

# Geminiviruses encode additional small proteins with specific subcellular localizations and virulence function

Pan Gong <sup>1,6</sup>, Huang Tan <sup>2,3,6</sup>, Siwen Zhao<sup>1</sup>, Hao Li<sup>1</sup>, Hui Liu<sup>4</sup>, Yu Ma<sup>2,3</sup>, Xi Zhang<sup>2,3</sup>, Junjie Rong<sup>2,3</sup>, Xing Fu<sup>2</sup>, Rosa Lozano-Durán <sup>2,5</sup>✉, Fangfang Li<sup>1</sup>✉ & Xueping Zhou <sup>1,4</sup>✉

Geminiviruses are plant viruses with limited coding capacity. Geminivirus-encoded proteins are traditionally identified by applying a 10-kDa arbitrary threshold; however, it is increasingly clear that small proteins play relevant roles in biological systems, which calls for the reconsideration of this criterion. Here, we show that geminiviral genomes contain additional ORFs. Using tomato yellow leaf curl virus, we demonstrate that some of these small ORFs are expressed during the infection, and that the encoded proteins display specific subcellular localizations. We prove that the largest of these additional ORFs, which we name V3, is required for full viral infection, and that the V3 protein localizes in the Golgi apparatus and functions as an RNA silencing suppressor. These results imply that the repertoire of geminiviral proteins can be expanded, and that getting a comprehensive overview of the molecular plant-geminivirus interactions will require the detailed study of small ORFs so far neglected.

<sup>1</sup>State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China.

<sup>2</sup>Shanghai Center for Plant Stress Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China.

<sup>3</sup>University of the Chinese Academy of Sciences, Beijing, China. <sup>4</sup>State Key Laboratory of Rice Biology, Institute of Biotechnology, Zhejiang University, Hangzhou, Zhejiang, China. <sup>5</sup>Department of Plant Biochemistry, Center for Plant Molecular Biology (ZMBP), Eberhard Karls University, Tübingen, Germany.

<sup>6</sup>These authors contributed equally: Pan Gong, Huang Tan. ✉email: [rosa.lozano-duran@zmbp.uni-tuebingen.de](mailto:rosa.lozano-duran@zmbp.uni-tuebingen.de); [lifangfang@caas.cn](mailto:lifangfang@caas.cn); [zzhou@zju.edu.cn](mailto:zzhou@zju.edu.cn)

Viruses are intracellular parasites that heavily rely on the host cell machinery to complete their infectious cycle. Most viruses have small genome sizes, with the concomitant limitation in coding capacity; in order to overcome the restrictions imposed by their reduced proteome, viruses have evolved to encode multifunctional proteins that efficiently target hub proteins in their host cells (reviewed in<sup>1,2</sup>). Nevertheless, higher numbers of virus-encoded proteins might enable more sophisticated infection mechanisms, and therefore maximization of the coding space would be expected to be an advantage to the pathogen.

Geminiviruses are a family of plant viruses with circular, single-stranded (ss) DNA genomes causing devastating diseases in crops around the globe. This family includes nine genera, based on host range, insect vector, and genome structure: *Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus*, *Turncurtovirus*, *Capulavirus*, and *Grablovirus*<sup>3</sup>; most species described to date belong to the genus *Begomovirus*. Members of this family have small genomes, composed of one or two DNA molecules of less than 3 Kb each, in which the use of coding space is optimized by bidirectional and partially overlapping open reading frames (ORFs): in one <3 Kb molecule, geminiviruses contain up to 7 ORFs, with a known maximum of 8 viral proteins per virus. The geminiviral infection cycle is complex, and multiple steps remain to be fully elucidated. Following transmission by an insect vector, the geminiviral DNA genome must be released from the virion and reach the nucleus, where it will be converted into a double-stranded (ds) DNA replicative intermediate; this dsDNA molecule will serve as template for the transcription of viral genes, including the replication-associated protein (Rep), which reprograms the cell cycle and recruits the host DNA replication machinery. Rolling-circle replication ensues, by which new ssDNA copies of the viral genome are produced. Eventually, the virus must move intracellularly, intercellularly, and systemically, invading new cells and making virions available for acquisition by the vector. In order to accomplish a successful infection, geminiviruses must tailor the cellular environment to favor their replication and spread; for this purpose, they modify the transcriptional landscape of the infected cell, re-direct post-transcriptional modifications, and interfere with hormone signaling, among other processes (reviewed in<sup>4–6</sup>), ultimately suppressing anti-viral defenses, creating conditions favorable to viral replication, and manipulating plant development. Although geminivirus-encoded proteins are described as multifunctional, how the plethora of tasks required for a fruitful infection can be performed by only 4–8 proteins is an intriguing biological puzzle. Whether members of this family encode additional small proteins, below the arbitrary 10 kDa threshold established following identification of the first geminivirus species, remains elusive.

Tomato yellow leaf curl virus (TYLCV) is a monopartite begomovirus and the causal agent of the destructive tomato leaf curl disease<sup>7</sup>. The TYLCV genome contains six known open reading frames (ORFs), encoding the capsid protein (CP)/V1 and V2 in the virion strand, and C1/Rep, C2, C3, and C4 in the complementary strand. Rep creates a cellular environment permissive for viral DNA replication and attracts the DNA replication machinery to the viral genome (reviewed in<sup>8</sup>); C2 suppresses post-transcriptional gene silencing (PTGS)<sup>9</sup>, protein ubiquitination<sup>10</sup> and jasmonic acid (JA) signaling<sup>10,11</sup>; C3 interacts with PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA), the NAC family transcription factor SINAC1, and the regulatory subunits of DNA polymerases  $\alpha$  and  $\delta$  to enhance viral replication<sup>12–14</sup>; C4 is a symptom determinant, interferes with the intercellular movement of PTGS, and hampers salicylic acid (SA)-dependent defenses<sup>9,15,16</sup>; the CP forms the viral capsid and is

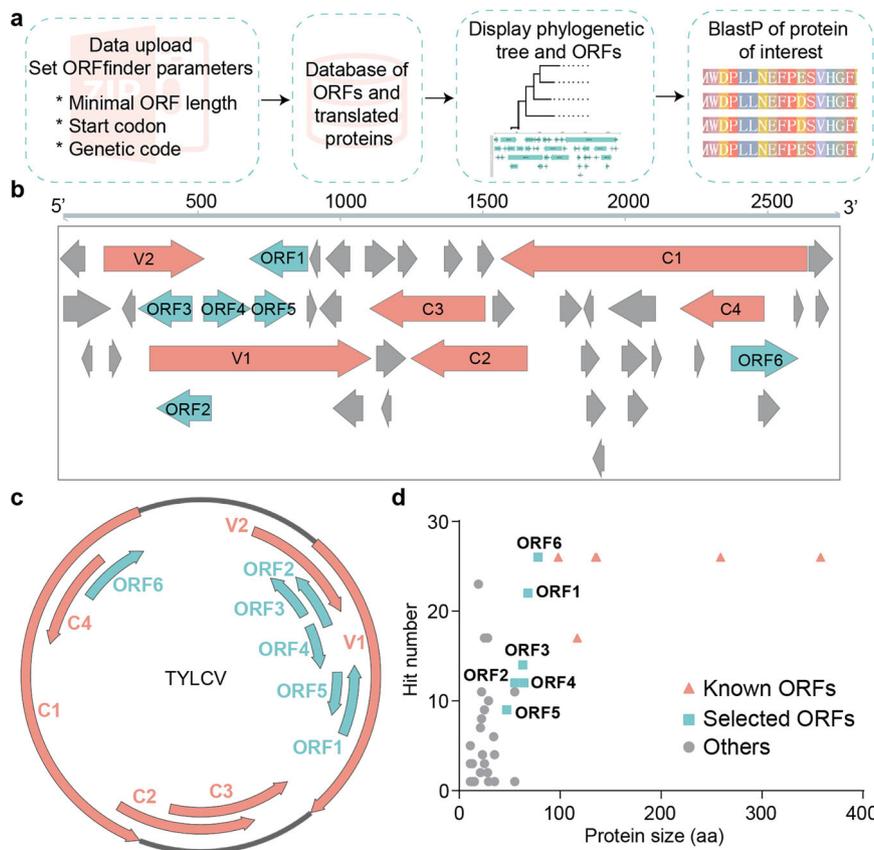
essential for the transmission by the insect vector, and shuttles the viral DNA between the nucleus and the cytoplasm<sup>17–24</sup>; V2 is a strong suppressor of PTGS as well as transcriptional gene silencing (TGS), and it mediates the nuclear export of CP<sup>25–29</sup>. Interestingly, a recent report identified 21 transcription initiation sites within the TYLCV genome by taking advantage of cap-snatching by rice stripe virus (RSV) in the experimental *Solanaceae* host *Nicotiana benthamiana*, suggesting that transcripts beyond those encoding these known ORFs might exist<sup>30</sup>. This idea is further indirectly supported by the fact that attempts at knocking-in tags in geminiviral genomes have so far been fruitless.

