Received: 11 December 2019 | Revised: 23 April 2020 | Accepted: 24 June 2020

DOI: 10.1111/jeb.13672

SPECIAL ISSUE

JOURNAL OF Evolutionary Biology eseb WILEY

Ecological fitting is the forerunner to diversification in a plant virus with broad host range

Adrián Peláez | **Michael J. McLeish** | **Ricky R. Paswan** | **Bhumika Dubay** | **Aurora Fraile** | **Fernando García-Arenal**

Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) and Instituto nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, UPM, Madrid, Spain

Correspondence
Fernando García-Arenal, Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) and Instituto nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, 28223 Pozuelo de Alarcón, UPM, Madrid, Spain.
Email: fernando.garciaarenal@upm.es

Funding information
ArimNet2 2015 joint Call, EMERAMB, Grant/Award Number: 618127; Plan Estatal de I+D+i, Ministerio de Economía y Competividad (MINECO), Grant/Award Number: BFU2015-64018-R and RTI2018-094302-B-I00; Erasmus Mundus Scholarship of the EU, BRAVE, Grant/Award Number: 2013-2536/001-001; Formación de Personal Investigador contract, Grant/Award Number: BES-2016-077810

Abstract

The evolution and diversification of ssRNA plant viruses are often examined under reductionist conditions that ignore potentially much wider biotic interactions. The host range of a plant virus is central to interactions at higher levels that are organized by both fitness and ecological criteria. Here we employ a strategy to minimize sampling biases across distinct plant communities and combine it with a high-throughput sequencing approach to examine the influence of four habitats on the evolution of *Watermelon mosaic virus* (WMV). Local, regional and global levels of genetic diversity that correspond to spatial and temporal extents are used to infer haplotype relationships using network and phylogenetic approaches. We find that the incidence and genetic diversity of WMV were structured significantly by host species and habitat type. A single haplotype that infected 11 host species of a total of 24 showed that few constraints on host species use exist in the crop communities. When the evolution of WMV was examined at broader levels of organization, we found variation in genetic diversity and contrasting host use footprints that broadly corresponded to habitat effects. The findings demonstrated that nondeterministic ecological factors structured the genetic diversity of WMV. Habitat-driven constraints underlie host use preferences.

KEYWORDS
community heterogeneity, facultative generalism, host shift, metagenomics, virus emergence

1 | INTRODUCTION

Watermelon mosaic virus (WMV) is a member of the largest genus of plant viruses, *Potyvirus* (family, *Potyviridae*), and known for causing substantial cucurbit crop yield loss across its global distribution (Lecoq & Desbiez, 2012). As with most members of the genus *Potyvirus*, WMV is nonpersistently transmitted by aphid species and has a reported host range of at least 160 dicotyledonous species of 23 families (http://www.dpvweb.net/). The translocation of plants and plant products by humans has increased connectivity among geographically distant populations in Europe and the rest of the world (Desbiez, Joannon, Wipf-Scheibel, Chandeysson, & Lecoq, 2009; Jones & Naidu, 2019). Genetic variants of WMV detected in cucurbits from France during the 2000s are hypothesized to be an invasive strain introduced from Asia (Desbiez et al., 2009). Invasions therefore contribute to the genetic variation of WMV

Peláez and McLeish contributed equally to this work.

observed in Europe, but other factors include recombination between invasive and local strains (Desbiez & Lecoq, 2008). The wide host range (Moreno et al., 2004) and variation in fitness of WMV across 48 host species has been demonstrated under controlled conditions (Lecoq et al., 2011). This suggests that in nature populations of noncucurbit hosts are also a major source of WMV genetic diversity. However, like for most plant viruses (Jones & Naidu, 2019), there have been no systematic studies investigating how the diversity of plant communities of agricultural ecosystems contribute to structuring the genetic diversity and evolution of WMV. In general, spatial and temporal variation in the proportion of resistant and susceptible hosts, and transmission, structure host–virus and virus–virus interactions and their evolution (Fabre, Rousseau, Mailleret, & Moury, 2012; Roossinck, 2015). Understanding whether the physical make up of agricultural ecosystems structures the incidence of WMV is important for anticipating the spread of invasive and local variants alike.

Genetic variation in plant viruses results from a combination of mutation, recombination and reassortment of genetic segments (Elena et al., 2011; García-Arenal, Fraile, & Malpica, 2001). The high mutation rate of single-stranded RNA (ssRNA) viruses, such as WMV, is hypothesized to be an evolutionary strategy that facilitates adaptation of host generalists (Bedhomme, Lafforgue, & Elena, 2012; Burch & Chao, 2000; Sanjuán, 2012). For example, the genetic diversity and evolution of *Tobacco mosaic virus* and *Cucumber mosaic virus* varied according to the host and host shifts (Schneider & Roossinck, 2001). Negative selection (i.e., purifying selection) that causes a decrease in genetic diversity is the rule of thumb in plant viruses, but may depend on the interaction between virus and host plant demographics (García-Arenal et al., 2001). Genetic drift associated with vector activity is influenced by ecological (i.e. distribution, abundance and interactions) factors (Fereres, 2015; Mauck, Bosque-Pérez, Eigenbrode, Moraes, & Mescher, 2012). Bottlenecks that arise as an ecological consequence in the form of transmission are therefore postulated as a key cause in the evolution of genetic diversity among virus populations. Genetic drift occurs as a result of random effects on subsequent generations, which can take place as a result of transmission. Additionally, virus fitness is expected to vary across available host plant genotypes or species (García-Arenal & Fraile, 2013), and experimental evidence supports across host fitness trade-offs (e.g., Bedhomme et al., 2012; Bera, Fraile, & García-Arenal, 2018). There is also evidence that trade-offs are not ubiquitous (Alizon, Hurford, Mideo, & Van Baalen, 2009; Elena, 2017), which suggests other factors contribute to the organization of plant–virus and virus–virus interactions at the community or ecosystem levels. For example, patterns of both global and local virus distributions reveal strong habitat specificity (Bernardo et al., 2018; Páez-Espino et al., 2016). Given negative selection, and negligible fitness costs of host-generalist viruses across multiple host species, the principal factors driving their evolution are expected to be ecological drift (i.e., demographic stochasticity among communities), dispersal and mutation rate (Bedhomme et al., 2012; Pagán, 2018). Additionally, ecological variation among host species that occupy plant communities of different habitats is expected to organize vector movement and the distribution of virus infections in an ecosystem (e.g., Borer, Seabloom, Mitchell, & Power, 2010). Temporal factors also contribute to the genetic variation among communities. For example, sharp changes in the prevalence of WMV over annual periods indicated large variations in population size are possible over relatively short periods of less than a year (Desbiez et al., 2019). Ecological variation among plant communities in agricultural ecosystems broadly translates to highly structured spatial and temporal distributions of host plants. This environmental heterogeneity, in turn, is expected to structure plant–virus genetic diversity (McLeish, Sacristán, Fraile, & Garcia-Arenal, 2017; McLeish, Sacristán, Fraile, & García-Arenal, 2019). In other words, ecological drift among communities and transmission among host species are expected to contribute to the effects of genetic drift and the evolution of genetic variation.

Watermelon mosaic virus is widespread in Spain and collected almost exclusively from melon crops (Juárez et al., 2013; Moreno et al., 2004) and infrequently from noncucurbit hosts (Malpica, Sacristán, Fraile, & García-Arenal, 2006). Much of the previous work on the genetic diversity of WMV has focused on infections of crop species, which ignores any variation in diversity that may be present noncrop host species. Species-rich agricultural ecosystems of central Spain (Médail & Diadema, 2009) are occupied by hundreds of noncucurbit species that represent a large metacommunity of potential hosts for viruses. The spatial proximity, or connectivity, of wild communities and crops in these agricultural ecosystems is hypothesized to promote the emergence of new diseases (Roossinck & García-Arenal, 2015). It is expected that increased connectivity among distinct plant communities, or habitats, increases the opportunity for the virus to encounter novel ecological conditions. Plant virus evolutionary studies of heterogeneous communities are best served by high-throughput approaches that efficiently detect a broad spectrum of genetic variants of viral species at ecosystem levels. Furthermore, strategic designs aim to reduce the sampling bias in the representation of host species and offer accurate means of quantifying genetic variance. The geographic provenance of samples and the date of their collection are also important for linking changes in multiple populations over time (Rambaut, Lam, Max Carvalho, & Pybus, 2016).

We aim to show that the genetic diversity of WMV and its incidence in an agricultural ecosystem in central Spain are structured by distinctions among plant communities. Four habitat categories were selected a priori before sampling, to represent the dominant land-cover types of the ecosystem. These cover types represent both wild and anthropic habitats. We combine a high-throughput sequencing (HTS) protocol to detect WMV with a validation approach using RT-PCR, to test the hypothesis that genetic diversity is structured by four habitat types of an agricultural ecosystem and underlies the evolution of WMV. We incorporate sequence data from other studies, to infer relationships between the local-level diversity uncovered by our study, with regional and global levels of WMV variation over the last 50 years.

