Extensive Recombination-Induced Disruption of Genetic Interactions Is Highly Deleterious but Can Be Partially Reversed by Small Numbers of Secondary Recombination Events

Adérito L. Monjane,[a,b] Darren P. Martin,[b] Francisco Lakay,[a] Brejnev M. Muhire,[b] Daniel Pande,[c] Arvind Varsani,[d,e,f] Gordon Harkins,[g] Dionne N. Shepherd,[a] Edward P. Rybicki[a,b]

Molecular and Cell Biology Department, University of Cape Town, Cape Town, South Africa[a]; Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa[b]; Department of Botany and Horticulture, Maseno University, Maseno, Kenya[c]; School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand[d]; Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA[e]; Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa[f]; South African National Bioinformatics Institute, MRC Unit for Bioinformatics Capacity Development, University of the Western Cape, Bellville, South Africa[g]

ABSTRACT

Although homologous recombination can potentially provide viruses with vastly more evolutionary options than are available through mutation alone, there are considerable limits on the adaptive potential of this important evolutionary process. Primary among these is the disruption of favorable coevolved genetic interactions that can occur following the transfer of foreign genetic material into a genome. Although the fitness costs of such disruptions can be severe, in some cases they can be rapidly recouped by either compensatory mutations or secondary recombination events. Here, we used a maize streak virus (MSV) experimental model to explore both the extremes of recombination-induced genetic disruption and the capacity of secondary recombination to adaptively reverse almost lethal recombination events. Starting with two naturally occurring parental viruses, we synthesized two of the most extreme conceivable MSV chimeras, each effectively carrying 182 recombination breakpoints and containing thorough reciprocal mixtures of parental polymorphisms. Although both chimeras were severely defective and apparently non-infectious, neither had individual movement-, encapsidation-, or replication-associated genome regions that were on their own "lethally recombinant." Surprisingly, mixed inoculations of the chimeras yielded symptomatic infections with viruses with secondary recombination events. These recombinants had only 2 to 6 breakpoints, had predominantly inherited the least defective of the chimeric parental genome fragments, and were obviously far more fit than their synthetic parents. It is clearly evident, therefore, that even when recombinationally disrupted virus genomes have extremely low fitness and there are no easily accessible routes to full recovery, small numbers of secondary recombination events can still yield tremendous fitness gains.

IMPORTANCE

Recombination between viruses can generate strains with enhanced pathological properties but also runs the risk of producing hybrid genomes with decreased fitness due to the disruption of favorable genetic interactions. Using two synthetic maize streak virus genome chimeras containing alternating genome segments derived from two natural viral strains, we examined both the fitness costs of extreme degrees of recombination (both chimeras had 182 recombination breakpoints) and the capacity of secondary recombination events to recoup these costs. After the severely defective chimeras were introduced together into a suitable host, viruses with between 1 and 3 secondary recombination events arose, which had greatly increased replication and infective capacities. This indicates that even in extreme cases where recombination-induced genetic disruptions are almost lethal, and 91 consecutive secondary recombination events would be required to reconstitute either one of the parental viruses, moderate degrees of fitness recovery can be achieved through relatively small numbers of secondary recombination events.

Genetic recombination is a major factor in the evolution of many viruses. While recombination can potentially provide viruses with some of the benefits of sexual reproduction, such as the capacity across generations to uncouple beneficial mutations from deleterious ones ([1–3]), it also provides viruses with access to the collective genetic resources of the vast numbers of both distantly related and completely unrelated viral and nonviral species with which they routinely come into contact.

It is obvious, however, that despite having the potential for truly promiscuous genetic exchanges, most recombining virus species show very little evidence of such exchanges. Specifically, there are only a few known cases of successful natural "long-distance" genetic transfers between viruses and their hosts ([4–6]) or between viruses in different families ([7–9]). The simple existence of a workable virus taxonomy with definable species and genera in-

dicates that there are likely significant constraints either on which viruses get opportunities to recombine or on which recombinant viruses survive for long enough to be sampled and characterized.

Whenever recombination occurs between distantly related in-

dividuals (for example, those in different species or genera), it will frequently yield progeny with decreased fitness (10–15). The degree to which such recombination is deleterious is expected to increase with both decreasing parental relatedness and the extent to which transferred genome sequences interact with other viral genome sequences (12, 16, 17). Specifically, it is apparent that viral genome regions that have few genetic interactions with other viral genome regions tend to continue functioning better in foreign genetic backgrounds than do those with extensive interaction networks (12, 18–20).

Genetic interactions could include nucleotide-nucleotide interactions within genomic secondary structures, amino acid interactions within and between proteins, nucleotide-amino acid interactions, and a range of other less direct interactions between, for example, enzymes within metabolic networks or regulatory molecules within signal transduction networks. The numbers of genetic interactions within small virus genomes are likely extremely large (21). In a range of different RNA viruses with genomes that are smaller than 10 kb, >10% of nonneutral and nonlethal single nucleotide polymorphism pairs display a detectable degree of genetic interaction; i.e., there is detectable epistasis between the sites in that the average fitness of the "parental" polymorphism pairs AA and aa (where, in these studies, "a" is a mutation from the wild-type state) is usually greater than that of the "recombinant" polymorphism pairs Aa and aA (22–25). When two distantly related viruses recombine, the new genomes that are produced will essentially contain large numbers of nonparental interacting Aa and aA polymorphism pairs, which, because of the generally unfavorable interactions between them (22–25), will be expected to decrease the fitness of recombinant genomes relative to that of their parental viruses.

Although maintenance of favorable genetic interactions appears to be a key determinant of whether newly emerged recombinants will survive, it is also evident that nonfatal recombination-induced disruptions of genetic interactions can be very rapidly repaired by either compensatory mutations (10, 13, 14) or secondary recombination events (26, 27). However, part of the reason why recombination-induced genetic disruptions appear to be so easily overcome is that studies investigating this phenomenon have focused on recombination events involving transfers of complete genes. Such exchanges would have no disruptive impacts on the genetic interactions that underlie the functional intraprotein amino acid contacts that determine proper protein folding (which certainly account for a large proportion of intragenome genetic interactions). It is perhaps unsurprising, therefore, that in the experimental systems that have investigated the repair of recombination-induced genetic disruption via secondary recombination events, many of the secondary events observed were essentially "reversion recombination events," in that they approximately reversed the primary recombination events (26, 27).

Here, we use the maize streak virus (MSV) (genus *Mastrevirus*, family *Geminiviridae*) model to investigate both the fitness costs of complex "difficult-to-reverse" recombination events that thoroughly mix all parental virus nucleotide polymorphisms and the capacity of secondary recombination events to recoup these costs. Starting with two MSV strains that differ at ~11% of genomic sites, we capitalized on the opportunities afforded by *de novo* whole-genome synthesis to generate two of the most extreme reciprocal recombinants that could possibly be produced from these two viruses by homologous recombination: genome chimeras containing alternating polymorphisms derived from each of the viruses. We show that despite these chimeras being severely defective and apparently incapable of producing symptomatic infections on their own, when they are present together within mixed infections, secondary recombinants arise that have dramatically increased infectivities and virulence.