Here, we report that geminiviral genomes contain additional ORFs besides the canonical ones described to date. These previously neglected ORFs frequently encode proteins that are phylogenetically conserved. Using the geminivirus TYLCV as an example, we show that some of these ORFs are transcribed during the viral infection, and that the proteins they encode accumulate in the plant cell and show specific subcellular localizations and distinctive features. Moreover, we demonstrate that one of these additional ORFs, which we have named V3 and is conserved in begomoviruses, is essential for full infectivity in *N. benthamiana* and tomato, and encodes a Golgi-localized protein that acts as a suppressor of PTGS and TGS. Taken together, our results indicate that geminiviruses encode additional proteins to the ones described to date, which may largely expand the geminiviral proteome and hence the intersection of these viruses with the host cell.

## Results

**The TYLCV genome contains additional conserved open reading frames.** It is increasingly clear that small proteins (<100 aa) are prevalent in eukaryotes, including plants, and have biological functions (reviewed in<sup>31–33</sup>). In order to explore whether geminiviral genomes may contain additional ORFs encoding small proteins of predictable functional relevance, we designed a tool that we called ViralORFfinder; this tool uses the ORFfinder from NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>) to identify ORFs in an inputted subset of DNA sequences (geminiviral genomes, in this case) and creates a small database with the translated protein sequences, which can be used to BLAST a protein of choice, therefore assessing conservation among the subset of selected species (Fig. 1a). The distribution of the protein of interest is then displayed in a phylogenetic tree of the inputted viral species, generated based on the DNA sequences provided. Using ViralORFfinder, additional ORFs can be consistently predicted in geminiviruses of different genera, as illustrated for bipartite begomoviruses (Supplementary Fig. 1; Supplementary Table 2), curtoviruses (Supplementary Fig. 2; Supplementary Table 3), and mastreviruses (Supplementary Fig. 3; Supplementary Table 4); the proteins encoded by some of these ORFs are conserved among species.

We then used ViralORFfinder to identify additional proteins encoded by monopartite begomoviruses causing tomato leaf curl disease isolated from different regions of the world (see “Methods” section; Supplementary Fig. 4; Supplementary Table 5), and identify those present in TYLCV and conserved in other species. As shown in Fig. 1b, 43 ORFs encoding proteins/peptides of >10 aa were identified in the TYLCV genome. Interestingly, a significant correlation can be found between the size of the encoded proteins and their representation in the selected subset of species, with the six larger proteins (>10 kDa) present in all of them (Supplementary Fig. 4b, pink). For further analyses, we selected 6 ORFs based on size and prevalence (Supplementary Fig. 4b, blue), named ORF1–6; the position of these ORFs in the



**Fig. 1** The TYLCV genome contains additional conserved open reading frames. **a** Working pipeline of ViralORFFinder, a web-based tool for ORF prediction and protein conservation analysis. **b** Schematic view of predicted ORFs ( $\geq 30$  nt) in the TYLCV genome. **c** Genome organization of TYLCV; arrows indicate ORFs. In **(b)** and **(c)**, pink arrows represent the six known ORFs (C1, C2, C3, C4, V1, and V2), while blue arrows represent the six additional ORFs described in this work (ORF1-6). **d** Correlation between the size of proteins encoded by the ORFs in the TYLCV genome and their representation in the selected subset of begomoviruses (see Supplementary Table 6). The TYLCV isolate used in these experiments is TYLCV-Alm.

TYLCV genome is shown in Fig. 1c. Of note, the proteins encoded by these ORFs are also conserved in other members of the *Begomovirus* genus, both bipartite and monopartite, infecting a broad range of hosts (Supplementary Fig. 5–7; Supplementary Table 6; Fig. 1d).

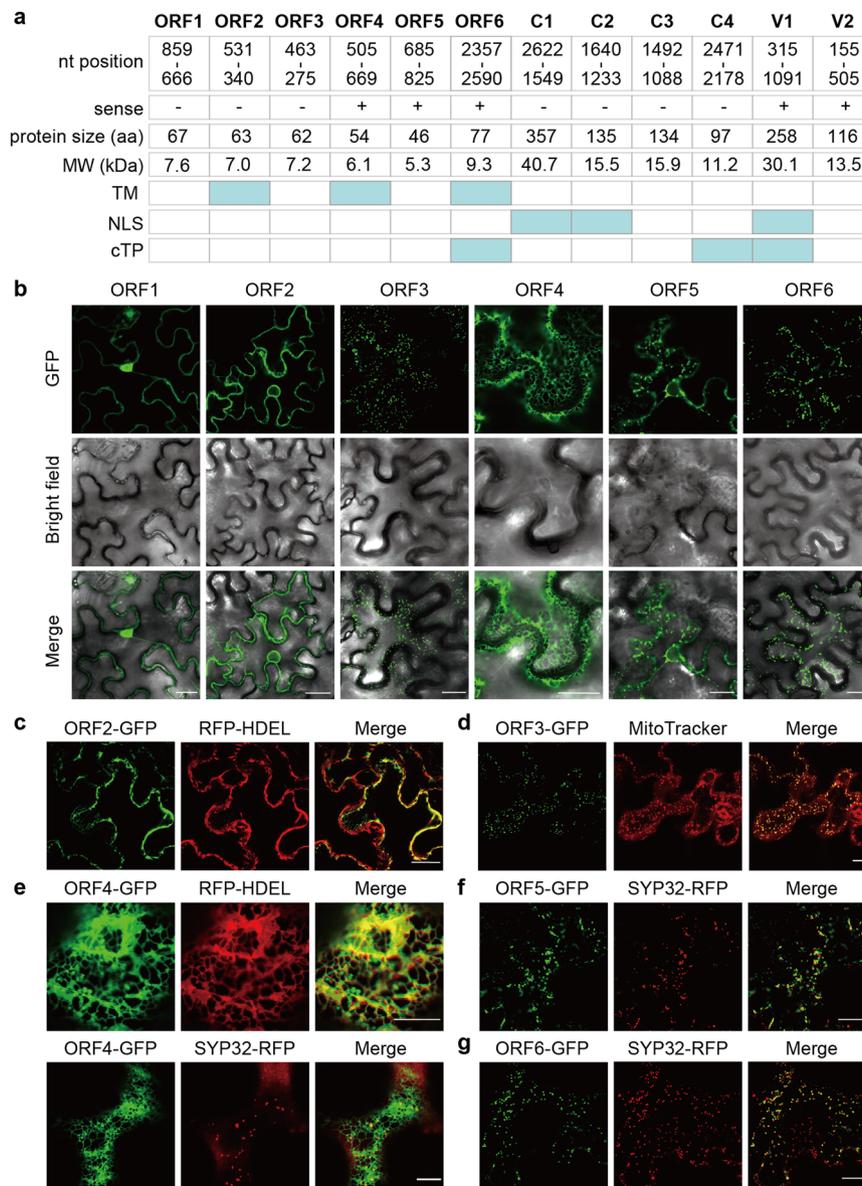
**Additional ORFs in TYLCV encode proteins with predicted domains and specific subcellular localizations.** With the aim of gaining insight into the properties of the proteins encoded by the additional ORFs from TYLCV, we investigated the presence in their sequence of predicted domains or signals, namely transmembrane domains (TM), nuclear localization signals (NSL), and chloroplast transit peptides (cTP). As shown in Fig. 2a, while none of these proteins contains an NLS, and only one of them contains a cTP, three of them contain a predicted TM, which is not present in any of the previously characterized proteins.

We then cloned ORF1-6, fused them to the GFP gene, and transiently expressed them in *N. benthamiana* leaves. Confocal microscopy indicates that these fusion proteins present specific subcellular localizations (Fig. 2b). Whereas the ORF1-encoded protein is mostly nuclear, co-expression with marker proteins or dyes unveils that the ORF2-encoded protein localizes in the endoplasmic reticulum (ER), as demonstrated by the co-localization with RFP-HDEL; the ORF3-encoded protein in mitochondria, as demonstrated by the co-localization with MitoTracker; the ORF4-encoded protein in the ER and the Golgi apparatus, as demonstrated by the partial co-localization with RFP-HDEL and SYP32-RFP; and the ORF5- and ORF6-encoded

proteins mostly in Golgi, as demonstrated by the partial co-localization with SYP32-RFP (Fig. 2c–g; Supplementary Fig. 8). The specific subcellular localization exhibited by each of these proteins suggests that, despite their small size (5.3–9.3 kDa; Fig. 2a), either their sequence contains the appropriate targeting signals, or they interact with plant proteins that enable their precise targeting in the cell.

A prerequisite for these ORFs to have a biological function is their expression in the context of the viral infection. Therefore, we cloned the ~500-bp TYLCV genomic sequence upstream of each ATG (pORF1-6) before the GFP reporter gene and tested its promoter activity in transiently transformed *N. benthamiana* leaves. As shown in Supplementary Fig. 9, none of these sequences could drive GFP expression in the absence of the virus, but pORF1, pORF2, pORF4, and pORF5 could when the virus was present; the sequence upstream of the C4 ORF was used as positive control, and could activate GFP expression both in the presence and absence of TYLCV. The promoter activity of these upstream sequences in infected cells strongly suggests that at least ORF1, 2, 4, and 5 are expressed during the viral infection.