2 | MATERIALS AND METHODS

2.1 | Collections and origin of WMV samples

Samples were collected in central Spain between the summer of 2015 and spring of 2016, at the sites indicated in Figure 1. We collected 3,476 individual plant specimens from 23 study sites that comprised communities of four habitats with distinct cover types (Appendix S1 and S2). Samples were taken systematically from the four habitats to minimize biases in the representation of plant species richness and abundance of these communities. Oak forest habitat (Oak) is variably subject to disturbances caused by either the production of fuel, recreational activities or light grazing. Oak is dominated by *Quercus* species with sclerophyllous shrubs and grasses. Patchy areas of successional scrubland of no particular use (Wasteland) are intermixed among cultivated areas of crop. Crops (Crop) are largely seasonal monocultures interspersed by other species, which are rotated or left fallow between seasons. The narrow borders of relatively permanent plant assemblages (Edge) that separate cultivated patches are generally left intact between seasonal cropping (Figure S1). Four sites each of Oak and Wasteland were sampled in autumn and spring. Edge and Crop with four and eleven sites respectively were sampled in spring, summer and autumn. A boustrophedonic transect method (a line taken alternately from right to left and from left to right, and so on) was used in all instances except for Edge that have highly linear configurations. In Edge and Crop, samples were collected from a 25 m × 2 m area at each site at each resampling. An herbarium was prepared during each collection for subsequent taxonomic assignment by an expert, with a total of 144 plant taxa identified.

Our approach combined HTS, for detecting a wider range of host species, with RT-PCR that permitted validation and the targeting of genetic regions of interest (McLeish, Fraile, & García-Arenal, 2020). A pilot study we conducted earlier indicated that RT-PCR can produce false negatives, which was associated with the virus titre in the total RNA extract, but allows nucleotide sequence determination of the chosen genomic region in all positive samples, which is not the case for HTS data. The HTS approach was less sensitive to plant sample heterogeneity than the RT-PCR approach. Our integrated approach met a balance between the bulk sampling efficiency of the HTS approach, and the targeted, validation method by RT-PCR. The pilot study produced an additional 36 *CP* sequences that were included in the genetic diversity and phylogenetic analyses.

2.2 | Nucleic acid extractions and sample preparation

Depending on the habit of the plant species, a number of leaves from different parts of the individual were collected, each collection of leaves from the individual representing a single sample. For total RNA extraction, we used the Agilent Plant RNA Isolation Mini Kit (Catalogue No.: 5185-5998), Spectrum™ Plant Total RNA Kit (Catalogue No.: STRN50) and the CTAB method (Chang, Puryear, & Cairney, 1993) depending on the plant species and leaf source. Optical density (OD) measurements of each extract were conducted to quantify the RNA concentration (ng/μl) and purity (at 260/280 and 260/230 wavelength ratios) using the NanoDrop® ND-1000. The extracted RNA was also run on a 1% agarose gel in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA) to check its integrity.

Up to seven individual RNA extract from samples of the same collection at the same site and for the same species were pooled to obtain a single preparation for HTS library preparation. For example, if there were only five individuals of a target species from a collection, then only five individuals were pooled. High-throughput sequencing libraries were outsourced to Centre de Regulació Genòmica (CRG; http://www.crg.eu/). All libraries were prepared from total RNA using the Truseq total RNAseq kit with an additional step of rRNA depletion using the Ribo-Zero kit. Paired-end reads of 150 nt were sequenced on either the Illumina HiSeq3000, 4000 or HiSeq2500 platforms. We estimated a depth of approximately 8.0 x 10^6 reads would be reasonable to detect high- and low-titre viruses with a proportion of 'contaminating' nucleic acids from other sources (Appendix S3). A sequencing depth of approximately 8.0 × 10^6 reads was generated for each library (pool).

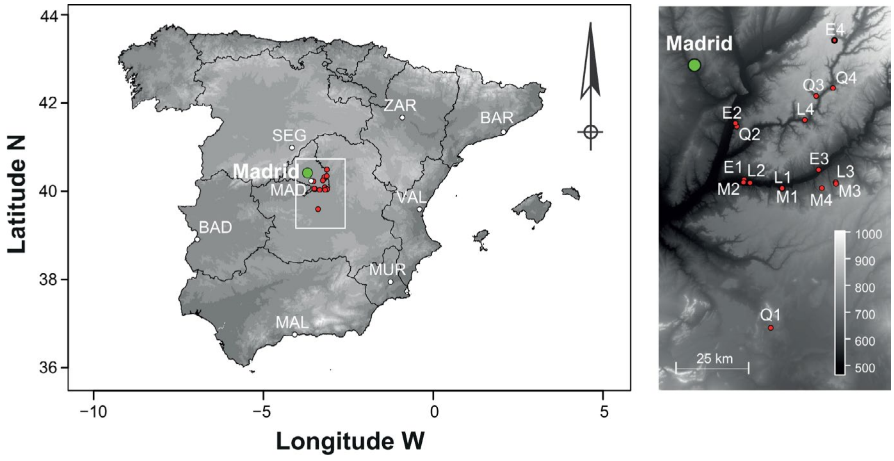

FIGURE 1 Map of Spain indicating the locations of the study sites analysed in this study. The area enclosed with the white rectangle line in central Spain has been enlarged on the right to show the sites sampled during this study (red points). White points indicate sampling sites of other studies that have been included in this study. Global collection sites not available. The scale is elevation in metres. M = Melon (Crop), E = Eriales (Wasteland), L = Linde (Edge), Q = *Quercus* (Oak)

2.3 | Detection by high-throughput sequencing

It was not practical to sequence all plant species collected in all habitats (*n* = 144), so we sequenced 276 HTS libraries that comprised the species that had 5 or more individuals in at least one of the four habitats (representative species). This totalled 104 plant taxa: 91 species, 11 identified to an unknown species within a known genus, with the remaining two as unknowns.

All HTS reads were provided with Phred quality scores greater than Q30. Adaptor trimming was conducted using *cutadapt* v1.8.3 (Martin, 2011). We used three steps to validate the presence of WMV in each of the HTS read libraries. The first step used a local Blast search against a reference database of ssRNA plant viruses sourced from NCBI's Viral Genome Browser (https://www.ncbi.nlm.nih.gov/genomes/). We used Blast query cut-off criteria of at least one match with the database, with percentage identity greater or equal to 90% (in anticipation of divergent strains); a query length greater or equal to 100 nt (as some reads were shorter than 150 nt after adaptor trimming), and a query coverage equal to 100%. The second step sought to assemble all genomes present in each of the read libraries. We considered assemblies that produced contigs greater than 300 or 500 nt as indicative of greater levels of certainty of true-positive detections compared to the Blast matches. The third validation step involved detection by RT-PCR using the individual total RNA extractions that were combined for each pool (see below). Local blast queries were conducted with Blast + version 2.2.29 (Camacho et al., 2009), and de novo metagenomic assemblies with IDBA-UD, the Iterative De Bruijn graph Assembler for sequencing data with highly Uneven Depth (Peng, Leung, Yiu, & Chin, 2012).

2.4 | Validation by RT-PCR and analyses of WMV incidence

The complete 843 nt of the *CP* gene of WMV was amplified with primers F2 5′ – GTGGCTTCTAAGCAAAGAYG – 3′ and R4 5′ – CTGTGACCAGTTTACCTAGYC – 3′ that amplify from position 8,828 (Accession Number NC_00626) in the nuclear inclusion b gene to position 9,859 in the untranslated region; and an 826 nt region of the *CP* region from position 9,032 to position 9,801 with primers WMV 826-CPF 5′ – TNGARAATTTGGATGYAGG – 3′ and WMV 826-CPR 5′ – CGACCCGAAATGCTAACTG – 3′. The PCR products were visualized by gel electrophoresis in 2.0% agarose after ethidium bromide staining. Incidence of WMV was calculated as the Number of RT-PCR-positive individuals/ total number of individual plant samples analysed. Differences of incidence among habitats were assessed using permutation *F* tests and *t* tests conducted with the R package *RVAideMemoire* (Hervé, 2020). The Benjamini & Hochberg *p*-value adjustment method for multiple comparisons and 1,000 permutations were used in all tests. The permutation tests are nonparametric and do not assume normally distributed errors. Incidence (unpaired) values are randomly assigned to either of two groups (i.e., habitats) and the statistic is computed for each iteration. The observed statistic is compared to the sampling (null) distribution to test (two-sided) the null hypothesis of no treatment effect.

2.5 | Genetic diversity and phylogenetic analyses

As our data are derived from multiple populations (here defined by multiple sampling sites), both polymorphic and fixed mutations were considered together in the genetic diversity and phylogenetic analyses. This approach provided information about the fate of mutations over a range of spatiotemporal scales. To evaluate habitat and host species affiliations and explore ambiguities in low-level population divergences, a network representation of genealogical relationships was inferred using the *HaploNet* function from the R package *pegas* (Paradis, 2010). The function implements a maximum parsimony approach, which constructs the haplotype network using the infinite sites model (Templeton, Crandall, & Sing, 1992).

We then compared groups of samples that reflected the widest possible genetic diversity and sampling period between 1972 and 2016. We sourced sequence data from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) for which the sampling date was available to conduct comparative analyses of genetic diversity at the local ('Local'), regional ('Regional') and global ('Global') levels (see Appendix S4). Our analyses included groups that comprised: (a) all samples validated from our study collected during the 2015–2016 period from *Cucumis melo* hosts and samples from wild host species ('Local'); (b) samples collected during 1995 and 1999 (Moreno et al., 2004) from *C. melo* in Spain ('Regional'); and (c) dated samples from around the world during the period 1972 to 2007 ('Global'), which were from crop species (sources referenced in Desbiez et al., 2009). For these groups, we estimated nucleotide and haplotype diversity (Tajima, 1983) differences within populations and used permutation with Tamura–Nei model correction for multiple hits (Tamura & Nei, 1993) to test differences between them. The permutation procedure resamples from the combined haplotypes over all populations, in order to test whether the observed differences in nucleotide/haplotype diversity between specific populations are greater than expected by chance (Alexander et al., 2016).

