Published in final edited form as:

Virus Res. 2013 September ; 176(0): 259–264. doi:10.1016/j.virusres.2013.06.016.

Zucchini yellow mosaic virus (ZYMV, Potyvirus): Vertical transmission, seed infection and cryptic infections

H.E. Simmons^{1,2}, J.P. Dunham³, K. E. Zinn², G.P. Munkvold¹, E.C. Holmes⁴, and A.G. Stephenson²

¹Seed Science Center, Iowa State University, Ames, IA 50011, USA

²Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

³Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90033, USA

⁴Sydney Emerging Infections & Biosecurity Institute, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia

Abstract

The role played by seed transmission in the evolution and epidemiology of viral crop pathogens remains unclear. We determined the seed infection and vertical transmission rates of zucchini yellow mosaic virus (ZYMV), in addition to undertaking Illumina sequencing of nine vertically transmitted ZYMV populations. We previously determined the seed-to-seedling transmission rate of ZYMV in *Cucurbita pepo* ssp. *texana* (a wild gourd) to be 1.6%, and herein observed a similar rate (1.8%) in the subsequent generation. We also observed that the seed infection rate is substantially higher (21.9%) than the seed-to-seedling transmission rate, suggesting that a major population bottleneck occurs during seed germination and seedling growth. In contrast, that two thirds of the variants present in the horizontally transmitted inoculant population were also present in the vertically transmitted populations implies that the bottleneck at vertical transmission may not be particularly severe. Strikingly, all of the vertically infected plants were symptomless in contrast to those infected horizontally, suggesting that vertical infection may be cryptic. Although no known virulence determining mutations were observed in the vertically infected samples, the 5' untranslated region was highly variable, with at least 26 different major haplotypes in this region compared to the two major haplotypes observed in the horizontally transmitted population. That the regions necessary for vector transmission are retained in the vertically infected populations, combined with the cryptic nature of vertical infection, suggests that seed transmission may be a significant contributor to the spread of ZYMV.

Keywords

zucchini yellow mosaic virus; *Potyvirus*; seed transmission; Illumina sequencing; Symptomless infection

© 2013 Elsevier B.V. All rights reserved.

Correspondence to: H. E. Simmons, Seed Science Center, Iowa State University, Ames, IA 50010. USA. hsimmons@iastate.edu; Tel: +1 515 294 6892; Fax: +1 515 294 2014.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Zucchini yellow mosaic virus (ZYMV; family *Potyviridae*) is a single-stranded positive-sense RNA virus that is an important pathogen of cucurbits (squash, melon and cucumbers). In addition to the distinctive yellow mottling of the leaves, the symptoms of ZYMV include stunting of the entire plant as well as fruit distortion and mottling (Desbiez and Lecoq, 1997). These symptoms can render the fruits unmarketable such that ZYMV can reduce agricultural yields by up to 94% (Blua and Perring, 1989). Cucurbits are one of the 15 most important agricultural crops in the United States, and their production is valued at approximately \$1.5 billion USD per year (Cantliffe et al., 2007) making this a viral pathogen of significant concern for agriculture.

The ZYMV genome consists of an ~9.6 Kb ZYMV open reading frame (ORF) encoding a single polyprotein precursor that is subsequently processed into ten proteins by three virally encoded proteases (Gal-On, 2007). An additional ORF, PIPO, within the P-3 coding region is translated in the +2 reading frame (Chung et al., 2008). The ZYMV 5' untranslated region (UTR) is thought to contain two regulatory regions that are believed to direct cap-independent translation (Niepel and Gallie, 1999) via interactions with the poly-A tail (Gallie, 2001). Importantly, ZYMV can be transmitted both horizontally by aphids and vertically by seeds, although the predominant method is by aphids in a non-persistent manner. A large number of aphid species (26) have been shown to be capable of transmitting the virus (Katis et al., 2006), although two species, *Myzus persicae* and *Aphis gossypii*, have the highest reported transmission efficiencies (41% and 35%, respectively) (Castle et al., 1992). Vector transmission occurs as a result of an interaction between the aphid stylet, and two viral proteins – the coat protein (CP) and the helper component protein (HC-Pro). Specifically, the DAG motif on the CP interacts with the PTK region of the HC-Pro, and a secondary motif on the HC-Pro (KLSC) interacts with the aphid stylet (Urcuqui-Inchima et al., 2001). Although ZYMV can be transmitted vertically by seed, this is less common than horizontal transmission via aphids, and we previously reported a seed-to-seedling transmission rate of only 1.6% (Simmons et al., 2011a).

