Direct and indirect viral associations predict coexistence in wild plant virus communities

Anna Norberg (✉ norberg.ami@gmail.com)

University of Zurich <https://orcid.org/0000-0002-3520-1043>

Hanna Susi

University of Helsinki <https://orcid.org/0000-0001-6814-9205>

Suvi Sallinen

University of Helsinki <https://orcid.org/0000-0001-7125-7186>

Pezhman Safdari

University of Helsinki

Nicholas Clark

University of Queensland <https://orcid.org/0000-0001-7131-3301>

Anna-Liisa Laine

University of Zurich

Article

Keywords: plant viruses, plant pathogens, community assembly, coexistence, co-occurrence networks, parasite coinfection, biotic interactions, Markov random fields, Conditional random fields, spatial analysis

Posted Date: July 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-741379/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title: Direct and indirect viral associations predict coexistence in wild plant virus communities

Running title: Viral associations predict community structure

Authors: Anna Norberg^{1,4*}, Hanna Susi^{2,4}, Suvi Sallinen ², Pezhman Safdari ², Nicholas J. Clark³ & Anna-Liisa Laine^{1,2}

¹ Department of Evolutionary Biology and Environmental Studies, University of Zürich, 8057 Zürich, Switzerland

² Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, PO BOX 65 00014 University of Helsinki, Finland

³ School of Veterinary Science, Faculty of Science, University of Queensland, Gatton, Queensland, 4343 Australia

⁴ These authors contributed equally: Anna Norberg, Hanna Susi

*Corresponding author

Email: A.N. norberg.ami@gmail.com, H.S. hanna.susi@helsinki.fi, S.S. suvi.sallinen@helsinki.fi, P.S. pezhman.safdari@helsinki.fi, N.J.C. nicholas.j.clark1214@gmail.com, A.-L.L. anna-liisa.laine@uzh.ch

Correspondence: Anna Norberg, norberg.ami@gmail.com, Winterthurerstrasse 190, 8057 Zurich, Switzerland, +41 44 635 6118

Keywords: plant viruses, plant pathogens, community assembly, coexistence, co-occurrence networks, parasite coinfection, biotic interactions, Markov random fields, Conditional random fields, spatial analysis

23   **Abstract**

24   Integration of community ecology with disease biology is viewed as a promising avenue for
25   uncovering determinants of pathogen diversity, and for predicting disease risks. Plant-
26   infecting viruses represent a vastly underestimated component of biodiversity with
27   potentially important ecological and evolutionary roles. We performed hierarchal spatial
28   analysis of wild plant populations to characterise the diversity and coexistence structure of
29   within-host virus communities, and their predictors. Our results show that these virus
30   communities are characterised by single infections of few, dominating virus taxa as well as
31   diverse, non-random coinfections. Using a novel graphical modelling framework we
32   demonstrate that after accounting for environmental heterogeneity at the level of both
33   individual host plants and populations, most virus co-occurrence patterns can be attributed to
34   virus-virus associations. Moreover, we show that conditioning variables changed virus
35   association networks especially through their indirect effects. This highlights a previously
36   underestimated mechanism of how human-driven environmental change can influence
37   disease risks.

38

39

40

41

42

43

44

45   **Introduction**

46   Viruses, like all organisms, occur as communities with other species¹. Importantly, the
47   diversity of co-occurring viruses may have profound impact on virus evolution and
48   epidemiology². Hence, understanding the conditions and scales at which virus communities
49   vary is critical when predicting how virus communities respond to environmental change or
50   to disease control measures³⁻⁶. Over the past decade metagenomic surveys exploring virus
51   diversity in wild hosts ranging from plants to insects and mammals⁷⁻¹¹ have revealed a
52   tremendous, largely undescribed, virus diversity across environments. It is becoming clear
53   that the diversity of virus taxa even within the same host can be highly variable, and that viral
54   co-occurrences are common in nature^{1,2,12,13}.

55   Predicting how pathogen communities respond to environmental change requires a
56   principled community ecology framework to disentangle possible drivers of coexistence
57   patterns^{14,15}. As Popovic *et al.*¹⁶ recently outlined, there are three main drivers of observed
58   species co-occurrence patterns (Figure 1). First, two species might share similar responses to
59   environmental variables such as temperature¹⁷⁻¹⁹. Second, two species may exhibit similar
60   responses to the occurrence of a third species: for example, some viruses require another
61   virus species to *mediate* their replication within a host^{20,21}, resulting in an indirect association
62   between the dependent species. Third, two species might exhibit a direct, biotic association: a
63   virus can for example facilitate the establishment of another^{22,23,22,23}. Hence, the sequence of
64   arrival²⁴, can have far reaching implications for the structure of pathogen communities²⁵.

65   Spatial structure and abiotic habitat conditions are considered important for structuring
66   pathogen communities²⁶ via their effect on both hosts and pathogens²⁷ (Figure 1). While
67   demographic stochasticity is important at smaller spatial scales²⁸, abiotic environmental

68 heterogeneity increases with increasing spatial scale, promoting greater coexistence through
69 species-specific environmental responses²⁹. Dispersal presumably decreases coexistence by
70 forcing species to interact and by homogenising intra- and interspecific interactions³⁰. Most
71 plant viruses are vector-dispersed, and thus the distribution of vectors can also influence the
72 structure of plant pathogen communities³¹. A special characteristic of pathogen communities
73 is that their immediate environment is a living organism in itself, and hence the influence of
74 the abiotic environment (e.g. weather) on pathogens can be either direct or mediated by the
75 host and/or vectors (Figure 1). Moreover, the distinction between abiotic and biotic effects
76 becomes blurred, as the interaction between a host and a pathogen can be considered both an
77 environmental effect as well as a biotic association (Figure 1). Pathogens can only occur
78 where they have susceptible hosts and thus spatial variation in resistance is expected to have
79 direct impacts on pathogen (co-)occurrence patterns³²⁻³⁴.

80 Ecological knowledge often relies on observational methods, but inferring signals of biotic
81 interactions from co-occurrence data is challenging³⁵. However, studying species associations
82 through the perspective of conditional probabilities helps to overcome some of these
83 challenges. Conditional probability refers to the probability that two species will be found
84 together *after* controlling for the other species in the network. Implementations for ecological
85 applications have been recently developed^{16,36,37}. *Markov random fields* (MRFs) are a group of
86 graphical network models which enable the estimation of conditional dependencies from
87 networks of interacting variable^{36,38}. In addition to analysing conditionally dependent species
88 co-occurrence patterns after accounting for the occurrences of all other species in the
89 community, also additional covariates can be included, resulting in *conditional random fields*
90 (CRFs)^{16,36,37,39}. MRFs and CRFs allow us to discover how species are associated with each

91 other and their environment, and importantly also how the environment influences these
92 associations⁴⁰.

