RESEARCH

Growth rate determines prokaryote-provirus network modulated by temperature and host genetic traits

Zhenghua Liu[1,2], Qingyun Yan[3], Chengying Jiang[4], Juan Li[5], Huahua Jian[6], Lu Fan[7], Rui Zhang[8], Xiang Xiao[6], Delong Meng[1], Xueduan Liu[1], Jianjun Wang[2*] and Huaqun Yin[1*]

Abstract

Background: Prokaryote-virus interactions play key roles in driving biogeochemical cycles. However, little is known about the drivers shaping their interaction network structures, especially from the host features. Here, we compiled 7656 species-level genomes in 39 prokaryotic phyla across environments globally and explored how their interaction specialization is constrained by host life history traits, such as growth rate.

Results: We first reported that host growth rate indicated by the reverse of minimal doubling time was negatively related to interaction specialization for host in host-provirus network across various ecosystems and taxonomy groups. Such a negative linear growth rate-specialization relationship (GrSR) was dependent on host optimal growth temperature (OGT), and stronger toward the two gradient ends of OGT. For instance, prokaryotic species with an OGT $\geq$ 40 °C showed a stronger GrSR (Pearson's $r = -0.525$, $P < 0.001$). Significant GrSRs were observed with the presences of host genes in promoting the infection cycle at stages of adsorption, establishment, and viral release, but nonsignificant with the presence of immune systems, such as restriction-modification systems and CRISPR-Cas systems. Moreover, GrSR strength was increased with the presence of temperature-dependent lytic switches, which was also confirmed by mathematical modeling.

Conclusions: Together, our results advance our understanding of the interactions between prokaryotes and proviruses and highlight the importance of host growth rate in interaction specialization during lysogenization.

Keywords: Host-virus interaction, Growth rate, Specialization, Temperature, Infection cycle, Genetic traits

Introduction

The structure of prokaryote-virus interaction networks is closely related to ecosystem functions [1], such as biogeochemical cycles [2]. Specific host-virus pairs in networks are primarily established on the basis of their gene-for-gene coevolution [1, 3], during which host serving as prey are often initiators to constantly update their antivirus (predator) mechanisms for survival [4]. This forwardly evolutionary strategy of hosts not only narrows their spectrum of viral susceptibilities but also enforces viruses to participate in an arms race and become specialists, especially at large phylogenetic scales of host-virus interactions [5]. Such coevolutionary dynamics enhance the heterogeneity of species specialization in cross-infection patterns [1], which contributes to nonrandom networks with nested, modular, or nested-modular structures [5–7]. However, previous studies have focused mostly on the

*Correspondence: jjwang@niglas.ac.cn; yinhuaqun_cs@sina.com

[1] Key Laboratory of Biometallurgy of Ministry of Education, School of Minerals Processing and Bioengineering, Central South University, Changsha 410006, China

[2] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

Full list of author information is available at the end of the article

host range or specificity of viruses in exploring specialized host-virus interactions [6–8], while the host side is understudied. Complementary to the findings about viruses, the study of host features such as interaction specialization for host could provide a novel insight into host-virus network structure.

Host specialized on the set of viruses is primarily dependent on the alignments between its physiological status and the viral interests [9]. Such specific matches could lead to the phenomenon of the presence of proviruses in prokaryotic genomes being associated with host life history traits [9, 10], aside from molecular interactions from receptor to immune recognition [11]. The host growth rate could be one of the most important traits in determining host-virus interaction outcomes [12–16], and its high variability across species could result in the heterogeneous distributions of proviruses within certain prokaryotic clades [9]. Following this clue, we proposed that host specialization in host-virus interactions will be linked to the host growth rate based on the three lines of indirect evidences. First, an improved growth rate of the host could be accompanied with a reduced range of resistance to viruses [17, 18], which is partly contributed by the tradeoff between immunity and growth rate [19].

Second, high-host density stimulated by a high-growth rate cloud provide more chances of host cells encountering more virus species and enhances the virus adsorption rate [20], which may increase the probability in coinfection or even superinfection with diverse viruses in the host and generally modelled by Kill-the-Winner dynamic [21]. Third, slow-growing species favoring lysogeny [9] could allow proviruses to evolve competition strategies (e.g., surface modification) in host for inhibiting further viral infection [22]. Inspired by above findings, we hypothesized that prokaryotic species attaining a fast growth rate (fast grower) may show less specificity on virus species than slow growers, that is, growth rate-specialization relationships (GrSRs) (Fig. 1).

We further expect that GrSRs could be influenced by the host growth environment and genetic traits [4, 23, 24]. For example, in high-temperature environment, thermal stress has stronger effects on the host than its viruses [23], which would lead to the host changes in outer-membrane protein expression [25] and immune activities [26]. These host responses alter the host susceptibility or specificity to viral infection, revealing that a temperature-dependent infection structure is primarily established by the host [23]. In saline environments,
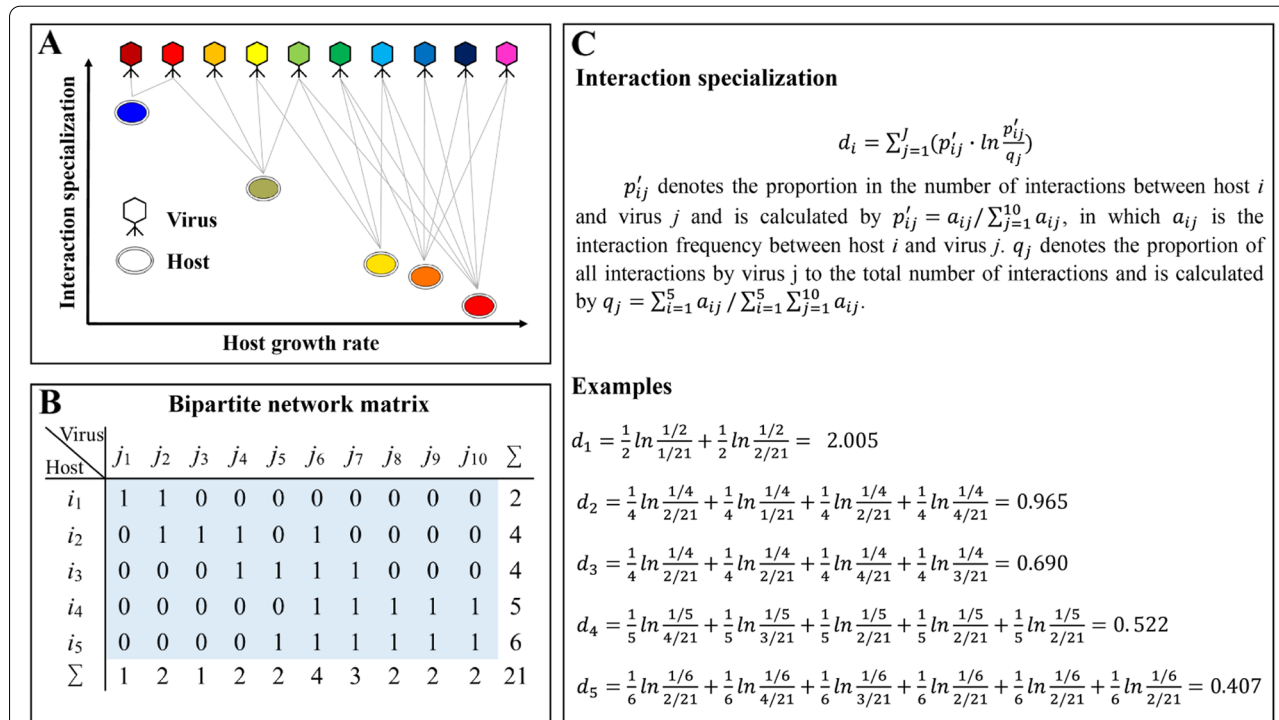

