Spatial structure and Lamarckian adaptation explain extreme genetic diversity at CRISPR locus.

Jan O. Haerter[1*], Kim Sneppen[1]

1 Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

∗ **E-mail: haerter@nbi.dk**

SUPPLEMENTARY MATERIAL

Analysis of the well-mixed bacteria-phage community

We supplement the main text by supporting additional analysis for the well-mixed state of a simple two-state system, which we then use to draw basic conclusions for a well-mixed CRISPR system with a varying level of immunity.

Imagine first a system consisting of susceptible ($\mathcal{S}$), resistant ($\mathcal{R}$) and infected ($\mathcal{I}$) bacteria (Fig. S1a, inset). The bacterial reproduction rate is taken as the elementary time-unit and we refer to all other timescales relative to it. Bacteria are allowed to transition from $\mathcal{S} \to \mathcal{R}$ ($\mathcal{R} \to \mathcal{S}$) at a rate $r \ll 1$ ($r' \ll 1$). Other relevant quantities are: (i) the phage reproduction rate w - where $1/w$ is the time between subsequent phage generations; (ii) the cost c of acquiring the defense system. We simply take the bacterial reproduction time to increase to $1 + c$ resulting in a growth rate $v(c) \equiv 1/(1 + c)$. The time-dependent number of susceptible bacteria $\mathcal{S}$ is given by

$$\frac{\partial \mathcal{S}}{\partial t} = n_b \mathcal{S}(1 - \mathcal{S} - \mathcal{R} - \mathcal{I}) - w n_a \mathcal{S}\mathcal{I} + (1 - v(c)) n_b \mathcal{S}\mathcal{R} - r n_b \mathcal{S} + v(c) r' n_b \mathcal{R} \tag{S1}$$

where n_b is the number of nearest neighbors the bacteria have access to and it will depend on the specific properties of the medium. n_a is the number of neighboring sites phages have access to after they are ejected from an infected host. This quantity in reality is determined by the effective volume that phage descendants are able to explore before being degraded. It increases with larger diffusion and smaller decay rate. The upper bound for n_a is the burst size (~ 100 [1, 2], depending on eco-system and phage type) while the theoretical lower bound for sustaining the population will be at least 1. The first term on the r.h.s. of Eq. S1 describes the growth of the $\mathcal{S}$ population where the space occupied by

other bacteria is not available for spreading. The second term is the interaction term between bacteria $\mathcal{S}$ and phage and the third term describes the competition between $\mathcal{R}$ and $\mathcal{S}$. Note that we allow $\mathcal{S}$ to grow into space previously occupied by $\mathcal{R}$ at a rate $1 - v(c) = c/(1 + c)$, hence gradually displacing the slower bacteria. The last two terms in Eq. S1 describe the sink and source of $\mathcal{S}$-bacteria due to occasional mutations.

The time-dependence of the phage-resistant $\mathcal{R}$-population is analogous except that the phage-interaction term now is diminished by an additional factor $1 - \delta$, with δ characterizing the bacterial immunity:

$$\frac{\partial \mathcal{R}}{\partial t} = v(c) n_b \mathcal{R}(1 - \mathcal{R} - \mathcal{S} - \mathcal{I}) - w n_a (1 - \delta)\mathcal{R}\mathcal{I} + (v(c) - 1) n_b \mathcal{R}\mathcal{S} - r' v(c) n_b \mathcal{R} + r n_b \mathcal{S} \tag{S2}$$

The phage can prey both on the susceptible ($\mathcal{S}$) and the resistant ($\mathcal{R}$) bacteria, but at different rates. The growth of the infected bacteria is

$$\frac{\partial \mathcal{I}}{\partial t} = w n_a \mathcal{I}(\mathcal{S} + (1 - \delta)\mathcal{R}) - w\mathcal{I} \tag{S3}$$

Perfect Immunity – Focusing on sustainable ecologies, we are interested in the steady-state solutions of the above equations. Since r and r' are generally much smaller than the typical bacterial reproduction rate (mutation is not a significant way to grow a population) we can neglect them in the solution to the well-mixed system. In the steady-state, the solution for perfectly immune $\mathcal{R}$-bacteria ($\delta = 1$) becomes:

$$\mathcal{S} = \frac{1}{n_a} , \tag{S4}$$

$$\mathcal{R} = -\frac{1}{n_a v(c)} + \frac{n_a}{n_b} \frac{w}{1 - v(c) + \frac{n_a}{n_b} w} \tag{S5}$$

and

$$\mathcal{I} = \frac{v - 1}{v - 1 - \frac{n_a}{n_b} w} , \tag{S6}$$

as shown as solid lines in Fig. S1a. Consequently, the susceptible bacteria population is constant and in particular independent of the cost c. Hence, unprotected bacteria are always available and constitute a substantial fraction of the total population. Also, the phage-infected population $\mathcal{I}$ remains finite for all positive values of the cost. Hence, phage can survive even if bacterial defense comes at essentially no cost. Only exactly zero cost would allow the bacteria to make the phage go extinct.

