



Spatial patterns in phage-*Rhizobium* coevolutionary interactions across regions of common bean domestication

Jannick Van Cauwenberghe^{1,2} · Rosa I. Santamaría¹ · Patricia Bustos¹ · Soledad Juárez¹ · Maria Antonella Ducci³ · Trinidad Figueroa Fleming⁴ · Angela Virginia Etcheverry⁴ · Víctor González¹

Received: 21 July 2020 / Revised: 14 January 2021 / Accepted: 21 January 2021 / Published online: 8 February 2021
© The Author(s) 2021. This article is published with open access, corrected publication 2021

Abstract

Bacteriophages play significant roles in the composition, diversity, and evolution of bacterial communities. Despite their importance, it remains unclear how phage diversity and phage-host interactions are spatially structured. Local adaptation may play a key role. Nitrogen-fixing symbiotic bacteria, known as rhizobia, have been shown to locally adapt to domesticated common bean at its Mesoamerican and Andean sites of origin. This may affect phage-rhizobium interactions. However, knowledge about the diversity and coevolution of phages with their respective *Rhizobium* populations is lacking. Here, through the study of four phage-*Rhizobium* communities in Mexico and Argentina, we show that both phage and host diversity is spatially structured. Cross-infection experiments demonstrated that phage infection rates were higher overall in sympatric rhizobia than in allopatric rhizobia except for one Argentinean community, indicating phage local adaptation and host maladaptation. Phage-host interactions were shaped by the genetic identity and geographic origin of both the phage and the host. The phages ranged from specialists to generalists, revealing a nested network of interactions. Our results suggest a key role of local adaptation to resident host bacterial communities in shaping the phage genetic and phenotypic composition, following a similar spatial pattern of diversity and coevolution to that in the host.

These authors contributed equally: Jannick Van Cauwenberghe, Rosa I. Santamaría

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-00907-z>.

✉ Jannick Van Cauwenberghe
jvancau@berkeley.edu

✉ Víctor González
vgonzal@ccg.unam.mx

¹ Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Mexico, Mexico

² Department of Integrative Biology, University of California, Berkeley, CA, USA

³ Instituto Nacional de Tecnología Agropecuaria, Universidad Nacional de Salta, Salta, Argentina

⁴ Facultad de Ciencias Naturales, Universidad Nacional de Salta, Salta, Argentina

Introduction

Bacterial viruses or bacteriophages are the most diverse and abundant biological entities on earth [1, 2]. They play a significant role in bacterial ecology and evolution, enabling horizontal gene transfer, and influencing bacterial diversity through their lytic or lysogenic cycles [3–6]. Nevertheless, phage biogeography, including patterns of dispersal, establishment, and community assembly, is very poorly understood [7], and its study could contribute to the advancement of microbiome engineering in agricultural and medical settings.

Most of the available knowledge about phage genetic diversity and spatiotemporal dynamics comes from metagenomic studies on marine cyanophages and gut microbiomes [8–10]. These studies have shown both cosmopolitan [1, 11–13] and habitat-specific phage lineages [14–21]. Given the predatory nature of phages, the presence or absence of suitable bacterial hosts shapes their distribution [8, 22–24].

Bacteriophages tend to be locally adapted to sympatric bacteria [25–28]. It is commonly predicted that local adaptation is a significant underlying factor of

compositional differences across phage communities [16, 24, 26, 27, 29]. Local adaptation is a process that results in a local population of a given species exhibiting higher fitness in its local environment than in allopatric populations and takes place when environmentally driven selection is more robust than migration [30, 31]. The implied lower fitness experienced by immigrants may limit gene flow and increase genetic differentiation across populations (i.e., “isolation-by-adaptation,” or more general “isolation-by-environment”) [32–35]. Host bacteria are a key environmental driver of phage local adaptation, or more precisely, “local coadaptation”, as phage adaptation (e.g., increased host range) readily invokes the counter- or coadaptation of bacteria (e.g., increased resistance) [25–27, 36–38].

While bacteriophages coadapt with their bacterial hosts, symbiotic bacteria coadapt with their eukaryotic hosts [39–41]. For instance, it is well established that rhizobia coadapt with legumes and form a mutualistic relationship in which they provide legumes with a steady supply of plant-usable nitrogen via nitrogen fixation [42, 43]. Aguilar et al. [44] found that Mesoamerican and Andean common bean (*Phaseolus vulgaris*) genotypes were preferentially associated with Mesoamerican and Andean rhizobia, respectively, indicating local coadaptation [44] across the two domesticated common bean gene pools [45, 46]. Furthermore, the *Rhizobium* communities were genetically differentiated, as the relative abundance of different types of the symbiotic *nodC* gene varied across the two bean gene pools [44]. In addition, in other legume-rhizobium systems, host legume population differentiation has led to rhizobium population differentiation [47, 48] and even local adaptation [49, 50].

Symbiotic organisms have significant reciprocal evolutionary effects on each other, which in turn affect third-party interactions [51–59]. Bacterial adaptation to plant hosts may affect bacteria-phage interactions (e.g., by altering the expression of surface receptors for plant interactions, which serve as anchor points of phages [60–62]). Here, we predicted that the genetic differentiation and local adaptation experienced by rhizobia across common bean gene pools shape the interaction and genetic differentiation of communities of phages infecting common bean-nodulating rhizobia. Even under the assumption of high phage dispersal capabilities, we expected communities of phages to be locally adapted, showing higher infection rates for sympatric rhizobia than for allopatric rhizobia. Hence, we aimed to elucidate whether the genomic identities and host ranges of phages infecting common bean-nodulating rhizobia are geographically structured. Our approach was to sample *Rhizobium* strains and associated bacteriophages from two common bean fields in Mexico and Argentina, corresponding to independent areas of common bean

domestication [45, 46]. Host species identity and bacteriophage genomic types were determined by Sanger sequencing and by both Sanger sequencing and whole-genome sequencing, respectively. Host range assessment results were analyzed for biogeographic signals and local adaptation. More specifically, we aimed to (i) assess bacteriophage diversity associated with common bean-nodulating rhizobia; (ii) compare the bacteriophage community composition across common bean gene pools; (iii) determine how bacteriophage host range pertains to bacteriophage provenance and the host species community composition; and (iv) obtain evidence of local adaptation of bacteriophages.

Materials and methods

Soil and *Rhizobium* sampling

The sites of rhizobia and phage sampling were common bean (*Phaseolus vulgaris*) agriculture fields located in Mexico (Tepoztlán and Yautepec, Morelos) and Argentina (Chicoana and Salta, Salta) (Fig. S1a). The bean fields were named after the municipalities in which they were located except for the ‘Salta’ field, which was located in Cerrillos near the border with the city of Salta. Sampling was done during spring in both regions to minimize effects of seasonal variation: Tepoztlán was sampled on March 2016 and Yautepec on May 2016 (Northern hemisphere spring); Argentinean fields were sampled on October 2016 (Southern hemisphere spring). The distance between Tepoztlán and Yautepec is c. 7.4 km; the distance between Salta and Chicoana is c. 20.8 km. In each bean field, we collected >1 L of rhizosphere soil from three plants separated by 10 m. Each soil sample was mixed and split into two aliquots, one used for phage isolation and one used to “trap” rhizobia in the root nodules of bean plants. In Tepoztlán, three bean plants (*P. vulgaris* var. Negro Veracruz) were collected directly from the field, and their nodules were used to isolate rhizobia. To trap rhizobia from Yautepec, we used *P. vulgaris* var. Negro Veracruz, while rhizobia from Argentina were caught using *P. vulgaris* var. Alubia Cerrillos. These cultivars belong to the Mesoamerican and Andean gene pools of domesticated common bean, respectively. Previously, axenically germinated seedlings were planted in pre-sterilized vermiculite pots and inoculated with a 100 mL soil sample. Nodules were harvested after 8–12 weeks of growth in a greenhouse under natural environmental conditions regarding light and temperature (Supplementary Method S1). *Rhizobium* strains were obtained from surface-sterilized nodules via the squashing method and streaked on PY_{Nal} agar plates (Supplementary method S2;

[63]). Three subculture steps were used to purify all isolated rhizobial strains. The isolates were stored at -80°C in 50% glycerol for long-term storage.

Isolation and purification of bacteriophages

Bacteriophages were isolated from soil samples via the enrichment protocol described previously [64] using both *Rhizobium* isolates from local sites (local collection, or LC) and *Rhizobium* from the laboratory collection (standard collection, or SC; Table S1). Briefly, dry sieved soil was suspended in PY-Nal medium at a ratio of 1:2 and incubated overnight at 30°C with shaking (250 rpm). Subsequently, the soil solution was centrifuged (10,000 g, 10 min, 4°C) to remove large particles, and chloroform was added to eliminate the remaining bacteria. Chloroform was removed by centrifugation, and the solution was filtered through a $0.22\ \mu\text{m}$ Millipore filter. The soil filtrate was used to inoculate one pair-member of each *Rhizobium* strain. The other pair-member served as an uninoculated control. LC rhizobia were inoculated with the filtrate from the soil of origin, while SC rhizobia were inoculated with soil filtrate that was pooled from each of the three soil samples per bean field. Cultures of 1 ml in PY_{Nal} were incubated at 30°C (250 rpm) to an optical density of 0.2 at 620 nm (OD_{620} ; Beckman DU650 spectrophotometer) in 96-deep-well plates. After 20 h of incubation, the cultures were centrifuged, and the supernatant was used to reinoculate new cultures of the same pairs of rhizobial strains. This enrichment process was repeated five times; each time, the OD_{620} value was compared between the control and the inoculated pair member. A decrease in the OD_{620} was indicative of cell lysis. The presence of phages was confirmed by plaquing a dilution series of the filtrate mixed with 2.5 mL of molten soft PY (0.65% agar) on lawns of rhizobia over solid PY_{Nal} plates, as in Carlson (2005) [65]. Lytic plaques were picked and inoculated again in the respective bacterial cultures. Two additional dilution series were performed to ensure the purity of the bacteriophage stocks. Finally, the phage dilution was treated with chloroform to remove the remaining bacteria and stored at 4°C .