A Bayesian approach was used to infer phylogenetic relationships at the Local level (*n* = 108) and in respect to the Regional and Global levels (*n* = 66). As recombination has been detected from genomic regions of WMV (Desbiez et al., 2009; Moreno et al., 2004), we used single breakpoint detection implemented in HyPhy (Pond, Frost, & Muse, 2005) and the Datamonkey version of GARD (Available at: http://www.datamonkey.org/GARD/), a Genetic Algorithm for Recombination Detection (Weaver et al., 2018), to identify any break points in the *CP* region used in this study. We also asked whether the haplotypes sampled for this study were subject to negative selection as inferred by other studies (Moreno et al., 2004; Desbiez et al., 2009). The mean of pairwise nonsynonymous to synonymous site ratios (ω) for unique haplotypes of each group was estimated. Measuring selection using within-population divergences may invalidate the d*N*/d*S* > 1 standard of positive

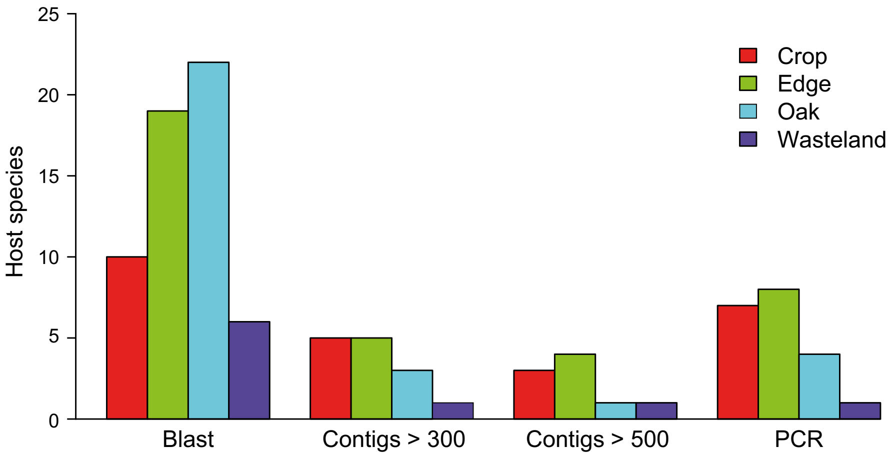

FIGURE 2 The realized host range of WMV in each habitat according to RT-PCR or high-throughput sequencing detection. Total host range for the RT-PCR detection was 14 species, and 51 species using the HTS method

selection (Kryazhimskiy & Plotkin, 2008) because the ratio may be due to segregating polymorphisms and not fixed difference between lineages. Therefore, we calculated the ratio over all unique haplotypes ($n = 74$), but also for haplotypes ($n = 42$) whose pairwise distance relationships (the number of sites that differed between each pair of sequences) were greater than the mean pairwise distance among all haplotypes. By maximizing the signal of divergence among haplotypes, we expect the ratio to reflect a higher proportion of fixed differences. The ratios were calculated using the Li (1993) method implemented with the function *dnds* in the R package *ape* (Paradis & Schliep, 2018).

In the presence of unresolved phylogenetic relationships expected from low-level population divergences, heterochronous sequence data can be informative of variation in evolutionary rates between sampling dates (Rambaut et al., 2016). The provision of collection dates by the online data sources permitted the evaluation of the linear regression between root-to-tip divergences (branch lengths) as inferred from the Bayesian maximum clade credibility consensus and the sampling date. The regression-based approach is not suitable for statistical hypotheses testing due to nonindependence of the terminal taxa in the tree. The regression generates residuals that are used to identify lineages that were on average more (positive residuals) or less (negative residuals) divergent than expected given the date of sampling. A positive regression slope would indicate that there was sufficient genetic change between sampling dates, otherwise, a negative slope indicates a weak or absent relationship between genetic divergence and time. A clade with more genetic divergence than expected by sampling date is indicative of processes such as recombination or hypermutation. We took a conservative approach and called sequences that were either greater than 0.01 or less than −0.01 of the residual value respectively. The calculation of the residuals and the regression were performed with TempEst (Rambaut et al., 2016).

Sequences of the *CP* gene were aligned using MUSCLE v3.8.31 (Edgar, 2004). We used the *modelTest* function in the R package *phangorn* (Schliep, 2011) to estimate substitution models and the log-likelihood, number of estimated parameters, AICc and BIC of all models tested (Posada, 2008). A general time-reversible DNA substitution model (GTR) with gamma distributed (+G) rates and a proportion of invariant sites (+I) was applied for the analysis. Four Markov chains were run for 10 million generations, sampling each chain every 1,000 trees. Posterior probabilities and mean branch lengths were derived from 7,500 post-burnin trees. We used MrBayes 3.2.7a (Huelsenbeck & Ronquist, 2001) to infer a Bayesian maximum credibility consensus phylogeny. Convergence and posterior parameter distributions were assessed using the MCMC Tracer Analysis Tool v1.7.1 (Rambaut, Suchard, Xie, & Drummond, 2014).

To assess the distribution of WMV among the four habitats, we inferred phylogeny–trait correlations using Bayesian Tip-association Significance (BaTS) tests (Parker, Rambaut, & Pybus, 2008). A random sample of 1,000 post-burnin trees from the posterior of the analysis described above was used to account for phylogenetic uncertainty expected at the population level. In the BaTS approach, the data are repeatedly simulated under the null hypothesis that characters at the tips are randomly distributed across the phylogeny. Theoretically, the resulting *p*-values follow a unit uniform distribution and the type 1 error of the test will be correct for all levels of statistical significance (Parker et al., 2008). We designated either a habitat or a host species trait for each taxon (tip) of the posterior set of trees (PSTs) in two separate analyses. The analysis generates the parsimony statistic (PS) and association index (AI) statistic. The true posterior distribution of each test statistic scored on the PSTs is compared against a null distribution. By placing the PS and AI statistics in a Bayesian framework, type I error and phylogenetic uncertainty in topology and branch lengths were accounted for. The inverse of the observed PS value is related to the strength of trait association on the terminal branches of the phylogeny and assumes that the gain/loss of the trait occurs parsimoniously. The AI statistic considers any imbalances in the topology of the phylogeny. A low AI values represent a strong phylogeny–trait association. The *p*-value of each statistic is the proportion of trees from the null distribution equal to, or more extreme than, the median posterior estimate from the PSTs (Parker et al., 2008).

