Electronic Supplementary Material

For "Unrestricted migration favours virulent pathogens in experimental meta-populations: Evolutionary genetics of a rapacious life history" by Eshelman et al.

Experimental Set-Up

The experimental set-up for the original evolution experiment is shown in Figure S1. More details can be found in Kerr *et al*. (2006).
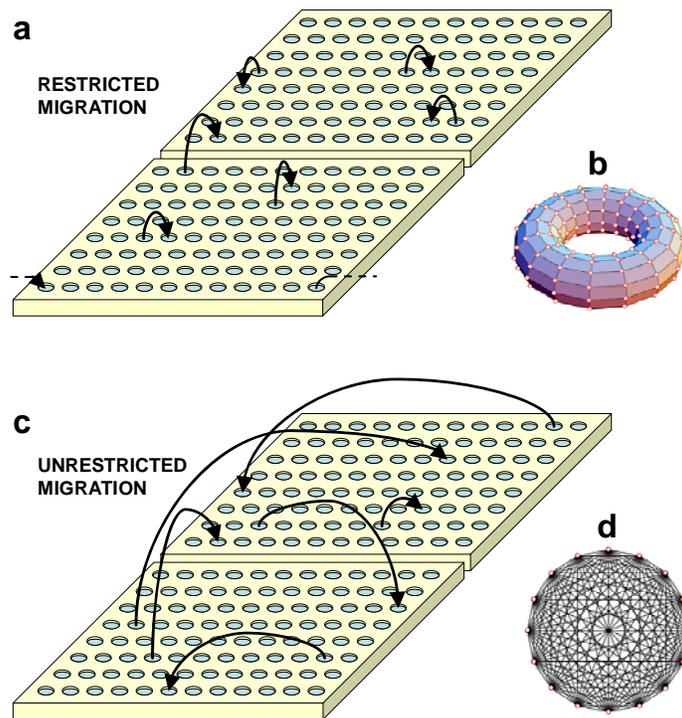
Figure S1. Experimental treatments. Each metapopulation consisted of two 96-well microtiter plates. The topology of migratory connections between wells distinguish the two treatments. (a) Under the Restricted Migration treatment, migrations take place between 'neighboring' wells only. Six migration events are shown for illustration. The metapopulation has wrap-around boundaries, so that every well has exactly four neighbors (thus, the well at the bottom right of the metapopulation can serve as a source of migrants to the well at the bottom left of the metapopulation as these wells are neighbors). (b) Topologically, the network of potential migratory connections in the Restricted treatment occupies a torus; wells are represented as red nodes and migratory connections are black edges. We note that only a fraction of possible migrations actually take place at each transfer. (c) Under the Unrestricted Migration treatment, migrations can take place between any two wells. Six migration events are shown for illustration. (d) The network of potential migratory connections in the Unrestricted treatment forms a complete graph. Because this is difficult to visualize with 192 nodes, we show the sub-network of 16 wells (red nodes) and their interconnections (black edges). Every well is connected to every other well. Again, only a fraction of possible migrations actually take place at each transfer, but the expected number of migrations is fixed across treatments.

Supplemental methods

Strains and media: Experimental evolution was performed with lytic phage T4 and *Escherichia coli* B (see Kerr *et al*. (2006) for strain details). Bacteria and phage were grown in minimal glucose liquid medium (MG) supplemented with streptomycin and novobiocin (the bacterial host was resistant to these antibiotics). For titer estimation, bacteria were plated on LB agar and phage were plated with abundant host in LB top agar over an LB hard agar base. For the competition assays, an *rII* T4 mutant and *E. coli* K-12 λ lysogen were used (Kerr *et al*. 2006). For the genetic engineering, TOP10 cells were used (Invitrogen).

Acclimation of bacteria and phage: For some of the assays, bacteria and phage were acclimated to the microtiter well environment. Bacterial populations were propagated over two 12 hour cycles in microtiter wells with MG medium. At each transfer, 20μL of fully grown culture was added to 180μL of fresh medium. Phage were amplified over a single 12 hour period in microtiter wells. To initialize the amplification, 20μL of fully grown bacteria was added to 180μL of MG medium with enough phage to achieve an initial MOI≈1. When well-based amplification of the phage did not produce a sufficiently high titer for the assay, amplification was carried out in a flask. For the adsorption assay, acclimated phage were separated from their bacterial hosts after amplification by adding the contents of the microtiter well to 30μL of chloroform, mixing with a pipette, and spinning the contents for 2 minutes at 1300 RPM in an Allegra X-15R centrifuge (Beckman-Coulter). Acclimated phage were then isolated from the supernatant.

