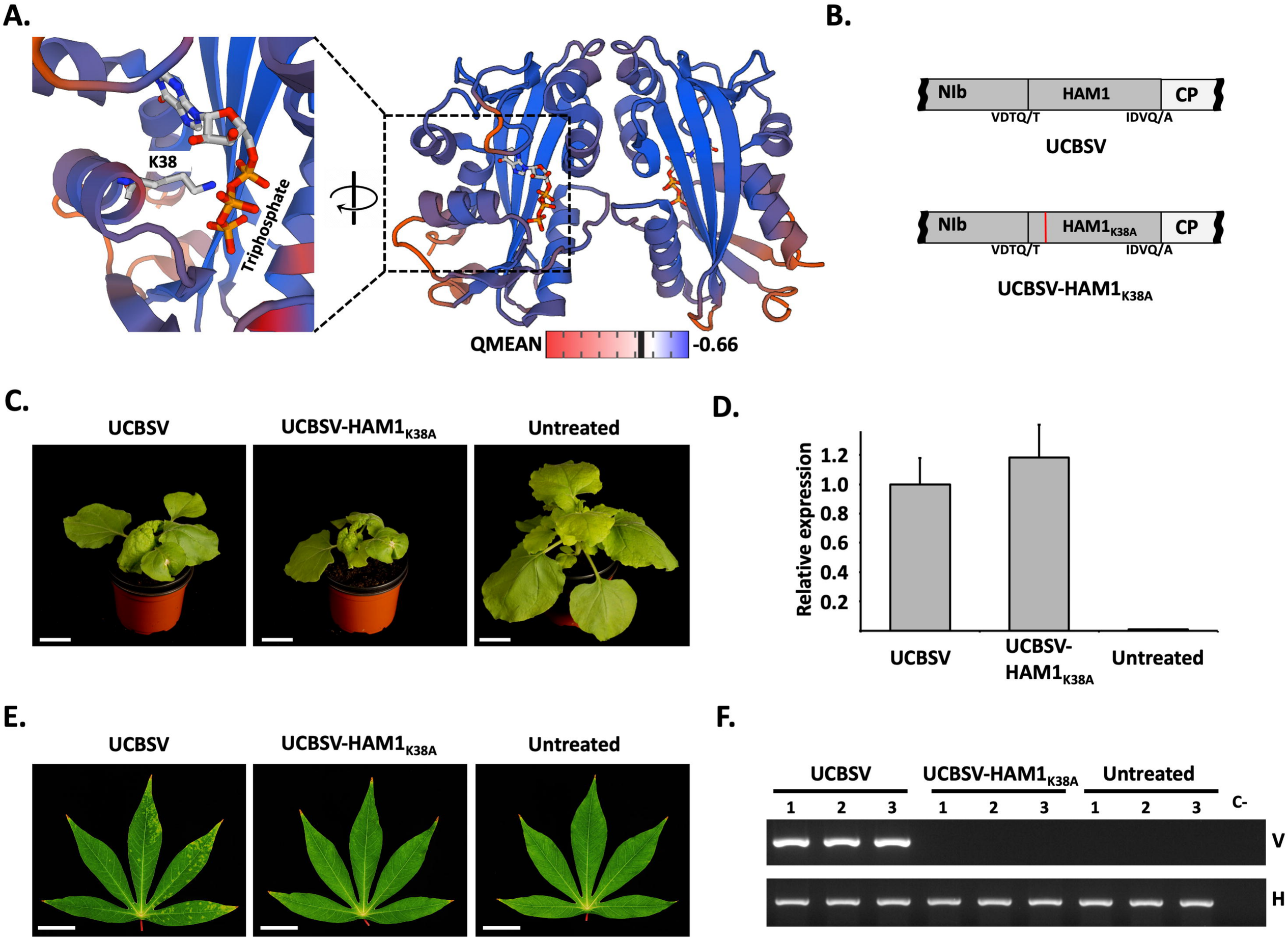
Figure 6. UCBSV HAM1$_{T1A}$ mutant, which undergoes an optimal cleavage at the NIb-HAM1 junction, evolves to display partial split. (A) Representative pictures of upper non-inoculated leaves, taken at 60 days post-inoculation, of cassava plants inoculated with the indicated 2xMyc-tagged versions of UCBSV. White bar = 4 cm. (B) Analysis by agarose gel electrophoresis of a fragment of the UCBSV genome (V) and a plant housekeeping gene (H) amplified by RT-PCR. RNA samples from upper non-inoculated leaves of 3 independent cassava plants inoculated with the indicated viruses and collected at 60 dpi, were used as template. (C) Chromatograms of Sanger sequencing results of the UCBSV genomic fragment of interest amplified by RT-PCR. RNA samples from upper non-inoculated leaves of a cassava plant inoculated with the UCBSV-HAM1$_{T1A}$-2xMyc mutant were used as template. Leaves for RNA preparation

