A theoretical analysis of how strain-specific viruses can control microbial species diversity

T. Frede Thingstad1, Selina Våge, Julia E. Storen, Ruth-Anne Sandaa, and Jarl Giske

Department of Biology, University of Bergen and Hjort Centre for Marine Ecosystem Dynamics, N-5020 Bergen, Norway

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved April 17, 2014 (received for review January 16, 2014)

Pelagic prokaryote communities are often dominated by the SAR11 clade. The recent discovery of viruses infecting this clade led to the suggestion that such dominance could not be explained by assuming SAR11 to be a defense specialist and that the explanation therefore should be sought in its competitive abilities. The issue is complicated by the fact that prokaryotes may develop strains differing in their balance between competition and viral defense, a situation not really captured by present idealized models that operate only with virus-controlled “host groups.” We here develop a theoretical framework where abundance within species emerges as the sum over virus-controlled strains and show that high abundance then is likely to occur for species able to use defense mechanisms with a low trade-off between competition and defense, rather than by extreme investment in one strategy or the other. The J-shaped activity–abundance community distribution derived from this analysis explains the high proportion low-active prokaryotes as a consequence of extreme defense as an alternative to explanations based on dormancy or death due to nutrient starvation.

In a commentary (1) to the recent description of viruses infecting the marine bacterium SAR11 (2), it was pointed out how strain diversification complicates the discussion of whether the success (in terms of abundance) of this clade is caused by its defensive or its competitive properties. The work presented here constitutes a formalization and expansion of this discussion.

Starting laboratory systems with the simple combination of a virus and a sensitive host strain frequently results in rapid takeover of resistant host mutants, the system thereby transiting to a new state characterized by low viral abundance and thus a low virus:host ratio (3). This led Weinbauer (4) to formulate the paradox that simple systems dominated by resistant hosts and a low virus:host ratio, as opposed to the high virus:host ratio and general host susceptibility to viral attack characteristic of natural prokaryote communities. Starting with both a sensitive and a resistant strain (5, 6), it has been shown how the sensitive strain becomes virus controlled and the resistant strain resource controlled.

Experimental host–virus systems have also been used to demonstrate the concept of antagonistic evolution (7, 8) where co-occurrence of viruses and hosts speeds up evolution both in the host (9) and in the virus (10), leading to an arms race with a characteristic timescale of days to weeks and bringing evolutionary processes into timescales otherwise characteristic of ecological processes. These dynamics are clearly reflected in the takeover of resistant host mutants and the subsequent diversification of host and virus strains in laboratory model systems and are thus closely related to the Weinbauer paradox discussed above.

The molecular defense systems used by the microbial hosts against viruses are diverse. Recent insight from sequencing different isolates of the same species of prokaryotes (defined by their 16S rDNA similarity) has led to the concept of a core genome common to all strains of a species, whereas other parts of the DNA vary between strains. The high frequency of genes affecting cell surface properties in these variable regions has been interpreted as an indication of a role in viral defense (11, 12). Other defense mechanisms such as the clustered regularly interspaced palindromic repeats (CRISPR) system in prokaryotes (13) also support the idea that viral defense is important as a fitness factor, stimulating evolution of sophisticated molecular defense systems. These systems presumably come at a cost of resistance (COR) in terms of reduced competitive ability, as would be expected if, e.g., modification of porins required for uptake of limiting nutrients (14, 15) or a high incidence of unfavorable mutations is involved. Other molecular mechanisms such as the CRISPR system may be speculated to have a different type of COR where there may be a significant cost in running the system, but the additional cost of adding a new recognition sequence could be small. Prokaryote species therefore probably differ, not only in competitive properties, but also in the trade-off defining how much they lose in competitiveness relative to the protection gained with the molecular defense system used.

Basic elements of a theory able to represent such relationships are present in the so-called killing-the-winner (KtW) model, but in its highly idealized original forms (16–18) this refers only to “host groups,” not really specifying whether these host groups represent species or strains. With defense systems diversifying the host population, mostly without affecting the 16S/18S rRNA genes, it is probably more correct to associate such host groups with strains than with species. Arguments based on the simple form of the KtW model therefore involve a clear risk of “comparing apples to pears” and a danger of vague or in the worst case wrong conclusions.