PCR and sequencing: A single phage plaque was mixed in 40μL of 0.01M Tris buffer in a microfuge tube and left at room temperature for 1.5 hours. 16μL of this phage mixture was placed in a PCR tube on ice with 24μL of a master mix, containing 3.2μL dNTPs (10mM), 4.0μL PCR buffer (10X), 1.4μL $MgCl_2$ (100mM), 1.0μL Taq polymerase (5U/μL, Genscript), 10.4μL of Milli-Q purified H_2O (Millipore), and 2.0μL of forward and reverse primers (10μM), which are listed under reactions 1 and 2 in Table 1. This mixture experienced 30 cycles of the following steps: 94°C for 1 minute, 54°C for 1 minute, and 72°C for 1 minute. 3.0μL of cleaned-up PCR product (QIAquick PCR Purification Kit, Qiagen) was added to a PCR tube on ice with 3.6μL of H_2O, 1.0μL of dilution buffer (5X), 2.0μL of 3.1 Big Dye (ABI), and 0.4μL of the primer of interest (10μM). After being briefly centrifuged, the mixture experienced a 2 minute hot-start at 96°C, followed by 25 cycles of the following steps: 96°C for 30 seconds, 50°C for 15 seconds, and 60°C for 4 minutes. After further processing (Sephadex dye-removal, drying and formamide resuspension), the sample was loaded on a PRISM 3100 genetic analyzer (ABI) and manufacturer instructions were followed. Sequences were analyzed with Sequencher (version 4.6).

Genetic engineering: A single phage plaque from an isolate containing the *rI* mutation of interest was mixed in 40μL of 0.01M Tris buffer in a microfuge tube and left at room temperature for 1.5 hours. 10μL of this phage mixture was placed in a PCR tube on ice

with 1μL dNTPs (10mM), 10μL Phusion Buffer (5X), 0.5μL Phusion polymerase (2U/μL, New England BioLabs), 23.5μL of H$_2$O, and 2.5μL of forward and reverse primers (10μM), which are listed under reaction 3 in Table 1. After 30 seconds at 98°C, this mixture experienced 25 cycles of the following steps: 98°C for 5 seconds, 63°C for 30 seconds, and 72°C for 1 minute. To add adenosine overhangs, the PCR product was gel purified using a QIAquick Gel Extraction Kit (Qiagen) and 15μL of the purified product was incubated with 0.5μL Taq polymerase, 2.5μL buffer (10X), 0.5μL dNTPs (10mM), and 6.5μL of H$_2$O at 72°C for 10 minutes. The amplicon was cloned into the pCRII-TOPO vector and TOP10 *E. coli* cells were subsequently transformed using the TOPO TA Cloning Kit (Invitrogen). The presence of the amplicon was confirmed following restriction digest with EcoRI and band confirmation using agarose gel electrophoresis. Bacteria containing the plasmid with insert were infected with wild-type T4. Double recombination between the insert and phage genome introduces the mutation into a common T4 background. The recombinants were identified by their large clear plaque phenotype. Presence of the correct mutation was confirmed by PCR and sequencing. Further PCR reactions (4-7 in Table 1) and sequencing were done to confirm that no inappropriate alterations occurred in the region of the original amplicon (see Fig. S2).