Fig. 1 Interaction specialization of host in host-virus network. **A** We hypothesized that interaction specialization for host in host-virus network is negatively related to host growth rate. It is noted that interaction specialization of host comprehensively considers both the number of different viral interactions and host range of viruses. Colors of triangles and circles represent different virus species and host species with gradient growth rates, respectively. **B** Adjacency matrix of host-virus interaction network (blue background) in plot **A**. **C** Formulation of interaction specialization of host in host-virus network, which is calculated by Shannon index *d* (see details in the 'Methods' section)

ironic strength not only affect the host physiological status [23] but also largely determines the infection activity of specialized viruses [27–30], implying the interactions of host-virus pairs could be constrained by environment salinity. Furthermore, host genetic traits are fundamental to their interaction specialization in infection networks during host-virus co-evolution. For instance, host resistance to viruses could shift with the evolution of some genetic traits, including the mutations in surface receptors, such as outer-membrane proteins and flagellum or pilin proteins [31], and the development of diverse immune systems, such as restriction-modification (RM) systems [32] and clustered regularly interspaced short palindromic repeats (CRISPR) and their associated protein (Cas) systems [33]. CRISPR-Cas systems, adaptive immune systems targeting specific viral sequences [34], could play a key role in shaping interaction specialization of host by mediating inefficient viral infection, or selectively inactivating viruses and allowing hosts to employ specific viruses during evolution for adaptation to diverse environments [35]. Despite the guidance of above findings, which and how environmental factors and host life history or genetic traits affect interaction specialization for host in prokaryote–virus networks are largely unknown and worth further investigations at global scale.

Compared with virulent viruses, proviruses as temperate viruses could increase host fitness such as growth rate [36] and temperature tolerance [37] by regulating gene expressions of host cell population at lysogenic state. This suggests that the interactions between host and proviruses are more dependent on host traits and environment factors than virulent viruses, and we therefore focused on the host-provirus interactions to investigate the drivers shaping interaction specialization for host. Here, we compiled 7656 species-level genomes from approximately 170,000 strain-level genomes of bacteria and archaea [38] across 15 ecosystems and 15 phyla and construct a host-provirus bipartite network based on the occurrences of proviruses in host genomes. We used Shannon index *d'* to quantify the interaction specialization for host within host-provirus network (Fig. 1). To uncover the role of host life history and genetic traits playing in its specialization in host-provirus interactions, we considered three main aspects of prokaryotic host: the growth rate estimated by the reverse of the minimal doubling time (*DT*), the optimal growth temperature (OGT), and the genes responsible for the infection cycle. We further established a mathematical model to explore the key determinants contributing to interaction specialization for host from the perspective of population dynamics. We had three main aims: (1) How does the host growth rate influences its interaction specialization across ecosystems and taxonomic groups? (2) What environmental factors or (3) genetic traits potentially modulate the effects of host growth rate on interaction specialization? Our results suggest that host growth rate plays a key role in shaping their interaction specialization during lysogenization, which is strengthened by high or low temperature and potentially constrained by molecular processes in the infection cycle.

Methods

Features of prokaryotes

We selected 7656 genomes as representative species from approximately 170,000 prokaryote strains complied by Madin et al. [38] and downloaded them from the NCBI according to accession numbers. The filtered genomes included 390 archaea and 7266 bacteria, and their completeness was greater than 90%. We considered species with the maximal genome size at the strain level as representative species. The natural environments were classified according to the protocols of Earth Microbiome Project Ontology (EMPO) [39]. Coding sequences in prokaryote genomes were also detected by Prokka v1.14.6 [40] with default parameters. The minimal doubling times were predicted based on ribosomal protein genes via the "predictGrowth" function in the "gRodon" v1.0.0 package [41], the reverse of which was used to indicate the host growth rate in units of doublings per day. The OGTs were predicted by software Tome v1.0 software [42] using species coding sequences with default parameters. Phylogenetic tree of all prokaryotic species was constructed by GToTree v1.6.12 software [43] based on 25 shared genes for bacteria and archaea (-H Bacteria_and_Archaea) with default parameters.

Proviruses

Proviruses in prokaryote genomes were identified by Phispy v4.1.22 based on AT and GC skews [44] with default parameters, and prophage genomes with lengths less than 3 kb were removed. To generate viral clusters at genus level, the remaining provirus genomes were clustered and taxonomically assigned by vConTACT2 v0.9.19 [45] against "ProkaryoticViralRefSeq201-Merged" reference via Diamond v0.9.24.125 [46] and ClusterONE (vcontact2 --rel-mode Diamond --pcs-mode MCL --vcs-mode ClusterONE). Provirus coding sequences were clustered by MMseqs2 v12.113e3 [47] with 70% identity and 70% coverage (mmseqs easy-linclust --min-seq-id 0.7 -c 0.7 --cov-mode 1). The completeness of provirus sequences were accessed by Checkv v0.8.1 software [48] with default parameters.

Surface receptor proteins

Host genes encoding phage protein receptors were searched for by DIAMOND against the phageReceptor database [49] with 50% identity.

Immunity systems

CRISPR arrays and Cas genes were identified by CRISPR-finder [50]. CRISPR-Cas systems were composed of CRISPR arrays and the closest Cas genes with distances less than 600 bp (bacteria: -cs -ccvr -vi 600; archaea: -ac -ccvr -vi 600). RM systems were identified by DIAMOND against the REBASE database [51] with 50% identity. The pair genes encoding restriction endonucleases and methylase were considered as a restriction-modification system when there were less than 5 genes between this pair genes.

Lytic switches

Genes encoding the repressors CI and Cro were searched for by DIAMOND against sequences collected from the UniProt database [52] with 50% identity. Genes encoding the repressors FpsA and FpsR were searched for against sequences provided by a previous study [53].

Species specialization of viral infections

A host-provirus network was constructed according to the coexistence of host and viral clusters. Since the inclusions of singletons and doubletons within host-provirus network would lead to a lot of perfect specialists during calculating interaction specialization d', we removed singletons and doubletons of hosts and viral clusters for downstream analyses. Interaction specialization of host was estimated by the standardized specialization index d' at host level within host-provirus network with I hosts × J viruses, which was carried out by the "dfun" function in the "bipartite" v2.16 package [54]. The value d' is derived from Shannon's diversity index [55], which represents how specialized a given species is in relation to available interacting partners [56]. It is given by following formula:

$$d_i = \sum_{j=1}^{J} \left(p'_{ij} \bullet \ln \frac{p'_{ij}}{q_j} \right) \text{ and}$$
$$d'_i = \frac{d_i - d_{min}}{d_{max} - d_{min}},$$

where d'_i represents standardized values derived from nonstandardized d_i of host i. p'_{ij} denotes the proportion in the number of interactions between host i and virus j, calculated by $p'_{ij} = a_{ij} / \sum_{j=1}^{J} a_{ij}$, in which a_{ij} is the interaction frequency between host i and virus j (presence: 1; absence: 0). q_j denotes the proportion of all interactions by virus j to the total number of interactions, given by $q_j = \sum_{i=1}^{I} a_{ij} / \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij}$. The value of d'_i ranges from 0 for extreme generalization to 1 for extreme specialization.

Mathematical model of viral infection dynamics regulated by temperature

The presented mathematical model (Fig. S1) is an extended combination of previous works [57, 58]. We considered a spatially homogeneous habitat of unitary volume with a maximum capacity C, in which B^+ and B^- are presented as the population density of nonsusceptible and susceptible cells to viruses, respectively, and P is the population density of viruses. For simplicity, we assumed that the growth of cells is described by the logistic function $\phi(N)$ multiplied by temperature fitness $\theta(T)$, where N and T are the population density of cells and environment temperature, respectively. We assumed that all viruses are ecological similar and adsorb to cells with an adsorption constant rate δ. The B^- infected by i viral species is presented as B_i^-, the probability of which is $Pr(X = i)$ and related to population density of cells and viruses. Here, we assumed B_0^- is lucky fellows escaping viral infection. The fraction α of B_i^- enters into lysogeny, whereas the remaining $1 - \alpha$ release β virus particles. The lysogen B_i^- could be induced to lyse and release β viral particles with an induction rate $\xi(T)$. For lysogen B_i^-, secondary infections are invalid and result in the loss of infecting viruses. The total cell population density is denoted by $N = B^+ + B^- + \sum_{i \neq 0} B_i^-$. It is noted that host death due to lysis was considered in the term of B^-. We assumed that there is a time delay (latent period) of τ between infection and lysis of lysogens.

