VIRUS EVOLUTION

Virus Evolution, 2022, **8(2)**, 1–16

DOI: https://doi.org/10.1093/ve/veac095
Advance access publication date: 3 October 2022
Research Article

Metagenomics show high spatiotemporal virus diversity and ecological compartmentalisation: Virus infections of melon, *Cucumis melo*, crops, and adjacent wild communities

Michael J. McLeish,[‡] Adrián D. Zamfir,[†,§] Bisola M. Babalola,[†] Adrián Peláez, Aurora Fraile, and Fernando García-Arenal[*,**]

Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM) and Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, and E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Campus de Montegancedo, UPM, 28223 Pozuelo de Alarcón, Madrid, Spain
[†]Equal contribution authors.
[‡]https://orcid.org/0000-0002-4525-5221
[§]https://orcid.org/0000-0002-5327-3200
[**]https://orcid.org/0000-0002-5327-3200
*Corresponding author: E-mail: fernando.garciaarenal@upm.es

Abstract

The emergence of viral diseases results from novel transmission dynamics between wild and crop plant communities. The bias of studies towards pathogenic viruses of crops has distracted from knowledge of non-antagonistic symbioses in wild plants. Here, we implemented a high-throughput approach to compare the viromes of melon (*Cucumis melo*) and wild plants of crop (Crop) and adjacent boundaries (Edge). Each of the 41-plant species examined was infected by at least one virus. The interactions of 104 virus operational taxonomic units (OTUs) with these hosts occurred largely within ecological compartments of either Crop or Edge, with Edge having traits of a reservoir community. Local scale patterns of infection were characterised by the positive correlation between plant and virus richness at each site, the tendency for increased specialist host use through seasons, and specialist host use by OTUs observed only in Crop, characterised local-scale patterns of infection. In this study of systematically sampled viromes of a crop and adjacent wild communities, most hosts showed no disease symptoms, suggesting non-antagonistic symbioses are common. The coexistence of viruses within species-rich ecological compartments of agro-systems might promote the evolution of a diversity of virus strategies for survival and transmission. These communities, including those suspected as reservoirs, are subject to sporadic changes in assemblages, and so too are the conditions that favour the emergence of disease.

Key words: plant–virus interactions; host range; transmission dynamics; reservoir; ecological network; emergence.

Introduction

The socioeconomic impact of virus diseases on crops (Nicaise 2014; Aranda and Freitas-Astúa 2017) has biased knowledge on plant–virus interactions towards those pathogenic to crops or, more recently, wild plants (Roossinck and García-Arenal 2015; Malmstrom and Alexander 2016; McLeish, Fraile, and García-Arenal 2020). The concept that viruses may not necessarily be antagonistic symbionts of plants, but commensal or even facultative mutualists (Roossinck 2005, 2011), challenges conventional views on the evolution of plant–virus interactions, which had assumed coevolution as a result of the negative effects of infection and defence on each partner's fitness (Woolhouse et al. 2002). Evidence for non-antagonistic plant–virus interactions arose in parallel with techniques that allowed massive analyses of nucleotide sequences. High-throughput sequencing (HTS) provides a means for the identification of the viruses that infect an individual host or a community of hosts with no bias towards those that cause diseases (Roossinck et al. 2010; Wu et al. 2015; Massart et al. 2017; Claverie et al. 2018; Ma et al. 2019). HTS was readily applied to the identification of the community of viruses (virome) associated with wild plants, allowing a detailed description of the virome of plant communities in a few non-agricultural ecosystems, relevant to analyses of infection dynamics and virus ecology and evolution (Bernardo et al. 2018; Muthukumar et al. 2009; Roossinck et al. 2010; Melcher and Grover 2011; Kamitani et al. 2019; Susi et al. 2019; Ma et al. 2021).

HTS analyses of crop viromes lagged behind those of wild ecosystems (Roossinck and García-Arenal 2015; Jones and Naidu 2019). Studies still do not abound and have focussed on virus discovery and diagnosis (Candresse et al. 2014), the temporal or

spatial variation of crop virus communities (Jones 2014; Xu et al. 2017; Souza et al. 2020), the comparison of virus communities in crops and wild relatives (Vélez-Olmedo et al. 2021), or the comparison of viromes of crops and weeds as potential sources of inoculum for crops (Alcalá-Briseño et al. 2019; Ma et al. 2020). Comparative HTS studies of virus communities in wild plant communities and crop communities are still scarce.

One major challenge for understanding disease systems is moving beyond the traditional focus on single-host–single-pathogen interactions: there is a need to understand host–pathogen interactions within complex multi-host–multi-pathogen communities (Woolhouse and Gowtage-Sequeria 2005; Johnson et al. 2013), and HTS offers an important tool for these studies (Claverie et al. 2018; Titcomb, Jerde, and Young 2019; Maclot et al. 2020; McLeish, Fraile, and García-Arenal 2020; Mcleish et al. 2021). Viruses, in general, exhibit high habitat specificity (Paez-Espino et al. 2016), where plant–virus interactions show strong ecological compartmentalisation within habitats (McLeish et al. 2017; Valverde et al. 2020). Ecological compartments are subgroups of taxa, within which many strong interactions occur, with few weak interactions between subgroups (Zhao et al. 2018). In agricultural ecosystems, anthropogenic disturbance has modified biodiversity (Tylianakis, Tscharntke, and Lewis 2007; Roossinck and García-Arenal 2015) and increased spatial heterogeneity and habitat fragmentation, which impacts species movement, that is, cross-species transmission and plant–virus encounter rates (Parratt, Numminen, and Laine 2016). Ecological compartmentalisation is, therefore, expected to generate conditions that influence virus evolution, diversification, and ultimately disease risk in crops (Bass et al. 2019). Specifically, human-driven biodiversity loss that results in ecosystem simplification is considered to favour novel virus–host encounters and novel transmission dynamics that lead to virus emergence (Stukenbrock and McDonald 2008; Jones et al. 2013; Roossinck and García-Arenal 2015). Conversely, wild plant communities are expected to support highly diverse virus communities and a lower prevalence of infection. This expectation derives from the assumption that host and parasite species richness are positively correlated because communities with more host species offer a higher number of niches for symbionts to exploit, as well as the specificity of host–parasite interactions (Lafferty 2012; Kamiya et al. 2014a). The assumptions and expectations above have rarely been tested with plants and viruses (Pagán et al. 2012; Rodelo-Urrego et al. 2013; Susi and Laine 2021), and analyses that jointly consider virus communities in crops and wild plant communities are required. This is the goal of this study, which is addressed by: (1) describing the virome of a crop; (2) comparing the virome of the crop with that of wild plants growing in the same space or adjacent to the crop field; and (3) analysing the variation of these viromes in space and time.

The crop chosen for these analyses is melon (*Cucumis melo*) in the agricultural region of south-central Spain, where virus infections of melon and wild plant communities have been analysed by our group using non-high-throughput tools (Sacristán, Fraile, and García-Arenal 2004; Malpica et al. 2006). We applied an HTS approach to the detection of viruses in wild plant communities that form edges between crop fields, as previous studies had suggested that edges were potential reservoir communities. An ecological compartment may function as a reservoir community, defined as an epidemiologically connected system that maintains infectious agents indefinitely and contributes to transmission (Ashford 2003). The edge habitat has higher plant and virus richness with relatively undisturbed assemblages compared to the crop field and supports a larger number of plant–virus and virus–virus interactions (McLeish et al. 2017, 2019). In this system, we use a metagenomic detection protocol for plant viruses to test the following hypotheses: (a) wild plant communities support a richer virus community than crop communities; and (b) crop and edge communities each form an ecological compartment with few weak interactions between them.

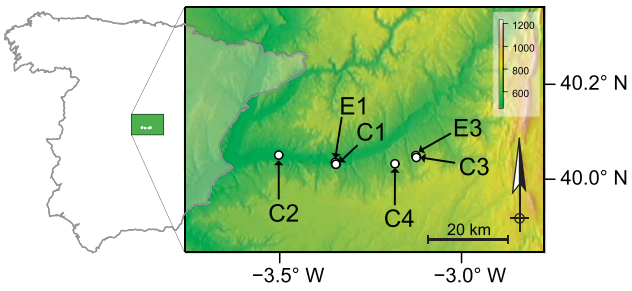

Figure 1. Location of study sites in south-central Spain. Crop sites (C) and Edge sites (E).

Materials and methods

Study area and sampling

We conducted thirteen plant collections between July 2015 and July 2016 in the Vega del Tajo-Tajuña region in south-central Spain (Fig. 1, Supplementary Table S1). Six sites were chosen a priori to represent plant communities of two distinct habitats in the region affected by different intensities of anthropic disturbance. Four sites where melon (*Cucumis melo* var. Piel de Sapo) was grown (Crop habitat sites C1, C2, C3, and C4, from here, Crop) were compared to relatively permanent communities that form the narrow borders (Edge habitat sites E1 and E3, here from Edge) that separate crop fields. The *C. melo* crops are annual monocultures in a habitat that is left fallow between seasons. Wild plants (weeds) grow intermixed with the crop in the Crop habitat. Edge experiences occasional disturbance by either burning or partial ploughing. At each site, fifty samples from a 25 m × 2 m area were collected systematically at each resampling according to fixed itineraries, regardless of whether plants showed symptoms of virus infection or not, as evaluated visually (Supplementary Methods and Mcleish et al. 2021). Sampling completeness at each site was estimated with the Chao 1 estimator of asymptotic species richness. Rarefaction curves and the number of expected species were estimated using the R packages *vegan* (Oksanen et al. 2020) and *iNEXT* (Hsieh, Ma, and Chao 2016).

Library preparation and sequencing

Rare species assumed to account for a negligible number of infections were removed from the sample of plant taxa sent for sequencing to minimise the potential for virus-negative libraries. Elimination of rare species is a common practice in studies of ecological interactions, as their inclusion may increase variances due to small sample sizes (Vázquez et al. 2005; Blüthgen et al. 2008), although the consequences of this procedure are a matter of debate in ecology (Poos and Jackson 2012). Species with five or more individuals in at least one of the two habitats were retained for HTS. Kendall's tau (τ) coefficient was used to test the correlation between species richness at each site before and after the rare species were omitted. Total RNA from individual plants of the species retained for HTS was extracted with the Agilent Plant RNA Isolation Mini Kit (Catalogue No.: 5185-5998), SpectrumTM Plant Total RNA Kit (Catalogue No.: STRN50), and

the hexadecyltrimethylammonium bromide method (Chang et al. 1993) depending on the plant species. To minimise contamination across samples, extractions were done in groups of samples from the same species and collection (i.e. same time and same site). Plastic bags from field collections were opened in the laboratory, and plant tissues were ground in liquid nitrogen with a mortar and pestle (one for each sample) and sterile glass powder before adding the extraction kit reactants. Mortars and pestles were washed in 0.2 M NaOH and autoclaved before being used. Working surfaces were also washed with 0.2 M NaOH between the manipulation of groups of samples from the same plant species and collection. RNA extracts from the same species and collection were pooled to obtain a single HTS library (Supplementary Table S2). High-throughput library preparation and sequencing were outsourced to the Centre de Regulació Genòmica (Barcelona, Spain). Sequencing was performed on different platforms according to the provider's changes in infrastructure. Insert sizes were 220 nt on average. Both read orientations of each pool were run in the same lane, and each pool was run in a separate lane. All libraries included a step for rRNA depletion using the RiboZero kit. Shotgun sequence libraries were simulated with the Grinder software (Angly et al. 2012) to estimate the number of reads required to detect operational taxonomic units (OTUs) from complex environmental samples. It was estimated that 8.0×10^6 paired-end reads (i.e. 16.0×10^6 reads in total) were sufficient to detect low titre viruses in libraries with a proportion of contaminating nucleic acids (Dubay 2017). Approximately 8.0×10^6 paired-end reads of either 125 or 150 nt per library were sequenced on Illumina HiSeq platforms. All reads were provided with Phred quality scores greater than Q30. Trimming of adaptor contamination was conducted using *cutadapt* v1.8.3 (Martin 2011). Details on the HTS platform used, chemistry employed, read length, and the fraction of reads that match viral references for each library are given in Supplementary Table S3.