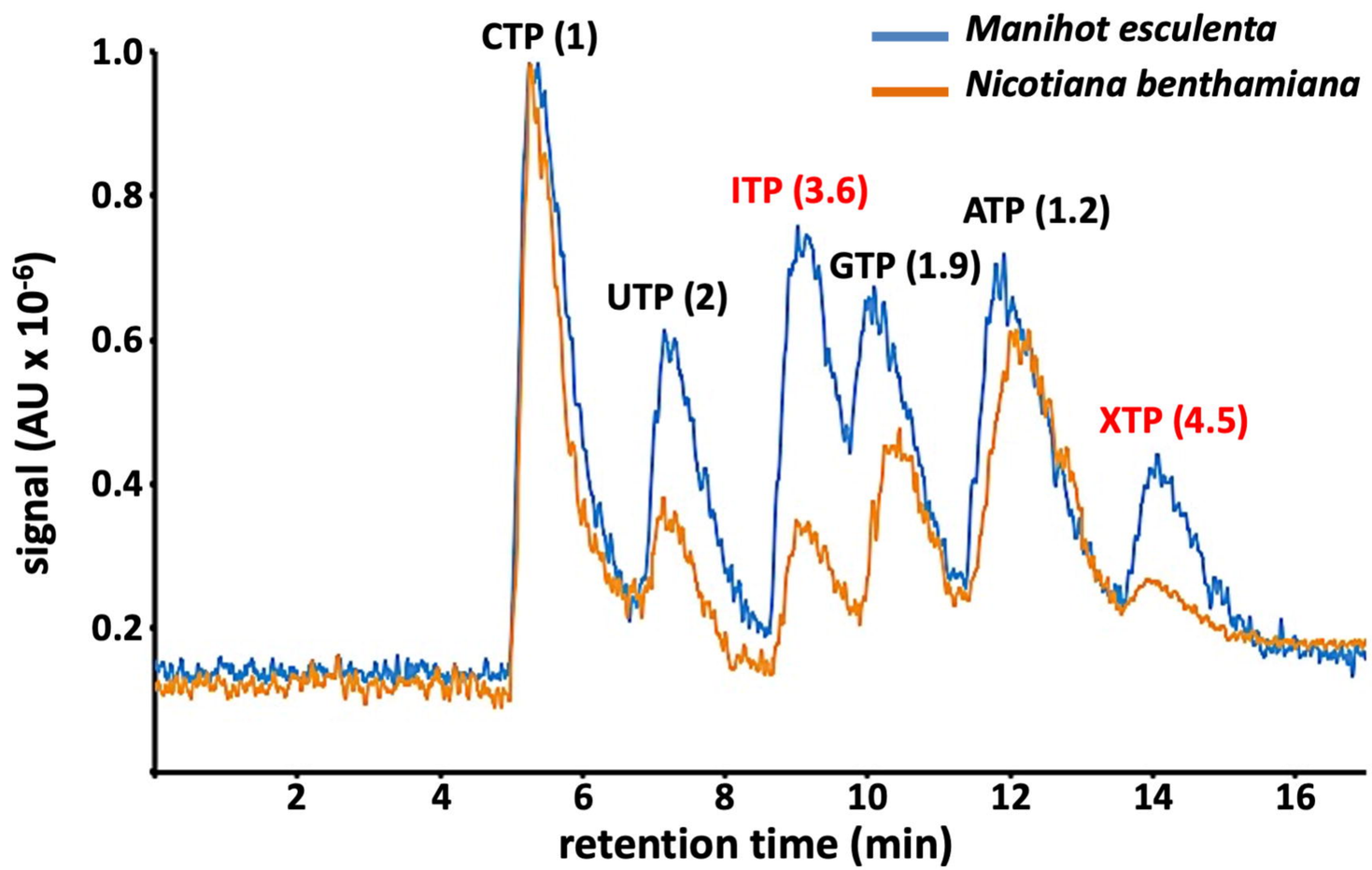
were harvested at 60 and 180 dpi. Residues derived from the original mutation and from the spontaneous second mutation are surrounded by a red circle. (D) Detection of Myc-tagged HAM1 and UCBSV CP by immunoblot analysis in samples from upper non-inoculated leaves of *N. benthamiana* plants infected with the indicated viruses. The positions of prestained molecular mass markers (in kilodaltons) run in the same gel is indicated to the right. Blot stained with Ponceau red showing the large subunit of the ribulose-1,5-bisphosphate carboxylase-oxygenase is included at the bottom as a loading control.

