JOURNAL OF
GENERAL VIROLOGY

RESEARCH ARTICLE
Aimone *et al*., *Journal of General Virology* 2021;102:001622
DOI 10.1099/jgv.0.001622

MICROBIOLOGY SOCIETY

OPEN ACCESS

Population diversity of cassava mosaic begomoviruses increases over the course of serial vegetative propagation

Catherine D. Aimone[1]†, Erik Lavington[2]†, J. Steen Hoyer[2], David O. Deppong[1], Leigh Mickelson-Young[1], Alana Jacobson[3], George G. Kennedy[4], Ignazio Carbone[5], Linda Hanley-Bowdoin[1] and Siobain Duffy[2,*]

Abstract

Cassava mosaic disease (CMD) represents a serious threat to cassava, a major root crop for more than 300 million Africans. CMD is caused by single-stranded DNA begomoviruses that evolve rapidly, making it challenging to develop durable disease resistance. In addition to the evolutionary forces of mutation, recombination and reassortment, factors such as climate, agriculture practices and the presence of DNA satellites may impact viral diversity. To gain insight into the factors that alter and shape viral diversity *in planta*, we used high-throughput sequencing to characterize the accumulation of nucleotide diversity after inoculation of infectious clones corresponding to African cassava mosaic virus (ACMV) and East African cassava mosaic Cameroon virus (EACMCV) in the susceptible cassava landrace Kibandameno. We found that vegetative propagation had a significant effect on viral nucleotide diversity, while temperature and a satellite DNA did not have measurable impacts in our study. EACMCV diversity increased linearly with the number of vegetative propagation passages, while ACMV diversity increased for a time and then decreased in later passages. We observed a substitution bias toward C→T and G→A for mutations in the viral genomes consistent with field isolates. Non-coding regions excluding the promoter regions of genes showed the highest levels of nucleotide diversity for each genome component. Changes in the 5′ intergenic region of DNA-A resembled the sequence of the cognate DNA-B sequence. The majority of nucleotide changes in coding regions were non-synonymous, most with predicted deleterious effects on protein structure, indicative of relaxed selection pressure over six vegetative passages. Overall, these results underscore the importance of knowing how cropping practices affect viral evolution and disease progression.

INTRODUCTION

Begomoviruses (genus *Begomovirus*, family *Geminiviridae*) are single-stranded DNA (ssDNA) viruses that cause serious diseases in many important crops worldwide [1]. They are characterized by their double icosahedral particles and whitefly (*Bemisia tabaci* Genn) vectors [2, 3]. Like other ssDNA viruses [4–9], begomoviruses have the potential to evolve rapidly because of their small genomes, large population sizes, short generation times and high substitution rates [10, 11]. Begomoviruses also readily undergo recombination [12], but mutation is the major driver of diversification of begomovirus populations [13, 14]. The highly polymorphic nature of begomovirus populations can lead to rapid adaptation and increased virulence [15, 16]. Plant virus evolution is also influenced by ecological conditions, such as agricultural practices and environmental stresses [17–20].

Cassava is a major vegetatively propagated root crop in Africa [21], whose production has been severely impacted by a rapidly evolving complex of 11 begomoviruses species, 9 of which are in Africa, which cause cassava mosaic disease

Received 19 March 2021; Accepted 05 June 2021; Published 26 July 2021

Author affiliations: [1]Department of Plant and Microbial Biology, North Carolina State University, Raleigh NC 27695, USA; [2]Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ 08901, USA; [3]Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA; [4]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, USA; [5]Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh NC 27695, USA.
***Correspondence:** Siobain Duffy, duffy@sebs.rutgers.edu

Keywords: cassava mosaic begomoviruses; vegetative propagation; viral diversity.

Abbreviations: ACMV, African cassava mosaic virus; CMB, cassava mosaic begomovirus; CMD, cassava mosaic disease; CP, coat protein; dsDNA, double-stranded DNA; EACMCV, East African cassava mosaic Cameroon virus; MP, movement protein; NSP, nuclear shuttle protein; PTGS, post-transcriptional gene silencing; RCA, rolling circle amplification; REn, replication enhancer protein; Rep, replication-associated protein; SEGS, sequence enhancing geminivirus symptoms; ssDNA, single-stranded DNA; TGS, transcriptional gene silencing.
Veg2: PRJNA667211 and Veg6: PRJNA667210.
†These authors contributed equally to this work
Three supplementary figures and eight supplementary tables are available with the online version of this article.

(CMD) [22]. Cassava mosaic begomoviruses (CMBs) have bipartite genomes that consist of two circular DNAs designated as DNA-A and DNA-B [23]. Both genome components, which together are ca. 5.5 Kb in size, display high nucleotide substitution rates of approximately 10^{-3} to 10^{-4} substitutions per site per year [10], which are similar to the rates reported for RNA viruses [24, 25]. DNA-A is necessary for viral replication, transcription and encapsidation, while DNA-B is required for viral movement [2, 3]. Both genome components contain divergent transcription units separated by a shared 5′ intergenic sequence or common region that contains the origin of replication and promoters for viral gene transcription [2, 3]. The viral replication origin includes a hairpin structure that contains the nick site for rolling circle replication and iteron motifs that function as origin recognition sequences [26, 27]. The origin motifs are conserved between cognate DNA-A and DNA-B components.

DNA-A encodes six proteins via overlapping genes, while DNA-B encodes two proteins on nonoverlapping genes [2, 3]. (Fig. S1, available in the online version of this article) for diagrams of viral clones and major functions of the viral proteins. Genes specified on the complementary DNA strand are designated as '*C*', while genes on the virion strand are indicated as '*V*'.) The replication-associated protein (Rep, *AC1*), which catalyses the initiation and termination of rolling circle replication [28] and functions as a DNA helicase [29], is the only viral protein essential for replication [30]. REn (*AC3*) greatly increases viral DNA accumulation by facilitating the recruitment of host DNA polymerases for viral replication [31]. Viral ssDNA generated during rolling circle replication can be converted to dsDNA [32] and reenter the replication cycle or be packaged into virions composed of the coat protein (CP, *AV1*) [33]. TrAP (*AC2*) is a transcription factor [34]. TrAP, AC4 and AV2 counter host defenses by interfering with post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS) (for review see [35]). African cassava mosaic virus (ACMV) DNA-A also includes an *AC5* ORF of unknown function that overlaps *AV1* [36, 37]. DNA-B encodes two proteins essential for movement. The movement protein (MP, *BC1*) is necessary for viral transport through the plasmodesmata into adjacent cells and systemically through the plant [38, 39]. The nuclear shuttle protein (NSP, *BV1*) is involved in viral DNA trafficking into and out the nucleus and across the cytoplasm to the cell periphery in coordination with MP [38–40].

CMBs often occur in mixed infections leading to synergy between the coinfecting viruses and increased symptom severity [12, 41]. In the 1990s and 2000s, synergy between ACMV and a recombinant CMB contributed to a severe CMD pandemic that spread from Uganda to other sub-Saharan countries and devastated cassava production [12, 42]. In response to the pandemic, many African farmers adopted cassava cultivars with the CMD2 locus, which confers resistance to CMBs [43, 44]. Cassava plants displaying severe, atypical CMD symptoms were observed in some fields after the widespread adoption of CMD2-resistant cultivars [45]. Subsequently, two DNAs, SEGS-1 and SEGS-2 (sequences enhancing geminivirus symptoms), were shown to produce similar symptoms when coinoculated with CMBs into both resistant and susceptible cassava cultivars [45]. SEGS-2, which occurs as episomes in CMB-infected cassava, virions and whiteflies, is thought to be a novel satellite [46]. SEGS-1 also forms episomes in CMB-infected cassava but is likely derived from a cassava genomic copy [45]. The uniqueness of SEGS-1 and SEGS-2 raises questions about how they interact with other viral components and might impact begomovirus diversity.

Although cassava production relies on vegetative propagation, little is known about how vegetative propagation, environmental factors, and the presence of the SEGS affect CMB evolution. To generate knowledge about the impact of these factors, we examined viral diversity in mixed infections of ACMV and East African cassava mosaic Cameroon virus (EACMCV) during serial vegetative propagation of cassava plants inoculated with only CMBs or coinoculated with SEGS and grown under controlled conditions at two temperatures. This study provided evidence that vegetative propagation, but not temperature or the presence of SEGS, had a significant impact on viral diversity over time.

METHODS

Vegetation propagation study with two passages (Veg2) at two Temperatures

Cassava plants (*Manihot esculenta cv*. Kibandameno) were propagated at 28 and 30 °C under a 12 h light/dark cycle representing the predicted 2 °C temperature shift in Africa by 2030 [47] and inoculated using a microsprayer to deliver plasmids (100 ng) containing partial tandem dimers of DNA-A or DNA-B of ACMV (accession numbers: MT858793.1 and MT858794.1) and EACMCV (accession numbers: MT856195 and MT856192) [48, 49]. Three plants were coinoculated with ACMV and EACMCV clones for each temperature treatment with each plant representing a biological replicate. Samples (1 mg) were collected at 28 days post-inoculation (days p.i.) from symptomatic tissue near the petiole of leaf 2 (relative to the plant apex), flash-frozen in liquid nitrogen, and stored at −80 °C until analysis. At 56 days p.i., a stem cutting with two nodes was generated from each biological replicate and transferred to fresh soil after treatment with rooting hormone (Garden Safe: TakeRoot Rooting Hormone). The stem cuttings were sampled at 28 days after propagation as described above and used for the next round of propagation at 56 days. Propagated plants originating from the same inoculated source plant represent a lineage. Leaf tissue collection and propagation continued for a total of two rounds at each temperature following the above protocol. Frozen leaf tissue was ground, total DNA was extracted using the MagMax Plant DNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA), and DNA <6 Kb in size was selected using the Blue Pippin Prep system (model no. BDQ3010, Sage Science, Beverly MA) prior to library prep [50].

Vegetative propagation study with six passages (Veg6)

Kibandameno plants were propagated from stem cuttings, grown at 28 °C, and inoculated as described above. Plants were inoculated with plasmid DNA (100 ng) corresponding to ACMV +EACMCV, ACMV +EACMCV+SEGS-1 (AY836366), or ACMV +EACMCV+SEGS-2 (AY836367). Each treatment was replicated three times. Leaf samples were collected, and stem cuttings were propagated as described above. This process was repeated six times with a total of seven rounds including the initial inoculated plant. Total DNA was isolated but was not size-fractionated because earlier serial experiments showed that sequencing total DNA samples produce sufficient read coverage for viral diversity analysis [50].

Characterization of viral infection

Symptoms (chlorosis, leaf deformation and stunting) were scored on a severity scale from 1 to 5 (scale: 1=no symptoms to 5=very severe) throughout the new growth on infected plants. The concentrations of ACMV DNA-A and EACMCV DNA-A were measured by quantitative PCR (qPCR) in total DNA samples (0.01 µg) extracted from cassava leaf tissue and analysed in 96-well plates on a Max3000P System (Stratagene, San Diego CA) following conditions provided in Aimone *et al.* [50].