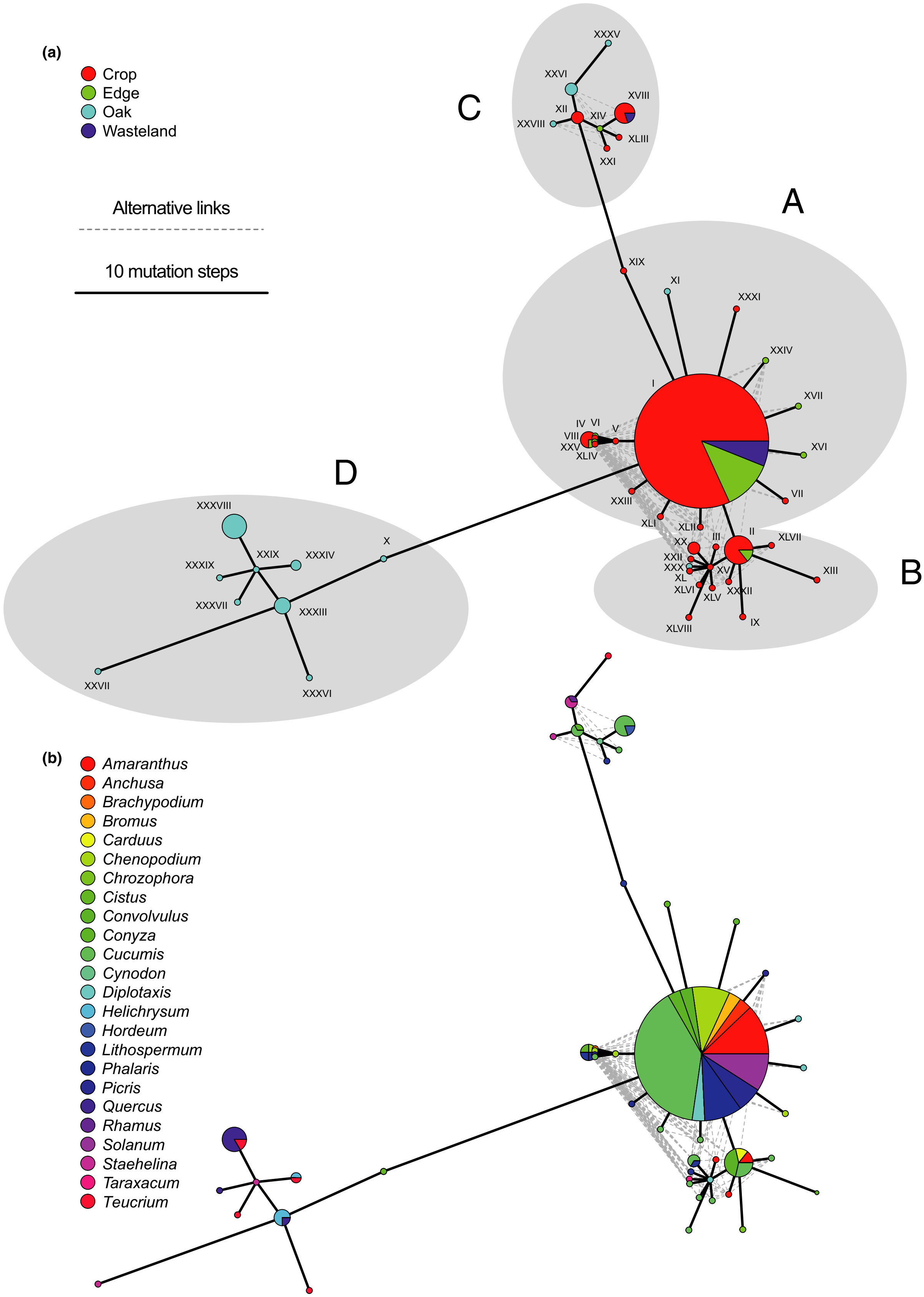

FIGURE 3 Maximum parsimony haplotype network constructed using the infinite sites model. Network (a) depicts the frequency of haplotypes associated with the four habitats, while network (b) shows host species associations. The groups (A–D) that we nominated are indicated with grey ellipses in network b only and are to assist cross-comparisons

3 | RESULTS

3.1 | Host range and habitat affiliations

As one study site was unavailable for sampling at one of the four time points, 31 of the 32 collections across the four habitats were completed. Out of a total of 276 sequencing libraries, 81 were positive for WMV under the used Blast query criteria. The local Blast queries of the ssRNA plant virus database indicated a host range of 51 species across the four habitats (Appendix S5). However, given that homologous regions among species related to WMV can produce false positives, there is uncertainty in the accuracy of this host range estimate (Figure 2). Apart from the Blast estimates of ('realized') host range in each habitat, the assembly and PCR methods of validation were consistent with wider ranges in the Crop and Edge compared to the other two habitats. The relatively wide host ranges in Edge and Oak as estimated using Blast might represent false positives (matches of reads of related viruses), low-titre infections or genetic variants not detected with our primers. Assemblies that produced contigs greater than 300 nt identified 12 host species and assemblies greater than 500 nt, eight species. The low read coverage of WMV present in each library inhibited the assembly of contigs >300 nt in most instances. The RT-PCR step detected 116 WMV positive individual samples and identified 17 host species out of the 51 initially detected using HTS. We also conducted RT-PCR assays for species that were not sent for HTS. From this group, an additional seven species were validated as WMV hosts by RT-PCR, which increased this estimate of host range to 24 species (Table 2).

There were 13 unique haplotypes in Oak, two in Wasteland, nine in Edge and 28 in Crop. The parsimony networks (Figure 3) revealed low-level divergence relationships between many single unique haplotypes (e.g., V, VI, VII, XVI, XVII, XXVIII) and the highest frequency class of haplotype I. Haplotype I was shared among Crop, Edge and Wasteland habitats (Figure 3a), with other examples between Edge and Crop (II) and Wasteland and Crop (XVIII). We nominated four groups (A–D) to assist in cross-comparisons between our analyses. There were ambiguous, alternative network relationships (indicative of homoplasies and convergence) within all groups except for D and a large number between A and B groups (Figure 3a). It is clear that the two most divergent clusters in the networks are associated with haplotypes in Oak habitat that were hosted by *Cistus salviifolius*, *Helichrysum stoechas*, *Quercus coccifera*, *Staehelina dubia*, and *Teucrium botrys* in group D; and, *Rhamnus lycioides S. dubia*, and *T. botrys* in group C. The occurrence of haplotype I on 11 host species (Figure 3b) showed the host range differences between Crop, Edge and Wasteland (Figure 2) of a single haplotype. *Cucumis melo* was host to a large number of haplotypes that clustered with each of three of the main sub-groups. Together, the networks show that the strongest structuring was between the most divergent haplotypes infecting hosts in the Oak habitat and the hosts of the other habitats. The high frequency of haplotype I on 11 host species (group A) indicated few constraints on host shifting between species. As in group A, the genetically close haplotypes of group C was indicative of transmission efficacy among habitats as well as host species.

3.2 | Genetic diversity of WMV

To contextualise the genetic diversity of samples from our study, we compared them with a number of populations sampled by others at different periods and locations in Spain and worldwide (Table 1). By splitting the haplotypes of our study into groups of either *C. melo* or wild host species, it allowed better balanced comparisons (as there was a paucity of haplotypes detected in Wasteland) and with the study that sampled *C. melo* only (Moreno et al., 2004). The wild host species as a group were infected by a higher diversity of WMV haplotypes compared to those that infected *C. melo* of this study. This difference corresponded to 15 haplotypes on 23 species compared to 7 haplotypes on *C. melo* (Table 1). There was significantly higher ($p = .029$) nucleotide diversity of WMV infecting wild species (15 haplotypes, $n = 76$ sequences) compared to those infecting *C. melo* (seven haplotypes, $n = 32$ sequences) in our sample. The nucleotide diversity of WMV in the wild species of our sample was also significantly higher than in melon collections over Spain (>17 haplotypes, $n = 44$ sequences) in 1995 ($p < .001$) or 1999 ($p = .005$) period (Moreno et al., 2004). The significantly higher estimates of nucleotide diversity of the Global group (22 haplotypes, $n = 22$ sequences) above all other groups ($p > .012$) indicated an increase in the average number of pairwise nucleotide differences per site between all possible pairs, that is a low frequency of haplotypes with intermediate divergences. By contrast, there were significantly higher estimates of haplotype diversity of WMV at the regional level across Spain in '95 and '99 ($p < .005$) and the rest of the world ($p < .012$) compared to the Wild and *C. melo* groups at the Local level. In general, the genetic diversity of a local population is expected to be lower than widely separated populations at the regional and global scales that are subject to relatively strong genetic drift and variable selective pressures. A low number of sampled individuals is expected to artificially inflate or deflate either haplotype or nucleotide diversity (Goodall-Copestake, Tarling, & Murphy, 2012). The disparity between the haplotype and nucleotide diversity estimates of the inter-study comparisons are consistent with differences in sampling effort and collections of different host species made at different spatial scales. Wild host species together supported substantial genetic variation that was comparable with that at much wider spatial scales.

3.3 | Incidence and habitat specificity of WMV

Permutation of habitat category for pairs of observations was used to generate a null distribution using 1,000 iterations and test the observed estimates of incidence (Tables 2 and 3). The *F* tests indicated differences in the distribution of infected host species between habitats. Oak had significantly different (two-sided test) variance in incidence among species compared to Edge (*F* test, $p = .006$) and

TABLE 1 Permutation tests of haplotype diversity and nucleotide diversity of WMV samples from our study and previous studies in Spain and worldwide between 1972 and 2016

	Wild 15 haplotypes	*C. melo* 7 haplotypes	*C. melo* ‘95 17 haplotypes	*C. melo* ‘99 21 haplotypes	Global 22 haplotypes
Nucleotide					
Wild	–	**.029**	**<.001**	**.005**	**.012**
C. melo	.010	–	**.028**	.418	**<.001**
C. melo ‘95	.024	.014	–	.169	**<.001**
C. melo ‘99	.015	.005	.009	–	**<.001**
Global	.013	.023	.037	.028	–
Diversity	.041	.031	.017	.026	.054
Haplotype					
Wild	–	.175	**.008**	**.002**	**.001**
C. melo	.053	–	**.003**	**.001**	**.000**
C. melo ‘95	.156	.209	–	.648	.507
C. melo ‘99	.181	.234	.025	–	.822
Global	.193	.246	.037	.012	–
Diversity	.807	.754	.963	.988	1.000

Note: Diversity difference between populations below diagonal, *p*-values corrected with Tamura–Nei model for multiple hits and shown above diagonal and significant in bold.

Wasteland (F test, $p = .006$), but not Crop. Likewise, Edge had significantly different variance in incidence among species compared to both Crop (F test, $p = .006$) and Oak (F test, $p = .036$). Incidence was significantly higher in Crop compared to Edge (t test, $p = .006$) and Wasteland (t test, $p = .006$). Crop supported the highest incidence followed by Oak, with Edge and Wasteland having the lowest incidence. The differences in incidence between Crop and Oak were not significantly different due to over-dispersion of incidence values in these two habitats. For species such as *Amaranthus* sp., *Chenopodium album*, *Convolvulus arvensis* and *Solanum nigrum* that occurred in Crop and at least one other habitat, incidence was higher in Crop. Together, each habitat supported different sets of host species, with 5 of 24 occurring in three and 8 of 24 in two habitats, which was consistent with the significant differences of the F tests. Incidence was correlated with host relative abundance in Crop ($R^2 = .91$, $F_{(1,8)} = 97.02$, $p < .001$), but not in the other habitats. This correlation was also significant ($R^2 = .66$, $F_{(1,7)} = 16.3$, $p = .005$) after the removal of *C. melo*. The permutation tests and the correlations between incidence and host abundance indicated habitat-driven differences in incidence, and evidence of host preferences for particular species in Edge, Oak and Wasteland that were not driven by abundance.