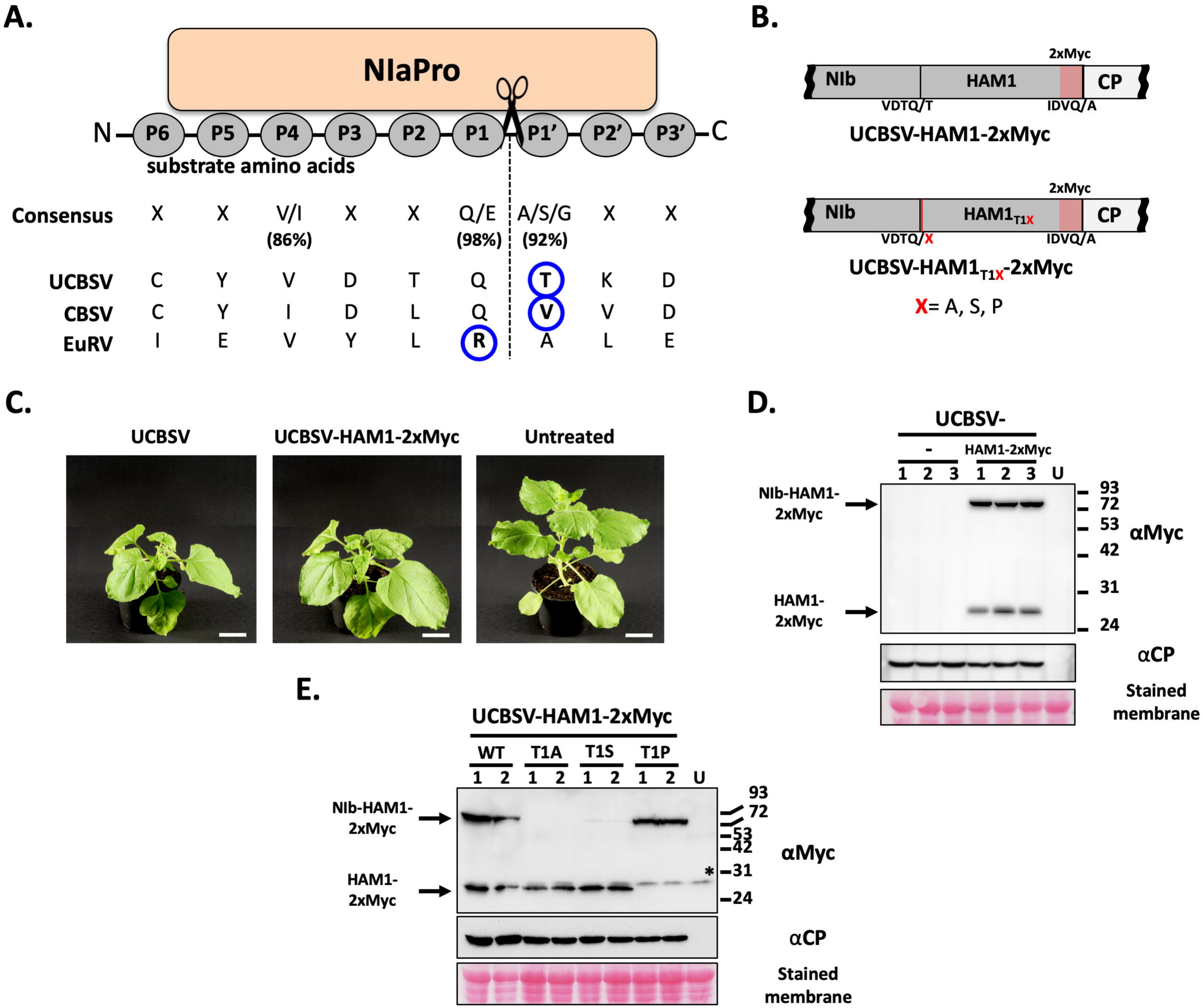
Figure 7. Suboptimal split of NIb-HAM1 is a general feature of potyvirids. Schematic representation of constructs based on pGWB702Ω and pGWB718 (Tanaka et al., 2011) used for these experiments. p35S: 35S promoter from cauliflower mosaic virus; tNOS: terminator from the NOS gene of *Agrobacterium tumefaciens;* NIb_C: NIb C-terminus; CP_N: CP N-terminus. (B) Detection of Myc-tagged proteins by immunoblot analysis in samples from *N. benthamiana* leaves expressing NIb_C-HAM1-CP_N versions in either the absence (-) or presence (+) of their cognate NIa. Viruses from which the transiently expressed proteins derived are indicated. Blots stained with Ponceau red showing the large subunit of the ribulose-1,5-bisphosphate carboxylase-oxygenase are included at the bottom as a loading control.