93 Here, we aim to determine the relative roles of spatial environmental heterogeneity and both
94 direct and indirect virus-virus associations in determining the coexistence structure of
95 within-host virus communities of naturally occurring *Plantago lanceolata* populations,
96 identified via deep sequencing of small RNAs from field collected plant samples⁴¹. To
97 overcome the challenges of inferring signals of biotic interactions from co-occurrence data³⁵,
98 we incorporate relevant environmental covariates at the scale of both viruses and the hosts,
99 analyse co-occurrence patterns through conditional probabilities, and use a reasonably-sized
100 data set of 400 plants, sampled hierarchically at biologically relevant scales (across a
101 population network and within populations), covering a wide environmental range.
102 Specifically, we ask:

103 Q1. What are the relevant spatial scales of virus diversity and co-occurrences, and can we
104 detect signals of coexistence mechanisms influencing the relationship between viral
105 diversity and co-occurrence at different spatial scales?

106 Q2. Do host and habitat characteristics, and spatial structure of the host populations
107 influence virus community structure, or can we explain the structure solely using the
108 direct and indirect associations between the viruses?

109 Q3. After accounting for the effects of the host and habitat characteristics, and spatial
110 patterns on virus distributions, are there any remaining non-random negative or positive
111 direct associations between viruses?

112 Q4. Do associations between viruses change when comparing conditional network models
113 (CRFs) with *only* host- and/or habitat-related and/or spatial explanatory variables

included in the model, demonstrating how these different sources of environmental heterogeneity explain the structure of the virus community?

Our results demonstrate that there are non-random co-occurrence patterns between viruses, which are only partly resolved by the host and habitat characteristics and spatial structure. We find that the majority of the explained virus co-occurrence patterns can be attributed to direct and indirect associations among the viruses.

Material and Methods

STUDY SYSTEM AND SAMPLING

Our study focused on viruses infecting natural populations of *Plantago lanceolata* in the Åland Island, SW of Finland. These populations have been monitored since 1993, and they form a highly fragmented network of ~4000 populations⁴². We used a stratification process to select 20 focal populations that were sampled in early June 2017 for their virus communities. We collected leaf samples from 20 randomly selected plants in each population, resulting in altogether 400 samples for small RNA sequencing. For more details on the study system and sampling, see Supplementary Material and Methods.

VIRUS DETECTION

For virus identification, RNA was extracted from the samples using phenol-chloroform-isoamylalcohol extraction⁴³. The small RNA (sRNA) was sequenced using Illumina HiSeq (2009). After quality check and adaptor removal, the obtained reads were analysed using the VirusDetect pipeline⁴⁵ to obtain Operational Taxonomical Units (OTUs). The OTUs were classified according to their host range and plant associated OTUs were selected for our analysis⁴⁶. The OTUs were named at the virus family level and are henceforward referred to

137 as “viruses”. We compiled a presence-absence matrix describing the virus community in each
138 host individual and host population.

139

140 *EXPLANATORY VARIABLES*

141 *Host- and habitat-related variables*

142 During sampling we recorded the locations of the plants with GPS and measured the size of
143 the plants (Table 1). We recorded signs of herbivore damage, and surveyed surrounding
144 vascular plant communities between 1930 June 2017 by counting the number of vascular
145 plant species within multiple one-square-meter quadrats in each population. We used the
146 vegetation data to calculate the Shannon diversity index⁴⁷ for each population (Table 1).

147 We estimated the proportion of agricultural area surrounding each study population from
148 Corine Land Cover (CLC, version 2020_20u1) with QGIS⁴⁸ by creating a one-kilometre buffer
149 zone around each study population following the patch borders and calculating the
150 proportion of 20 m × 20 m pixels falling under agricultural land use category within this buffer
151 zone (Table 1). We also estimated the coverage of *P. lanceolata* foliage in square meters in
152 each population. We quantified the connectivity of a host population with respect to other
153 populations by calculating the Euclidian distances between populations and calibrating these
154 measures by the species dispersal capacity⁴⁹.

155 We obtained weather observations for the study populations from the Finnish Meteorological
156 Institute⁵⁰. We calculated the number of severe winter days during the winter before the
157 sampling season (2016-2017), and the sum of temperatures of the effective summer days
158 during previous summer (2016). See Table 1 and Supplementary Material and Methods for
159 more details, including biological justification for the selected variables.

160 *Spatial variables*

161 We included spatial variables implemented as Moran's eigenvector maps (MEMs)⁵¹. We
162 calculated MEMs based on the GPS coordinates of the sampled host plants, with the
163 assumption of positive autocorrelation. We estimated the significance of individual
164 eigenvectors with a permutation test for the Moran's I statistic and included all the
165 eigenvectors with Moran's I > 0.7, as well as the last significant one, thus representing both
166 coarse and fine scale signal of significant spatial autocorrelation. We used the R environment
167 ⁵² and packages 'spdep' and 'adespatial'.

168

169 *STATISTICAL METHODS*

170 *Descriptive analyses*

171 We began with descriptive analyses and illustration of the virus co-occurrence structure. We
172 assessed the nestedness of the data by organising the virus community matrix based on
173 overlap in virus presences among plants and decreasing fill⁵³. We calculated C-scores⁵⁴ for the
174 virus community at the level of individual host plants and populations (Figure 2). We
175 calculated the numbers of pairwise virus co-occurrence combinations in host plants, as well
176 as the pairs of viruses that never co-occur. We used the full dataset of 25 viruses (See Results)
177 for these descriptive analyses.

178 We quantified the relationship between the cumulative virus richness and co-occurrence
179 patterns with respect to increasing area sampled by calculating the mean species–area–curve
180 and mean coexistence–curve⁵⁵. Both curves were constructed by randomly selecting one
181 sampled host plant, and then increasing the spatial scale and sample by including the next
182 closest plant to the species richness or co-occurrence calculation. We calculated the mean

183 curves by repeating this 100 times, selecting a different initial plant for each round. We also
184 calculated the maximum number of co-occurring virus pairs from the number of taxa
185 observed thus far. For more details, see Supplementary Material and Methods and
186 Supplementary Results Figure S2-3.

187 *Markov Random Field networks*

188 To understand the role of host- and habitat-related variables and spatial configuration of the
189 hosts, as well as virus interactions (Figure 1, Table 1) in explaining the network structure, we
190 fitted both Unconditional (MRF) and Conditional Markov Random Field models (CRF) to the
191 virus community data³⁶. Markov Random Fields (MRFs) are graphical models, which can
192 represent complex distributions as network graphs. These networks consist of *nodes* and
193 *edges*, corresponding to the observed variables within the data, and to the probabilistic
194 interactions between variables that need to be estimated. The edge associations are
195 undirected, meaning that the effect of one node on another is reciprocal. If there is no edge
196 between two nodes in the estimated graph, these nodes are conditionally independent from
197 one another, whereas if there is an edge, these nodes are conditionally dependent, *after*
198 accounting for the other node effects in the graph model⁵⁶. Conditional Random Fields allow
199 for these dependencies among nodes to be further conditional on other covariates^{36,56}. Hence,
200 the values for the edge associations can change in the presence of these covariates, and the
201 resulting graph model illustrates the pairwise associations between viruses in host plants,
202 conditional not only on the rest of the virus community, but also on the covariates included in
203 the model (Table 1).