We used a posterior sample of trees to account for phylogenetic uncertainty in the inference of habitat or host trait–tip correlations (Table 4). The result of the BaTS analyses indicated significant host and habitat effects, but the habitat phylogeny–trait association was stronger. The AI and PS statistics for habitat and host traits were both significant ($p < .0001$). The inverse of the PS statistic is proportional to the strength of the tip-character association and indicated that habitat associations (1/19.1 = 0.052) were stronger than host associations (1/61.7 = 0.016). The AI statistic measures the imbalance of internal nodes of a phylogeny, where low values indicated that the phylogeny–habitat associations (2.37) were stronger than the phylogeny–host associations (6.72). The association between related haplotypes and habitat was therefore stronger than for host species.

3.4 | Phylogenetic analysis of WMV

To contextualise genetic diversity variation in the ecosystem, it was useful to compare with WMV haplotypes from other studies (Moreno et al., 2004; Desbiez et al., 2008). Breakpoints indicative of recombination were not detected in the sequences of the *CP* gene of our data set. Evidence of negative selection and conservation of the CP was consistent with other studies. The mean pairwise d*N*/d*S* in groups of haplotypes from our study was $\omega = 0.136$ for unique haplotypes from wild host species, $\omega = 0.085$ for *C. melo* hosts, $\omega = 0.131$ over all hosts and $\omega = 0.134$ for a set of haplotypes with higher than mean pairwise divergences.

The inclusion of haplotypes sampled from outside our study area resolved the reticulated relationships between groups A and B evident in the parsimony networks (Figure 3). However, the Bayesian maximum clade credibility consensus (Figure 4) shows poor backbone support for some deep stem clades (posterior probabilities <0.80). These polytomies indicated unresolved deep divergences among well-supported derived clades including haplotypes from Europe, the Middle East, Central and Southeast Asia, Australasia, and the Americas. The evolutionary relationships of ancestors of groups A to C (shown in Figure 3) and haplotypes from Pakistan (AB218280) and

TABLE 2 Observed incidence (%) of WMV and sample size of host (*n*) for each habitat collected during this study

Host taxon	Crop	*n*	Edge	*n*	Oak	*n*	Waste.	*n*	Total inc.	Total *n*
Amaranthus sp.	53.9	13	0	19	-	0	-	0	21.875	32
Anchusa undulata	-	0	0	2	-	0	3.5	29	3.2	31
Brachypodium retusum	-	0	5.0	20	-	0	-	0	5.0	20
Bromus sp.	-	0	0	8	0	12	3.1	32	1.9	52
Carduus bourgeanus	-	0	5.9	17	0	8	0	53	1.3	78
Chenopodium album	44.4	18	6.7	15	-	0	-	0	27.3	33
Chrozophora tinctoria	33.3	3	-	0	-	0	-	0	33.3	3
Cistus salviifolius	-	0	-	0	7.1	28	-	0	7.1	28
Convolvulus arvensis	38.9	18	0	60	-	0	0	21	7.1	99
Conyza canadensis	-	0	20.0	5	-	0	-	0	20.0	5
Cucumis melo	63.2	57	-	0	-	0	-	0	63.2	57
Cynodon dactylon	-	0	6.7	15	-	0	-	0	6.7	15
Diplotaxis erucoides	33.3	3	7.5	40	-	0	0	27	5.7	70
Helichrysum stoechas	-	0	-	0	80.0	5	-	0	80.0	5
Hordeum matritense	-	0	-	0	0	2	4.2	24	3.8	26
Lithospermum arvense	22.7	22	0	7	-	0	-	0	17.2	29
Phalaris minor	66.7	6	0	6	-	0	-	0	33.3	12
Picris echioides	100.0	1	14.3	28	-	0	0	11	12.5	40
Quercus coccifera	-	0	-	0	12.7	63	-	0	12.7	63
Rhamnus lycioides	-	0	-	0	2.5	40	-	0	2.5	40
Solanum nigrum	27.3	11	0	4	-	0	-	0	20.0	15
Staehelina dubia	-	0	-	0	28.6	21	-	0	28.6	21
Taraxacum officinale	-	0	-	0	50.0	2	0	48	2.0	50
Teucrium botrys	-	0	-	0	85.7	7	-	0	85.7	7
Total	48.0	152	4.9	246	14.9	188	1.2	245	14.0	831

Note: The hyphen ('-'), indicates no record.

TABLE 3 Permutation *F* tests and *t* tests of differences in incidence between the four habitats

	Crop	Edge	Oak	Wasteland
Crop	–	**.006**	.24	**.006**
Edge	**.006**	–	**.036**	.055
Oak	.116	.054	–	**.039**
Wasteland	**.006**	.116	.054	–

Note: p-values of *F* tests are above the diagonal and *t* tests below. Significant differences indicated with bold text. Benjamini & Hochberg adjusted *p*-values are shown.

France (EU660585), respectively, were unresolved. Similarly, divergences between haplotypes of group A from those sampled in 1995 and '99 in Spain had relatively weak (~0.80) posterior support. These weakly supported divergences included recombinants from France (EU660578), Italy (EU660590) and around the world. An infection of *Picris* (L2V_27, group B) clustered with samples collected from *C. melo* in 1995 (VAL95) and indicated the persistence of a closely related haplotype detected in a wild host of Edge, for over 20 years.

The high frequency of unresolved divergences in the consensus tree was consistent with the reticulated relationships shown in the haplotype network (Figure 3). Unresolved relationships in gene trees of this nature are a consequence of saturation of substitutions (too much homoplasy) and indicative of localization of recombination events to specific clades (Carbone, Liu, Hillman, & Milgroom, 2004). Recombination produces networks rather than bifurcating relationships and form cyclical 'splits'. We used split decomposition (Dopazo, Dress, & Von Haeseler, 1993) to further investigate the ambiguities of the unresolved lineages in our Bayesian inference. The differences between the bifurcations inferred from the Bayesian analysis and splits networks (Appendix S6) were indicative of recombination events (Figure 4) that delimited our four groups A–D.

Clades that diversified faster or slower than expected on average are indicative of variation in evolutionary rates between sampling dates. Conservative cut-off thresholds were set above 0.01 and below −0.01 of the residual error values of the root-to-tip distance and sampling date regression (Figure 5). The analysis produced a negative regression slope that indicated a weak relationship between genetic divergence and sampling times. The evolutionary rate

TABLE 4 Bayesian Tip-association Significance (BaTS) tests for the strength of either habitat or host species traits and 1,000 posterior phylogenies

	Statistic	Obs. Mean	L. 95% CI	U. 95% CU	Null mean	L. 95% CI	U. 95% CI	*p*-value
Host	AI	6.724	5.465	7.899	8.385	8.076	8.662	**<.0001**
	PS	61.691	59.000	64.000	73.456	71.926	74.576	**<.0001**
Habitat	AI	2.374	1.475	3.255	4.905	4.482	5.274	**<.0001**
	PS	19.048	18.000	20.000	36.999	35.439	38.081	**<.0001**

Note: The observed parsimony statistic (PS) value is inversely related to the strength of trait association on the terminal branches of the phylogeny. A low association index (AI) value represents a strong phylogeny–trait association. Higher values of the monophyletic clade (MC) size is positively correlated with the strength of the phylogeny–trait associations.

characteristics of WMV in lineages of groups B and D were not characteristic of other lineages (Figure 4). Group B diversified faster than expected on average and comprised WMV that infected nine species in Crop, Edge and Oak. More than expected divergence of group B was consistent with rapid evolutionary change and its association with a clade that included recombinant samples collected in France and Italy in 2000. Group D also comprised host species in Oak infected by lineages with more than expected divergence on average. Group A was part of a deep polytomy that also included clades with recombinants. Faster than expected divergences supported other evidence of localization of recombination among clades that contributed to the evolution of WMV between 1972 and 2016.

4 | DISCUSSION

We combined HTS and traditional genetics approaches to look at the evolution of WMV in an agricultural ecosystem. To minimize biases in the detection of genetic diversity, we strategically chose and sampled four habitats in an agricultural ecosystem. Our conclusions were based on the detection of WMV using HTS and on the analyses of observations that had been validated using RT-PCR. The validated data set is therefore expected to yield conservative estimates of host range and genetic diversity (Figure 2). The main findings were the high incidence and concentration of haplotypes in Crop compared to the less anthropic habitats. The most divergent set of haplotypes occurred in the most natural habitat, Oak and produced a strong habitat effect. However, habitat and host effects were not mutually exclusive.