To control for homogeneity of sequence quality, we analysed the detection of plant genes that are expected to be conserved among plant taxa (Expósito-Rodríguez et al. 2008). Four conserved genes encoding DnaJ-like protein, elongation factor 1-alpha, glyceraldehide-3-phosphate dehydrogenase, and ribosomal protein L2 were detected in 81, 100, 99.7, and 92.8 per cent of 306 analysed libraries (the number of libraries for this quality control exceeded those analysed in this study), respectively, indicating a comparable quality for sequence detection across samples.

Metagenomic detection protocol

We assembled read sequence libraries that had undergone FastQC and trimming of adapter contamination and then, with the contigs generated, conducted BLAST queries against local databases. The assemblies were implemented with IDBA v.1.1.1 (Peng et al. 2012). The Iterative De Bruijn graph Assembler for sequencing data with highly Uneven Depth algorithm was used for the metagenomic assemblies, which are optimised for read sequence data with highly uneven depth. The composition of contigs generated from the assemblies depended on the amount of contamination by non-viral reads in a given library. The subtraction, or not, of non-viral reads before assembly biased the detection of true positive OTUs. The most reproducible method involved eliminating the assembly step from the pipeline but was instead used as a guide to the effectiveness of the pipeline in detecting true positive OTUs (see Supplementary Information). Local BLAST queries were conducted with BLAST+ version 2.2.29 (Camacho et al. 2009) to detect virus OTUs. The queries were conducted against a database of plant virus genomic references of positive- and negative-sense single-stranded RNA (ssRNA), double-stranded RNA (dsRNA), and double- and single-stranded DNA (dsDNA and ssDNA) available from NCBI Viral Genome Browser (https://www.ncbi.nlm.nih.gov/genomes/ accessed December 2018). Uncertainty in quantifying plant–virus ecological interactions derives from the sensitivity of the method used to detect infections and the presence of false positives produced by sequence homology between different taxa. Thus, our aim was to decrease the frequency of false positive detections without excluding true positive virus OTUs that have large intraspecific sequence divergence with a reference, and to standardise the BLAST query matches across libraries. The high proportion of unknown virus species expected in wild plant communities (Bernardo et al. 2018) prevents assigning taxonomy correctly to all viruses present in the sample using annotated sequence references. The presence of unknown viruses increases the likelihood that homologous regions between different species (i.e. the query and reference) will lead to false positive taxonomic assignments. In addition, the detection of a virus OTU by BLAST may be based on either thousands of read sequences when the virus is present in high titres or incidence (libraries being derived from pooled samples), or on a few reads when a virus is present in low titres or incidence. True positive virus OTUs with few reads may be excluded when a threshold number of reads is required to substantiate the match with a reference. Rather than setting such a threshold, we implemented an *in silico* selection procedure that did not rely on the number of reads detected to verify the credibility of OTUs. Reads were retained through the following pipeline (Fig. 2) only if: (1) the query coverage was 100 per cent; (2) the alignment length was greater or equal to 125 nt; (3) where each read-pair of a given library matched a single reference genome only; (4) and matched with a single region of that reference; and (5) where the difference between the maximum and minimum of query start positions relative to a reference genome had a span of more than 1 per cent of the length of the reference (effectively making it a criterion for a minimum of at least one read-pair match per OTU detection); and (6) the matches were paired-reads only.

Steps 1 and 2 of the *in silico* selection procedure were implemented to optimise sequence similarity between read and reference while allowing for mismatches and gaps expected from the presence of viruses that are divergent from the available references. The query coverage (sometimes 'query cover') describes how much a reference sequence is covered by a read query. We required that the whole read query spanned the corresponding positions of the reference, that is, at a query coverage of 100 per cent. The highest percentage query coverage was selected to minimise false positive detections by avoiding high percentage-identity matches that can occur even if only a fraction of the read query corresponds with the reference. The presence of divergent taxa expected from environmental samples means that indels can be introduced to match between query and reference. To be sensitive to the presence of indels and retain divergent virus taxa through the selection procedure, we used an alignment length cut-off at a minimum of 125 nt. The alignment length parameter is the number of nucleotides in the high scoring pair of the matching section of DNA sequences that BLAST returns and includes gaps and mismatches. The alignment length will be the same as the query length when there are no indels between the read query and reference sequences. This step allowed relaxed sequence similarity between query and reference, given maximum query coverage of reads of either 125 or 150 nt in length. Uncertainty in true positive detections and correct taxonomic assignment is also a function of the quality and quantity of matches at each

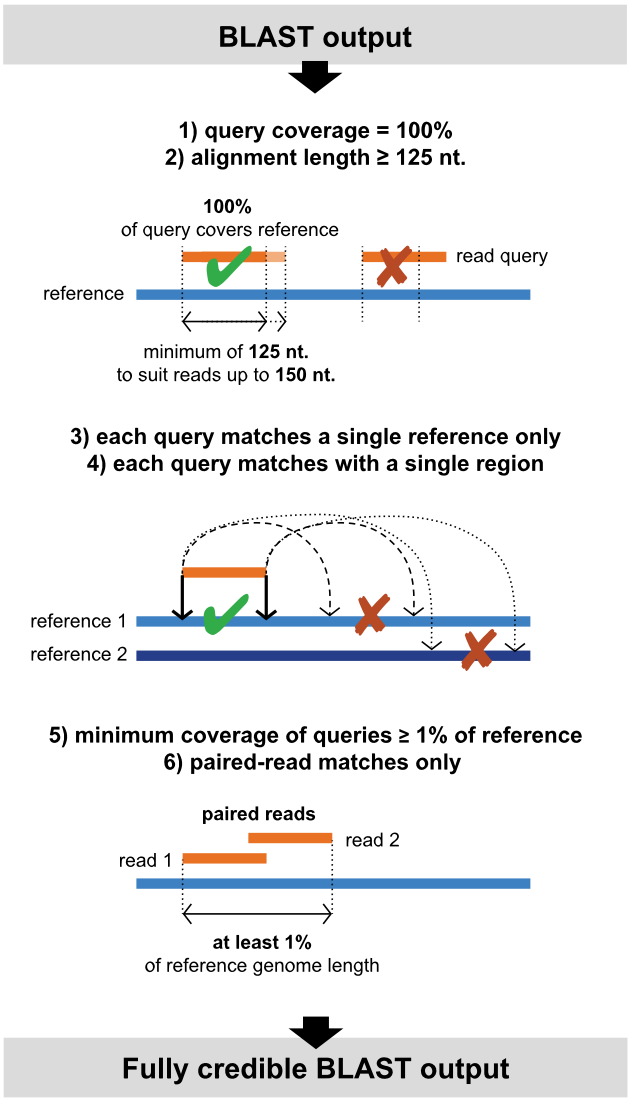

Figure 2. A flowchart explaining the *in silico* selection procedure implemented to standardise the detection of OTUs across read libraries.

region of the reference (vertical coverage), and across its length (horizontal coverage). The remaining Steps 3–5 further optimised read–reference homology, took advantage of read-pairing, and set a minimum threshold of horizontal coverage of the reference genome. Our method was not intended to identify completely novel viruses, as detecting unknown-unknowns—viruses that are completely novel and share no sequence similarity with other known viruses (Stobbe and Roossinck 2014)—is beyond the practicalities of community-level studies. Instead, we focussed on the detection of known (annotated) virus species and unknown variants of them that fulfil all six selection steps.

RT-PCR detection

The presence in plant samples of RNA of seven viruses with unexpectedly large host ranges according to HTS data [tobacco mild green mosaic virus (TMGMV), pepper mild mottle virus (PMMoV), rubus chlorotic mottle virus (RuCMV), tomato aspermy virus (TAV), pelargonium zonate spot virus (PZSV), plum pox virus (PPV), and turnip mosaic virus (TuMV)] and of one virus known to have a large host range (cucumber mosaic virus, CMV) was analysed further by RT-PCR using specific primers that amplified regions of 70–148 nucleotides in the coat protein gene (Supplementary Table S4). In all cases, primer sequences were a consensus of the sequences for that taxon in the databases. For RT-PCR, the complementary DNA (cDNA) was synthesised with the Maxima First Strand cDNA Synthesis Kit (ThermoFisher Scientific, Lithuania) on 0.5 μg of the pooled RNA extracts that had been sent for HTS. The PCR was performed with the Supreme NZYTaq II polymerase kit (NZYtech, Lisboa, Portugal) with the following cycle conditions: initial denaturation at 95°C for 3 minutes, forty cycles of denaturation at 94°C for 30 seconds, annealing temperature based on each primer pair for 30 seconds, and extension at 72°C for 30 seconds. Lastly, a final extension of 72°C for 7 minutes was carried out. The PCR products were visualised by gel electrophoresis in 2.0 per cent agarose after ethidium bromide staining.

Network topology

The fully selected OTUs were retained to generate plant–virus interaction networks. To compare variation in community interactions, we calculated the mean node degree (k), modularity (Q) (Newman 2006), and nestedness (nestedness metric based on overlap and decreasing fill (NODF)] of network topology. Each node in the network represents an OTU or host species, and for each node, the number of links describes its 'degree' k (Blüthgen et al. 2008). Networks have been found to be divided into modules, and the quality function called modularity (Q) is a spectral algorithm for community detection (Newman 2006). Nestedness is a metric used to describe patterns of species interactions in bipartite networks. Higher values of nestedness have been interpreted as an indication of weaker reciprocal specialisation that is required for coevolution (Blüthgen et al. 2008). Maximum nestedness indicates that each virus OTU is found in a subset of the hosts in which other OTUs with wider host ranges occur (Almeida-Neto and Ulrich 2011). As the networks differed in the number of nodes and links, z-scores were used to standardise the Q and NODF indices (Miyauchi and Kawase 2016) for comparisons. To test whether the observed index value differed from chance expectations, a null set of 1,000 networks of the same size as the observed was simulated, and a t-test was conducted to compare the observed index value with the mean of the null set.

We estimated network modules (Dormann and Strauss 2014) as approximations of compartments since modules represent groups of species with many interactions within them and few interactions with species in other groups (Fortuna et al. 2010). Network indices and modules were estimated from undirected (i.e. symmetric associations) graphs, as traits of both viruses and plants determine plant–virus interactions. All network analyses were conducted with the R (R Core Team 2018) packages *igraph* (Csardi and Nepusz 2006) and *bipartite* (Dormann, Gruber, and Fründ 2008).

Results

Sample completeness and detection of virus OTUs

The collections produced 645 individual samples of sixty-seven plant species (twenty-two families) distributed over Crop and Edge. Rarefaction curves of collections from each site (Supplementary Fig. S1) approached an asymptote, which indicated sufficient sample completeness. Site E3 of Edge was under-sampled compared to the other sites that all had Chao1 estimates of species richness close to the observed value (Supplementary Table S5). After rare species were omitted, the sample was reduced to fifty-one species (eighteen families) and 614 samples. Species

Table 1. Virus OTUs detected from all host species of sites of Crop habitat for the fully selected reads (C1 and C3) and for the >4-reads dataset (C1 (>4), C3 (>4)), which included *C. melo* and wild plants in the same field.

Virus (OTU)[a]	Family	Reference genome of virus OTU	Abbr.	C1	C3	C1 (>4)	C3 (>4)
Alphaendornavirus 1	Endornaviridae	Cucumis melo endornavirus	CmEV	1	1	1	1
Anulavirus 1	Bromoviridae	Pelargonium zonate spot virus	PZSV	0	1	0	1
Aureusvirus 2	Tombusviridae	Pothos latent virus	PoLV	1	0	ND	ND
Cucumovirus 1	Bromoviridae	Cucumber mosaic virus	CMV	1	1	1	1
Cucumovirus 2	Bromoviridae	Gayfeather mild mottle virus	GMMV	1	0	1	0
Cucumovirus 3	Bromoviridae	Tomato aspermy virus	TAV	1	0	1	0
Potyvirus 2	Potyviridae	Bean common mosaic virus	BCMV	1	0	1	0
Potyvirus 3	Potyviridae	Cowpea aphid-borne mosaic virus	CABMV	1	0	1	0
Potyvirus 4	Potyviridae	Calla lily latent virus	CLLV	1	0	1	0
Potyvirus 5	Potyviridae	Fritillary virus Y	FVY	1	0	1	0
Potyvirus 6	Potyviridae	Iranian johnsongrass mosaic virus	IJGMV	1	0	1	0
Potyvirus 12	Potyviridae	Soybean mosaic virus	SbMV	1	0	1	0
Potyvirus 16	Potyviridae	Telosma mosaic virus	TelMV	1	0	1	0
Potyvirus 17	Potyviridae	Tomato necrotic stunt virus	TNSV	1	0	ND	ND
Potyvirus 18	Potyviridae	Turnip mosaic virus	TuMV	0	1	ND	ND
Potyvirus 19	Potyviridae	Watermelon mosaic virus	WMV	1	1	1	1
Potyvirus 20	Potyviridae	Yam bean mosaic virus	YBMV	1	0	ND	ND
Potyvirus 21	Potyviridae	Yam mosaic virus	YMV	1	0	1	0
Potyvirus 22	Potyviridae	Zucchini yellow mosaic virus	ZYMV	1	0	ND	ND
Sobemovirus 1	Solemoviridae	Rubus chlorotic mottle virus	RuCMV	1	1	0	1
Tobamovirus 3	Virgaviridae	Pepper mild mottle virus	PMMoV	1	1	ND	ND
Tobamovirus 6	Virgaviridae	Tobacco mild green mosaic virus	TMGMV	1	1	0	1
Tobamovirus 12	Virgaviridae	Youcai mosaic virus	YoMV	0	1	ND	ND

[a]OTUs names are the genus names of the reference sequence plus an ordinal. ND = not detected.
1 indicates detection of a virus OTU in a given site, and 0 indicates a non-detection. Abbr. = Abbreviation.

richness between each site before and after rare species reduction was positively correlated (Kendall rank correlation, $z = 2.141$, P value = 0.032, $\tau = 0.786$).