204 The modelling framework is described in detail by Clark *et al.*³⁶. Briefly, we modelled the log-
205 odds of detecting virus *i* given covariate *x* and occurrence of virus *j* with

$$\log\left(\frac{P(y_i = 1|y_{\setminus i}, x)}{1 - P(y_i = 1|y_{\setminus i}, x)}\right) = \alpha_{i0} + \beta_i^T x + \sum_{j:j \neq i} (\alpha_{ij0} + \beta_{ij}^T x) y_j ,$$

where y_i is the vector of presences and absences of virus i ; $y_{\setminus i}$ denotes the presences and absences of all other viruses except i ; α_{i0} is the virus-level intercept; and β_i^T is the effect of covariate x on the occurrence probability of virus i . Parameters α_{ij0} and β_{ij}^T represent the associations between viruses, conditional on the occurrences of all the non-focal viruses (other than the focal virus i).

We used data on viruses with at least 10 occurrences (16 viruses) in the entire virus community matrix (i.e. minimum prevalence of 2.5% of sampling units). To understand how environmental characteristics and the host affect the virus community, we included several explanatory, conditioning variables in the model (Table 1), describing: 1) The level of *spatial autocorrelation* of the host populations (implemented as Moran's eigenvectors) 2) *habitat-related* characteristics, namely the *quality* of the habitat of the host plants (the connectedness (S) of the focal *P. lanceolata* population to other populations, agricultural land use (percentage of the surrounding landscape) and the Shannon diversity of the local plant community, which have been demonstrated to influence virus occurrences in this system⁵⁷, and as the *weather* conditions of local populations (severity of the previous winter and temperature sum over the effective summer days during previous summer); as well as 4) *host-related* characteristics of the focal host plants (host population size, host plant size and signs of herbivory). See Table 1 for full details.

Altogether our dataset used for modelling consisted of 16 virus taxa and 16 explanatory variables (Table 1), resulting in 272 coefficients in each regression. To avoid overfitting, regularisation has been implemented in the method through least absolute shrinkage and selection operator (LASSO), forcing some regression coefficients to zero, and thus performing

229 variable selection and reducing the risk of overfitting³⁶. To achieve an undirected network
 230 and symmetry within the coefficients of conditional dependence (i.e. $\alpha_{ij0} = \alpha_{ji0}$ and $\beta_{ij}^T =$
 231 β_{ji}^T) we take the mean of the corresponding estimates, which is the default setting of the
 232 applied algorithm³⁶.

233 We fitted six model variants in total: 1) an Unconditional Markov Random Field model
 234 (referred to as 'MRF'), with only virus occurrences included, 2) a Conditional Markov Random
 235 Field model (CRF) with only habitat- and host-related (collectively referred as
 236 'environmental', see Table 1) variables included as additional constraints ('CRFenv'), 3) a CRF
 237 model with only host-related variables included as additional constraints ('CRFhost'), 4) a
 238 CRF model with only spatial variables and variables related to habitat (quality and weather)
 239 included as additional constraints ('CRFhabitat'), 5) a CRF model with only spatial variables
 240 included as additional constraints ('CRFspat'), 6) a Conditional Markov Random Field model
 241 with both all environmental as well as spatial variables included as additional constraints
 242 ('CRFfull'). We will refer to the variants (2-6) collectively as 'CRF models' or 'CRFs'.

243 We evaluated the model fit by calculating the Area Under Curve values⁵⁸ using the full data
 244 set. We used four-fold cross validation to estimate model generality by comparing predicted
 245 and observed outcomes simultaneously for all taxa. To account for parameter uncertainty of
 246 the final model, we modelled 100 bootstrapped replicates for the model. If the 90%
 247 confidence interval of bootstrapped coefficients did not overlap with zero, we considered the
 248 variable to have a statistically significant effect. To test whether the viral associations were
 249 phylogenetically conservative, we compared the direct associations drawn from all our
 250 network models to the phylogenetic relationships of the viruses, constructed from taxonomy,
 251 by conducting a Mantel test between the matrices.

252 All results were produced with R (version 4.0.2)⁵², and packages ‘vegan’⁵⁹, ‘MFRcov’³⁶,
253 ‘igraph’⁶⁰, along with their dependencies. An R package called ‘meta17-network’ including the
254 analytical pipeline, data, and documentation for full reproduction of the results can be found
255 in Github (aminorberg/meta17network-pkg).

256
257 Results

258 *Description of the data*

259 The plant small RNA sequencing yielded on average 18 222 557 reads (min 11148864, max
260 63 508 829 and sd 3 944 297) per sample. VirusDetect assembled 512 908 contigs in total
261 (min 39 nt, mean 63 nt, max 5408 nt and sd 45 nt). There were 5504 contigs (min 40 nt, mean
262 90 nt, max 1232 nt) with hits to known virus taxa⁶¹ with average sequence similarity of 82%
263 (min 23%, max 100%). Our analyses are focused on the 25 identified plant-associated viruses.
264 Out of the 400 sampled host plants, four samples were discarded due to missing explanatory
265 variable data. Thus, we had 396 sampled plants, resulting in 9 900 unique virus observations.
266 The prevalence of the individual viruses varied from 0.5 to 36% in the whole data. Of the 396
267 plants, 29% were uninfected, 32% hosted a single infection, and 39% of the plants hosted
268 multiple infections, of which 7% consisted of five or more viruses (Figure 2A).

269 *Descriptive analyses*

270 The virus communities exhibit a significantly nested structure: the mean C-score over the
271 viruses was 290.52 at the individual plant level (with the maximum of 87155 possible
272 checkerboard units); and 2.97 at the population level (with the maximum of 890 possible
273 checkerboard units) (Figure 2B). Almost all viruses (24 OTUs) were present in the most
274 virus-rich host population (population 861; Figure 2B and 2D), while other populations

275 hosted smaller subset of the viruses, with the simplest metacommunity consisting of six
276 viruses (population 3222; Figure 2B and 2C).

277 All the viruses co-occurred with another virus at least once (Figure 3A). The most abundant
278 viruses, Closteroviridae and Caulimoviridae, occurred as single infections in 39 and 33 host
279 plants, respectively, which made up nearly one third of their total occurrences (133 and 122,
280 respectively). The number of unique virus co-occurrence combinations (of two viruses or
281 more) in the whole data set was 111. Only Avsunoviroideae and Alphasatellitidae never co-
282 occurred (of which the latter was so rare that it was not included in the network models, see
283 below).

284 The species-area curve shows that all 25 viruses have been encountered when on average 264
285 plants have been sampled (Figure 3B). The difference between the species-area and
286 coexistence-area curves shows how the cumulative sum of co-occurring viruses is lower than
287 the cumulative viral richness would predict, until the coexistence curve reaches its maximum
288 at 387 sampled host plants. For example, when 100 host plants have been sampled, the
289 number of viruses observed was 20, which would enable 190 co-occurring pairs of viruses
290 while the average number of observed co-occurring virus pairs at that point was 168.