MATERIALS AND METHODS

Design and construction of synthetic MSV genomes. We designed *in silico* two MSV genomes based on the genetic sequence of a maize-adapted MSV strain A (MSV-A) variant, MSV-MatA (GenBank accession number AF329881) (28), and a *Digitaria*-adapted MSV strain B variant, MSV-VW (GenBank accession number AF239960) (29). The genome sequences of these two MSVs differ at 277 polymorphic sites, and generation of a pairwise alignment of their homologous nucleotides required the insertion of 23 gapped sites (representing insertion/deletion events) into one or the other of the sequences. In this alignment, polymorphic sites and gaps were separated by regions containing between 1 and 67 conserved nucleotides.

In our design of the synthetic MSV genomes, to ensure that we systematically and evenly distributed the polymorphisms from the MSV-A and MSV-B isolates across two viral genomes containing a genomic backbone of nucleotide sites shared by the MSV-A and MSV-B isolates, we followed these simple rules: (i) in each synthetic MSV genome (named sMSV1 and sMSV2), the placement of a polymorphism derived from, say, MSV-A would be followed by another from MSV-B whenever the two polymorphic sites were separated by at least two conserved nucleotides; (ii) when two or more contiguous polymorphic sites were inserted between conserved regions made up of at least two nucleotides, all the contiguous polymorphisms would be transferred as a unit into either sMSV1 or sMSV2; and (iii) whenever single or multiple polymorphic sites were separated by a single conserved nucleotide, all contiguous polymorphisms would be transferred as a single unit into either sMSV1 or sMSV2. Although recombination breakpoints could have been made randomly, the design that we used ensured that recombinants contained the maximum number of effective recombination breakpoints (i.e., breakpoints that segregate parental polymorphisms) while at the same time retaining the biological plausibility of the recombination events (i.e., the simulated sequence transfers could conceivably have occurred by homologous recombination). Following these design guidelines, we generated two complete, reciprocally recombinant MSV genomes consisting of 182 individual segments of alternating MSV-A- and MSV-B-derived sequence (i.e., the recombinants had 182 recombination breakpoints).

In silico, the sMSV1 and sMSV2 genomes were arranged end-to-end as a single construct, with an extra copy of the virion strand origin of replication (a site within the long intergenic region [LIR]) of sMSV2 placed at the start of the construct so that, in effect, the synthesis of this construct by Epoch Life Science resulted in an MSV dimer (named M-V50V-M20M-V200) with three LIR regions cloned within pBluescript II SK(+) (see Fig. S1 in the supplemental material).

Construction of MSV-A-based chimeras containing synthetic genes. In order to test the individual impact of multiple synthetic recombination events on the functioning of genome regions involved in movement (movement protein gene [*mp*] region; V2), encapsidation (coat protein gene [*cp*] region; V1), and replication (the replication-associated gene region and flanking virion and complementary-strand replication origins [LIR-*rep*-SIR {short intergenic region}, LIR, and C1-C2-SIR]), we transferred these regions from sMSV1 and sMSV2 into the genetic background of MSV-A to construct eight distinct MSV-A-based chimeric viruses (described in more detail in Text S1 in the supplemental material).

Construction of agroinfectious clones. Agroinfectious clones of MSV-B (30) and MSV-A (28) were described previously. The eight MSV-A-based chimeras mentioned above and two synthetic viruses cloned into pSK+ were dimerized as previously described (29) (see Text S1 in the supplemental material). However, due to the presence of an ScaI restriction site in *cp* of sMSV2, all MSV genomes containing this particular gene

were dimerized by first separating the MSV genome from its pSK+ vector using BamHI (Fermentas, USA). Once isolated, each MSV genome was self-ligated by using T4 DNA ligase (Fermentas, USA) and subsequently cloned as a dimer into pSK+ digested with BamHI (Fermentas, USA). Using EcoRI and XbaI, all the dimers were then extracted from pSK+ and transferred into similarly digested pBI121 (Clontech, USA) to produce infectious clones, as previously described (31). The sMSV1 and sMSV2 genomes arranged contiguously in the M-V50V-M20M-V200 construct were made infectious by using a similar approach.

Agroinoculation-based infectivity and pathogenicity assays in maize. Three-day-old seedlings of the MSV-sensitive maize genotype Sweetcorn cv. Golden Bantam (GB; Millington Seed Co., USA) were agroinfected separately with sMSV1, sMSV2, M-V50V-M20M-V200, MSV-A-based chimeric viruses, and recombinant viruses arising during evolution experiments, as previously described (26). As a proxy measure of virus fitness (32–34), we (i) quantified chlorotic leaf areas occurring on leaves 4, 5, and 6 of symptomatically infected plants, as previously described (35, 36); (ii) determined infection frequencies; and (iii) determined mean times until the development of visible disease symptoms. The infectivity of MSV-A-based chimeras was also assayed by agroinoculation except that instead of Golden Bantam, we used a moderately resistant maize genotype (cv. Star 7714; Starke Ayres, South Africa) and determined only infection frequencies. The use of this moderately resistant genotype enabled better differentiation between the infection frequencies of the various viruses investigated.

Viral DNA isolation, cloning, and sequencing. Viral DNA was isolated from symptomatic leaf material at between 20 and 127 days (median, 47 days) postinoculation (p.i.) by using the Extract-n-Amp kit (Sigma-Aldrich, USA), followed by rolling-circle amplification as previously described (37, 38). Amplified DNA was digested with the restriction enzyme BamHI (Fermentas, USA) to generate ~2.7-kb monomeric MSV genomes, which were gel purified (GFX; GE Healthcare), ligated into BamHI-digested pGEM-3Zf(+) (Promega, USA) using T4 DNA ligase (Fermentas, USA), and used to transform competent E. cloni cells (Lucigen Corporation, USA). From each symptomatic plant, we randomly selected at least one (and in some cases up to 20) cloned genome for sequencing (Macrogen, South Korea), using universal M13 forward and reverse sequencing primers and previously described MSV-specific primers (39).

These isolated viruses were named in such a way that the number immediately after the letter "S" refers to the plant from which the virus was obtained and the number after the dash (as in "S7-30") refers to the specific clone selected for full-genome sequencing. Genetic distances (*p*-distances) between the sequenced genomes were determined by using Mega5 (40) with pairwise deletion of gaps.

Quantitative PCR-based analysis of viral replication efficiency. To assess viral replication efficiencies, particle bombardment of maize suspension cells followed by a quantitative real-time PCR assay was performed as described previously (34, 41, 42).