Phylogeny and taxonomic *Rhizobium* species identification

Two chromosomal genes, *dnaB* and *recA*, were partially amplified by colony PCR. The primers and PCR protocols used are described in Table S2. The PCR products were purified using the Exo-SAP cleanup protocol [66]. The purified products were sent to Macrogen for sequencing (Macrogen Inc, Seoul, Korea). Sanger sequences were edited and assembled in Genius Pro v. 6.1.2. Multiple alignments were performed using MUSCLE [67], followed

by manual correction to remove ambiguously aligned regions. Phylogenetic trees were reconstructed and edited with MEGA 7 using the maximum likelihood (ML) method based on the GTR + G + I model and 1000 bootstrap replicates. Rhizobia were clustered into sequence types (STs) based on 100% sequence identity of *dnaB* and *recA* sequences.

Genome sequencing

We obtained the genome sequences of 100 phages from the collection chosen according to their host range differences. Phage genomic DNA was purified from phage stocks propagated in the corresponding host *Rhizobium* strains. Host DNA and RNA were eliminated using DNase and RNase. Subsequently, phages were precipitated using a PEG-8000/NaCl solution. After centrifugation (10,000 g, 20 min, 4°C), the pellet was suspended in Tris-EDTA buffer. Proteins were hydrolyzed using 4% SDS and proteinase K and precipitated by adding 3 M potassium acetate. Phage genomic DNA was precipitated with 100% isopropanol and washed with 70% ethanol twice.

Phage genome sequencing was performed with Illumina technology in a Nextgen 500 system (Unidad Universitaria de Secuenciación Masiva de DNA (UUSMD-UNAM). Genomes were assembled from trimmed [68] sequence reads using the Spades v. 3.13.1 [69], Velvet v. 1.2.10 [70] and Phred/Phrap/Consed v. 23.0 [71] software packages.

The remaining 96 phages were identified via PCR and Sanger sequencing of phage marker genes used to distinguish between phage genomic types (PGTs). Each of these genes was shared among members of the corresponding PGT, and showed little to none nucleotide sequence similarity to genes of other PGTs; this was assessed by the presence-absence profile obtained with the BPGA software [72]. The extracted sequences were used to design primers with Primer3 [73]. The selected genes, primers, and PCR protocol are described in Table S2.

Comparative genomics

The average nucleotide identity (ANI) of all pairs of phage genomes was calculated with pyani v.0.2.9 using the ANIm MuMmer method [74, 75]. PGTs were clustered based on 80% nucleotide identity and 60% coverage of the smaller genome. The assignment of PGTs to the phage morphological families of tailed bacterial viruses (Siphoviridae, Myoviridae, and Podoviridae) was performed using the VirFam server (<http://biodev.cea.fr/virfam/>) [76]. PGTs assigned to Microviridae were recognized by their short genome length and through BLASTn searches against the virus database of the NCBI (<https://www.ncbi.nlm.nih.gov/genome/viruses/>).

Rhizobium susceptibility and host range assessment

The infectivity of all isolated phages was evaluated by the spot test procedure [77] in the rhizobia isolated from all four bean fields. Although this method of assessing the spotting host range could overestimate the lytic properties of phages and underestimate infections due to low phage titers, different adsorption efficiencies, or alternative infection modes (e.g., pseudolysogeny) [78–81], the spotting of phages on lawns is an affordable method for testing a large number of pairwise interactions [82]. *Rhizobium* ‘lawns’ in double-agar-layer plates were spotted with 5 µl of each bacteriophage solution prepared from the respective phage *Rhizobium* lysates. After overnight incubation at 30 °C, the plates were assessed for lysis. At least three replicates of each *Rhizobium*-phage combination were performed to ensure reproducibility, and at least two replicates with the same lytic or resistant phenotype were considered to indicate a positive or negative result. Spots resulting from lysis with a translucent appearance rather than a transparent appearance were recorded as ‘partial lysis’ but were treated equivalently to transparent spots for the statistical and BiMat analyses. The binary interaction matrix is available from Github (github.com/jvancau/interactiondata).

With this information, a matrix was constructed based on Bray–Curtis dissimilarity calculated with the *vegdist* function of the *vegan* package in R [83]. To compare the phenotypic composition with the genetic composition, we clustered the bacterial hosts showing >80% similarity in terms of their susceptibility to phages into *Rhizobium* phenotype groups (RPGs, Table S3). Then, phages that infected a common range of hosts (Bray–Curtis similarity >80%) formed phage phenotype groups (PPGs, Table S3).

Network structure of phage-bacterium interactions

To analyze the modularity and nestedness properties of the phage-bacterium bipartite network, we employed the BiMat program [84], which maximizes the similarities in the bacteria-phage lytic interaction matrix. The program ran in the MATLAB environment and was used according to the author’s start guide [84] (<https://www.github.com/cesar7f/BiMat>). Modularity was tested using the Adaptive Brim algorithm, and nestedness was tested using the NODF (Nestedness metric based on Overlap and Decreasing Fill) and NTC (Nestedness Temperature Calculator) algorithms. Statistical analysis was performed using 1000 replicates and the equiprobable null model.

Local adaptation

Rhizobium susceptibility rates were calculated for each *Rhizobium* strain as the number of phages able to infect the

strain divided by the total number of tested phages. Similarly, phage infection rates were calculated as the proportion of rhizobia that a given phage could infect among the total number of strains tested. These values were calculated for sympatric and allopatric phage-strain combinations. The statistical significance of the differences between sympatric and allopatric susceptibility/infection rates was tested using a generalized linear model in R with a quasi-binomial model [85]. Lower susceptibility rates with sympatric phages than with allopatric phages indicate rhizobium local adaptation; higher infection rates on sympatric rhizobia than on allopatric rhizobia indicate phage local adaptation.

Phage local adaptation was also calculated according to Vos et al. [27] as the mean difference in the mean of the rates of sympatric (S) and allopatric (A) phage infections.

Statistical analyses

To test the significance of compositional differences in *Rhizobium* STs and PGTs across common bean fields, we performed PERMANOVA tests based on Jaccard and Bray–Curtis distances with 999 permutations using the *ecodist* and *vegan* packages [83]. Principal coordinates of neighbor matrices (PCNM), which are orthogonal spatial variables derived from a spatial distance matrix, were calculated from the geographical coordinates using the ‘*pcnm*’ function of the *vegan* package for R [86]. The first PCNM value was used as a proxy for spatially related variation across the two regions and was fit on a principal coordinates analysis (PCoA) using *envfit* (‘*vegan*’) with 9999 permutations. This approach was employed to assess the significance of the distance between the two regions in explaining the compositional differences among bean fields. For each distance matrix type, the correlations among all datasets of the genetic and phenotypic composition across bean fields (ST, PGT, RPG and PPG) were assessed using Mantel tests (999 permutations, *vegan* package).

A Mantel test was also used to correlate *Rhizobium* and phage genetic distances with the corresponding susceptibility range and host range distances across all isolated rhizobia and phages. PERMANOVAs were used to test the significance of the effects of the origin and genetic or taxonomic identity of rhizobia and phages on their susceptibility range and host range, respectively.

Phage and *Rhizobium* nucleotide accessions

Phage genome sequences are available from GenBank with IDs MN988459 to MN988558. *Rhizobium* nucleotide sequences are available under: MT756388 – MT756428 (*dnaB*) and MT756429 – MT756469 (*recA*).

Results

Rhizobium and phage community sample composition

We studied 229 *Rhizobium* strains isolated from four agricultural plots in Central Mexico (Tepoztlán and Yautepec) and Northwest Argentina (Salta and Chicoana) (Fig. S1a–b). The rhizobial strains from each site (LC, local collections) were employed to trap phages from the soil of the same locality by the enrichment method [64]. In parallel, the same soils were pooled for each plot and used to search for phages using a SC of 94 *Rhizobium* strains of diverse geographic origins maintained in our laboratory (Table S1). A total of 196 phages were obtained with this protocol, 110 from LC and 86 from SC (Fig. S1b).

Genetic differentiation of *Rhizobium* populations between agricultural fields

Our first aim was to investigate whether the collected *Rhizobium* strains represent geographically structured populations. Phylogenetic analysis of the partial *recA-dnaB* sequences of 229 rhizobial strains identified *R. etli* as the predominant species at Mexican sampling sites (74.6%), while *R. phaseoli* was dominant in Argentina (80%) (Fig. 1; Fig. S2a). According to the nucleotide variations in *recA* and *dnaB*, all of the isolated rhizobia were grouped into 41 chromosomal STs (Table S4). PERMANOVA tests employing two distance measures (Jaccard and Bray–Curtis) indicated that the rhizobial communities differed significantly in terms of genetic composition and the relative abundance of genotypes (STs) among different bean fields and regions (Fig. 2A–C; Table S4–S5). The geographic distance between the regions, represented by a PCNM vector, was correlated with the ST composition among the rhizobia isolated from the four common bean fields (Table S5). At the species level, *R. etli* and *R. phaseoli* STs differed significantly among the sites of origin of the common bean fields. Between regions, only the *R. etli* populations differed in terms of the ST composition, whereas the *R. phaseoli* ST composition changed only marginally across regions (Table S5). The abundance of ST-5 in *R. phaseoli*, the only ST of the 41 total STs found across the four common bean fields, might explain this last result (Fig. 2C; Table S4).