Recent refinements of the KtW model have introduced trade-offs between host growth rate and viral defense that influence the rank-abundance distribution patterns of viral and microbial host populations within communities (19). The simplifying one-to-one relationship between hosts and viruses assumed in the original versions has also recently been replaced by a nested interaction matrix allowing for viruses with broader host ranges (20, 21). Evolutionary aspects of the host–virus arms race have been analyzed in a chemostat setting (22).

With this basis, we use a simple chemostat arms race model as the setting for a discussion of how competitive and defensive mechanisms work together to control strain-level diversification of an initially clonal host population exposed to a lytic virus. This work presents the first detailed analysis to the authors' knowledge of how species-level diversity is a property emerging from competitive and defensive abilities at the organism level in a microbial system where the diversity-generating mechanism is strain-specific viral lysis. The theoretical analysis constitutes a general case treatment of the important special case question of what properties may make SAR11, a subphyllum within the Alphaproteobacteria, so dominant in the pelagic environment. The resulting conceptual framework connects differences in the molecular defense mechanism to ecosystem-level properties such as diversity and activity. It also suggests a reinterpretation of the concept of dormancy in aquatic microbial communities.

Significance

This work presents the first detailed analysis to the authors’ knowledge of how species-level diversity is a property emerging from competitive and defensive abilities at the organism level in a microbial system where the diversity-generating mechanism is strain-specific viral lysis. The theoretical analysis constitutes a general case treatment of the important special case question of what properties may make SAR11, a subphyllum within the Alphaproteobacteria, so dominant in the pelagic environment. The resulting conceptual framework connects differences in the molecular defense mechanism to ecosystem-level properties such as diversity and activity. It also suggests a reinterpretation of the concept of dormancy in aquatic microbial communities.

Author contributions: T.F.T. designed research; T.F.T., S.V., J.E.S., R.-A.S., and J.G. performed research; and T.F.T., S.V., J.E.S., R.-A.S., and J.G. wrote the paper.

The authors declare no conflict of interest.

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved April 17, 2014 (received for review January 16, 2014)

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: frede.thingstad@bio.uib.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400909111/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1400909111

PNAS Early Edition | 1 of 6
allows us to subsequently apply the same principles to the balance between two or more coexisting species and derive the structure of the mature community that will establish from a given initial set of seeding species. The analysis is formulated in general terms but with numerical examples aiming to fit the community of marine prokaryotes.

**Strain Generation in an Idealized Arms Race**

A steady-state chemostat community can be invaded by a new organism only if its net growth rate is positive at the concentration of limiting substrate present in the chemostat, i.e., faster than the dilution rate when there are no lytic viruses or other loss mechanisms for the invader. With a virus-free clonal population \(H_0\) in the chemostat, only a mutant growing faster than the parent strain at the existing concentration can thus invade. If, however, we introduce a lytic virus \((V_0)\) able to infect \(H_0\), there is a new steady state possible where the abundance \(H_0\) of the parent strain is controlled by the virus (Methods),

\[
H_0 = \frac{D + \delta_0}{(m_00 - 1)/\mu_0} \tag{1}
\]

where \(m_0, \mu_0, \delta_0, \) and \(D\) denote the burst size of virus \(j\) reproducing on host \(i\), the effective adsorption coefficient for virus \(j\) on host \(i\), the decay rate of virus \(j\), and the dilution rate, respectively (symbols are summarized in Table 1). Hosts with strong defense mechanisms able to minimize \(\beta_0\) and/or \(m_0\), preferably in combination with a high viral decay rate \(\delta_0\), will thus be able to maintain a high abundance of the parent strain. For the set of numerical parameters for species A (Table 1), \(H_0 = 1.9 \times 10^7\) cells L\(^{-1}\).