Table 1: PCR primers

Reaction	Amplified Genes	Forward Primer	Reverse Primer
1	t	ggcctgaaaaagaacagaagccttgg	gggttttgagggtgtatatcg
2	rl	ccactttgtgaaaagtcgctg	taccttgataaagttaaggccg
3	mobD.5; rl.-1; rl; rl.1	atgcgttcctcatctgcttt	tgaagtcgtttctcgcattg
4	mobD.5; rl.-1; rl	atgcgttcctcatctgcttt	taccttgataaagttaaggccg
5	rl; rl.1	ccactttgtgaaaagtcgctg	tgaagtcgtttctcgcattg
6	mobD.4	caaactgttcgcgtgttagc	cgttataatatgtgtgatgaagttgc
7	tk (partial)	ggctcaattgatgcacctgt	tgaatgctggaaaatctgctt
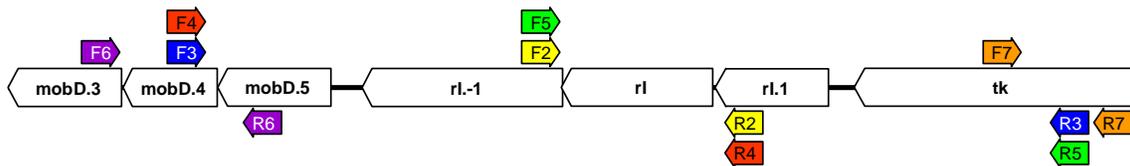
Figure S2. A section of the T4 genome spanning position 58155 to 60344 (genes are shown as pylons). The primers used for PCR and sequencing in the process of engineering the *rl* mutants are shown as arrows (numbers refer to the reactions in Table 1).

Mutations at the rI locus

We found a total of six different mutations at the *rI* locus through sequencing 128 random isolates from various metapopulations. Table 2 provides information about each of these mutations and Table 3 gives the number isolates with each mutation found in each replicate metapopulation in our Unrestricted and Restricted Migration treatments. Note that a total of 16 random isolates were sequenced from every replicate metapopulation. We see that about ¾ of the mutations at this locus involve single base deletions or insertions and nearly all of these occur in homopolymeric repeat regions (see Table 2). These frameshift mutations dramatically affect the amino acid sequence of the RI protein: Figure S3 shows the protein sequences for the two most common mutations, which are deletions. It is likely that such proteins are non-functional. An intriguing hypothesis is that these homopolymers are hypermutable, such that T4 can "snap" into and out of virulent phenotypes through reversible frameshift mutations. Such a bet-hedging strategy may be advantageous when conditions that influence the value of virulence continually change. Future work will explore the hypothesis that *rI* is a contingency locus.

Table 2: Information about rI Mutations

Mutation name	Mutation type	Genome location	Ancestor sequence	Mutant sequence	Amino acid change(s)	Length of RI protein
rIA	deletion	59400	AAA	AA:	(frameshift)	49
rIB	deletion	59299	AAAAA	AAAA:	(frameshift)	75
rIC	insertion	59406.1	TTTT:	TTTTT	(frameshift)	30
rID	deletion	59390	AATCT	AA:CT	(frameshift)	49
rIE	ns substitution	59261	GGC	CGC	G78R	97
rIF	ns substitution	59249	GCA	CCA	A82P	97

Table 3: Origin of rI Mutants

Mutation name	Number of *rI* Mutants								
	Unrestricted Migration				Restricted Migration				
	Rep 1	Rep 2	Rep 3	Rep 4	Rep 1	Rep 2	Rep 3	Rep 4	
rIA	1	2		4	1				8
rIB		1		1					2
rIC	1								1
rID							1		1
rIE		1							1
rIF		1							1
	2	5	0	5	1	0	1	0	Totals

a. wt

`MALKATALFAMLGLSFVLSPSIEA`NVDPHFDKFMESGIRHVYMLFENKSVESSEQFYSFMRTTYKNDPCSSDFECIERGAEMAQSYARIMNIKLETE

b. rIA

`MALKATALFAMLGLSFVLSPSIEA`NVDPHFDNLWNLVLGTFICFLKIKA

c. rIB

`MALKATALFAMLGLSFVLSPSIEA`NVDPHFDKFMESGIRHVYMLFENKSVESSEQFYSFMRTTYKMTRALLILNV

Figure S3. Amino acid sequences of the two most frequent *rI* mutants. (a) The amino acid sequence of the wild type. The business end of the antiholin is the C-terminal domain (28) (unboxed portion). (b) Over 90% of this domain is altered by the frameshift in the *rIA* mutant (new amino acids in red). (c) Under 50% of this domain is altered by the frameshift in the *rIB* mutant (new amino acids in orange). Both frameshifts cause premature termination of the protein. The N-terminal domain (boxed), which is important in the secretion of the antiholin in membrane-tethered form and its subsequent release into the periplasm (34), is unaffected by both mutations.