Given the above assumptions, the model of viral infection dynamics regulated by temperature is given by delay differential equations, as follows:

$$\frac{dB^+(t)}{dt} = \underbrace{\varphi_+(N(t))\theta_+(T)B^+(t)}_{\text{Cell growth}},$$

$$\frac{dB^-(t)}{dt} = \underbrace{\varphi_-(N(t))\theta_-(T)B^-(t)}_{\text{Cell growth}} - \underbrace{\delta B^-(t)P(t)}_{\text{Infecting cell}} + \underbrace{\delta Pr(0|P(t), B^-(t), N(t))B^-(t)P(t)}_{\text{Cells escaping infection}},$$

$$\frac{dB_i^-(t)}{dt} = \underbrace{\varphi_-(N(t))\theta_-(T)B_i^-(t)}_{\text{Cell growth}} + \underbrace{\alpha\delta Pr(i|P(t), B^-(t), N(t))B^-(t)P(t)}_{\text{Lysogenization}} - \underbrace{\xi(T)B_i^-(t)}_{\text{Induction}}, i \neq 0,$$

$$\frac{dP(t)}{dt} = \underbrace{\sum_{i \neq 0}(1-\alpha)\delta\beta Pr(i|P(t), B^-(t), N(t)) B^-(t)P(t)}_{\text{Lysis}} + \underbrace{\sum_{i \neq 0} \beta\xi(T)B_i^-(t)}_{\text{Induction}} - \underbrace{\delta N(t)P(t)}_{\text{Adsorption}},$$

$$\frac{dN(t)}{dt} = \frac{dB^+(t)}{dt} + \frac{dB^-(t)}{dt} + \frac{dB_i^-(t)}{dt}.$$

The growth rates of B^+ and B^- are described by logistic functions of $\phi_+(N(t)) = V_+(1 - N(t)/C)$ and $\phi_-(N(t)) = V_-(1 - N(t)/C)$, respectively. The maximal growth rates of B^+ and B^- are presented by V_+ and V_-, respectively.

The temperature fitness of B^+ and B^- is described by Gaussian functions [59] of $\theta_+(T) = \exp\left(-\frac{(T-T_+)^2}{2\sigma_+^2}\right)$

and $\theta_-(T) = \exp\left(-\frac{(T-T_-)^2}{2\sigma_-^2}\right)$, respectively. The OGTs of B^+ and B^- are presented by T_+ and T_-, respectively. The temperature niche breadths of B^+ and B^- are presented by σ_+ and σ_-, respectively.

We assumed that host cells are infected by i viruses at time t occurs as a series of Bernoulli trials with the probability of success equal to p. Considering that the number of Bernoulli trials n is equal to viral particles and approximately infinite, the probability of a cell infected by i viruses at time t could be calculated by the Poisson formula and given by $\Pr(X=i) = \frac{\lambda^i}{i!}e^{-\lambda}, \lambda = np$. Here, λ is the average number of viruses infecting a host cell and fluctuating with the ratio of virus population density (P) to total cell population density (N). Since merely susceptible cells could be infected by viruses, the average number of viruses infecting a susceptible cell should be multiplied by the ratio of susceptible cell population density B^- to the total cell population density N and formulized by $\lambda = \left(\lambda_0 + \frac{P}{N}\right)B^-/N$. Here, λ_0 is the basic average number of viral species infecting host cell.

The temperature-dependent switch of lysogeny induction is described by a sigmoid function. We considered that there are heat- [60] and cold-activated switches [61], which are formulated by $f_h(T) = \frac{1}{1+\exp\left(k(T_h-T)\right)}$ and $f_c(T) = \frac{1}{1+\exp\left(k(T-T_c)\right)}$. T_h and T_c represent the temperatures of a half of the lysogeny events switching to lysis activated by heat and cold temperature, respectively. The parameter of k is a constant. Subsequently, the induction rate is described by $\xi(T) = \xi_0 \max(f_h(T), f_c(T))$, where ξ_0 is the induction rate constant.

Species specialization of host d is calculated by the following equation:

$$S = \sum_{i\neq0} B_i^-$$
$$d = \frac{1}{\sum_{i\neq0} \frac{B_i^- \times i}{S}}$$

where S is the population density of lysogens.

All constant parameters and initial states were listed in Tables S1 and S2, respectively. The system of delay differential equations was solved by Simon Wood's solv95 program, which was conducted with "PBSddesolve" package [62] in R v3.6 environment [63]. We deposited the codes of the model on Github repository https://github.com/zhenghualiu/GrSRs.

Statistical analyses

The linear relationships between host growth rate and interaction specialization d' (GrSR) were estimated by a linear regression model [64]. To control the effects of phylogenetic structure in linear model for each phylum,

phylogenetically corrected linear regression was further conducted by phylolm function with the Brownian motion model (BM) and the Pagel's *delta* model from "phylolm" v2.6.4 package [65]. For multiple testing analyses, P values were adjusted by Benjamini-Hochberg correction [66]. To detect the changes in GrSRs along with the OGT gradient, we used moving window approaches [67, 68] to locate sudden changes in the strength of relationships between host growth rate (log10-scaled) and d'. Specifically, dataset was divided into overlapping bins by 10°C (e.g., 10–20°C, 11–21°C, and 12–22°C until 70–80°C) and 20°C moving windows (e.g., 10–30°C, 11–31°C, and 12–32°C until 60–80°C) of OGT between 10 and 80°C. The Poisson distribution of the multiplicity of infections in lysogens was estimated by the maximum-likelihood method [69], and its significance was assessed by goodness-of-fit tests [70]. The above analyses were conducted in the R v3.6 environment [63].

Results and discussion

In total, we identified 11,798 provirus sequences in 5283 and 166 species-level genomes for bacteria and archaea, respectively. The number of proviruses in genomes followed the Poisson distribution with an average of 1.55 (Fig. S2). The genomic length of provirus ranged from 3.01 to 224.01 kb and averaged 32.94 ($\pm$ 19.66, SD) kb. We accessed the completeness of provirus genomes by CheckV [48] and found that 2549 sequences had completeness > 50% and 3447 sequences had completeness $\leq$ 50% while the others could not be determined. Based on gene-sharing network constructed by vContact2 [45], all provirus sequences could be classified into 1581 viral clusters, and 5.88% of which representing 7.65% of provirus sequences could be assigned with order-level taxonomies.