For simplicity, we assume a nonrespired limiting substrate like orthophosphate, present in the reservoir at concentration \(S_R\), creating a carrying capacity \(H_T = Q^{-1}S_R\) for the total number of individuals in the community (Methods). The \(Q\) and \(S_R\) chosen for the one- and two-species examples (Table 1) correspond to a carrying capacity \(H_T = 4 \times 10^8\) cells L\(^{-1}\). With \(H_0\) virus controlled, there is an unused resource \((H_T - H_0)Q\) that will remain as free phosphate and allow for the possible establishment of a resistant host mutant \(H_1\). As a simple model we assume a fitness reduction associated with the mutation from \(H_0\) to \(H_1\) in the form of a reduction in maximum growth rate \(\mu_0^{max}\) by a factor \(\nu\) (0 < \(\nu\) ≤ 1) so that \(\mu_1^{max} = \mu_0^{max}\nu\).

At steady state, it is now the immune mutant \(H_1\) that must grow at a rate equal to the dilution rate \(D\). For this to occur the culture concentration of limiting substrate must increase from \(S_0\) to \(S_1\) and \(H_0\) will therefore now grow faster than the dilution rate \(D\). This increase in growth rate is compensated by viral lysis. Sufficient viruses are required to produce a lysis rate compensating for the growth rate difference between \(H_0\) and \(H_1\); viral abundance thus becomes an indicator of the COR associated with the host’s defense mechanism. We now have a situation with

<p>| Table 1. Symbols with description and numerical values used in examples with one or two (separated by “/”) species |</p>
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>Chemostat dilution rate</td>
<td>0.1 h(^{-1})</td>
<td>Arbitrarily chosen</td>
</tr>
<tr>
<td></td>
<td>Reservoir concentration of limiting substrate</td>
<td>20 nM</td>
<td>Small value chosen to give a community size (H_T) suitable for illustrative purposes</td>
</tr>
</tbody>
</table>

Properties of host species determining the geometry of Fig. 2

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q)</td>
<td>Cell quota of limiting element</td>
<td>(5 \times 10^{-8}) nmol P</td>
<td>Approximate P content of a bacterium with 25 fg C cell and a molar CP ratio of 50</td>
</tr>
<tr>
<td>(\mu_0^{max})</td>
<td>Maximum specific growth rate of strain 0</td>
<td>1.0/0.5 h(^{-1})</td>
<td>Nutrient affinity constant for phosphate (e.g., ref. 29)</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>Slope of growth curve as substrate concentration approaches zero (nutrient affinity)</td>
<td>0.75/1.5 \times 10^{-9}</td>
<td></td>
</tr>
<tr>
<td>(\nu)</td>
<td>Fractional decrease in (\mu_0^{max}) for each step in (i)</td>
<td>0.9/0.95 d.l.</td>
<td>Lennon et al. (30) suggest that COR values may be up to 20% ((\nu = 0.8))</td>
</tr>
</tbody>
</table>

Properties of host–virus interaction

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_0)</td>
<td>Effective adsorption coefficient for the interaction between strain 0 and virus 0</td>
<td>(1.5 \times 10^{-10}) L h(^{-1})</td>
<td>Based on ref. 31</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>&quot;Memory&quot; in viruses of ability to infect previous host strains</td>
<td>0.5/0.9 d.l.</td>
<td>From ref. 32</td>
</tr>
<tr>
<td>(\rho)</td>
<td>Fractional decrease in (\beta) for each step in increased host range</td>
<td>0.92/0.9 d.l.</td>
<td>Arbitrarily chosen for illustration purposes</td>
</tr>
<tr>
<td>(\delta_0)</td>
<td>Decay rate of viruses</td>
<td>0.012 h(^{-1})</td>
<td></td>
</tr>
<tr>
<td>(m_0)</td>
<td>Burst size</td>
<td>40 d.l.</td>
<td>E.g., ref. 33 gives a range of 6–140</td>
</tr>
</tbody>
</table>

Emergent properties

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_T)</td>
<td>Carrying capacity for the community</td>
<td>(4 \times 10^8) and (2 \times 10^8) L(^{-1})</td>
<td>(= S_R Q^{-1}). High value used for the multispecies case</td>
</tr>
<tr>
<td>(n_d)</td>
<td>No. host strains that can establish before (\mu_0^{max}) decreases below (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n_r)</td>
<td>No. host strains that can establish before host population exceeds (H_T)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

State variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S_i)</td>
<td>Culture concentration of limiting substrate when host strain (i) grows at (\mu_0(S_i) = D)</td>
<td>nmol P</td>
<td>Italic is used to denote the variable whereas nonitalic is used as the name of the species/strain. E.g., (H_i) is the abundance of strain (H_i).</td>
</tr>
<tr>
<td>(H_i)</td>
<td>Abundance of host cells of strain (i)</td>
<td>L(^{-1})</td>
<td></td>
</tr>
<tr>
<td>(V_j)</td>
<td>Abundance of viruses of strain (j)</td>
<td>L(^{-1})</td>
<td></td>
</tr>
</tbody>
</table>

Subscripts before the symbol denote species; subscripts after the symbol, when relevant, denote strains. d.l., dimensionless.
viruses control the original strain H₀ and resource control of the resistant mutant H₁ as in the classical experiments of Bohannan and Lenski (5, 6), whereas the abundance of the virus population V₀ is a result of the cost of resistance in the mutation from H₀ to H₁. V₀ can be calculated from the requirement that growth equals the sum of losses to dilution and viral lysis for H₀,

\[ V₀ = \frac{c₀}{ρ₀₀}, \]

where \( c₀ = μ₀ - D \) (Methods) and represents the actual COR, i.e., the actual growth rate difference between H₀ and H₁. With the parameters chosen for species A (Table 1), this gives \( V₀ = 6.6 \times 10^{0} \text{ L}^{-1} \), corresponding to a virus:host ratio of 1.65 in the state with H₁ established. Note how, in this framework, a species is defined by a set of parameters defining the properties of its undefended parent strain (\( μ_{max} \), \( α \)) and the three factors (\( i \), \( μ \), \( β \)) defining how its competitive and defensive properties change in successive mutants. The subsequent evolutionary steps expected in such a chemostat would be a sequence of alternating additions of new immune host mutants followed by new virus mutants with their host spectrum expanded to include the last new host strain. The corresponding sequence of steady states would alternate between states where all host strains are virus controlled and contain unused limiting resources and states with one resource-controlled resistant strain. A model where host mutations leading to new resistant host strains Hₙ are most likely to occur in host Hₙ₋₁, susceptible to one virus only (Fig. 1), where the subsequent virus Vₙ able to attack Hₙ is most likely to arise from a mutation in virus Vₙ₋₁ that already has the ability to attack Hₙ’s closest relative Hₙ₋₁, leads to the structure in Fig. 1. With H₀ susceptible to all viruses and H₋₁ attacked only by one (Vᵥₙ₋₁), this structure combines viruses attacking a single host strain (V₀) with broad host-range viruses (e.g., Vₙ) and corresponds to what Flores et al. (23) have termed “nested” infection. Because this pattern has been shown to occur in experimental (23) and natural (24) systems of prokaryotes and their viruses, arms race models such as the one discussed above may be seen as a possible hypothesis for how such observed patterns evolve in nature (21).

Each step of this evolution contains what might be seen as a “remaining resource”: When all strains are virus controlled, this remaining resource will be in the form of a free unused part of the limiting nutrient, creating a niche for an immune host mutant to invade. As this immune mutant establishes, the previously free nutrient will be tied up in the biomass of the new resource-controlled immune strain, creating a niche for a new viral mutant able to attack it and thus a subsequent transition back to a situation with all host strains again virus controlled, but with a reduced pool of remaining resources. The remaining resource will thus diminish as successive steps bind an increasing part of the total limiting resource \( S_R \) in new susceptible host strains. Further evolution subsequent to the last step depleting the remaining resources would have to occur as rearrangements within a mature community and would depend on optimization of existing or replacement of established strains, rather than as an exploitation of remaining resources. The term “mature community” is subsequently used to refer to the situation where the remaining resources have been depleted. Evolutionary processes on longer timescales are not considered here.