Approximately 20% of viral plant pathogens are known to be seed-transmitted, and it is possible that one-third will eventually be shown to be transmitted in this manner (Johansen et al., 1994). Although seed transmission is commonplace in the *Potyviridae*, the number of virions entering the seed or even the mechanism by which the virus enters the seed is unknown. There is, however, some evidence in pea seed-borne mosaic virus that the virus may directly invade the embryo via the suspensor (Wang and Maule, 1994). In this case it is thought that the virus travels from the maternal cells in the micropyle to the endospermic cytoplasm and embryonic suspensor to the embryo (Roberts et al., 2003). Evidence for the indirect invasion of the embryo via invasion of reproductive meristematic tissue early in plant development has been demonstrated in barley streak mosaic virus (Carroll, 1981). Given that seed transmission is most likely a mechanism by which viruses persist between crop seasons, and that even a seed transmission rate as low as 0.001 can initiate an epidemic (Ryder, 1973), understanding how seed transmission shapes viral genetic diversity is fundamental to agriculture. It is also possible that the success of seed transmission is dictated by temporal or spatial shifts in the virus population as it infects the seed prior to causing infection in the subsequent seedling.

Natural populations of RNA viruses rapidly generate genetic diversity because of a combination of high mutation rates, rapid replication, and large population sizes (Duffy et al., 2008). However, there are at least three bottlenecks that are thought to restrict this genetic diversity in plant viral populations: (i) vector acquisition/transmission, (ii) systemic movement within the host plant, and (iii) during seed (vertical) transmission. For example, it

has been demonstrated that genetic variation is reduced as a result of inter-plant transmission (Ali et al., 2006; Hall et al., 2001), with the number of virions transmitted per event by aphids estimated to be on the order of 1-3 for potato virus Y and cucumber mosaic virus (Betancourt et al., 2008; Moury et al., 2007). Within-plant systemic movements include cell-to-cell in a tissue, as well as movements through the vascular system. Current estimates of cell-to-cell movement are approximately six virions for the first cell-to-cell movement and five for subsequent movements in soil-borne wheat mosaic virus (Miyashita and Kishino, 2010), with similar numbers reported for tobacco mosaic virus (TMV) (Gonzalez-Jara et al., 2009). However, this bottleneck appears to be dependent on the stage of infection. In cauliflower mosaic virus (CaMV) the number of virions that successfully replicate after cell entry is low in the early stages of infection (two virions), increases during infection (13 virions), and drops back to initial levels as host senescence occurs (Gutierrez et al., 2010). Estimates of the founding population of TMV in a new leaf after systemic infection are 2-20 (Sacristan et al., 2003), and approximately four Wheat streak mosaic virions were found in a new tiller (French and Stenger, 2003). Similarly, it was determined that the number of mutants decreased as a function of distance from the inoculated leaf (Li and Roossinck, 2004). Population bottlenecks also appear to be dependent on the stage of infection, with the greatest effect during early and late infection, yet relatively relaxed during mid-infection (Gutierrez et al., 2012). However, despite this body of knowledge there are currently no data on the extent and consequences of the genetic bottleneck(s) associated with seed transmission.

To investigate the nature and evolutionary impact of vertical transmission in ZYMV, including the effect of the bottleneck imposed on genetic diversity, we performed Illumina (i.e. deep) sequencing of nine vertically transmitted ZYMV populations, four from the first generation of vertical transmission and five from the subsequent generation (vertical to vertical transmission). In addition, the seed transmission rate of the second generation of vertically transmitted ZYMV was determined, as was the seed infection rate and germination rate of seeds harvested from ZYMV infected fruits. The host plant used, *Cucurbita pepo* ssp. *texana*, is thought to be the progenitor of domestic squashes (Decker and Wilson, 1987) and is considered to be the optimal host for the maintenance of ZYMV (Gal-On, 2007).

2. Methods

2.1 ZYMV experimental design and sample collection

Approximately 6,000 seeds (count estimated by weight) were harvested from ZYMV infected *Cucurbita pepo* ssp. *texana* plants, that displayed typical viral symptoms, from a 0.4 ha experimental field at The Pennsylvania State University Agricultural Research Farm at the end of the 2008 growing season. To ensure that any infection was the result of embryonic infection as opposed to virus particles remaining on the seed coat, the seeds were extracted in 4% hydrochloric acid and washed in a 10% bleach solution. The seeds were subsequently germinated in flats in a greenhouse at The Pennsylvania State University. Of the 6,000 seeds planted, 3,195 germinated and infection status of 2,910 seedlings was ascertained by visual inspection. Two seedlings exhibited extremely slight symptoms. Despite the low symptom severity these were tested for ZYMV via RT-PCR; as both tested positive they were submitted for Illumina sequencing (FG1 and FG2). Given the lack of symptom expression, visual inspection was abandoned and the remaining 283 seedlings were batched into groups of ten and tested via RT-PCR for ZYMV.