A host–provirus network was constructed based on the occurrence of proviruses in host genomes and used to further quantify interaction specialization of host (see details in the 'Methods' section). Interaction specialization of host was determined by the Shannon diversity index d and standardized by min-max scaling (d') [71]. It is noted that specialization d' (interaction specialization of host) comprehensively considers both the number of different viral interactions and host range of virus, and a higher value of d' indicates that host is more specialized to interact with viruses (Fig. 1). Since the singletons and doubletons would introduce 61% of potentially spurious proviruses into host-provirus network due to their lack of genome completeness (Fig. S3). If we included them in the calculation of interaction specialization d', it led to 31% of the most specialized host with an extreme d' value of 1 and
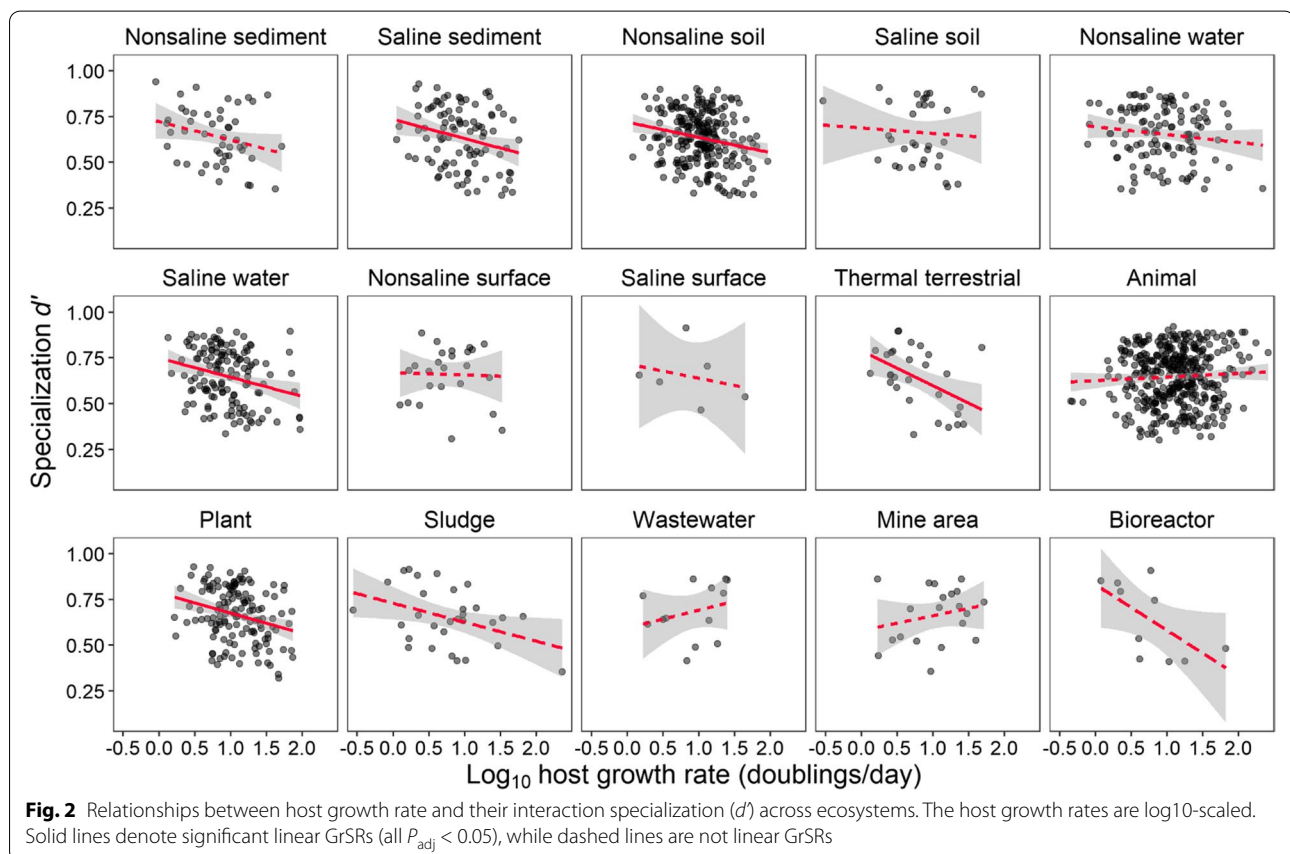
caused potential bias due to the non-normal distribution of *d'* (Fig. S4A). In contrast, the exclusion of singletons and doubletons could result in an approximately normal distribution of *d'* (Fig. S4B). Thus, to ensure the reliability in downstream analyses, we did not include the singletons and doubletons of viral clusters and host in host-provirus network. Consequently, we got a host–provirus bipartite network of 3115 hosts × 548 viral clusters with 4339 links, where each host species averagely interacted with 1.39 (± 0.68) viral clusters. Across the 39 phyla of bacteria and archaea, specialization *d'* averaged 0.64 (± 0.15) and was significantly correlated with the reverse of viral cluster number, with a weak correlation of −0.045 (df = 3054, $P = 0.013$; Fig. S5). Furthermore, specialization *d'* was significantly correlated with the copy number of the 5S rRNA, 16S rRNA, and 23S rRNA genes (all $P < 0.001$; Fig. S5), which indicates that interaction specialization of host is closely linked to the host growth rate ($r = −0.119$, $P < 0.001$; Fig. S5).

Host growth rate shapes interaction specialization

We found that there were significant linear relationships between host growth rate and specialization *d'* across various ecosystems (Figs. 2 and S6) and phyla (Fig. S7).
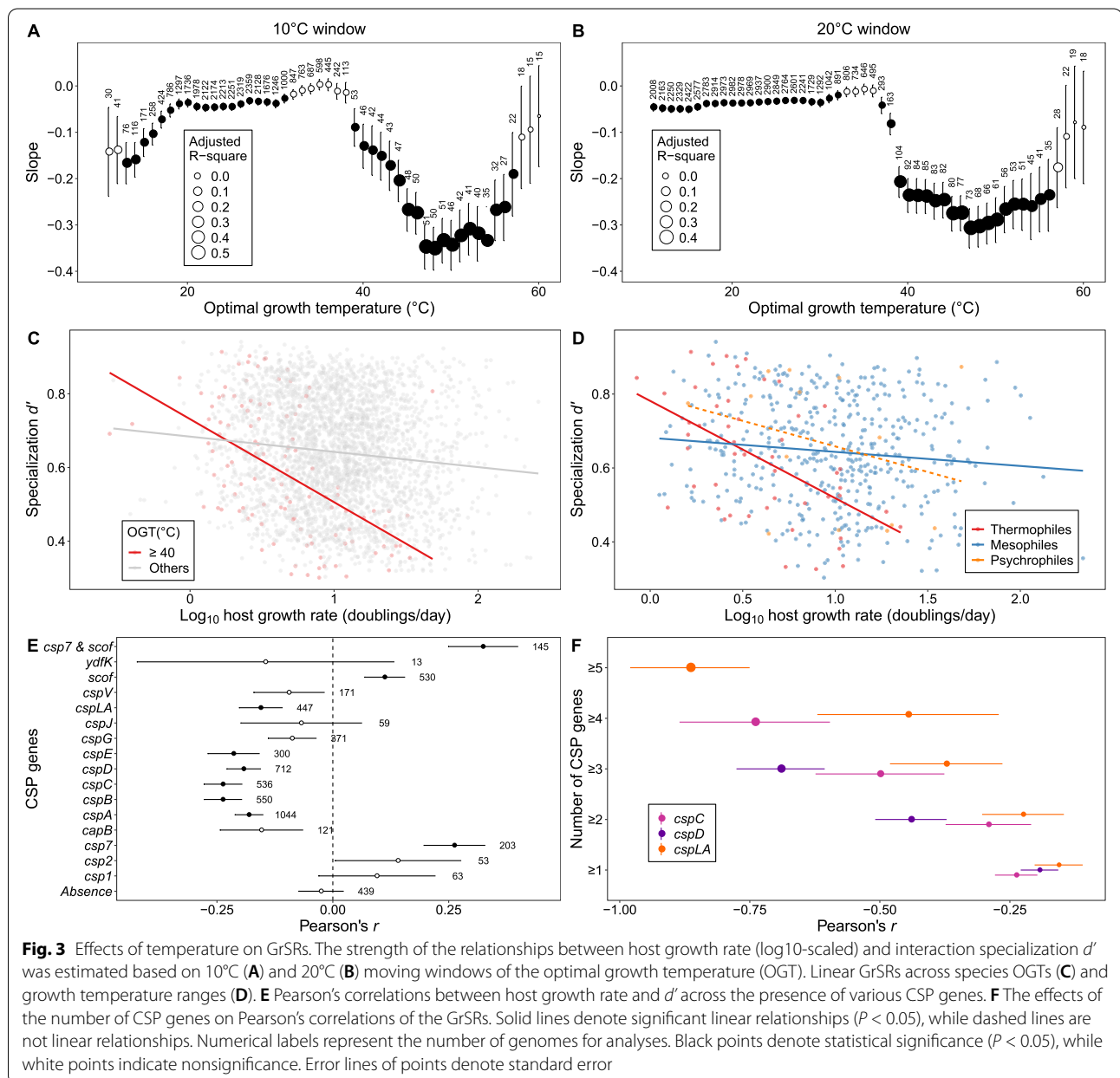
For instance, the negative linear growth rate-specialization relationships (GrSRs) were observed in natural environments such as terrestrial thermal systems ($R^2_{adj} = 0.19$, $P < 0.05$; Figs. 2 and S6). There was also significantly negative GrSRs in various phyla, such as Deinococcus-Thermus ($R^2_{adj} = 0.468$, $P < 0.05$; Fig. S7, Table S3). Notably, one exception of a positive GrSR occurred for the abundant phylum of Actinobacteria ($P < 0.05$; Fig. S7), which may be caused by its higher susceptibility to diverse viruses than other phyla [72]. Moreover, we included phylogenetic structure in linear regressions and still found significant ($P < 0.05$) relationships between host growth rate and their interaction specialization *d'* in host-provirus network (Table S4) for abundant phyla such as Actinobacteria ($n = 745$), Delta/epsilon subdivisions ($n = 60$), Firmicutes ($n = 776$) and Proteobacteria ($n = 1077$), which may be contributed by the phylogeny signal of growth rate (supplementary material Note 1). Taken together, these results indicate that GrSRs are negative in some natural environments and specific taxonomies, and thus, in which fast growers are likely to show less specificity to virus clusters than slow growers.