Computation of DNA folding distances among sequences. We inferred ensembles of plausible secondary structures for MSV-A, MSV-B, sMSV1, sMSV2, and various recombinants and sMSV1 mutants arising during the evolution experiment using the hybrid-ss-min component of the UNAFold computer program (43). Each genome was folded as circular single-stranded DNA (ssDNA) at 25°C in a 0 M magnesium–1 M sodium solution, and we considered all suboptimal structures with free energies within 5% of the minimum free energy as credible structures. A "fold distance" was calculated to determine the degree to which the DNA secondary structure of each of the genomes differed from those of four reference genomes: MSV-A, MSV-B, sMSV1, and sMSV2.

For each genome folded, we represented all predicted base-pairing interactions in a consensus matrix whose i,j entry contained values of 1 if, in any of the suboptimal structures, the nucleotide at the ith position was predicted to be paired with the nucleotide at jth position and 0 if nucleo-

tides at the ith and jth positions were not predicted to be paired in any of the structures.

By considering only entries where MSV-A and MSV-B consensus matrices had different values (i.e., entries showing base pair interactions in one matrix but not in the other), we computed the fold distance, say, between sequence X and MSV-B as the number of base pairs in the X consensus matrix that were not present in the MSV-B consensus matrix plus the number of base pairs in the MSV-B consensus matrix that were not present in the X consensus matrix. The fold distance between sequence X and the remainder of the reference genome folds was similarly calculated.

Statistical and mathematical analyses. Using the percent pairwise genetic distance between recombinants and MSV-A, we calculated a two-tailed *P* value by using a Mann-Whitney U test to determine whether there were significant differences in the overall genomic similarity of recombinant viruses to MSV-A. Using this test, we also assessed the statistical significance of the differences observed for percent chlorotic leaf area and the number of days to symptom development seen after infection of GB seedlings with four recombinant viruses arising during the evolution experiments.

Euclidian distances from the midpoint of the MSV-A and MSV-B genetic/fold distances were calculated for each virus (with the sign of the distance indicating whether they were on the MSV-A or MSV-B side of the midpoint, respectively), and it was determined whether the 95% confidence intervals of genetic/fold distances from this midpoint included zero. Whereas a value significantly higher than zero would indicate a significant enrichment among the analyzed viruses of MSV-A-derived nucleotide sequence/structural fold characteristics, a value significantly lower than zero would indicate a significant degree of enrichment for MSV-B-derived nucleotide sequence/structural fold characteristics.

RESULTS AND DISCUSSION

The synthetic MSV chimeras are severely defective. The genomes of the two MSV strains used to produce the pair of synthetic recombinant viruses differ at 278 out of 2,689 genome sites. These sites might conceivably be distributed by a minimum of 182 consecutive homologous recombination crossover events between two reciprocal recombinant genomes (where each event involves recombination breakpoints falling at sites where there is at least one conserved nucleotide between two polymorphic nucleotide sites). We synthesized these reciprocal MSV-A/MSV-B genome chimeras (named sMSV1 and sMSV2) and tested their viability.

Although neither of the two synthetic viruses produced symptomatic infections when agroinoculated into 3-day-old maize seedlings (approximately 150 seedlings were inoculated with each chimera), sMSV2 was detectably capable of autonomous replication in cultured maize cells (Mann-Whitney two-tailed *P* value of 0.0002, compared to a nonreplicating construct) (Fig. 1).

While confirming our prediction that sMSV1 and sMSV2 would display high degrees of recombination-induced genetic disruption (i.e., the AA and aa parental virus genetic configurations had higher average fitnesses than did the Aa and aA reciprocal recombinant genetic configurations), we wanted to determine which regions of the chimeric genomes had incurred the most damaging disruptions. In order to pinpoint these regions, we separately transferred the encapsidation-associated (coat protein [*cp*]), cell-to-cell movement-associated (movement protein [*mp*]), systemic movement-associated (movement protein and coat protein [*mp-cp*]), and replication-associated (short intergenic region, replication-associated protein gene, and LIR [SIR-*rep*-LIR]) genome fragments of sMSV1 and sMSV2 into the genomic background of the MSV-A parental isolate (the one best
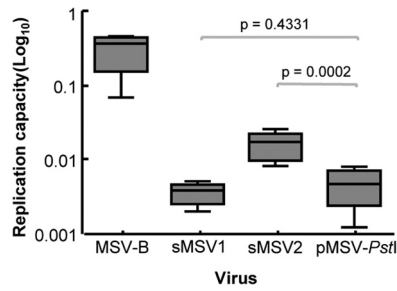
FIG 1 Replication capacities of sMSV1 and sMSV2 in cultured maize cells. Box-and-whisker plots show the results of a quantitative real-time PCR-based analysis of the replication capacities of MSV-B, the synthetic MSV genomes (sMSV1 and sMSV2), and a nonreplicating genome (pMSV-PstI, included as a negative control [53]) relative to that of MSV-A (which, although not shown in the figure, is considered to have a replication capacity of 1). In each plot, the box extends from the 25th to the 75th percentiles, and the whiskers represent the 10th and 90th percentiles. Outliers are identified as filled black circles. Of the synthetic viruses, only sMSV2 displayed a degree of replication that was detectably higher than that of the nonreplicating genome.

adapted to infecting maize) and compared these eight partially synthetic genomes in a symptomatic infection assay (Fig. 2).

All of the partially synthetic genomes were capable of producing symptomatic infections (Fig. 2), indicating that none of the individual synthesized genome regions were by themselves solely responsible for the apparent inability of sMSV1 and sMSV2 to symptomatically infect maize.

We noted, however, that there were clear differences in the infectivities of the eight partially synthetic genomes. Whereas the systemic movement-associated genome fragments (the *mp-cp* gene module) of both sMSV1 and sMSV2 and the replication-associated genome fragment (the SIR-*rep*-LIR module) of sMSV2 were all severely defective (yielding only 9%, 3%, and 8% symptomatic infection frequencies, respectively, within the MSV-A genetic background), the cell-to-cell movement-associated (*mp*) fragments of both synthetic viruses and the replication-associated genome fragment of sMSV1 were reasonably functional (yielding >54% infection frequencies within the MSV-A genetic background).

These results also enabled us to objectively compare degrees of functionality between pairs of homologous synthetic genome regions and indicated that whereas sMSV2 probably had the least defective cell-to-cell movement-associated genome fragment (the *mp* module, which had a 76% versus a 65% infection frequency), sMSV1 had the least defective replication-associated (with a 54% versus an 8% infection frequency) and encapsidation-associated (with an 80% versus a 36% infection frequency) genome fragments.