Statistical analysis of rI frequency

The number of *rI* mutants discovered in each treatment is given in Table 3. Let the number of isolates from replicate metapopulation i in the Unrestricted Migration treatment be given by n_i^{U} and the number of isolates from replicate metapopulation i in the Restricted Migration treatment be given by n_i^{R} (we note that $n_i^{\mathrm{U}} = n_i^{\mathrm{R}} = 16$ for all i in Table 3). Here we consider two models:

Model A: The number of *rI* mutants in the set of isolates from replicate metapopulation i from the Unrestricted Migration treatment is a binomially distributed random variable with parameters f_{U} and n_i^{U}. The number of *rI* mutants in the set of isolates from replicate metapopulation i from the Restricted Migration treatment is a binomially distributed random variable with parameters f_{R} and n_i^{R}.

Model B: The number of *rI* mutants in the set of isolates from replicate metapopulation i from the Unrestricted Migration treatment is a binomially distributed random variable with parameters f and n_i^{U}. The number of *rI* mutants in the set of isolates from replicate metapopulation i from the Restricted Migration treatment is a binomially distributed random variable with parameters f and n_i^{R}.

Here, the *f* parameters represent the frequency of *rI* mutants. We note that Model B is nested within Model A. Specifically, Model A is Model B when $f_{\mathrm{U}} = f_{\mathrm{R}} = f$ (which forms

the hypothesis we will test). If there are r replicates (r=4 in our case), the maximum likelihood estimates for the f parameters are:

$$\hat{f}_U = \frac{\sum_{i=1}^{r} m_i^{U}}{\sum_{i=1}^{r} n_i^{U}}, \qquad \text{[S1]}$$

$$\hat{f}_R = \frac{\sum_{i=1}^{r} m_i^{R}}{\sum_{i=1}^{r} n_i^{R}}, \qquad \text{[S2]}$$

$$\hat{f} = \frac{\sum_{i=1}^{r} \left(m_i^{U} + m_i^{R} \right)}{\sum_{i=1}^{4} \left(n_i^{U} + n_i^{R} \right)}, \qquad \text{[S3]}$$

where the m parameters give the number of rI mutants among the isolates. Given these MLE values, the likelihood ratio can be computed:

$$LR = 2\ln \frac{\ell_A(\hat{f}_U, \hat{f}_R | data)}{\ell_B(\hat{f} | data)},$$

where ℓ_A and ℓ_B are the likelihood of models A and B, respectively. This LR test statistic is χ^2 distributed with one degree of freedom (for a sufficiently large sample size).

For the initial set of evolved isolates (where $n_i^{U} = n_i^{R} = 4$), we have $\hat{f}_U = 0.25$, $\hat{f}_R = 0.0625$, $\hat{f} = 0.078125$, and LR=2.26147. Thus, we do not reject the hypothesis $f_U = f_R = f$ (p=0.1326). However, after including the additional isolates (the full data shown in Tables 2 and 3), we have $\hat{f}_U = 0.1875$, $\hat{f}_R = 0.03125$, $\hat{f} = 0.109375$, and LR=8.8032. Here, we reject the hypothesis $f_U = f_R = f$ (p=0.003). Thus, with the full data set, the frequency of rI mutants in the Unrestricted Migration treatment is significantly greater than the frequency in the Restricted Migration treatment.

We can also analyze the data using a generalized linear mixed effects approach, where presence of the rI mutation in an isolate is the response variable (0 or 1), the migration treatment is a fixed factor, replicate is a random factor, and we assume that errors are binomially distributed. Using this approach, we find that the rI mutant frequency in the Unrestricted Migration treatment is significantly greater than the rI mutant frequency in the Restricted Migration treatment (z=-2.0676, p=0.03868).

Optimal foraging theory and optimal latent period

To start, we imagine an organism foraging in a patchy environment. The travel time between patches is d. For a residence time of t in a patch, the organism obtains $g(t)$ calories (with $g(0) = 0$, $g'(t) > 0$, and $g''(t) < 0$ for all $t \geq 0$). Here we will investigate the residence time, t^*, that maximizes the rate of energetic gain $r(t) = g(t)/(d+t)$.