Interestingly, we found that temperature could influence the GrSRs in host-provirus interaction network, which was supported by the following three lines of

Fig. 2 Relationships between host growth rate and their interaction specialization (*d'*) across ecosystems. The host growth rates are log10-scaled. Solid lines denote significant linear GrSRs (all $P_{adj} < 0.05$), while dashed lines are not linear GrSRs

evidence. First, the slope or strength of negative linear GrSR in terrestrial thermal systems (slope = −0.19 ± 0.07, *P* = 0.01) was generally larger than that in other relatively low-temperature environments (all slopes < 0.12, *P* < 0.05; Figs. 2 and S6). Second, the GrSR strength indicated by the linear slope increased toward high or low OGTs. For example, the result of multiple linear regression analyses showed that growth rate (*P* < 0.001) and OGT (*P* = 0.005) had independent effects on interaction specialization *d'*, while their interaction term was nonsignificant (*P* = 0.803). Moreover, moving window analyses showed that the GrSR rapidly decreased until reaching a plateau phase when OGT gradually increased from 37

to 40°C, (Fig. 3A, B). Consequently, the GrSR strength in host species with OGTs ≥ 40 °C (−0.23 ± 0.04, *P* < 0.001) was stronger than that in other species (−0.04 ± 0.01, *P* < 0.001; Fig. 3C). This temperature range of 37 to 40 °C has been empirically shown to have a wide physiological significance in host-virus interactions by elevating the synthesis of heat shock protein [73] and cell flagellum protein [74], decreasing the activities of the CRISPR-Cas systems in mesophilic bacterium [26] and introducing the lysogeny-to-lysis transition [60]. These physiological statuses are conducive to promoting the infection cycle and thus may strengthen GrSRs. Similarly, the GrSRs gradually decreased when OGT decrease from 20 to 13°C

Fig. 3 Effects of temperature on GrSRs. The strength of the relationships between host growth rate (log10-scaled) and interaction specialization *d'* was estimated based on 10°C (**A**) and 20°C (**B**) moving windows of the optimal growth temperature (OGT). Linear GrSRs across species OGTs (**C**) and growth temperature ranges (**D**). **E** Pearson's correlations between host growth rate and *d'* across the presence of various CSP genes. **F** The effects of the number of CSP genes on Pearson's correlations of the GrSRs. Solid lines denote significant linear relationships (*P* < 0.05), while dashed lines are not linear relationships. Numerical labels represent the number of genomes for analyses. Black points denote statistical significance (*P* < 0.05), while white points indicate nonsignificance. Error lines of points denote standard error

(Fig. 3A). In this temperature range, the expression of cold shock protein [75] and the outer-membrane receptor OmpF was gradually upregulate [25]. Third, the significantly negative GrSRs were stronger in thermophiles (−0.26 ± 0.06, $P < 0.001$; Fig. 3D) than other groups, which was supported by that Deinococcus–Thermus with an OGT range of 50.9 to 66.8 °C have larger GrSR slope (−0.13 ± 0.05, $P = 0.05$; Table S3, Fig. S7) than other phyla, such as Bacteroidetes and Proteobacteria (Fig. S7).

Molecular mechanisms underlying GrSRs
To mechanistically understand the temperature-dependent GrSRs, we explored potential molecular processes from the perspectives of heat or cold shock responses and the key stages in the infection cycle. Our results showed that heat shock proteins (HSPs) or cold shock proteins (CSPs), protecting cells from lysis damage under high or low temperature stress [76, 77], respectively, potentially mediated the coupling of GrSRs. For HSPs, the presence of their genes (e.g., *hsp40*, *hsp70*, and *hsp100*) resulted in significantly negative GrSRs (Figs. S8 and S9), which were probably contributed by their chaperone activities for lysogeny development [78] and lysis [79].

For CSPs, the absence of their genes showed a non-significant GrSRs ($P > 0.05$; Figs. 3E and S10), while the presence of their genes, such as *csp7*, *cspA*, *cspB*, and *scof*, generally led to significant GrSRs (all $P < 0.05$; Figs. 3E and S10). Furthermore, different CSP genes would cause negative or positive GrSRs. For example, on the one hand, GrSRs were significantly negative for the presence of *cspC*, *cspD*, and *cspLA* (all $P < 0.05$; Figs. 3F and S10), and the higher number of these three CSP genes would result in stronger negative GrSRs (Figs. 3F and S10), which may be associated with the upregulated expression of CSP genes during viral infections [80]. On the other hand, GrSRs were positive for the presence of *csp7* and *scof* genes ($P < 0.05$; Figs. 3E and S10) that were primarily found in Actinobacteria genomes, which implies that both two genes are key genetic traits for the positive GrSR among Actinobacteria. However, at the order level, the abundant groups with *csp7* genes, including Micrococcales, Rhizobiales, Streptomycetales, and Streptosporangiales (all $P > 0.1$; Fig. S11), have no significant GrSR, while Micrococcales with *scof* showed a significantly positive GrSR ($P < 0.05$; Fig. S12). These results imply that the positive GrSR could be taxonomically dependent regarding specific CSP genes, which needs to be further supported by studying more genomes. Overall, we revealed that HSPs and CSPs contribute to the coupling of GrSRs, which may be due to their molecular chaperone activities assisting viral reproduction.

Considering that lysis–lysogeny decision of lysogens is obviously influenced by temperature [53, 60], we further investigated how GrSRs are constrained by the genes responsible for the processes in the infection cycle, including key stages of virus adsorption, establishment, and release. For the adsorption stage, host receptors are diverse and play a key role in the specific matching of the host-viral pair [81]. Throughout all receptor proteins, we found that only the presence of genes encoding the flagellum protein FliC or FljB and pilin protein MshA would lead to a significantly negative GrSR (all $P < 0.05$; Fig. 4, Table S5), while this was not the case for outer membrane proteins ($P > 0.05$; Table S5). Compared with the outer membrane proteins, the structural proteins of flagellum and pilus not only serve as viral receptors but also enhance cell motility and form biofilm formation [82] that increase host-virus encounters [16]. For the establishment stage, most of host defense systems generally decoupled GrSRs, including RM systems of types I and III ($P > 0.05$; Figs. 4 and S13) and CRISPR-Cas systems of types I, II, and III ($P > 0.05$; Figs. 4 and S14). This is likely because host defense prevents virus insertion into the host genome and blocks the infection cycle, which is partly supported by the effects of some anti-CRISPR proteins, which resulted in significant GrSRs (all $P < 0.05$; Figs. 4 and S15). For the release stage, considering temperature-dependent GrSRs, we primarily focused on the temperature-dependent lytic switches, such as the CI/Cro repressor [60] and newly identified FpsR [53, 61] for high (≥ 36 °C) and low temperature inductions (≤4 °C), respectively. As expected, the presence of repressor genes, including *cI/cro* or *fpsR*, led to a significantly negative GrSR ($P < 0.05$; Figs. 4 and S16), while their absence decoupled GrSRs ($P > 0.05$; Fig. S16). Together, these results highlighted that GrSRs were strengthened by the molecular processes in promoting the infection cycle at the stages of adsorption, establishment, and viral release ($R^2_{adj} = 0.182$, $P < 0.001$; Fig. S17), but were decoupled by immune systems.