These attempts to untangle the individual contributions of the various synthesized genome regions to the decreased viability of sMSV1 and sMSV2 indicated that, within the context of mixed infections between these barely viable viruses, it might be possible for small numbers of secondary recombination events to produce genomes with increased fitness. Specifically, we realized that although the synthetic viruses collectively contained every nucleotide necessary to reconstitute the highly maize-adapted MSV-A genotype, it was very unlikely that the 91 consecutive secondary recombination events (involving 182 precisely located breakpoints) needed to achieve this reconstitution would ever occur. What these experiments instead suggested is that much simpler sMSV1 and sMSV2 secondary recombinants (i.e., with <10 breakpoints), such as those containing *mp* of sMSV2 and the remainder of their genomes from sMSV1, could conceivably be considerably more fit than either of the synthetic genomes.

Mixed infections of sMSV1 and sMSV2 yield secondary recombinant viruses. The production of sMSV1-sMSV2 mixed infections presented a significant technical problem. Not only would it be necessary to ensure that both of these barely viable viruses successfully initiated systemic infections within the same plant, but recombination between these genomes would additionally require that they reliably coinfect the same cells. Our solution to this problem was to produce a single infectious clone, called M-V50V-M20M-V200 (see Fig. S1 in the supplemental material), which contained sMSV1 and sMSV2 arranged in tandem, with each genome bounded by its own virion strand origin of replication. It was expected that, as in standard agroinoculation systems (44), both sMSV1 and sMSV2 would be replicationally released together from the M-V50V-M20M-V200 construct following its
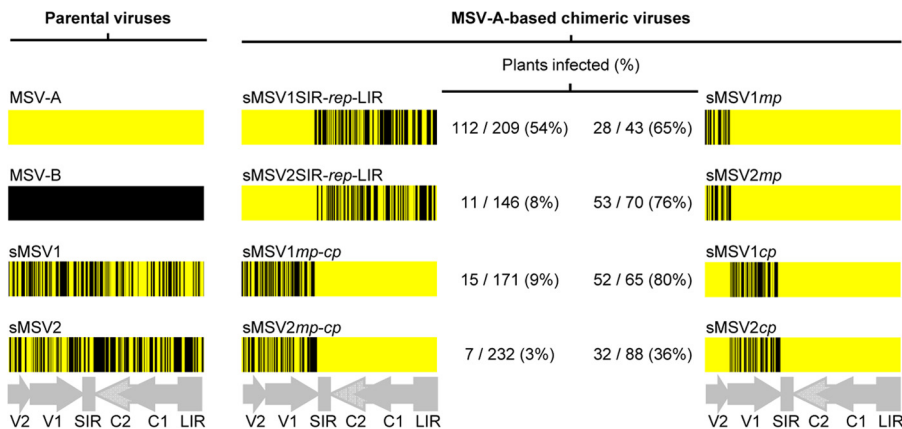
FIG 2 Infectivities of synthetic/partially synthetic recombinant viruses. Starting with two natural MSV isolates, MSV-A and MSV-B (represented in yellow and black, respectively), polymorphisms differentiating these isolates were systematically mixed to produce two synthetic genome sequences, sMSV1 and sMSV2. Genome segments associated with intercellular virus movement (*mp*; V2), systemic virus movement (*mp-cp*; V2 and V1), encapsidation (*cp*; V1), and replication (LIR-*rep*-SIR and LIR-C1-C2-SIR) were transferred from synthetic viruses into the MSV-A genomic background. The infection frequency of each recombinant genome in maize cv. Star 7714 is indicated to the right of each construct.
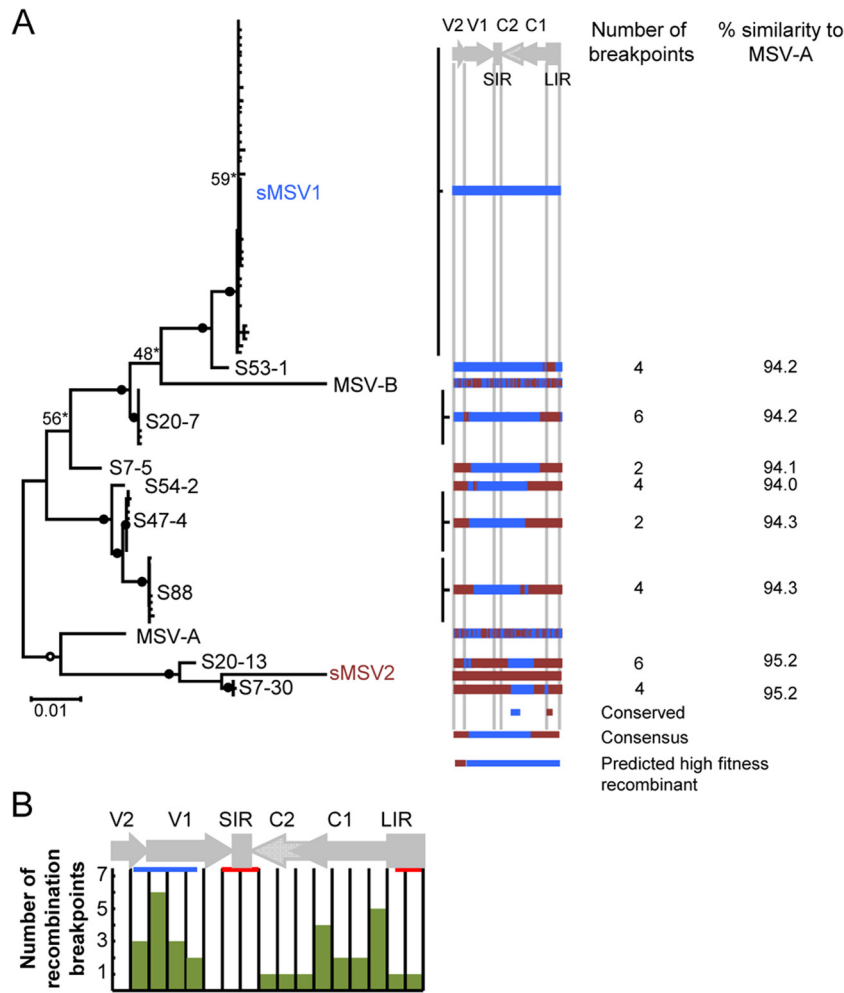
FIG 3 Relationships and recombination patterns displayed by viruses arising during sMSV1 and sMSV2 coinfections in maize. (A) Neighbor-joining tree for visualization of the degrees of sequence similarity to sMSV1 (in blue type) and sMSV2 (in brown type) displayed between recombinant and mutant MSV genomes sampled from symptomatic mixed infections. Except where indicated with an asterisk, branches occurring in ≥70% and ≥90% of 1,000 full neighbor-joining bootstrap replicates are indicated by open and filled circles, respectively. The positions of MSV-A and MSV-B are also shown. Short branches within the sMSV1 clade represent sMSV1 mutants that arose during mixed infections (discussed further in Text S2 in the supplemental material). To the right of the phylogenetic tree is a linearized genome schematic showing the pattern of recombination and the number of breakpoints in each group of recombinant MSVs, where genomic regions in blue are derived from sMSV1 and regions in brown are derived from sMSV2. The genomic regions that are conserved or form the majority consensus of all recombinant viruses are shown. Also, the genomic regions of sMSV1 and sMSV2 that were experimentally determined to be the least dysfunctional among these viruses (Fig. 2) are shown. (B) Linearized MSV genome schematic divided into 17 approximately equal segments. The total number of recombinants with breakpoints within each segment is represented by a bar graph. The red and blue horizontal bars indicate the approximate regions of prominent recombination hot spots and cold spots, respectively, identified in field-isolated MSV genomes (46). V2, movement protein gene; V1, coat protein gene; SIR, short intergenic region; C1/C2, replication-associated protein gene; C1, *repA* gene; LIR, long intergenic region.