The derivative of r with respect to residence time is found by using the quotient rule:

$$\frac{dr}{dt} = \frac{g'(t) \cdot (d+t) - g(t)}{(d+t)^2} \qquad \text{[S4]}$$

The maxima, minima, and zero-slope inflection points of r are found by setting this derivative equal to zero. This occurs when $g'(t) = g(t)/(d+t)$. This equality has a solution where the instantaneous rate of calorie intake equals the long-term rate of calorie intake. Let us assume that this equality does have a solution—labeled t^*. Consider the second derivative of r:

$$\frac{d^2 r}{dt^2} = \frac{g''(t) \cdot (d+t)^2 - 2g'(t) \cdot (d+t) + 2g(t)}{(d+t)^3}. \qquad \text{[S5]}$$

At t^*, we know $g'(t^*) \cdot (d+t^*) = g(t^*)$. Therefore

$$\left. \frac{d^2 r}{dt^2} \right|_{t=t^*} = \frac{g''(t^*)}{d+t^*} < 0, \qquad \text{[S6]}$$

which means that t^* gives a local maximum for $r(t)$.

Because $g'(t) > 0$ and $g''(t) < 0$ for all $t > 0$, we know that $\gamma(t) = g'(t)$ is an invertible function (and this inverse is well-defined for at least some positive range of its independent variable). Let $\gamma^{-1} = \Gamma$. We know that (where defined) $\Gamma'(x) < 0$. In particular, $\Gamma(r(t^*)) = t^*$ and the derivative of Γ at $r(t^*)$ is negative. By the marginal value theorem, we can express the optimal latent period as follows:

$$t^* = \Gamma\left(\frac{g(t^*)}{d+t^*}\right) \qquad \text{[S7]}$$

By taking the derivative of both sides with respect to d (the dispersal time) and simplifying, we have:

$$\frac{\partial t^*}{\partial d} = \Gamma'\left(\frac{g(t^*)}{d+t^*}\right) \cdot \left(\frac{-g(t^*)}{(d+t^*)^2}\right) \qquad \text{[S8]}$$

Since $\Gamma'[g(t^*)/(d+t^*)] < 0$ and $-g(t^*)/(d+t^*)^2 < 0$, then $\partial t^* / \partial d > 0$, which means that as the dispersal time increases, the optimal latent period also increases.

Now, let $g(t) = a \cdot h(t)$, where a is a positive scalar and $h(t)$ has the same properties (with regards to the signs of its derivatives) as $g(t)$. By taking the derivative of t^* with respect to a we can discern the effect of increasing the "quality" of the patch on optimal residence time. We have

$$\frac{\partial t^*}{\partial a} = \Gamma'\left(\frac{g(t^*)}{d+t^*}\right) \cdot \left(\frac{h(t^*)}{d+t^*}\right) \quad\quad \text{[S9]}$$

Since $\Gamma'[g(t^*)/(d+t^*)] < 0$ and $h(t^*)/(d+t^*) > 0$, then $\partial t^* / \partial a < 0$, which means that as patch quality increases, the optimal residence time decreases.

For our "foraging phage," the time from host lysis to subsequent infection is D (this includes the time for the phage particle to diffuse to its new host and bind irreversibly). For simplicity, we assume that D is a constant (of course, this assumption is violated under our experimental conditions). The latent period, L, is similar to the residence time above. Here we assume that the phage gets an offspring burst size of $B(L)$. Previous work (Bull et al. 2004, Hutchison & Sinsheimer 1966, Josslin 1970, Wang 2006) has shown that B is an increasing function of L. For instance, if E is the eclipse period, one reasonable description of burst size is:

$$B(L) = \begin{cases} 0 & \text{if } L \leq E \\ m(L-E) & \text{if } L > E \end{cases} \quad\quad \text{[S10]}$$

where $m>0$. How does the relationship in [S10] jibe with our data? There are only three strains considered in our engineered strains (the ancestor and two *rI* mutants). However, there is no obvious misfit between our data and the above function (see Figure S4).