In 2009, 180 *C. pepo* ssp. *texana* seedlings were transplanted into a 0.4 ha plot at our field site, and in late June one plant was inoculated with ZYMV using a portion of a leaf collected during the 2008 field season. The other portion of the leaf was saved for Illumina

sequencing. ZYMV then spread through the field population via aphid transmission. The fruits from ZYMV-infected plants were collected at the end of the 2009 season. Seeds collected from visibly infected plants were extracted and cleaned as described above, and seeds from each plant were pooled. These seeds were planted in flats in the greenhouse and allowed to germinate, and a leaf was collected from each seedling at the third true leaf stage. A total of 2,336 samples were collected and pooled into batches of ten for RNA extraction and RT-PCR determination of ZYMV using the methods described above. If a batch tested positive for ZYMV, each individual within that group was then tested for ZYMV. Two of the 36 ZYMV positive samples were submitted for Illumina sequencing (FG3 and FG4).

Nine of the 36 vertically ZYMV infected plants from the greenhouse were grown in a garden plot that was isolated by distance (approximately 10 miles) from wild and cultivated cucurbit species. The seedlings were transplanted and covered by cages made from anti-aphid netting (ProtekNet 120gr, Dubois Agroinnovation) to prevent the plants from becoming infected with aphid-transmitted ZYMV. Given that the bee pollinators were also excluded by the netting and that *C. pepo* is monoecious (i.e. has separate male and female flowers on the same plant) the female flowers were hand-pollinated by gently touching an anther from a male flower to the stigma of the female.

Once the fruits reached maturation the seeds were harvested and cleaned as described above, stored for 5 months, and then planted in flats in the greenhouse and allowed to germinate. From the resulting 1,010 germinations, leaf samples were harvested and then stored at -80°C. A portion of each sample was pooled into batches of ten for extraction and RT-PCR. Twelve of these groups tested positive ZYMV and individual testing was undertaken on each sample for these groups. A total of 18 ZYMV positive samples were found, five of which were submitted for Illumina sequencing (SG1, SG2, SG3, SG4 and SG5).

2.2 Germination test

Germination testing was undertaken in the Seed Testing Lab at Iowa State University, using the Association of Official Seed Analysts (AOSA)-approved germination test. Seeds were placed on moistened rolled towels, 50 per replication with eight replications for a total of 400 seeds. These were placed in buckets, covered with plastic wrap, and allowed to germinate for two weeks with a diurnal cycle of eight hours at 30°C with light, and 16 hours at 20°C in the dark. The seeds were scored three times over the two-week period (day 7, day 11 and day 14), and the towels remoistened at days 7 and eleven.

2.3 RNA isolation, RT-PCR and sequencing of seedlings

The methods used for RNA isolation, RT-PCR and Illumina sequencing of the seedling samples are as previously described (Simmons et al., 2012) with the exception that the E.Z.N.A.® RNA isolation kit (Omega bio-tek) was used to isolate RNA from frozen leaf samples.

2.4 RNA Extraction, first strand synthesis and qPCR of seed samples

The ZYMO Direct-zol™ RNA MiniPrep w/ TRI-Reagent® RNA extraction kit was used following the manufacturer's protocol to extract RNA from seed. First strand cDNA synthesis was performed using a Superscript III First-strand Synthesis kit (Invitrogen) following the protocol provided by the supplier. To determine the ZYMV seed infection rate, RNA extractions were performed on 96 individual seeds harvested from ZYMV infected *C. pepo* fruits. After extraction the 96 samples were pooled into eight groups of 12 and performed cDNA synthesis and qPCR on these. If a group tested positive for ZYMV, each individual within that group was then tested for ZYMV. Through this protocol 21 of the 96 seeds were found to be positive.