*Agrobacterium tumefaciens*-mediated delivery into suitable plant cells.

We agroinoculated approximately 300 MSV-susceptible 3-day-old maize seedlings with M-V50V-M20M-V200, and by 100 days postinoculation, 41 of these seedlings had clearly developed MSV infection symptoms. We cloned and sequenced 108 genomes from these 41 plants, with between 1 and 29 genomes sampled per plant.

Along with 42 sMSV1-sMSV2 recombinant genomes displaying eight distinct recombination patterns that were sampled from 6 of the 41 plants (Fig. 3A), 25 nonrecombinant sMSV1 genomes and 42 sMSV1 mutant genomes were also sampled. A detailed analysis of sMSV1 mutants is provided in Texts S1 and S2 and Fig. S2 in the supplemental material.

Although four of the recombinant genomes were sampled from four different plants, there were two instances in which two distinct recombinants were found in the same plants (i.e., recombinants S20-7 and S20-13 and recombinants S7-30 and S7-5) (Fig. 3A).

Secondary recombination events do not occur at known recombination hot spots. The recombinant genomes displayed clear evidence of between 2 and 6 recombination breakpoints (Fig. 3A). Previous studies involving both natural and laboratory-derived geminiviruses that are related to MSV indicated that recombination breakpoints tend to cluster within noncoding regions of the genome. This pattern is probably due to both the mechanistic predisposition for recombination breakpoints in geminiviruses to occur at replication origins and selective forces disfavoring the survival of recombinants encoding chimeric proteins (18, 45–49).
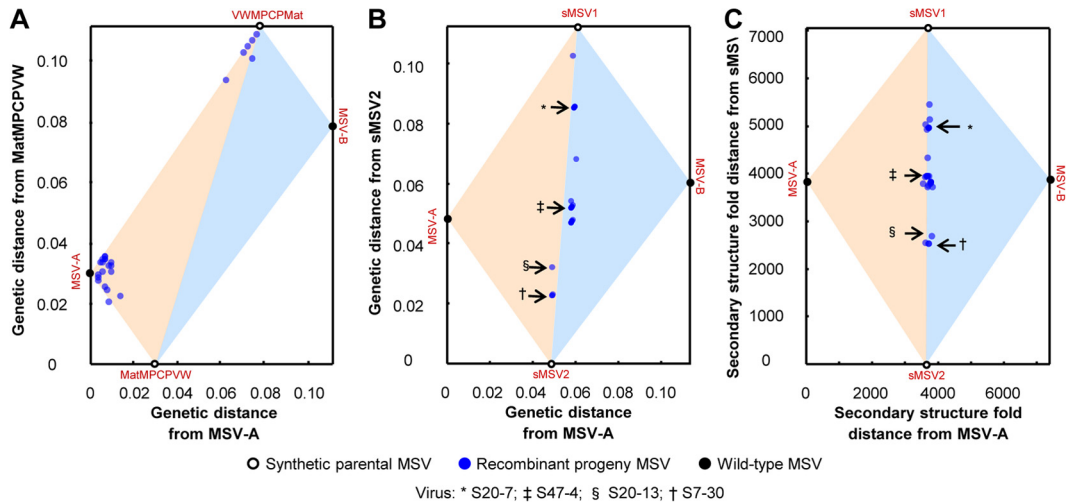
FIG 4 Genetic distances (A and B) and secondary-structure folding distances (C) of recombinant viruses from MSV-A, MSV-B, sMSV1, and sMSV2. The recombinants shown are those obtained after coinfection of maize cv. Golden Bantam with the reciprocally recombinant viruses MatV1V2VW and VWV1V2Mat (A) (data taken from a previous study examining the efficiency of secondary recombination to restore fitness deficits in recombinants with only two breakpoints [27]) or sMSV1 and sMSV2 (B). Whereas panels A and B indicate genetic distances between the recombinant viruses and the four reference viruses (one reference virus at each corner of the colored rhombus), panel C indicates estimated genomic secondary-structure folding distances between recombinants and the reference viruses. Whereas there is very clear evidence of recombinants converging on the MSV-A genotype (which is presumed to be the best recombination solution) in panel A, there is no evidence of this convergence in panels B and C. Infectivity increases relative to the infectivities of sMSV1 and sMSV2 were verified for the recombinants indicated with the symbols * (S20-7), ‡ (S47-4), § (S20-13), and † (S7-30) (Fig. 5).

However, within the recombinants obtained here, we found no evidence of a bias favoring recombination breakpoints occurring preferentially in noncoding regions ($P = 0.149$) (see Table S1 in the supplemental material).

In addition, recombination breakpoint distributions found in both field-isolated geminiviruses and geminiviruses arising during recombination experiments display well-defined hot and cold spots (26, 45, 46, 48–50) that can strongly influence the efficiency of adaptive evolution via recombination (27). Here, however, we observed little evidence of recombination occurring at known recombination hot spots in the MSV genome and, in fact, found a number of breakpoints falling within known recombination cold spots (Fig. 3B).

This finding suggested that particular genomic features, such as secondary structures within ssDNA (51), recombination-prone nucleotide motifs, or DNA methylation patterns, that predispose natural geminivirus genomes to accumulate breakpoints at certain sites more than at others might have been disrupted within the synthetic genomes. It is also possible, however, that strong selection favoring the survival of rare recombinants with breakpoints outside recombination hot spots simply skewed the distribution of breakpoints evident within the secondary recombinants observed here.