291 *Markov Random Field networks*

292 The CRF models (with additional conditioning variables) clearly outperformed the MRF (with
293 only virus occurrences included): the AUC for MRF was 0.69 while for the CRFs it varied
294 between 0.87 and 0.89. Based on cross-validation, there were no pronounced discrepancies
295 between the different CRFs, but the overall performance of the CRFs was better than that of
296 the MRF model: the 50% quantile for the mean for predicting both true positives and
297 negatives correctly for the MRF was 0.76, whereas the corresponding value for the CRF
298 variants was around 0.91. The MRF predicted more false positives, whereas the CRFs

299 predicted more false negatives. The mean values for different performance measures are
300 reported in the Supplementary Results Table S1.

301 To understand the changes in the network resulting from the addition of conditioning
302 variables, we compared the virus-virus-associations between viruses based on the MRF and
303 the different CRF variants. The MRF revealed mostly positive associations between the
304 viruses (Figure 4A). After including spatial, habitat and host-related variables (Table 1), some
305 of the associations between the viruses diminished or disappeared, and all of the conditional
306 associations were positive (Figure 4B). The number of significant virus co-occurrences
307 captured by the MRF model was 50 (Figure 4A). The corresponding number for the CRFfull
308 model was 16 (Figure 4B). The CRFs incorporating subsets of conditioning variables
309 identified intermediate amounts of associations: 30 for CRFhost, 38 for CRFspat, 28 for
310 CRFhabitat, and 18 for CRFenv.

311 Although several associations could be explained exchangeably with habitat- or host-related
312 variables, many associations were also explained solely by either habitat- or host-related
313 variables (Figure 4C-D). For example, Bromoviridae showed a high number of associations
314 with other viruses (Figure 4A), but was not explained by host-related or spatial variables
315 (Figure 4C and E). However, several of its strong associations with other viruses were
316 explained by the habitat-related effects (Figure 4D). The 11 association links captured by the
317 CRFfull model were captured by all the other conditional and unconditional model variants as
318 well (Figure 4B). We will refer to these associations as 'permanent'. In this network,
319 Bromoviridae and especially Secoviridae appeared as hubs, with five association links to
320 other viruses. These permanent associations represented direct interactions between viruses
321 that could not be explained with indirect effects of the rest of the virus community nor any
322 combination of additional conditioning variables.

323 Next, to understand how host- and habitat-related variables and spatial configuration of the
324 hosts influence virus community structure, we investigated the direct effects of the additional
325 conditioning variables. All the significant direct effects of the environmental and spatial
326 variables were for either Caulimoviridae or Geminiviridae (Table 2): e.g., increasing
327 agricultural land use in the surrounding landscape increased the occurrence probability of
328 Caulimoviridae, and host population size predicted higher occurrence probability for
329 Geminiviridae.

330 None of the indirect effects of the additional conditioning variables influenced the
331 associations between viruses so that the direction of the direct virus-virus association would
332 change from positive to negative or vice versa. However, all conditioning variables except host
333 plant size and agricultural land use had some effect(s) on some viral associations (Table 2). In
334 terms of the number of virus-virus-associations influenced, the most influential indirect
335 effects were the spatial structure of the host populations (MEMs) and host population
336 connectivity. All the effects of the first, coarse scale spatial variable (MEM1) were positive,
337 whereas the effects of the spatial variables at increasingly finer scales (MEM2-4) were all
338 negative. Increasing connectivity of the host population had both negative and positive effects
339 on the virus-virus associations: for example, higher connectivity lowered the occurrence
340 probability of Avsunoviridae in the presence of Bromoviridae (and vice versa, symmetrically).
341 There were altogether eight significant herbivory-related effect, all of them positive (Table 2).

342

343 **Discussion**

344 Here, we use community modelling to uncover determinants of wild plant virus diversity, a
345 vastly unexplored component of biodiversity⁶². We show that virus communities exhibit a
346 clear nested structure with abundant co-occurrences, both pairwise and higher-order. The

347 observed coexistence patterns are mediated by host plant characteristics, abiotic
348 environment and spatial structure, as well as associations among the viruses. Many present-
349 day threats posed by infectious diseases involve interactions that are manifested across
350 nested scales of biological organisation¹⁴, and our findings shed light on how plant virus
351 coexistence is maintained at different spatial scales.

352 The simplest within-host viral communities detected in our data consisted of single infections,
353 but nearly half (46%) of the infected plants hosted multiple infections. Single infections were
354 typically observed only for the few most common viruses, and half of all the viruses never
355 occurred as single infections. Our findings are in line with other studies that have found high
356 levels of coinfection^{63–65}. However, our approach enabled the detection of virus communities
357 in a substantial number of host individuals, with capacity to detect viromes consisting of up to
358 24 distinct taxa. This suggests that virus communities are highly variable within and among
359 hosts.

360 From the difference between species-area and coexistence curves, we see that a larger
361 number of hosts is needed to maintain the coexistence patterns that could be derived from the
362 overall virus richness. Indeed, increasing environmental heterogeneity and varying responses
363 of viruses to this heterogeneity promote coexistence by reducing competition, as predicted by
364 the classic species sorting paradigm, which has been shown to be influential for microbial
365 communities (see e.g. Székely & Langenheder 2014 and refs. within). Combined with the
366 weakening effects of demographic stochasticity and the homogenising effects of dispersal,
367 these mechanisms lead to more stable coexistence at increasing spatial scales^{55,67}.

368 We found that the conditional network models (CRFs), outperformed the model incorporating
369 only associations between viruses (MRF) in explaining virus community structure. The
370 differences between the CRF model variants were less pronounced, as seen from their equal

371 model fit and performance. Hence, we conclude that there is environmental variation and
372 processes operating at different spatial scales that influence the wild virus community
373 coexistence structure both directly and indirectly. The environmental characteristics and the
374 spatial variables explain community structure almost interchangeably. Previously, e.g. within-
375 host diversity of pathogens has been shown to increase with latitude, while pathogen
376 turnover follows an opposite trajectory, suggesting limited transmission in lower latitudes⁶⁸.

377 Our analysis revealed spatial variables to influence associations between viruses, resulting in
378 indirect effects on their co-occurrence. These viruses are vector-dispersed, and most likely
379 dispersal-limited at larger spatial scales⁶⁹, leading to less homogenisation at these scales.
380 However, at the scale of a host populations, variation among host genotypes⁷⁰ and
381 demographic stochasticity (through e.g. herbivory, as seen from the indirect effects of
382 herbivore damage on the associations between viruses) are expected to be important, as well
383 as abiotic interactions mediating coexistence ⁷¹.