In total, 526 extractions were pooled to create ninety-six libraries, which comprised forty-one plant species (some samples of Poaceae collected prior to flowering were collapsed into a single taxon) of up to seventeen families (i.e. there were unidentified plant species). The mean number of samples per library was 5.48 (range 1–46, Supplementary Table S2).

Ninety-three of the ninety-six libraries sequenced were positive for at least one virus. After the application of the first two selection criteria to the raw BLAST query output, 150 OTUs were detected (Supplementary Table S6). The metavirome was dominated by positive-sense ssRNA viruses [(+)ssRNA, 128 OTUs], with the rest comprised of negative-sense ssRNA [(–)ssRNA, $n = 5$ OTUs], dsRNA (two OTUs), ssDNA (sixteen OTUs), and reverse-transcribing double-stranded DNA viruses (nine OTUs). We identify OTUs by the name of the matched genus followed by an ordinal (e.g. Tobamovirus i). For each OTU indicated in the table, the taxon name and the NCBI accession to which it matched during the BLAST query is given (Tables 1–3). After application of the remaining selection criteria, 104 OTUs were detected (Tables 1–3 and Supplementary Tables S7 and S8) from 506 individual extractions from ninety-three libraries that represented forty-one plant species (minimum seventeen families). There were only three libraries in which viruses were not detected, but these libraries were of plant species from which viruses were detected in other libraries (*Solanum nigrum, Picris echioides*, and *Lithospermum arvense*). There was no correlation between the number of OTUs detected per library and the number of RNA extracts from individual samples pooled, neither when the first two (not shown) nor the full set of selection criteria (Supplementary Fig. S2) were applied.

Of the 104 OTUs, fifteen were observed in both Crop and Edge, nine only in Crop, and eighty only in Edge (Supplementary Table S7). There were twenty-two OTUs detected in *C. melo* at Crop sites C1–C4 (twenty-one in sites C1 and C3). The best-represented genera in the fully selected metavirome of Crop and Edge together were *Polerovirus* (twenty-six OTUs), *Potyvirus* (twenty-two OTUs) and *Tobamovirus* (twelve OTUs) with (+)ssRNA genomes, and *Curtovirus* (5 OTUs) with ssDNA. High numbers of OTUs were from insect-transmitted genera, mostly by aphids (fifty-nine OTUs in six genera), but also by contact/soil-transmitted (seventeen OTUs in five genera). Other transmission modes (Supplementary Table S10) were by other insects, including whiteflies (two OTUs, two genera), planthoppers (one OTU), leafhoppers (three OTUs, two genera), Coleoptera (two OTUs, two genera), by mites (one OTU), and by fungi (five OTUs, three genera) (Tables 1–3).

The virome of *C. melo* varied among the four Crop sites (Table 2), with only Cucumovirus 1 and Alphaendornavirus 1 detected from all Crop sites. *Cucumis melo* samples from sites C1 and C2 were those that shared more OTUs (thirteen OTUs). Six OTUs, Cucumovirus 3, Cucumovirus 2, Potyvirus 22, Tobamovirus 12, Anulavirus 1, and Tritimovirus 1, were detected in only one site. The *C. melo* virome varied largely between each of the C1–C2 and C3–C4 pairs of sites (Fig. 1, Supplementary Fig. S3), indicating a spatial pattern between virus communities from the eastern and western extents of the study area.

The data in Tables 1–3 and Supplementary Table S8 show unexpected virus OTU–host plant species associations. Thus, for eight (+)ssRNA viruses, seven with unexpectedly large host ranges and one with a known large host range, detection by HTS was confirmed by RT-PCR. Amplification confirmed the presence of the analysed virus species in between 82.3 per cent (for PMMoV) and 100 per cent (for PPV and TuMV) of the analysed libraries in which reads matched the full set of selection criteria (Supplementary Table S9). The data in Supplementary Table S9 show that detections based on a few (2–4) reads may show true positive virus–plant associations. Also, RT-PCR detections occurred in libraries in

Table 2. Viruses detected only in *Cucumis melo* in sites of Crop habitat for the fully selected reads (C1–C4) and for the >4-reads dataset [C1 (>4) to C4 (>4)].

Virus (OTU)[a]	Family	Reference genome of virus OTU	Abbr.	C1	C2	C3	C4	C1 (>4)	C2 (>4)	C3 (>4)	C4 (>4)
Alphaendornavirus 1	Endornaviridae	*Cucumis melo* endornavirus	CmEV	1	1	1	1	1	1	1	1
Anulavirus 1	Bromoviridae	Pelargonium zonate spot virus	PZSV	0	0	1	0	ND	ND	ND	ND
Cucumovirus 1	Bromoviridae	Cucumber mosaic virus	CMV	1	1	1	1	1	1	1	1
Cucumovirus 2	Bromoviridae	Gayfeather mild mottle virus	GMMV	1	0	0	0	1	0	0	0
Cucumovirus 3	Bromoviridae	Tomato aspermy virus	TAV	1	0	0	0	1	0	0	0
Potyvirus 2	Potyviridae	Bean common mosaic virus	BCMV	1	1	0	0	1	1	0	0
Potyvirus 3	Potyviridae	Cowpea aphid-borne mosaic virus	CABMV	1	1	0	0	1	1	0	0
Potyvirus 4	Potyviridae	Calla lily latent virus	CLLV	1	1	0	1	1	1	0	0
Potyvirus 5	Potyviridae	Fritillary virus Y	FVY	1	1	0	0	1	0	0	0
Potyvirus 6	Potyviridae	Iranian johnsongrass mosaic virus	IJGMV	1	1	0	0	1	0	0	0
Potyvirus 12	Potyviridae	Soybean mosaic virus	SbMV	1	1	0	1	1	1	0	1
Potyvirus 16	Potyviridae	Telosma mosaic virus	TelMV	1	1	0	0	1	0	0	0
Potyvirus 17	Potyviridae	Tomato necrotic stunt virus	TNSV	1	1	0	0	ND	ND	ND	ND
Potyvirus 19	Potyviridae	Watermelon mosaic virus	WMV	1	1	0	1	1	1	0	1
Potyvirus 20	Potyviridae	Yam bean mosaic virus	YBMV	1	1	0	0	ND	ND	ND	ND
Potyvirus 21	Potyviridae	Yam mosaic virus	YMV	1	1	0	0	1	1	0	0
Potyvirus 22	Potyviridae	Zucchini yellow mosaic virus	ZYMV	1	0	0	0	ND	ND	ND	ND
Sobemovirus 1	Solemoviridae	Rubus chlorotic mottle virus	RuCMV	0	0	1	1	0	0	1	0
Tobamovirus 3	Virgaviridae	Pepper mild mottle virus	PMMoV	1	0	1	1	ND	ND	ND	ND
Tobamovirus 6	Virgaviridae	Tobacco mild green mosaic virus	TMGMV	1	0	1	1	0	0	1	1
Tobamovirus 12	Virgaviridae	Youcai mosaic virus	YoMV	0	0	1	0	ND	ND	ND	ND
Tritimovirus 1	Potyviridae	Wheat streak mosaic virus	WSMV	0	0	0	1	ND	ND	ND	ND

[a]OTUs names are the genus names of the reference sequence plus an ordinal. ND = not detected.
1 indicates detection of a virus OTU in a given site, and 0 indicates a non-detection. Abbr. = Abbreviation.

Table 3. Viruses detected from all host species of sites of Edge habitat for the fully selected reads (E1 and E3) and for the >4-reads dataset [E1 (>4), E3 (>4)].

Virus (OTU)[a]	Family	Reference genome of virus OTU	Abbr.	E1	E3	E1 (>4)	E3 (>4)
Alphaendornavirus 1	Endornaviridae	*Cucumis melo* endornavirus	CmEV	0	1	ND	ND
Alphanecrovirus 1	Tombusviridae	Olive latent virus 1	OLV1	0	1	0	1
Alphanecrovirus 2	Tombusviridae	Olive mild mosaic virus	OMMV	0	1	0	1
Alphanecrovirus 3	Tombusviridae	Potato necrosis virus	PNV	0	1	ND	ND
Anulavirus 1	Bromoviridae	Pelargonium zonate spot virus	PZSV	1	1	1	1
Aureusvirus 1	Tombusviridae	Cucumber leaf spot virus	CLSV	1	0	1	0
Aureusvirus 2	Tombusviridae	Pothos latent virus	PoLV	1	0	1	0
Aureusvirus 3	Tombusviridae	Yam spherical virus	YSV	1	0	1	0
Becurtovirus 1	Geminiviridae	Beet curly top Iran virus	BCTIV	1	1	0	1
Becurtovirus 2	Geminiviridae	Spinach curly top Arizona virus	SCTAV	0	1	0	1
Begomovirus 1	Geminiviridae	Tomato leaf curl New Delhi virus	ToLCNDV	0	1	0	1
Betacarmovirus 1	Tombusviridae	Cardamine chlorotic fleck virus	CCFV	0	1	ND	ND
Bromovirus 1	Bromoviridae	Brome mosaic virus	BMV	0	1	ND	ND
Carmovirus 1	Tombusviridae	Turnip crinkle virus	TCV	0	1	0	1
Caulimovirus 1	Caulimoviridae	Cauliflower mosaic virus	CaMV	1	0	ND	ND
Caulimovirus 2	Caulimoviridae	Lamium leaf distortion associated virus	LLDAV	0	1	ND	ND
Caulimovirus 3	Caulimoviridae	Soybean mild mottle pararetrovirus	SbMMoV	1	0	1	0
Crinivirus 1	Closteroviridae	Sweet potato chlorotic stunt virus	SPCSV	0	1	0	1
Cucumovirus 1	Bromoviridae	Cucumber mosaic virus	CMV	1	1	1	1
Cucumovirus 2	Bromoviridae	Gayfeather mild mottle virus	GMMV	1	1	1	1
Cucumovirus 3	Bromoviridae	Tomato aspermy virus	TAV	1	1	1	1
Curtovirus 1	Geminiviridae	Beet curly top virus	BCTV	0	1	0	1
Curtovirus 2	Geminiviridae	Beet mild curly top virus	BMCTV	0	1	0	1
Curtovirus 3	Geminiviridae	Beet severe curly top virus	BSCTV	0	1	0	1
Curtovirus 4	Geminiviridae	Pepper yellow dwarf virus	PeYDV	0	1	0	1
Curtovirus 5	Geminiviridae	Spinach severe curly top virus	SpSCTV	0	1	0	1
Cytorhabdovirus 1	Rhabdoviridae	Barley yellow striate mosaic virus	BYSMV	1	0	1	0
Deltapartitivirus 1	Partitiviridae	Raphanus sativus cryptic virus 3	RsCV3	0	1	ND	ND
Ilarvirus 1	Bromoviridae	Parietaria mottle virus	PMoV	1	1	0	1
Ilarvirus 2	Bromoviridae	Strawberry necrotic shock virus	SNSV	0	1	ND	ND
Luteovirus 1	Tombusviridae	Barley yellow dwarf virus-GAV	BYDV_GAV	1	1	1	1
Luteovirus 2	Tombusviridae	Barley yellow dwarf virus-MAV	BYDV_MAV	1	1	1	1
Luteovirus 3	Tombusviridae	Barley yellow dwarf virus-PAS	BYDV_PAS	1	1	1	1