Secondary recombinants converge on a predicted high-fitness genotype. In previously reported recombination experiments involving mixed infections of simple MSV-A/MSV-B reciprocal chimeras in maize, it was noted that recombination very frequently and repeatedly reconstituted genomes resembling MSV-A, a sequence representing an optimal "high-fitness solution" in maize (26) (Fig. 4A). However, the eight genetically distinct recombinants retrieved here were not significantly more MSV-A- or MSV-B-like than could be accounted for by chance (95% confidence interval of the mean genetic distance to MSV-A of −0.00313 to 0.00082, where positive and negative values indi-

cate genetic similarities trending toward MSV-A and MSV-B, respectively) (Fig. 4B; see also Table S2 in the supplemental material).

This difference between previously reported MSV recombination experiments involving reciprocal chimeras with only 2 breakpoints and our experiment involving reciprocal chimeras with 182 breakpoints probably reflects the fact that, despite the theoretical possibility of reconstituting an MSV-A genome through 91 successive recombination events (with each event necessarily involving two breakpoints), there are either very few or no incremental recombination steps from sMSV1 and sMSV2 toward the MSV-A genotype that would yield fitness gains greater than that accessible on a completely different evolutionary trajectory.

It is interesting, therefore, that the one aspect of the recombinants recovered in our experiment which mirrors that seen in previously reported MSV-A/MSV-B recombination experiments is that they are apparently converging on a particular high-fitness solution (Fig. 3A). Crucially, this solution is very close to that predicted during our efforts to determine which components of the sMSV1 and sMSV2 genomes were most defective, that is, a virus containing the cell-to-cell movement-associated genome region of sMSV2 (*mp*) with the remainder of its genome being derived from sMSV1 (Fig. 3A).

Secondary recombinants do not detectably recover wild-type genomic secondary structures. Since other MSV evolution experiments have identified extremely strong selection pressures favoring the restoration of mutationally disrupted genomic ssDNA secondary structures (34, 52), we sought to determine whether, rather than converging on the MSV-A sequence, the secondary recombinant viruses displayed any evidence of an overall tendency toward converging on the nucleic acid secondary structures of either the MSV-B or MSV-A genotype.

We therefore computationally inferred secondary structures of MSV-A, MSV-B, sMSV1, sMSV2, and the secondary recombinant

viruses (Fig. 4C; see also Table S3 in the supplemental material) and calculated fold distances between all these and four reference genomes: MSV-A, MSV-B, sMSV1, and sMSV2. Relative to the reference sequences, these fold distances accounted for both base-pairing interactions that were predicted to have been disrupted within the recombinant genomes and novel (and therefore potentially maladaptive) base-pairing interactions that were created within these genomes.

We found that relative to the folds of sMSV1 and sMSV2, the genomic secondary structures of the recombinant viruses displayed no significant tendencies toward acquiring genomic secondary structures that were either more MSV-A- or more MSV-B-like (95% confidence interval of the mean structural fold distance of -65.99 to 7.09, where positive and negative values indicate structural folding trending toward those of the MSV-A and MSV-B genomes, respectively). This indicates that selection favoring the restoration of wild-type secondary structures has not obviously favored the survival of the secondary recombinants detected here.

Recombinant viruses are more infectious than both sMSV1 and sMSV2. In order to test whether the observed recombination events were adaptive, we constructed infectious clones for four (S7-30, S20-7, S20-13, and S47-4) of the eight distinct recombinant viruses retrieved from sMSV1 and sMSV2 coinfection experiments (Fig. 3) and tested the infectivity of these clones by agroinoculation of 3-day-old MSV-sensitive maize seedlings. These particular clones were selected for this test because they contained no extra mutations; i.e., differences in their viability relative to sMSV1 and sMSV2 would be entirely attributable to the recombination events that had occurred to produce these viruses. As a proxy for fitness, we measured infection frequencies, days until symptom development, and percent chlorotic leaf areas, as previously described (35, 36).

All of the secondary recombinants were clearly viable and induced symptomatic infections in maize, something that their parental viruses were apparently unable to do. As expected, however, all of them induced symptoms that appeared later than those caused by the MSV-A and MSV-B isolates, for which symptoms took on average 7 days to appear (data not shown). Of the four secondary recombinants, S47-4 had the highest infection frequency (70%) (Fig. 5A), whereas S20-7 had the lowest (15%) (Fig. 5A). S47-4 also induced symptoms significantly earlier than did either S20-7 ($P = 0.007$ by Mann-Whitney U test) (Fig. 5B) or S7-30 ($P = 0.05$ by Mann-Whitney U test). We also found statistical evidence for a difference in the percent chlorotic leaf area when we compared the amounts of leaf chlorosis induced by S47-4 versus S20-13 ($P = 0.01$) and S20-13 versus S20-7 ($P = 0.006$).

Conclusion. Using an MSV model, we have demonstrated that while extreme degrees of recombination can be highly deleterious, much of the damage caused to intragenome genetic interactions by this type of recombination can be rapidly mitigated by small numbers of secondary recombination events.

The two synthetic reciprocally recombinant viral genomes that we investigated contained thorough mixtures of polymorphisms derived from two divergent MSV isolates. Although these chimeras represented recombinants with the maximum number of breakpoints achievable through homologous recombination between these viruses, it is highly implausible that anything resembling these chimeras would ever naturally occur. Despite both chimeras being highly defective, they displayed complementary
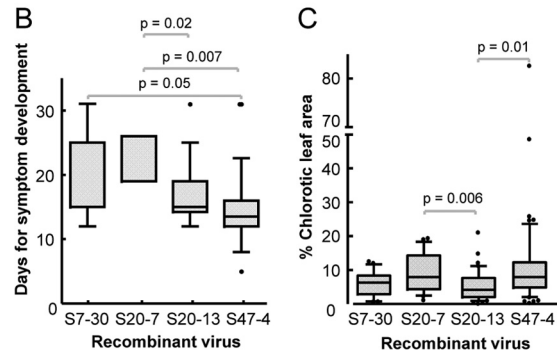
FIG 5 Infectivity and symptom phenotypes of sMSV1/sMSV2 recombinants. The infectivities and virulences of four recombinant viruses arising during sMSV1 and sMSV2 coinfections were assessed by measuring their infection frequencies (A), the number of days postinoculation before chlorotic streak symptoms appeared (B), and the mean chlorotic leaf areas that they induced in infected plants on leaves 4, 5, and 6 (with the 95% confidence intervals of these estimates indicated by error bars) (C).

genomic defects, which suggested that small numbers of secondary recombination events between them could produce genomes with considerably increased fitness. This prediction was confirmed when the synthetic viruses were given the opportunity to recombine within a series of mixed infections.