384 After accounting for the effects of the host and habitat characteristics, as well as spatial
385 structure on virus distributions, non-random, significant positive associations between
386 viruses remain. Although these associations are based on purely observational co-occurrence
387 patterns, given the spatial scale of our sampling and our method for detection that targets the
388 host's defence response⁴¹, we consider these indicative of biotic interactions between the
389 viruses within a host plant⁷². Positive associations between species along with microhabitat
390 preferences have been suggested to promote the (co)existence of rare species⁷³. The
391 associations among coinfecting viruses may involve positive effects on replication^{74,75}, or even
392 obligate dependencies as some viruses require their specific helper virus in order to complete
393 their lifecycle⁷⁶. Viruses may also suppress host immunity allowing subsequent infections to
394 escape recognition by host immunity⁷⁷. As expected, the inclusion of explanatory variables

395 reduced the number of direct associations between viruses identified by the models,
396 suggesting that shared environmental responses play an important role in the assembly of the
397 communities (see e.g. ^{78,79}). After accounting for host- or habitat-related or spatial variables,
398 support for several direct associations (e.g. Tombusviridae-Alphaflexiviridae,
399 Betaflexiviridae- Potyviridae and Tombusviridae-Betaflexiviridae) was no longer detected.

400 Importantly, we found that the inclusion of the additional conditioning variables changed the
401 association networks especially through their indirect effects. For example, we found a
402 significant indirect effect of the coarsest-scale spatial variable (MEM1) on the association
403 between Fimoviridae and Alphaflexiviridae, while the direct effect of this spatial variable was
404 not significant. Hence, the occurrence probability of Alphaflexiviridae is jointly affected by the
405 occurrence of Fimoviridae as well as spatial structure. Due to symmetry, Fimoviridae is
406 similarly affected jointly by the occurrence of Alphaflexiviridae and spatial structure. These
407 indirect significant effects were more frequent (in total 32 effects) in comparison to direct
408 effects (in total 10). The indirect links can be indicative of either biotic interactions between
409 viruses only manifesting under certain environmental conditions, or these environmental
410 conditions having an effect only in interaction with the other virus⁷¹. Such indirect effects
411 have traditionally been challenging to detect, yet not accounting for them can lead to over- or
412 underestimation of signals of biotic interactions in co-occurrence data³⁵. Assessing species
413 network structure at different spatial scales, along with other environmental variables, can
414 provide insight to our understanding of community stability and functioning in relation to the
415 different processes that are at play on each spatial scale⁸⁰.

416 Although our virus community data set along with its environmental explanatory variables is
417 extensive and of high quality, our results are limited by our sample size and its effects on the
418 parameter estimation of our modelling method. We use regularisation to avoid overfitting, but

we note that the estimated parameters could in theory change with more data included. However, as seen from the species-area and coexistence-area curves, our sampling effort captures the detected virus diversity already before all the samples have been included, indicating a promising sample size.

Our results demonstrate that natural plant virus communities are characterised by single infections of few, dominating virus taxa as well as diverse, non-random coinfections. Virus diversity can be explained by coexistence-promoting mechanisms, some of which we could tease apart with our modelling. We show that host and habitat characteristics, as well as spatial structure, resolve some of the observed co-occurrence patterns, to some degree interchangeably. Importantly, we find that some virus-virus associations are mediated by either host or habitat characteristics, or the spatial structure of the host populations. However, a substantial part of the explained virus co-occurrence patterns can be attributed to positive, direct associations among the viruses. Moreover, we show that additional conditioning variables changed virus association networks especially through their indirect effects. Thus, our study contributes to increasing understanding on how plant virus coexistence and thus biodiversity is maintained at different spatial scales. Our results highlight a previously underestimated mechanism of how human-driven environmental change can influence disease risks by changing biotic associations between viruses that are conditional on their environment.

438

439 **References**

- 440 1. Díaz-Muñoz, S. L. Viral coinfection is shaped by host ecology and virus-virus
441 interactions across diverse microbial taxa and environments. *Virus Evol.* **3**, 1–14 (2017).
- 442 2. Alcaide, C., Rabadán, M. P., Moreno-Pérez, M. G. & Gómez, P. Implications of mixed viral
443 infections on plant disease ecology and evolution. *Adv. Virus Res.* **106**, 145–169 (2020).

- 444 3. Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W. & Holt, R. D. A framework for
445 community interactions under climate change. *Trends Ecol. Evol.* **25**, 325–331 (2010).
- 446 4. Malmstrom, C. M., Melcher, U. & Bosque-Pérez, N. A. The expanding field of plant virus
447 ecology: Historical foundations, knowledge gaps, and research directions. *Virus Res.*
448 **159**, 84–94 (2011).
- 449 5. Massart, S. *et al.* A framework for the evaluation of biosecurity, commercial, regulatory,
450 and scientific impacts of plant viruses and viroids identified by NGS technologies. *Front.*
451 *Microbiol.* **8**, (2017).
- 452 6. Halliday, F. W., Rohr, J. R. & Laine, A. L. Biodiversity loss underlies the dilution effect of
453 biodiversity. *Ecol. Lett.* **23**, 1611–1622 (2020).
- 454 7. Wren, J. D. *et al.* Plant virus biodiversity and ecology. *PLoS Biol.* **4**, 0314–0315 (2006).
- 455 8. Roossinck, M. J. Lifestyles of plant viruses. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1899–
456 1905 (2010).
- 457 9. Ng, T. F. F. *et al.* Exploring the diversity of plant DNA viruses and their satellites using
458 vector-enabled metagenomics on whiteflies. *PLoS One* **6**, (2011).
- 459 10. Roossinck, M. J. Plant Virus Metagenomics: Biodiversity and Ecology. *Annu. Rev. Genet.*
460 **46**, 359–369 (2012).
- 461 11. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus
462 diversity, spillover and emergence. *Nat. Rev. Microbiol.* **18**, 461–471 (2020).
- 463 12. Munson-Mcgee, J. H. *et al.* A virus or more in (nearly) every cell: Ubiquitous networks of
464 virus-host interactions in extreme environments. *ISME J.* **12**, 1706–1714 (2018).
- 465 13. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus–host
466 interactions resolved from publicly available microbial genomes. *Elife* **4**, 1–20 (2015).
- 467 14. Johnson, P. T. J., de Roode, J. C. & Fenton, A. Why infectious disease research needs
468 community ecology. *Science (80-.).* **349**, 1259504 (2015).
- 469 15. Seabloom, E. W. *et al.* The community ecology of pathogens: Coinfection, coexistence
470 and community composition. *Ecol. Lett.* **18**, 401–415 (2015).
- 471 16. Popovic, G. C., Warton, D. I., Thomson, F. J., Hui, F. K. C. & Moles, A. T. Untangling direct
472 species associations from indirect mediator species effects with graphical models.
473 *Methods Ecol. Evol.* **10**, 1571–1583 (2019).
- 474 17. Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E. & Mittler, R. Abiotic and biotic stress
475 combinations. *New Phytol.* **203**, 32–43 (2014).
- 476 18. Obrepalska-Stęplowska, A. *et al.* Effect of temperature on the pathogenesis,

