Mutation, selection, and quantitative genetic architecture of susceptibility to bacterial pathogens in *C. elegans*

Lindsay M. Johnson[1,2,*], Sayran Saber[1,*], Md. Monjurul Islam Rifat[1], Sydney Rouse[1], and Charles F. Baer[1,3,†]

1 – Dept. of Biology, University of Florida, Gainesville, FL 32611 USA

2 – Present address: Resilience Biotechnologies, Inc., Alachua, FL, 32615 USA

3 - University of Florida Genetics Institute, Gainesville, FL 32611 USA

* - These authors contributed equally

† - Communication to: cbaer@ufl.edu

Abstract

Understanding the evolutionary and genetic underpinnings of susceptibility to pathogens is of fundamental importance across a wide swathe of biology. Much theoretical and empirical effort has focused on genetic variants of large effect, but pathogen susceptibility often appears to be a polygenic complex trait. Here we investigate the quantitative genetics of survival over 120 hours of exposure ("susceptibility") of *C. elegans* to three bacterial pathogens of varying virulence, along with the non-pathogenic OP50 strain of *E. coli*. We compare the genetic (co)variance input by spontaneous mutations accumulated under minimal selection to the standing genetic (co)variance in a set of ~50 wild isolates. Three conclusions emerge. First, with one exception, mutations increase susceptibility to pathogens, and susceptibility is uncorrelated with fitness in the absence of pathogens. Second, the orientation in trait space of the heritable (co)variance of wild isolates is sufficiently explained by mutation. However, pathogen susceptibility is clearly under purifying, apparently directional, selection of magnitude similar to that of competitive fitness in the MA conditions. The results provide no evidence for fitness tradeoffs between pathogen susceptibility and fitness in the absence of pathogens, nor that balancing selection is important in maintaining genetic variation for susceptibility to these bacterial pathogens.

Introduction

Infection by pathogens is historically the leading cause of human mortality, and pathogens continue to exact an enormous societal toll. Moreover, the relationship between pathogens and their hosts holds an exalted place in evolutionary biology [1, 2]. It is well-established that susceptibility of individuals to the harmful effects of pathogens often has a genetic basis. In some cases, the genetic basis of variation in host susceptibility to pathogens can be attributed to variants at one or two loci. For example, variants at a few loci in the human genome (e.g., *HBB*, *ABO*, *G6PD*) are famously implicated in resistance to malaria [3]. The notion that *particular* (multilocus) host genotypes confer resistance or susceptibility to *particular* pathogen genotypes has had outsize influence in evolutionary biology, because "Red Queen" models of the evolution of sex and recombination are predicated on fluctuating epistasis [4], with negative frequency-dependent selection resulting from a host-pathogen arms-race providing the most obvious plausible scenario [2].

Often, however, the heritable basis of susceptibility to a particular pathogen appears to be polygenic, with a substantial fraction of unexplained heritability [5-8], sometimes even in cases with segregating large-effect loci (e.g., malaria [9]). Pathogens may thus impose significant selection even in the absence of Red Queen dynamics. It seems reasonable, then, to consider "susceptibility to pathogen X" as a generic complex trait; in some cases, there will be loci of large effect that explain much of the variance in susceptibility, but other times there will not.

Two fundamental issues in understanding the biology of complex traits, are (i) what is the rate of input of genetic variance (and covariance) by mutation? and (ii) what is the relationship of the trait(s) with fitness? Taken together, these two considerations present a hierarchy of evolutionary hypotheses. The simplest explanation for any pattern of (co)variation is neutral evolution *sensu stricto*, i.e., the pattern can be sufficiently explained by neutral mutation and random genetic drift. If the pattern cannot be plausibly explained by neutral

58 evolution, the next-simplest explanation is purifying selection against unconditionally deleterious
59 mutations, plus drift. Only if those two explanations prove unsatisfactory need more
60 complicated explanations be considered; for example, scenarios of balancing or positive
61 selection.

62 It is a bedrock principle of neutral theory that polymorphism within a group is expected to
63 be proportional to divergence between groups if the only evolutionary forces involved are
64 mutation, drift, and uniform purifying selection [10-12]. For DNA sequences, the proportionality
65 is between nucleotide diversity ($\theta \approx 4N_e\mu$ in diploids) and divergence between taxa ($D=2\mu t$),
66 where N_e is the genetic effective population size, μ is the per-nucleotide mutation rate and t is
67 the time since divergence, in generations. From classical population genetics theory, the
68 frequency of a deleterious allele at mutation-selection balance (MSB) is $\hat{q} \approx \frac{\mu}{s}$, where s is the
69 strength of purifying selection [13] (in the case of near-complete selfing, as in *C. elegans*, the
70 relevant selection coefficient is the homozygous effect). The strength of purifying selection
71 and/or the mutation rate may differ between genes, but the proportionality is not affected.

72 For quantitative traits, the analog of the per-nucleotide mutation rate, μ, is the mutational
73 variance $V_M=U\alpha^2$, where $U = \sum_i \mu_i$ is the genomic mutation rate and α^2 is the average squared
74 effect of a new mutation on the trait [11, 14, 15]. Different traits represent the quantitative
75 genetic analog to different genes at the molecular level. For a neutral trait at mutation-drift
76 equilibrium, the genetic variation segregating within a population $V_G=2N_eV_M$, and populations will
77 diverge asymptotically at rate $2V_M$ [11]. For traits under purifying selection (directional or
78 stabilizing), at MSB $V_G \approx \frac{V_M}{s}$, where s is the average strength of purifying selection against a
79 mutant allele that affects the trait [15-17].