Mathematical modeling to explain GrSRs
To explore how temperature and population dynamic features modulate the GrSRs, we developed a mathematical model of the population dynamic in host-viral systems using realistic parameters (Fig. S1, Tables S1 and S2) based on the previous frameworks of delay differential equations [57, 58]. Briefly, host-virus systems contained five compartments including susceptible cells, nonsusceptible cells, viruses, infected cell in lysogeny, and lytic state. The transitions between compartments dependent on host-virus infection dynamic in the infection cycle. We assumed that the processes of multiple infection are Bernoulli trails and the induction rate of lysogens is

Fig. 4 Effects of genes on the negative GrSRs throughout the infection cycle, including the stages of adsorption, establishment, and viral release. The negative GrSRs were significantly ($P < 0.05$) strengthened by the presence of genes responsible for host viral receptors of flagellum and pili, temperature-dependent lytic switches, and phage anti-CRISPR systems but decoupled by host immune systems, including the CRISPR-Cas and RM systems. CRISPR-Cas system: clustered regularly interspaced short palindromic repeats arrays (CRISPR) and CRISPR-associated protein (Cas) system. RM system, restriction-modification system; HSP, heat shock protein; CSP, cold shock protein

temperature dependent. The temperature thresholds for heat and cold inductions were set as 37°C and 4°C (Fig. S18), respectively. It should be noted that the host growth rate in the modeling system was dependent on the environmental temperature and was not the theoretical maximum value.

To simplify the questions, interaction specialization of host in modeling system was estimated by the reverse of number of viral species. Our simulation results confirmed the empirical observations showing that thermophiles have a stronger GrSR than mesophiles (Figs. 3D and 5), which may be caused by insufficient virus particles to infect susceptible cells due to the low induction rate at medium temperature. Meanwhile, there was a lag of GrSR curves between thermophiles and psychrophiles toward high host growth rate (Fig. 5), which was consistent with the empirical results that GrSRs in potential psychrophiles harboring CSPs were lagger than those in thermophiles with OGTs ≥ 40°C (Fig. S19). This phenomenon implied that the coupling of GrSRs at low temperature may require a fast growth rate as compensation when the species OGT is greater than the environment temperature of cold induction. Furthermore, GrSR strength were regulated by the temperature-dependent lytic switches. When such lytic switches were removed

from the host-virus system, the strengths of GrSR were decreased for thermophiles and psychrophiles (Fig. 5).

Challenges in testing growth rate-specialization relationships

Based on the solid results of both genomic and modeling analyses, we found that the linear GrSRs are widely existed in Earth's ecosystems, and its strength is influenced by environment types, host phylogenetic lineages, and traits, which implies that the host-virus interaction specialization at ecosystem levels may be further constrained by environmental factors, such as temperature and nutrient [56]. To confirm and extend these findings, there are at least four challenges to be considered in future studies. First, the GrSR could be further supported by the inclusion of unculturable strains by culture-independent sequencing technology, such as combined use of flow cytometry and single cell sequencing. All lysogenic genomes in this study were culturable isolates compiled from the presently largest dataset [38], which gives an unprecedented scale to investigate the prokaryote-provirus interaction network. Nevertheless, the GrSR also needs an extension to unculturable strains, which may provide novel insights into infection structure due to the unculturable host relying on intercellular metabolic
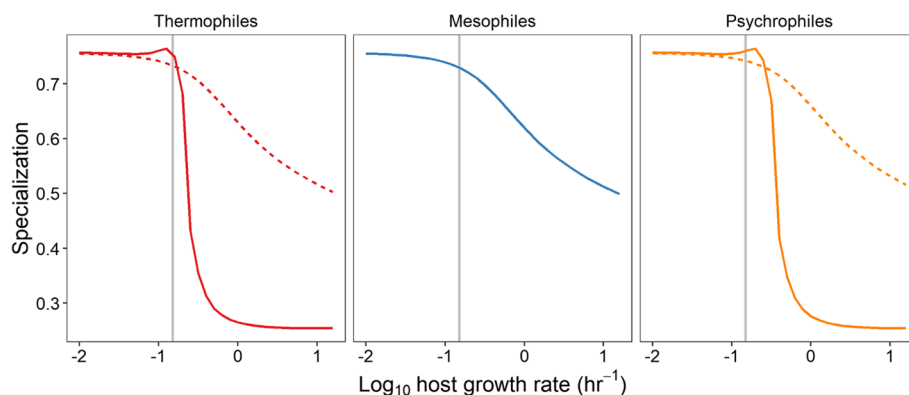
Fig. 5 Numerical solutions for the growth rate-specialization relationships (GrSRs) of thermophiles, mesophiles, and psychrophiles. The solid curves represent GrSRs simulated by the mathematical model with a temperature-dependent lytic switch, while dashed curves have no this lytic switch. Thermophiles: OGT = 45°C, environmental temperature = 37°C. Mesophiles: OGT = 30°C, environmental temperature = 30°C. Psychrophiles: OGT = 20°C, environmental temperature = 4°C. Gray vertical lines are the growth rate of nonsusceptible cell population. All parameters used to run the model are listed in Tables S1 and S2

networks [83]. Notably, we have conducted similar analyses for metagenome assembled genomes (MAGs) collected from diverse habitats covering all of Earth's continents and oceans [84]. Compared to isolate genomes, we found that the results of MAGs were far from reliable due to the following reasons: (1) The reconstruction of MAGs is largely dependent on tetranucleotide frequency, which would leave out many fragments of provirus genome in MAGs. (2) The incompleteness and contamination of MAGs would decrease the prediction accuracy of maximal growth rate and optimal growth temperature. Thus, we expected that metagenome assembled genomes were not suitable for exploring the relationships between host growth rate and their interaction specialization *d'* among unculturable strains.

Second, the host maximal growth rate could be estimated by culture-dependent approaches but it is highly variant even under optimal growth conditions and in the absence of interspecific competition [85]. Our results are mainly based on the maximum host growth rate, which was theoretically predicted by the reverse of the minimal doubling time based on the number of genes encoding ribosomal proteins and the codon usage frequency [85]. In experimental systems, we could use the actual host growth rate in a given environment, like our modeling system, to test whether it has a similar relation to interaction specialization of host during lysogenization. However, we should note that the experimental measured maximum host growth rates may largely dependent on the culture conditions and thus would be different from theoretical maximum host growth rates.

Third, it is challenging to establish and maintain a coculture system containing host species across a large phylogenetic scale. We used a host-provirus network consisting of 3115 host species across 39 phyla to investigate the GrSRs and found that growth rate-specialization relationships were phylogenetically scaled and easier to be found at the higher taxonomic level. For instance, such relationships were significant for 44% of phyla (4 out of 9), 35% of classes (7 out of 12), 19% of orders (10 out of 53), 9% of families (8 out of 88), and 5% of genera (3 out of 56). Thus, GrSR is likely to be found for the groups with common genetic traits rather than specifically taxonomic groups.

Finally, gene manipulations could be used to verify the effects of host genetic traits on GrSRs but it is high cost to develop genetic manipulation system for a large number of species. Alternatively, host strains with a gradient of growth rates belonging to a certain taxonomic group that display GrSRs in genomic analyses could be employed to explore more genes or molecular processes influencing GrSRs. Overcoming the above challenges is key to uncovering the global pattern in the prokaryote-provirus interaction network, which will lead the future researches in host-virus interaction specialization on during host-virus coevolution.

Conclusions

Exploring the drivers shaping prokaryote-virus network pattern is one of the most important issues in virus ecology. Our study highlights that the prokaryote growth rate play a key role in determining interaction specialization for host in prokaryote-provirus network. The negative growth rate-specialization relationships are widespread

in the Earth's microbiome. Meanwhile, such relationships are temperature-dependent and strengthened by the presence of host genetic traits promoting the infection cycle at the stages of adsorption, establishment, and viral release, but are decoupled by immune systems. Overall, these results help us to uncover the determinants of prokaryote-virus interactions and mechanistically understand their interaction specialization.