- 477 accumulation of viral and satellite RNAs and on plant proteome in peanut stunt virus
478 and satellite RNA-infected plants. *Front. Plant Sci.* **6**, 1–14 (2015).
- 479 19. Alcaide, C., Sardanyés, J., Elena, S. F. & Gómez, P. Increasing temperature alters the
480 within-host competition of viral strains and influences virus genetic variability. *Virus*
481 *Evol.* 2020.07.06.190173 (2021) doi:10.1093/ve/veab017.
- 482 20. Syller, J. Facilitative and antagonistic interactions between plant viruses in mixed
483 infections. *Mol. Plant Pathol.* **13**, 204–216 (2012).
- 484 21. Pirone, T. P. & Blanc, S. Helper-Dependent Vector Transmission of Plant Viruses. *Annu.*
485 *Rev. Phytopathol.* **34**, 227–247 (1996).
- 486 22. Six, E. W. & Klug, C. A. C. Bacteriophage P4: a satellite virus depending on a helper such
487 as prophage P2. *Virology* **51**, 327–344 (1973).
- 488 23. Waterhouse, P. M. & Murrant, A. F. Further evidence on the nature of the dependence of
489 carrot mottle virus on carrot red leaf virus for transmission by aphids. *Ann. Appl. Biol.*
490 **103**, 455–464 (1983).
- 491 24. Fukami, T. Historical contingency in community assembly: integrating niches, species
492 pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* **46**, 1–23 (2015).
- 493 25. Karvonen, A., Jokela, J. & Laine, A.-L. Importance of Sequence and Timing in Parasite
494 Coinfections. *Trends Parasitol.* **35**, 109–118 (2019).
- 495 26. Bergner, L. M. *et al.* Demographic and environmental drivers of metagenomic viral
496 diversity in vampire bats. *Mol. Ecol.* **29**, 26–39 (2020).
- 497 27. Makiola, A. *et al.* Land use is a determinant of plant pathogen alpha- but not beta-
498 diversity. *Mol. Ecol.* **28**, 3786–3798 (2019).
- 499 28. Tilman, D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of
500 resource competition, invasion, and community assembly. *Proc. Natl. Acad. Sci. U. S. A.*
501 **101**, 10854–61 (2004).
- 502 29. Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary*
503 *Approaches*. (University of Chicago Press, 2003).
- 504 30. Snyder, R. E. & Chesson, P. Local dispersal can facilitate coexistence in the presence of
505 permanent spatial heterogeneity. *Ecol. Lett.* **6**, 301–309 (2003).
- 506 31. Schröder, M. L., Glinwood, R., Ignell, R. & Krüger, K. The role of visual and olfactory plant
507 cues in aphid behaviour and the development of non-persistent virus management
508 strategies. *Arthropod. Plant. Interact.* **11**, 1–13 (2017).
- 509 32. Jousimo, J. *et al.* Ecological and evolutionary effects of fragmentation on infectious

disease dynamics. *Science* (80-.). **344**, 1289–1293 (2014).

33. Carlsson-Granér, U. & Thrall, P. H. Host resistance and pathogen infectivity in host populations with varying connectivity. *Evolution* (N. Y). **69**, 926–938 (2015).

34. Paplauskas, S., Brand, J. & Auld, S. K. J. R. Ecology directs host–parasite coevolutionary trajectories across *Daphnia*–microparasite populations. *Nat. Ecol. Evol.* (2021) doi:10.1038/s41559-021-01390-7.

35. Blanchet, F. G., Gazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* (2020) doi:10.1111/ele.13525.

36. Clark, N. J., Wells, K. & Lindberg, O. Unravelling changing interspecific interactions across environmental gradients using Markov random fields. *Ecology* **99**, 1277–1283 (2018).

37. Harris, D. J. Inferring species interactions from co-occurrence data with Markov networks. *Ecology* **97**, 3308–3314 (2016).

38. Sutton, C. & McCallum, A. An introduction to conditional random fields. *Found. Trends Mach. Learn.* **4**, 267–373 (2011).

39. Azaele, S., Muneeppeerakul, R., Rinaldo, A. & Rodriguez-Iturbe, I. Inferring plant ecosystem organization from species occurrences. *J. Theor. Biol.* **262**, 323–329 (2010).

40. Clark, N. J., Kerry, J. T. & Fraser, C. I. Rapid winter warming could disrupt coastal marine fish community structure. *Nat. Clim. Chang.* **2040**, (2020).

41. Kreuze, J. F. *et al.* Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* **388**, 1–7 (2009).

42. Ojanen, S. P., Nieminen, M., Meyke, E., Pöyry, J. & Hanski, I. Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): Survey methods, data management, and long-term population trends. *Ecol. Evol.* **3**, 3713–3737 (2013).

43. Chang, S., Puryear, J. & Cairney, J. A Simple and Efficient Method for Isolating RNA from Pine Trees. *Plant Mol. Biol. Report.* **11**, 113–116 (1993).

44. Cuellar, W. J. *et al.* Elimination of antiviral defense by viral RNase III. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10354–10358 (2009).

45. Zheng, Y. *et al.* VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* **500**, 130–138 (2017).

46. Hulo, C. *et al.* ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**, 576–582 (2011).

- 543 47. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 623–656
544 (1948).
- 545 48. QGIS Development Team. QGIS Geographic Information System. (2019).
- 546 49. Hanski, I. *Metapopulation Ecology. Oxford Series in Ecology and Evolution* (Oxford
547 University Press, Oxford, UK, 1999).
- 548 50. Aalto, J., Pirinen, P. & Jylhä, K. New gridded daily climatology of Finland: permutation-
549 based uncertainty estimates and temporal trends in climate. *J. Geophys. Res. Atmos.* **121**,
550 (2016).
- 551 51. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework
552 for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* **196**, 483–
553 493 (2006).
- 554 52. R Core Team. R: A language and environment for statistical computing. (2020).
- 555 53. Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. & Ulrich, W. A consistent
556 metric for nestedness analysis in ecological systems: Reconciling concept and
557 measurement. *Oikos* **117**, 1227–1239 (2008).
- 558 54. Stone, L. & Roberts, A. The Checkerboard Score and Species Distributions. *Oecologia* **85**,
559 74–79 (1990).
- 560 55. Hart, S. P., Usinowicz, J. & Levine, J. M. The spatial scales of species coexistence. *Nat.*
561 *Ecol. Evol.* **1**, 1066–1073 (2017).
- 562 56. Cheng, J., Levina, E., Wang, P. & Zhu, J. A sparse Ising model with covariates. *Biometrics*
563 **70**, 943–953 (2014).
- 564 57. Susi, H. & Laine, A.-L. Agricultural land use disrupts biodiversity mediation of virus
565 infections in wild plant populations. *New Phytol.* **230**, 2447- (2020).
- 566 58. Hanley, J. A. & Mcneil, B. J. The Meaning and Use of the Area under a Receiver
567 Characteristic. *Radiology* **143**, 29–36 (1982).
- 568 59. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-6. (2019).
- 569 60. Csardi, G. & Nepusz, T. The igraph software package for complex network research.
570 *InterJournal, Complex Syst.* (2006).
- 571 61. Gorbalenyai, A. E. & Siddell, S. G. Recognizing species as a new focus of virus research.
572 *PLoS Pathog.* **17**, 1–7 (2021).
- 573 62. Roossinck, M. J. The big unknown: Plant virus biodiversity. *Curr. Opin. Virol.* **1**, 63–67
574 (2011).
- 575 63. Rey, M. E. C. *et al.* Diversity of dicotyledenous-infecting geminiviruses and their