This arms race model contains two types of limits to strain richness \( n \), one dilution driven and one resource driven. The dilution-driven limit \( (n_d) \) is reached when \( μ_{max} = D \), i.e., when the COR of strain \( n \) has become so high that further defense prevents a growth rate balancing the dilution rate. The resource-driven limit \( (n_r) \) is the richness reached when the remaining resources are depleted. We restrict the discussion to the case \( n_r < n_d \). Assuming the same fractional reduction \( i \) in maximum growth rate for each new mutation step, mutant \( n \) will have a maximum growth rate

\[ μ_{max} = μ₀^{n} \nu^{n} \]

From which \( n_d \) can be calculated (Methods). For species A (Table 1 and Fig. 2A), \( n_d = 21 \).

The culture concentration \( S₀ \) of the limiting substrate at step \( n \) is defined by the requirement that the immune strain \( n \) must grow at the dilution rate (Methods)

\[ S₀ = \frac{μ_{max} \nu^{n} D}{μ₀ μ_{max} \nu^{n} - D} \]

Where \( α \) is the \( n \)-independent slope of the Michalis–Menten-shaped growth rate curve for substrate concentration approaching zero.

The sequence of solutions to Eq. 4 as the arms race progresses \((n) \) increases) is illustrated in Fig. 2A. As new strains with increasing COR establish, the previous strains with lower COR grow increasingly faster, requiring compensation by more viruses (or more efficient viruses).

To calculate the resource-controlled limit \( n_r \), we need to calculate the abundance \( H_i \) of each of the virus-controlled strains, \( i = 0 \ldots n_r \). Host abundances can in analogy to Eq. 1 be derived from the requirement that all virus populations are produced at a rate balancing the sum of loss by dilution and by decay (Methods). Assuming for simplicity that all decay rates are equal \( (δ_i = δ₀ \text{ for all } i) \) and all burst sizes are equal \( (mᵣ = m₀ \text{ for all } 0 ≤ i, j < n_r) \), the general matrix equation for determination of the vector \( H \) of host abundances becomes (Methods)

\[ H = \frac{δ₀ + D}{m₀ - 1} (βᵣ^{-1})^{-1} U, \]

Where \( U \) is a column vector with all \( n_r \) elements equal to 1, and \( βᵣ \) is the upper triangular matrix of effective adsorption coefficients with \( β_j \) as the coefficient for the interaction of host \( j \) with virus \( i \). Whereas each component of \( H \) is determined by the host–virus interactions (and the dilution rate), the number \( n_r \) of virus-controlled strains that can establish (the dimensionality of the \( β \) matrix) is reached when the elements \( H_i \) sum up to the resources available, i.e., when \( \sum_{k=0}^{n_r} H_k = Q^{-1} S_R \). The dimensionality \( n_r \) of the \( β \) matrix thus depends on the reservoir concentration \( S_R \).

Abundance in the virus strains \( V \) can be calculated from the requirement that each host strain \( H_i \), \( 0 ≤ i < n_r \), has a growth rate balancing the sum of dilution and its loss to viral lysis. Solving (Methods) these requirements for the vector \( V \) of viral abundances gives

\[ V = βᵣ^{-1} c(n_r), \]

where \( c_i(n_r) = (μ_i (S_R) - D) \) is the difference in growth rate between strain \( i \) and the last established, immune strain \( n \), growing at \( μ₀ = D \). Eq. 6 is the \( n_r \)-dimensional matrix analogy to Eq. 2. Note that the dependence on \( n_r \) signifies that not only the
dimensionality, but also the individual elements in the COR vector $c(n_r)$ change with $n_r$.