Diversity of phage communities

To assess phage diversity, we obtained the complete genome sequences of 100 out of 196 phages (Table S6). The phage genomes displayed a wide range of lengths (from 4.8

to 207.6 kb; median 54.4) and GC contents (from 41 to 61%; median 57%). They were clustered into 29 PGTs (defined by ANIm, see Methods), 18 of which had two or more individual genomes, and 11 were singletons. Within PGTs, the genomes exhibited nucleotide variation ranging from 85.8 to 99.7%, with a coverage of ~64.1 to 100% (Fig. S3). In addition, 90 phages among the remaining 96 phages were assigned to PGTs by the PCR identification of specific phage genes conserved within the members of the 29 PGTs, followed by Sanger sequencing (see Methods) (Table S2).

Phages were also classified into morphological families using VirFam predictions [76]. Most of them (26/29) belonged to the order Caudovirales, represented by the families Podoviridae (12 PGTs), Siphoviridae (8 PGTs), and Myoviridae (6 PGTs). Three PGTs were identified as members of the Microviridae family.

Most phage families were present at the four sampling sites, but the Salta community was dominated by Myoviridae (69%) and the Chicoana community by Siphoviridae (62%) (Fig. S2b; Table S7). The Microviridae family (F02), defined by small-genome phages (4.8–6.2 kb), was dominant in Tepoztlán (60%), whereas it showed low abundance or was absent in the other populations (0–23%; Table S7).

Phage population differentiation across agricultural sites

Following the differences in the rhizobial ST composition per sampling site, we found that the composition of PGTs also differed significantly among common bean fields and regions (Fig. 2 D–F, Table S7). The differences were significantly correlated with geographic distance (Table S5). The phage communities were also significantly different between the Mexican bean fields within regions based on Jaccard distances ($F_{1,5} = 3.7$, $p = 0.024$), but they were not significantly different between Mexican bean fields based on Bray–Curtis distances ($F_{1,5} = 3.7$, $p = 0.055$) or between Argentinian bean fields (Jaccard: $F_{1,5} = 1.1$, $p = 0.384$; Bray–Curtis: $F_{1,5} = 2.2$, $p = 0.149$). Mantel tests showed that the genetic composition differences among phage communities were significantly correlated with the differences in the *Rhizobium* community genetic composition among bean fields (Table S5). Some PGTs coexisted at two of the sampling sites, whereas PGTs rarely coexisted at three sites and never at four (Fig. 2F). Moreover, 52% of the 29 PGTs occurred solely in one bean field, and 69% were restricted to a particular region. A spatial pattern distinction was also shown by the ANI values, as the average ANI of allopatric phages belonging to the same PGTs was 88%. In comparison, the average ANI of sympatric phages belonging to the same PGT was 96%.

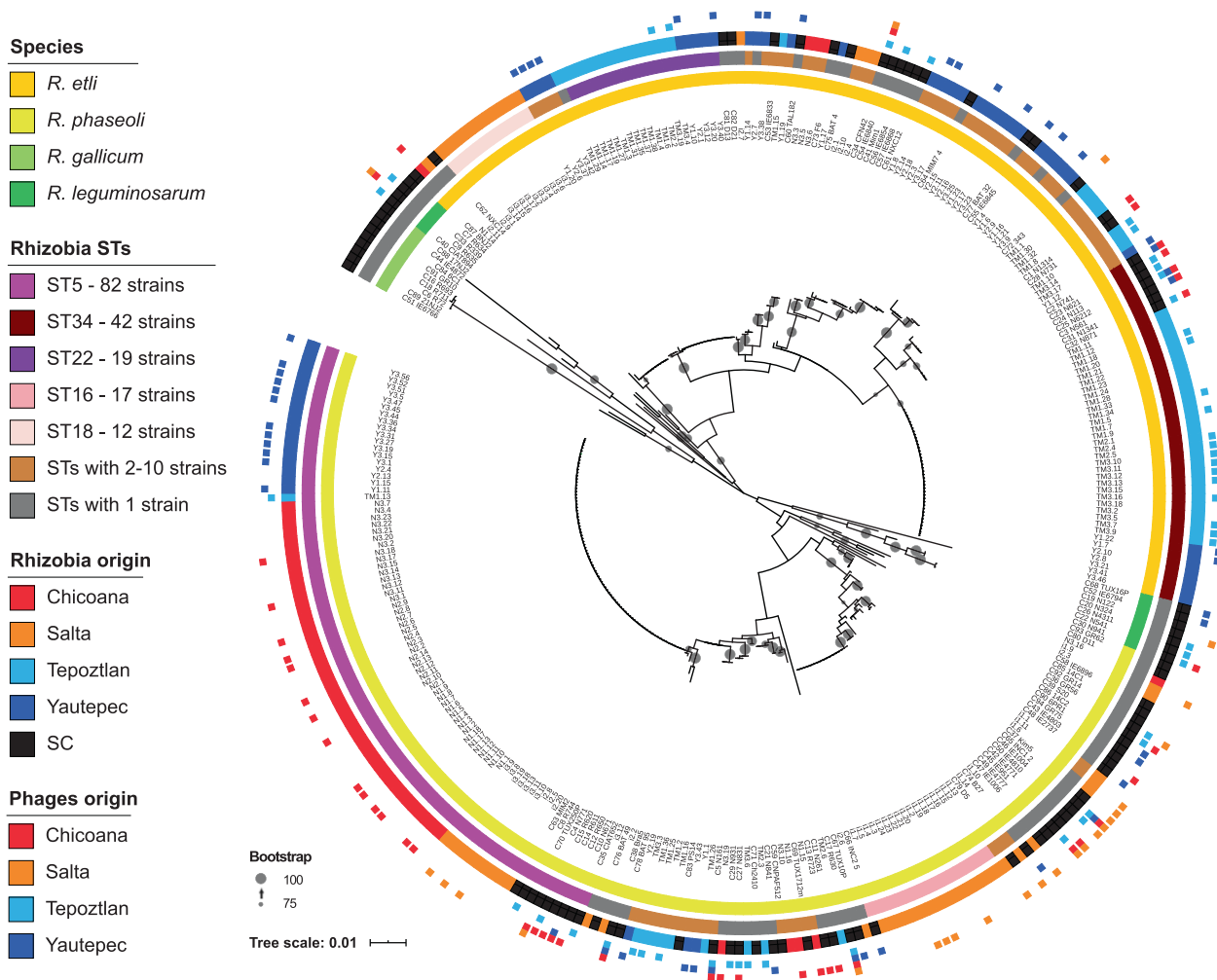


Fig. 1 Phylogenetic tree of all rhizobia collected from Mexico and Argentina (LC, $n = 229$) and from the standard laboratory collection (SC, $n = 94$). The tree was constructed using the maximum likelihood method and is based on the concatenated sequences of two chromosomal genes (*dnaA-recA*) (see Methods). The bar scale indicates the number of nucleotide substitutions per site. Insets on the left side explain the contents of the four concentric circles. From the inner

to the outermost circles: taxonomic classification of rhizobia, *Rhizobium* chromosomal STs (sequence types; no STs were assigned to SC strains), field of origin of the strains, and squares indicating the origin of the phages isolated using the corresponding strain. Bootstrap values are shown with proportionally-sized gray circles on the tree branches; smallest circles equal 75% and the largest circles equal a value of 100%.

Structure of the phage-bacterium interaction network

To define the rhizobium-phage interactions within and between the four communities, we tested the infectivity of 196 phages against 229 *Rhizobium* strains by the spot assay (see methods). We registered the following phenotypes: complete lysis (transparent spots), partial lysis (translucent spots), and resistance (absence of lysis), in three independent experiments.

A total of 44,884 interactions were examined in triplicate experiments; 19,474 plaques showed full or partial lytic phenotypes, recorded as positive interactions (1 in the bipartite network of Fig. 3) [84, 87]. The *Rhizobium*-resistant phenotypes were recorded as 0 in the corresponding binary matrix (Fig. 3).

Overall, the BiMat network showed high connectance (0.43), indicating that there were ~4/10 effective phage-bacterium interactions in the community, and weak modular differentiation ($Q_b = 0.21$; Fig. 3A). Three large modules, with high internal connectance in comparison with the outside modules, were detected (Fig. 3A). This means that to a certain extent the rhizobia and phages in these modules preferentially interact with each other rather than with members the other modules. Phages from Mexico (Tepoztlán and Yautepec) generally interacted best with *Rhizobium* isolates from Mexico. It was less common for these phages to infect *Rhizobium* from Argentina. These results suggest large-scale modularity dominated by sympatric interactions. However, some exceptions were observed; for instance, phages from Chicoana were very promiscuous, since a