In the limit $c \to 0$, the fraction of resistant bacteria is maximal ($\mathcal{R} \to 1 - 1/n_a$) but the susceptible bacteria remain constant ($\mathcal{S} \to 1/n_a$). Increasing c then leads to a monotonic decrease of $\mathcal{R}$ and monotonic increase of $\mathcal{I}$ until $\mathcal{R}$ finally vanishes at $c_0 = (n_a - 1)\frac{n_a}{n_b} w/(1 + \frac{n_a}{n_b} w)$.

Imperfect Immunity – Real immune systems, e.g. the CRISPR mechanism, are more likely to be somewhat incomplete. Such imperfect immunity can be characterized by a probability $\delta < 1$ of survival for the $\mathcal{R}$-population (sink term on r.h.s. of Eq. S2). Eq. S3 then yields:

$$\mathcal{S} + (1 - \delta)\mathcal{R} = 1/n_a \ . \tag{S7}$$

Hence, the population of $\mathcal{S}$-bacteria is now in fact reduced as compared to $\delta = 1$. The steady-state solutions of $\mathcal{S}$, $\mathcal{R}$, and $\mathcal{I}$ now become somewhat more complicated. Therefore, we only plot their cost dependence (Fig. S1a and b). Generally, the total bacteria population decreases with decreasing δ, while $\mathcal{R}$ increases at small cost but decreases at larger cost. The value of c_0 where $\mathcal{R}$ vanishes moves to smaller cost as δ decreases. In contrast, $\mathcal{S}$ decreases at small cost but saturates to $1/n_a$ at $c = c_0$. Hence, $\mathcal{S}$ is now no longer constant. In general, decreasing δ leads to a reduced coexistence regime with respect to cost of resistance c. In the limit of $\delta \to 0$ coexistence vanishes entirely (Fig. S1b). This means that sub-populations with very similar phage-defense chances generally cannot coexist while those with differing properties generally can.

Compare this discussion also with previous work [3], where similar conclusions were reached, however within a somewhat different model. Noticeably, for cost below a certain threshold ($\simeq 0.2\ c_0$ in Fig. S1b) it in fact pays off for the resistant bacteria to expose itself by lowering its immunity partially as it in this way uses an increased phage population to eliminate its susceptible competitor, a well-known effect previously termed *apparent competition* [4]. As a consequence of this indirect competition between two bacterial strains the coexistence range narrows with lowering δ, and in fact vanishes with $\delta \to 0$ (Fig. S1b). At small values of cost, the more resistant species entirely dominates while at slightly larger cost the susceptible species takes over.

Sustainable co-existance of susceptible and resistance bacteria with smoothly varying resistance

To give a simple argument for the existence of a stable steady state solution for a well-mixed CRISPR system with n_p phage species, we assume the two extreme sub-populations (complete resistance and complete sensitivity) given by the above discussion. We then show that these two extremes essentially constitute absorbing states for a given bacterium with variable n_{res}, i.e. in the vicinity of both extremes there will be a gradient in the bacterium's growth rate restoring its resistancy to the respective absorbing state.

Defining c now as the cost of a bacterium carrying a single spacer, the average growth rate $g(n_{res})$ for a given bacterium as a function of the number of its spacers n_{res} is

$$g(n_{res}) = -\mathcal{I}\left(1 - \frac{n_{res}}{n_p}\right) + \mathcal{S}\cdot v_{\mathcal{S}} + \mathcal{R}\cdot v_{\mathcal{R}} + \frac{1-\mathcal{I}-\mathcal{S}-\mathcal{R}}{1+n_{res}c} \tag{S8}$$

with the relative velocities

$$v_{\mathcal{S}} \equiv -\frac{n_{res}c}{1+n_{res}c} \tag{S9}$$

and

$$v_{\mathcal{R}} \equiv \frac{(n_p - n_{res})c}{(1+n_{res}c)(1+n_pc)}\ . \tag{S10}$$

For the range of coexisting $\mathcal{S}$ and $\mathcal{R}$ shown in Fig. S1b, $g(n_{res})$ is a smooth, non-positive convex function (Fig. S2) which vanishes only at zero and n_p and exhibits a minimum between these two points.