To analyze the steady state further, we want to be able to vary two aspects of the nested structure of the $\beta$ matrix: (i) a possible fractional reduction ($\rho$) in $\beta_{ij}$ for each increasing step in $j$, i.e., that viruses able to infect increasingly defensive host strains have a cost in the form of a reduced effective adsorption coefficient, and (ii) a possible fractional reduction ($\sigma$) in $\beta_{ij}$ with each decreasing step in $i$ away from $j$, i.e., a possible loss of “memory” where a new virus mutant loses some of its ability to infect previous hosts. In analogy with the modeled decrease in $p_{ij}^{\text{max}}$ (Eq. 3), we thus assume the elements of $\beta$ to be given as

$$\beta_{ij} = \beta_{i0}(\sigma^{j-i}/\rho^i); \quad i=0...n_r-1, \quad j=i...n_r-1.$$  

[7]

$\rho$ and $\sigma$ thus represent a horizontal and a vertical attenuation coefficient, respectively, for the strength of the interactions in the interaction matrix as illustrated for the case $n_r = 3$:

$$\beta = \begin{bmatrix} 1 & \sigma \rho^1 & \sigma^2 \rho^2 \\ 0 & \rho^1 & \sigma^2 \\ 0 & 0 & \rho^2 \end{bmatrix}.$$  

[8]

Note how this model contains the simple one-virus–one-host model as a special case when the memory coefficient $\sigma = 0$, making all off-diagonal elements zero.

Because the transpose $\beta^T$ is a lower triangular matrix, Eq. 5 can be solved with backward substitution (Methods) to give

$$H_0 = \frac{(\delta_0 + D)}{(m_0 - 1)\rho_0}.$$  

[9a]

and

$$H_k = \frac{\delta_0 + D}{(m_0 - 1)\rho_0} (1 - \sigma \rho)^k, \quad k = 1...n_r - 1.$$  

[9b]

Whereas $H_1$ may be larger or smaller than $H_0$, depending on the factor $(1 - \rho)/\rho$ (Methods), $H_k$ is a monotonously increasing function with $k$ for all $k > 1$, implying an increase in abundance as strains become more defensive and slow growing (Fig. 3A). Also, $H_k$ increases as $\rho$ or $\sigma$ decreases, implying that efficient host defense or inefficient viruses (small $\rho$, $\sigma$) give high abundance of individuals within strains.

Summing up abundance over all strains (Methods) gives the total abundance, which for $n_r$ strains equals the carrying capacity

$$H_T = H_0 \left[ 1 + \frac{(1 - \sigma \rho)}{(1 - \rho)} \frac{(1 - \rho^{n_r - 1})}{\rho^{n_r - 1}} \right].$$  

[10]

Solving Eq. 10 for $n_r$, we get the richness of sensitive strains that can establish before the remaining immune strain $n_r$ becomes too small to carry a new virus mutant. This gives (Methods)

$$n_r = \text{Int} \left[ 1 + \frac{\text{log}(1 + ((1 - \rho)/(1 - \sigma \rho))H_T/H_0 - 1))}{\text{log}(\rho^{-1})} \right].$$  

[11]

For species A (Table 1) this gives $n_r = 20$. High values of $\sigma$ and/or $\rho$ imply low defense and/or efficient viruses and therefore lower abundance per strain. This leaves room for more strains and therefore higher strain diversity $n_r$ (Fig. S1).

With the upper triangular form of the $\beta$ matrix, Eq. 6 can be solved with backward substitution (Methods) to give the $V$ vector (Fig. 3B), and the total viral abundance can be found by summing up its elements:

$$V_T = \sum_{k=0}^{n_r-1} V_k = \rho_0^{-1} \left( c_0(n_r) + (1 - \sigma \rho) \sum_{k=0}^{n_r-1} c_k(n_r) \rho^k \right).$$  

[12]

For our choice of parameters, the virus abundance has an optimum for strains with an intermediate host range (Fig. 3B). The position of this optimum is sensitive to the memory factor $\sigma$ with the optimum shifted toward lower viral strain number, i.e., toward viruses with a smaller host range attacking the faster-growing hosts for small values of $\sigma$ (Methods).

In our example the virus:host ratio of the mature community of strains in the one-species case is 58.

### The Multispecies Case

The competitive aspects of cases with more than one species are illustrated in Fig. 2B and C. The strains from each species will form a family of growth curves as illustrated for the two-species case (Fig. 2B). For the multispecies case (Fig. 2C) each family of curves is for visual clarity represented by the growth curve of the parent strain only. As in the one-species case, the strains established in the mature community will be determined by the sequence of intersections between the vertical
line \( S = S_n \), and the respective growth curves, terminated downward by the horizontal line \( \mu = D \). The position of the \( S = S_n \) line is constrained by the requirement that the sum of individuals over all established strains of all species equals the carrying capacity. The two environmental conditions, dilution and resource availability, thus both influence the structure of the mature community.