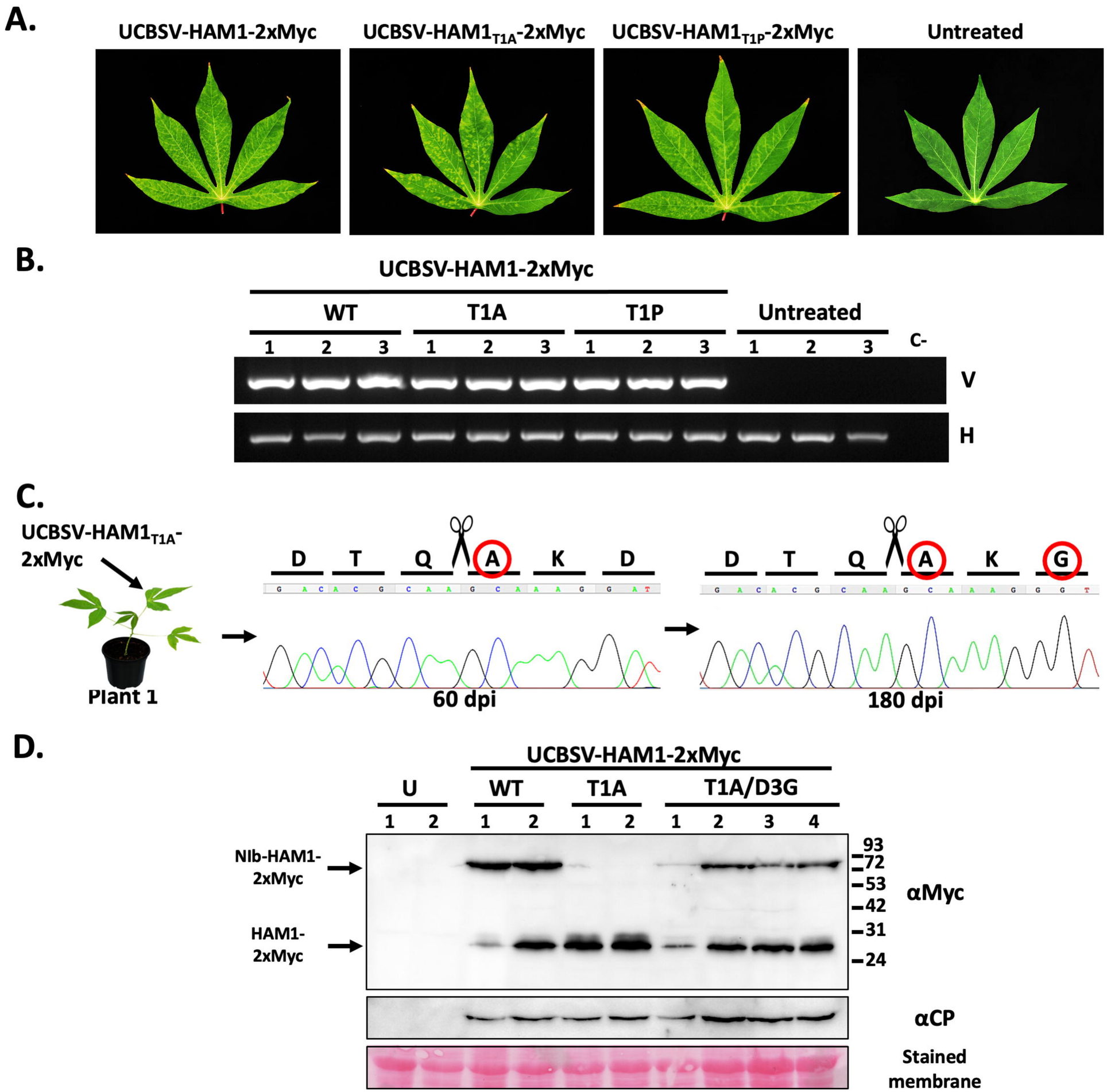

A.
p35S
intron
6K1
6K2
tNOS
P1
P3
CI
NIa
NIb
HAM1
CP
pLX-UCBSVi
P3N-PIPO
NIb
HAM1
GFP
CP
pLX-UCBSVi-GFP
B.
UV light
Untreated
UCBSV
UCBSV-GFP
C.
UCBSV
UCBSV-GFP
U
1
2
3
1
2
3
CP
GFP
Stained membrane
D.
1172 nt
NIb
HAM1
CP
UCBSV
VDTQ/T
IDVQ/A
1916 nt
NIb
HAM1
GFP
CP
UCBSV-GFP
VDTQ/T
IDVQ/A
IDVQ/A
UCBSV-GFP
Passage 1
U
-
1
1-1
1-2
1500
1000
500
(a)
(b)
(a)
NIb
VD
CP
VQ/A
(b)
NIb
VDTQ/T
CP
IDVQ/A