**The V3 protein from TYLCV is a Golgi-localized silencing suppressor required for full infection.** ORF6 is the largest of the newly described ORFs in the TYLCV genome, and the protein it encodes displays the highest degree of conservation in a selected subset of 26 representative begomoviruses (Supplementary Fig. 7). Therefore, we decided to further characterize this ORF as a proof-of-concept of the potential biological roles of additional



**Fig. 2 Additional open reading frames in the TYLCV genome encode proteins with predicted domains and specific subcellular localizations. a**

Nucleotide position of the six additional ORFs and the six known ORFs in the TYLCV genome, size (in aa) and predicted molecular weight (MW; in kDa) of the corresponding encoded proteins, and domains or signals predicted in the protein sequence. TM: transmembrane domain; NLS: nuclear localization signal; cTP: chloroplast transit peptide. **b** Subcellular localization of the proteins encoded by ORF1-6 fused to GFP at their C-terminus transiently expressed in *N. benthamiana* leaves. Scale bar: 25  $\mu$ m. **c** Co-localization of ORF2-GFP with the ER marker RFP-HDEL. Scale bar: 25  $\mu$ m. **d** Co-localization of ORF3-GFP with the mitochondrial stain MitoTracker Red. Scale bar: 25  $\mu$ m. **e** Co-localization of ORF4-GFP with the ER marker RFP-HDEL and the cis-Golgi marker SYP32-RFP. Scale bar: 10  $\mu$ m. **f** Co-localization of ORF5-GFP with the cis-Golgi marker SYP32-RFP. Scale bar: 25  $\mu$ m. **g** Co-localization of ORF6-GFP with the cis-Golgi marker SYP32-RFP. Scale bar: 25  $\mu$ m. These experiments were repeated at least three times with similar results; representative images are shown. The TYLCV isolate used in these experiments is TYLCV-Alm.

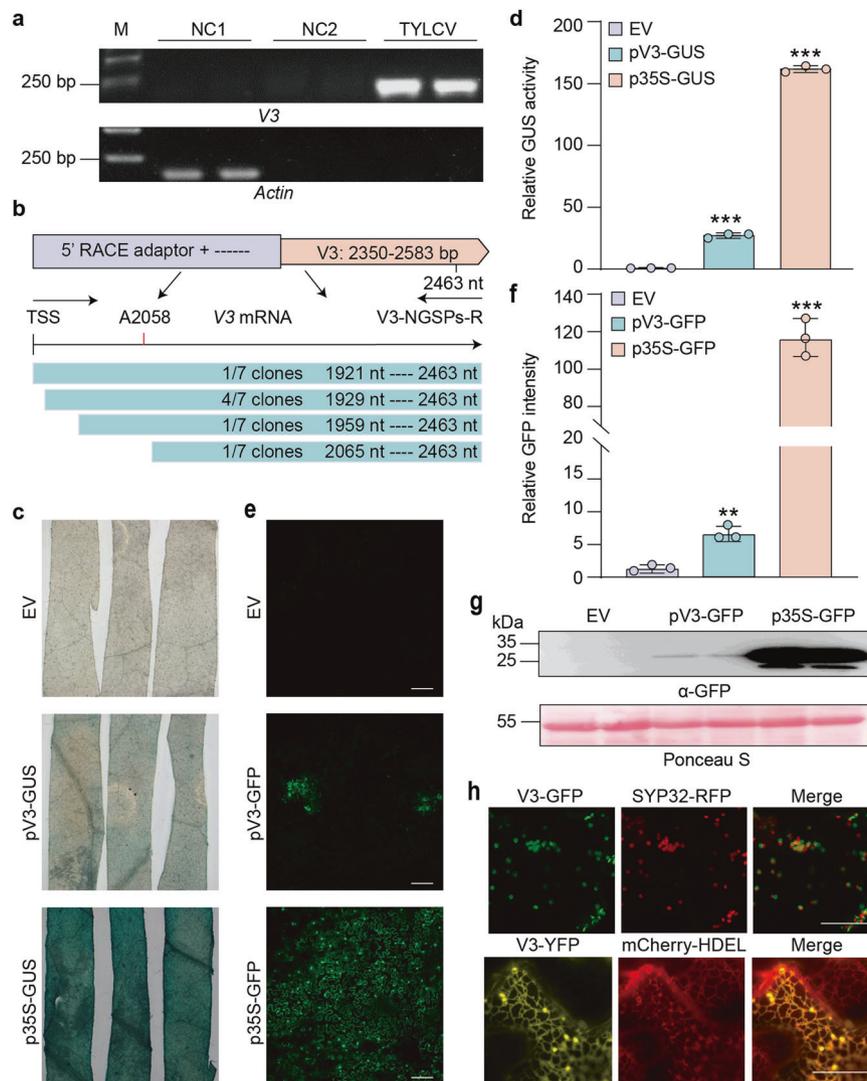
ORFs. Hereafter, ORF6 will be referred to as V3, since it is the third ORF on the viral strand in the TYLCV genome.

The V3 ORF is located in positions 2350–2583 of the TYLCV genome (TYLCV-BJ; Fig. 1d), and encodes a 77-amino acid protein. In different begomovirus species, the V3 ORF ranges from 87 to 234 nt, and the protein it encodes presents a high degree of similarity, with 6 residues completely conserved (Supplementary Fig. 7).

In order to determine whether V3 is transcribed during the viral infection, we checked if the corresponding transcript was present in TYLCV-infected samples: as shown in Fig. 3a, the V3 transcript was found upon TYLCV infection, but not in uninfected plants. 5' rapid amplification of cDNA ends (RACE)

was then used to identify the transcriptional initiation site of V3, which was found to be located between 176 and 421 nt upstream of the start codon (Fig. 3b). Interestingly, most of the sites identified by RACE (7 out of 11) are close to A2058, which was previously isolated by cap-snatching of RSV<sup>30</sup>.

The finding that a transcript corresponding to the V3 ORF can be identified in infected samples strongly suggests that the sequence upstream of this ORF must act as a promoter. Since the 500 bp fragment previously tested (Supplementary Fig. 9) did not show promoter activity, we tested a larger, 833-nt sequence upstream of the V3 start codon, which was cloned before the GUS or GFP reporter genes. As shown in Fig. 3c, d, this viral sequence could activate GUS expression, leading to detectable GUS activity,