Although predictable, this result was not obvious because despite the two synthetic viral genomes collectively containing every polymorphism necessary to produce an MSV-A-like genotype with an extremely high degree of infectivity in maize, the secondary recombinant genomes that emerged during our experiment displayed no evidence of convergence upon this genotype. This indicates that when confronted with an optimization problem, in this case increased MSV fitness in maize, adaptive evolution through recombination might be efficient only at finding a reasonably good solution that is easily accessible, in this case replacing the cell-to-cell movement-associated genome region of one chimera with that of the other. For better, but more complex, solutions, i.e., the simultaneous repair of disrupted genetic interactions and the enrichment of maize-adaptive MSV-A polymorphisms that were distributed between the two synthetic genomes, it might be difficult for recombination to randomly generate enough genome mixtures to ensure that some of these will have appreciably increased fitness. One obvious reason why solving such problems with recombination might be so hard is that in viruses such as MSV, recombination breakpoints might simply not happen frequently enough to finely shuffle closely linked polymorphisms so as to bring together those that are most adaptive.

ACKNOWLEDGMENTS

REFERENCES

1. **Felsenstein J.** 1974. The evolutionary advantage of recombination. Genetics **78**:737–756.
2. **Keightley PD, Otto SP.** 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature **443**:89–92. http://dx.doi.org/10.1038/nature05049.
3. **Rice WR.** 2002. Experimental tests of the adaptive significance of sexual recombination. Nat. Rev. Genet. **3**:241–251. http://dx.doi.org/10.1038/nrg760.
4. **Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Peng Y, Yi X, Jiang D.** 2011. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. J. Virol. **85**:9863–9876. http://dx.doi.org/10.1128/JVI.00828-11.
5. **Krupovic M, Forterre P.** 2011. Microviridae goes temperate: microvirus-related proviruses reside in the genomes of Bacteroidetes. PLoS One **6**:e19893. http://dx.doi.org/10.1371/journal.pone.0019893.
6. **Belyi VA, Levine AJ, Skalka AM.** 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the Parvoviridae and Circoviridae are more than 40 to 50 million years old. J. Virol. **84**:12458–12462. http://dx.doi.org/10.1128/JVI.01789-10.
7. **Gibbs MJ, Weiller GF.** 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. Proc. Natl. Acad. Sci. U. S. A. **96**:8022–8027. http://dx.doi.org/10.1073/pnas.96.14.8022.
8. **Diemer GS, Stedman KM.** 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biol. Direct **7**:13. http://dx.doi.org/10.1186/1745-6150-7-13.
9. **Saunders K, Stanley J.** 1999. A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: evidence for interfamilial recombination between plant DNA viruses. Virology **264**:142–152. http://dx.doi.org/10.1006/viro.1999.9948.
10. **Rokyta DR, Wichman HA.** 2009. Genic incompatibilities in two hybrid bacteriophages. Mol. Biol. Evol. **26**:2831–2839. http://dx.doi.org/10.1093/molbev/msp199.
11. **Martin DP, Rybicki EP.** 2002. Investigation of Maize streak virus pathogenicity determinants using chimaeric genomes. Virology **300**:180–188. http://dx.doi.org/10.1006/viro.2002.1458.
12. **Martin DP, Van Der Walt E, Posada D, Rybicki EP.** 2005. The evolutionary value of recombination is constrained by genome modularity. PLoS Genet. **1**:e51. http://dx.doi.org/10.1371/journal.pgen.0010051.
13. **Springman R, Kapadia-Desai DS, Molineux IJ, Bull JJ.** 2012. Evolutionary recovery of a recombinant viral genome. G3 (Bethesda) **2**:825–830. http://dx.doi.org/10.1534/g3.112.002758.
14. **Bull JJ, Springman R, Molineux IJ.** 2007. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. Mol. Biol. Evol. **24**:900–908. http://dx.doi.org/10.1093/molbev/msm006.
15. **Jain R, Rivera MC, Lake JA.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. U. S. A. **96**:3801–3806. http://dx.doi.org/10.1073/pnas.96.7.3801.
16. **Galli A, Kearney M, Nikolaitchik OA, Yu S, Chin MPS, Maldarelli F, Coffin JM, Pathak VK, Hu W-S.** 2010. Patterns of human immunodeficiency virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. J. Virol. **84**:7651–7661. http://dx.doi.org/10.1128/JVI.00276-10.
17. **Escriu F, Fraile A, García-Arenal F.** 2007. Constraints to genetic exchange support gene coadaptation in a tripartite RNA virus. PLoS Pathog. **3**:e8. http://dx.doi.org/10.1371/journal.ppat.0030008.
18. **Lefeuvre P, Lett J-M, Reynaud B, Martin DP.** 2007. Avoidance of protein fold disruption in natural virus recombinants. PLoS Pathog. **3**:e181. http://dx.doi.org/10.1371/journal.ppat.0030181.
19. **Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefeuvre P, Martin DP, Robertson DL, Negroni M.** 2009. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. PLoS Pathog. **5**:e1000418. http://dx.doi.org/10.1371/journal.ppat.1000418.
20. **Woo J, Robertson D, Lovell S.** 2014. Constraints from protein structure and intra-molecular coevolution restrict recombination of HIV-1. Virology **454–455**:34–39. http://dx.doi.org/10.1016/j.virol.2014.01.029.
21. **Elena SF, Solé RV, Sardanyés J.** 2010. Simple genomes, complex interactions: epistasis in RNA virus. Chaos **20**:026106. http://dx.doi.org/10.1063/1.3449300.
22. **Parera M, Perez-Alvarez N, Clotet B, Martínez MA.** 2009. Epistasis among deleterious mutations in the HIV-1 protease. J. Mol. Biol. **392**:243–250. http://dx.doi.org/10.1016/j.jmb.2009.07.015.
23. **Van Opijnen T, Boerlijst MC, Berkhout B.** 2006. Effects of random mutations in the human immunodeficiency virus type 1 transcriptional promoter on viral fitness in different host cell environments. J. Virol. **80**:6678–6685. http://dx.doi.org/10.1128/JVI.02547-05.
24. **Sanjuán R, Moya A, Elena SF.** 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. Proc. Natl. Acad. Sci. U. S. A. **101**:15376–15379. http://dx.doi.org/10.1073/pnas.0404125101.
25. **Lalić J, Elena SF.** 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. Heredity (Edinb.) **109**:71–77. http://dx.doi.org/10.1038/hdy.2012.15.
26. **Van der Walt E, Rybicki EP, Varsani A, Polston JE, Billharz R, Donaldson L, Monjane AL, Martin DP.** 2009. Rapid host adaptation by extensive recombination. J. Gen. Virol. **90**:734–746. http://dx.doi.org/10.1099/vir.0.007724-0.
27. **Monjane AL, van der Walt E, Varsani A, Rybicki EP, Martin DP.** 2011. Recombination hotspots and host susceptibility modulate the adaptive value of recombination during maize streak virus evolution. BMC Evol. Biol. **11**:350. http://dx.doi.org/10.1186/1471-2148-11-350.
28. **Martin DP, Willment JA, Billharz R, Velders R, Odhiambo B, Njuguna J, James D, Rybicki EP.** 2001. Sequence diversity and virulence in Zea mays of Maize streak virus isolates. Virology **288**:247–255. http://dx.doi.org/10.1006/viro.2001.1075.
29. **Willment JA, Martin DP, Rybicki EP.** 2001. Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data. J. Virol. Methods **93**:75–87. http://dx.doi.org/10.1016/S0166-0934(00)00299-8.
30. **Willment JA, Martin DP, Van der Walt E, Rybicki EP.** 2002. Biological and genomic sequence characterization of Maize streak virus isolates from wheat. Phytopathology **92**:81–86. http://dx.doi.org/10.1094/PHYTO.2002.92.1.81.
31. **Martin D, Rybicki E.** 2000. Improved efficiency of Zea mays agroinoculation with Maize streak virus. Plant Dis. **84**:1096–1098. http://dx.doi.org/10.1094/PDIS.2000.84.10.1096.
32. **Lucy AP, Boulton MI, Davies JW, Maule AJ.** 1996. Tissue specificity of Zea mays infection by maize streak virus. Am. Phytopathol. Soc. **9**:22–31.
33. **Schnippenkoetter WH, Martin DP, Willment JA, Rybicki EP.** 2001. Forced recombination between distinct strains of Maize streak virus. J. Gen. Virol. **82**:3081–3090. http://vir.sgmjournals.org/content/82/12/3081.long.
34. **Shepherd DN, Martin DP, McGivern DR, Boulton MI, Thomson JA, Rybicki EP.** 2005. A three-nucleotide mutation altering the Maize streak virus Rep pRBR-interaction motif reduces symptom severity in maize and partially reverts at high frequency without restoring pRBR-Rep binding. J. Gen. Virol. **86**:803–813. http://dx.doi.org/10.1099/vir.0.80694-0.
35. **Martin DP, Rybicki EP.** 1998. Microcomputer-based quantification of maize streak virus symptoms in Zea mays. Phytopathology **88**:422–427. http://dx.doi.org/10.1094/PHYTO.1998.88.5.422.
36. **Martin DP, Willment JA, Rybicki EP.** 1999. Evaluation of maize streak virus pathogenicity in differentially resistant Zea mays genotypes. Phytopathology **89**:695–700. http://dx.doi.org/10.1094/PHYTO.1999.89.8.695.
37. **Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T.** 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase. J. Virol. Methods **116**:209–211. http://dx.doi.org/10.1016/j.jviromet.2003.11.015.
38. **Shepherd DN, Martin DP, Lefeuvre P, Monjane AL, Owor BE, Rybicki EP, Varsani A.** 2008. A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. J. Virol. Methods **149**:97–102. http://dx.doi.org/10.1016/j.jviromet.2007.12.014.
39. **Owor BE, Shepherd DN, Taylor NJ, Edema R, Monjane AL, Thomson JA, Martin DP, Varsani A.** 2007. Successful application of FTA Classic Card technology and use of bacteriophage phi29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. J. Virol. Methods **140**:100–105. http://dx.doi.org/10.1016/j.jviromet.2006.11.004.
40. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739. http://dx.doi.org/10.1093/molbev/msr121.