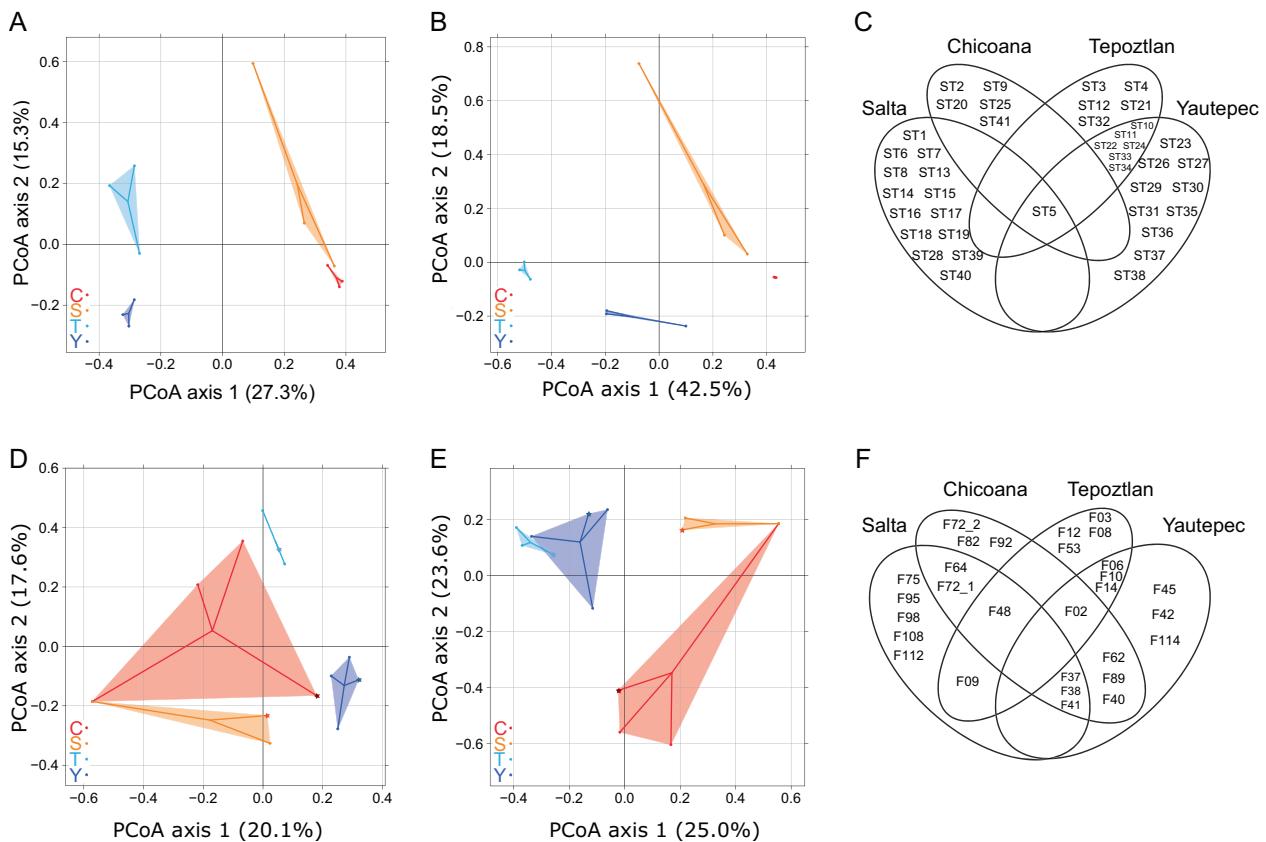


Fig. 2 Differentiation of *Rhizobium* and phage communities. **A** and **B** PCoA plots illustrating the differences in chromosomal composition between *Rhizobium* communities across common bean populations based on Jaccard distances (presence-absence) and Bray–Curtis distances (relative abundance), respectively. **D** and **E** PCoA plots showing the differences in the phage genomic type composition between phage communities across common bean populations based

on Jaccard distances and Bray–Curtis distances. Bean fields are indicated by different colors: Tepoztlán (T), light blue; Yauatepec (Y), dark blue; Salta (S), orange; Chicoana (C), red. Phages isolated by inoculating pooled samples from a given location into rhizobia from the laboratory’s standard collection of rhizobia (SC) are indicated with a star symbol. **C** and **F** Venn diagrams of the distribution of *Rhizobium* chromosomal STs and phage genomic types (PGTs), respectively.

significant number of Mexican strains were cross-infected by these Argentinian phages (Fig. 3A). This may in part explain the overall nested structure shown by the BiMat matrix, quantified by the independent NODF (0.68) and NTC (0.73) algorithms. This suggests that the phage communities consist of a range of specialist to generalist phages (Flores et al. 2013). This is visualized in Fig. 3B, where phages on the left side are more generalist, while further to the right phages become increasingly specialist.

***Rhizobium* susceptibility and phage host range**

Although all *Rhizobium* strains were closely related in the *recA-dnaB* phylogenetic tree, they varied widely in their susceptibility range, with some rhizobia being infected by ~10.2–74.0% of the phages (average rate of infection = 43.4%). Hence, we sought to investigate whether the susceptibility of rhizobia was associated with geographic origin, taxonomic affiliation or genetic identity. When we compared the variation in the susceptibility range between

individual rhizobia, we found that the susceptibility of rhizobia was significantly different not only between species ($F_{1,217} = 10.3$, $p = 0.001$) (Fig. 4A) but also among bean fields ($F_{3,217} = 31.0$, $p = 0.001$), and we observed the interaction of the two factors ($F_{3,217} = 6.4$, $p = 0.001$; Fig. 4A). The last finding indicates that the effect of *Rhizobium* species identity on the susceptibility range depends on the geographic origin of the rhizobia. When species were split into STs, the *Rhizobium* susceptibility range was found to differ significantly among bean fields ($F_{3,178} = 53.5$, $p = 0.001$) and STs ($F_{40,178} = 6.1$, $p = 0.001$), although their interaction was not significant ($F_{6,178} = 0.9$, $p = 0.542$) due to the limited geographic spread of most STs. Susceptible strains were clustered in 48 RPGs (*Rhizobium* Phenotype Groups; see methods). Twenty-nine RPGs were singletons, whereas 19 were formed by more than two strains (maximum of 40) (Table S3). Thirty-two RPGs belonged to *R. etli*, twelve RPGs corresponded to *R. phaseoli* (RPG 4, 13, 19, 20, 22, 23, 24, 27, 31, 32, 33 & 37), and the other four RPGs belonged to both species (RPG 1, 2, 6, and 11). The

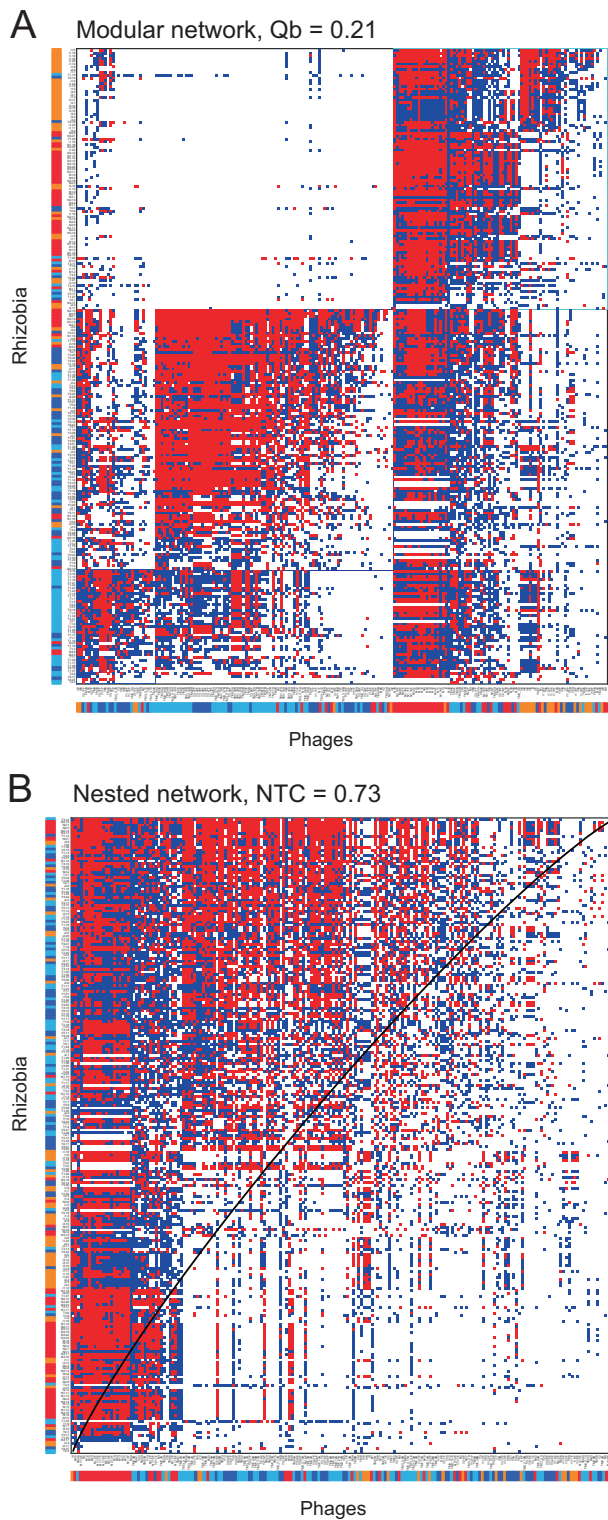


Fig. 3 Structure of phage-*Rhizobium* interactions. Phage-bacterium infection network of interactions between 196 phages (columns) and 229 *Rhizobium* strains (rows) performed with BiMat [84]. Either full (red) or partial (blue) lytic interactions were recorded as positive interactions in at least 2/3 of independent tests; blank cells indicate the absence of interaction. In the bottom row, colored bars indicate the bean field of origin of the phages: Tepoztlán (T, Mexico) = light blue; Yautepec (Y, Mexico) = dark blue; Salta (S, Argentina) = orange; Chicoana (C, Argentina) = red. The first column on the right indicates the site of origin of rhizobia with the same colors of the bars used for the phages. **A** Modular sorting of the interaction data, visualizing the presence of three modules. **B** Nested sorting of the interaction data, visualizing the spectrum of generalists to specialist phages. The isocline (black line) represents the division line in a perfectly nested matrix between an area of interactions and an area of no interactions. Name labels can be identified in the pdf version of this document, and the dataset is available as supplementary data online.

range similarity were significantly correlated ($r = 0.31$, $p = 0.001$).

We examined the host range of phages by the spot method, to identify those with similar infection spectra. Phage infection rates, or the proportion of hosts that a given phage isolate could infect, varied considerably from 2.2 to 92.6% (average infection rate 43.4%). All but three phages were able to infect both *R. etli* and *R. phaseoli*; however, the phages were able to infect only 51% of STs on average. The host range of the phages was significantly affected by the bean field of origin ($F_{3,142} = 23.0$, $p = 0.001$), phage genomic type (PGT; $F_{27,142} = 6.9$, $p = 0.001$), and the interplay between these factors ($F_{14,142} = 2.9$, $p = 0.001$). The last finding indicates that the effect of PGT on the host range depends on the geographic origin of the phage. Similar results were obtained when the phages were grouped by taxonomic family: phage host range was significantly affected by the bean field of origin ($F_{3,172} = 14.2$, $p = 0.001$), family ($F_{3,172} = 8.4$, $p = 0.001$), and the interplay between them ($F_{8,172} = 3.9$, $p = 0.001$) (Fig. 4B).