Making the simplifying assumption that viruses do not cross-infect between species, the \( \beta \) matrix for the multispecies case becomes a compound matrix with upper-triangular submatrices along the diagonal, each representing a species. The dimension of each submatrix is determined by the requirement that the last strain of that species is one just to grow equal to or faster than \( D \) at substrate concentration \( S_n \). Community composition and culture concentration \( S_n \) at maturity can then be determined using an iterative algorithm.

Starting from an initial guess \( S' \) for \( S_n \), the number \( n' \) of strains of species \( x \) growing at a rate \( \geq D \) can be calculated (Methods). Knowing \( n' \), the abundance \( H'_T \) of species \( x \) can be calculated as in Eq. 10, but summing up to the abundance of the species (Methods).

Summing up \( H'_T \) over all established species finally gives the total community abundance \( H'_T(S) \) that corresponds to \( S' \). Because \( H'_T(S) \) is a monotonically increasing function of \( S' \), the direction for iterative adjustment of \( S' \) is known and \( S_n \) and community composition can be found by iterating until the largest value of \( H'_T(S) \) \( \leq H_T \) is reached.

With \( S_n \) known and the established strains identified, the growth rates and therefore also the COR vector \( c(n) \) are known and virus abundances for the mature community can be found from the equivalent to Eq. 6.

In our numerical example for two species (Table 1), species B is dominant with \( 3.14 \times 10^8 \) compared with \( 0.86 \times 10^8 \) A individuals per liter. This is a combined result of the number of strains established for each species \([6 \text{ for } A \text{ and } 22 \text{ for } B \text{ (of which one is the single resistant strain)}\)], but also due to the high abundance of the most defensive strains of B (Fig. 3A). In our example the two species differ considerably in the value of the memory factor \( \sigma \) (0.5 and 0.9 for A and B, respectively). As in the single-species case, this shifts the optimum of the virus distributions, in our case so much that whereas the low-memory \( (\sigma = 0.5) \) viral population of host species A is dominated by narrow host-range viruses (those with low viral strain number) attacking the fast-growing A strains, the long-memory \( (\sigma = 0.9) \) viral population belonging to host species B is dominated by viruses with broad host range (high viral strain number) (Fig. 3B).

To illustrate the multispecies case we generated a seeding community of 13 species, using random number generation (Table S1). Four of the 13 species (F, I, L, and M) did not have a combination of \( \rho_{\text{max}} \) and \( \alpha \), allowing them to establish in the given environment \( (D = 0.1 \text{ h}^{-1}, H_T = 2 \times 10^9 \text{ h}^{-1}) \) (Fig. 2C). The \( \beta \) matrix of the mature community is illustrated graphically in Fig. S2. To capture the essence of how the abundance within the established species depends on the parameter values, we defined a trade-off index for species \( x \) as \( \tau(x) \) by dividing the reduction in effective adsorption coefficient from parent strain 0 to the first mutant strain 1 by the corresponding reduction in maximum growth rate. To get a dimensionless number we also divided by the ratio \( \rho_0/\mu_{\text{max}} \).

This trade-off index captured 75% of the variability in abundance (Fig. 4A). Illustrative cases include species E, which combines a high growth rate of the parent strain (Fig. 2C) with a low trade-off index (Fig. 4A) and high species abundance. This can be compared with species G, also with a comparably high growth rate for the parent strain, but with a high trade-off, and therefore ending up with a very low abundance (Fig. 4A). Species K, despite an only moderate competitiveness of the parent strain (Fig. 2C), has a low trade-off and ends up as the dominant species (Fig. 4A).