4.1 | Habitat affects virus host use and diversification

The distribution of incidence (Table 2) and the trait association analyses (Table 4) supported stronger habitat than host effects on transmission and diversification. Ecological factors that constrain transmission among habitats were evident in the differentiation between haplotypes in Oak and the anthropic habitats. Importantly, incidence outside Crop was contingent on species preferences associated with each habitat, and in Edge, Oak and Wasteland, not conditional on host relative abundance. The haplotype network (Figure 3) shows two divergent groups C and D, and large differences between Oak and the other habitats. Although the network delineates haplotypes of Oak (D) from a group of closely related haplotypes mainly of Crop and also Wasteland, there are divergent haplotypes of Oak within this group (C) as well. Constraints on virus diversification that correspond to habitat, in particular Oak, are evident in the phylogeny. The clusters of clades (Figure 4) that each corresponded to the different studies showed that haplotypes were structured according to spatial and temporal factors. The phylogenetic over-dispersion and mixing of these clades also indicated close relationships among them. Even though we did not detect recombination in sequences of the *CP* gene from our sample, the over-dispersion of samples that were identified by other studies as recombinant strains as well as the presence of polytomies, suggests that recombination in WMV has been pervasive. Overall, ecological distinctions shaped the diversification of WMV by increasing variation of genetic diversity between the habitats of the ecosystem. Between-habitat processes did not appear to be limited by genetic constraints on host species use, but by other properties of the habitats. Habitat effects result in a subset of possible biotic interactions at any given time, which are often defined as the 'realized' component of the ecological niche (Wells & Clark, 2019). Habitat constraints on host species (Figure 2) use might restrict the transmission of haplotypes at local scales and facilitate the persistence of a diversity of strains for long periods in the environment. Compositional heterogeneity among habitats of agricultural ecosystems affects the availability of hosts and influences the means by which viruses move among habitats.

Theoretically, fitness maxima of host-generalist viruses are hypothesized to be lower than in specialists, but equivalent across a wider number of hosts (Elena, Fraile, & García-Arenal, 2014). Equivalence in fitness across host species has been supported by the observation that plant viruses exhibit specialist-like strategies that resemble facultative generalism among wild host populations (Malpica et al., 2006; McLeish et al., 2017). Under this form of resource use, the wide host ranges of WMV may be very narrow on any particular host without incurring inhibitive fitness costs in a particular habitat. The composition of a Crop community is different to an Edge community and will offer different degrees of

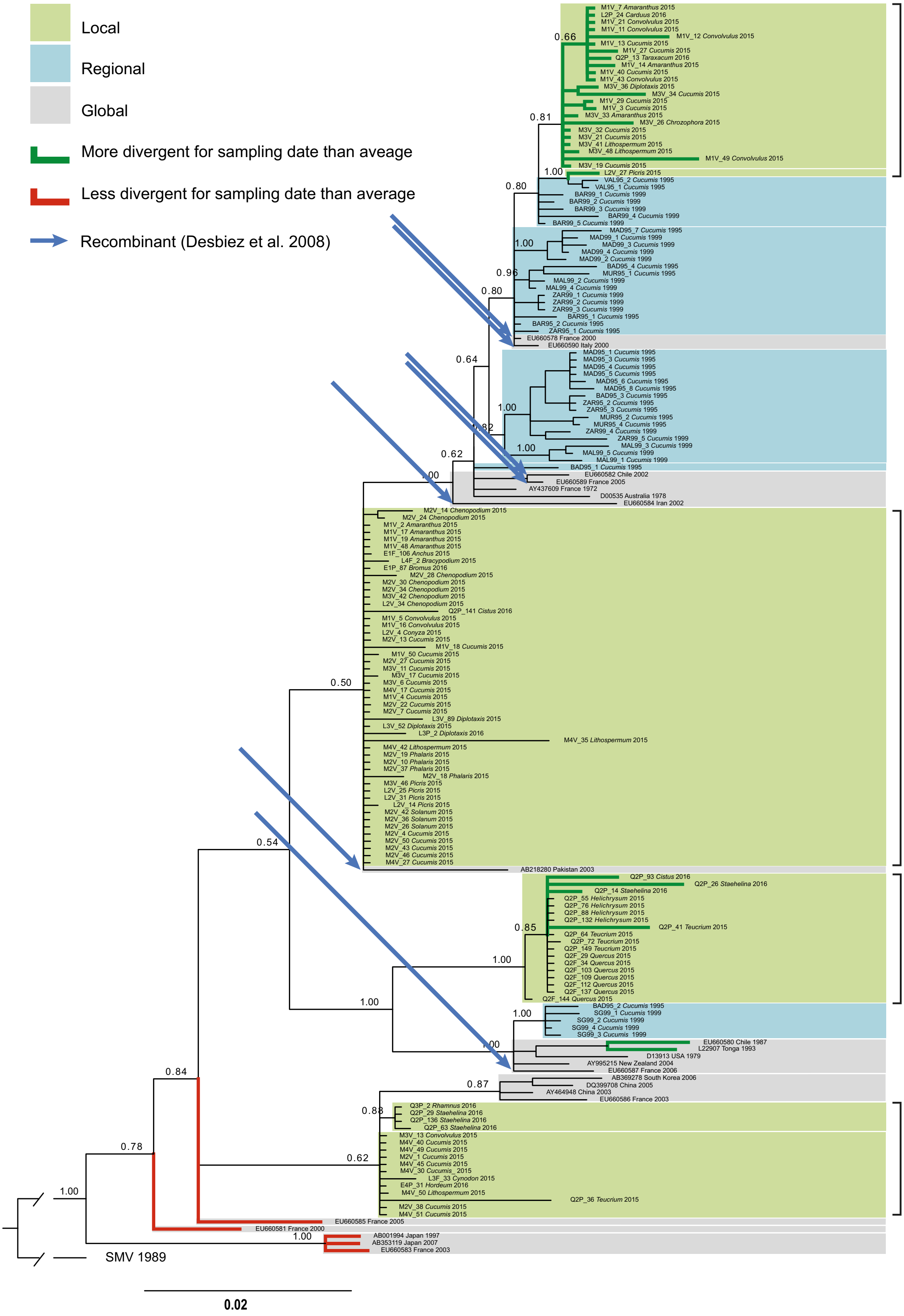

FIGURE 4 Bayesian maximum clade credibility consensus phylogeny of the coat protein gene of WMV. The groups (A–D) that we nominated are indicated along the right side to assist cross-comparisons. Outgroup: *Soybean Mosaic virus* (SMV: Genbank Accession D00507)

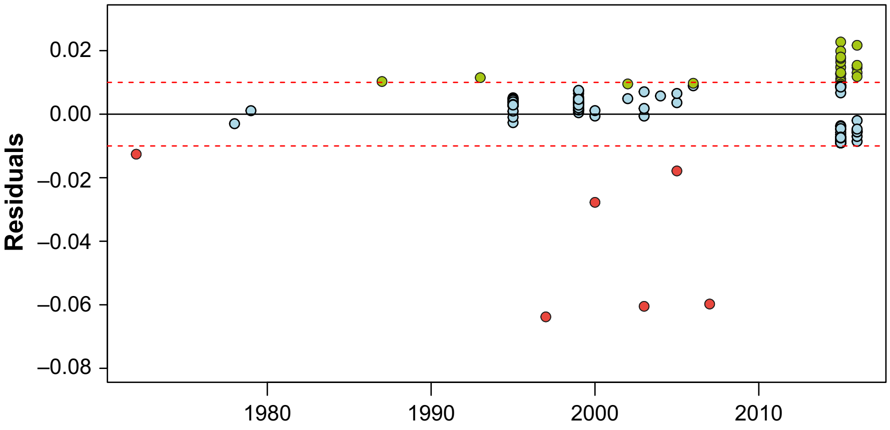

FIGURE 5 Residual plots of root-to-tip genetic distance correlations with the sampling date at three levels based on Bayesian maximum clade credibility consensus phylogenies of the coat protein gene. Green points in the positive residual area above the dashed red lines indicate samples that are relatively more divergent, and red points in the negative residual area below the line indicate samples that were relatively less divergent

ecological apparency (roughly, the availability of plants to vector species; de Lucena, de Medeiros, de Lima Araújo, Alves, & de Albuquerque, 2012), and each habitat is expected to have unique functional characteristics (McLeish et al., 2019). For instance, a species may act as a maintenance host (e.g., *S. dubia*), or as a principal host, like *C. melo*. This distinction is dependent on the habitat and on preferences for hosts within it, and may be irrespective of their relative abundance. The evolution of facultative generalist viruses depends on associations within subsets of available hosts that offer a range of fitness options. Constraints on the spatial and temporal variation of maximal fitness options among hosts will be observed as habitat and/or host specificity.

4.2 | Opportunity to evolve in agricultural ecosystems

The reticulate relationships among haplotypes shown in the networks (Figure 3, Appendix S6), and the poorly resolved divergences present in the backbone of the phylogeny (Figure 4), are indicative of either recombination or rapid diversification (Desbiez & Lecoq, 2008; Holmes, Worobey, & Rambaut, 1999; Ohshima et al., 2002; Posada & Crandall, 2002). Furthermore, the reticulated relationships shown in the networks were only partially resolved in the phylogenetic inference that included haplotypes sampled at wider spatial and temporal scales (Figure 4). Evidence of homoplasy from the networks and of phylogenetic over-dispersion of recombinant strains suggests localization of recombination events to specific clades: that is viral evolution of a recombinant strain driven by local-level dynamics (Carbone et al., 2004). Frequent recombination in ancestral lineages might have had unique consequences for genetic variation in each descendent clade. Although we did not detect breakpoints in our *CP* sequences, there is evidence from other studies of recombination outside the *CP* region of the WMV genome (Desbiez et al., 2009; Moreno et al., 2004). This suggests the phylogenetic signal among haplotypes was mixed due to recombination outside the CP. To test this possibility, we assembled the metagenomes of WMV from the HTS data and detected two putative breakpoints in the helper component protease (*HC-Pro*) gene and one in the viral genome-linked protein (*VPg-Pro*) region of the nuclear inclusion a (*NIa*) gene. Additionally, the analysis of root-to-tip divergences (Figure 5) indicated a lack of temporal signal in the amount of divergence given the date of sampling. Such heterochronous sequence data are indicative of either recombination or variation in evolutionary rates (Rambaut et al., 2016). Group B experienced greater divergence than expected on average across seven host species mainly in Crop (Figure 3). Similarly, the greater than expected divergences of group D involved haplotypes on three species in Oak. The differences in divergence characteristics between groups B/D, with higher than expected divergences, and A/C, with no uncharacteristic divergences, suggest ecological factors underlie the evolution of WMV.