Library preparation

Sequencing libraries for Veg2 and Veg6 were generated using EquiPhi polymerase (Thermo Fisher Scientific, Waltham, MA, USA) for rolling circle amplification (RCA) and the Nextera XT kit (Illumina, San Diego, CA, USA) with unique dual index sequences for library preparation [50]. The protocol was modified to include two RCA reactions for each sample. Each RCA reaction was diluted to 5 ng ml^{-1}, and 2 µl of each reaction were combined. The combined RCA reactions were diluted to 0.2 ng/µl (1 ng total in 5 µl) and used to construct two technical replicate libraries. Libraries of the inoculum plasmids for ACMV and EACMCV (1 ng of plasmid DNA in 5 µl) were also prepared using the Nextera XT kit. The libraries were pooled in equimolar amounts and sequenced on an Illumina NovaSeq 6000 S4 lane to generate 150 bp, paired-end reads.

Analysis of raw reads

Raw sequencing data were processed according to Aimone *et al.* [50], and the workflow is available on Galaxy (ViralSeq, https://cassavavirusevolution.vcl.ncsu.edu/). Raw Illumina data is available from the NCBI Sequence Read Archive for Veg2 (PRJNA667211) and Veg6 (PRJNA667210).

SNP filtering and functional analysis

SNPs were detected using VarScan [51] and filtered for all SNPs present in both of the sequencing technical replicates and at a frequency ≥ 3% in at least one replicate. SNPs present at ≥ 3% frequency in both technical replicates and present in more than one passage were selected for functional analyses. SNPs were categorized by intergenic region, protein, functional domain and known motif using SNPeff [52]. SNPs in coding regions were categorized as synonymous or non-synonymous. The SIFT tool D algorithm was used to predict the effects of single amino acid substitutions caused by non-synonymous codons, with a substitution scored as damaging (≤ 0.05) or tolerated (> 0.05) [53]. Predictions are based on a scaled probability matrix built by the SIFT algorithm using protein sequence alignments between the queried protein sequence and proteins in known databases, in this case [non-redundant protein] [53]. In the overlaps between coding regions, if a SNP was called using VarScan and identified by protein region using SNPeff in one coding region, its impact on the second coding region was categorized as synonymous or non-synonymous. SNPs in the common regions of the DNA-A and DNA-B of ACMV or EACMCV were compared to one another and the reference sequence of the infectious clone for each viral genome component used in the propagation studies. The alignments were performed using a Smith–Waterman sequence alignment [54] using SnapGene software (from Insightful Science; available at snapgene.com). A Mann–Whitney Test in R was used to calculate the difference between the frequency of SNPs observed in the historical database and SNPs that occurred in Veg6 experiment.

Nucleotide diversity and Tajima's D

Nucleotide diversity was calculated from SNP frequencies per nucleotide position by the formula $\pi = \sum_{ij} x_i x_j \pi_{ij}$ [55] using custom Python scripts. When calculating Tajima's D, total read coverage averaged across all SNPs for a given region was used as a proxy for sample size [56]. Sliding window calculations of π were calculated by custom Python scripts with a window size of 300 bp and a step of 10 nucleotides and are reported at the central position of the window.

Experimental effect analyses

The effect size of experimental design variables was analysed in Python by linear regression using the statsmodels module (v0.12.0) with genomic nucleotide diversity as the response [57]. Details of the model and results are discussed below.

Analysis of nucleotide substitution bias

Filtered SNPs were combined by experiment, species and component across all passages. Each unique substitution was used to generate observed counts. Substitution counts for each pair of nucleotides were tested by a χ^2 test of independence on a 2×2 contingency table of observed and expected counts. To generate the expected counts for each pair, we assumed that substitutions in both directions were equally likely, divided observed counts in half, and adjusted for nucleotide proportions in the reference sequence [10]. For example, considering the test for A↔T substitution bias in ACMV DNA-A in the Veg2 experiment, we observed a total of 5 A→T and 13 T→A substitutions. Counts of A and T in the ACMV DNA reference are 743 and 795, respectively, yielding expectations of 8.7 A→T and 9.3 T→A.

Table 1. Veg2, ANOVA results of nucleotide diversity as a response to experimental variables. Model R^2=0.834. Model P=0.00115

	Degrees of freedom	Sum of squares	Mean square	*F*	*P*
Species	1	7.78	7.78	0.395	0.5496
Species: segment	2	27.45	13.73	0.697	0.5296
Passage	1	658.75	658.75	33.453	*0.0007*
Residual	7	137.84	19.69		

Data and analysis availability

Data in the form of VarScan outputs for each of the models with associated metadata along with custom scripts used for analyses are available at https://wwwgithubcom/elavington/PIRE.

RESULTS

Experimental variable effects

We tested the effects of three treatments (temperature, vegetative propagation and the presence of SEGS) on nucleotide diversity of ACMV and EACMCV in two experiments designated as Veg2 and Veg6. Veg2 tested the effect of temperature and included bombard plants (Passage 1, P1) and plants from two rounds of propagation (P2 and P3). Veg6 included clonally bombard plants (P1) and six subsequent rounds of propagation (P2-P7) that tested the effect of SEGS presence and vegetative propagation. Each experiment was conducted with three bioreplicates (independently inoculated plant lineages) per treatment. Symptoms decreased over passages for both Veg2 (P1-P3; Fig. S2a) and Veg6 (P1-P6) (Fig. S2b). Viral titre decreased for ACMV-A (Fig. S2c, d) but not for EACMCV-A (Fig. S2e, f) for Veg2 and Veg6. Best fit linear regression models included passage rounds, viral species (ACMV and EACMCV) and segment (DNA-A or DNA-B) nested in species (Tables 1 and 2, full experimental models are included in Tables S4-1 and S4-2.) Given the lack of evidence for a significant effect for several experimental treatments (replicate, temperature and SEGS) and the failure of the full model to pass diagnostic tests of normality (Jarque–Bera test) and autocorrelation (Durbin–Watson test) [58], we grouped samples by generating a mean variant frequency. For example, to group SEGS treatments (no SEGS,+SEGS-1,+SEGS-2), we averaged allele frequencies across all three SEGS treatments for each SNP that shared the same passage, species, segment and lineage. This had the effect of reducing the number of samples while generating a data set and model that passed diagnostic tests. Sample grouping consistent with this model was used for the rest of the analyses presented in this study. Both Veg2 and Veg6 grouped data were fit to the model:

$$y = \alpha + R_i + S_j + S_j\left(G_k\right) + \varepsilon_{ijk}$$

With *R* for the initial bombardment and each ith passage number, *S* for species, *G* for segment nested in species, intercept α and error term ε. The overall models were significant for both Veg2 (P=0.00115, R^2=0.834), and Veg6 (P=0.00537, R^2=0.438). Only passage was significant in both Veg2 and Veg6, accounting for 94 and 79% of the variance explained by the model, respectively (as eta-squared calculated from Tables 1 and 2).

Viral diversity changes over time

After determining that passage number (P1-P3 for Veg2; P1-P7 for Veg6) was the only factor that significantly impacted viral diversity quantitatively, we investigated how the diversity changed through successive rounds of vegetative propagation across the ACMV and EACMCV genome components. Sliding windows of π were calculated across segments accounting for their circular architecture and information retained after grouping read counts. π, which is the average pairwise difference between sequences in a sample [55], was examined on a population level and not by plant lineage, due to lineage failing to pass the test of normality and autocorrelation (Tables S4-1 and S4-2). Maximum π values were higher in the Veg6 experiment than the Veg2 experiment, suggesting that more rounds of vegetative propagation allowed for greater accumulation of nucleotide diversity (cf. Figs 1 and 2). Comparison of P1, P2 and P3 between Veg2 and Veg6 revealed that the patterns of diversity are similar (Figs 1a, c and S3a, c), except for a decrease in π in the ACMV *AV1* gene at P3 of Veg6 (cf. Figs 1b and 3b). The patterns of increasing and decreasing π along genome components varied across passages, with several regions increasing in diversity across passages in Veg6 (Fig. S3a, c). Variation along the genome

Table 2. Veg6, ANOVA results of nucleotide diversity as a response to experimental variables. Model R^2=0.438. Model P=0.00537

	Degrees of freedom	Sum of squares	Mean square	*F*	*P*
Species	1	231.4	231.4	2.06	0.165
Species: segment	2	149.2	74.6	0.66	0.525
Passage	1	1632.6	1632.6	14.51	*0.0009*
Residual	23	2587.0	112.5		

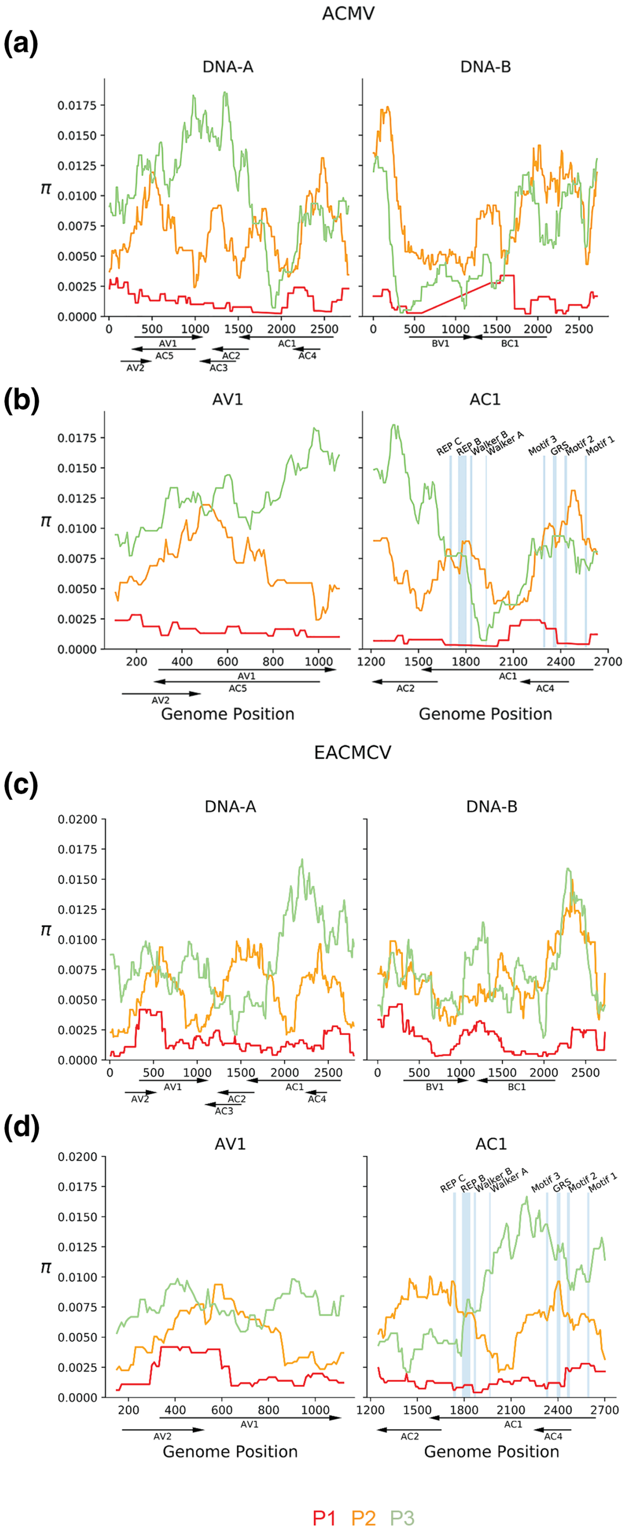


Fig. 1. ACMV and EACMCV nucleotide diversity in the Veg2 experiment. (a) Sliding window analysis of nucleotide diversity (*π*) of ACMV DNA-A and DNA-B and (c) EACMCV DNA-A and DNA-B. Red to green represents the nucleotide diversity across the genome of inoculated plants (p1) and two vegetative propagations (p2 and p3). Enhanced views of the nucleotide diversity of the *AV1* and *AC1* open reading frames during P1-P3 for ACMV-A (b) and EACMCV-A (d). Blue lines mark the locations of codons encoding functional motifs in the Rep protein, i.e. Rep C, Rep B, Walker B, Walker A [63], Motif 3 [62], GRS [64], Motif 2 [62] and Motif 1 [83]. The motifs are shown to scale. Genome coordinates (nt), the positions of open reading frames and their directions of transcription are shown below each graph.