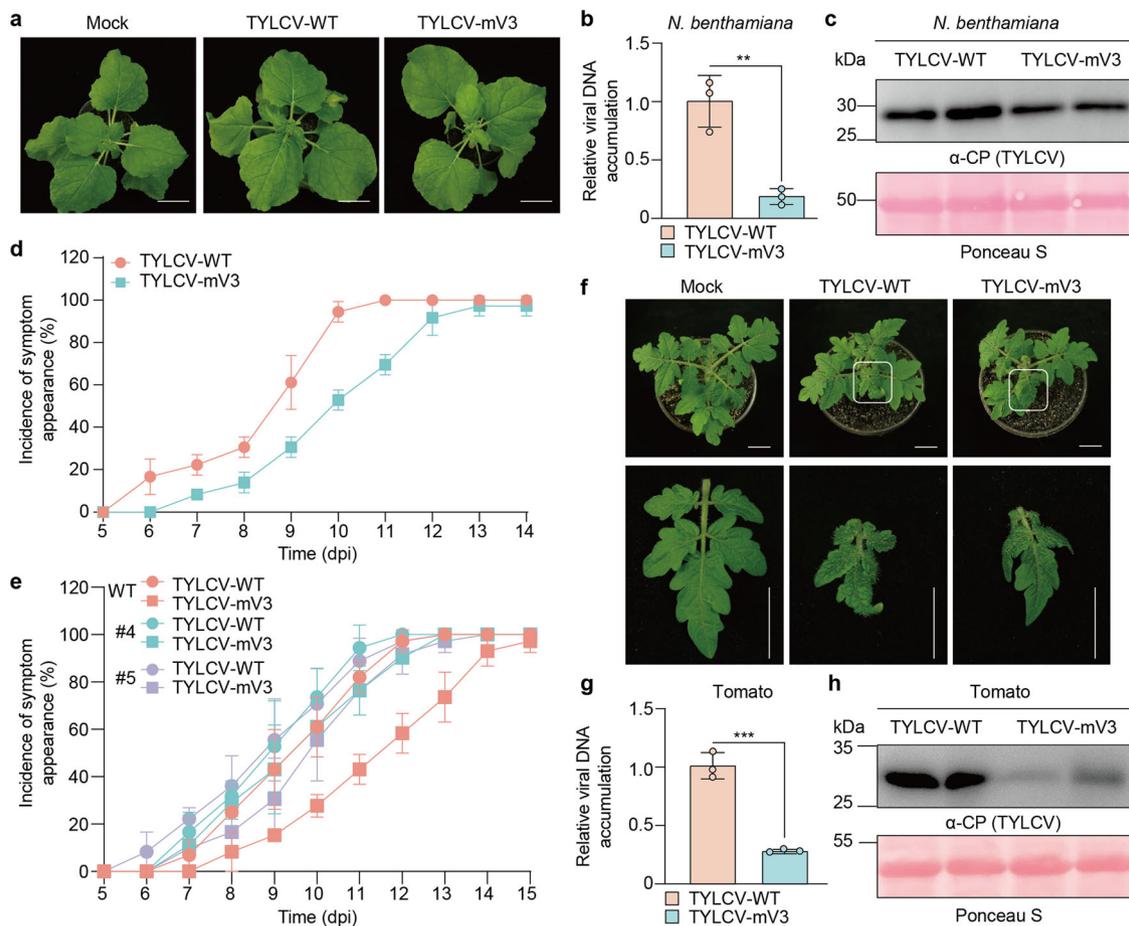


**Fig. 3 ORF6/V3 is expressed during the infection and encodes a Golgi-localized protein.** **a** RT-PCR analysis of V3 transcripts from TYLCV-infected or uninfected *N. benthamiana* plants. M: DNA ladder marker. NC1: negative control 1 (reverse-transcription of total RNA extracted from uninfected plants with RT Primers). NC2: negative control 2 (reverse-transcription of total RNA extracted from uninfected plants with V3-specific primers). TYLCV: reverse-transcription of total RNA extracted from TYLCV-infected plants with V3-specific primers. This experiment was repeated three times with similar results; representative images are shown. **b** Transcriptional start site analysis of TYLCV V3 by 5' RACE. TSS: transcription start site. A2058: V3 TSS captured by RSV cap-snatching (Lin et al., 2017). **c** Activity of pV3 promoter (and p35S promoter as positive control) in promoter-GUS fusions in transiently transformed *N. benthamiana* leaves at 2 dpi. EV: empty vector. This experiment was repeated three times with similar results; representative images are shown. **d** Quantification of relative GUS activity in samples from (c). Data are the mean of three independent biological replicates. Error bars represent SD. An asterisks indicate a statistically significant difference according to unpaired Student's *t* test (two-tailed), \*\*\*  $p < 0.001$ . **e** Activity of pV3 promoter (and p35S promoter as positive control) in promoter-GFP fusions in transiently transformed *N. benthamiana* leaves at 2 dpi. Scale bar: 100  $\mu\text{m}$ . EV: empty vector. This experiment was repeated three times with similar results; representative images are shown. **f** Quantification of relative GFP intensity in samples from (e). Data are the mean of three independent biological replicates. Error bars represent SD. Asterisks indicate a statistically significant difference according to unpaired Student's *t* test (two-tailed), \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . **g** Western blot analysis of GFP protein from (e). Ponceau S staining of the large RuBisCO subunit serves as loading control. EV: empty vector. This experiment was repeated three times with similar results; representative images are shown. **h** Co-localization of V3-GFP with the cis-Golgi maker SYP32-RFP (upper panel) and co-localization of V3-YFP with the ER maker mCherry-HDEL (lower panel). Scale bar: 20  $\mu\text{m}$ . The TYLCV isolate used in these experiments is TYLCV-BJ. The original data from all experiments and replicates can be found in the Source data file.

and it could also drive expression of GFP (Fig. 3e–g), albeit more weakly than the 35S promoter. Taken together, these results confirm that the V3 ORF can be expressed in planta from its genomic context.

Given that protein function is tightly linked to spatial location, we then decided to analyse the subcellular localization of the V3 protein in detail. As can be seen in Fig. 3h, V3 is a Golgi-localized protein; this localization does not change in the presence of the

virus (Supplementary Fig. 10a). A closer observation confirms that V3 is localized to cis-Golgi, as shown by the co-localization of V3-GFP with the markers Man49-mCherry and SYP32-RFP (Fig. 3h; Supplementary Fig. 10b); nevertheless, the YFP fusions YFP-V3 and V3-YFP can also partially co-localize with the endoplasmic reticulum (ER), as indicated by their partial co-localization with the ER marker mCherry-HDEL or RFP-HDEL (Fig. 3h; Supplementary Fig. 10c, d), which may



**Fig. 4 V3 is required for full TYLCV infection in *N. benthamiana* and tomato.** **a** Symptoms of *N. benthamiana* plants inoculated with wild-type TYLCV (TYLCV-WT), a V3 null mutant TYLCV (TYLCV-mV3), or mock-inoculated (pCAMBIA2300 empty vector), at 10 dpi. Bar = 2 cm. **b** Viral DNA accumulation in TYLCV-WT- and TYLCV-mV3-infected plants in **(a)**, measured by qPCR. Data are the mean of three independent biological replicates. Error bars represent SD. An asterisk indicates a statistically significant difference according to unpaired Student's *t* test (two-tailed), \*\*\*  $p < 0.01$ . 25S rRNA was used as internal reference. **c** Western blot showing TYLCV CP accumulation in systemic leaves of TYLCV-WT- and TYLCV-mV3-infected plants from **(a)**. Ponceau S staining of the large RuBisCO subunit serves as loading control. This experiment was repeated three times with similar results; representative images are shown. **d, e** Incidence of symptom appearance in WT **(d)** or V3 transgenic **(e)** *N. benthamiana* plants infected with TYLCV-WT or TYLCV-mV3. Data are the mean of three independent biological experiments. Error bars represent SD. At least 8 plants were used per viral genotype and experiment. For images of symptoms in the V3 transgenic lines, see Supplementary Fig. 10d. **f** Symptoms of tomato plants inoculated with WT TYLCV (TYLCV-WT), a V3 null mutant TYLCV (TYLCV-mV3), or mock-inoculated (pCAMBIA2300 empty vector), at 10 dpi. Bar = 2 cm. **g** Viral DNA accumulation in TYLCV-WT- and TYLCV-mV3-infected plants from **(f)**, measured by qPCR. Data are the mean of three independent biological replicates. Error bars represent SD. An asterisk indicates a statistically significant difference according to unpaired Student's *t* test (two-tailed), \*\*\*  $p < 0.001$ . 25S rRNA was used as internal reference. **h** Western blot showing TYLCV CP accumulation in systemic leaves of TYLCV-WT- and TYLCV-mV3-infected plants from **(f)**. Ponceau S staining of the large RuBisCO subunit serves as loading control. This experiment was repeated three times with similar results; representative images are shown. The TYLCV isolate used in these experiments is TYLCV-BJ. The original data from all experiments and replicates can be found in the Source data file.

reflect the transition of the protein from the ER to the Golgi apparatus.

With the aim of assessing the biological relevance of the V3 protein for the TYLCV infection, we next generated a mutated infectious clone carrying a T2351C substitution in the V3 ATG, hence impairing the production of the V3 protein. Since the V3 ORF overlaps with the Rep/C1 and C4 ORFs, nt replacements in the start codon of the V3 ORF necessarily affect the protein sequence of the resulting Rep or C4 proteins; the chosen change results in a I89V substitution in the Rep/C1 protein, with no change in C4. This mutant infectious clone is hereafter referred to as TYLCV-mV3.

TYLCV-mV3 was then inoculated into *N. benthamiana* plants, and its performance compared to that of the wild-type (WT) virus (TYLCV-WT). At 10 days post-inoculation (dpi), TYLCV-

mV3-infected plants displayed mild leaf curling symptoms and presented lower viral DNA load compared to plants inoculated with TYLCV-WT (Fig. 4a, b), which correlated with a lower accumulation of CP (Fig. 4c). These differences apparently result from a delay in the infection, measured as symptom appearance (Fig. 4d). To evaluate the potential functional impact of the I89V substitution in Rep/C1 on the viral infection and disentangle this effect to that derived from the absence of V3, transgenic *N. benthamiana* lines expressing V3-YFP under a 35S promoter were generated, and a complementation assay was performed. Of note, the V3-expressing plants do not display obvious developmental abnormalities (Supplementary Fig. 11a); the expression of V3-YFP was confirmed by qRT-PCR and western blot (Supplementary Fig. 11b, c). As shown in Fig. 4e and Supplementary Fig. 11d, transgenic expression of V3-YFP could fully

complement the lack of V3 in the TYLCV-mV3 clone, measured as incidence and severity of symptom appearance, indicating that the slower progression of the infection observed upon inoculation with the V3 null mutant is due to the lack of this protein, and not to a suboptimal performance of Rep-I89V. Next, we evaluated the virulence of TYLCV-mV3 on tomato, the virus' natural host. As previously observed in *N. benthamiana*, the lack of V3 resulted in lower viral load and CP accumulation, and milder symptoms at 10 dpi (Fig. 4f–h), confirming that V3 plays a relevant role in the viral infection that is not restricted to *N. benthamiana*.

Heterologous expression from a potato virus X (PVX)-derived vector and quantification of the impact on PVX pathogenicity is a widely used approach to test virulence activity of viral genes of interest. Confirming a contribution of V3 to virulence, the presence of this gene in PVX-V3 led to an exacerbation of disease symptoms in inoculated *N. benthamiana* plants compared to PVX alone at 10 and 30 dpi, with a concomitant higher accumulation of the PVX CP at 30 dpi (Supplementary Fig. 12);  $\beta$ C1, a symptom determinant encoded by tomato yellow leaf curl China betasatellite (TYLCCNB), was used as positive control. Infection by PVX-V3, however, did not lead to H<sub>2</sub>O<sub>2</sub> accumulation or cell death (Supplementary Fig. 12c).