Based on Bray-Curtis dissimilarity <20%, 139 phages (70% of the total) were clustered into 24 PPGs; the remaining phages (57 phages) exhibited a unique PPG. Within the PPGs, two profiles accounted for 53% of the phages, whereas 22 profiles exhibited fewer than ten phages from similar PPGs. The PPG composition across common bean fields was significantly correlated with the RPG composition based on Bray-Curtis distances, but not based on Jaccard distances (Table S8). The PPG composition was also significantly correlated with PGT composition (Table S8). Accordingly, host range similarity was considerably correlated with phage ANI ($r = 0.29$, $p = 0.001$).

Rhizobium-phage local adaptation

The above results suggest that phage communities are adapted to *Rhizobium* in their areas of origin. To examine this issue, we estimated the infectivity rates of phage isolates

most abundant RPGs (1–4) were composed of the frequently identified STs of *R. phaseoli* (ST-5) and *R. etli* (ST10 and ST34). We found that the RPG composition was strongly correlated with the ST composition (Table S8). In addition, *Rhizobium* genetic distance and susceptibility

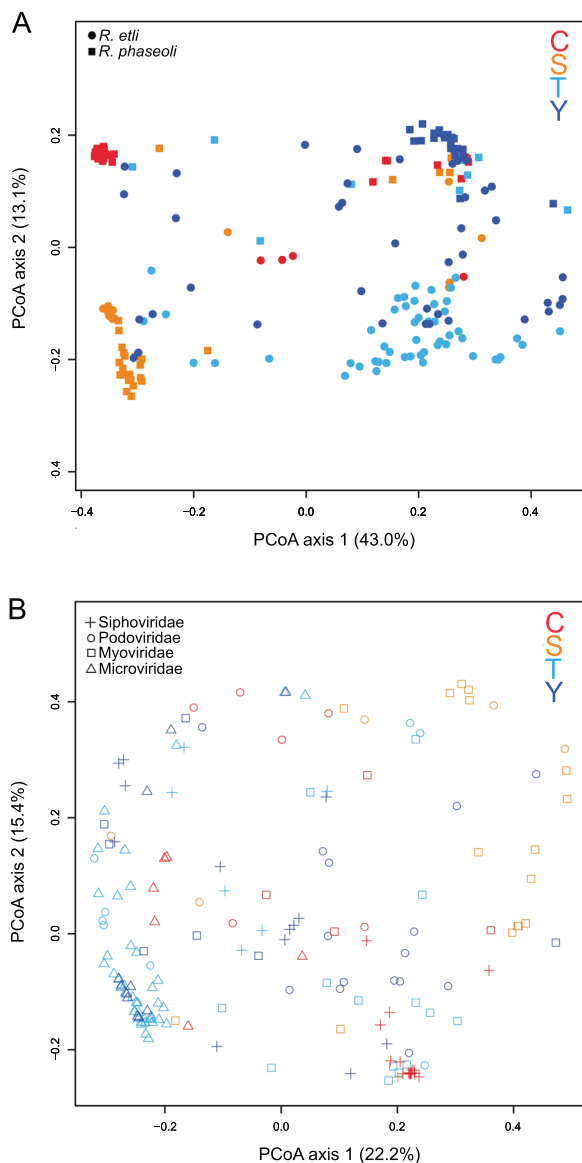


Fig. 4 *Rhizobium* susceptibility range and phage host range similarity. Principal coordinates analysis (PCoA) plot showing the Bray–Curtis dissimilarity in the *Rhizobium* susceptibility range among common bean fields and *Rhizobium* species in (A) and the Bray–Curtis dissimilarity in the phage host range among regions and phage taxonomic families in (B). Each axis explains a certain fraction of dissimilarity, given within parentheses. Different *Rhizobium* species and phage taxonomic families are represented by symbols. The bean fields of origin are indicated by different colors: Tepoztlán (T), light blue; Yautepec (Y), dark blue; Salta (S), orange; Chicoana (C), red.

and the susceptibility rates of *Rhizobium* isolates in sympatric and allopatric combinations. Despite ample variation, phage infection rates were significantly greater for sympatric infections (mean = 0.55 ± 0.01 ; CV = 55%; $\sigma^2 = 0.092$) than for allopatric infections (mean = 0.39 ± 0.02 ; CV = 64%; $\sigma^2 = 0.062$; $F_{1,390} = 23.2$, $p = 2.06e-6$) (Fig. 5A), suggesting a trend of local adaptation. This was true for both phages isolated using the SC (SC phages; $F_{1,170} = 5.7$, $p = 0.019$) and phages isolated using local rhizobia (LC phages;

$F_{1,218} = 20.1$, $p = 1.17e-05$). However, sympatric phage infection rates were significantly higher than allopatric infection rates for Mexican communities (0.52 ± 0.02 versus 0.35 ± 0.02 ; $F_{1,252} = 30.2$, $p = 9.49e-08$), but the difference was not statistically significant for Argentinean communities (0.60 ± 0.04 versus 0.47 ± 0.04 ; $F_{1,136} = 3.5$, $p = 0.062$) (Fig. 5B). The same observations were made both for phages isolated using the SC (SC; Mexican: $F_{1,94} = 7.1$, $p = 9.30e-03$; Argentinean: $F_{1,74} = 1.0$, $p = 0.323$) as for phages isolated using local rhizobia (LC; Mexican: $F_{1,156} = 26.4$, $p = 8.17e-07$; Argentinean: $F_{1,60} = 2.7$, $p = 0.104$). This was largely due to the fact that the phages isolated from Chicoana (Argentina) showed no differences between the rates of infection of local rhizobia and those of nonlocal rhizobia isolated from the other three sites (Fig. 5E). Although the Mexican phages were locally adapted overall, phage cross-infection was detected at similar rates between the communities of Tepoztlán and Yautepec (Mexico), indicating nonadaptive phage differentiation between these communities (Fig. 5E). Indeed, the infection rates of Mexican phages in Argentinian rhizobia were significantly lower, suggesting an effect of geographic distance on local adaptation (Fig. 5B; Fig. 5E).

On average, the rhizobia were more susceptible to sympatric phages (mean 0.54 ± 0.01 ; CV = 36.9%; $\sigma^2 = 0.039$) than to allopatric phages (mean 0.40 ± 0.01 ; CV = 49.7%; $\sigma^2 = 0.040$), indicating that the bacterial populations were maladapted to the local phages ($F_{1,456} = 46.7$, $p = 2.72e-11$; Fig. 5C). Similar results were obtained regardless of whether susceptibility rates were calculated for each *Rhizobium* isolate using only infections by phages isolated by using the SC (SC phages; $F_{1,456} = 38.9$, $p = 9.99e-10$) or phages isolated by using local rhizobia (LC phages; $F_{1,456} = 45.9$, $p = 3.85e-11$). All of these results suggest that rhizobia are maladapted to coexisting phages. However, the high susceptibility of rhizobia to local phages appeared to hold only for Argentinian *Rhizobium* communities (0.57 ± 0.02 versus 0.30 ± 0.02 ; $F_{1,212} = 64.3$, $p = 7.20e-14$; SC: $F_{1,212} = 102.5$, $p < 2.2e-16$; LC: $F_{1,212} = 42.3$, $p = 5.64e-10$), since the Mexican *Rhizobium* isolates showed no significant differences in susceptibility to sympatric and allopatric phage infection (0.51 ± 0.02 versus 0.49 ± 0.01 ; $F_{1,242} = 1.2$, $p = 0.268$; SC: $F_{1,242} = 0.4$, $p = 0.508$; LC: $F_{1,242} = 2.0$, $p = 0.159$) (Fig. 5D; Fig. 5F). When we considered the susceptibility of *Rhizobium* to sympatric phages or allopatric phages by site, the Chicoana *Rhizobium* community was found to be very susceptible to its own phages but resistant to phages from the other sites (Fig. 5F). In contrast, the rhizobial strains from Tepoztlán, Yautepec, and Salta were very susceptible to Chicoana phages (Fig. 5F), which partly explains the lack of significant differences in susceptibility to sympatric and allopatric phage infection by Mexican *Rhizobium* isolates.