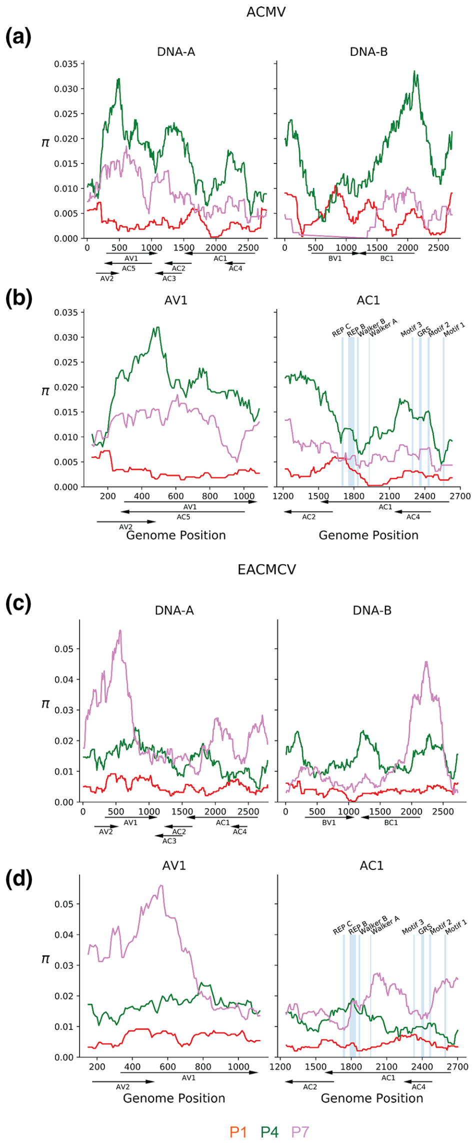


Fig. 2. Veg6 ACMV and EACMCV nucleotide diversity sliding windows. (a) Sliding window analysis of nucleotide diversity (π) of ACMV DNA-A and DNA-B and (c) EACMCV DNA-A and DNA-B. Red to pink represents the nucleotide diversity across the genome of inoculated plants (p1) and two vegetative propagations (p4 and p6). Enhanced views of the nucleotide diversity of the *AV1* and *AC1* open reading frames during P1, P4, and P7 for ACMV-A (b) and EACMCV-A (d). Blue lines mark the locations of codons encoding functional motifs in the Rep protein, i.e. Rep C, Rep B, Walker B, Walker A [63], Motif 3 [62], GRS [64], Motif 2 [62] and Motif 1 [83]. The motifs are shown to scale. Genome coordinates (nt), the positions of open reading frames and their directions of transcription are shown below each graph.

components was most apparent for EACMCV, which showed the highest level of nucleotide diversity in P7 for both genome components (Figs 1c and S3c). The nucleotide diversity for EACMCV-A at P6 was intermediate between that of P7 and P4/P5, and the diversity of EACMCV-B was similar for P4-P6. In contrast, both genome components of ACMV displayed the highest levels of diversity at P4 and P6, with P5 and P7 showing lower levels (Figs 2a and S3a).

The DNA-A components of ACMV and EACMCV exhibited similar patterns of diversity in the passages showing the highest levels of nucleotide diversity, i.e. P4 for ACMV and P7 for EACMCV in the V6 experiment (Fig. 2a, c). The highest peaks of nucleotide diversity were over *AV2* and *AV1* (Fig. 2a, c). In ACMV-A, nucleotide diversity covered a broader region of *AV1* overlapping with *AC5* encoded on the opposite strand. Peaks of diversity were also observed over other overlapping regions of ACMV-A: *AC1*/*AC4* and *AC2*/*AC3* (Fig. 2a, b). In EACMCV, two peaks of diversity occurred over non-overlapping coding regions of *AC1,* with overlapping regions showing moderate diversity(Fig. 2c, d). It was unexpected that nucleotide diversity appeared to peak in regions where viral genes overlap in different reading frames, in which codon wobble positions would be constrained. Chi-square tests confirmed that diversity was not constrained by the number of overlapping protein coding regions within ACMV-A or EACMCV-A ($P > 0.21$, Table S4-3). The DNA-B components of ACMV and EACMCV also displayed similar patterns of nucleotide diversity (Fig. 2a, c). High levels of diversity were seen in the 5′ intergenic region, and *BC1* had higher levels compared to *BV1*.

Test of evolution/demographic changes

After examining viral diversity on the genomic level, we examined the potential impact of more frequent individual variants. We examined the effect of varying minimum variant frequency thresholds on Tajima's D, which infers selection and/or demographic events (population size changes not due to selection) and is sensitive to rare variants [59]. Tajima's D is the standardized difference of two different ways to calculate the expected nucleotide diversity. The caveat with using Tajima's D is that purely demographic changes in the population can result in extreme values of Tajima's D. Our study has an advantage to this point over the problem of disentangling selection and demographic histories in wild populations: given our experimental design we have a good understanding of what demographic effects are possible. We also do not use d*N*/d*S* as this measure is not well suited to the short evolutionary time frame and likely small population sizes found in these experiments. We filtered SNPs by varying the minimum frequency for each of the Veg2 and Veg6 experiments (Fig. 3). In Fig. 3, SNP frequencies were grouped by passage and segment nested within species (ACMV-A, ACMV-B, EACMCV-A, EACMCV-B) for each experiment. General interpretations of Tajima's D typically hold for minimum variant frequencies of 2–4% for SNP analysis [59]. Hence, we used a minimum variant frequency of 3% in our analysis (Fig. 3, dotted vertical line) to rule out the influence of very rare variants and simplify the analysis of potential functional implications of variance. In the Veg2 experiment, Tajima's D > 2 was significant ($P < 0.01$) in P2 for ACMV DNA-B and in P3 for ACMV DNA-A, ACMV DNA-B, and EACMCV DNA-B (Fig. 3a, b), but not for EACMCV DNA-A in any passage (Fig. 3b). In contrast, the later passages of the Veg6 experiment were positively significant for all segments (Fig. 3c, d), but the magnitudes of Tajima's D for all segments varied across passages in a nonlinear fashion similar to the nucleotide diversity profiles (Fig. S3). The marginal significance ($0.01 < P < 0.05$) of a positive Tajima's D in the Veg2 and Veg6 experiments indicates more intermediate-frequency variants than expected and does not support strong selective sweeps in either experiment. We note our data do not support strong selective sweeps or detectable population bottlenecks even though our methods are not as conservative as those in other pool-seq genetic diversity analytics [56].

Biases in nucleotide substitutions

High mutation rates of ssDNA viruses have been attributed to deamination and oxidative damage [10, 60]. Thus, we investigated whether nucleotide substitution bias was detectable within the timeframe of our experiments. We accounted for multiple testing by the Benjamini–Hochberg method with FDR=0.01 [61]. Substitutions were biassed toward C→T and G→A for all components in both experiments ($P < 10^{-6}$). A bias toward C→A was observed in both experiments and toward G→T for all but ACMV DNA-B in the Veg6 experiment ($P < 10^{-3}$, see Table S4-4).

Changes in viral diversity at the codon level

To examine the impact of nucleotide diversity on amino acid codons, we looked at SNPs that passed our 3% threshold, occurred in both technical replicates, and were present in more than one passage [see Table S4-5 (Veg2) and S4-6 (Veg6) for a list of all of the SNPs found in at least two passages]. ACMV had 125 codon changes, while EACMCV had 97 changes, with codon changes occurring in all ORFs of both viruses (Fig. 4a, b). When adjusted for ORF length, *AC1* of both viruses and ACMV *BV1* contained the fewest SNPs, with the other ORFs showing similar levels of SNPs (Fig. 4c, d). We also observed more non-synonymous SNPs than synonymous across both genomes (ACMV: N:80 s:49; EACMCV N:54 s:42). SIFT analysis predicted that the proportion of amino acid substitutions that negatively impact protein function was greater for EACMCV (65%) than for ACMV (52%, Table S4-6). Detrimental amino acid substitutions were predicted to occur in all viral ORFs. Some ORFs had higher number of non-synonymous mutations, with *AV2*, *AC4*, *AC5* of ACMV and *AC4* of EACMCV having the highest (Fig. 4a, b). When adjusted for ORF length, *AC1* and *BV1* had the lowest numbers and fraction of non-synonymous changes for both viruses. Rep is a complex enzyme that catalyses multiple reactions necessary for viral replication [28] and, as such, may be less tolerant to amino acid substitution. In addition to viral DNA trafficking, NSP is engaged in a number of host interactions that may be sensitive to amino acid changes [40].

We asked how the SNPs impacted codons in known functional domains and motifs of the Rep protein (annotated in Figs 1 and 2). The *AC1* ORFs of both viruses include long stretches devoid of polymorphisms, and the majority of changes were synonymous. ACMV *AC1* had synonymous SNPs in the DNA cleavage Motif 3 [62] and the Walker A helicase motif, while EACMCV *AC1* had a synonymous change in the Walker B helicase motif [63]. The sequence of the EACMCV *AC2* promoter element that overlaps the Walker B motif was maintained in the mutant Tyr257Tyr [63]. However, there were several non-synonymous changes distributed throughout the DNA binding/cleavage, oligomerization and DNA helicase domains of their Rep proteins [29] (Table S4-6). Most notably, there was a SNP in EACMCV *AC1* that resulted in a Phe75Val codon change at a highly conserved position in the GRS motif [64].