It has been previously established that a high correlation exists between the ability of a viral protein to suppress RNA silencing and its capacity to enhance the severity of the PVX infection. RNA silencing is conserved in eukaryotes, and is considered the main anti-viral defense mechanism in plants<sup>34–36</sup>. Supporting this notion, virtually all plant viruses described to date encode at least one protein with RNA silencing suppression activity<sup>36,37</sup>. With the aim to test if V3 can suppress PTGS, we transiently expressed GFP from a 35S promoter in leaves of transgenic 16c *N. benthamiana* plants, harboring a 35S:GFP cassette<sup>38</sup>, in the presence or absence of V3; the well-described silencing suppressor P19 from tomato bushy stunt virus (TBSV) was used as positive control. At 4 dpi, fluorescence had already substantially decreased when no viral protein was co-expressed, but was maintained in the samples with P19 or Myc-V3 (Fig. 5a). Western blot and qRT-PCR were used to confirm that both the GFP protein as well as the corresponding mRNA accumulated to higher levels in tissues expressing P19 or Myc-V3 (Fig. 5b, c). At 20 dpi, systemic leaves of 16c plants inoculated with the constructs to express either of the viral proteins remained green, while fluorescence had disappeared in control plants as a result of systemic silencing (Fig. 5a).

The ability of V3 to suppress PTGS was further confirmed by expressing this protein from a PVX-based vector; in this case,  $\beta$ C1, which also functions as silencing suppressor, was used as positive control. Transient co-transformation of a PVX infectious clone together with a 35S:GFP cassette in leaves of 16c plants led to weak fluorescence in the infiltrated tissues at 7 dpi, and systemic silencing at 20 dpi; in stark contrast, co-transformation with PVX- $\beta$ C1 or PVX-V3 resulted in the maintenance of strong fluorescent signal at 7 dpi, and absence of systemic silencing (Fig. 5d). Neither  $\beta$ C1 nor V3 enhanced the local accumulation of PVX, as indicated by the accumulation of the PVX CP protein (Fig. 5e), hence ruling out an indirect effect of these proteins on the endogenous ability of PVX to suppress silencing. Therefore, our results demonstrate that V3 from TYLCV can effectively suppress PTGS in plants.

Another level of RNA silencing is TGS, which acts through methylation of DNA at cytosine residues; TGS also acts as an anti-viral response in plants, and is particularly relevant against geminiviruses, which replicate their DNA genomes in the nucleus of the infected cell<sup>36,39</sup>. To investigate whether V3 can also act as a TGS suppressor, we inoculated 16-TGS plants, in which the GFP transgene is silenced due to methylation of the 35S

promoter<sup>40</sup>, with PVX or PVX-V3; PVX- $\beta$ C1 was used as positive control, since  $\beta$ C1 can also act as a TGS suppressor. At 10 dpi, green fluorescence could be observed in systemic leaves of 16-TGS plants inoculated with PVX- $\beta$ C1 and PVX-V3, as opposed to mock- or PVX-inoculated plants (Fig. 5f); this fluorescence persisted at 28 dpi (Fig. 5g). Visual assessment was confirmed at the molecular level by western blot (Fig. 5h, i), indicating that V3 can also suppress TGS in the host plant.

To confirm the effect of V3 on DNA methylation, the level of genome-wide methylation in transgenic V3-YFP plants was examined by digestion with a methylation-dependent restriction enzyme, *McrBC*<sup>41</sup>. As presented in Fig. 5j, genomic DNA from two independent V3-YFP lines was completely digested by the methylation-independent restriction endonuclease *DraI*, but only partially digested by *McrBC*, in sharp contrast to the genomic DNA from WT plants. This indicates a lower level of DNA methylation in transgenic plants expressing V3, further supporting a function of this viral protein as a TGS suppressor.

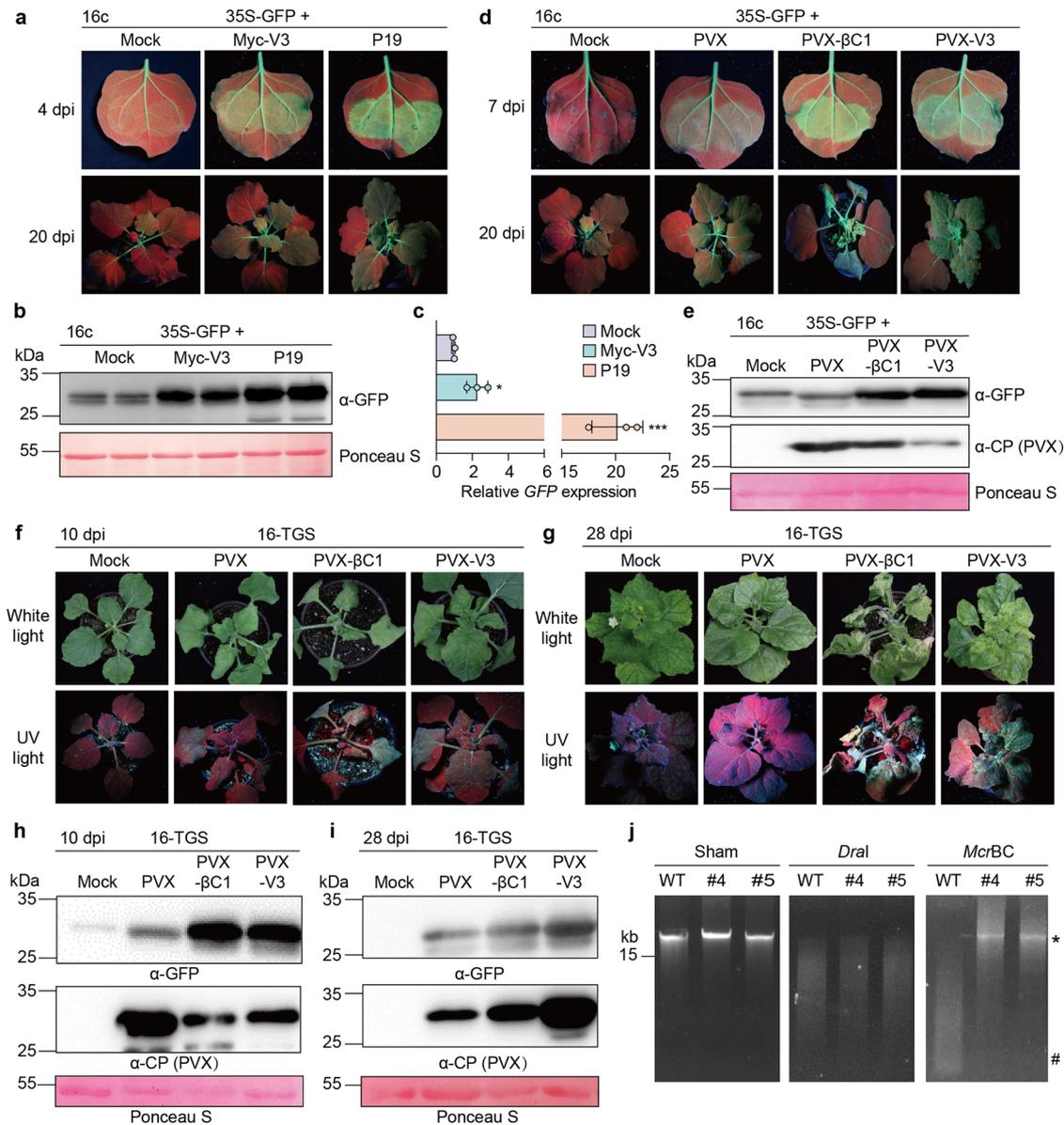
In summary, our results demonstrate that V3 is a newly described protein encoded by TYLCV, which preponderantly localizes in the cis-Golgi, significantly contributes to virulence, and functions as a suppressor of both PTGS and TGS.

## Discussion

Geminivirus-encoded proteins have so far been identified taking into consideration an arbitrary threshold of 10 kDa, below which potential proteins were discarded. However, it is increasingly clear that small proteins and peptides play relevant roles in biological systems, hence calling for the reconsideration of this criterion. Here, we show that geminiviral genomes contain additional ORFs beyond those previously described, at least some of which are conserved within members of a given genus, hinting at functional relevance. Using TYLCV as a model, we demonstrate that at least some of these conserved additional small ORFs are expressed during the infection, and that the proteins they encode localize in specific subcellular compartments. Interestingly, in this subset of selected proteins novel localizations, not described for any of the other TYLCV-encoded proteins, are represented, including the Golgi apparatus (ORF4, 5, and 6) and mitochondria (ORF 3). Also of note, some of these proteins (ORF2, 4, and 6) harbor transmembrane domains, which are not present in any of the “canonical” proteins from TYLCV. These results indicate that the repertoire of geminiviral proteins can be expanded, and that the new additions to the viral proteomes will most likely perform additional virulence functions and/or employ alternative molecular mechanisms to those exhibited by the previously described proteins. We might still be, therefore, far from getting a comprehensive overview of the plant-geminivirus molecular interaction landscape, which will require the detailed study of potentially multiple small ORFs that have so far been neglected.