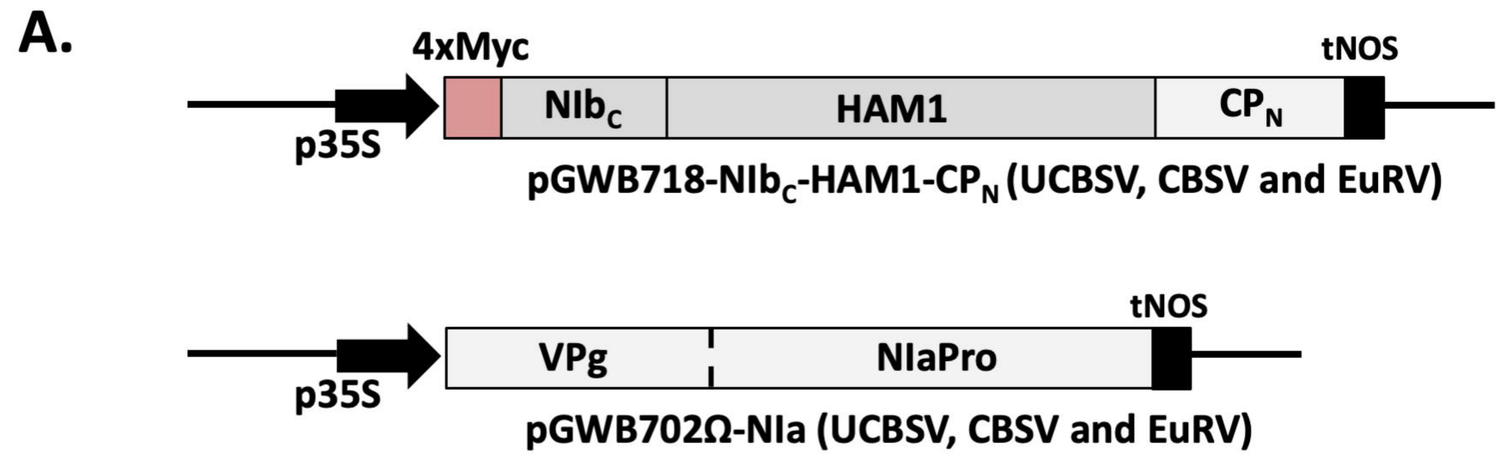

A.
NIb
HAM1
CP
VDTQ/T
IDVQ/A
UCBSV
NIb
CP
VDTQ/A
UCBSV-ΔHAM1
B.
UCBSV
UCBSV-ΔHAM1
Untreated
C.
Relative expression
1.2
1.0
0.8
0.6
0.4
0.2
UCBSV
UCBSV-ΔHAM1
Untreated
D.
UCBSV
UCBSV-ΔHAM1
Untreated
E.
UCBSV
UCBSV-ΔHAM1
Untreated
1
2
3
1
2
3
1
2
3
C-
V
H