Supplementary Information

Additional file 1: Table S1. Model default parameters. **Table S2.** Model initial conditions. **Table S3.** Linear regression model for host growth rate and their virus specificity across different phyla. **Table S4.** Phylogenetic linear regression model for host growth rate and their virus specificity across different phyla. **Table S5.** Effects of the host surface receptors on the GrSRs estimated by linear regression analyses. **Figure S1.** Schematic diagram of modeling host-virus population dynamic. There are five compartments including susceptible cell (B^-, B_0^-), non-susceptible cell (B^+), infected host in lysogeny (B_i^-, $i \neq 0$) and lytic state (L_i) states, virus (P). Arrows denote the transitions between compartments. More details can be saw in the Methods. **Figure S2.** Distribution of the viral cluster number across all species genomes. Such distribution follows the Poisson distribution with an expect value of 1.55. Bar plot are observed values and blue points are predicted values. **Figure S3.** The proportions of proviruses with unavailable genome completeness after the removal of the nodes with low network degrees in the bipartite networks. We considered only the proviruses with unavailable genome completeness for all proviruses or the proviruses of host-provirus networks, and found that the proportion of these proviruses decreased until the network degree $\geq$ 3. The degrees of 1 and 2 are referring the nodes of singletons and doubletons, respectively. **Figure S4.** The distribution of interaction specialization d' for host in the whole host-provirus network (A) and the host-provirus network without singletons and doubletons (B), and the skewness and kurtosis of frequency distribution in the panel B (C) and the panel A (D). Notably, the skewness and kurtosis of frequency distribution indicate the distribution in the panel B is closer to normal distribution than that in the panel A. **Figure S5.** The pairwise correlations among variables. The diagonal plots are the frequency distributions of variables. The plots in upper triangular matrix show the Pearson's correlation between variables. d': interaction specialization for host. GS: genome size (Mb). GC: GC-content %. CDS: the number of gene coding sequences. 5S rRNA: the number of 5S ribosome RNA. 16S rRNA: the number of 16S ribosome RNA. 23S rRNA: the number of 23S ribosome RNA. 1/NVC: the reverse of number of viral clusters. Gr: host growth rate (doublings/day). OGT: optimal growth temperature (°C). *: P < 0.05. **: P < 0.01. ***: P < 0.001. **Figure S6.** Slopes of the GrSRs across different ecosystems. The slope of GrSR for all environments was estimated by a mixed effect model: $d' \sim \log_{10}(1/DT) + (1 + \log_{10}(1/DT) \mid \text{ecosystem})$. The slope of GrSR for each environment was estimated by linear model: $d' \sim \log_{10}(1/DT)$. Black points denote significant linear relationships (all P_{adj} < 0.05), while white points are nonsignificant. Point size represents adjust R-square. **Figure S7.** The GrSRs across phyla. Solid lines are significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S8.** Effects of the genes encoding heat shock proteins (HSP) on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S9.** The GrSR slopes across the number of genes encoding heat shock proteins (HSP). Numerical labels represent the number of genomes for analyses. Black points denote significant linear relationships (P_{adj} < 0.05), while grey points are nonsignificant. Error lines denote standard error. **Figure S10.** Effects of the genes encoding cold shock proteins (CSP) on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S11.** Effects of the *csp7* genes on GrSRs across main orders. Dashed lines denote nonsignificant linear relationships (all P >

0.1). **Figure S12.** Effects of the *scoF* genes on GrSRs across various orders. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S13.** Effects of the RM systems on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S14.** Effects of the CRISPR-Cas systems on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S15.** Effects of the anti-CRISPR systems on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S16.** Effects of the temperature-dependent lytic switches on GrSRs. Solid lines denote significant linear relationships (P_{adj} < 0.05), while dashed lines are nonsignificant. **Figure S17.** GrSRs for those species harboring the genes involving in the molecular processes in promoting the infection cycle at the stages of adsorption, establishment and viral release (R^2_{adj} = 0.182, P < 0.001, df = 154). Those species harbor at least one gene encoding heat shock protein (*hsp20*, *hsp40*, *hsp70* or *hsp100*), at least two genes encoding cold shock protein (*cspC*, *cspD*, or *cspLA*), at least one gene encoding receptor protein (*fliC*, *fljB* or *mshA*), at least one pair of genes encoding heat lytic switch (*cI* and *cro*) and at least one gene encoding cold lytic switch (*fpsR*). **Figure S18.** Temperature-dependent induction. It is controlled by two lytic switches including the repressors of CI/Cro (for high temperature at 37°C) and FpsR (for low temperature at 4 °C), which is simulated by a Sigmond function. **Figure S19.** Significant linear GrSRs of the potential psychrophiles harboring at least three CSP genes and the thermophiles with OGT larger than 40°C. All P_{adj} < 0.05.

Authors' contributions

H.Y. and J.W. conceived and designed the works. Z.L. and Q.Y. performed the bioinformatic analyses. Z.L. and J.W. performed the statistical analyses. Z.L., J.W., and H.J. constructed the mathematical model. Z.L. wrote the original draft manuscript. J.W., H.Y., C.J., J.L., H.J., L.F., R.Z., X.X., D.M., and X.L. reviewed and edited the manuscript. The authors read and approved the final manuscript.

Availability of data and materials

The codes and datasets generated during the current study are available in the online repository, https://github.com/zhenghualiu/GrSRs.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Key Laboratory of Biometallurgy of Ministry of Education, School of Minerals Processing and Bioengineering, Central South University, Changsha 410006, China. [2]State Key Laboratory of Lake Science and Environment, Nanjing

Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China. [3]Environmental Microbiomics Research Center, School of Environmental Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China. [4]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. [5]College of Agronomy, Hunan Agricultural University, Changsha 410125, China. [6]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. [7]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China. [8]State Key Laboratory of Marine Environmental Science, The Institute of Marine Microbes and Ecospheres, Xiamen University, Xiamen 361102, China.