In this work, we selected the largest of the additional ORFs found in TYLCV, which we name V3, to be used as a proof-of-concept of the potential functionality of previously overlooked viral small proteins. Strikingly, the lack of V3 negatively impacted the viral infection in two different hosts, *N. benthamiana* and tomato, demonstrating that V3 has a biological function. However, V3 is not essential, since a V3 null mutant virus can still accumulate and establish a systemic infection.

The V3 protein is mostly localized in the cis-Golgi, a novel localization for a geminivirus-encoded protein. Remarkably, V3 can suppress both PTGS and TGS, an ability that may underlie its virulence-promoting effect on TYLCV and PVX. One intriguing question is how V3 can exert this effect from the Golgi apparatus. Interestingly, connections have been drawn between the



**Fig. 5** V3 functions as a suppressor of PTGS and TGS. **a** Transgenic 16c *N. benthamiana* plants co-infiltrated with constructs to express GFP (35S-GFP) and Myc-V3, P19 (as positive control), or mock (empty vector, as negative control) at 4 dpi (upper panel) or 20 dpi (lower panel) under UV light. **b** Western blot showing the GFP accumulation in inoculated leaves from **(a)** at 4 dpi. The corresponding Ponceau S staining of the large RuBisCO subunit serves as a loading control. This experiment was repeated three times with similar results; representative images are shown. **c** Relative GFP mRNA accumulation in inoculated leaves from **(a)** at 4 dpi measured by qRT-PCR. Data are the mean of three independent biological replicates. Error bars represent SD. An asterisk indicates a statistically significant difference according to unpaired Student's *t* test (two-tailed), \*  $p < 0.5$ , \*\*\*  $p < 0.001$ . *NbActin2* was used as the internal reference. **d** Transgenic 16c *N. benthamiana* plants co-infiltrated with constructs to express GFP (35S-GFP) and PVX, PVX-V3, PVX-βC1 (as positive control), or mock (infiltration buffer, as negative control) at 7 dpi (upper panel) or 20 dpi (lower panel) under UV light. **e** Western blot showing the GFP accumulation in inoculated leaves from **(d)** at 7 dpi. The corresponding Ponceau S staining of the large RuBisCO subunit serves as loading control. This experiment was repeated three times with similar results; representative images are shown. **f, g** Symptoms of 16-TGS *N. benthamiana* plants infected with PVX, PVX-V3, PVX-βC1 (as positive control), or mock-inoculated under white light or UV light at 10 dpi **(f)** and 28 dpi **(g)**. **h, i** Western blot showing accumulation of GFP and PVX CP in systemically infected leaves from **(f)** and **(i)**. The corresponding Ponceau S staining of the large RuBisCO subunit serves as loading control. This experiment was repeated three times with similar results; representative images are shown. **j** DNA methylation analysis by restriction enzyme digestion in V3-YFP transgenic *N. benthamiana* plants. Genomic DNA extracted from WT *N. benthamiana* or two independent V3-YFP transgenic lines (#4 and #5) was digested with the methylation-dependent restriction enzyme *McrBC* and the methylation-insensitive enzyme *Dral*. "Sham" indicates a mock digestion with no enzyme added. The positions of undigested input genomic DNA is indicated with an asterisk; the position of the *McrBC*-digested products is indicated with a hashtag. This experiment was repeated three times with similar results; representative images are shown. The original data from all experiments and replicates can be found in the Source data file.

endomembrane system and RNA silencing (reviewed in Kim et al., 2014). In the model plant *Arabidopsis thaliana*, electron microscopy unveiled an enrichment of the Argonaute protein AGO1, a central player in PTGS, in close proximity to Golgi<sup>42</sup>; another Argonaute protein, AGO7, which has been recently shown to play a role in anti-viral defense<sup>43</sup>, also co-purifies with membranes and concentrates in cytoplasmic bodies linked to the ER/Golgi endomembrane system<sup>44</sup>. It seems therefore plausible that this subcellular localization is permissive for a direct targeting of RNA silencing, although further experiments will be necessary to uncover the exact molecular mechanism underlying this activity of V3.

PTGS and TGS are arguably the main plant defense mechanisms against geminiviruses. This idea is supported by the fact that, despite limited coding capacity, a given geminivirus species can produce several proteins that target these processes. TYLCV encodes at least three proteins capable of acting as PTGS suppressors, namely C2, C4, and V2<sup>9,16,29</sup>, and at least two, Rep and V2, capable of suppressing TGS<sup>25–27,45</sup>; these proteins exert their functions through non-overlapping mechanisms. Only one of the viral proteins, V2, has been described as a simultaneous suppressor of PTGS and TGS, as observed for V3. All of these silencing suppressors encoded by TYLCV are essential for the infection, although this may be due to their contribution to additional virulence activities, enabled by their multifunctional nature. Similarly, it is possible that V3 exerts additional, yet-to-be described functions during the viral infection.

Why a given geminivirus species needs multiple proteins targeting the same pathway is a thought-provoking question. Since the viral infection is a process, the temporal dimension must be considered: geminiviral genes can be classified as early or late, depending on the timing of their expression, with the strongest PTGS and TGS suppressor, V2, being a late gene, probably due to the requirement of another viral protein, C2, to activate its expression. The V3 promoter was active in the absence of the infection, which suggests that it can be expressed as an early gene. The early expression of V3 would guarantee the availability of a PTGS and TGS suppressor during the time between the synthesis of the dsDNA replicative intermediate and the expression of V2 later in the cycle, enhancing the effectiveness of viral accumulation and spread. Nevertheless, further work will be required to acquire a full understanding of the potential breadth of functions exerted by V3 and of the underpinning molecular mechanisms.

## Methods

**Plant materials.** *N. benthamiana* and *Solanum lycopersicum* (tomato) plants were grown in a growth chamber with 60% relative humidity and a 16 h:8 h light:dark, 25 °C:18 °C regime. The transgenic GFP 16c line was kindly provided by David C. Baulcombe (University of Cambridge, UK)<sup>38</sup>; 16-TGS plants were described previously<sup>46</sup>; the transgenic RFP-H2B line was kindly shared by Michael M. Goodin (University of Kentucky, USA)<sup>47</sup>.

The *Agrobacterium tumefaciens* strain EHA105 containing the pEarleygate101:V3-YFP construct was used to generate 35S:V3-YFP transgenic *N. benthamiana* lines by leaf disc transformation as previously described<sup>48</sup>. In brief, the pEarleygate101:V3-YFP vector was transformed into the *A. tumefaciens* EHA105 strain and co-cultured with *N. benthamiana* leaf discs. Transformants were selected on medium containing 0.005% Basta (Coolaber, China) and then transferred to shoot regeneration medium for rooting culture. Transgenic plants were transferred to soil and screened by PCR with specific primers (Supplementary Table 1). The accumulation of the V3 protein and the corresponding mRNA were confirmed by western blot and qRT-PCR, respectively.

**Agroinfiltration and viral inoculation.** For *A. tumefaciens*-mediated transient expression in *N. benthamiana*, plasmids were transformed into the EHA105 strain by the freeze-thaw method (Fig. 3a–h lower panel, 4, 5, Supplementary Fig. 10b, c, Supplementary Fig. 12) or to the GV3101 strain through electroporation (Figs. 2, 3h upper panel, Supplementary Fig. 8, Supplementary Fig. 9, Supplementary Fig. 10a, d). *Agrobacterium* cultures were resuspended in infiltration buffer (10 mM MgCl<sub>2</sub>, 10 mM MES (pH 5.6), and 100 μM acetosyringone) to an OD<sub>600</sub> = 0.1–0.5, then infiltrated into the adaxial side of four-week-old *N. benthamiana* leaves with a

needle-less syringe. For viral inoculation, two-week-old *N. benthamiana* plants or tomato plants at the two-leaf stage were infiltrated with *Agrobacterium* cultures carrying the TYLCV-BJ (MN432609) infectious clone. For experiments that required co-infiltration, *Agrobacterium* suspensions carrying different constructs were mixed at 1:1 ratio before infiltration.

**Plasmid construction.** Viral open reading frames (ORFs) from the TYLCV (TYLCV-Alm, Accession No. AJ489258) genome were cloned in the pENTR™/D-TOPO™ vector (Thermo Scientific) (for ORF1, ORF2, ORF4, ORF5, ORF6/V3) or the pDONR™/Zeo vector (Thermo Scientific) (for ORF3) without a stop codon. The binary plasmids to express GFP-fused viral proteins were generated by sub-cloning (Gateway LR reaction, Thermo Scientific) the viral ORFs from the corresponding entry vectors into pGWB505<sup>49</sup>. To generate the constructs to express V3-YFP, YFP-V3, or Myc-V3, the full-length V3 ORF was obtained from TYLCV (TYLCV-BJ, Accession No. MN432609) and recombined into the binary destination vectors pEarleygate101, pEarleygate104, or pEarleygate203, respectively<sup>50</sup>. Please note that the V3/ORF6 protein sequences from TYLCV-Alm and TYLCV-BJ are identical.