(continued)

Table 3. (Continued)

Virus (OTU)[a]	Family	Reference genome of virus OTU	Abbr.	E1	E3	E1 (>4)	E3 (>4)
Luteovirus 4	Tombusviridae	Barley yellow dwarf virus-PAV	BYDV_PAV	1	1	1	1
Luteovirus 5	Tombusviridae	Soybean dwarf virus	SbDV	0	1	0	1
Necrovirus 1	Tombusviridae	Beet black scorch virus	BBSV	1	0	1	0
Necrovirus 2	Tombusviridae	Tobacco necrosis virus A	TNVA	0	1	0	1
Necrovirus 3	Tombusviridae	Tobacco necrosis virus D	TNVD	0	1	ND	ND
Nucleorhabdovirus 1	Rhabdoviridae	Datura yellow vein virus	DYVV	1	0	1	0
Polerovirus 1	Solemoviridae	African eggplant yellowing virus	AEYV	1	1	ND	ND
Polerovirus 2	Solemoviridae	Barley virus G	BaVG	0	1	0	1
Polerovirus 3	Solemoviridae	Beet chlorosis virus	BChV	1	1	0	1
Polerovirus 4	Solemoviridae	Beet mild yellowing virus	BMYV	1	1	1	1
Polerovirus 5	Solemoviridae	Brassica yellows virus	BrYV	1	1	1	1
Polerovirus 6	Solemoviridae	Beet western yellows virus	BWYV	1	1	1	1
Polerovirus 7	Solemoviridae	Cucurbit aphid-borne yellows virus	CABYV	0	1	0	1
Polerovirus 8	Solemoviridae	Cotton leafroll dwarf virus	CLRDV	0	1	0	1
Polerovirus 9	Solemoviridae	Chickpea chlorotic stunt virus	CpCSV	0	1	0	1
Polerovirus 10	Solemoviridae	Carrot red leaf virus	CtRLV	1	0	1	0
Polerovirus 11	Solemoviridae	Cereal yellow dwarf virus-RPS	CYDV_RPS	1	0	1	0
Polerovirus 12	Solemoviridae	Cereal yellow dwarf virus-RPV	CYDV_RPV	1	0	1	0
Polerovirus 13	Solemoviridae	Ixeridium yellow mottle virus 1	IYMoV1	1	0	ND	ND
Polerovirus 14	Solemoviridae	Melon aphid-borne yellows virus	MABYV	0	1	0	1
Polerovirus 15	Solemoviridae	Maize yellow dwarf virus 2	MYDV2	0	1	ND	ND
Polerovirus 16	Solemoviridae	Pepo aphid-borne yellows virus	PABYV	0	1	0	1
Polerovirus 17	Solemoviridae	Phasey bean mild yellows virus	PBMYV	0	1	0	1
Polerovirus 18	Solemoviridae	Potato leafroll virus	PLRV	0	1	0	1
Polerovirus 19	Solemoviridae	Pepper vein yellows virus	PVYV	0	1	0	1
Polerovirus 20	Solemoviridae	Suakwa aphid-borne yellows virus	SABYV	1	1	ND	ND
Polerovirus 21	Solemoviridae	Sauropus yellowing virus	SaYV	1	0	1	0
Polerovirus 22	Solemoviridae	Sugarcane yellow leaf virus	ScYLV	1	0	1	0
Polerovirus 23	Solemoviridae	Turnip yellows virus	TuYV	1	1	1	1
Polerovirus 24	Solemoviridae	Tobacco vein distorting virus	TVDV	0	1	0	1
Polerovirus 25	Solemoviridae	Wheat leaf yellowing-associated virus	WLYaV	1	0	1	0
Polerovirus 26	Solemoviridae	Wheat yellow dwarf virus-GPV	WYDV	1	0	1	0
Potyvirus 1	Potyviridae	Asparagus virus 1	AV1	0	1	0	1
Potyvirus 4	Potyviridae	Calla lily latent virus	CLLV	0	1	ND	ND
Potyvirus 6	Potyviridae	Iranian johnsongrass mosaic virus	IJGMV	1	0	1	0
Potyvirus 7	Potyviridae	Johnsongrass mosaic virus	JGMV	0	1	0	1
Potyvirus 8	Potyviridae	Lettuce mosaic virus	LMV	0	1	0	1
Potyvirus 9	Potyviridae	Maize dwarf mosaic virus	MDMV	1	1	1	0
Potyvirus 10	Potyviridae	Plum pox virus	PPV	1	0	ND	ND
Potyvirus 11	Potyviridae	Potato virus Y	PVY	1	0	ND	ND
Potyvirus 12	Potyviridae	Soybean mosaic virus	SbMV	0	1	ND	ND
Potyvirus 13	Potyviridae	Scallion mosaic virus	ScaMV	0	1	0	1
Potyvirus 14	Potyviridae	Sugarcane mosaic virus	SCMV	1	0	1	0
Potyvirus 15	Potyviridae	Sorghum mosaic virus	SrMV	1	0	1	0
Potyvirus 18	Potyviridae	Turnip mosaic virus	TuMV	1	1	0	1
Potyvirus 19	Potyviridae	Watermelon mosaic virus	WMV	1	1	1	1
Sobemovirus 1	Solemoviridae	Rubus chlorotic mottle virus	RuCMV	1	1	1	1
Solendovirus 1	Caulimoviridae	Tobacco vein clearing virus	TVCV	0	1	ND	ND
Tobamovirus 1	Virgaviridae	Bell pepper mottle virus	BPeMV	1	0	ND	ND
Tobamovirus 2	Virgaviridae	Cucumber green mottle mosaic virus	CGMMV	1	0	ND	ND
Tobamovirus 3	Virgaviridae	Pepper mild mottle virus	PMMoV	1	1	1	1
Tobamovirus 4	Virgaviridae	Rehmannia mosaic virus	RheMV	1	1	1	0
Tobamovirus 5	Virgaviridae	Tomato brown rugose fruit virus	TBRFV	1	0	1	0
Tobamovirus 6	Virgaviridae	Tobacco mild green mosaic virus	TMGMV	1	1	1	1
Tobamovirus 7	Virgaviridae	Tobacco mosaic virus	TMV	1	1	1	1
Tobamovirus 8	Virgaviridae	Tomato mottle mosaic virus	ToMMV	1	0	ND	ND
Tobamovirus 9	Virgaviridae	Tomato mosaic virus	ToMV	1	1	1	0
Tobamovirus 10	Virgaviridae	Tropical soda apple mosaic virus	TSAMV	0	1	ND	ND
Tobamovirus 11	Virgaviridae	Wasabi mottle virus	WMoV	0	1	ND	ND
Tobamovirus 12	Virgaviridae	Youcai mosaic virus	YoMV	0	1	0	1
Varicosavirus 1	Rhabdoviridae	Lettuce big-vein associated virus	LBVaV	1	0	1	0
Varicosavirus 2	Rhabdoviridae	Lettuce yellow mottle virus	LYMoV	1	1	1	0

[a]OTUs names are the genus names of the reference sequence plus an ordinal.
1 indicates detection of a virus OTU in a given site, and 0 indicates a non-detection. Abbr. = Abbreviation; ND = not detected.

which the analysed viruses were detected by the two first selection criteria but not by the full selection criteria (0 reads in Supplementary Table S9). On these data, the efficiency of detection of the eight viruses was tested. A binomial mixed regression model with positive or negative RT-PCR detections ($n = 209$) as a response and the number of fully selected read hits as a predictor with a random factor for the detection classes of either two or four reads or more than four reads were used to test whether HTS detections based on few reads were associated with a higher frequency of false positives. The relationship between OTU detection by RT-PCR and the number of read hits was not significant ($z = 1.071$, $P = 0.284$), with the detection class contributing no variance to the binomial logistic regression model. These results indicate that many unexpected virus–host associations are real. Also, the high number of reads in many of these unexpected associations (Supplementary Table S8) indicates virus infection rather than contamination by viruses with highly stable particles such as tobamoviruses or sobemoviruses (see TMGMV in library PV062 from *Cucumis melo*, as an example).

Still, we cannot discard that an unknown fraction of plant–virus associations are due to false positive detections caused by either any of the caveats described in the Metagenomic detection protocol (see Material and Methods), sample cross-talk during sample handling and sequencing, or other possible causes. Examples could be the infection of dicotyledon hosts by CYDV or BYDV, or the detection of CmEV in hosts unrelated to *C. melo*, which were based on very few reads (Supplementary Table S8). Thus, the data set was also reanalysed after omitting all detections based on four or less reads (54 per cent of virus OTUs–plant interactions), to assess the influence of arbitrary cut-off thresholds on our conclusions of differences between crop and wild communities and the presence of ecological compartments (see below). After the application of the six selection criteria to this >4-read database, eighty virus OTUs were detected (Tables 1–3) from libraries that represented forty plant species (Supplementary Table S8). These eighty OTUs had genomes of (+)ssRNA ($n = 66$), (–)ssRNA ($n = 4$), dsRNA ($n = 1$), ssDNA ($n = 8$), or reverse transcribed dsDNA ($n = 1$).

The analyses conducted with >4-read detections showed that there were eight OTUs associated exclusively with Crop, sixty-four OTUs with Edge, and eight OTUs with both habitats. The best-represented genera in the >4-read data of Crop and Edge together were *Polerovirus* (twenty-two OTUs) and *Potyvirus* (17 OTUs), and most OTUs were transmitted by aphids (forty-eight OTUs in four genera) (Tables 1–3 and Supplementary Table S10). The virome based on >4-read data of *C. melo* varied among the four Crop sites (Table 2), with only Cucumovirus 1 and Alphaendornavirus 1 detected from all Crop sites, as in the selected read analysis. *Cucumis melo* samples from sites C1 and C2 were those that shared more OTUs (eight OTUs).

Plant–virus interaction networks

Networks were created to assess the size of the virus assemblage associated with Crop and Edge communities (i.e. Hypothesis a), determine the degree to which viruses shared host species present in Edge and Crop communities (Supplementary Table S11), and estimate the number of interactions between the habitats (i.e. Hypothesis b). Pairs of Crop and Edge sites were compared between two locations approximately 20 km distant (Fig. 1).

The network generated by the aggregation of plant–virus interactions observed in all collections at sites C1–C4 indicated that the central role *C. melo* has as a host in Crop habitat (Fig. 3). Other

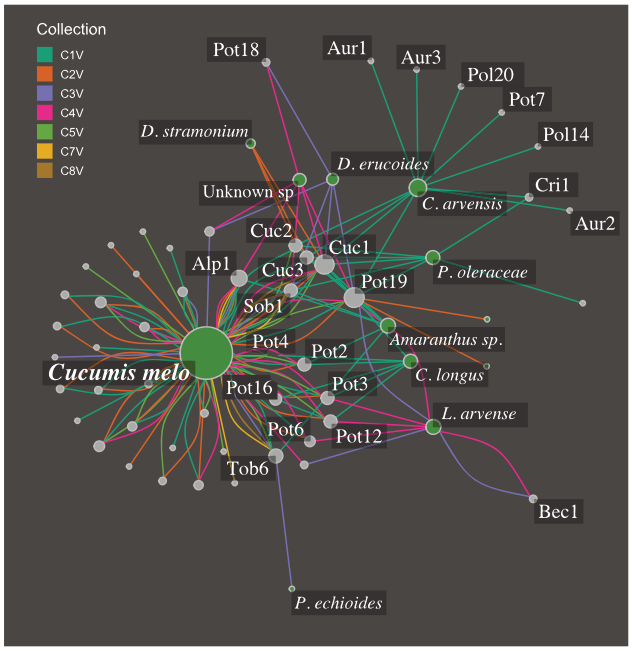

Figure 3. Plant–virus interaction network from all collections made at Crop sites C1–C4. Virus OTUs as detected under the first two selection criteria. Green nodes represent host species and grey nodes represent virus OTUs. Node size is proportional to the square root of the number of interactions detected in either virus or host. The network highlights: (1) the small proportion of nodes that have many interactions, and a larger proportion with very few interactions; (2) the small proportion of interactions not exclusively with *C. melo* were largely confined to a single collection (C1V); and (3) many of the same OTUs were detected in multiple collections. In most collections, there were typically a large proportion of interactions between many of the same OTUs and *C. melo* only. Exemplar nodes highlighted. OTU codes give first three letters of the genus and ordinal given in Supplementary Table S7.

interactions of Crop varied largely according to site and collection. Most OTUs of Crop (Table 1) infected *C. melo* (Table 2), the exceptions being Aureusvirus 2 at site C1 and Potyvirus 18 at site C3. This pattern of interactions, including the host and virus OTUs with nodes that are dominant in the network, was also retrieved when the >4-read data set was analysed (Supplementary Fig. S4). Nine wild host species were observed at Crop sites. *Cucumis melo, Cyperus longus, Portulaca oleraceae*, and an unidentified species were observed only in Crop and not in Edge. The remaining thirty-four host species were observed only in Edge (Supplementary Table S11).