- 576 associated DNA molecules in Southern Africa, including the South-west Indian Ocean
577 Islands. *Viruses* **4**, 1753–1791 (2012).
- 578 64. Tugume, A. K., Mukasa, S. B. & Valkonen, J. P. T. Mixed infections of four viruses, the
579 incidence and phylogenetic relationships of Sweet potato chlorotic fleck virus
580 (Betaflexiviridae) isolates in wild species and sweetpotatoes in Uganda and evidence of
581 distinct isolates in East Africa. *PLoS One* **11**, 1–27 (2016).
- 582 65. Al Rwahnih, M., Daubert, S., Golino, D. & Rowhani, A. Deep sequencing analysis of RNAs
583 from a grapevine showing Syrah decline symptoms reveals a multiple virus infection
584 that includes a novel virus. *Virology* **387**, 395–401 (2009).
- 585 66. Székely, A. J. & Langenheder, S. The importance of species sorting differs between
586 habitat generalists and specialists in bacterial communities. *FEMS Microbiol. Ecol.* **87**,
587 102–112 (2014).
- 588 67. Levine, J. M. & Hart, S. P. The Dimensions of Species Coexistence. in *Unsolved Problems in*
589 *Ecology* (eds. Dobson, A., Tilman, D. & Holt, R. D.) 145–159 (Princeton University Press,
590 2020). doi:10.2307/j.ctvs9fh2n.16.
- 591 68. Seabloom, E. W., Borer, E. T., Mitchell, C. E. & Power, A. G. Viral diversity and prevalence
592 gradients in North American Pacific Coast grasslands. *Ecology* **91**, 721–732 (2010).
- 593 69. Pleydell, D. R. J. *et al.* Estimation of the dispersal distances of an aphid-borne virus in a
594 patchy landscape. *PLoS Comput. Biol.* **14**, 1–24 (2018).
- 595 70. Sallinen, S., Norberg, A., Susi, H. & Laine, A. L. Intraspecific host variation plays a key
596 role in virus community assembly. *Nat. Commun.* **11**, 1–11 (2020).
- 597 71. Kozanitas, M., Osmundson, T. W., Linzer, R. & Garbelotto, M. Interspecific interactions
598 between the Sudden Oak Death pathogen *Phytophthora ramorum* and two sympatric
599 *Phytophthora* species in varying ecological conditions. *Fungal Ecol.* **28**, 86–96 (2017).
- 600 72. Wintermantel, W. M., Cortez, A. A., Anchieta, A. G., Gulati-Sakhuja, A. & Hladky, L. L. Co-
601 infection by two criniviruses alters accumulation of each virus in a host-specific manner
602 and influences efficiency of virus transmission. *Phytopathology* **98**, 1340–1345 (2008).
- 603 73. Calatayud, J. *et al.* Positive associations among rare species and their persistence in
604 ecological assemblages. *Nat. Ecol. Evol.* **4**, 40–45 (2020).
- 605 74. Pruss, G., Ge, X., Shi, X. M., Carrington, J. C. & Vance, V. B. Plant viral synergism: The
606 potyviral genome encodes a broad-range pathogenicity enhancer that transactivates
607 replication of heterologous viruses. *Plant Cell* **9**, 859–868 (1997).
- 608 75. Taiwo, M. A., Kareem, K. T., Nsa, I. Y. & D'A Hughes, J. Cowpea viruses: Effect of single

- 609 and mixed infections on symptomatology and virus concentration. *Viol. J.* **4**, 1–5
 610 (2007).
- 611 76. DaPalma, T., Doonan, B. P., Trager, N. M. & Kasman, L. M. A systematic approach to
 612 virus-virus interactions. *Virus Res.* **149**, 1–9 (2010).
- 613 77. González-Jara, P. *et al.* A single amino acid mutation in the Plum pox virus helper
 614 component-proteinase gene abolishes both synergistic and RNA silencing suppression
 615 activities. *Phytopathology* **95**, 894–901 (2005).
- 616 78. Leathwick, J. R., Elith, J. & Hastie, T. Comparative performance of generalized additive
 617 models and multivariate adaptive regression splines for statistical modelling of species
 618 distributions. *Ecol. Modell.* **199**, 188–196 (2006).
- 619 79. Ovaskainen, O. & Soininen, J. Making more out of sparse data: hierarchical modeling of
 620 species communities. *Ecology* **92**, 289–295 (2011).
- 621 80. Galiana, N. *et al.* The spatial scaling of species interaction networks. *Nat. Ecol. Evol.* **2**,
 622 782–790 (2018).
- 623 81. Dray, S. *et al.* adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-
 624 14. (2021).

626 **Acknowledgements**

627 We thank Krista Raveala, Mikko Jalo, Pauliina Hyttinen, and Mikko Immonen for assisting in
 628 the field work, Laura Häkkinen for assisting in the laboratory work, and Elina Numminen for
 629 helping with the use of the host plant population database. We also thank Guillaume Blanchet
 630 for helpful discussions. This work was funded by the European Research Council (4100097
 631 RESISTANCE) and Academy of Finland (334276) to A.-LL, Academy of Finland (321441) to HS,
 632 and LUOVA doctoral program fellowship to SS.

634 **Author Contributions:** Original idea of the study by A.-LL. H.S. and A.-LL. planned the study
 635 design. H.S. and S.S. collected the data, H.S. and P.S. conducted the bioinformatic work. A.N.

636 analysed the data and wrote the first draft of the manuscript. N.J.C. provided comments on the
637 statistical modelling. All authors contributed substantially to manuscript revisions.

638

639 **Data accessibility statement:** Upon acceptance, the data along with the whole analytical
640 pipeline leading to our results will be archived in Zenodo, and a DOI will be provided. You
641 may already find the data, along with all the processing steps, in Github
642 (aminorberg/meta17network-pkg), based on which this release will be.

643

644 **Competing Interests:** The authors declare no competing interests.

645

646

647

648

649

650

651

652

653

654

655

657

658 **Figure 1. A schematic representation of direct and indirect associations between**

659 **viruses in the context of their hosts and environment.** Within plant populations,

660 individual plants host viruses. The viruses have associations with each other, either through

661 direct mutualistic or antagonistic interactions (bold font), or indirect associations mediated

662 by another virus, shared vector, and/or host immunological responses (small font). The

663 environment can influence the (co-)occurrences either directly or through host effects at the

664 level of host populations or individuals, or via the vectors (capital font). The solid bold arrow

665 represents the direct association between two viruses, whereas the dashed arrows show the

666 confounding indirect associations. The additional conditioning explanatory variables included

667 in our constrained models are illustrated. The indirect effects (mediator species, changes and

668 variation in host immune response) are included in the models implicitly through spatial

669 variables and the model structure.