The qPCR primers were designed based on the reference strain (GenBank accession no. NC_003224.1) using Primer Express® Software version 3.0 (Applied Biosystems). The primers fall within the CI protein; forward primer: 5'-GGACAGTGC GACTATAGCTTCAA-3' and reverse primer: 5'-TTTAACCGCGAATTGCGTATC-3'. The PCR mix was as follows: 12.5µl SYBER green, 0.15µl of each primer, 5.35µl of PCR grade water and 2.0µl of template for a total reaction volume of 20µl. The PCR was carried out in an Applied Biosystems StepOnePlus™ Real-Time PCR-System and the cycling conditions were as follows: holding temperature of 95°C for 5 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. The standard curve was produced by creating a dilution series of cDNA stock of ZYMV ranging from 531.7ng/µl to 20.5ng/µl. We were unable to detect the 2.43 ng/µl dilution. The amount of DNA present in the seed samples ranged from 11.3 ng/µl to 60.3 ng/µl as compared to the standard curve.

2.5 PCR Cloning and Sanger Sequencing of CP sequences

The methods used for PCR cloning and Sanger sequencing of the coat protein region of the samples SG1 and SG2 are as previously described (Simmons et al., 2011b).

2.6 Alignment of raw reads, read accuracy and variant calling

Sequence alignments were performed using the Burrows Wheeler aligner (BWA) version 0.6.2 allowing 10 mismatches (Li and Durbin, 2009). BAM to SAM file conversion and filtering was performed with Samtools version 0.1.18 (Li et al., 2009). Varscan (version 2.3.2) (Koboldt et al., 2012) was used to call the minor mutational variants. Both strand bias and low quality scores have been demonstrated to increase the number of miscalls obtained while mapping Illumina reads (Minoche et al., 2011; Nakamura et al., 2011). To account for strand bias any variants found had to be validated in both strands to be considered a true variant. To control for mapping quality any sites that had a phred-scaled quality score less than 20 as compared with the Illumina supplied control (PhiX 174) were excluded. According to Illumina, the inferred base call accuracy is 99% with this quality score. To ensure that any false positives were eliminated, and to account for methodological errors introduced as a result of the experimental procedures an extremely conservative approach was used when calling the minor variants. Specifically, the requirements used to determine minor variants were that they occurred at greater than 100X coverage, had a frequency of 1% or greater, and possessed a minimum of 10 reads at a particular nucleotide position. The 5' UTR (nt 1-138) region was found to be highly variable in comparison to the rest of the genome and therefore considered as a separate region with respect to analysis. ShoRAH (version 0.6) (Zagordi et al., 2011) was used to determine the haplotype frequencies in the 5'UTR region. In this case, the default parameters were used, we only considered haplotypes with a posterior probability equal to one, and to eliminate false positives we only considered haplotypes that had a coverage of at least 100x or greater. The genome wide SNP frequency plot was generated with GraphPad Prism version 6.0b for Mac OS X, GraphPad Software (La Jolla California USA) www.graphpad.com.

2.7 Sequence analysis

The ZYMV CP sequences were manually aligned using Se-AI (2.0a11; kindly provided by Andrew Rambaut, University of Edinburgh). To estimate the number of nonsynonymous (d_N) and synonymous substitutions (d_S) per site (ratio d_N/d_S) the Single Likelihood Ancestor Counting (SLAC) algorithm employing the MG94 × HY85_3 × 4 substitution model in HyPhy (Pond et al., 2005) was used. The 5'UTR haplotypes were aligned using SeaView (version 4) (Gouy et al., 2010). The annotation and effects of SNPs and indels were determined using SnpEff version 3.1 (Cingolani et al., 2012).

All the consensus nucleotide sequences generated here have been submitted to GenBank and assigned accession numbers KC665627-KC665635 and JQ716413.

3. Results and Discussion

We previously reported the seed transmission rate in the first generation of vertical transmission (i.e. seed-to-seedling) to be 1.6% (Simmons et al., 2011a). Strikingly, we found an almost identical rate for the second generation of seed to seedling transmission rate (1.8%), and we determined the seed infection rate to be 21.9%. Clearly, these data indicate that the bottleneck imposed on the viral population during vertical transmission is relatively severe such that only a subset of virally infected seeds are capable of successfully initiating infections in subsequent generations. This could be due to a number of reasons. It appears that viral infection of the mother plant is affecting the germination rate such that seeds harvested from infected fruits tend to germinate at a lower rate than those from healthy fruits. We ascertained that the germination rate of seeds extracted from fruits from infected parents was 22.5% (90/400) versus 87.5% (350/400) for those harvested from non-infected parents. For the seeds harvested from ZYMV infected fruit the percentage of abnormal seeds was 2.75% (11/400), and the percentage of seeds that did not germinate was 74.75% (299/400). For the seeds harvested from healthy fruit the percentage of abnormal seeds was 5% (20/400) and the percentage of seeds that did not germinate was 7.5% (30/400). Similar rates have been reported for alfalfa mosaic virus where the higher incidence of seed infection (20.6%) versus seedling infection (7.3%) was believed to be due to the higher incidence of virions in the seed coat versus the embryo (Pesic and Hiruki, 1986). It is also possible that the viral titers in the seeds are simply too low to consistently initiate effective infections in the subsequent generation, and these titers as determined by qPCR were several orders of magnitude lower than in the same quantity of leaf sample (seed range 11.3ng/ul -60ng-ul compared to the ZYMV infected leaf sample range 2,000 ng/ul – 3,400 ng/ul). Alternatively, the viral population may be severely constrained by the host plant such that only a subset of the viral population is transmitted from the seed to the seedling, or host defense mechanisms in the seed may be eliminating the viral population, or preventing it from being transmitted from the ephemeral cotyledons to the persistent seedling organs.