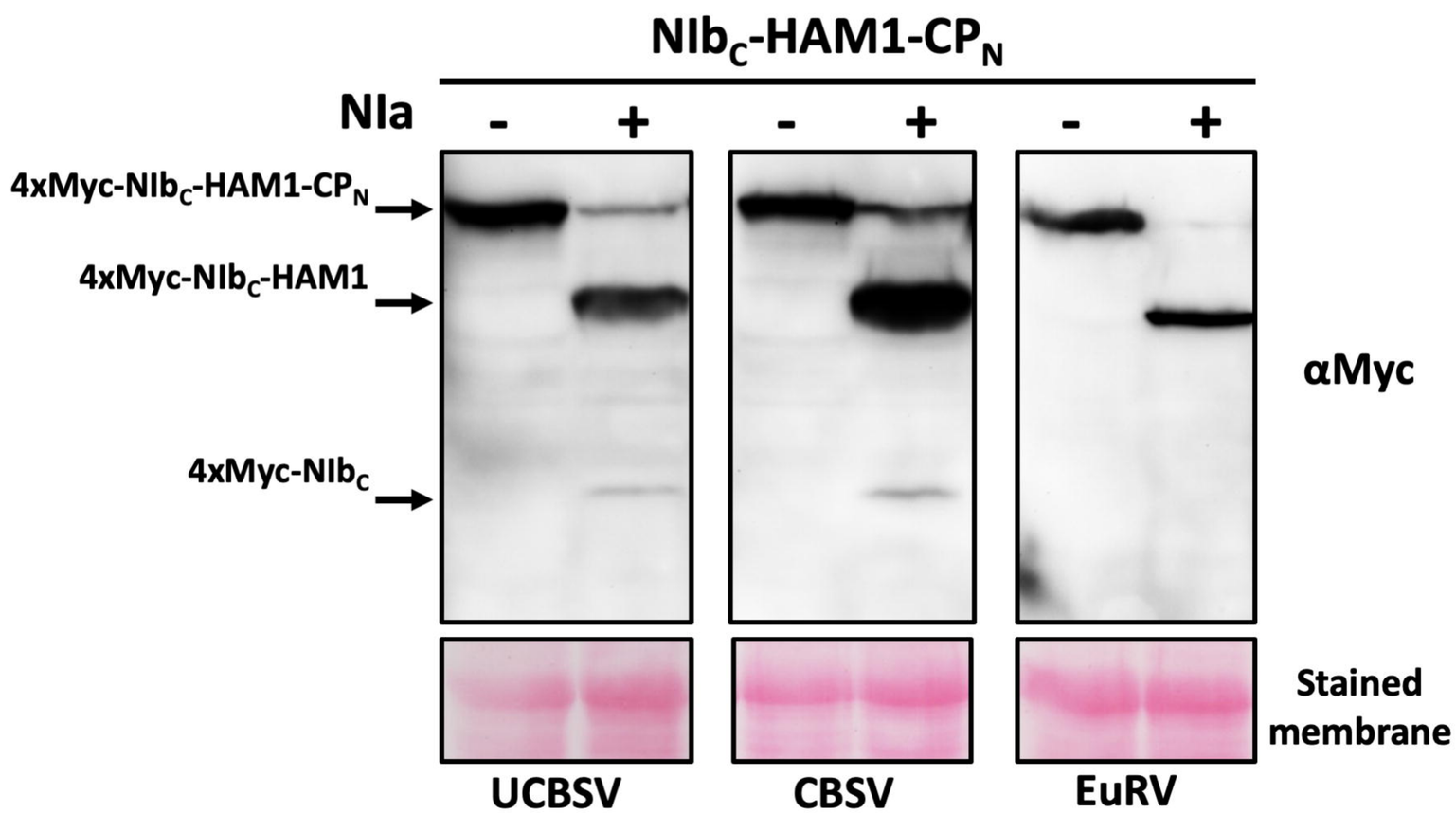

A.
K38
Triphosphate
QMEAN
-0.66
B.
NIb
HAM1
CP
VDTQ/T
IDVQ/A
UCBSV
NIb
HAM1K38A
CP
VDTQ/T
IDVQ/A
UCBSV-HAM1K38A
C.
UCBSV
UCBSV-HAM1K38A
Untreated
D.
Relative expression
1.2
1.0
0.8
0.6
0.4
0.2
UCBSV
UCBSV-HAM1K38A
Untreated
E.
UCBSV
UCBSV-HAM1K38A
Untreated
F.
UCBSV
UCBSV-HAM1K38A
Untreated
1 2 3 1 2 3 1 2 3 C-
V
H


Manihot esculenta
Nicotiana benthamiana
CTP (1)
UTP (2)
ITP (3.6)
GTP (1.9)
ATP (1.2)
XTP (4.5)
signal (AU x 10-6)
1.0
0.8
0.6
0.4
0.2
2
4
6
8
10
12
14
16
retention time (min)


A.
NIaPro
N
P6
P5
P4
P3
P2
P1
P1'
P2'
P3'
C
substrate amino acids
Consensus X X V/I X X Q/E A/S/G X X
(86%) (98%) (92%)
UCBSV C Y V D T Q T K D
CBSV C Y I D L Q V V D
EuRV I E V Y L R A L E
B.
2xMyc
NIb
HAM1
CP
VDTQ/T
IDVQ/A
UCBSV-HAM1-2xMyc
2xMyc
NIb
HAM1T1X
CP
VDTQ/X
IDVQ/A
UCBSV-HAM1T1X-2xMyc
X= A, S, P
C.
UCBSV
UCBSV-HAM1-2xMyc