670

671

672

673

674

675

676

677

678

679

680

681

682

Figure 2. Virus communities by population. In A, each pie chart represents the virus community of a host plant population, overlayed on the map of Åland. Infection load is shown with greyscale, as indicated in the legend. Sliced portions of same shade of grey show plants with equal number of viruses, but in different combinations. B) Nestedness of the virus communities when aggregated to host population level. Two most contrasting communities are shown in more detail in C and D. In these barplots, the within-host virus community composition of a virus-rich plant host population 861 (D) and virus-poor population 3222 (C) is shown with different colours indicating different virus taxa, as shown in legend.

Figure 4. Virus networks. A) Markov random field co-occurrence network, representing direct pairwise associations between viruses, after accounting for the rest of the community (MRF). B) Constrained Markov random field co-occurrence network, representing direct pairwise associations between viruses, after accounting for the rest of the community as well as all environmental and spatial effects (CRFfull). C) Direct associations *explained* by host-related variables (CRFhost); D) direct associations *explained* by habitat-related variables (CRFhabitat); and E) direct associations *explained* by spatial variables (CRFspat). The viruses closer in space have stronger positive associations with each other and edge thickness is scaled by the strength of association. In summary, when the associations in panels C-D are subtracted from all the associations in panel A, the associations in panel B are left. Hence, the dashed lines in C-D represent associations that disappear from panel A after adding the explanatory variable group in question, resulting in the permanent associations illustrated with solid lines in panel B.

710 **Table 1. Explanatory conditioning variables.** See Supplementary Material and Methods for
711 more thorough description and justification of the explanatory variables.

Type		Name	Explanation	No. of unique values		
Spatial	1) Spatial auto-correlation		Moran’s eigenvectors (MEMs)	Spatial effects due to e.g. unmeasured environmental gradients and dispersal of the viruses and/or their vectors, implemented as distance-based Moran’s eigenvectors ⁵¹ , calculated ⁸¹ from the spatial locations of the sampled host plants.	394 (MEM1) 394 (MEM2) 396 (MEM3) 396 (MEM4)	
Environmental	2) Habitat-related	Quality	Connectivity	Connectivity of the population with respect to other populations, measured as the Euclidian distances separating populations and calibrated by the species dispersal capacity ⁴⁹	20	
			Local plant diversity	Shannon diversity index	20	
			Agricultural land use	% of the surrounding area, calculated as a proportion of 20m × 20m pixels from a one-km radius from the population falling under this land use category	20	
			Host population size	Amount of host plant within the focal population (m ²)	16	
		Weather	Winter severity	Number of severe winter days during winter 2017 (days)	9	
			Effective summer day temperature sum	Sum over temperatures of the effective summer days in 2016 (°C)	12	
		3) Host-related	Signs of herbivory	Suck-/bitemarks	Suck- and/or bitemarks produced by herbivores, recorded during the survey (0/1)	2
				Holes	Holes produced by herbivores, recorded during the survey (0/1)	2
	Moth			Moth pupa, recorded in the survey (0/1)	2	
	Miner			Leaf miner damage, recorded in the survey (0/1)	2	
	Thrips			Thrips damage, recorded in the survey (0/1)	2	
	Spittle			Spittle bug eggs, recorded in the survey (0/1)	2	
	Host size		Plant size	Plant size = $n \times A$, where n is the number of leaves, and leaf area is calculated using the equation of ellipse area: $A = \pi ab$, where a is a half axis of the width of the longest leaf, and b is the half axis of the length of the longest leaf	387	

712

713 **Table 2. The direct effects of the additional explanatory variables on the virus**
714 **occurrence probabilities and the indirect effects of the additional explanatory**
715 **variables through their influence on the association between a focal pair of viruses.**
716 Effects $\geq |0.25|$ are highlighted with **bold** font. All mean coefficient estimates are significant
717 (based on bootstrapping and a 90% confidence interval).

	Covariate	Virus A	Virus B	Coefficient mean
Direct effects	Connectivity	Geminiviridae	-	-0.15
	Host population size	Geminiviridae	-	1.0
	Local plant diversity	Caulimoviridae	-	-0.30
	Agricultural land use	Caulimoviridae	-	0.05
	Spittle	Caulimoviridae	-	1.54
	Moran's eigenvector 1	Geminiviridae	-	0.07
	Moran's eigenvector 2	Closteroviridae	-	0.05
	Moran's eigenvector 3	Caulimoviridae	-	-0.07
	Moran's eigenvector 4	Caulimoviridae	-	-0.15
	Moran's eigenvector 4	Geminiviridae	-	0.01
Indirect effects	Connectivity	Tymoviridae	Secoviridae	0.06
		Tombusviridae	Secoviridae	-0.21
		Pospiviroidae	Tymoviridae	0.23
		Bromoviridae	Avsunviroidae	-0.64
	Host population size	Tymoviridae	Secoviridae	0.01
	Local plant diversity	Tymoviridae	Secoviridae	0.24
	Winter severity	Bromoviridae	Tymoviridae	-0.25
		Geminiviridae	Pospiviroidae	-0.09
	Effective summer day	Geminiviridae	Virgaviridae	0.25
	Miner	Geminiviridae	Alphaflexiviridae	0.56
		Bromoviridae	Tymoviridae	0.99
	Spittle	Tombusviridae	Secoviridae	2.13
		Geminiviridae	Alphaflexiviridae	0.98
	Holes	Alphaflexiviridae	Virgaviridae	2.02
		Bromoviridae	Tymoviridae	0.37
		Caulimoviridae	Alphaflexiviridae	0.08
	Suck-/bitemarks	Bromoviridae	Avsunviroidae	1.5
	Moran's eigenvector 1	Betaflexiviridae	Secoviridae	0.40
		Tombusviridae	Secoviridae	0.07
		Potyviridae	Betaflexiviridae	0.28
		Fimoviridae	Alphaflexiviridae	0.65
		Geminiviridae	Pospiviroidae	0.13
		Tombusviridae	Bromoviridae	0.19
		Bromoviridae	Pospiviroidae	0.17
	Moran's eigenvector 2	Alphaflexiviridae	Virgaviridae	-0.1
		Pospiviroidae	Tymoviridae	-0.81
		Partitiviridae	Caulimoviridae	-0.14
	Moran's eigenvector 3	Caulimoviridae	Closteroviridae	-0.04
		Caulimoviridae	Partitiviridae	-0.35
		Geminiviridae	Bromoviridae	-0.26
	Moran's eigenvector 4	Partitiviridae	Caulimoviridae	-0.05
		Tymoviridae	Pospiviroidae	-0.81

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement.pdf](#)