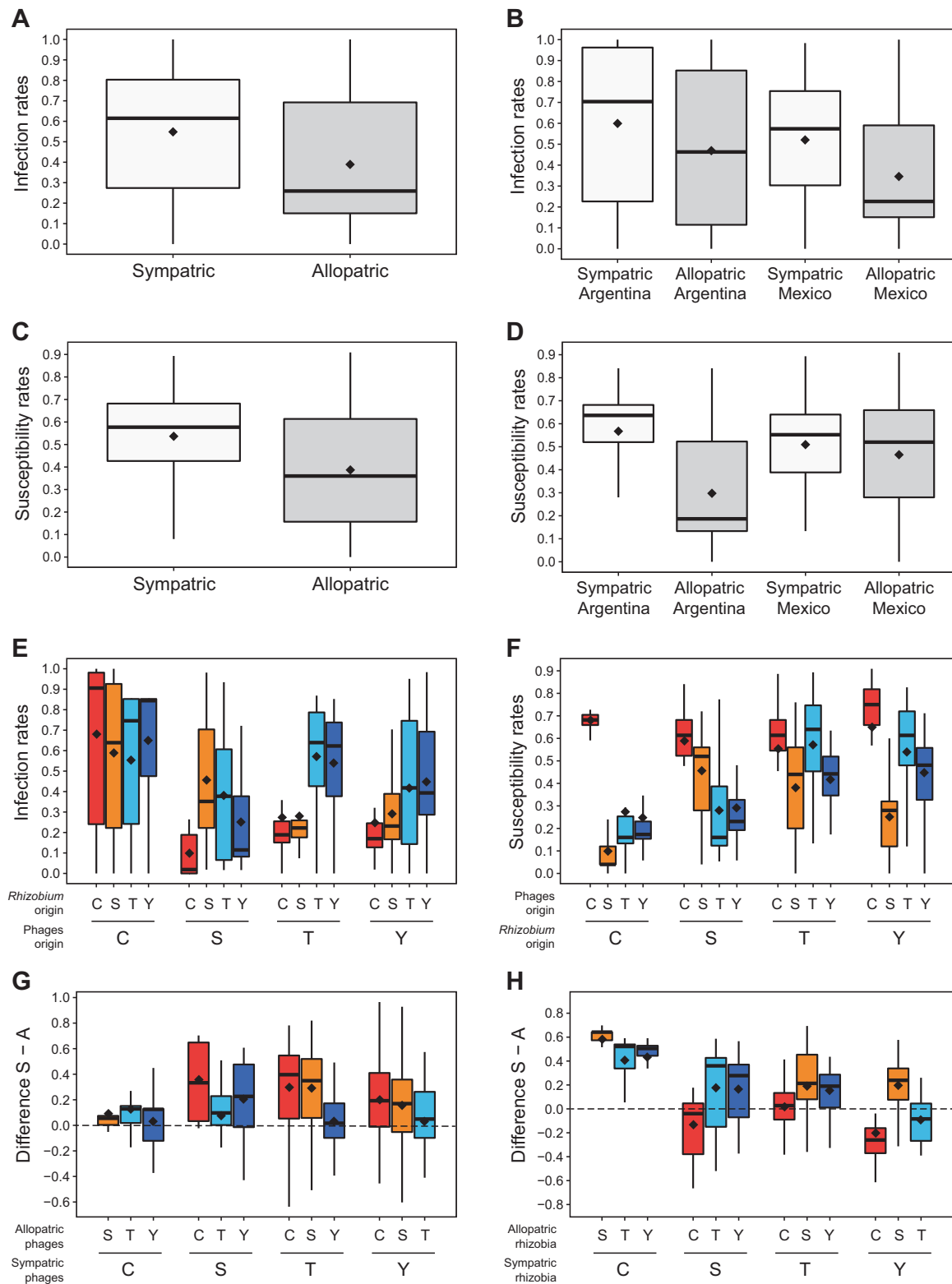


Fig. 5 Box plots of phage infection rates and *Rhizobium* susceptibility rates indicating phage local adaptation and *Rhizobium* maladaptation. Phage infection rates between sympatric versus allopatric combinations averaged across all locations (A), averaged across populations within regions (B), or among populations (E). *Rhizobium* susceptibility rates between sympatric versus allopatric combinations averaged across all locations (C), averaged across populations within

regions (D), or among populations (F). The mean is given for each box plot (black diamond). Differences between the means of sympatric (“S”) versus allopatric (“A”) phage infection rates (G) and the rates of the susceptibility of *Rhizobium* (H) are also shown. The origins of the phage and *Rhizobium* samples are indicated by T = Tepoztlán (Mexico), Y = Yautepec (Mexico), S = Salta (Argentina), and C = Chicoana (Argentina).

In most cases, the difference between the mean sympatric rates of phage infection (S) and the mean allopatric rates of infection (A) indicated phage local adaptation in the four rhizobium-phage communities (Fig. 5G). In contrast, resident rhizobia were more susceptible to sympatric than to allopatric phages, except for the rhizobia from Yautepec that were also efficiently infected by Tepoztlán phages (Fig. 5H). The four populations of rhizobia were susceptible to the Chicoana phages, but the Chicoana rhizobia were mostly resistant to allopatric phages (Fig. 5H).

Discussion

Overall, our data indicate that the genetic and phenotypic diversity of phages and their *Rhizobium* hosts is spatially structured and that phages are adapted to their local host communities. Previous research has shown that rhizobia are spatially structured and can locally adapt to their legume hosts and other local environmental factors [49, 50, 88, 89]. For instance, across the areas of common bean domestication, rhizobia receive a greater competitive benefit when nodulating sympatric common beans (*Phaseolus vulgaris*) than nonnative common bean varieties [44]. In turn, our results show that sympatric phage communities have locally adapted to these rhizobia.

We found that each of the four phage communities analyzed was dominated by a particular taxonomic family, prominently Microviridae in Tepoztlán, Myoviridae in Salta, and Siphoviridae in Chicoana. Moreover, phage genome sequencing revealed high genetic diversity within taxonomic families. Phage genetic diversity varied considerably within and between communities, and phages were clustered into PGTs. Approximately 52% of the 29 PGTs occurred solely in a single phage community (PGT). Similarly, Scola et al. [90] found that 66.4% of Namib Desert soil phage OTUs were exclusive to a single sampling site. Other PGTs (31%) occurred in both Mexican and Argentinian bean fields, with an average ANI of 88% across regions. Moreover, phages of the same PGT exhibited an 8% higher ANI within regions than between regions on average. Highly similar phages have previously been found across multiple distant aquatic ecosystems around the world, in human virome samples [91, 92] and even in more distinct ecosystems [11]. The results indicate an emerging pattern in which a higher fraction of PGT members present a limited geographic range, while a significant minority of relatively closely related phages are distributed globally [16, 90, 93]. It remains unclear to what extent such observations are due to dispersal limitation [8]. The *Rhizobium* communities also showed spatial structure. *R. etli* was the predominant species nodulating common bean in the analyzed agricultural fields in Mexico, and *R. phaseoli*

predominated in the Argentinian fields, although both species were mainly characterized by spatially restricted STs.

Rhizobia from Tepoztlán were isolated from nodules collected directly from the bean field, while the rhizobia from the other bean fields were obtained by soil inoculation of growing beans in the greenhouse (see methods). This may have contributed to the differences observed in the composition of the rhizobia-phage communities. However, the latter is a common method for trapping rhizobia in rhizobium research to accurately sample rhizobium diversity [94, 95]. The specific technique used here was a slight variation of the method used in Van Cauwenberghe et al. [96]. Briefly, they found that rhizobial diversity and composition, was mostly identical to what was found using nodules collected directly from the same sampling sites in previous studies [97, 98]. Although the time of sampling was during the spring of 2016, the Yautepec samples were obtained 3 months later than those from Tepoztlán. This may have caused a confounding effect regarding to the compositional differences between these two communities. However, our analyses found only a significant difference in presence-absence (based on Jaccard dissimilarity) of phages between these two fields, which is probably also affected by various environmental differences (e.g., a 430 m altitude difference) and local historical contingencies. Moreover, we did not find significant differences in phage relative abundance (Bray–Curtis) and rhizobium presence-absence and relative abundance between the *Rhizobium* communities of Tepoztlán and Yautepec. Since chloroform was used during the phage isolation process, it may have excluded the membrane-containing phages or the recently described Autolykiviridae [82]. Although the results presented here suggest a minor role of these potential sources of variation, further specific experiments addressing them should be done. In particular, to extend the model to chloroform sensitive phages, that were not considered in this work.

Phage community (PGT) spatial patterns were correlated with the compositional differences among *Rhizobium* (ST) communities. Similar correlations between host-phage communities have been seen in aquatic systems [16, 17, 22, 99–101], which seems to verify the common assumption that the relative abundance of phages within a community depends largely on host abundance and susceptibility [8, 23]. However, the presence of susceptible rhizobium lineages did not imply the presence of specific phages or vice versa. For example, the omnipresent ST-5 was susceptible to all but two members of the phage lineages, but most phage lineages were spatially limited. Similarly, F06 phages could infect members of all STs, yet their presence was limited to Mexican bean fields. Although our study provides detailed genetic information on phages, the *Rhizobium* lineages were broadly defined by

housekeeping gene markers (*recA*, *dnaB*). It is expected that much of the still-uncharacterized phenotypic and genetic microdiversity within bacterial species [102] may explain the spatial heterogeneity of the PGT composition and host range patterns better than the presence or absence of a suitable host (defined by species or ST).

Host range breadth varied considerably among phages, from generalists to specialists, resulting in a nested structure of the inferred whole cross-interaction network. Based on the extensive analysis of experimental cross-infection data, Flores et al. [103] concluded that nestedness is the characteristic profile in most cases. A modular network structure may be significant at large phylogenetic scales [87], but genotype-to-genotype interactions are most frequent within narrow phylogenetic ranges and result from coevolutionary processes of susceptibility and resistance. High host genetic similarity may underlie the nested structure of our network.

Nevertheless, we found that phage-rhizobium interactions were significantly affected by the genetic identity of both phages and rhizobia as well as their geographic origin. This may explain the detection of three large modules with high internal connectivity and suggests ongoing local adaptation. Indeed, we showed that phages infecting common bean-nodulating rhizobia experienced higher infection rates in sympatric rhizobia than in allopatric rhizobia; hence, they were locally adapted. Furthermore, sympatric phages showed more similar host ranges than allopatric phages, and sympatric rhizobia shared similar susceptibility ranges. Only a few field studies have provided evidence that phages locally adapt to their bacterial hosts in nature [27, 87, 104]. Although the *Rhizobium* communities were generally maladapted to the local phages, they may be adapted to their local environment as a whole. A nodule may provide an isolated niche for rhizobia where they may survive competition with and the antagonistic effects of other bacteria or, more directly, phages. In free-living conditions, depredation by phages may change the population structure of rhizobia. Phages are usually considered to be slightly ahead of their host in their coevolutionary arms race due to the higher selective pressure they experience and their greater evolutionary potential [27, 105]. Although local bacterial adaptation to phages has been described multiple times in coevolutionary in vitro experiments [36, 106, 107], evidence in nature appears to be lacking [26]. This discrepancy is probably due to the relatively high availability of resources to hosts in vitro, which sways the arms race to the benefit of the host [108].