We sequenced a total of ten ZYMV samples: four from the first generation, five from the second, and the inoculant population. Overall, we achieved an average coverage depth of 7,974X (range 505X – 32,081X), and successfully sequenced 99.11% of the genome (range 98.91% -99.46%). Excluding the highly variable 5'UTR (i.e. nt position 139 onwards), there were a total of 106 mutations, 96 of which were single nucleotide polymorphisms (SNPs), nine were frameshift mutations and the remaining mutation was a stop (ochre) (Fig 1). Sixty-eight of the mutations were found in one sample only, 43 were found in the first generation only, and 25 in the second generation. The remaining 38 mutations were shared between at least two seedlings. Of the 38 shared mutations, 21 were found in both the first and second-generation seedlings, 12 were found in the second generation only and five in the first generation only. A total of 46/96 (48%) of the SNPs were non-synonymous and 18 of these were found in multiple samples, with the remaining 28 found in a single seedling only (Table S1). Sixteen of the 24 variants (excluding the 5'UTR) found in the inoculant were also found in the vertically transmitted populations.

The majority of SNPs in RNA viral populations are likely to be deleterious and as such are expected to be transient and subsequently purged from the population (Sanjuan et al., 2004). Interestingly, the ratio of nonsynonymous to synonymous substitutions (i.e. d_N/d_S) in the CP region of two populations (FG1 and FG2) of vertically transmitted ZYMV was 2.17 (range 0.86-4.39). This value is substantially higher than those estimated previously for clones of this virus (i.e. $d_N/d_S < 1.0$) (Simmons et al., 2008), and tentatively compatible with some

localized positive selection, although this is difficult to assess over such a short time span. In addition, the percentage of nonsynonymous mutations that were shared between samples (47%) increased in the vertically transmitted populations compared to the study on the horizontally transmitted populations in which only 35% of the nonsynonymous mutations were shared between plants (Simmons et al., 2012). Interestingly, 16 of the 24 variants found in the inoculant population were present in the vertically transmitted populations suggesting that the bottleneck at this stage may not be as severe as previously predicted. None of the regions needed for aphid transmission varied in any of the vertically transmitted samples, as was also the case for the regions necessary for replication and movement within the host plant. Clearly, additional work is required to reveal the fitness of these mutations.

All of our vertically transmitted populations resulted in symptomless infection; this is in contrast to our horizontally infected plants, all of which displayed typical ZYMV symptoms. Although this could be taken to imply that there are fixed differences in point mutations (particularly amino acid changes) between the two viral populations, none were observed here. In addition, although specific mutations are known to determine virulence in ZYMV, for instance, FRNK to FINK (Gal-On, 2000), FRNK to FRNK(X)₁₂ (Desbiez et al., 2010), R₉₁₇W (Desbiez et al., 2003) and R₁₈₀I, F₂₀₅L and E₃₁₆N (Lin et al., 2007), none of these mutations were observed here. However, we did observe a large number of haplotypes in the 5'UTR that did not occur in the rest of the genome, including a number of deletion mutations. In the highly variable 5'UTR region, there were only three samples that had sufficient coverage for the entire region – FG1 (average 45,392X), FG2 (average 1,336X) and SG4 (average 2,295X) – reflecting the fact that Illumina sequencing coverage tends to be lower in the 5'UTR region (Wang et al., 2011). FG1 and SG4 both had 26 haplotypes in this region and FG2 had 31. This is far greater than in the coding sequence in which there was only one major haplotype. The inoculant population, which is a horizontally transmitted population, had two major haplotypes in the 5'UTR (Fig. 2). All of the haplotypes from the seed transmitted viral populations shared a large deletion (nt 90 -133), relative to the NCBI reference (NC_003224.1), and a smaller deletion in the same region relative to the horizontally transmitted population (nt 98-127) (see Figure 2).