We also looked at the potential effects of codon changes in other viral proteins. More SNPs were associated with ACMV *AC2* than EACMCV *AC2* (ACMV: 15, EACMCV: 5, Fig. 4). The amino acid changes were distributed across regions of the *AC2* protein associated with DNA binding [65], suppression

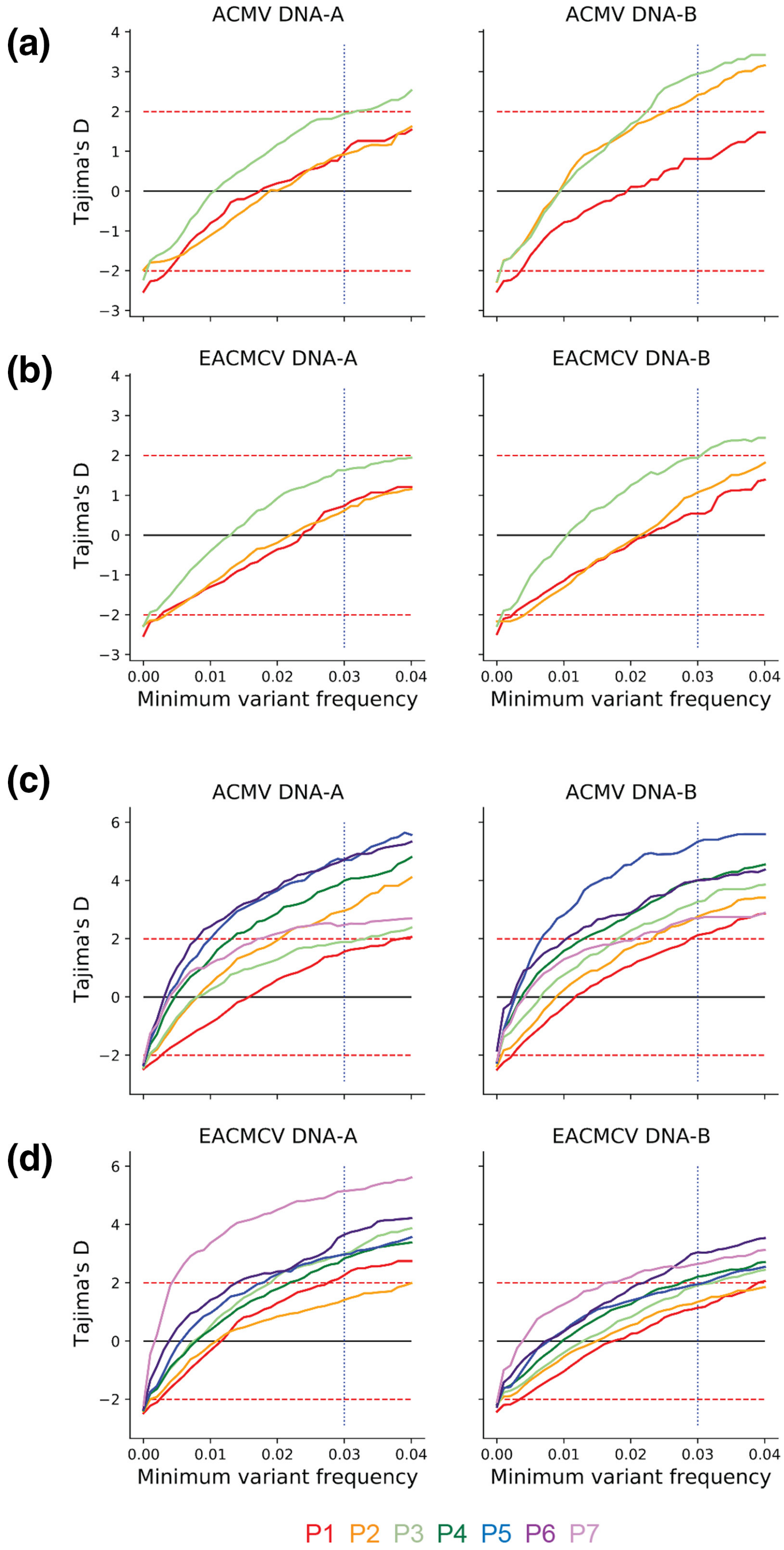


Fig. 3. Tajima's D analysis by passage and minimum variant frequency cutoff. Tajima's D analysis passage (p) for (a) ACMV DNA-A and DNA-B and (b) EACMCV DNA-A and DNA-B in the Veg2 experiment. Tajima's D by passage for (c) ACMV DNA-A and DNA-B and (d) EACMCV DNA-A and DNA-B in the Veg6 experiment. Dotted red horizontal lines represent thresholds for significant Tajima's D values [59]. The vertical blue dotted line represents the minimum variant frequency of 3% used in this study. The lines representing each passage are colour coded, as shown at the bottom.

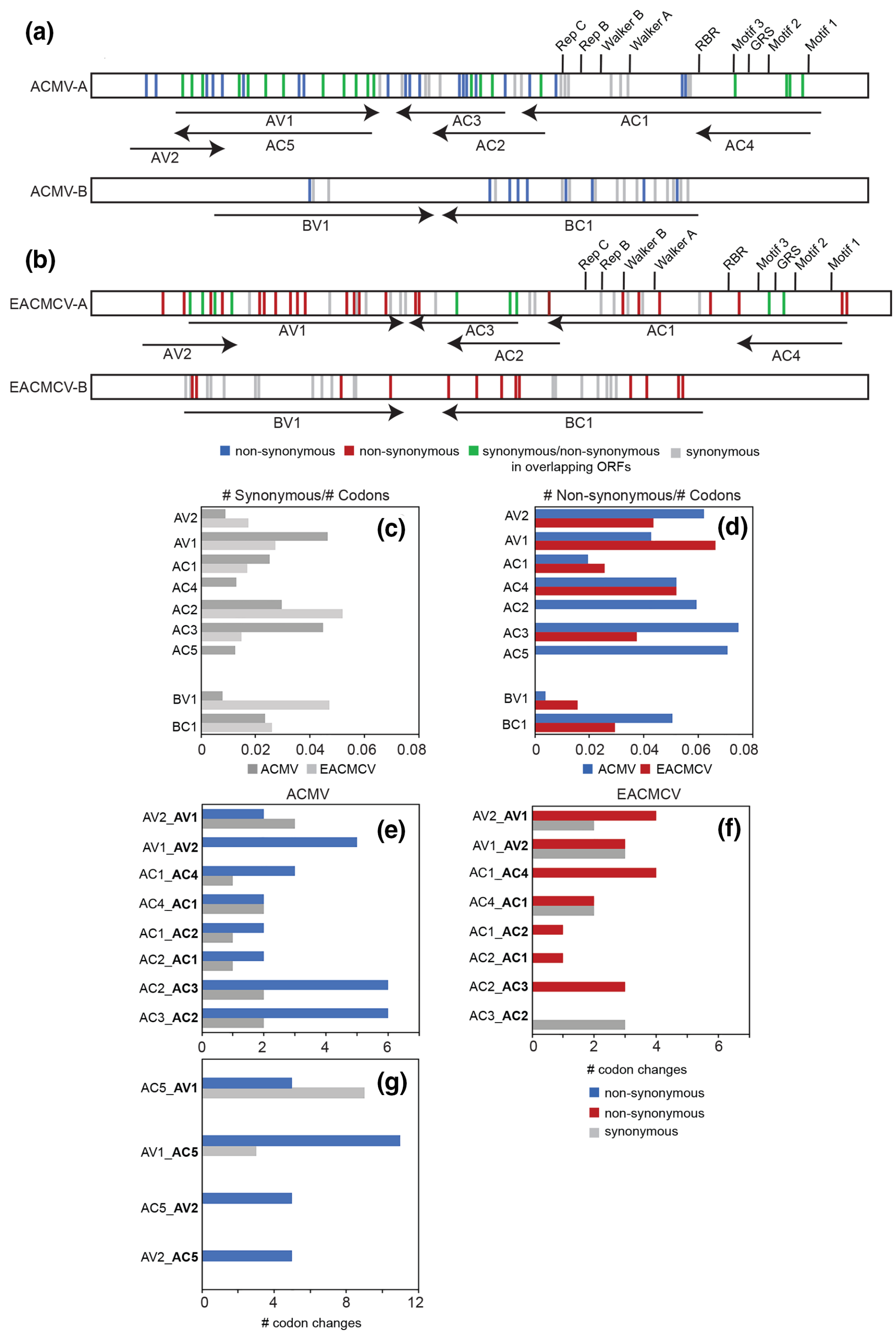


Fig. 4. Synonymous and non-synonymous codon changes in ACMV and EACMCV. Diagram of the locations of synonymous codon changes (grey), unassigned changes in overlapping open reading frames (ORFs; green), and non-synonymous changes for (a) ACMV (blue) and (b) EACMCV (red). The number of synonymous (c) and non-synonymous (d) codon changes normalized to the total number of codons in the ORF for ACMV (dark grey and blue) and EACMCV (light grey and red). The number if synonymous (grey) and non-synonymous (ACMV-blue, EACMCV-red) codon changes in overlapping coding sequences in ORFs overlapping within ACMV excluding *AC5* (e), EACMCV (f), and regions within ACMV that overlap *AC5* (g). The overlapping ORF is designated first and the ORF assessed for the codon change is designated second in bold.

Table 3. Number of SNPs observed in the historical database for Veg6 experiment by genome component

	Acmv-a	Acmv-b	Eacmcv-a	Eacmcv-b
Synonymous	5	1	3	0
Non-synonymous	14	2	3	4

of PTGS [66], and transcriptional activation [67]. ACMV *AC3* was also associated with more SNPs than EACMCV *AC3* (ACMV: 16, EACMCV: 7, Fig. 4), and most of the SNPs in ACMV *AC3* resulted in non-synonymous codon changes, consistent with the capacity of REn to accommodate amino acid substitutions at many positions throughout the protein [68]. The *AC3* gene contained the only codon change with a SNP in both viral genomes, i.e. Pro77 changing to Tyr in ACMV and to His in EACMCV. In contrast, the number of SNPs in *AV1* was similar for both viruses, with approximately half of the SNPs resulting in codon changes. The non-synonymous codon changes impacted amino acids associated with CP nuclear localization, multimerization and ssDNA binding [69], as well as whitefly transmission of begomoviruses [70–72]. There were non-synonymous changes throughout the *BC1* gene of both viruses, some of which introduced amino acid changes at positions in the MP implicated in oligomerization [73] and subcellular targeting [74]. One of the few non-synonymous codon changes in *BV1* was in the NSP nuclear export signal [75].

We also examined codon changes in regions where ORFs overlap on DNA-A. ACMV had more non-synonymous codon changes than synonymous ones in the region of *AV1* that overlaps with *AV2* and *AC5* (Fig. 4e, g), but no such bias was observed in the *AV2*/*AV1* overlapping region in EACMCV (Fig. 4f). Both viruses showed a tendency for non-synonymous changes to accumulate in the *AC4* ORF (Fig. 4e, f), which occurs entirely within *AC1*. Similar results have been reported for Tomato leaf deformation virus and Tomato yellow leaf curl virus [76, 77]. ACMV *AC2* and *AC3* both had a higher proportion of non-synonymous codon changes in their overlapping region (Fig. 4e), but this was not seen for EACMCV, which had non-synonymous codon changes in *AC3* but not *AC2* (Fig. 4f).

Some of the SNPs detected in the Veg6 experiment are also present in a historical database of CMB sequences available from GenBank (Table S4-7), as ascertained by querying recently described multiple alignments [78]. Collectively, 19 Veg6 SNPs for ACMV DNA-A occurred 451 times in the set of 851 DNA-A sequences in GenBank (http://zenodo.org/record/4029589). Six EACMCV DNA-A Veg6 SNPs occurred 20 times in the same set of sequences, three ACMV DNA-B SNPs occurred five times (in a set of 104 sequences; http://zenodo.org/record/3964979), and four EACMCV DNA-B SNPs occurred one time each (in a set of 243 sequences; http://zenodo.org/record/3965023) [79, 80]. The SNPs represented in the historical database represent 9% of the total SNPs identified in the Veg6 study with 15, 6, 8 and 5% of the total SNPs identified for ACMV-A, ACMV-B, EACMCV-A and EACMCV-B, respectively. This indicates that the majority of the SNPs observed in our studies are novel. The SNPs observed in our experiments, those currently in the historical database occurred at a higher allele frequency than those not represented in the historical database for ACMV-A at P 3–7, ACMV-B at P5-6, EACMV-A at P5 and P7, and EACMCV-B at P5 (*P*-value <0.05, Table S4-8). While we cannot rule out random genetic drift, these results beg further investigation of SNPs found in nature where, which may have been under positive selection at certain passages in our experiments (Table S4-8). Yet, a majority of the changes found in the historical database for each genome segment were non-synonymous, except for in EACMCV-A where roughly equal numbers of synonymous and non-synonymous changes were observed (Table 3). Of the non-synonymous changes, 40 % were predicted to be detrimental based on SIFT analysis [53] (Table S4-6).