By defining the derivative of $g(n_{res})$ with respect to n_{res}, one yields the differential growth rate

$$g'(n_{res}) \equiv \frac{\partial g(n_{res})}{\partial n_{res}} = \frac{c^2(n_{res}(2+c\cdot n_{res}) - n_p)}{(1+c\cdot n_{res})^2(1+2c\cdot n_p)}\ . \tag{S11}$$

In the one-sided limits $n_{res}\to 0$ and $n_{res}\to n_p$ at $n_{res}=0$ and $n_{res}=n_p$, respectively, we yield

$$\lim_{n_{res}\to 0} g'(n_{res}) = -\frac{c^2 n_p}{1+2c\cdot n_p} \tag{S12}$$

and

$$\lim_{n_{res}\to n_p} g'(n_{res}) = \frac{c^2 n_p}{1+3c\cdot n_p + 2c^2 n_p^2}\ . \tag{S13}$$

For finite positive cost c and numbers of phage species n_p, $\lim_{n_{res}\to 0} g'(n_{res})$ is always negative while $\lim_{n_{res}\to n_p} g'(n_{res})$ is always positive. n_{res} is hence always drawn to the extreme ends of the distribution (compare Fig. 2, main article).

That said, note that in real communities the limit $n_{res} \to n_p$ is essentially impossible to reach if bacteria of initially empty CRISPR-arrays were exposed to a large number of phage species. This would require migration of spacer-sequences against the *restoring force* from $n_{res} = 0$ to $n_{res} = n_p$ and this process would have to be completed before new phage species are produced by mutations. In a well-mixed lab system - where cultures could be initialized with empty CRISPR-arrays and very few phages, the polarized situation, as described above, could in principle be reproduced by *adiabatically* adding new phages and allowing the system to adjust to each insertion. Concludingly, real well-mixed systems will only exhibit an exponentially decaying distribution of spacer lengths near $n_{res} = 0$.

Spatial Modeling

In the following we provide several extensions of the spatial modeling provided in the main text: We explore (i) the transient behavior of phage and spacer diversity, (ii) phage-side fitness costs as associated with variations in their virulence, (iii) temporal fluctuations in populations densities.

(i) Time-dependent population dynamics

In Fig. S3 we present transient behavior of the bacteria-phage community. We contrast the spatial and the well-mixed case and initialize the system with bacteria that have an empty CRISPR-array (no spacers). We then introduce a single phage species which is allowed to spread on the bacteria. The bacteria can acquire CRISPR spacers at a rate r when attacked by the phage, allowing them to subsequently become immune. Also, a phage can mutate at a given rate r_{ph} by which it becomes a new species, i.e. any spacers present in the bacterial CRISPR system do not allow defense against this new phage species. In Fig. S3 we show the mean number of spacers carried by the bacteria, n_{res}, and the number of distinct phage species in the system, n_p. The figure shows the time-dependence of these quantities in units of system updates. A system update means that on average each site has been updated once. The number of distinct phage species initially increases but eventually saturates to a value between 40 and 50, both for the spatial and the well-mixed system. However, the bacterial reaction to the growing n_p is rather different in the two cases: While the spatially-structured bacteria acquire more and more spacers and finally also saturate (in

the figure between 3 and 4 spacers) the well-mixed bacteria only acquire more spacers until n_p reaches 15 and then decay to a value close to zero. This is explained by the payoff of incorporating a spacer: In a well-mixed system the payoff essentially is proportional to $1/n_p$, as the probability of encountering the same phage species again scales reciprocally with the number of species. On the other hand, in the spatially structured system the chances are much higher and nearly independent of n_p, as phages are constrained to the local environment (by migration).

(ii) Fitness-costs in the phage population

So far we have only addressed fitness variations in the bacterial population, as resulting from the acquisition of CRISPR-spacers. However, also the phage population may be subject to variations in fitness, as reflected by its infection parameters. As a point in case, we imagine different phage species that are able to generate different burst sizes, i.e. the number of offspring generated from an event of successful infection and subsequent burst of the bacterium.

In our model, the default is for a burst to result in phage-offspring spreading to all neighboring sites (i.e. four sites on a square lattice). We mimic the effect of varying burst size by an infection-associated probability, p_{bs}, which controls the probability of a particular phage species to spread to a given neighbor. E.g. when $p_{bs} = 1/2$, every neighbor only is infected at a probability 1/2 resulting on average in 2 generated offspring-sites for the given lattice of four nearest-neighbors. When phages mutate as before, they now are both assigned a new unique index and update the parameter p_{bs} by choosing a random number from a uniform distribution between 0 and 1. While this sampling will allow all p_{bs} to be represented equally, different values of p_{bs} will however grant the corresponding phage species different fitness with large values constituting a competitive advantage due to faster duplication. The uniform input distribution may hence be expected to result in a distribution skewed towards large p_{bs} in the evolving system.