41. **Owor BE, Martin DP, Rybicki EP, Thomson JA, Bezuidenhout ME, Lakay FM, Shepherd DN.** 2011. A rep-based hairpin inhibits replication of diverse maize streak virus isolates in a transient assay. J. Gen. Virol. **92:**2458–2465. http://dx.doi.org/10.1099/vir.0.032862-0.

42. **Ruschhaupt M, Martin DP, Lakay F, Bezuidenhout M, Rybicki EP, Jeske H, Shepherd DN.** 2013. Replication modes of Maize streak virus mutants lacking RepA or the RepA-pRBR interaction motif. Virology **442:**173–179. http://dx.doi.org/10.1016/j.virol.2013.04.012.

43. **Markham NR, Zuker M.** 2008. Software for nucleic acid folding and hybridization. Bioinformatics **453:**1–28. http://dx.doi.org/10.1007/978-1-60327-429-6_1.

44. **Grimsley N, Hohn B, Hohn T, Walden R.** 1986. "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. Proc. Natl. Acad. Sci. U. S. A. **83:**3282–3286. http://dx.doi.org/10.1073/pnas.83.10.3282.

45. **Owor BE, Martin DP, Shepherd DN, Edema R, Monjane AL, Rybicki EP, Thomson JA, Varsani A.** 2007. Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. J. Gen. Virol. **88:**3154–3165. http://dx.doi.org/10.1099/vir.0.83144-0.

46. **Varsani A, Shepherd DN, Monjane AL, Owor BE, Erdmann JB, Rybicki EP, Peterschmitt M, Briddon RW, Markham PG, Oluwafemi S, Windram OP, Lefeuvre P, Lett JM, Martin DP.** 2008. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. J. Gen. Virol. **89:**2063–2074. http://dx.doi.org/10.1099/vir.0.2008/003590-0.

47. **Lefeuvre P, Martin DP, Harkins G, Lemey P, Gray AJA, Meredith S, Lakay F, Monjane A, Lett J-M, Varsani A, Heydarnejad J.** 2010. The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathog. **6:**e1001164. http://dx.doi.org/10.1371/journal.ppat.1001164.

48. **Martin DP, Lefeuvre P, Varsani A, Hoareau M, Semegni J-Y, Dijoux B, Vincent C, Reynaud B, Lett J-M.** 2011. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. PLoS Pathog. **7:**e1002203. http://dx.doi.org/10.1371/journal.ppat.1002203.

49. **Albuquerque LC, Inoue-Nagata AK, Pinheiro B, Resende RO, Moriones E, Navas-Castillo J.** 2012. Genetic diversity and recombination analysis of sweepoviruses from Brazil. Virol. J. **9:**241. http://dx.doi.org/10.1186/1743-422X-9-241.

50. **Lefeuvre P, Lett J-M, Varsani A, Martin DP.** 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. **83:**2697–2707. http://dx.doi.org/10.1128/JVI.02152-08.

51. **García-Andrés S, Tomás DM, Sánchez-Campos S, Navas-Castillo J, Moriones E.** 2007. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. Virology **365:**210–219. http://dx.doi.org/10.1016/j.virol.2007.03.045.

52. **Shepherd DN, Martin DP, Varsani A, Thomson JA, Rybicki EP, Klump HH.** 2006. Restoration of native folding of single-stranded DNA sequences through reverse mutations: an indication of a new epigenetic mechanism. Arch. Biochem. Biophys. **453:**108–122. http://dx.doi.org/10.1016/j.abb.2005.12.009.

53. **Palmer K.** 1997. Investigations into the use of Maize streak virus as a gene vector. Ph.D. thesis. University of Cape Town, Cape Town, South Africa.