Evolutionary or ecological changes that led to an increase in lineage diversification, population density, habitat-use or epidemiology (e.g., transmission) have been referred to as *ecological opportunity* (Yoder et al., 2010). The central tenet of the concept expresses that the development of biodiversity stems from availability of empty and varied ecological niches and underutilized resources (Wellborn & Langerhans, 2015). It has been shown that ecological opportunity in the form of host heterogeneity or the colonization of new habitats, coincided with increased diversification of potyviruses (Gibbs et al., 2017; González, Butković, & Elena, 2019). The successional communities that assemble with each annual crop and the enormous plant diversity of the Mediterranean Basin represent high-dimensional resources for utilization by WMV. In this respect, a wide host range is an important ecological trait and central to the diversification of plant virus lineages because it determines the breadth of 'opportunity' to unutilised resources. The variation in realized host range between the habitats (Figure 2) provides evidence of both constraints and opportunity in resource use.

Our evidence suggests that habitat-structured local populations of WMV will enhance genetic drift, and together with ecological opportunity, will facilitate the maintenance of genetic diversity across many host species in the ecosystem at the local scale. In a wider context, haplotypic diversity was higher at both regional and global levels, but not nucleotide diversity. Although this was partially explained by variation in sampling effort at different spatial scales used in the different studies, the effects of local-level processes on natural populations may produce such patterns. High recombination rates and other diversifying agents (Carbone et al., 2004) as well as large population sizes (Cvijović, Good, & Desai, 2018) may also shape genetic diversity. The role of negative selection plus recombination in RNA virus diversification has been reported in a number of systems (Chare & Holmes, 2006; López et al., 2011; Martín et al., 2009). Crops are only one of a variety of vegetation types that make up anthropic agricultural ecosystems, which also include natural plant communities. The shifting mosaics of successional communities promote opportunistic and stochastic (nondeterminant) encounters between divergent strains and novel hosts. Divergent strains maintained in reservoir communities by a high diversity of wild plant species have regular opportunities to aggregate in abundant annual crop species.

In summary, our results showed that ecological factors underlie the diversification of WMV apart from evolutionary change that might correspond to specific host traits. A strain that has co-evolved with its community (Hall, Ashby, Bascompte, & King, 2020) is indirectly inferred from its community network structure (McLeish et al., 2019). A haplotype that is evolutionarily constrained within such a network may expand out of its local habitat because a phenotype suited to life in another community just happens to evolve in the original habitat. Interactions among co-occurring species without adaptive evolution, or *ecological fitting*, may occur for a genotype that is 'serendipitously robust' to conditions of the new habitat and evolve in the absence of selective pressures (Agosta & Klemens, 2008; Janzen, 1985). At odds with ecological opportunity, which does not exclude adaptive diversification, ecological fitting presumes that there are many fine-scale and contradictory selective pressures acting at the same time over the range of the species. The results we present here show that, in addition to or instead of adaptive evolution, ecological fitting needs to be considered for understanding plant virus diversification and emergence.

ACKNOWLEDGMENTS

This work was funded in part by Plan Estatal de I+D+i, Ministerio de Economía y Competitividad, (MINECO), Spain (grants BFU2015-64018-R and RTI2018-094302-B-I00) and by ArimNet2 2015 joint Call (project 618127, EMERAMB). AP was supported by a Formación de Personal Investigador contract (BES-2016-077810), and RRP and BD were supported by an Erasmus Mundus Scholarship of the EU (2013-2536/001-001, BRAVE).

CONFLICT OF INTEREST

The authors declare that they have no conflicting interests.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jeb.13672.

ORCID

Michael J. McLeish https://orcid.org/0000-0002-4525-5221
Bhumika Dubay https://orcid.org/0000-0002-7886-0823

REFERENCES

Agosta, S. J., & Klemens, J. A. (2008). Ecological fitting by phenotypically flexible genotypes: Implications for species associations, community assembly and evolution. *Ecology Letters*, *11*, 1123–1134. https://doi.org/10.1111/j.1461-0248.2008.01237.x

Alexander, A., Steel, D., Hoekzema, K., Mesnick, S. L., Engelhaupt, D., Kerr, I., … Baker, C. S. (2016). What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Molecular Ecology*, *25*, 2754–2772. https://doi.org/10.1111/mec.13638

Alizon, S., Hurford, A., Mideo, N., & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. *Journal of Evolutionary Biology*, *22*, 245–259. https://doi.org/10.1111/j.1420-9101.2008.01658.x

Bedhomme, S., Lafforgue, G., & Elena, S. F. (2012). Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Molecular Biology and Evolution*, *29*, 1481–1492. https://doi.org/10.1093/molbev/msr314

Bera, S., Fraile, A., & García-Arenal, F. (2018). Analysis of fitness trade-offs in the host range expansion of an RNA virus, tobacco mild green mosaic virus. *Journal of Virology*, *92*, e1268–e1318. https://doi.org/10.1128/JVI.01268-18

Bernardo, P., Charles-Dominique, T., Barakat, M., Ortet, P., Fernandez, E., Filloux, D., … Roumagnac, P. (2018). Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. *The ISME Journal*, *12*, 173–184. https://doi.org/10.1038/ismej.2017.155

Borer, E. T., Seabloom, E. W., Mitchell, C. E., & Power, A. G. (2010). Local context drives infection of grasses by vector-borne generalist viruses. *Ecology Letters*, *13*, 810–818. https://doi.org/10.1111/j.1461-0248.2010.01475.x

Burch, C. L., & Chao, L. (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature*, *406*, 625–628. https://doi.org/10.1038/35020564

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Carbone, I., Liu, Y. C., Hillman, B. I., & Milgroom, M. G. (2004). Recombination and migration of *Cryphonectria hypovirus 1* as inferred from gene genealogies and the coalescent. *Genetics*, *166*, 1611–1629. https://doi.org/10.1534/genetics.166.4.1611

Chang, S., Puryear, J., & Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*, *11*, 113–116. https://doi.org/10.1007/BF02670468

Chare, E. R., & Holmes, E. C. (2006). A phylogenetic survey of recombination frequency in plant RNA viruses. *Archives of Virology*, *151*, 933–946.0675-x. https://doi.org/10.1007/s00705-005-0675-x

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*, 1235–1278. https://doi.org/10.1534/genetics.118.301058

de Lucena, R. F. P., de Medeiros, P. M., de Lima Araújo, E., Alves, A. G. C., & de Albuquerque, U. P. (2012). The ecological apparency hypothesis and the importance of useful plants in rural communities from Northeastern Brazil: An assessment based on use value. *Journal of Environmental Management*, *96*, 106–115. https://doi.org/10.1016/j.jenvman.2011.09.001

Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., & Lecoq, H. (2009). Emergence of new strains of Watermelon mosaic virus in South-eastern France: Evidence for limited spread but rapid local population shift. *Virus Research*, *141*, 201–208. https://doi.org/10.1016/j.virusres.2008.08.018

Desbiez, C., & Lecoq, H. (2008). Evidence for multiple intraspecific recombinants in natural populations of Watermelon mosaic virus (WMV, Potyvirus). *Archives of Virology*, *153*, 1749–1754. https://doi.org/10.1007/s00705-008-0170-2

Desbiez, C., Verdin, E., Moury, B., Lecoq, H., Millot, P., Wipf-Scheibel, C., … Huseynova, I. (2019). Prevalence and molecular diversity of the main viruses infecting cucurbit and solanaceous crops in Azerbaijan. *European Journal of Plant Pathology*, *153*, 359–369. https://doi.org/10.1007/s10658-018-1562-0

Dopazo, J., Dress, A., & Von Haeseler, A. (1993). Split decomposition: A technique to analyze viral evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(21), 10320–10324. https://doi.org/10.1073/pnas.90.21.10320.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Elena, S. F. (2017). Local adaptation of plant viruses: Lessons from experimental evolution. *Molecular Ecology*, *26*, 1711–1719. https://doi.org/10.1111/mec.13836

Elena, S. F., Bedhomme, S., Carrasco, P., Cuevas, J. M., de la Iglesia, F., Lafforgue, G., … Zwart, M. P. (2011). The evolutionary genetics of emerging plant RNA viruses. *Molecular Plant-Microbe Interactions*, *24*, 287–293. https://doi.org/10.1094/MPMI-09-10-0214

Elena, S. F., Fraile, A., & García-Arenal, F. (2014). Evolution and emergence of plant viruses. *Advances in Virus Research*, *88*, 161–191. https://doi.org/10.1016/B978-0-12-800098-4.00003-9