When the network included the Edge sites E1 and E3, an assemblage of viruses that infected *C. melo* exclusively formed a sub-network of a larger set of interactions observed in Edge (Fig. 4). A number of OTUs that were part of a second assemblage central to the entire network infected *C. melo* as well as a diversity of wild plants (thirty-eight species) in Edge. A third assemblage of OTUs infected a large number of wild plants at both sites of Edge only (e.g. Sobemovirus 1, Luteovirus 1–4, and Anulavirus 1), but not necessarily the same host at each site (Fig. 4, Supplementary Table S8). The same pattern of interactions, including the dominant host and virus OTUs nodes, was shown in the less densely populated networks derived from the >4-read data (Supplementary Fig. S5). Thus, the observed structure of the virus OTU–plant infection network either at Crop or Crop and Edge sites is robust to a large variation in OTU interactions (i.e. detection protocol).

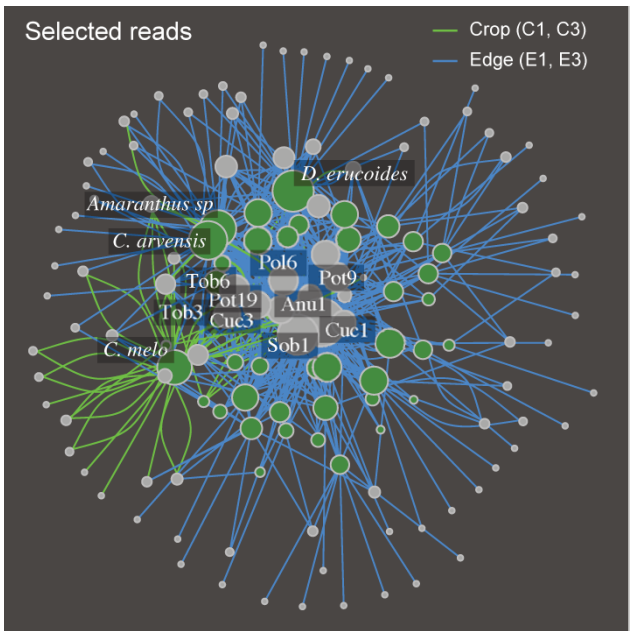

Figure 4. Plant–virus interaction network at Edge (E1 and E3) and Crop (C1 and C3) sites. Virus OTUs ($n = 103$) as detected under the full set of selection criteria. Green nodes represent host species ($n = 40$), and grey nodes represent virus OTUs, with the size proportional to the number of infections. Links (i.e. 638 interactions) are coloured according to the habitat, Edge or Crop, in which the interaction occurs. The network highlights the many interactions of a small proportion of nodes and a larger proportion with very few interactions. Exemplar nodes highlighted. OTU codes give first three letters of the genus and ordinal given in Supplementary Table S7.

Principal component analyses of either host species or virus OTUs by sites (Supplementary Fig. S6) indicated a strong correspondence between assemblages of Crop (C1 and C3), and a large distinction between those of Edge (E1 and E3). The strong correspondence between Crop sites was due to the occurrence of *C. longus, P. oleraceae*, and *C. melo*, and the assembly of OTUs that specialised on hosts of Crop (Supplementary Fig. S6). Taken together, the specialisation of virus assemblages on hosts of either habitat indicated a community-level effect on virus host use.

The OTUs of viruses that infected *C. melo* exhibited realised host ranges (RHRs; Supplementary Table S12) of 1–37 (1–32 in the >4-read data set) host species observed at sites C1, C3, E1, and E3. OTUs in the genera *Anulavirus, Cucumovirus, Sobemovirus*, and *Tobamovirus* (except Tobamovirus 12) had the widest ranges of at least ten species. All *Potyvirus* OTUs, apart from Potyvirus 19, with a host range of 17, had ranges up to three host species. Anulavirus 1 had the widest range of thirty-seven species, followed by Cucumovirus 1 with 34, and Sobemovirus 1 with 29 species. The OTUs observed only in Crop communities were specialists with narrow host ranges, for instance, Potyvirus 2 (realised host range, RHR = 1), Potyvirus 3 (RHR = 2), Potyvirus 5 (RHR = 1), Potyvirus 21 (RHR = 1), and Potyvirus 22 (RHR = 1) (Supplementary Table S13). Except for Potyvirus 3, these OTUs' only host was *C. melo*. Other OTUs detected in Crop [Potyvirus 4 (RHR = 2), Alphaendornavirus 1 (RHR = 5), Potyvirus 6 (RHR = 3), and Potyvirus 12 (RHR = 3)] with less narrow host ranges were also detected in Edge. Finally, other OTUs detected in Crop and Edge had some of the widest host ranges. These OTUs were transmitted by aphids (Cucumovirus 1, RHR = 34; Cucumovirus 2, RHR = 10; Cucumovirus 3, RHR = 19; or Potyvirus 19, RHR = 17), pollen (Anulavirus 1, RHR = 37), or contact/soil (Sobemovirus 1, RHR = 29; Tobamovirus 3, RHR = 18; and Tobamovirus 6, RHR = 17). By contrast, OTUs that were observed only in Edge had either broad (for instance, Polerovirus 5, RHR = 14; Ilarvirus 1, RHR = 20; Potyvirus 9, RHR = 16; Polerovirus 6, RHR = 17; Polerovirus 4, RHR = 12; and Polerovirus 23, RHR = 20) or narrow (below 3, Potyvirus 8, Potyvirus 11, and Carmovirus 1) host ranges. These data showed no obvious association between host range and genome class or mode of transmission. Such associations were not observed in the >4-read data set.

Spatial and temporal changes in virus and host assemblages and interactions

The networks of Edge sites E1 and E3 indicated distinct and dynamic patterns of seasonal plant–virus interactions and changes in species assemblies through time between sites, either when derived from the full data (Fig. 5) set or the >4-read data set (Supplementary Fig. S7 and Supplementary Fig. S8). A Chi-square test between the number of network nodes (plants and viruses) and links (plant–virus interactions) by site and season was not significant ($\chi^2_{(5,\ 758)} = 9.247$, $P = 0.0996$) and indicated that the proportions of the number of species (nodes) to the number of interactions (links) were not dependent on site or season. The proportion of the number of virus OTUs (Nv) to host species (Nh) by site and season was also non-significant ($\chi^2_{(5,\ 226)} = 9.909$, $P = 0.0779$), despite being more than 3:1 at E3 in spring but almost 1:1 at E1. The same results were obtained with the >4-read dataset (not shown). However, these non-significant dependencies with season and site coincided with variation in the mean node degree (k) (Supplementary Table S14). The relative node degree minima in summer and maxima in autumn indicated that the number of interactions coincided with increases in the number of both virus and host species in autumn.

The relationships between the number of interactions and virus and plant richness were characterised by changes in the specificity of interactions through the seasons. Modularity (Q) and nestedness (NODF) were both significantly greater than expected by chance for all networks of each site and season. Although the relationship of the host and virus assemblage between each site and across seasons was weak, the variation in standardised (z-score) Q and NODF indices through the seasons at each site indicated a trend towards both increasing modularity and nestedness between spring and autumn (Supplementary Fig. S8a), which indicated an increasing tendency for specialist viruses to interact within subsets of host species that generalist viruses interact with. Such spring–autumn changes did not occur in the sparsely populated >4-read network at E1. Noticeable decreases in Q during the summer at site E3, and sharp increases in NODF in the autumn at sites E1 and E3, indicated strong changes to network topology in the space of a season. Seasonal variation in community composition corresponded with changes in the dominant OTUs and host species (i.e. those with the most infections). For instance, at site E3 in spring, *Diplotaxis erucoides* and *Papaver rhoeas*, and Cucumovirus 1 and Anulavirus 1, had relatively central roles in network structure. During the summer, no host had a dominant role, while Cucumovirus 1, but not Anulavirus 1, had a narrower host range than in the previous season. In autumn, *Amaranthus* sp. was the dominant host species and Cucumovirus 1 again assumed a central role in the network with Anulavirus 1, Sobemovirus 1, Tobamovirus 3, and Potyvirus 19. Thus, evidence of virus transmission dynamics showed local-scale patterns of infection dependent on the community and season. These seasonal patterns were also observed in the >4-read data set (Supplementary Fig. S8b).

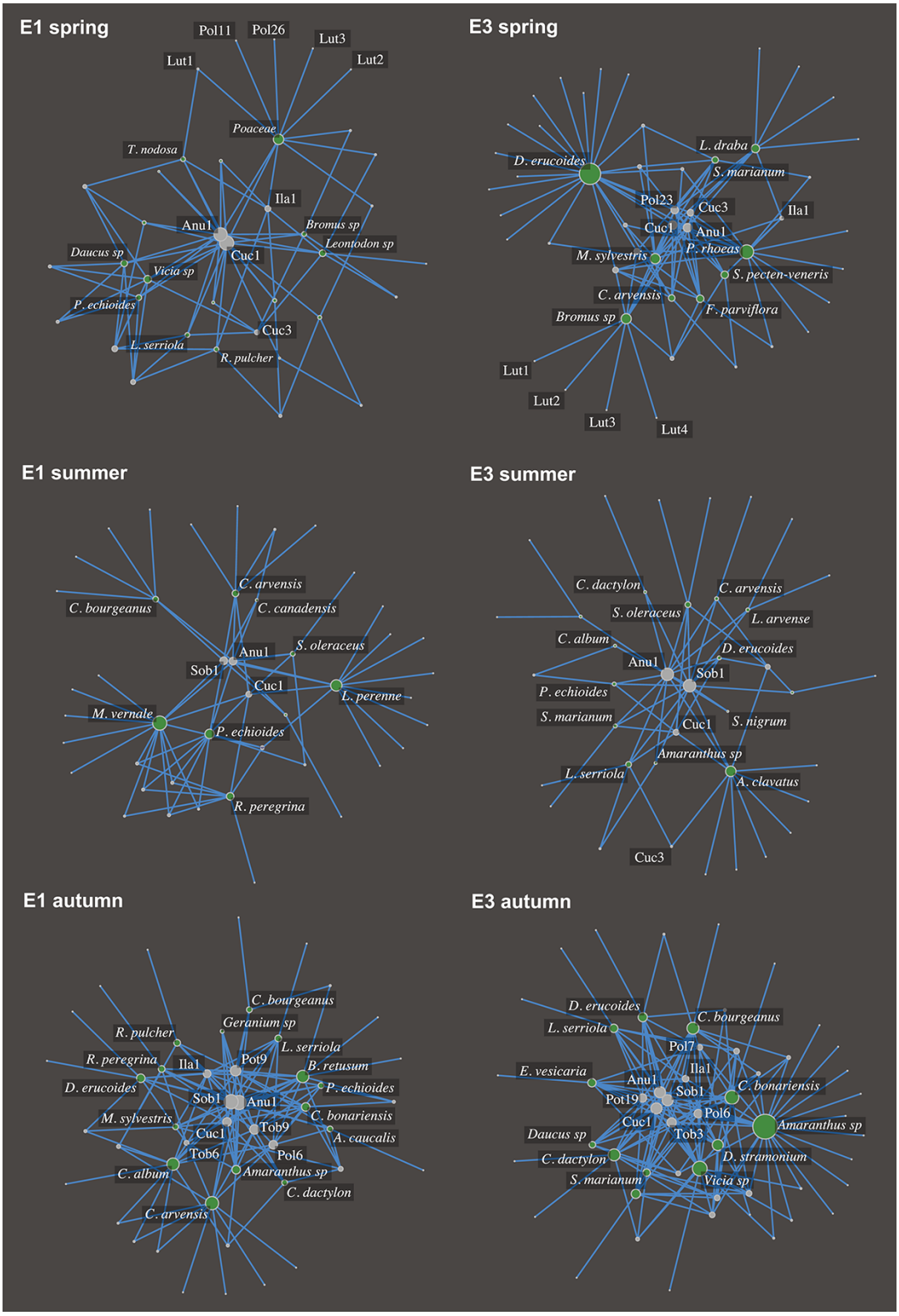

Figure 5. Plant–virus interactions are not static in time or space. Variation in ecological networks from collections made at Edge sites E1 (left) and E3 (right) by season (spring, summer, an autumn). Virus OTUs as detected under the full set of selection criteria. Green nodes represent host species and grey nodes represent virus OTUs. The network highlights the variability in OTU connectivity with hosts among sites and seasons, and the variability in node importance/centrality in the network. Exemplar nodes highlighted. OTU codes give first three letters of the genus and ordinal given in Supplementary Table S7.