The different parameter combinations result in different growth rate spectra for the different species (Fig. 4C), but the combination of increasing number of species (Fig. 4A) and increasing abundance per strain (Fig. 4C) in the lower end of the growth rate spectrum enhances the J-shaped form of the growth rate vs. abundance spectrum, corresponding to a huge dominance of slow-overfast-growing individuals (Fig. 4B).

Discussion

The conceptual importance of distinguishing between species and strains is best illustrated by how this resolves the Weinbauer paradox (4). The simple and complex systems now correspond to the early and mature states of the host–virus arms race, dominated by a resistant strain or composed of a series of sensitive strains, respectively.

A host species is in our model defined by two separate sets of traits defining its competitive \( (\mu_{\text{max}}, \alpha, \nu) \) and its defensive \( (\beta) \) properties, respectively. The two sets have different consequences for the system’s mature state: The competitive traits determine the number of strains of each species, whereas the defensive traits determine the abundance of individuals within each strain. The environment \( (D, S_0) \) interacts with these traits by influencing both which species that can establish the largest number of strains and the total number of individuals that can establish before the remaining resources are depleted.

Numerical dominance of a species thus becomes a function of both its competitive and its defensive abilities. The costs of these strategies are represented in the parameters describing the stepwise reductions \( (\nu) \) in growth rate and the gain in protection by the reductions in viral efficiency \( (\rho) \).

With a seeding community of \( N \) different potentially invading species, the topography of Fig. 2C would be determined by \( 3N + 2 \) parameters: the three competitive traits \( (\mu_{\text{max}}, \alpha, \nu) \) for each of the \( N \) species plus the two environmental parameters \((D, S_0)\). The result is, however, conceptually simple with the \( N \) families of growth curves and their position relative to the two lines \( \nu = D \) and \( \nu = S_0 \), determining community composition in a transparent manner. Also the \( \beta \) matrix has interesting properties pointing to interesting generalizations (Methods).

The seeding community in Table S1 was generated without assuming any correlations constraining the trade-off between parameters. This gave a range of different trade-offs between competition and defense for the generated species (Fig. 4A).

Biological insight allowing one to constrain this trade-off or allocate specific trade-offs to specific species would be an important expansion of the present model. Consider, e.g., how a CRISPR system hypothetically could create a correlation between \( \mu_{\text{max}} \) and \( \nu \) if there is a high species-associated running cost, but a small strain-related cost in adding new recognition sequences. Species investing
in CRISPR defense systems would under these conditions invade late in the community succession toward maturation, but once established, the growth rate differences between strains would be small, and strains belonging to the same species would tend to cluster along the growth rate axis. A defense system with a small running cost but a large price for each additional virus would give a different pattern with strains of different species widely scattered along the growth rate axis. Relevant experimental information includes spot area measurements on microautoradiographs (e.g., ref. 25) producing J-shaped activity spectra as predicted by our model (also ref. 19). Interestingly, the observed spectrum for SAR11 resembles that of the rest of the heterotrophic prokaryotic community (22), as our model would predict whether species have comparable $\mu_{\text{max}}$ values. However, in experiments where viral infection pressure has been reduced, previously rare species become dominant (26), as one would expect from this model if there were large species differences in $\mu_{\text{max}}$.

A classic discussion in microbial ecology is whether the low- or inactive cells in natural populations of heterotrophic prokaryotes are dormant or dead (e.g., ref. 27 and references therein). As an alternative to substrate specialists starving due to lack of suitable substrates, the proposed model explains low activity as the consequence of unavoidable costs of resistance. This allows for coexistence of high- and low-activity hosts, even if the hosts are substrate generalists as suggested from molecular evidence for their use of carbon sources (28).

**Conclusion**

Within the highly idealized framework of this model, the primary key to numerical dominance is a low trade-off between competition and defense. The secret behind numerical success as for SAR11 is thus perhaps not to be found in extreme competitive or defensive abilities, but maybe more subtly in a (set of) defensive mechanism(s) that can operate at the molecular level without a large trade-off toward competitive ability.

**ACKNOWLEDGMENTS.** This work was financed by the European Union FP7 through the European Research Council Advanced Grant 250254 Microbial Network Organisation and by the Department of Biology, University of Bergen.

---