Changes in common region sequences

We examined nucleotide changes that occurred during multiple passages in the noncoding sequences of DNA-A and DNA-B. There were 17 and 19 changes in the 5′ intergenic sequences of ACMV DNA-A and DNA-B, respectively, while EACMCV had 12 and 35 changes in DNA-A and DNA-B, respectively (Table S4-6). We also observed seven nucleotide changes in the 3′ intergenic region of EACMCV-B, but none for the other viral components, which all have very short 3′ intergenic regions with bidirectional polyadenylation signals [81].

The 5′ intergenic regions of DNA-A and DNA-B contain a conserved sequence designated as the common region that contains the origin of replication and divergent promoter sequences (Figs 5 and 6). Changes in the common region occurred primarily outside of known conserved *cis* elements involved in replication (iterons and the stem loop [26, 27]) and transcription (TATA box, CLE element and CCAAT box [27, 82]). However, we observed an A→G change in the putative TATA box of the ACMV *AV2* promoter (Table S4-6). We also detected G→C transversions in the stem sequences downstream of the nick site of both ACMV DNA-A and ACMV DNA-B (Fig. 5b), but we did not find compensatory changes in the upstream stem sequences. Given that the stem structure is necessary for viral replication [26], it was surprising that the 3′-stem SNPs were maintained in the population at a frequency of ca. 5 % in at least two passages of the Veg6 experiment. We observed a G→A transition at position 2626 in EACMCV-B. Comparison to the historical database indicated that the 5 G residues starting at position 2626 constitute a third conserved iteron in the origin of replication (Fig. 6b, Table S4-7). However, the G at position 2626 is followed by 5 G residues, suggesting they can also serve as an origin iteron when G-2626 is mutated [83].

We also found that the common regions of the DNA-A components of both ACMV and EACMCV acquired mutations such that they more closely resembled the sequences of their cognate DNA-B common regions (Figs 5b and 6b). This

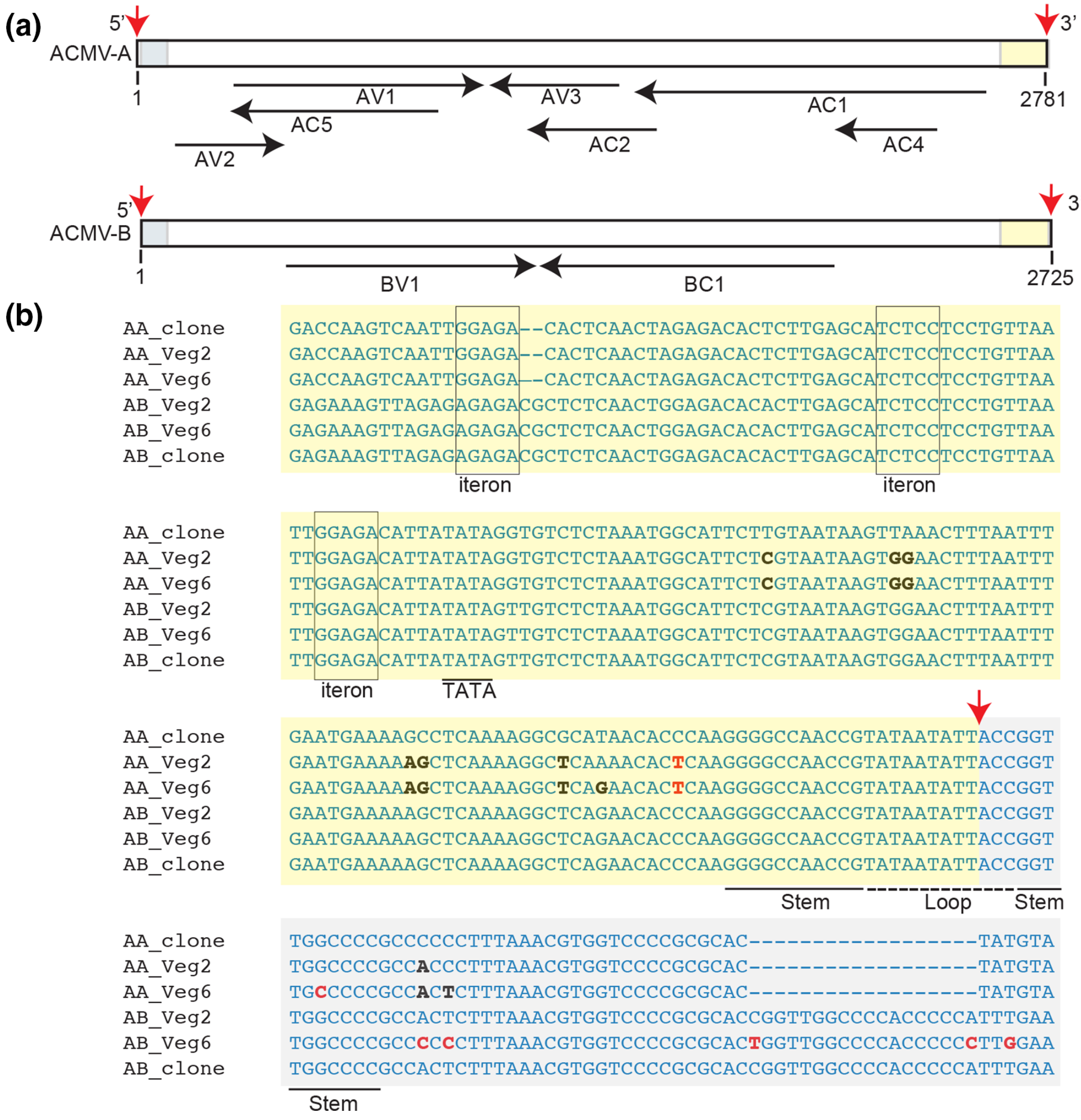
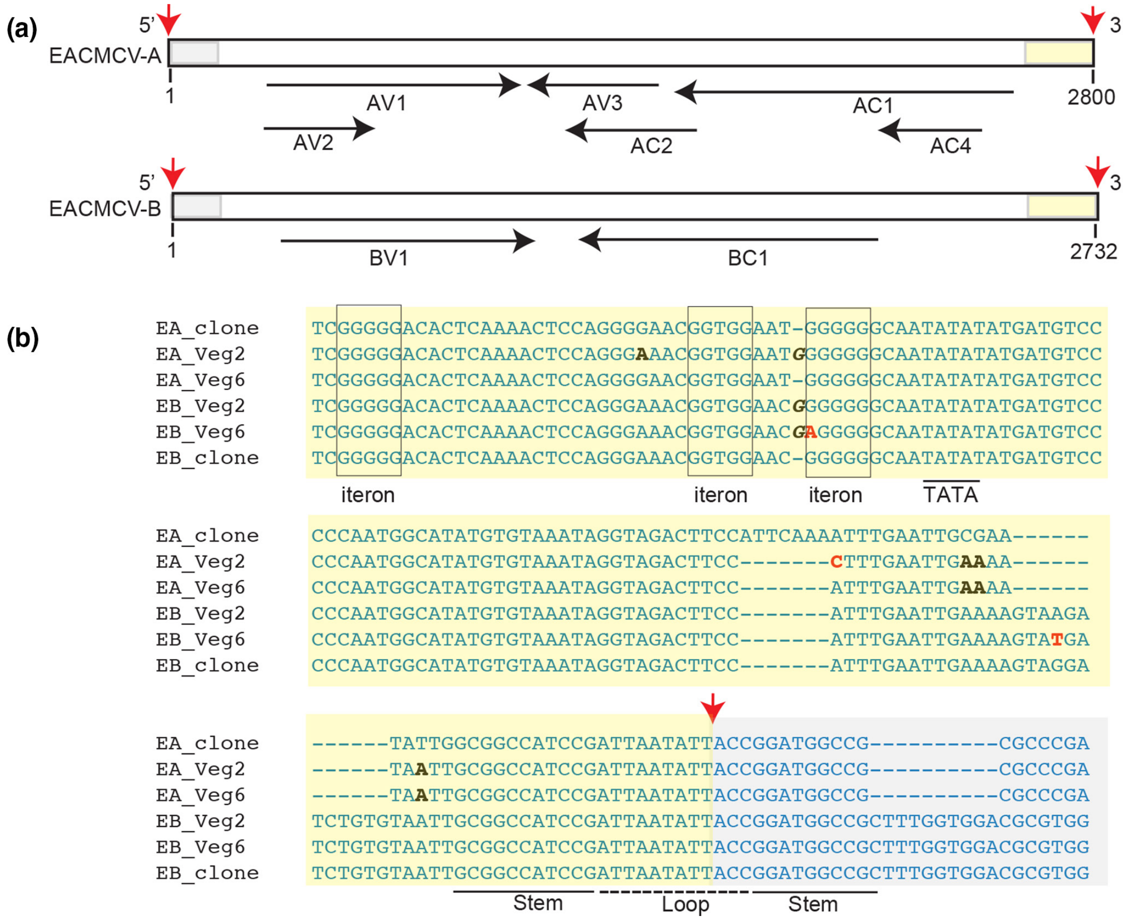