Discussion

In this study, we used an HTS-based metagenomic approach to characterise the virome of an important crop species, the melon *C. melo*, and that of the wild plants growing in its proximity. The aim of the study was not virus discovery but, rather, to understand multi-host–multi-pathogen interactions in complex

communities. This community ecology perspective reveals virus resource usage and its effect on virus evolution, an underexplored topic (Woolhouse and Gowtage-Sequeria 2005; Johnson et al. 2013). We thus limited our scope to a part of the virome, i.e. the one that could be identified in reference to described viruses. Metagenomic studies may be based on different procedures aimed at the enrichment of viral sequences or on total RNA extracts, each with advantages, disadvantages, and associated detection biases (Lacroix et al. 2016). We chose to sequence the total RNA of plant extracts that should not introduce major taxonomic biases but may bias against low titre or low-incidence viruses. Bias related to differences in virus incidence may also derive from differences in sample size across hosts, which is unavoidable as host sample size reflected host abundance (see Material and Methods). A source of potential heterogeneity in the data might derive from differences in the quality of RNA extracts associated to plant taxonomy and/or phenology. Detection of messenger RNAs of four conserved plant genes (see Material and Methods) showed comparable quality for sequence detection across samples. Our procedure of pooling RNA extracts of samples from the same plant species and collection for HTS resulted in variation in the number of samples per library that could affect the homogeneity of virus OTU detection across libraries. There was no relationship between the number of virus OTUs detected and the number of individual samples pooled per library, which indicated that this heterogeneity should not result in a significant bias. Uncertainty in the detection of plant–virus interactions derives from the sensitivity of the method used to detect infections, and the presence of false positives may be produced by various mechanisms, such as the incorrect assignment of unknown taxa due to sequence homology with different taxa and contamination due to sample cross-talk during the whole process of sample handling and sequencing, which we cannot discard despite careful manipulation of samples to minimise this possibility. To decrease the frequency of false positive detections, we developed a stringent selection pipeline. In addition, for a set of eight generalist viruses, detection by HTS was confirmed by RT-PCR with specific primers, which showed that HTS detections based on few reads were often true positives by RT-PCR. A binomial logistic regression model showed that detection by RT-PCR was no less efficient in libraries in which few reads of the analysed virus OTU were detected. However, we cannot discard that an unknown fraction of the HTS-detected plant–virus interactions was due to false positive detection. Thus, we analysed the dataset of all virus–OTUs interactions based on the full selection criteria and a parallel data set in which the large fraction (54 per cent) of detections based on four or less reads were eliminated (>4-read data set). The number of OTUs in the >4-read data was affected less by a reduction in number than by the number of interactions that were observed for each of the OTUs. This discrepancy showed that the detection by HTS of a given OTU in a community was better than the detection of all the interactions it was associated with and attributable to low virus titre more than to sequence homology errors. The results of all analyses were consistent for the two data sets, showing that the observed patterns of virus–plant interactions that we discuss in this section were robust to a large variation of OTUs detections.

The representation of the virome with ecological networks offered a convenient means to connect the resource use of virus species with habitat variation (Blüthgen et al. 2008). The sampling scheme used was designed to address the environmental heterogeneity and to capture between- and within-habitat variability in plant and virus richness. Sub-sampling the plant samples to limit the frequency of rare species, which reduces the sample size of those sent for sequencing, potentially eliminates interactions However, the sample size needed to accurately represent higher-order processes in a network does not require an exhaustive sample of individual species (Hegland et al. 2010). Previous work showed two homogeneous clusters of plant species collections that corresponded to distinct assemblies of Crop and Edge habitats (Mcleish et al. 2021). Similarly, here we report that co-occurrence of the 104 fully selected virus OTUs in both habitats was infrequent: only fifteen OTUs were shared between habitats. We acknowledge that a proportion of true positives may be omitted by our virus detection procedure and that false positives may be retained, as discussed earlier. These potential drawbacks should be outweighed by the extensive sample size and data aggregation at fine- (i.e. replicate and repeat collections) to broad-scale experimental units; i.e. habitats of an ecosystem (Feest 2006). Uncertainty derived from non-exhaustive sampling and biases in the specificity and sensitivity of virus detection by HTS, is relaxed when inferring higher-order processes of communities (Hegland et al. 2010). Our sampling strategy emphasised a standardised comparative approach (Feest 2006) to draw conclusions about community-level interactions and differences among habitats.

RNA extracts of all plant species sent for HTS had virus sequences, in agreement with the frequent association of symbionts with hosts in other systems (Roossinck 2011; Fleming-Davies et al. 2015). Viruses are ubiquitous components in every ecosystem and life-form, and most virus–host interactions produce no disease symptoms (Roossinck and Bazán 2017), which is consistent with our observations where most plants were symptomless. Reports on viruses being antagonistic or mutualistic symbionts of plants according to environmental conditions (Gibbs 1980; Xu et al. 2008; Hily et al. 2016; Aguilar et al. 2020) and on virus evolution towards mutualism under conditions of abiotic stress for the host (González et al. 2021) suggest that mutualistic symbioses between plants and viruses are common and are important to host ecology and the evolution of both partners.

Our procedures were able to recover OTUs from all plant virus genome classes, in proportions that corresponded to expectations according to the frequency of the different genome classes among described plant virus species (King et al. 2012; Koonin et al. 2020). In addition, as the degree distribution of OTUs from the network that comprised all detections in Crop and Edge (Fig. 4) was right-skewed, it showed that most viruses had few interactions, which suggests that our virome is not biased towards viruses with a high prevalence in the ecosystem. Note that since we blast against virus databases, any bias in the databases will be reflected in our data set. It is noteworthy that our observations revealed plant–virus interactions unexpected from current knowledge on plant virus host ranges (e.g. VIDE data base, https://www.cabi.org/isc). For instance, virus OTUs not associated with cucurbit hosts (e.g. Anulavirus 1, Cucumovirus 3, Potyvirus 2, Potyvirus 3, Potyvirus 18, etc.) were found infecting *C. melo* (Table 2). We expect a high fraction of such unexpected associations to be real, as indicated by informatic and experimental procedures aimed at reducing false positive detections, but OTUs might represent several distinct species or several strains of the same species. Still, false positives may occur, and some associations based on a low number of reads may be spurious. Examples could include the infection of dicotyledon hosts by BYDV or of non-cucurbit hosts by CmEV. This should not be an issue for the purposes of comparing OTU virome richness between the habitats, as the comparative analyses of the full and the >4-read data sets show. Unexpected associations may be explained by the limited knowledge on plant virus host ranges,

which derives mostly from experimental inoculations or from field data on viruses causing diseases in crops (Gibbs et al. 2020). Most of the 'unexpected' infections in *C. melo* must be asymptomatic, as only 22.8 per cent ($n = 236$) of the sampled *C. melo* plants showed virosis-like symptoms, such as mosaics, lamina distortion, or growth reduction, and then go unnoticed in studies that focus on symptomatic plants. Interestingly, in addition to differences among habitats, virus communities also varied in composition at smaller scales among the sites of Crop (Supplementary Table S15 and Fig. S3).

Plant species richness at Edge sites E1 and E3 was more than triple that of Crop sites C1–C4, and virus OTUs richness in Edge was four times that of Crop (Supplementary Table S5), indicating a positive relationship between virus and plant richness that held at the level of site (Likelihood ratio $\chi^2_{(1)} = 77.82$, $P < 0.001$). Higher virus richness in Edge supports Hypothesis a that wild plant communities have more species-rich virus communities than plant communities associated with the crop. Many studies, mostly with animal hosts, have supported the positive correlation between host and parasite richness (Lafferty 2012; Kamiya et al. 2014a, 2014b), a relationship not yet analysed, to our knowledge, with plants and viruses. The lower virus OTU and plant richness in Crop than Edge might be related to the higher turnover of plant communities in Crop, which re-assemble after each fallow period. In addition, although 44 per cent of plant species in Crop are biannual or perennial, all of them are functionally annual, while in Edge, the fraction of biannual and perennials accounts for 56 per cent of relative abundance. However, the effects of such turnover on virus communities are not obvious. Plant community re-assembly need not necessarily translate into a reduction of virus richness, as subsequent plant assemblages at individual sites are each dependent on regional species pools. This would be expected to introduce a random effect on colonisation and could result in sinks for infection by different virus communities compared to previous ones. Drivers of parasite richness include other factors such as coinfection (Johnson et al. 2013), abiotic factors of climate, and habitat age and history of use (Mitchell et al. 2010; Preisser 2019; Song and Proctor 2020).

We estimated network modules as an approximation of ecological compartments. Compartmentalisation of interactions agreed with Hypothesis b and was consistent with other work (Hernandez et al. 2021). The modules estimated from Crop and Edge sites C1, C3, E1, and E3 (Fig. 4, Supplementary Fig. S9) comprised one main module with the majority of plant–virus interactions (seventy-four virus OTUs and thirty-three host species) and several smaller ones, each with less than four host species. The largest module comprised interactions that occurred at a level of organisation higher than the plant community (site) and across host species present in both habitats. In other words, Crop was connected by a few virus OTUs of Edge (i.e. Tobamovirus 3, Cucumovirus 1, Anulavirus 1, Sobemovirus 1, Cucumovirus 3, Tobamovirus 6, and Potyvirus 19) that had wide host ranges (Fig. 4). The networks from the Crop sites C1 and C3 showed two main modules. The first comprised mostly OTUs that specialised on *C. melo* (Supplementary Fig. S10). In the other module that did not include *C. melo*, several OTUs that infect *C. melo*, such as Cucumovirus 1 or Potyvirus 19, infected more than one weed species. At a higher level of organisation, five of the twenty-three OTUs of Crop sites C1 and C3 infected hosts from Edge. This module structure suggests that Edge is a possible inoculum reservoir for Crop and that weeds are possible inoculum sources for the crop. However, while Edge fulfils the definition of a community reservoir (Ashford 2003), as these communities are relatively undisturbed, Crop does not, as these are seasonal assemblies unlikely to maintain virus populations indefinitely (Supplementary Table S4). The observed modules were not organised according to the taxonomy of viruses or hosts but reflected the spatial organisation of plant–virus interactions between sites (Supplementary Fig. S3). Plant–virus interactions in Crop are weakly connected to those in Edge despite close spatial proximity between plant communities in these habitats but vary according to spatially discrete sites within a habitat. Therefore, local plant–virus interactions emerge from processes that propagate at both the ecosystem and local spatial and temporal scales (McLeish et al. 2019). The largest diversity of strategies for transmission and persistence occurred in Edge, and some of these strategies that enabled a relatively small number of virus OTUs to utilise both habitats might reflect traits that increase the likelihood of long-term persistence and of emergence in new hosts or communities.