However, CRISPR as an adaptive immune system will then be challenged more strongly by the dominant species, i.e. those with larger values of p_{bs}, which will again deplete this sub-population more strongly than those species with low values of p_{bs}. In Fig. S4 we present model results comparing again the spatial and the well-mixed system.

Fig. S4a shows the spatial pattern generated by CRISPR-bacteria interacting with evolving phages as described above. Again, a rather heterogeneous pattern results w.r.t. variations in CRISPR spacer

numbers. Considering again the spacer number distribution for the spatial and well-mixed case (Fig. S4b), qualitatively similar curves to those without fitness variations (Fig. 2, main text) are obtained. Overall, the analysis of phage-fitness variations shows that in the spatially-structured system a distribution of spacer numbers n_p again results, with intermediate values most likely. For the well-mixed system the extreme state of $n_p = 0$ dominates independent of fitness variations.

To quantify how the phage fitness distribution effects the overall phage diversity, we again produce plots of the transient behavior, in the spirit of Fig. S3. Fig. S4c shows the time-dependence of n_p for variable and fixed p_{bs}. Generally, diversity decreases when phages are allowed to lower their burst size by mutations (indicated by vertical arrows in the figure). This is clear when remembering that mutations leading to lower burst size are generally unsuccessful and these phages quickly die out. However, phages in a well-mixed system are punished much more when they lower their burst size, evident from the stronger reduction of diversity as shown in the plot.

Further, we consider the distribution of phage characteristics: The dashed black curve in Fig. S4d shows the distribution of p_{bs} for the spatially-structured system, as obtained by counting the number of corresponding phage species present in the system. The distribution is roughly exponentially increasing as a function of p_{bs}, reflecting the more rapid spreading of high-p_{bs} phages and their out-competing of phages with lower values of p_{bs}. We again contrast with the well-mixed model. Generally, the distribution in the well-mixed system is more sharply increasing towards large p_{bs}. Note also the cross-over of the two curves, possibly a result of the feedback in the spatially-structured system where low-p_{bs} phages are locally out-competed by those of larger p_{bs}, an effect acting on longer time scales in the well-mixed system.

To monitor how the CRISPR system reacts to the varying fitness of the predator, we correspondingly show the time-dependence of the bacterial immunity (quantified by the mean value of n_{res}, Fig. S4e). Again, there is no strong difference in the reaction to the phage population. The spatial system yields intermediate resistance levels - reflecting the locally unchanged diversity - while the well-mixed system yields values of n_{res} close to zero. Also, we sample all spacers in the bacterial population and collect the values of p_{bs} corresponding to the encoded phage species (Fig. S4f, solid black curve). Comparing the slopes of the two black curves in Fig. S4d and f shows that CRISPR in fact reacts most strongly to rapidly spreading phages (high p_{bs}) but allows weak phages higher infection rates (relatively less spacers at small p_{bs}).

(iii) Fluctuations in population density

In the main text we state that for the CRISPR system and sufficiently large system sizes, phage densities of the individual species become comparable. We justify this by a negative feedback induced by CRISPR, i.e. the presence of an above average phage density of a given species leads to the production of more spacers in response to the elevated density. The given phage species will subsequently have a reduced probability to infect a given bacterium. This effect should lead to a restoring of population densities towards the mean.

To show that this is in fact the case, we perform several simulations with a fixed number of phage species $n_p = 20$: For both the well-mixed and the spatial model, we shown time series for systems of linear dimension $L = 200$ and $L = 400$ (shown in Fig. S5a,b). In all simulations, none of the phage species go extinct and their densities fluctuate around a common mean. Initially weak species (low density) quickly increase and converge to the same mean. Larger system sizes lead to a further reduction of the variations in population density.

Comparing with a system with de-activated CRISPR (no spacer-acquisition allowed) population densities fluctuate widely and extinctions are common (Fig. S5c,d). The absence of CRISPR means that the feedback is removed and phage population densities undergo an unrestrained drift, conserving only the overall population total of the phage population, not that of the individual species.

References

1. M. Delbrueck. The burst size distribution in the growth of bacterial viruses (bacteriophages). *J. Bacteriol.*, 50:131–135, 1945.

2. K. E. Wommack and R. R. Colwell. Virioplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews*, 64:69–113, 2000.

3. B.R. Levin, F. M. Stewart, and L. Chao. Resource-Limited Growth, Competition, and Predation: A Model and Experimental Studies with Bacteria and Bacteriophage. *The American Naturalist*, 111:3–24, 1977.

4. R. Holt. Predation, Apparent Competition, and the Structure of Prey Communities. *Theor. Pop. Biology*, 12:197–229, 1977.