Previous work with other potyviruses has linked insertions and deletions in the 5'UTR region to symptom development. For example, plum pox virus mutants with a sequence deletion of nt 127-145 in the 5'UTR exhibited mild symptoms (Simon-Buela et al., 1997). Similarly, a seven nt insertion in the 5'UTR region in bean common mosaic virus resulted in more severe symptoms in asparagus bean (Zheng et al., 2002). Low sequence similarity has been observed in the 5'UTR region among members of the *Potyviridae* (Kneller et al., 2006) as well as within ZYMV isolates (Lin et al., 2001). Although our data imply that the symptomless nature of vertical infection in this viral pathogen may be due to mutations in the 5'UTR region, this clearly needs to be determined through functional studies. Indeed, it is also possible that the lack of disease symptoms could be due to environmental factors. For example, with pea early browning mosaic virus, seeds planted in the greenhouse produced symptomless plants, whereas seeds planted in the field displayed typical viral symptoms, indicating that environmental conditions may play a role in symptom development (Bos and van der Want, 1962). This is of epidemiological importance especially for crops that are cultivated in greenhouse settings. We previously determined that aphids are capable of infecting healthy *C. pepo* plants with ZYMV after feeding on vertically infected plants in the greenhouse (Simmons et al., 2011a), but what has yet to be determined is how symptom development would progress if these plants were subsequently planted in the field.

Given the economic impact of the potyviruses on agriculture, understanding how the genetic bottleneck experienced during vertical transmission contributes to epidemics is an essential component to controlling this and other viral pathogens. Overall, our findings suggest that

although a significant population bottleneck is imposed on the viral population during seed transmission, significant genetic diversity is transmitted via seeds. Hence, vertically infected seedlings may be instrumental in contributing to the worldwide dissemination of this pathogen, especially given the cryptic nature of the infection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Science Foundation Doctorial Dissertation Grant Program no. 1010881, the Biotechnology Risk Assessment Program Grant no. 2009-33120-20093 from the USDA National Institute of Food and Agriculture, and the USDA-FAS, Technical Assistance for Specialty Crops, grant #2011-23. JPD is supported by NIH RO1 EUREKA GM098741. ECH is supported by an NHMRC Australia Fellowship.

We thank Shelby Fleischer for kindly providing the netting used to make the cages. We thank Tony Omeis for greenhouse assistance as well as the use of the Biology Greenhouse, and R. Oberheim and his staff for use of the Horticulture Farm at the Penn State Agriculture Experiment Station at Rock Springs, PA.

References

- Ali A, Li HY, Schneider WL, Sherman DJ, Gray S, Smith D, Roossinck MJ. Analysis of genetic bottlenecks during horizontal transmission of cucumber mosaic virus. *Journal of virology*. 2006; 80(17):8345–8350. [PubMed: 16912285]
- Betancourt M, Fereres A, Fraile A, Garcia-Arenal F. Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *Journal of virology*. 2008; 82(24):12416–12421. [PubMed: 18842732]
- Blua MJ, Perring TM. Effect of zucchini yellow mosaic-virus on development and yield of cantaloupe (*Cucumis-Melo*). *Plant Disease*. 1989; 73(4):317–320.
- Bos L, van der Want JPH. Early browning of pea, a disease caused by a soil- and seed-borne virus. *Tijdschr PltZiekt*. 1962; 68:368.
- Cantliffe DJ, Shaw NL, Stoffella PJ. Current trends in cucurbit production in the US. *Proceedings of the IIIrd International Symposium on Cucurbits*. 2007; (731):473–478.
- Carroll, TW. Seedborne viruses: virus-host interactions. In: Maramorosch, K.; Harris, KF., editors. *Plant Diseases and Vectors: Ecology and epidemiology*. Academic Press; New York: 1981. p. 293-317.
- Castle SJ, Perring TM, Farrar CA, Kishaba AN. Field and laboratory transmission of watermelon mosaic virus-2 and zucchini yellow mosaic-virus by various aphid species. *Phytopathology*. 1992; 82(2):235–240.
- Chung BYW, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(15): 5897–5902. [PubMed: 18408156]
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly*. 2012; 6(2):80–92. [PubMed: 22728672]
- Decker DS, Wilson HD. Allozyme variation in *Cucurbita pepo* complex: *C. pepo* var. *overifera* vs. *C. texana*. *Systematic Botany*. 1987; 12:263–273.
- Desbiez C, Gal-On A, Girard M, Wipf-Scheibel C, Lecoq H. Increase in zucchini yellow mosaic virus symptom severity in tolerant zucchini cultivars is related to a point mutation in P3 protein and is associated with a loss of relative fitness on susceptible plants. *Phytopathology*. 2003; 93(12): 1478–1484. [PubMed: 18943611]