In a seminal article, B. D. Harrison proposed that viruses that share features related to strategies of survival and dispersal should be associated with a particular type of plant community. He hypothesised that crop plant communities, but not wild communities, would favour the prevalence of viruses that spread by contact due to high host density—cultivated plant-adapted (CULPAD) viruses (Harrison 1981). Other groups of viruses, transmitted by vectors in persistent manners, referred to as wild plant-adapted (WILPAD) viruses, were hypothesised to infect crop plants incidentally. Other groups, including non-persistently vectored viruses, would infect crops and wild plant communities. Our observations mostly support these hypotheses. Virus OTUs of *Caulimoviridae, Closteroviridae, Geminiviridae, Rhabdoviridae, Solemoviridae*, and *Tombusviridae* with WILPAD traits were detected in Edge only, and the aphid non-persistently transmitted cucumoviruses and potyviruses were found in Edge and Crop, as predicted. However, a much larger diversity of the contact-transmitted tobamoviruses, with CULPAD traits, was found in Edge rather than in Crop. The co-occurrence of virus species within ecological compartments suggests coexistence, where persistence is indefinite for multiple species. Coexistence may result from niche differentiation (Leibold and McPeek 2006), which is consistent with the evolution of diverse strategies for survival and transmission within ecological compartments (Mauck and Chesnais 2020).

The networks of Edge sites E1 and E3 showed dynamic plant–virus interactions that varied according to season and site (Fig. 5). Both modularity and nestedness changed from spring through to autumn at both sites, which seemed to be partly independent of changes in the number and identity of virus and host species. Modules are expected to change when the number of species changes, which affects their relative abundance and preferences as hosts (Dormann and Strauss 2014). Changes to network topology through the seasons suggest changes that also affect the structure, function, and evolution of communities (Honjo et al. 2020; Segar et al. 2020). For example, in a study over 4 years, interactions of *Arabidopsis thaliana* with five viruses were highly dependent on site and season (Pagán et al. 2010). The fact both nestedness and modularity changed through the seasons suggests changes in specialised (modularity) as well as structured (nestedness) plant–virus interactions associated with resource availability. The coexistence of nestedness and modularity in networks has been explained as a consequence of processes that occur at different spatial scales (Valverde et al. 2020), or those relevant to ecological or evolutionary timescales (Cai et al. 2020). For example, a study that used a large database of temporally-resolved plant–pollinator networks showed that species turnover and phenology both modulated network nestedness and modularity at

broader temporal scales, such as across seasons (Schwarz et al. 2020). Our study differed from previous studies (Vázquez et al. 2009) as species richness, the number of interactions (k), modularity, and nestedness were more-or-less all positively correlated. This contrast with other studies might be explained by changes in the types of symbiosis over space and time, where mutualistic strategies may evolve from parasitic ones (Shapiro and Turner 2018), or where the ecological roles of virus strategies for survival are altered (Brum et al. 2016). Within compartments, the interaction of viruses with multiple potential hosts and with other viruses will necessarily affect virus evolution, condition their host range, and adaptation to the abiotic environment and the primary environment of the host.

The potential of hosts to function as reservoirs for infection has been viewed as a property of populations (Haydon et al. 2002) and species (Cronin et al. 2010) that vary in susceptibility and competence. Our observations support Hypothesis a that wild plants are reservoirs for viruses. Edge and its connectivity with Crop (Fig. 4) fits the concept of an ecological reservoir system with its component host populations in which an infectious agent survives indefinitely (Ashford 2003). The occurrence of virus OTUs that infect hosts in Edge and weeds of Crops suggests weeds have a role as a bridge reservoir (Viana et al. 2014; Caron et al. 2015) by providing a link through which viruses are transmitted from maintenance communities (Edge) to target hosts (*C. melo*). If the initial infections that occur early in the assembly of Crop communities are independent of Edge, they would potentially facilitate or inhibit subsequent infections by viruses transmitted from plant communities of Edge. The function of a habitat as a reservoir, therefore, might be sporadic as the structures of ecological compartments change within the system. Note that data on virus occurrence in Crop and in Edge at different seasons suggest inoculum fluxes occur both ways, as suggested by 5 OTUs occurring in Crop in summer and in Edge in autumn, after the crop is removed.

The results of this study are relevant to understanding how multi-host–multi-virus interactions in heterogeneous environments and host resource use may affect virus transmission dynamics. These dynamics are not static in space or time, and hence, the shifting resource space and constraints on virus–virus and plant–virus interactions shape virus diversity and evolution. Importantly, the practices used to conduct agriculture are the very processes that further complicate the prediction of virus dynamics and evolution. Theory and empirical evidence (Borrett 2013) show that only a small proportion of species or groups of species in a network are responsible for most of the behaviour of the (eco)system. Both agricultural ecosystems and less-disturbed ecosystems comprise communities that are exposed to the same natural forces, which all have characteristics and functions that are transitory in time and space (Whittaker 1962). Our findings support the hypothesis that the states of an agricultural ecosystem, like any other, that are conducive to emergent diseases will also be transitory and directly or indirectly associated with changes to a small fraction of species. Identifying those species that traverse ecological compartments and directly or indirectly influence the epidemiology of other agents in multiple assemblages will be key in unlocking better management schemes in agriculture.

Data availability

The datasets generated and/or analysed during the current study are available from the Dryad Digital Repository (available at, https://doi.org/10.5061/dryad.kd51c5b5v).

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Funding

This work was funded in part by Plan Estatal de I+D+i, Ministerio de Economía y Competividad, Spain (grants BFU2015-64018-R and RTI2018-094302-B-I00) and by ArimNet2 2015 joint Call (project 618127, EMERAMB). A.P. was supported by a Formacion de Personal Investigador contract (BES-2016-077810). A.Z. and B.B. were supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 813542T INEXTVIR. We thank Dr Rosario Gavilán, Departamento de Farmacología, Farmacognosia y Botánica, Facultad de Farmacia, Universidad Complutense, Madrid, Spain, for the taxonomic identification of plant species. We also acknowledge the contribution of detailed recommendations and time spent by three anonymous reviewers, which improved the quality of the manuscript.

Conflict of interest: None declared.

Author contributions

F.G.A. and A.F. developed and designed the project. M.M. and A.P. assisted with sample collection and identification. A.Z. and B.B. assisted with experimental data generation. A.Z., B.B., and A.P. assisted with descriptive statistics. M.M. conceived of, analysed, and wrote the manuscript, and all authors assisted with drafting the final manuscript. All authors have read and approved the manuscript.

References

Aguilar, E. et al. (2020) 'Virus Infection Induces Resistance to Pseudomonas Syringae and to Drought in Both Compatible and Incompatible Bacteria–host Interactions, Which Are Compromised under Conditions of Elevated Temperature and CO_2 Levels', *Journal of General Virology*, 101: 122–35.

Alcalá-Briseño, R. I. et al. (2019) 'Network Analysis of the Papaya Orchard Virome from Two Agroecological Regions of Chiapas, Mexico', *Msystems*, 5: e00423–19.

Almeida-Neto, M., and Ulrich, W. (2011) 'A Straightforward Computational Approach for Measuring Nestedness Using Quantitative Matrices', *Environmental Modelling and Software*, 26: 173–8.

Angly, F. E. et al. (2012) 'Grinder: A Versatile Amplicon and Shotgun Sequence Simulator', *Nucleic Acids Research*, 40: e94.

Aranda, M. A., and Freitas-Astúa, J. (2017) 'Ecology and Diversity of Plant Viruses, and Epidemiology of Plant Virus-induced Diseases', *Annals of Applied Biology*, 171: 1–4.

Ashford, R. W. (2003) 'When Is a Reservoir Not a Reservoir?', *Emerging Infectious Diseases*, 9: 1495–6.

Bass, D. et al. (2019) 'The Pathobiome in Animal and Plant Diseases', *Trends in Ecology & Evolution*, 34: 996–1008.

Bernardo, P. et al. (2018) 'Geometagenomics Illuminates the Impact of Agriculture on the Distribution and Prevalence of Plant Viruses at the Ecosystem Scale', *The ISME Journal*, 12: 173–84.

Blüthgen, N. et al. (2008) 'What Do Interaction Network Metrics Tell Us about Specialization and Biological Traits?', *Ecology*, 89: 3387–99.

Borrett, S. R. (2013) 'Throughflow Centrality Is a Global Indicator of the Functional Importance of Species in Ecosystems', *Ecological Indicators*, 32: 182–96.

Brum, J. R. et al. (2016) 'Seasonal Time Bombs: Dominant Temperate Viruses Affect Southern Ocean Microbial Dynamics', *The ISME Journal*, 10: 437–49.

Cai, W. et al. (2020) 'Mutualistic Networks Emerging from Adaptive Niche-based Interactions', *Nature Communications*, 11: 1–10.

Camacho, C. et al. (2009) 'BLAST+: Architecture and Applications', *BMC Bioinformatics*, 10: 1–9.

Candresse, T. et al. (2014) 'Appearances Can Be Deceptive: Revealing a Hidden Viral Infection with Deep Sequencing in a Plant Quarantine Context', *PLoS ONE*, 9: e102945.

Caron, A. et al. (2015) 'Bridge Hosts, a Missing Link for Disease Ecology in Multi-host Systems', *Veterinary Research*, 46: 1–11.

Chang, S., Puryear, J., and Cairney, J. (1993) 'A simple and efficient method for isolating RNA from pine trees', *Plant molecular biology reporter*, 11: 113–6.

Claverie, S. et al. (2018) 'From Spatial Metagenomics to Molecular Characterization of Plant Viruses: A Geminivirus Case Study', *Advances in Virus Research*, 101: 55–83.

Cronin, J. P. et al. (2010) 'Host Physiological Phenotype Explains Pathogen Reservoir Potential', *Ecology Letters*, 13: 1221–32.

Csardi, G., and Nepusz, T. (2006) 'The Igraph Software Package for Complex Network Research', *InterJournal, Complex Systems*, 1695: 1–9.

Dormann, C., Gruber, B., and Fründ, J. (2008) 'Introducing the Bipartite Package: Analysing Ecological Networks', *Interaction*, 1:0.2413793.

Dormann, C. F. and Strauss, R. (2014) 'A Method for Detecting Modules in Quantitative Bipartite Networks', *Methods in Ecology and Evolution*, 5: 90–8.

Dubay, B. (2017) *Developing Methodology for Elucidating Host Range Evolution in Plant RNA Viruses*. Universidad Politécnica de Madrid.

Expósito-Rodríguez, M. et al. (2008) 'Selection of Internal Control Genes for Quantitative Real-time RT-PCR Studies during Tomato Development Process', *BMC Plant Biology*, 12: 1–12.

Feest, A. (2006) 'Establishing Baseline Indices for the Quality of the Biodiversity of Restored Habitats Using a Standardized Sampling Process', *Restoration Ecology*, 14: 112–22.

Fleming-Davies, A. E. et al. (2015) 'Effects of Host Heterogeneity on Pathogen Diversity and Evolution', *Ecology Letters*, 18: 1252–61.

Fortuna, M. A. et al. (2010) 'Nestedness versus Modularity in Ecological Networks: Two Sides of the Same Coin?', *Journal of Animal Ecology*, 79: 811–7.

Gibbs, A. (1980) 'A Plant Virus that Partially Protects Its Wild Legume Host against Herbivores', *Intervirology*, 13: 42–7.

Gibbs, A. J. et al. (2020) 'The Potyviruses: An Evolutionary Synthesis Is Emerging', *Viruses*, 12: 132.

González, R. et al. (2021) 'Plant Virus Evolution under Strong Drought Conditions Results in a Transition from Parasitism to Mutualism', *Proceedings of the National Academy of Sciences of the United States of America*, 118: e2020990118.

Harrison, B. D. (1981) 'Plant Virus Ecology: Ingredients, Interactions and Environmental Influences', *Annals of Applied Biology*, 99: 195–209.

Haydon, D. T. et al. (2002) 'Emerging Infectious Diseases—Identifying Reservoirs of Infection a Conceptual and Practical Challenge', *Emerging Infectious Diseases*, 8: 1468–73.

Hegland, S. J. et al. (2010) 'How to Monitor Ecological Communities Cost-efficiently: The Example of Plant-pollinator Networks', *Biological Conservation*, 143: 2092–101.

Hernandez, D. J. et al. (2021) 'Environmental Stress Destabilizes Microbial Networks', *The ISME Journal*, 15: 1722–34.

Hily, J. M. et al. (2016) 'Environment and Host Genotype Determine the Outcome of a Plant-virus Interaction: From Antagonism to Mutualism', *New Phytologist*, 209: 812–22.

Honjo, M. N. et al. (2020) 'Seasonality of Interactions between a Plant Virus and Its Host during Persistent Infection in a Natural Environment', *The ISME Journal*, 14: 506–18.

Hsieh, T. C., Ma, K. H., and Chao, A. (2016) 'iNEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (Hill Numbers)', *Methods in Ecology and Evolution*, 7: 1451–6.