To generate the construct to express SYP32-RFP (as a cis-Golgi marker), the gene encoding SYNTAXIN OF PLANTS 32 (SYP32) was amplified from *A. thaliana* cDNA, cloned into the pENTR™/D-TOPO™ vector (Thermo Scientific), and sub-cloned into pGWB554<sup>49</sup> using a Gateway LR reaction (Thermo Scientific). The construct to express RFP-HDEL is from<sup>51</sup>, and the ones to express mCherry-HDEL and Man49-mCherry are from<sup>52</sup>.

To generate the constructs used for the analysis of promoter activity, the 500-nt sequence upstream of the ORF1, ORF2, ORF4, ORF5, ORF6/V3, and C4 ATG, or the 570-nt sequence upstream of the ORF3 ATG, were PCR-amplified and cloned into the pDONR™/Zeo vector (Thermo Scientific). These sequences were then sub-cloned into pGWB504<sup>49</sup> by a Gateway LR reaction to generate pORF-GFP. In addition, the 833-nt sequence upstream of the V3/ORF6 ATG was PCR-amplified and cloned into pINT121-GUS digested with *Hind*III and *Bam*HI to generate pINT121-V3-GUS, or into pCHF3-GFP digested with *Eco*RI and *Sac*I to generate pCHF3-V3-GFP (pV3-GFP) using In-Fusion Cloning according to the manufacturer's instructions. The full-length V3 ORF was inserted into the PVX vector digested with *Cla*I and *Sal*I to generate PVX-V3. The pCHF3-35S-GFP, pCHF3-p19, PVX-βC1, and a PVX-based expression vector for PTGS suppression assays have been described previously<sup>53,54</sup>, as has the PVX-based expression vector PVX-βC1 for TGS suppression assays<sup>55</sup>. All primers used in this study can be found in Supplementary Table 1.

**Sequence analysis.** The ViralORFfinder platform (see Code availability) was constructed by Shiny, an R package used to build interactive web applications (<https://shiny.rstudio.com>). The NCBI ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to identify ORFs for each uploaded virus, and the R package Gviz<sup>56</sup> was used for visualization; for Supplementary Figs. 1 and 3, 1.2-mer genomic sequences were used. To investigate the conservation of an ORF-encoded protein of choice, this tool creates a small database with the translated sequences from the inputted viral sequences, and BLASTp is used to identify proteins with high identity (e-value ≤ 0.05). Phylogenetic trees were obtained by the R package DECIPHER and visualized by ggtree. Multiple sequence alignments were constructed by ClustalW and visualized by the R package ggmsa (<https://cran.r-project.org/web/packages/ggmsa/vignettes/ggmsa.html>). Names and NCBI accession numbers of virus species used in this work are listed in Supplementary Tables 2–6.

**Prediction of domains or signals in protein sequences.** The prediction of transmembrane domains (TM) was performed by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) and Phobius (<https://phobius.sbc.su.se/>). The prediction of nuclear localization signal (NSL) was performed by cNLS Mapper ([http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS\\_Mapper\\_form.cgi](http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi)<sup>57,58</sup>). The prediction of chloroplast transit peptide (cTP) was performed by ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>).

**Confocal microscopy.** Confocal microscopy was performed using a Leica TCS SP8 point scanning confocal microscope (Figs. 2, 3h upper panel, Supplementary Figs. 8, 9, 10a, and 10d) or Zeiss LSM980 confocal microscope (Carl Zeiss) (Figs. 3e, 3h lower panel, Supplementary Fig. 10b, c), with the preset settings for GFP (Ex: 488 nm, Em: 500–550 nm), RFP (Ex: 561 nm, Em: 570–620 nm), YFP (Ex: 514 nm, Em: 515–570 nm), or mCherry (Ex: 594 nm, Em: 597–640 nm). For co-localization imaging, the sequential scanning mode was used.

**Mitochondrial staining.** To visualize mitochondria, staining with 250 nM MitoTracker® Red CMXRos (Invitrogen) was used. The chemical was infiltrated 10–30 min before imaging. The stock solution (1 mM) was prepared by dissolving the corresponding amount of MitoTracker® in dimethylsulfoxide (DMSO). The working solution was prepared by diluting the stock solution in water or infiltration buffer (10 mM MgCl<sub>2</sub>, 10 mM MES (pH 5.6), and 100 μM acetosyringone). The MitoTracker® red fluorescence was imaged using a Leica TCS SP8 point scanning confocal microscope with the following settings: Ex: 561 nm, Em: 570–620 nm.

**DNA and RNA extraction and qPCR/qRT-PCR.** Total DNA was extracted from infected plants using the CTAB method. Total RNA was extracted from collected plant leaves using TaKaRa MiniBEST Universal RNA Extraction Kit (Takara, Japan). 1 µg of total RNA was reverse-transcribed into cDNA using PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time) (Takara, Japan). qPCR or qRT-PCR was performed using TB Green® Premix Ex Taq™ II (Takara, Japan). 25S RNA and *NbActin2* were used as internal references for DNA and RNA normalization, respectively.

**5' rapid amplification of cDNA ends (RACE).** Total RNA extracted from TYLCV (TYLCV-BJ, Accession No. MN432609)-infected *N. benthamiana* plants was used for 5' RACE with SMARTer RACE 5'/3' Kit (Takara, Japan) according to the manual booklet.

**3,3'-diaminobenzidine (DAB) staining.** For DAB staining, systemic leaves of infected plants were incubated in DAB solution (1 mg/mL, pH 3.8) for 10 h at 25 °C, then boiled for 5–10 min and decolorized in 95% ethanol.

**Protein extraction and western blotting.** Total protein was isolated from infiltrated leaf patches with protein extraction buffer (containing 50 mM Tris-HCl (pH 6.8), 4.5% (m/v) SDS, 7.5% (v/v) 2-Mercaptoethanol, 9 M carbamide). Immunoblotting was performed with primary mouse polyclonal antibodies, followed by anti-mouse IgG HRP-linked antibodies (1:5000; Cell Signaling Technology, Cat#: 7076, USA); primary antibodies used are as follows: anti-GFP (1:5000; ROCHE, Cat#: 11814460001, USA), and custom-made anti-PVX CP (1:5000) and anti-TYLCV CP (1:5000)<sup>59,60</sup>.

## Data availability

All data generated or analyzed during this study are included in this article (and its supplementary information files). Source data are provided with this paper.

## Code availability

The code for developing ViralORFinder can be found on GitHub at <https://github.com/mayupsc/vORFinder2.0> (<https://doi.org/10.5281/zenodo.4950738>)<sup>61</sup>

Received: 1 March 2021; Accepted: 21 June 2021;