(a)
5'
3'
ACMV-A
1
2781
AV1
AV3
AC1
AC5
AC2
AC4
AV2
5'
3'
ACMV-B
1
2725
BV1
BC1
(b)
AA_clone GACCAAGTCAATTGGAGA--CACTCAACTAGAGACACTCTTGAGCATCTCCTCCTGTTAA
AA_Veg2 GACCAAGTCAATTGGAGA--CACTCAACTAGAGACACTCTTGAGCATCTCCTCCTGTTAA
AA_Veg6 GACCAAGTCAATTGGAGA--CACTCAACTAGAGACACTCTTGAGCATCTCCTCCTGTTAA
AB_Veg2 GAGAAAGTTAGAGAGAGACGCTCTCAACTGGAGACACACTTGAGCATCTCCTCCTGTTAA
AB_Veg6 GAGAAAGTTAGAGAGAGACGCTCTCAACTGGAGACACACTTGAGCATCTCCTCCTGTTAA
AB_clone GAGAAAGTTAGAGAGAGACGCTCTCAACTGGAGACACACTTGAGCATCTCCTCCTGTTAA
iteron
iteron
AA_clone TTGGAGACATTATATAGGTGTCTCTAAATGGCATTCTTGTAATAAGTTAAACTTTAATTT
AA_Veg2 TTGGAGACATTATATAGGTGTCTCTAAATGGCATTCTCGTAATAAGTGGAACTTTAATTT
AA_Veg6 TTGGAGACATTATATAGGTGTCTCTAAATGGCATTCTCGTAATAAGTGGAACTTTAATTT
AB_Veg2 TTGGAGACATTATATAGTTGTCTCTAAATGGCATTCTCGTAATAAGTGGAACTTTAATTT
AB_Veg6 TTGGAGACATTATATAGTTGTCTCTAAATGGCATTCTCGTAATAAGTGGAACTTTAATTT
AB_clone TTGGAGACATTATATAGTTGTCTCTAAATGGCATTCTCGTAATAAGTGGAACTTTAATTT
iteron
TATA
AA_clone GAATGAAAAGCCTCAAAAGGCGCATAACACCCAAGGGGCCAACCGTATAATATTACCGGT
AA_Veg2 GAATGAAAAAGCTCAAAAGGCTCAAAACACTCAAGGGGCCAACCGTATAATATTACCGGT
AA_Veg6 GAATGAAAAAGCTCAAAAGGCTCAGAACACTCAAGGGGCCAACCGTATAATATTACCGGT
AB_Veg2 GAATGAAAAAGCTCAAAAGGCTCAGAACACCCAAGGGGCCAACCGTATAATATTACCGGT
AB_Veg6 GAATGAAAAAGCTCAAAAGGCTCAGAACACCCAAGGGGCCAACCGTATAATATTACCGGT
AB_clone GAATGAAAAAGCTCAAAAGGCTCAGAACACCCAAGGGGCCAACCGTATAATATTACCGGT
Stem
Loop
Stem
AA_clone TGGCCCCGCCCCCCTTTAAACGTGGTCCCCGCGCAC-----------------TATGTA
AA_Veg2 TGGCCCCGCCACCCTTTAAACGTGGTCCCCGCGCAC-----------------TATGTA
AA_Veg6 TGCCCCCGCCACTCTTTAAACGTGGTCCCCGCGCAC-----------------TATGTA
AB_Veg2 TGGCCCCGCCACTCTTTAAACGTGGTCCCCGCGCACCGGTTGGCCCACCCCCATTTGAA
AB_Veg6 TGGCCCCGCCCCCCTTTAAACGTGGTCCCCGCGCACTGGTTGGCCCACCCCCCTTGGAA
AB_clone TGGCCCCGCCACTCTTTAAACGTGGTCCCCGCGCACCGGTTGGCCCACCCCCATTTGAA
Stem

Fig. 5. ACMV common regions undergo sequence convergence. (a) Linear maps of ACMV DNA-A and ACMV DNA-B. The maps were linearized in the common region at the cleavage site in the top strand of the viral origin of replication. Red arrows mark the 3′-OH and the 5′-P of the nick site [101]. The common region upstream (yellow) and downstream (grey) of the nick site are marked. The open reading frames and directions of transcription are shown by the black arrows below. (b) ACMV DNA-A and ACMV DNA-B sequences showing their common regions in the circularized genomic form. The labelling is the same as in (a), with the nick site indicated by a red arrow and the upstream and downstream sequences marked by yellow and grey shading, respectively. The iterons (boxed) and the hairpin motif (stem: underlined; loop: dotted line) involved for the initiation of viral replication are marked. The TATA box (underlined) for complementary sense transcription is also labelled. SNPs showing convergence of the common region sequences of DNA-A and DNA-B are in black typeface, and other SNPs are in red typeface.

occurred in all bioreplicates of each passage in both Veg2 and Veg6 experiments. The ACMV DNA-A common region had SNPs at 11 positions of which nine matched ACMV-B (Fig. 5b). ACMV DNA-B had five SNPs, three at equivalent positions as ACMV-A substitutions. The EACMCV DNA-A common region had five SNPs with four matching EACMCV DNA-B, and EACMCV DNA-B had two other SNPs (Fig. 6b).

We observed multiple deletions in the alignments of the ACMV DNA-A and EACMCV DNA-A common regions, while their cognate B components maintained the same length without indel mutations throughout all of the propagation experiments. There was a 7 bp deletion in EACMCV DNA-A in both the Veg2 and Veg6 experiments that resulted in a sequence match with EACMCV DNA-B, another mutation that made the DNA-A intergenic region better resemble that



```
EA_clone   TCGGGGGACACTCAAAACTCCAGGGGAACGGTGGAAT-GGGGGGCAATATATATGATGTCC
EA_Veg2    TCGGGGGACACTCAAAACTCCAGGGAAACGGTGGAATGGGGGGGCAATATATATGATGTCC
EA_Veg6    TCGGGGGACACTCAAAACTCCAGGGGAACGGTGGAAT-GGGGGGCAATATATATGATGTCC
EB_Veg2    TCGGGGGACACTCAAAACTCCAGGGAAACGGTGGAACGGGGGGGCAATATATATGATGTCC
EB_Veg6    TCGGGGGACACTCAAAACTCCAGGGAAACGGTGGAACGAGGGGGCAATATATATGATGTCC
EB_clone   TCGGGGGACACTCAAAACTCCAGGGAAACGGTGGAAC-GGGGGGCAATATATATGATGTCC

EA_clone   CCCAATGGCATATGTGTAAATAGGTAGACTTCCATTCAAAATTTGAATTGCGAA------
EA_Veg2    CCCAATGGCATATGTGTAAATAGGTAGACTTCC-------CTTTGAATTGAAAA------
EA_Veg6    CCCAATGGCATATGTGTAAATAGGTAGACTTCC-------ATTTGAATTGAAAA------
EB_Veg2    CCCAATGGCATATGTGTAAATAGGTAGACTTCC-------ATTTGAATTGAAAAGTAAGA
EB_Veg6    CCCAATGGCATATGTGTAAATAGGTAGACTTCC-------ATTTGAATTGAAAAGTATGA
EB_clone   CCCAATGGCATATGTGTAAATAGGTAGACTTCC-------ATTTGAATTGAAAAGTAGGA

EA_clone   ------TATTGGCGGCCATCCGATTAATATTACCGGATGGCCG----------CGCCCGA
EA_Veg2    ------TAATTGCGGCCATCCGATTAATATTACCGGATGGCCG----------CGCCCGA
EA_Veg6    ------TAATTGCGGCCATCCGATTAATATTACCGGATGGCCG----------CGCCCGA
EB_Veg2    TCTGTGTAATTGCGGCCATCCGATTAATATTACCGGATGGCCGCTTTGGTGGACGCGTGG
EB_Veg6    TCTGTGTAATTGCGGCCATCCGATTAATATTACCGGATGGCCGCTTTGGTGGACGCGTGG
EB_clone   TCTGTGTAATTGCGGCCATCCGATTAATATTACCGGATGGCCGCTTTGGTGGACGCGTGG
```

Fig. 6. EACMCV common regions undergo sequence convergence. (a) Linear maps of EACMCV DNA-A and EACMCV DNA-B. The maps were linearized in the common region at the cleavage site in the top strand of the viral origin of replication. Red arrows mark the 3′-OH and the 5′-P of the nick site [101]. The common region upstream (yellow) and downstream (grey) of the nick site are marked. The open reading frames and directions of transcription are shown by the black arrows below. (b) EACMCV DNA-A and EACMCV DNA-B sequences showing their common regions in the circularized genomic form. The labelling is the same as in (a), with the nick site indicated by a red arrow and the upstream and downstream sequences marked by yellow and grey shading, respectively. The iterons (boxed) and the hairpin motif (stem: underlined; loop: dotted line) involved for the initiation of viral replication are marked. The TATA box (underlined) for complementary sense transcription is also labelled. SNPs showing convergence of the common region sequences of DNA-A and DNA-B are in black typeface, and other SNPs are in red typeface. Insertion is indicated by italics.

of the cognate DNA-B (Fig. 6b). The deletion was accompanied by CG→AA changes 11–12 bp downstream. Mapping of the paired-end reads established that the 7 bp deletion and associated nucleotide changes in EACMCV DNA-A were not due to erroneous mapping of EACMCV DNA-B reads. Moreover, the deletion and nucleotide changes were not detected by Illumina sequencing of the plasmid controls, confirming that the variant was not present at low frequency in the EACMCV DNA-A inoculum. Together, these results suggested that the intergenic regions of ACMV DNA-A and EACMCV DNA-A sequences were less fit than those of their DNA-Bs, and there was selective pressure causing convergent evolution through inter-segment recombination with the more optimal sequences that resemble their cognate DNA-B sequences.

DISCUSSION

Viruses exist as populations of related sequence variants [15]. This reservoir of genetic diversity enables plant viral populations to change rapidly in response to environmental conditions, agricultural practices and different hosts [17–20]. Thus, understanding viral diversity and the external parameters that impact diversity is essential to develop durable disease resistance strategies. We characterized the genetic diversity of ACMV and EACMCV after inoculation of cassava plants with cloned viral sequences. We examined the effects of vegetative propagation as an agriculture practice, temperature and the presence of exogenous DNA sequences on the nucleotide diversity of the two CMBs. Our studies revealed that repeated vegetative propagation of infected cassava increased genome-wide nucleotide diversity of CMBs without detectable bottlenecks. Our experimental design starting from defined clones allowed us to confidently perform fine-scale analyses of genetic diversity using π, and to detect signatures of selection using Tajima's D involving large population sizes.

In our primary analysis of environmental factors, we found that vegetative propagation had the largest, and only significant, impact on CMB nucleotide diversity (Tables 1 and 2). Multiple rounds of vegetative propagation of potato tubers infected with potato virus Y (PVY) also resulted in an overall increase in nucleotide diversity, and vegetative propagation had a larger effect on PVY diversity than vector transmission [84]. Diversity profiles through passages varied depending on the PVY strain, with some strains increasing linearly and other strains peaking after the first vegetative passage and then decreasing [84]. Our studies uncovered differences between CMB species, in that ACMV nucleotide diversity peaked at P4 and P6 and then decreased, while EACMCV diversity increased linearly through passages (Figs 2 and S3a). Vegetative propagation was also the main factor leading to recombination events and generation of new geminivirus species in sweet potato [85, 86]. Together, these results show how transmission mode impacts viral populations and evolution in root and tuber crops. Our results underscore the importance of discarding infected plants and providing access to virus-free planting material instead of using vegetative propagation to reduce virus spread and the emergence of new viral variants.

Surprisingly, increasing temperature had no significant impact on CMB nucleotide diversity (Table S4-2). A 10 °C shift has been associated with an increase in potexvirus nucleotide diversity in tomato, and the selection of SNPs in a strain-dependent manner [87]. We used a 2 °C temperature shift, based on predicted increases for global warming in Africa by 2030 [47]. The different outcomes in the two studies are likely due to the fivefold difference in the temperature shifts but could also reflect differences in how DNA vs. RNA viruses or cassava vs. tomato plants respond to elevated temperatures. Nevertheless, our results suggest that the predicted 2 °C temperature shift in Africa, may not be a main driver of CMB diversity [47].

Some begomovirus satellites have been associated with elevated levels of viral DNA and suppression of host DNA methylation and silencing pathways [88], and as such have the potential to impact viral diversity. Although it is not known if SEGS-1 or SEGS-2 function in a similar manner, they have several features in common with satellites [46]. We were unable to detect an effect of either SEGS-1 or SEGS-2 on viral nucleotide diversity in coinoculation experiments with CMBs (Table S4-1). However, these results do not rule out that other types of begomovirus satellites may impact viral diversity.