- Desbiez C, Girard M, Lecoq H. A novel natural mutation in HC-Pro responsible for mild symptomatology of Zucchini yellow mosaic virus (ZYMV, Potyvirus) in cucurbits. *Arch Virol*. 2010; 155(3):397–401. [PubMed: 20112122]
- Desbiez C, Lecoq H. Zucchini yellow mosaic virus. *Plant Pathology*. 1997; 46(6):809–829.
- Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*. 2008; 9(4):267–276.
- French R, Stenger DC. Evolution of wheat streak mosaic virus: Dynamics of population growth within plants may explain limited variation. *Annual Review of Phytopathology*. 2003; 41:199–214.
- Gal-On A. A point mutation in the FRNK motif of the potyvirus helper component-protease gene alters symptom expression in cucurbits and elicits protection against the severe homologous virus. *Phytopathology*. 2000; 90(5):467–473. [PubMed: 18944551]
- Gal-On A. Zucchini yellow mosaic virus: insect transmission and pathogenicity - the tails of two proteins. *Molecular plant pathology*. 2007; 8(2):139–150. [PubMed: 20507486]
- Gallie DR. Cap-independent translation conferred by the 5' leader of tobacco etch virus is eukaryotic initiation factor 4G dependent. *Journal of virology*. 2001; 75(24):12141–12152. [PubMed: 11711605]
- Gonzalez-Jara P, Fraile A, Canto T, Garcia-Arenal F. The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *Journal of virology*. 2009; 83(15):7487–94. [PubMed: 19474097]
- Gouy M, Guindon S, Gascuel O. SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*. 2010; 27(2):221–224. [PubMed: 19854763]
- Gutierrez S, Yvon M, Pirolles E, Garzo E, Fereres A, Michalakakis Y, Blanc S. Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. *PLoS pathogens*. 2012; 8(11):e1003009. [PubMed: 23133389]
- Gutierrez S, Yvon M, Thebaud G, Monsion B, Michalakakis Y, Blanc S. Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS pathogens*. 2010; 6(9):e1001113. [PubMed: 20862320]
- Hall JS, French R, Hein GL, Morris TJ, Stenger DC. Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology*. 2001; 282(2):230–236. [PubMed: 11289805]
- Johansen E, Edwards MC, Hampton RO. Seed transmission of viruses: current perspectives. *Annual review of phytopathology*. 1994; 32:363–386.
- Katis NI, Tsitsipis JA, Lykouressis DP, Papapanayotou A, Margaritopoulos JT, Kokinis GM, Perdakis DC, Manoussopoulos IN. Transmission of zucchini yellow mosaic virus by colonizing and non-colonizing aphids in Greece and new aphid species vectors of the virus. *Journal of Phytopathology*. 2006; 154(5):293–302.
- Kneller ELP, Rakotondrafara AIM, Miller WA. Cap-independent translation of plant viral RNAs. *Virus Research*. 2006; 119(1):63–75. [PubMed: 16360925]
- Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L, Wilson R. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. 2012; 22(10):1101–1111. [PubMed: 22988059]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–1760. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 2009; 25:2078–9. [PubMed: 19505943]
- Li H, Roossinck MJ. Genetic bottlenecks reduce population variation in an experimental RNA virus population. *Journal of virology*. 2004; 78(19):10582–10587. [PubMed: 15367625]
- Lin SS, Hou RF, Yeh SD. Complete genome sequence and genetic organization of a Taiwan isolate of zucchini yellow mosaic virus. *Botanical Bulletin of Academia Sinica*. 2001; 42(4):243–250.
- Lin SS, Wu HW, Jan FJ, Hou RF, Yeh SD. Modifications of the helper component-protease of zucchini yellow mosaic virus for generation of attenuated mutants for cross protection against severe infection. *Phytopathology*. 2007; 97(3):287–96. [PubMed: 18943647]

- Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*. 2011; 12(11)
- Miyashita S, Kishino H. Estimation of the size of genetic bottlenecks in cell-to-cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *Journal of virology*. 2010; 84(4):1828–1837. [PubMed: 19955302]
- Moury B, Fabre F, Senoussi R. Estimation of the number of virus particles transmitted by an insect vector. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(45):17891–17896. [PubMed: 17971440]
- Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, Altaf-Ul-Amin M, Ogasawara N, Kanaya S. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*. 2011; 39(13)
- Niepel M, Gallie DR. Identification and characterization of the functional elements within the tobacco etch virus 5' leader required for cap-independent translation. *Journal of virology*. 1999; 73(11): 9080–9088. [PubMed: 10516014]
- Pesic Z, Hiruki C. Differences in the incidence of alfalfa mosaic virus in seed coat and embryo of alfalfa seed. *Canadian Journal of Plant Pathology*. 1986; 8(1):39–42.
- Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005; 21(5):676–679. [PubMed: 15509596]
- Roberts IM, Wang D, Thomas CL, Maule AJ. Pea seed-borne mosaic virus seed transmission exploits novel symplastic pathways to infect the pea embryo and is, in part, dependent upon chance. *Protoplasma*. 2003; 222(1-2):31–43. [PubMed: 14513309]
- Ryder EJ. Seed transmission of lettuce mosaic virus in mosaic resistant lettuce. *J Am Soc Hortic Sci*. 1973; 98:610–614.
- Sacristan S, Malpica JM, Fraile A, Garcia-Arenal F. Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. *Journal of virology*. 2003; 77(18): 9906–9911. [PubMed: 12941900]
- Sanjuan R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(22):8396–8401. [PubMed: 15159545]
- Simmons HE, Dunham JP, Stack JC, Dickins BJA, Pagan I, Holmes EC, Stephenson AG. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *Journal of General Virology*. 2012; 93:1831–1840. [PubMed: 22592263]
- Simmons HE, Holmes EC, Gildow FE, Bothe-Goralczyk MA, Stephenson AG. Experimental Verification of Seed Transmission of zucchini yellow mosaic virus. *Plant Disease*. 2011a; 95(6): 751–754.
- Simmons HE, Holmes EC, Stephenson AG. Rapid evolutionary dynamics of zucchini yellow mosaic virus. *Journal of General Virology*. 2008; 89:1081–1085. [PubMed: 18343852]
- Simmons HE, Holmes EC, Stephenson AG. Rapid turnover of intra-host genetic diversity in zucchini yellow mosaic virus. *Virus Research*. 2011b; 155(2):389–396. [PubMed: 21138748]
- Simon-Buela L, Guo HS, Garcia JA. Long sequences in the 5' noncoding region of plum pox virus are not necessary for viral infectivity but contribute to viral competitiveness and pathogenesis. *Virology*. 1997; 233(1):157–162. [PubMed: 9201225]
- Urcuqui-Inchima S, Haenni AL, Bernardi F. Potyvirus proteins: a wealth of functions. *Virus Research*. 2001; 74(1-2):157–75. [PubMed: 11226583]
- Wang DW, Maule AJ. A model for seed transmission of a plant-virus - genetic and structural-analyses of pea embryo invasion by pea seed-borne mosaic-virus. *The Plant cell*. 1994; 6(6):777–787. [PubMed: 12244258]
- Wang WX, Wei Z, Lam TW, Wang JW. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. *Scientific Reports*. 2011; 1
- Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. *BMC Bioinformatics*. 2011; 12(1):119. [PubMed: 21521499]

Zheng H, Chen J, Chen J, Adams MJ, Hou M. Bean common mosaic virus isolates causing different symptoms in asparagus bean in China differ greatly in the 5'-parts of their genomes. Archives of Virology. 2002; 147(6):1257–1262. [PubMed: 12111434]

Highlights

We obtained deep sequence data from nine vertically transmitted ZYMV populations

The 5'UTR region is variable, with 26 major haplotypes observed in this region

Vertically infected plants are symptomless unlike those infected horizontally

Seed infection higher than seed transmission rate, possible population bottleneck

Figure 1. Distribution of mutations across the ZYMV genome from the first (top) and second (bottom) generations of vertically transmitted populations. The length of the ticks indicates the frequency of that particular mutation, and the schematic of the genome indicates in which gene region the mutations occurred.

Figure 2.

Alignment of the haplotypes from the horizontally transmitted sample (inoculant), one of the vertically transmitted samples (FG1), and the NCBI reference (NC_003224.1). This sequence alignment shows the large deletion (nt 98-127) found in all 26 5'UTR haplotypes from the vertically transmitted sample (last 26 sequences), and the partial deletion found in both haplotypes from the inoculant (first two sequences). The NCBI Reference is the third sequence.