Untreated
D.
UCBSV-
-
HAM1-2xMyc
1 2 3 1 2 3 U
NIb-HAM1-2xMyc
HAM1-2xMyc
93
72
53
42
31
24
αMyc
αCP
Stained membrane
E.
UCBSV-HAM1-2xMyc
WT T1A T1S T1P
1 2 1 2 1 2 1 2 U
NIb-HAM1-2xMyc
HAM1-2xMyc
93
72
53
42
31
24
*
αMyc
αCP
Stained membrane


A.
UCBSV-HAM1-2xMyc
UCBSV-HAM1T1A-2xMyc
UCBSV-HAM1T1P-2xMyc
Untreated
B.
UCBSV-HAM1-2xMyc
WT
T1A
T1P
Untreated
1 2 3 1 2 3 1 2 3 1 2 3
C-
V
H
C.
UCBSV-HAM1T1A-2xMyc
Plant 1
D T Q A K D
D T Q A K G
60 dpi
180 dpi
D.
UCBSV-HAM1-2xMyc
U
WT
T1A
T1A/D3G
1 2 1 2 1 2 1 2 3 4
Nib-HAM1-2xMyc
HAM1-2xMyc
93
72
53
42
31
24
αMyc
αCP
Stained membrane


A.
4xMyc
tNOS
NIb$_C$
HAM1
CP$_N$
p35S
pGWB718-NIb$_C$-HAM1-CP$_N$ (UCBSV, CBSV and EuRV)
tNOS
VPg
NIaPro
p35S
pGWB702Ω-NIa (UCBSV, CBSV and EuRV)


B.
NIb$_C$-HAM1-CP$_N$
NIa
- + - + - +
4xMyc-NIb$_C$-HAM1-CP$_N$
4xMyc-NIb$_C$-HAM1
4xMyc-NIb$_C$
αMyc
Stained membrane
UCBSV
CBSV
EuRV