We used Tajima's D to examine the patterns of nucleotide diversity through time and to gain insight into whether the CMB genomes were undergoing selection during vegetative propagation. Tajima's D compares the average number of pairwise differences with the number of segregating sites and determines selection or bottleneck pressure based on deviations from constant population size. Our analysis uncovered a significant positive change in variant frequency for both ACMV and EACMCV (Fig. 3), indicative of an increase in the number of variants during vegetative propagation. However, the marginal significance is consistent with the generation of a population of intermediate variants from a founder event, i.e. inoculation of a cloned viral sequence, and that most variants are not under positive selection. The lack of selection pressure is consistent with the high level of variation in the amount of nucleotide diversity across passages (Figs 2 and S3a). However, the intermediate variant population may undergo selection with more time, more passages, or most importantly, if the diverse population is exposed to a novel selection pressure. A similar phenomenon was observed for PVY SNP populations associated with five rounds of vegetative propagation of field-grown, infected potato [84]. These studies found that very few variants became fixed in the PVY population due to positive or negative selection, indicating that a small number of viral variants contributed to each new population after propagation [84]. In our study, the variation in nucleotide diversity patterns across the CMB genomic components from one passage to the next (Fig. S3a) is consistent with a small fraction of viral variants propagated to the next generation with no mutations fixed in the population for more than three passages.

We observed more non-synonymous changes than synonymous changes in the genomes of both ACMV and EACMCV (Fig. 4c, d), which is consistent with the higher possibility of non-synonymous mutations in coding regions by random chance. However, a large fraction of the non-synonymous codon changes also observed in both viruses were predicted to cause detrimental amino acid substitutions using SIFT analysis, which has been correlated with a lack of purifying selection [89]. Interestingly, symptom scores declined across the passages (Fig. S2a, b), potentially due to the accumulation of non-synonymous codon changes reducing virulence. However, EACMCV-A titre was stable across all passages (Fig. S2e, f), and ACMV-A titre declined and then stabilized before the reduction in symptom severity (Fig. S2c, d), suggesting that non-synonymous codon changes did not alter viral DNA accumulation over time. This is consistent with the observation that the strongest effects of purifying selection (fewest non-synonymous changes, large stretches without changes) was observed in *AC1*, not *AV1,* as has been observed in field isolates of EACMV [10]. This difference is likely a direct result of the vegetative transmission mode being tested here; many plant viruses have strong purifying selection on their capsid proteins because of their interactions with both plant and insects [90]. During vegetative propagation of CMBs there is no selective pressure from a whitefly vector, and we see that relaxation in the higher numbers of non-synonymous changes in *AV1* compared to *AC1*. While these results rule out strong purifying selection acting on CMB viral ORFs, we cannot rule out balancing selection within the timeframe of our studies based on the results of Tajima's D. The presence of SNPs from our experiment in field-collected sequences in GenBank means that some of the variation we observe is not so deleterious as to never be isolated in nature, but we saw no evidence that this subset of SNPs was under more positive selection in our vegetative passaging. Because field populations of CMBs are not exhaustively sequenced, it remains unclear what fraction of the SNPs in this experiment have and would persist and thrive in the field.

Unlike the viral ORFs, we observed multiple nucleotide substitutions in the common regions of the A components of ACMV and EACMCV that were present in all bioreplicates, resulting in sequences that more closely resembled the common regions of their cognate B components. The common region includes *cis*-acting sequences that are necessary for replication. When an infectious clone is made, a single sequence is cloned out of the viral population in the source plant, and that sequence must have a functional origin to be viable infectious clone. This is in stark contrast to viral sequences carrying detrimental codon mutations, which can be complemented *in trans* by other viral components carrying functional promoters and ORFs [91]. However, the cloned sequence may not have an efficient origin and, thus, would be under selective pressure to better compete with more efficient origin sequences. This type of competition was seen in origin mutants of Tomato golden mosaic virus (TGMV) DNA-B in protoplast replication assays [26]. Hence, the convergence of CMB DNA-A common region sequences toward DNA-B may reflect evolutionary selection of a more efficient origin through sequentially acquired intersegment recombination, which would explain the multiple similar substitutions more easily than sequential acquisition of independent mutations. Examples of intersegment (inter-component) recombination of common regions have been observed during passaging or long-term isolation in *Nicotiana benthamiana* plants under laboratory conditions with tomato mottle virus, bean dwarf mosaic virus, cotton leaf curl virus and African cassava mosaic virus [92–94], but the directionality has previously always been a DNA-B common region becoming more like DNA-A. Intersegment recombination of the common region has also been frequently observed in the ssDNA phytopathogenic nanoviruses, which have six or more genome components [95].

Nucleotide substitution bias has been observed for geminivirus sequences in historical datasets [10] and in experimental populations examined within single plants infected for up to 5 years [96, 97]. Our results showed that nucleotide substitution biases are readily detected in a 2 month timeframe for ACMV and EACMCV, and that this short time is sufficient to detect nucleotide substitution biases in both transitions (C→T and G→A), consistent with historical sequences of East African cassava mosaic virus [10] and experimental infection of Tomato yellow leaf curl China virus [97] and one transversion (G→T, previously reported for experimental populations [96]) for both viruses. It has been proposed that the nucleotide substitution bias is due to oxidative damage [10, 96]. The repeated pattern of nucleotide substitution bias in our experiments points to oxidative damage [98]. The C→T and G→A transitions could be due to hydroxylation mediated by reactive oxygen species, possibly by oxidative deamination of cytosine in the case of C→T mutations [99], and the G→T transversions could be from oxidative damage converting guanine to 8-oxoguanine, which base pairs with adenine leading to a G-C base pair mutating to a T-A base pair [100]. Our SNP nucleotide biases match the substitution biases of ssDNA virus historical data sets, showing that the same molecular patterns can be observed over months as decades.

This study provides evidence that farming practices, such as vegetative propagation, can have a large impact on genetic diversity across CMB genome components. This study also presents evidence that CMB genes are under relaxed selection pressure during vegetative propagation, with *AC1* and *BV1* under the strongest purifying selection pressure. This is not the case for the intergenic region, especially in the convergence of the DNA-A common region sequences toward DNA-B, likely reflecting selection of a more efficient origin. This pattern is consistent with mutated viral genomes being able to complement protein functions *in trans*, but there being strong selection for optimal origin of replication function, which cannot be complemented. Understanding how CMBs and other ssDNA viruses are evolving through experimental evolution studies such as this will ultimately help inform and improve strategies for disease management.

Funding information

This work was supported by the National Science Foundation grant OISE-1545553 to L.H.-B, G.G.K. and S.D. C.D.A. was supported with NSF Graduate Research Fellowship and an NCSU Provost Fellowship.

Acknowledgements

We thank Mary Beth Dallas for her help growing cassava plants. Infectious clones were generously provided by Vincent Fondong (Addgene plasmids 159 134 to 159137; https://www.addgene.org/browse/article/28211870/). We thank the NC State University Genome Science Laboratory for their support. The Cassava Virus Evolution Galaxy server (https://cassavavirusevolution.vcl.ncsu.edu/) is hosted in the NC State University Virtual Computing Lab, and we thank Matthew Gronke, Saahil Chawande, and Jim White for server setup and maintenance. This work also used the Amarel cluster maintained by the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey.

Author contributions

C.A. and E.L. were responsible for writing and preparing the manuscript, experimental design, collection, and data analysis. J.S.H. was responsible for experimental design and data analysis. D.D. and L.M-Y. conducted library preparation. L.H-B., S.D., A.L.J., I.C. and G.G.K. were responsible for experimental design and manuscript preparation.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Rojas MR, Macedo MA, Maliano MR, Soto-Aguilar M, Souza JO, *et al*. World management of geminiviruses. *Annu Rev Phytopathol* 2018;56:637–677.
2. Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol* 2013;11:777–788.
3. Kumar RV. Plant antiviral immunity against geminiviruses and viral counter-defense for survival. *Front Microbiol* 2019;10:1460.

4. Denhardt DT, Silver RB. An analysis of the clone size distribution of phi-X-174 mutants and recombinants. *Virology* 1966;30:10–19.
5. Fersht AR. Fidelity of replication of phage phi X174 DNA by DNA polymerase III holoenzyme: spontaneous mutation by misincorporation. *Proc Natl Acad Sci U S A* 1979;76:4946–4950.
6. Raney JL, Delongchamp RR, Valentine CR. Spontaneous mutant frequency and mutation spectrum for gene A of phiX174 grown in E. *Environ Mol Mutagen* 2004;44:119–127.
7. Shackelton LA, Parrish CR, Truyen U, Holmes EC. High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* 2005;102:379–384.
8. Firth C, Charleston MA, Duffy S, Shapiro B, Holmes EC. Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *J Virol* 2009;83:12813–12821.
9. Grigoras I, Timchenko T, Grande-Pérez A, Katul L, Vetten HJ, *et al*. High variability and rapid evolution of a nanovirus. *J Virol* 2010;84:9105–9117.
10. Duffy S, Holmes EC. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* 2009;90:1539–1547.
11. Rocha CS, Castillo-Urquiza GP, Lima ATM, Silva FN, Xavier CAD, *et al*. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J Virol* 2013;87:5784–5799.
12. Pita JS, Fondong VN, Sangaré A, Otim-Nape GW, Ogwal S, *et al*. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* 2001;82:655–665.
13. Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, *et al*. The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evol* 2017;3:vex005.
14. Lefeuvre P, Moriones E. Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Curr Opin Virol* 2015;10:14–19.
15. Elena SF, Sanjuán R. Virus evolution: Insights from an experimental approach. *Annu Rev Ecol Evol Syst* 2007;38:27–52.
16. Safari M, Roossinck MJ. How does the genome structure and lifestyle of a virus affect its population variation? *Curr Opin Virol* 2014;9:39–44.
17. DeFilippis VR, Villarreal LP. An introduction to the evolutionary ecology of viruses. *Viral Ecology* 2000:125–208.
18. Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, *et al*. Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *ISME J* 2018;12:173–184.
19. Bergès SE, Vile D, Vazquez-Rovere C, Blanc S, Yvon M, *et al*. Interactions between drought and plant genotype change epidemiological traits of cauliflower mosaic virus. *Front Plant Sci* 2018;9:703.
20. Roberts KE, Hadfield JD, Sharma MD, Longdon B. Changes in temperature alter the potential outcomes of virus host shifts. *PLoS Pathog* 2018;14:e1007185.
21. FAOSTAT. FAO Statistical Databases. In: *Food and Agriculture Organization (FAO) of the United Nations*. Rome, Italy, 2016. http://faostatfaoorg/
22. Legg JP, Lava Kumar P, Makeshkumar T, Tripathi L, Ferguson M, *et al*. Cassava virus diseases: biology, epidemiology, and management. *Adv Virus Res* 2015;91:85–142.
23. Stanley J, Gay MR. Nucleotide sequence of of cassava latent virus DNA. *Nature* 1983;301:2660–2662.
24. Sanjuán R. From molecular genetics to phylodynamics: Evolutionary relevance of mutation rates across viruses. *PLoS Pathog* 2012;8:e1002685.
25. Hicks AL, Duffy S. Cell tropism predicts long-term nucleotide substitution rates of mammalian RNA viruses. *PLoS Pathog* 2014;10:e1003838.
26. Orozco BM, Hanley-Bowdoin L. A DNA structure is required for geminivirus replication origin function. *J Virol* 1996;70 1:148–158.
27. Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology* 1994;203:90–100.
28. Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, *et al*. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci USA* 1995;92:3879–3883.
29. Clérot D, Bernardi F. DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. *J Virol* 2006;80:11322–11330.
30. Elmer JS, Brand L, Sunter G, Gardiner WE, Bisaro DM, *et al*. Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication. *Nucleic Acids Res* 1988;16:7043–7060.
31. Wu M, Wei H, Tan H, Pan S, Liu Q, *et al*. Plant DNA polymerases alpha and delta mediate replication of geminiviruses. *bioRxiv* 2020.
32. Jeske H, Lütgemeier M, Preiß W. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *The EMBO Journal* 2001;20:6158–6167.
33. Hipp K, Grimm C, Jeske H, Böttcher B. Near-atomic resolution structure of a plant geminivirus determined by electron cryomicroscopy. *Structure* 2017;25:1303–1309.
34. Sunter G, Bisaro DM. Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell* 1992;4:1321–1331.
35. Loriato VAP, Martins LGC, Euclydes NC, Reis PAB, Duarte CEM, *et al*. Engineering resistance against geminiviruses: A review of suppressed natural defenses and the use of RNAI and the CRISPR/CAS system. *Plant Sci J Exp Bot* 2020;292:110410.
36. Li F, Xu X, Huang C, Gu Z, Cao L, *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* 2015;208:555–569.
37. Stanley J, Gay MR. Nucleotide sequence of cassava latent virus DNA. *Nature* 1983;301:260–262.
38. Noueiry AO, Lucas WJ, Gilbertson RL. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell* 1994;76:925–932.
39. Sanderfoot AA, Lazarowitz SG. Cooperation in viral movement: the geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. *Plant Cell* 1995;7:1185–1194.
40. Martins LGC, Raimundo GAS, Ribeiro NGA, Silva JCF, Euclydes NC, *et al*. A Begomovirus nuclear shuttle protein-interacting immune hub: Hijacking host transport activities and suppressing incompatible functions. *Front Plant Sci* 2020;11:398.
41. Fondong VN, Pita JS, MEC R, de Kochko A, Beachy RN, *et al*. Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *J Gen Virol* 2000;81:287–297.
42. Deng D, Otim-Nape WG, Sangare A, Ogwal S, Beachy RN, *et al*. Presence of a new virus closely related to East African cassava mosaic geminivirus, associated with cassava mosaic outbreak in Uganda. *African J Root Tuber Crops* 1997;2:23–28.
43. Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, *et al*. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res* 2014;186:87–96.
44. Akano A, Dixon A, Mba C, Barrera E, Fregene M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor Appl Genet* 2002;105:521–525.