Fabre, F., Rousseau, E., Mailleret, L., & Moury, B. (2012). Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist*, *193*, 1064–1075. https://doi.org/10.1111/j.1469-8137.2011.04019.x

Fereres, A. (2015). Insect vectors as drivers of plant virus emergence. *Current Opinion in Virology*, *10*, 42–46. https://doi.org/10.1016/j.coviro.2014.12.008

García-Arenal, F., & Fraile, A. (2013). Trade-offs in host range evolution of plant viruses. *Plant Pathology*, *62*, 2–9. https://doi.org/10.1111/ppa.12104

García-Arenal, F., Fraile, A., & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology*, *39*, 157–186. https://doi.org/10.1146/annurev.phyto.39.1.157

Gibbs, A. J., Ohshima, K., Yasaka, R., Mohammadi, M., Gibbs, M. J., & Jones, R. A. (2017). The phylogenetics of the global population of potato virus Y and its necrogenic recombinants. Virus. *Evolution*, *3*, vex002. https://doi.org/10.1093/ve/vex002

González, R., Butković, A., & Elena, S. F. (2019). Role of host genetic diversity for susceptibility-to-infection in the evolution of virulence of a plant virus. Virus. *Evolution*, *5*, vez024. https://doi.org/10.1093/ve/vez024

Goodall-Copestake, W. P., Tarling, G. A., & Murphy, E. J. (2012). On the comparison of population-level estimates of haplotype and nucleotide diversity: A case study using the gene cox1 in animals. *Heredity*, *109*(1), 50–56. https://doi.org/10.1038/hdy.2012.12

Hall, A. R., Ashby, B., Bascompte, J., & King, K. C. (2020). Measuring coevolutionary dynamics in species-rich communities. *Trends in Ecology and Evolution*, *35*(6), 539–550. https://doi.org/10.1016/j.tree.2020.02.002

Hervé, M. (2020). *RVAideMemoire: Testing and plotting procedures for biostatistics*. R package version 0.9-75. Retrieved from https://CRAN.R-project.org/package=RVAideMemoire

Holmes, E. C., Worobey, M., & Rambaut, A. (1999). Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution*, *16*, 405–409. https://doi.org/10.1093/oxfordjournals.molbev.a026121

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*, 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Janzen, D. (1985). Dan Janzen's thoughts from the tropics 1: On ecological fitting. *Oikos*, *45*, 308–310.

Jones, R. A. C., & Naidu, R. A. (2019). Global dimensions of plant virus diseases: Current status and future perspectives. *The Annual Review of Virology*, *6*, 387–409. https://doi.org/10.1146/annurev-virology-092818-015606

Juarez, M., Legua, P., Mengual, C. M., Kassem, M. A., Sempere, R. N., Gómez, P., … Aranda, M. A. (2013). Relative incidence, spatial distribution and genetic diversity of cucurbit viruses in eastern Spain. *Annals of Applied Biology*, *162*, 362–370. https://doi.org/10.1111/aab.12029

Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics*, *4*(12), e1000304. https://doi.org/10.1371/journal.pgen.1000304

Lecoq, H., & Desbiez, C. (2012). Viruses of cucurbit crops in the Mediterranean region: an ever-changing picture. In G. Loebenstein, & N. Katis (Eds.), *Advances in virus research* (Vol. *84*, pp. 67–126). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-394314-9.00003-8

Lecoq, H., Fabre, F., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., Schoeny, A., & Desbiez, C. (2011). Search for factors involved in the rapid shift in *Watermelon mosaic virus* (WMV) populations in South-eastern France. *Virus Research*, *159*, 115–123. https://doi.org/10.1016/j.virusres.2011.05.004

Li, W. H. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution*, *36*, 96–99. https://doi.org/10.1007/BF02407308

López, C., Aramburu, J., Galipienso, L., Soler, S., Nuez, F., & Rubio, L. (2011). Evolutionary analysis of tomato Sw-5 resistance-breaking isolates of Tomato spotted wilt virus. *Journal of General Virology*, *92*, 210–215. https://doi.org/10.1099/vir.0.026708-0

Malpica, J. M., Sacristán, S., Fraile, A., & García-Arenal, F. (2006). Association and host selectivity in multi-host pathogens. *PLoS One*, *1*, e41. https://doi.org/10.1371/journal.pone.0000041

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10–12 https://doi.org/10.14806/ej.17.1.200

Martín, S., Sambade, A., Rubio, L., Vives, M. C., Moya, P., Guerri, J., … Moreno, P. (2009). Contribution of recombination and selection to molecular evolution of *Citrus tristeza virus*. *Journal of General Virology*, *90*, 1527–1538. https://doi.org/10.1099/vir.0.008193-0

Mauck, K., Bosque-Pérez, N. A., Eigenbrode, S. D., Moraes, C. M., & Mescher, M. C. (2012). Transmission mechanisms shape pathogen effects on host–vector interactions: Evidence from plant viruses. *Functional Ecology*, *26*, 1162–1175. https://doi.org/10.1111/j.1365-2435.2012.02026.x

McLeish, M. J., Fraile, A., & García-Arenal, F. (2020). Trends and gaps in forecasting plant virus disease risk. *Annals of Applied Biology*, *176*, 102–108. https://doi.org/10.1111/aab.12553

McLeish, M., Sacristán, S., Fraile, A., & Garcia-Arenal, F. (2017). Scale dependencies and generalism in host use shape virus prevalence. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20172066. https://doi.org/10.1098/rspb.2017.2066

McLeish, M., Sacristán, S., Fraile, A., & García-Arenal, F. (2019). Coinfection organizes epidemiological networks of viruses and hosts and reveals hubs of transmission. *Phytopathology*, *109*, 1003–1010. https://doi.org/10.1094/PHYTO-08-18-0293-R

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, *36*, 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Moreno, I. M., Malpica, J. M., Dıaz-Pendón, J. A., Moriones, E., Fraile, A., & Garcıa-Arenal, F. (2004). Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology*, *318*, 451–460. https://doi.org/10.1016/j.virol.2003.10.002

Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., … Chen, J. (2002). Molecular evolution of *Turnip mosaic virus*: Evidence of host adaptation, genetic recombination and geographical spread. *Journal of General Virology*, *83*, 1511–1521. https://doi.org/10.1099/0022-1317-83-6-1511

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., … Kyrpides, N. C. (2016). Uncovering Earth's virome. *Nature*, *536*, 425–430. https://doi.org/10.1038/nature19094

Pagán, I. (2018). The diversity, evolution and epidemiology of plant viruses: A phylogenetic view. *Infection, Genetics and Evolution*, *65*, 187–199. https://doi.org/10.1016/j.meegid.2018.07.033

Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420. https://doi.org/10.1093/bioinformatics/btp696

Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parker, J., Rambaut, A., & Pybus, O. G. (2008). Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution*, *8*, 239–246. https://doi.org/10.1016/j.meegid.2007.08.001

Peng, Y., Leung, H. C., Yiu, S. M., & Chin, F. Y. (2012). IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Pond, S. L. K., Frost, S. D. W., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*, 676–679. https://doi.org/10.1007/0-387-27733-1_6

Posada, D. (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, *25*, 1253–1256. https://doi.org/10.1093/molbev/msn083

Posada, D., & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution*, *54*, 396–402. https://doi.org/10.1007/s00239-001-0034-9

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus. *Evolution*, *2*, vew007. https://doi.org/10.1093/ve/vew007

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*. Retrieved from http://beast.bio.ed.ac.uk/Tracer

Roossinck, M. J. (2015). Plants, viruses and the environment: Ecology and mutualism. *Virology*, *479*, 271–277. https://doi.org/10.1016/j.virol.2015.03.041

Roossinck, M. J., & García-Arenal, F. (2015). Ecosystem simplification, biodiversity loss and plant virus emergence. *Current Opinion in Virology*, *10*, 56–62. https://doi.org/10.1016/j.coviro.2015.01.005

Sanjuán, R. (2012). From molecular genetics to phylodynamics: Evolutionary relevance of mutation rates across viruses. *PLoS Path*, *8*, e1002685. https://doi.org/10.1371/journal.ppat.1002685

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schneider, W. L., & Roossinck, M. J. (2001). Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *Journal of Virology*, *75*, 6566–6571. https://doi.org/10.1128/JVI.75.14.6566-6571.2001

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*, 437–460.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, *132*, 619–633.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Pond, S. L. K. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*, 773–777. https://doi.org/10.1093/molbev/msx335

Wellborn, G. A., & Langerhans, R. B. (2015). Ecological opportunity and the adaptive diversification of lineages. *Ecology and Evolution*, *5*, 176–195. https://doi.org/10.1002/ece3.1347

Wells, K., & Clark, N. J. (2019). Host specificity in variable environments. *Trends in Parasitology*, *35*, 452–465. https://doi.org/10.1016/j.pt.2019.04.001

Yoder, J. B., Clancey, E., Des roches, S., Eastman, J. M., Gentry, L., Godsoe, W., … Harmon, L. J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, *23*, 1581–1596. https://doi.org/10.1111/j.1420-9101.2010.02029.x

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Peláez A, McLeish MJ, Paswan RR, Dubay B, Fraile A, García-Arenal F. Ecological fitting is the forerunner to diversification in a plant virus with broad host range. *J Evol Biol*. 2021;34:1917–1931. https://doi.org/10.1111/jeb.13672