Published online: 13 July 2021

## References

1. Brito, A. F. & Pinney, J. W. Protein-protein interactions in virus-host systems. *Front Microbiol* **8**, 1557 (2017).
2. King, C. R., Zhang, A., Tessier, T. M., Gameiro, S. F. & Mymryk, J. S. Hacking the cell: network intrusion and exploitation by adenovirus E1A. *mBio* **9**, e00390–18 (2018).
3. Zerbini, F. M. et al. ICTV Virus taxonomy profile: geminiviridae. *J. Gen. Virol.* **98**, 131–133 (2017).
4. Aguilar, E., Garnelo Gomez, B. & Lozano-Duran, R. Recent advances on the plant manipulation by geminiviruses. *Curr. Opin. Plant Biol.* **56**, 56–64 (2020).
5. Kumar, R. V. Plant antiviral immunity against geminiviruses and viral counter-defense for survival. *Front Microbiol* **10**, 1460 (2019).
6. Liu, X. et al. Protein modification: a critical modulator in the interaction between geminiviruses and host plants. *Plant Cell Environ.* <https://doi.org/10.1111/pce.14008> (2021).
7. Basak, J. Tomato yellow leaf curl virus: A serious threat to tomato plants world wide. *J. Plant Pathol. Microbiol.* **7**, 4 (2016).
8. Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D. & Mansoor, S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol* **11**, 777–788 (2013).
9. Luna, A. P., Morilla, G., Voinnet, O. & Bejarano, E. R. Functional analysis of gene-silencing suppressors from tomato yellow leaf curl disease viruses. *Mol. Plant Microbe Interact.* **25**, 1294–1306 (2012).
10. Lozano-Duran, R. et al. Geminiviruses subvert ubiquitination by altering CSN-mediated derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. *Plant Cell* **23**, 1014–1032 (2011).
11. Rosas-Diaz, T., Macho, A. P., Beuzon, C. R., Lozano-Duran, R. & Bejarano, E. R. The C2 protein from the geminivirus tomato yellow leaf curl Sardinia virus decreases sensitivity to jasmonates and suppresses jasmonate-mediated defences. *Plants (Basel)* **5**, 8 (2016).
12. Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. & Bejarano, E. R. Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology* **312**, 381–394 (2003).
13. Selth, L. A. et al. A NAC domain protein interacts with tomato leaf curl virus replication accessory protein and enhances viral replication. *Plant Cell* **17**, 311–325 (2005).
14. Wu, M. et al. Plant DNA polymerases alpha and delta mediate replication of geminiviruses. *Nat. Commun.* **12**, 2780 (2021).
15. Medina-Puche, L. et al. A defense pathway linking plasma membrane and chloroplasts and co-opted by pathogens. *Cell* **182**, 1109–1124 (2020). e25.
16. Rosas-Diaz, T. et al. A virus-targeted plant receptor-like kinase promotes cell-to-cell spread of RNAi. *Proc. Natl Acad. Sci. USA* **115**, 1388–1393 (2018).
17. Azzam, O. et al. Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. *Virology* **204**, 289–296 (1994).
18. Diaz-Pendon, J. A. et al. Tomato yellow leaf curl viruses: menage a trois between the virus complex, the plant and the whitefly vector. *Mol. Plant Pathol.* **11**, 441–450 (2010).
19. Gotz, M. et al. Implication of Bemisia tabaci heat shock protein 70 in begomovirus-whitefly interactions. *J. Virol.* **86**, 13241–13252 (2012).
20. Kunik, T., Palanichelvam, K., Czosnek, H., Citovsky, V. & Gafni, Y. Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *Plant J.* **13**, 393–399 (1998).
21. Ohnesorge, S. & Bejarano, E. R. Begomovirus coat protein interacts with a small heat-shock protein of its transmission vector (Bemisia tabaci). *Insect Mol. Biol.* **18**, 693–703 (2009).
22. Palanichelvam, K., Kunik, T., Citovsky, V. & Gafni, Y. The capsid protein of tomato yellow leaf curl virus binds cooperatively to single-stranded DNA. *J. Gen. Virol.* **79**, 2829–2833 (1998).
23. Rojas, M. R. et al. Functional analysis of proteins involved in movement of the monopartite begomovirus, tomato yellow leaf curl virus. *Virology* **291**, 110–125 (2001).
24. Rubinstein, G. & Czosnek, H. Long-term association of tomato yellow leaf curl virus with its whitefly vector Bemisia tabaci: effect on the insect transmission capacity, longevity and fecundity. *J. Gen. Virol.* **78**, 2683–2689 (1997).
25. Wang, B. et al. V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. *J. Gen. Virol.* **95**, 225–230 (2014).
26. Wang, B., Yang, X., Wang, Y., Xie, Y. & Zhou, X. Tomato yellow leaf curl virus V2 interacts with host histone deacetylase 6 to suppress methylation-mediated transcriptional gene silencing in plants. *J. Virol.* **92**, e00036–18 (2018).
27. Wang, L. et al. A virus-encoded protein suppresses methylation of the viral genome through its interaction with AGO4 in the Cajal body. *eLife* **9**, e55542 (2020).
28. Zhao, W. et al. Tomato yellow leaf curl virus V2 protein plays a critical role in the nuclear export of V1 protein and viral systemic infection. *Front Microbiol* **11**, 1243 (2020).
29. Zrachya, A. et al. Suppressor of RNA silencing encoded by tomato yellow leaf curl virus-Israel. *Virology* **358**, 159–165 (2007).
30. Lin, W. et al. The cap snatching of segmented negative sense RNA viruses as a tool to map the transcription start sites of heterologous co-infecting viruses. *Front Microbiol* **8**, 2519 (2017).
31. Eguen, T., Straub, D., Graeff, M. & Wenkel, S. MicroProteins: small size-big impact. *Trends Plant Sci.* **20**, 477–482 (2015).
32. Hsu, P. Y. & Benfey, P. N. Small but mighty: functional peptides encoded by small ORFs in plants. *Proteomics* **18**, e1700038 (2018).
33. Murphy, E., Smith, S. & De Smet, I. Small signaling peptides in Arabidopsis development: how cells communicate over a short distance. *Plant Cell* **24**, 3198–3217 (2012).
34. Ding, S. W. RNA-based antiviral immunity. *Nat. Rev. Immunol.* **10**, 632–644 (2010).
35. Ding, S. W., Li, H., Lu, R., Li, F. & Li, W. X. RNA silencing: a conserved antiviral immunity of plants and animals. *Virus Res* **102**, 109–115 (2004).
36. Jin, Y., Zhao, J. H. & Guo, H. S. Recent advances in understanding plant antiviral RNAi and viral suppressors of RNAi. *Curr. Opin. Virol.* **46**, 65–72 (2020).
37. Burgyan, J. & Havelda, Z. Viral suppressors of RNA silencing. *Trends Plant Sci.* **16**, 265–272 (2011).
38. Ruiz, M. T., Voinnet, O. & Baulcombe, D. C. Initiation and maintenance of virus-induced gene silencing. *Plant Cell* **10**, 937–946 (1998).
39. Wang, C. et al. Epigenetics in the plant-virus interaction. *Plant Cell Rep.* **38**, 1031–1038 (2019).
40. Yang, X. et al. Suppression of methylation-mediated transcriptional gene silencing by βC1-SAHH protein interaction during geminivirus-beta satellite infection. *PLoS Pathog.* **7**, e1002329 (2011).
41. Stewart, F. J., Panne, D., Bickle, T. A. & Raleigh, E. A. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* **298**, 611–622 (2000).
42. Derrien, B. et al. Degradation of the antiviral component ARGONAUTE1 by the autophagy pathway. *Proc. Natl. Acad. Sci. USA* **109**, 15942–15946 (2012).
43. Zheng, X., Fahlgren, N., Abbasi, A., Berry, J. C. & Carrington, J. C. Antiviral ARGONAUTES against turnip crinkle virus revealed by image-based trait analysis. *Plant Physiol.* **180**, 1418–1435 (2019).

44. Jouannet, V. et al. Cytoplasmic Arabidopsis AGO7 accumulates in membrane-associated siRNA bodies and is required for ta-siRNA biogenesis. *EMBO J.* **31**, 1704–1713 (2012).
45. Rodriguez-Negrete, E. et al. Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *N. Phytol.* **199**, 464–475 (2013).
46. Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G. & Bisaro, D. M. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J Virol.* **83**, 5005–5013 (2009).
47. Martin, K. et al. Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *Plant J.* **59**, 150–162 (2009).
48. Li, F. et al. The AC5 protein encoded by mungbean yellow mosaic India virus is a pathogenicity determinant that suppresses RNA silencing-based antiviral defenses. *N. Phytol.* **208**, 555–569 (2015).
49. Nakagawa, T. et al. Improved gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* **71**, 2095–2100 (2007).
50. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
51. Liu, Y. et al. EBS7 is a plant-specific component of a highly conserved endoplasmic reticulum-associated degradation system in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **112**, 12205–12210 (2015).
52. Nelson, B. K., Cai, X. & Nebenfuhr, A. A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**, 1126–1136 (2006).
53. Li, F. et al. The AC5 protein encoded by mungbean yellow mosaic India virus is a pathogenicity determinant that suppresses RNA silencing-based antiviral defenses. *New Phytol.* **208**, 555–569 (2015).
54. Xiong, R. et al. Characterization and subcellular localization of an RNA silencing suppressor encoded by rice stripe tenuivirus. *Virology* **387**, 29–40 (2009).
55. Yang, X. et al. Suppression of methylation-mediated transcriptional gene silencing by  $\beta$ C1-SAHH protein interaction during geminivirus-betasatellite infection. *PLoS Pathog* **7**, e1002329 (2011).
56. Hahne, F. & Ivanek, R. Visualizing genomic data using gviz and bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
57. Kosugi, S. et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284**, 478–485 (2009).
58. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. USA* **106**, 10171–10176 (2009).
59. Wu, J. & Zhou, X. Production and application of monoclonal antibodies against potato virus X. *J. Zhejiang Univ. B* **31**, 608–612 (2005).
60. Wu, J. X., Shang, H. L., Xie, Y. & Zhou, X. P. Monoclonal antibodies against the whitefly-transmitted tomato yellow leaf curl virus and their application in virus detection. *J. Integr. Agric.* **11**, 263–268 (2012).
61. Ma, Y. Geminiviruses encode additional small proteins with specific subcellular localizations and virulence function. Repository: <https://doi.org/10.5281/zenodo.4950738>. (2021).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (31930089 and 31720103914), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB27040206), and the Shanghai Center for Plant Stress Biology from the Chinese Academy of Sciences. The authors thank all members of Rosa Lozano-Duran's lab and Alberto Macho's lab for fruitful discussions, Xinyu Jian, Aurora Luque, and the PSC Cell Biology Facility for technical assistance, Alberto Macho for critical reading of the manuscript, and Dr. Michael M. Goodin (University of Kentucky, USA) and Prof. David Baulcombe (University of Cambridge) for kindly sharing materials.

## Author contributions

Conceptualization: R.L.-D., F.L., X.Z.; Investigation: P.G., H.T., S.Z., H.L., H.L., Y.M., X.Z., J.R., X.F.; Writing: P.G., H.T., R.L.-D., F.L., X.Z.; Visualization: P.G., H.T.; Supervision: R.L.-D., F.L., X.Z.; Funding acquisition: R.L.-D., F.L., X.Z.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24617-4>.

**Correspondence** and requests for materials should be addressed to R.L.-D., F.L. or X.Z.

**Peer review information** *Nature Communications* thanks Hernan Garcia Ruiz and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021