In our model, the degree of local adaptation was spatially inconsistent. Argentinean phages (mainly from Chicoana) infected approximately as many local as nonlocal rhizobia, while Mexican phages were more infectious in local rhizobia than in nonlocal rhizobia. Spatial asymmetry in phage local adaptation is believed to be the result of the effects of

nutrients on phage-host encounter rates, mutation rates and the cost of resistance [38, 109] or the local mode of coevolutionary dynamics (i.e., arms race or fluctuating selection [110]). Although we did not detect local phage maladaptation and we assume that the differences in productivity across the sampled actively cultivated bean fields might be relatively minor compared to those in media based in vitro experiments, these studies show how environmental differences create spatially different intensities of phage local adaptation. Indeed, the spatial heterogeneity of environmental factors results in a geographic mosaic of different evolutionary pressures [111]. Local adaptation to various environmental conditions can undermine the colonization success of allopatric individuals and limit gene flow (i.e., “isolation-by-adaptation,” or more general “isolation-by-environment” [33, 35, 109, 111]). Zhang & Buckling [34] found that host bacteria grown in the presence of phages in heterogeneous environments were limited in their ability to migrate across environments as a result of maladaptation. The limiting effect of local adaptation on phage migration has not been tested explicitly, although it has long been predicted [24, 109].

The spatial structure in the genetic composition of phage communities is probably due to the interplay of a variety of factors (e.g., historical contingencies, abiotic selection, genetic drift, and, potentially, dispersal limitation [8, 15]). Our results indicate that the presence of suitable hosts may play a role in shaping phage biogeography and that suitability is determined not only by the genetic identity of the host but also by local adaptation. The spatial patterns are analogous to those observed in *Rhizobium*-common bean interactions and suggest that the local adaptation of rhizobia to common bean may have shaped the spatial differences in the phage-rhizobium interactions. Through isolation-by-adaptation, local adaptation may reinforce spatial patterns in the PGT composition. Strong local adaptation of phages has been found across much shorter distances than in our present study [27, 104], and it is as yet unclear to what extent phage local adaptation leads to limited migration and at which scale this may occur. At smaller scales, spatial heterogeneity is probably under greater pressure due to higher viral migrant densities. However, across broad scales, local adaptation may be a significant barrier to successful long-distance phage migration.

Acknowledgements PAPIIT-UNAM IN209817 and the CCG-UNAM budget to VG funded the work. JVC received a Postdoctoral Scholarship from DGAPA-UNAM (2016-2018) and was partly supported by NSF grants DEB-1457508 and IOS-1759048, both awarded to E.L. Simms. We thank Gabriela Guerrero, José Espíritu, Alfredo Hernández, and Víctor del Moral for bioinformatics support, Alfonso Leija and Georgina Hernández for their help in greenhouse experiments, and Mario Marquina for providing access to the agricultural fields at Tepoztlán (Finca Xochitlamila) and Yauatepec. We thank Ellen Simms, Olga María Pérez Carrascal, and Ellie Harrison for comments on previous versions of the paper. We also thank two anonymous

reviewers whose suggestions helped to improve this paper. Special thanks go to Joshua Weitz for his advice on the BiMat application.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.* 2005;13:278–84.
- Hatfull GF. Dark matter of the biosphere: the amazing world of bacteriophage diversity. *J Virol.* 2015;89:8107–10.
- Bouvier T, Del Giorgio PA. Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ Microbiol.* 2007;9:287–97.
- Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H. Phage as agents of lateral gene transfer. *Curr Opin Microbiol.* 2003;6:417–24.
- Howard-Varona C, Hargreaves KR, Solonenko NE, Markillie LM, White RA, Brewer HM, et al. Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *ISME J.* 2018;12:1605–18.
- Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol.* 2004;6:1–11.
- Thurber RV. Current insights into phage biodiversity and biogeography. *Curr Opin Microbiol.* 2009;12:582–7.
- Chow C-ET, Suttle CA. Biogeography of viruses in the sea. *Annu Rev Virol.* 2015;2:41–66.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature.* 2016;537:689–93.
- Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. Φ CrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat Commun.* 2018;9:4781.
- Breitbart M, Miyake JH, Rohwer F. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett.* 2004;236:249–56.
- Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun.* 2014;5:4498.
- Jameson E, Mann NH, Joint I, Sambles C, Mühlhling M. The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. *ISME J.* 2011;5:1713–21.
- Delong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N, et al. Community genomics among stratified microbial assemblages in the ocean's interior. *Science.* 2006;311:496–503.
- Finke JF, Suttle CA. The environment and cyanophage diversity: insights from environmental sequencing of DNA polymerase. *Front Microbiol.* 2019;10:167.
- Hanson CA, Marston MF, Martiny JB. Biogeographic variation in host range phenotypes and taxonomic composition of marine cyanophage isolates. *Front Microbiol.* 2016;7:983.
- Huang S, Zhang S, Jiao N, Chen F. Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. *Appl Environ Microbiol.* 2015;81:441–52.
- Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJE, Martiny JBH. Marine cyanophages exhibit local and regional biogeography. *Environ Microbiol.* 2013;15:1452–63.
- Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature.* 2016;536:425–30.
- Winter C, Matthews B, Suttle CA. Effects of environmental variation and spatial distance on bacteria, archaea and viruses in sub-polar and arctic waters. *ISME J.* 2013;7:1507–18.
- Luo E, Aylward FO, Mende DR, Delong EF. Bacteriophage distributions and temporal variability in the ocean's interior. *mBio* 2017;8:e01903–17.
- Brum JR, Ignacio-espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, et al. Patterns and ecological drivers of ocean viral communities. *Science.* 2015;348:1261498.
- Dennehy JJ. What ecologists can tell virologists. *Annu Rev Microbiol.* 2014;68:117–35.
- Held NL, Whitaker RJ. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ Microbiol.* 2009;11:457–66.
- Ashby B, Boots M. Multi-mode fluctuating selection in host–parasite coevolution. *Ecol Lett.* 2017;20:357–65.
- Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* 2014;38:916–31.
- Vos M, Birkett PJ, Birch E, Griffiths RI, Buckling A. Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* 2009;325:833.
- Gomez P, Buckling A. Coevolution with phages does not influence the evolution of bacterial mutation rates in soil. *ISME J.* 2013;7:2242–4.
- Kraemer SA, Boynton PJ. Evidence for microbial local adaptation in nature. *Mol Ecol.* 2017;26:1860–76.
- Kawecki T, Ebert D. Conceptual issues in local adaptation. *Ecol Lett.* 2004;7:1225–41.
- Lenormand T. Gene flow and the limits to natural selection. *Trends Ecol Evol.* 2002;17:183–9.
- Nosil P, Egan SP, Funk DJ. Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. *Evolution.* 2008;62:316–36.
- Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol.* 2013;22:5983–99.
- Zhang Q-G, Buckling A. Migration highways and migration barriers created by host–parasite interactions. *Ecol Lett.* 2016;19:1479–85.
- Wang IJ, Bradburd GS. Isolation by environment. *Mol Ecol.* 2014;23:5649–62.
- Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc Biol Sci.* 2002;269:931–6.
- Kunin V, He S, Warnecke F, Peterson SB, Garcia Martin H, Haynes M, et al. A bacterial metapopulation adapts locally to