45. Ndunguru J, De León L, Doyle CD, Sseruwagi P, Plata G, *et al*. Two novel DNAs that enhance symptoms and overcome CMD2 resistance to cassava mosaic disease. *J Virol* 2016;90:4160–4173.

46. Aimone CD, De Leon L, Dallas MM, Ndunguru J, Ascencio-Ibanez JT, *et al*. A new type of satellite associated with cassava mosaic begomoviruses. *bioRxiv* 2021.

47. Jarvis A, Ramirez-Villegas J, Campo BVH, Navarro-Racines C. Is cassava the answer to African climate change adaptation? *Trop Plant Biol* 2012;5:9–29.

48. Hoyer JS, Fondong VN, Dallas MM, Aimone CD, Deppong DO, *et al*. Deeply sequenced infectious clones of key *Cassava begomovirus* isolates from Cameroon. *Microbiol Resour Announc* 2020;9:e00802-20.

49. Fondong VN, Chen K. Genetic variability of East African cassava mosaic Cameroon virus under field and controlled environment conditions. *Virology* 2011;413:275–282.

50. Aimone CD, Hoyer JS, Dye AE, Deppong DO, Duffy S, *et al*. An improved experimental pipeline for preparing circular SSDNA viruses for next-generation sequencing. *bioRxiv* 2020.

51. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, *et al*. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–576.

52. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.

53. Kumar P, Henikoff S, PC N. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 2009;4:1073–1081.

54. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;147:195–197.

55. Nei M, WH L. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 1979;76:5269–5273.

56. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 2011;27:3435–3436.

57. Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python 2010.

58. DCMaEA P. *Introduction to Linear Regression Analysis*. 2nd ed. Wiley, 1992.

59. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989;123:585–595.

60. Harkins GW, Delport W, Duffy S, Wood N, Monjane AL, *et al*. Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. *Virol J* 2009;6:104.

61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57:289–300.

62. Orozco BM, Hanley-Bowdoin L. Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *J Biol Chem* 1998;273:24448–24456.

63. George B, Ruhel R, Mazumder M, Sharma VK, Jain SK, *et al*. Mutational analysis of the helicase domain of a replication initiator protein reveals critical roles of Lys 272 of the B' motif and Lys 289 of the *β*-hairpin loop in geminivirus replication. *J Gen Virol* 2014;95:1591–1602.

64. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibañez JT, *et al*. Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol* 2011;85:1182–1192.

65. Sung YK, Coutts RH. Potato yellow mosaic geminivirus AC2 protein is a sequence non-specific DNA binding protein. *FEBS Lett* 1996;383:51–54.

66. Trinks D, Rajeswaran R, Shivaprasad PV, Akbergenov R, Oakeley EJ, *et al*. Suppression of RNA silencing by a geminivirus nuclear protein, AC2, correlates with transactivation of host genes. *J Virol* 2005;79:2517–2527.

67. Wang H, Buckley KJ, Yang X, Buchmann RC, Bisaro DM. *Adenosine kinase* inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. *J Virol* 2005;79:7410–7418.

68. Settlage SB, See RG, Hanley-Bowdoin L. Geminivirus C3 protein: replication enhancement and protein interactions. *J Virol* 2005;79:9885–9895.

69. Qin S, Ward BM, Lazarowitz SG. The bipartite Geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA. *J Virol* 1998;72:9247–9256.

70. Kheyr-Pour A, Bananej K, Dafalla GA, Caciagli P, Noris E, *et al*. Watermelon chlorotic stunt virus from the Sudan and Iran: Sequence comparisons and identification of a whitefly-transmission determinant. *Phytopathology* 2000;90:629–635.

71. Caciagli P, Piles M, Marian D, Vecchiati M, Masenga V, *et al*. Virion stability is important for the circulative transmission of *Tomato yellow leaf curl Sardinia virus* by *Bemisia tabaci*, but virion access to salivary glands does not guarantee transmissibility. *J Virol* 2009;83:5784–5795.

72. Höhnle M, Höfer P, Bedford ID, Briddon RW, Markham PG, *et al*. Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible *Abutilon mosaic* virus isolate. *Virology* 2001;290:164–171.

73. Kleinow T, Nischang M, Beck A, Kratzer U, Tanwir F, *et al*. Three C-terminal phosphorylation sites in the *Abutilon mosaic* virus movement protein affect symptom development and viral DNA accumulation. *Virology* 2009;390:89–101.

74. Zhang SC, Ghosh R, Jeske H. Subcellular targeting domains of *Abutilon mosaic* geminivirus movement protein BC1. *Arch Virol* 2002;147:2349–2363.

75. Ward BM, Lazarowitz SG. Nuclear export in plants. Use of geminivirus movement proteins for a cell-based export assay. *Plant Cell* 1999;11:1267–1276.

76. Melgarejo TA, Kon T, Rojas MR, Paz-Carrasco L, Zerbini FM, *et al*. Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *J Virol* 2013;87:5397–5413.

77. X-l Y, M-n Z, Y-j Q, Xie Y, X-p Z. Molecular variability and evolution of a natural population of tomato yellow leaf curl virus in Shanghai, China. *Journal of Zhejiang University SCIENCE B* 2014;15:133–142.

78. Crespo-Bellido A, Hoyer JS, Dubey D, Jeannot RB, Duffy S. Interspecies recombination has driven the macroevolution of cassava mosaic begomoviruses. *J Virol* 2021:JVI0054121.

79. Dubey D, Hoyer JS, Duffy S. *Multiple Alignment of ACMV and ACMBFV Dna-B Sequences Zenodo*. 2020.

80. Crespo-Bellido A, Duffy S. Multiple sequence alignment of DNA-A sequences from ACMBFV, ACMV, CMMGV, EACMCV, EACMKV, EACMMV, EACMV, EACMZV, SACMV, ICMV, SLCMV (11 species) Zenodo. 2020.

81. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Crit Rev Plant Sci* 1999;18:71–106.

82. Cantú-Iris M, Pastor-Palacios G, Mauricio-Castillo JA, Bañuelos-Hernández B, Avalos-Calleros JA, *et al*. Analysis of a new begomovirus unveils a composite element conserved in the CP gene promoters of several *Geminiviridae* genera: Clues to comprehend the complex regulation of late genes. *PloS one* 2019;14:e0210485.

83. Argüello-Astorga GR, Ruiz-Medrano R. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 2001;146:1465–1485.

84. da Silva W, Kutnjak D, Xu Y, Xu Y, Giovannoni J, *et al*. Transmission modes affect the population structure of potato virus Y in potato. *PLoS Pathog* 2020;16:e1008608.

85. Paprotka T, Boiteux LS, Fonseca MEN, Resende RO, Jeske H, *et al*. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Res* 2010;149:224–233.

86. Nhlapo TF, Rees DJG, Odeny DA, Mulabisana JM, MEC R. Viral metagenomics reveals sweet potato virus diversity in the eastern and western Cape provinces of South Africa. *S Afr J Bot* 2018;117:256–267.

87. Alcaide C, Sardanyés J, Elena SF, Gómez P. Increasing growth temperature alters the within-host competition of viral strains and influences virus genetic variation. *bioRxiv* 2020.

88. Nawaz-ul-Rehman MS, Fauquet CM. Evolution of geminiviruses and their satellites. *FEBS Lett* 2009;583:1825–1832.

89. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 2000;15:496–503.

90. Chare ER, Holmes EC. Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J Gen Virol* 2004;85:3149–3157.

91. Fondong VN. Geminivirus protein structure and function. *Mol Plant Pathol* 2013;14:635–649.

92. Roberts S, Stanley J. Lethal mutations within the conserved stem-loop of African cassava mosaic virus DNA are rapidly corrected by genomic recombination. *J Gen Virol* 1994;75:3203–3209.

93. Hou YM, Gilbertson RL. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *J Virol* 1996;70:5430–5436.

94. Liu Y, Robinson DJ, Harrison BD. Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *Journal of General Virology* 1998;79:1501–1508.

95. Martin DP, Biagini P, Lefeuvre P, Golden M, Roumagnac P, *et al*. Recombination in eukaryotic single stranded DNA viruses. *Viruses* 2011;3:1699–1738.

96. van der Walt E, Martin DP, Varsani A, Polston JE, Rybicki EP. Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. *Virol J* 2008;5:104.

97. Ge L, Zhang J, Zhou X, Li H. Genetic structure and population variability of tomato yellow leaf Curl China virus. *J Virol* 2007;81:5902–5907.

98. Frederico LA, Kunkel TA, Shaw BR. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* 1990;29:2532–2537.

99. Kreutzer DA, Essigmann JM. Oxidized, deaminated cytosines are a source of C --> T transitions in vivo. *Proc Natl Acad Sci U S A* 1998;95:3578–3582.

100. David SS, O'Shea VL, Kundu S. Base-excision repair of oxidative DNA damage. *Nature* 2007;447:941–950.

101. Koonin EV, Ilyina TV. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J Gen Virol* 1992;73:2763–2766.