A phage lineage with latent period L will increase at the rate $G = [B(L)]^{1/(D+L)}$. The latent period that maximizes G will also maximize $\ln(G) = \ln[B(L)]/(D+L)$ (as ln is a monotonic increasing function). The derivative of $\ln(G)$ with respect to L is

$$\frac{d\ln(G)}{dL} = \frac{\dfrac{d\ln[B(L)]}{dL} \cdot (D+L) - \ln B(L)}{(D+L)^2} \quad\quad \text{[S11]}$$

.

Again, the maxima, minima, and zero-slope inflection points of $\ln(G)$ are found by setting this derivative equal to zero. This occurs when

$$\frac{d\ln[B(L)]}{dL} = \frac{\ln[B(L)]}{(D+L)} \quad\quad \text{[S12]}$$
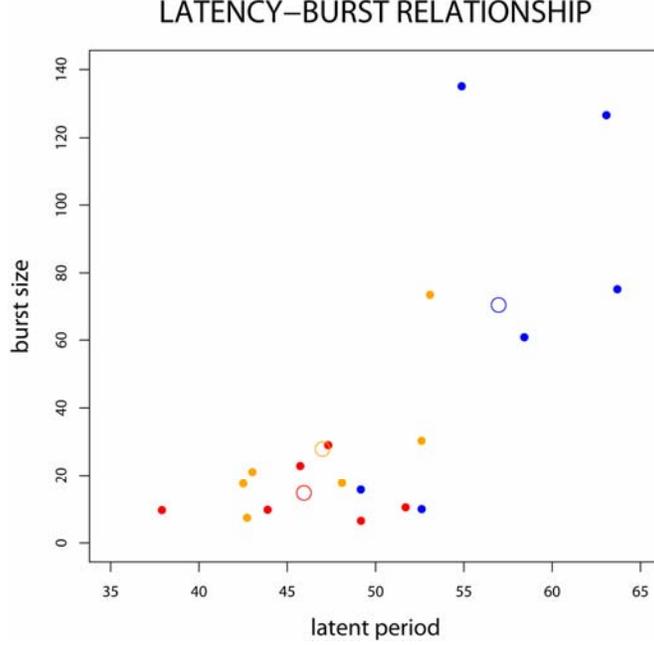
Figure S4. The relationship between burst size and latent period. Solid circles give data from one-step assays and open circles are the mean values across replicates. The red points refer to the *rIA* engineered mutant, the orange points refer to the *rIB* engineered mutant and the blue points refer to the ancestor phage.

The equality given in [S12] has a solution where the instantaneous rate of log-progeny production equals the long-term rate of log-progeny production. Let us assume that this equality does have a solution—labeled L^*. Consider the second derivative of $\ln(G)$:

$$\frac{d^2 \ln(G)}{dL^2} = \frac{\dfrac{d^2 \ln[B(L)]}{dL^2} \cdot (D+L)^2 - 2 \cdot \left(\dfrac{d \ln[B(L)]}{dL} \cdot (D+L) - \ln[B(L)] \right)}{(D+L)^3} \qquad [S13]$$

At L^*, we know $(D+L^*) \cdot d \ln[B(L^*)]/dL = \ln[B(L^*)]$, therefore,

$$\left. \frac{d^2 \ln(G)}{dL^2} \right|_{L=L^*} = \frac{1}{(D+L^*)} \cdot \left(\frac{d^2 \ln[B(L^*)]}{dL^2} \right) \qquad [S14]$$

Further, we have

$$\frac{d^2 \ln[B(L)]}{dL^2} = \frac{B''(L) \cdot B(L) - [B'(L)]^2}{B^2(L)} \qquad [S15]$$

9

Generally, we will assume $B(L) \geq 0$, $B'(L) \geq 0$, $B''(L) \leq 0$, for all $L>0$ (or wherever these derivates are well-defined). This is the case for [S10], where the derivatives are defined. If we assume a few strict inequalities hold at $L = L^*$: $B(L^*) > 0$, $B'(L^*) > 0$, $B''(L^*) \leq 0$, then $d^2 \ln[B(L^*)]/dL^2 < 0$. Therefore,

$$\left. \frac{d^2 \ln(G)}{dL^2} \right|_{L=L^*} < 0,$$

and L^* will be a local maximum.

Using an approach similar to that outlined for the optimal forager, we can show that as inter-host dispersal time increases or host quality decreases, the optimal latent period for the phage increases.