- phage predation despite global dispersal. *Genome Res.* 2008;18:293–7.
38. Lopez Pascua L, Gandon S, Buckling A. Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. *J Evol Biol.* 2012;25:187–95.
 39. Baumann P. Biology of endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol.* 2005;59:155–89.
 40. Levy A, Gonzalez IS, Mittelviehhaus M, Clingenpeel S, Paredes SH, Miao J, et al. Genomic features of bacterial adaptation to plants. *Nat Genet.* 2018;50:138–50.
 41. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science* 2005;307:1915–20.
 42. Heath KD, Tiffin P. Context dependence in the coevolution of plant and rhizobial mutualists. *Proc Biol Sci.* 2007;274:1905–12.
 43. Koch M, Delmotte N, Rehrauer H, Vorholt JA, Pessi G, Hennecke H. Rhizobial adaptation to hosts, a new facet in the legume root-nodule symbiosis. *Mol Plant Microbe Interact.* 2010;23:784–90.
 44. Aguilar OM, Riva O, Peltzer E. Analysis of *Rhizobium elii* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *Proc Natl Acad Sci.* 2004;101:13548–53.
 45. Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *N Phytol.* 2013;197:300–13.
 46. Koenig R, Gepts P. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. *Theor Appl Genet.* 1989;78:809–17.
 47. Melkonian R, Moulin L, Béna G, Tisseyre P, Chaintreuil C, Heulin K, et al. The geographical patterns of symbiont diversity in the invasive legume *Mimosa pudica* can be explained by the competitiveness of its symbionts and by the host genotype. *Environ Microbiol.* 2014;16:2099–111.
 48. Tian CF, Young JPW, Wang ET, Tamimi SM, Chen WX. Population mixing of *Rhizobium leguminosarum* bv. *viciae* nodulating *Vicia faba*: the role of recombination and lateral gene transfer. *FEMS Microbiol Ecol.* 2010;73:563–76.
 49. Burdon JJ, Thrall PH. Spatial and temporal patterns in coevolving plant and pathogen associations. *Am Nat.* 1999;153:S15–S33.
 50. Van Cauwenberghe J, Visch W, Michiels J, Honnay O. Selection mosaics differentiate *Rhizobium*-host plant interactions across nitrogen environments. *Oikos* 2016;125:1755–61.
 51. Guimarães PR, Pires MM, Jordano P, Bascompte J, Thompson JN. Indirect effects drive coevolution in mutualistic networks. *Nature* 2017;550:511–4.
 52. Heath KD, Lau JA. Herbivores alter the fitness benefits of a plant–rhizobium mutualism. *Acta Oecol.* 2011;37:87–92.
 53. Rogers HS, Buhle ER, HilleRisLambers J, Fricke EC, Miller RH, Tewksbury JJ. Effects of an invasive predator cascade to plants via mutualism disruption. *Nat Commun.* 2017;8:6–13.
 54. Delmas E, Besson M, Brice MH, Burkle LA, Dalla Riva GV, Fortin MJ, et al. Analysing ecological networks of species interactions. *Biol Rev.* 2019;94:16–36.
 55. Gaiarsa MP, Guimarães PR. Interaction strength promotes robustness against cascading effects in mutualistic networks. *Sci Rep.* 2019;9:1–7.
 56. Sih A, Crowley P, McPeck M, Petranka J, Strohmeier K. Predation, competition, and prey communities: a review of field experiments. *Annu Rev Ecol Syst.* 1985;16:269–311.
 57. Parratt SR, Barrès B, Penczykowski RM, Laine AL. Local adaptation at higher trophic levels: contrasting hyperparasite–pathogen infection dynamics in the field and laboratory. *Mol Ecol.* 2017;26:1964–79.
 58. Hatcher MJ, Dick JTA, Dunn AM. How parasites affect interactions between competitors and predators. *Ecol Lett.* 2006;9:1253–71.
 59. Hutchinson MC, Bramon Mora B, Pilosof S, Barner AK, Kéfi S, Thébault E, et al. Seeing the forest for the trees: putting multi-layer networks to work for community ecology. *Funct Ecol.* 2019;33:206–17.
 60. Koskella B, Taylor TB. Multifaceted impacts of bacteriophages in the plant microbiome. *Annu Rev Phytopathol.* 2018;56:361–80.
 61. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol.* 2010;8:317–27.
 62. Evans TJ, Ind A, Komitopoulou E, Salmond GPC. Phage-selected lipopolysaccharide mutants of *Pectobacterium atrosepticum* exhibit different impacts on virulence. *J Appl Microbiol.* 2010;109:505–14.
 63. Perez Carrascal OM, Vaninsberghe D, Juárez S, Polz MF. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. *Environ Microbiol.* 2016;18:2660–76.
 64. Santamaría RI, Bustos P, Sepúlveda-Robles O, Lozano L, Rodríguez C, Fernández JL, et al. Narrow-host-range bacteriophages that infect *Rhizobium elii* associate with distinct genomic types. *Appl Environ Microbiol.* 2014;80:446–54.
 65. Carlson K. Working with bacteriophages: common techniques and methodological approaches. In: Kutter E, Sulakvelidze A (eds). *Bacteriophages: biology and applications*. Boca Raton, FL: CRC Press; 2005. p. 437–94.
 66. Werle E, Schneider C, Renner M, Völker M, Fiehn W. Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Res.* 1994;22:4354–5.
 67. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
 68. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20.
 69. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-Cell sequencing. *J Comput Biol.* 2012;19:455–77.
 70. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18:821–9.
 71. Gordon D, Green P. ConSeq: a graphical editor for next-generation sequencing. *Bioinformatics* 2013;29:2936–7.
 72. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci Rep.* 2016;6:24373.
 73. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3 — new capabilities and interfaces. *Nucleic Acids Res.* 2012;40:e115.
 74. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci.* 2009;106:19126–31.
 75. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2016;8:12–14.
 76. Lopes A, Tavares P, Petit M, Guérois R, Zinn-Justin S. Automated classification of tailed bacteriophages according to their neck organization. *BMC Genom.* 2014;15:1027.
 77. Hyman P, Abedon ST. Phage host range and efficiency of plating. In: Clokie MRJ, Kropinski AM (eds). *Bacteriophages, methods and protocols*. Vol. I: Isolation, characterization, and interactions. Totowa, NJ: Humana Press; 2009. p. 175–202.
 78. Hyman P, Abedon ST. Bacteriophage host range and bacterial resistance. *Adv Appl Microbiol.* 2010;70:217–48.

79. Holmfeldt K, Solonenko N, Howard-Varona C, Moreno M, Malmstrom RR, Blow MJ, et al. Large-scale maps of variable infection efficiencies in aquatic Bacteroidetes phage-host model systems. *Environ Microbiol.* 2016;18:3949–61.
80. Ishizawa H, Kuroda M, Morikawa M, Ike M. Evaluation of environmental bacterial communities as a factor affecting the growth of duckweed *Lemna minor*. *Biotechnol Biofuels.* 2017;10:1–10.
81. Cenens W, Makumi A, Mebrhatu MT, Lavigne R, Aertsen A. Phage–host interactions during pseudolysogeny. *Bacteriophage* 2013;3:e25029.
82. Kauffman KM, Hussain FA, Yang J, Arevalo P, Brown JM, Chang WK, et al. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature.* 2018;554:118–22.
83. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Glinn D, et al. Community Ecology Package. <https://cran.r-project.org>, <https://github.com/vegandevs/vegan>. 2019.
84. Flores CO, Poisot T, Valverde S, Weitz JS. BiMat: a MATLAB package to facilitate the analysis of bipartite networks. *Methods Ecol Evol.* 2016;7:127–32.
85. Consul PC. A simple urn model dependent on predetermined strategy. *Sankhyā Indian J Stat Ser B.* 1974;36:391–9.
86. Borcard D, Legendre P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Modell.* 2002;153:51–68.
87. Flores CO, Valverde S, Weitz JS. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME J.* 2013;7:520–32.
88. Porter SS, Chang PL, Conow CA, Dunham JP, Friesen ML. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J.* 2016;11:248–62.
89. Greenlon A, Chang PL, Damtew ZM, Muleta A, Carrasquilla-Garcia N, Kim D, et al. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc Natl Acad Sci.* 2019;116:15200–9.
90. Scola V, Ramond JB, Frossard A, Zablocki O, Adriaenssens EM, Johnson RM, et al. Namib desert soil microbial community diversity, assembly, and function along a natural xeric gradient. *Micro Ecol.* 2018;75:193–203.
91. Short CM, Suttle CA. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl Environ Microbiol.* 2005;71:480–6.
92. Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, et al. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat Microbiol.* 2019;4:1727–36.
93. Culley AI, Steward GF. New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. *Appl Environ Microbiol.* 2007;73:5937–44.
94. Miranda-Sánchez F, Rivera J, Vinuesa P. Diversity patterns of Rhizobiaceae communities inhabiting soils, root surfaces and nodules reveal a strong selection of rhizobial partners by legumes. *Environ Microbiol.* 2016;18:2375–91.
95. Bontemps C, Rogel MA, Wiechmann A, Mussabekova A, Moody S, Simon MF, et al. Endemic *Mimosa* species from Mexico prefer alphaproteobacterial rhizobial symbionts. *N Phytol.* 2016;209:319–33.
96. Van Cauwenberghe J, Lemaire B, Stefan A, Efroze R, Michiels J, Honnay O. Symbiont abundance is more important than pre-infection partner choice in a *Rhizobium* – legume mutualism. *Syst Appl Microbiol.* 2016;39:345–9.
97. Van Cauwenberghe J, Michiels J, Honnay O. Effects of local environmental variables and geographical location on the genetic diversity and composition of *Rhizobium leguminosarum* nodulating *Vicia cracca* populations. *Soil Biol Biochem.* 2015;90:71–9.
98. Van Cauwenberghe J, Verstraete B, Lemaire B, Lievens B, Michiels J, Honnay O. Population structure of root nodulating *Rhizobium leguminosarum* in *Vicia cracca* populations at local to regional geographic scales. *Syst Appl Microbiol.* 2014;37: 613–21.
99. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the ‘core’ and ‘flexible’ Pacific Ocean Virome. *ISME J.* 2015;9:472–84.
100. Mühlhling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, et al. Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ Microbiol.* 2005;7:499–508.
101. Sun Y, Zhang S, Long L, Dong J, Chen F, Huang S. Genetic diversity and cooccurrence patterns of marine cyanopodoviruses and picocyanobacteria. *Appl Environ Microbiol.* 2018;84: e00591–18.
102. Chase AB, Arevalo P, Brodie EL, Polz MF, Karaoz U, Martiny JBH. Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. *mBio.* 2019;10:e02361–19.
103. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host – phage interactions. *Proc Natl Acad Sci.* 2011;108:E288.
104. Koskella B, Thompson JN, Preston GM, Buckling A. Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *Am Nat.* 2011;177:440–51.
105. Koskella B, Parr N. The evolution of bacterial resistance against bacteriophages in the horse chestnut phyllosphere is general across both space and time. *Philos Trans R Soc B Biol Sci.* 2015;370:20140297.
106. Morgan AD, Gandon S, Buckling A. The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* 2005;437:253–6.
107. Gómez P, Paterson S, De Meester L, Liu X, Lenzi L, Sharma MD, et al. Local adaptation of a bacterium is as important as its presence in structuring a natural microbial community. *Nat Commun.* 2016;7:12453.
108. Zhang Q-G, Buckling A. Resource-dependent antagonistic coevolution leads to a new paradox of enrichment. *Ecology* 2016;97:1319–28.
109. Lopez-Pascua LDC, Buckling A. Increasing productivity accelerates host-parasite coevolution. *J Evol Biol.* 2008;21:853–60.
110. Gurney J, Aldakak L, Betts A, Gougat-Barbera C, Poisot T, Kaltz O, et al. Network structure and local adaptation in co-evolving bacteria–phage interactions. *Mol Ecol.* 2017;26: 1764–77.
111. Thompson JN. *The geographic mosaic of coevolution*. Chicago, IL: Uni. Chicago Press; 2005.