References

1. Weitz JS, et al. Phage-bacteria infection networks. Trends Microbiol. 2013;21(2):82–91.
2. Jover LF, et al. The elemental composition of virus particles: implications for marine biogeochemical cycles. Nat Rev Microbiol. 2014;12(7):519–28.
3. Agrawal A, Lively CM. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol Ecol Res. 2002;4(1):91–107.
4. Rostol JT, Marraffini L. (Ph)ighting phages: how bacteria resist their parasites. Cell Host Microbe. 2019;25(2):184–94.
5. Beckett SJ, Williams HT. Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Interface Focus. 2013;3(6):20130033.
6. Valverde S, et al. Coexistence of nestedness and modularity in host-pathogen infection networks. Nat Ecol Evol. 2020;4(4):568–77.
7. Jover LF, Cortez MH, Weitz JS. Mechanisms of multi-strain coexistence in host-phage systems with nested infection networks. J Theor Biol. 2013;332:65–77.
8. Roux S, et al. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife. 2015;4:e08490.
9. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J. 2016;10(11):2744–54.
10. Wang X, et al. Phage combination therapies for bacterial wilt disease in tomato. Nat Biotechnol. 2019;37(12):1513–20.
11. Rothenburg S, Brennan G. Species-specific host–virus interactions: implications for viral host range and virulence. Trends Microbiol. 2020;28(1):46–56.
12. Thingstad T, Lignell R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat Microb Ecol. 1997;13(1):19–27.
13. Rodriguez-Brito B, et al. Viral and microbial community dynamics in four aquatic environments. ISME J. 2010;4(6):739–51.
14. Yoshida M, et al. Ecological dynamics of the toxic bloom-forming cyanobacterium Microcystis aeruginosa and its cyanophages in freshwater. Appl Environ Microbiol. 2008;74(10):3269–73.
15. Knowles B, et al. Lytic to temperate switching of viral communities. Nature. 2016;531(7595):466–70.
16. Silveira CB, Rohwer FL. Piggyback-the-winner in host-associated microbial communities. NPJ Biofilms Microbiomes. 2016;2:16010.
17. Gómez P, Buckling A. Bacteria-phage antagonistic coevolution in soil. Science. 2011;332(6025):106–9.
18. Avrani S, Lindell D. Convergent evolution toward an improved growth rate and a reduced resistance range in Prochlorococcus strains resistant to phage. Proc Natl Acad Sci U S A. 2015;112(17):E2191–200.
19. Weissman JL, et al. Immune loss as a driver of coexistence during host-phage coevolution. ISME J. 2018;12(2):585–97.
20. Luque A, Silveira CB. Quantification of lysogeny caused by phage coinfections in microbial communities from biophysical principles. mSystems. 2020;5(5):e00353–20.
21. Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr. 2000;45(6):1320–8.
22. Bondy-Denomy J, et al. Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 2016;10(12):2854–66.
23. Mojica KD, Brussaard CP. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. FEMS Microbiol Ecol. 2014;89(3):495–515.
24. Wang J, et al. Embracing mountain microbiome and ecosystem functions under global change. New Phytol. 2022. https://doi.org/10.1111/nph.18051.
25. Leon-Velarde CG, et al. Yersinia enterocolitica-specific infection by bacteriophages TG1 and varphiR1-RT is dependent on temperature-regulated expression of the phage host receptor OmpF. Appl Environ Microbiol. 2016;82(17):5340–53.
26. Hoyland-Kroghsbo NM, Munoz KA, Bassler BL. Temperature, by controlling growth rate, regulates CRISPR-Cas activity in Pseudomonas aeruginosa. mBio. 2018;9(6):e02184–18.
27. Keynan A, et al. Marine transducing bacteriophage attacking a luminous bacterium. J Virol. 1974;14(2):333–40.
28. Kukkaro P, Bamford DH. Virus–host interactions in environments with a wide range of ionic strengths. Environ Microbiol Rep. 2009;1(1):71–7.
29. Bettarel Y, et al. Ecological traits of planktonic viruses and prokaryotes along a full-salinity gradient. FEMS Microbiol Ecol. 2011;76(2):360–72.
30. Cordova A, et al. Osmotic shock and the strength of viral capsids. Biophys J. 2003;85(1):70–4.
31. Seed KD, et al. Evolutionary consequences of intra-patient phage predation on microbial populations. Elife. 2014;3:e03497.
32. Tock MR, Dryden DT. The biology of restriction and anti-restriction. Curr Opin Microbiol. 2005;8(4):466–72.
33. Makarova KS, et al. Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol. 2011;9(6):467–77.
34. Makarova KS, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020;18(2):67–83.
35. Zheng Z, et al. The CRISPR-Cas systems were selectively inactivated during evolution of Bacillus cereus group for adaptation to diverse environments. ISME J. 2020;14(6):1479–93.
36. Wang X, et al. Cryptic prophages help bacteria cope with adverse environments. Nat Commun. 2010;1:147.
37. Sy BM, Lan R, Tree JJ. Early termination of the Shiga toxin transcript generates a regulatory small RNA. Proc Natl Acad Sci U S A. 2020;117(40):25055–65.
38. Madin JS, et al. A synthesis of bacterial and archaeal phenotypic trait data. Sci Data. 2020;7(1):170.
39. Thompson LR, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457–63.
40. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
41. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci. 2021;118(12):e2016810118.
42. Li G, et al. Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. ACS Synth Biol. 2019;8(6):1411–20.
43. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019;35(20):4162–4.
44. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 2012;40(16):e126.
45. Jang HB, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37(6):632–9.
46. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.
47. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026–8.
48. Nayfach S, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39(5):578–85.
49. Zhang Z, et al. Phage protein receptors have multiple interaction partners and high expressions. Bioinformatics. 2020;36(10):2975–9.

50. Symeonidi E, et al. CRISPR-finder: a high throughput and cost-effective method to identify successfully edited Arabidopsis thaliana individuals. Quant Plant Biology. 2021;2:e1.

51. Roberts RJ, et al. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015;43(D1):D298–9.

52. Consortium, U. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(D1):D204–12.

53. Jian H, et al. Multiple mechanisms are involved in repression of filamentous phage SW1 transcription by the DNA-binding protein FpsR. J Mol Biol. 2019;431(6):1113–26.

54. Dormann CF, Gruber B, Fründ J. Introducing the bipartite package: analysing ecological networks. Interaction. 2008;1(0.2413793):8–11.

55. Blüthgen N, Menzel F, Blüthgen N. Measuring specialization in species interaction networks. BMC Ecol. 2006;6(1):1–12.

56. Hu A, et al. Quantifying microbial associations of dissolved organic matter under global change. bioRxiv. 2021.

57. Pleska M, et al. Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. Nat Ecol Evol. 2018;2(2):359–66.

58. Egilmez HI, et al. Temperature-dependent virus lifecycle choices may reveal and predict facets of the biology of opportunistic pathogenic bacteria. Sci Rep. 2018;8(1):9642.

59. Hawkes CV, Keitt TH. Resilience vs. historical contingency in microbial responses to environmental change. Ecol Lett. 2015;18(7):612–25.

60. Bednarz M, et al. Revisiting bistability in the lysis/lysogeny circuit of bacteriophage lambda. PLoS One. 2014;9(6):e100876.

61. Meng C, et al. The thermo-regulated genetic switch of deep-sea filamentous phage SW1 and its distribution in the Pacific Ocean. FEMS Microbiol Lett. 2020;367(12):fnaa094.

62. Schnute, J., A. Couture-Beil, R. Haigh, A users guide to the R package PBSddesolve. Technical report, 2013. Version 1.10.

63. Team, R.C. R language definition. Vienna: R foundation for statistical computing; 2000.

64. Chambers J, Hastie T. Linear models. Chapter 4 of statistical models in S. New York: Wadsworth & Brooks/Cole; 1992. https://doi.org/10.1201/9780203738535.

65. Ho, L.S.T., et al., Package 'phylolm'. 2016. See http://cran.r-project.org/web/packages/phylolm. Accessed Feb 2018.

66. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289–300.

67. Cohen JM, et al. The thermal mismatch hypothesis explains host susceptibility to an emerging infectious disease. Ecol Lett. 2017;20(2):184–93.

68. Cohen JM, et al. Impacts of thermal mismatches on chytrid fungus Batrachochytrium dendrobatidis prevalence are moderated by life stage, body size, elevation and latitude. Ecol Lett. 2019;22(5):817–25.

69. Ripley BD. Modern applied statistics with S. New York: Springer; 2002. https://doi.org/10.1007/978-0-387-21706-2.

70. Ziegel ER. Visualizing categorical data. Technometrics. 2001;43(4):498.

71. Bluthgen N, Menzel F, Bluthgen N. Measuring specialization in species interaction networks. BMC Ecol. 2006;6:9.

72. Mavrich TN, Hatfull GF. Bacteriophage evolution differs by host, lifestyle and genome. Nat Microbiol. 2017;2:17112.

73. Young DB, Garbe TR. Heat shock proteins and antigens of Mycobacterium tuberculosis. Infect Immun. 1991;59(9):3086–93.

74. Friman VP, et al. High temperature and bacteriophages can indirectly select for bacterial pathogenicity in environmental reservoirs. PLoS One. 2011;6(3):e17651.

75. Kaan T, Jürgen B, Schweder T. Regulation of the expression of the cold shock proteins CspB and CspC in Bacillus subtilis. Mol Gen Genet. 1999;262(2):351–4.

76. Roncarati D, Scarlato V. Regulation of heat-shock genes in bacteria: from signal sensing to gene expression output. FEMS Microbiol Rev. 2017;41(4):549–74.

77. Horn G, et al. Structure and function of bacterial cold shock proteins. Cell Mol Life Sci. 2007;64(12):1457–70.

78. Champ S, et al. Chaperone-assisted excisive recombination, a solitary role for DnaJ (Hsp40) chaperone in lysogeny escape. J Biol Chem. 2011;286(45):38876–85.

79. Chamakura KR, Tran JS, Young R. MS2 Lysis of Escherichia coli depends on host chaperone DnaJ. J Bacteriol. 2017;199(12):e00058–17.

80. Luo P, et al. Strand-specific RNA-Seq analysis provides first insight into transcriptome response of Vibrio alginolyticus to phage infection. Mar Genomics. 2018;38:5–8.

81. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363(4):fnw002.

82. Guttenplan SB, Kearns DB. Regulation of flagellar motility during biofilm formation. FEMS Microbiol Rev. 2013;37(6):849–71.

83. Pande S, Kost C. Bacterial unculturability and the formation of intercellular metabolic networks. Trends Microbiol. 2017;25(5):349–61.

84. Nayfach S, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39(4):499–509.

85. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci U S A. 2021;118(12):e2016810118.

Publisher's Note