Johnson, P. T. J. et al. (2013) 'Host and Parasite Diversity Jointly Control Disease Risk in Complex Communities', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 16916–21.

Jones, B. A. et al. (2013) 'Zoonosis Emergence Linked to Agricultural Intensification and Environmental Change', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 8399–404.

Jones, R. A. C. (2014) 'Trends in Plant Virus Epidemiology: Opportunities from New or Improved Technologies', *Virus Research*, 186: 3–19.

Jones, R. A. C., and Naidu, R. A. (2019) 'Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives', *Annual Review of Virology*, 6: 387–409.

Kamitani, M. et al. (2019) 'A Survey on Plant Viruses in Natural Brassicaceae Communities Using RNA-Seq', *Microbial Ecology*, 78: 113–21.

Kamiya, T. et al. (2014a) 'Host Diversity Drives Parasite Diversity: Meta-analytical Insights into Patterns and Causal Mechanisms', *Ecography*, 37: 689–97.

——— et al. (2014b) 'What Determines Species Richness of Parasitic Organisms? A Meta-analysis across Animal, Plant and Fungal Hosts', *Biological Reviews*, 89: 123–34.

King, A. M. Q. et al. (2012) *Virus Taxonomy. Classification and Nomenclature of Viruses*. Academic Press: New York.

Koonin, E. V. et al. (2020) 'Global Organization and Proposed Megataxonomy of the Virus World', *Microbiology and Molecular Biology Reviews*, 84: 1–33.

Lacroix, C. et al. (2016) 'Methodological Guidelines for Accurate Detection of Viruses in Wild Plant Species', *Applied and Environmental Microbiology*, 82: 1966–75.

Lafferty, K. D. (2012) 'Biodiversity Loss Decreases Parasite Diversity: Theory and Patterns', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367: 2814–27.

Leibold, M. A., and McPeek, M. A. (2006) 'Coexistence of the Niche and Neutral Perspectives in Community Ecology', *Ecology*, 87: 1399–410.

Ma, Y. et al. (2021) 'Leaf-associated Fungal and Viral Communities of Wild Plant Populations Differ between Cultivated and Natural Ecosystems', *Plant-Environment Interactions*, 2: 1–13.

——— et al. (2020) 'Metagenomic Analysis of Virome Cross-talk between Cultivated Solanum Lycopersicum and Wild Solanum Nigrum', *Virology*, 540: 38–44.

——— et al. (2019) 'Phytovirome Analysis of Wild Plant Populations: Comparison of Double-Stranded RNA and Virion-Associated Nucleic Acid Metagenomic Approaches', *Journal of Virology*, 94: 1–14.

Maclot, F. et al. (2020) 'Illuminating an Ecological Blackbox: Using High Throughput Sequencing to Characterize the Plant Virome across Scales', *Frontiers in Microbiology*, 11: 1–16.

Malmstrom, C. M., and Alexander, H. M. (2016) 'Effects of Crop Viruses on Wild Plants', *Current Opinion in Virology*, 19: 30–6.

Malpica, J. M. et al. (2006) 'Association and Host Selectivity in Multihost Pathogens', *PLoS ONE*, 1: e41.

Martin, M. (2011) 'Cutadapt Removes Adapter Sequences from High-throughput Sequencing Reads', *EMBnet.journal*, 17: 10.

Massart, S. et al. (2017) 'A Framework for the Evaluation of Biosecurity, Commercial, Regulatory, and Scientific Impacts of Plant Viruses and Viroids Identified by NGS Technologies', *Frontiers in Microbiology*, 8: 45.

Mauck, K. E., and Chesnais, Q. (2020) 'A Synthesis of Virus-vector Associations Reveals Important Deficiencies in Studies on Host and Vector Manipulation by Plant Viruses', *Virus Research*, 285: 197957.

McLeish, M. et al. (2017) 'Scale Dependencies and Generalism in Host Use Shape Virus Prevalence', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20172066.

——— et al. (2019) 'Coinfection Organizes Epidemiological Networks of Viruses and Hosts and Reveals Hubs of Transmission', *Phytopathology*, 109: 1003–10.

McLeish, M. J., Fraile, A., and García-Arenal, F. (2020) 'Trends and Gaps in Forecasting Plant Virus Disease Risk', *Annals of Applied Biology*, 176: 102–8.

——— (2021) 'Population Genomics of Plant Viruses: The Ecology and Evolution of Virus Emergence', *Phytopathology*, 111: 32–9.

Mcleish, M. J. et al. (2021) 'Structuring of Plant Diversity across Agricultural Landscape Mosaics: The Importance of Community Connectivity and the Scale of Effect in Disease Ecology', *BMC Ecology and Evolution*, 21: 173.

Melcher, U., and Grover, V. (2011) 'Genomic Approaches to Discovery of Viral Species Diversity of Non-cultivated Plants'. In C. Caranta and M. A. Aranda (eds) *Recent Advances in Plant Virology*, pp. 321–42. Norfolk: Caister Academic Press.

Mitchell, C. E. et al. (2010) 'Controls on Pathogen Species Richness in Plants' Introduced and Native Ranges: Roles of Residence Time, Range Size and Host Traits', *Ecology Letters*, 13: 1525–35.

Miyauchi, A. and Kawase, Y. (2016) 'Z-score-based Modularity for Community Detection in Networks', *PLoS ONE*, 11: 1–17.

Muthukumar, V. et al. (2009) 'Non-cultivated Plants of the Tallgrass Prairie Preserve of Northeastern Oklahoma Frequently Contain Virus-like Sequences in Particulate Fractions', *Virus Research*, 141: 169–73.

Newman, M. E. J. (2006) 'Modularity and Community Structure in Networks', *Proceedings of the National Academy of Sciences, U S A*, 103: 8577–82.

Nicaise, V. (2014) 'Crop Immunity against Viruses: Outcomes and Future Challenges', *Frontiers in Plant Science*, 5: 1–18.

Oksanen, J. et al. 2020. Vegan: Community Ecology Package.

Paez-Espino, D. et al. (2016) 'Uncovering Earth's Virome', *Nature*, 536: 425–30.

Pagán, I. et al. (2010) '*Arabidopsis thaliana* as a Model for the Study of Plant-virus Co-evolution', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365: 1983–95.

——— et al. (2012) 'Effect of Biodiversity Changes in Disease Risk: Exploring Disease Emergence in a Plant-virus System', *PLoS Pathogens*, 8: 47.

Parratt, S. R., Numminen, E., and Laine, A. L. (2016) 'Infectious Disease Dynamics in Heterogeneous Landscapes', *Annual Review of Ecology, Evolution, and Systematics*, 47: 283–306.

Peng, Y. et al. (2012) 'IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth', *Bioinformatics*, 28: 1420–28.

Poos, M. S., and Jackson, D. A. (2012) 'Addressing the Removal of Rare Species in Multivariate Bioassessments: The Impact of Methodological Choices', *Ecological Indicators*, 18: 82–90.

Preisser, W. (2019) 'Latitudinal Gradients of Parasite Richness: A Review and New Insights from Helminths of Cricetid Rodents', *Ecography*, 42: 1315–30.

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna <https://www.R-project.org>.

Rodelo-Urrego, M. et al. (2013) 'Landscape Heterogeneity Shapes Host-parasite Interactions and Results in Apparent Plant-virus Codivergence', *Molecular Ecology*, 22: 2325–40.

Roossinck, M. J. (2005) 'Symbiosis versus Competition in Plant Virus Evolution', *Nature Reviews. Microbiology*, 3: 917–24.

——— (2011) 'The Good Viruses: Viral Mutualistic Symbioses', *Nature Reviews. Microbiology*, 9: 99–108.

Roossinck, M. J., and Bazán, E. R. (2017) 'Symbiosis: Viruses as Intimate Partners', *Annual Review of Virology*, 4: 123–39.

Roossinck, M. J., and García-Arenal, F. (2015) 'Ecosystem Simplification, Biodiversity Loss and Plant Virus Emergence', *Current Opinion in Virology*, 10: 56–62.

Roossinck, M. J. et al. (2010) 'Ecogenomics: Using Massively Parallel Pyrosequencing to Understand Virus Ecology', *Molecular Ecology*, 19: 81–8.

Sacristán, S., Fraile, A., and García-Arenal, F. (2004) 'Population Dynamics of Cucumber Mosaic Virus in Melon Crops and in Weeds in Central Spain', *Phytopathology*, 94: 992–8.

Schwarz, B. et al. (2020) 'Temporal Scale-dependence of Plant–pollinator Networks', *Oikos*, 129: 1289–302.

Segar, S. T. et al. (2020) 'The Role of Evolution in Shaping Ecological Networks', *Trends in Ecology & Evolution*, 35: 454–66.

Shapiro, J. W., and Turner, P. E. (2018) 'Evolution of Mutualism from Parasitism in Experimental Virus Populations', *Evolution*, 72: 707–12.

Song, Z., and Proctor, H. (2020) 'Parasite Prevalence in Intermediate Hosts Increases with Waterbody Age and Abundance of Final Hosts', *Oecologia*, 192: 311–21.

Souza, T. A. et al. (2020) 'A Temporal Diversity Analysis of Brazilian Begomoviruses in Tomato Reveals A Decrease in Species Richness between 2003 and 2016', *Frontiers in Plant Science*, 11: 1–13.

Stobbe, A. H., and Roossinck, M. J. (2014) 'Plant Virus Metagenomics: What We Know and Why We Need to Know More', *Frontiers in Plant Science*, 5: 1–4.

Stukenbrock, E. H., and McDonald, B. A. (2008) 'The Origins of Plant Pathogens in Agro-ecosystems', *Annual Review of Phytopathology*, 46: 75–100.

Susi, H. et al. (2019) 'Diverse and Variable Virus Communities in Wild Plant Populations Revealed by Metagenomic Tools', *PeerJ*, 7: e6140.

Susi, H., and Laine, A. L. (2021) 'Agricultural Land Use Disrupts Biodiversity Mediation of Virus Infections in Wild Plant Populations', *New Phytologist*, 230: 2447–58.

Titcomb, G. C., Jerde, C. L., and Young, H. S. (2019) 'High-throughput Sequencing for Understanding the Ecology of Emerging Infectious Diseases at the Wildlife-human Interface', *Frontiers in Ecology and Evolution*, 7: 1–7.

Tylianakis, J. M., Tscharntke, T., and Lewis, O. T. (2007) 'Habitat Modification Alters the Structure of Tropical Host-parasitoid Food Webs', *Nature*, 445: 202–5.

Valverde, S. et al. (2020) 'Coexistence of Nestedness and Modularity in Host–pathogen Infection Networks', *Nature Ecology and Evolution*, 4: 568–77.

Vázquez, D. P. et al. (2009) 'Uniting Pattern and Process in Plant-animal Mutualistic Networks: A Review', *Annals of Botany*, 103: 1445–57.

——— et al. (2005) 'Species Abundance and the Distribution of Specialization in Host–parasite Interaction Networks', *Journal of Animal Ecology*, 74: 946–55.

Vélez-Olmedo, J. B. et al. (2021) 'Low Virus Diversity and Spread in Wild *Capsicum* spp. Accessions from Ecuador under Natural Inoculum Pressure', *Archives of Virology*, 166: 1447–53.

Viana, M. et al. (2014) 'Assembling Evidence for Identifying Reservoirs of Infection', *Trends in Ecology & Evolution*, 29: 270–9.

Whittaker, R. H. (1962) 'Classification of Natural Communities', *The Botanical Review*, 28: 1–239.

Woolhouse, M. E. J., and Gowtage-Sequeria, S. (2005) 'Host Range and Emerging and Reemerging Pathogens', *Emerging Infectious Diseases*, 11: 1842–7.

Woolhouse, M. E. J. et al. (2002) 'Biological and Biomedical Implications of the Co-evolution of Pathogens and Their Hosts', *Nature Genetics*, 32: 569–77.

Wu, Q. et al. (2015) 'Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms', *Annual Review of Phytopathology*, 53: 425–44.

Xu, C. et al. (2017) 'Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing', *Journal of Virology*, 91: 1–14.

Xu, P. et al. (2008) 'Virus Infection Improves Drought Tolerance', *New Phytologist*, 180: 911–21.

Zhao, L. et al. (2018) 'Identifying Compartments in Ecological Networks Based on Energy Channels', *Ecology and Evolution*, 8: 309–18.