80 Obviously, no trait is an island [18], which is to say that the evolutionary fate of an allele
81 x that affects trait Z depends not only on its effect on Z (plus drift) but also on its pleiotropic
82 effects on all other selected traits, and also on the cumulative effects of alleles with which x is in

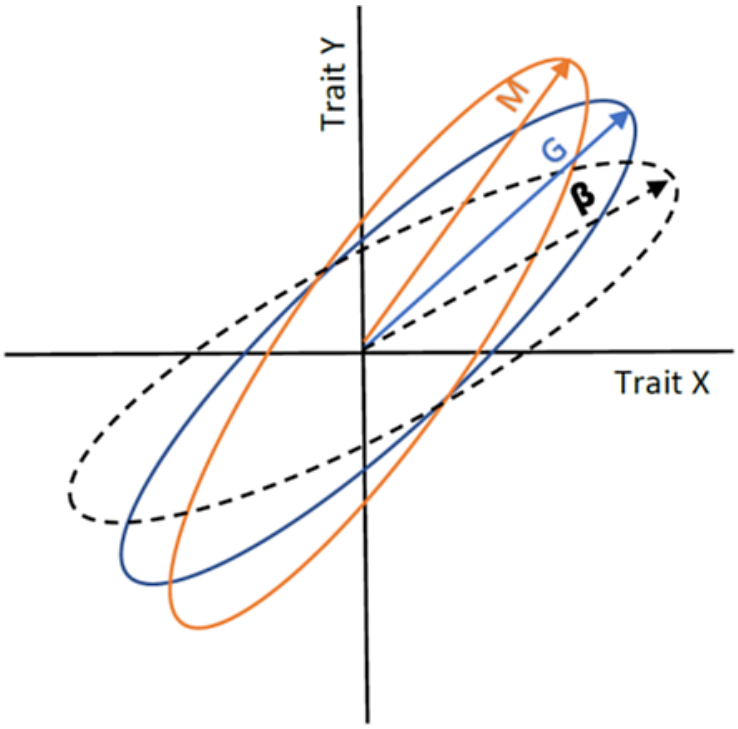
linkage disequilibrium. The relationship $V_G \approx \frac{V_M}{s}$ implicitly includes both the direct selective effects of mutations on the trait in question, but also the pleiotropic effects of selection on correlated traits.

The multivariate extensions of V_G and V_M are the standing genetic and mutational variance-covariance matrices, **G** and **M**, of which the diagonal elements are the variances, $V_{G,i}$ and $V_{M,i}$ respectively, and the off-diagonal elements are the genetic/mutational covariance between traits *i* and *j*, $Cov_{G,ij}$ and $Cov_{M,ij}$ [19]. The covariances capture the cumulative effects of pleiotropy and linkage disequilibrium.

In a finite population, random genetic drift removes genetic (co)variance at rate $(1 - \frac{1}{2N_e})$ per generation. Thus, for neutral traits at mutation-drift equilibrium, the expectation is $\mathbf{G}=2N_e\mathbf{M}$ with random mating and $4N_e\mathbf{M}$ with complete selfing [11]. For traits under strong (quasi-deterministic) selection, the evolution of **G** can be (approximately) captured in the relationship $\Delta\mathbf{G} = \mathbf{G}\left(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\beta}^{\boldsymbol{T}}\right)\mathbf{G} + 2\mathbf{M} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\left(\boldsymbol{C}_{ij} - \boldsymbol{C}'_{ij}\right)$ [19, 20]. The first term in parentheses describes the composite effects of selection, both directional ($\boldsymbol{\beta}$) and quadratic (the matrix of stabilizing and correlational selection gradients, $\boldsymbol{\gamma}$), and the second term in parentheses describes the decay by recombination of genetic covariance resulting from linkage disequilibrium built up by selection and mutation, where r_{ij} is the recombination rate between loci *i* and *j* (the effects of LD are usually sufficiently weak to where they can be ignored [20]). At MSB, $\Delta\mathbf{G}=0$, so the deviation between **G** and **M** represents the cumulative effects of selection (**Figure 1**). Unfortunately, there is no stochastic theory of the evolution of **G** equivalent to that for discrete loci [21, 22], so the hypothesis of neutral evolution as an underlying cause of an observed deviation between **G**-matrices (e.g., different populations or species), or between **G** and **M**, cannot be formally tested [23]. However, a significant difference between **G** and **M** is sufficient to reject the hypothesis that mutation alone is a sufficient predictor of **G**; at minimum,

107 genetic drift has caused **G** to evolve a different genetic architecture than expected from
108 mutation in the absence of other evolutionary forces.

109 Here we report results of a set of experiments in which we (i) estimate the cumulative
110 effects of spontaneous mutations on the susceptibility of two strains of *C. elegans* to three
111 bacterial pathogens, and (ii) estimate the standing genetic (co)variance for susceptibility to
112 those same three pathogens from a set of ~50 *C. elegans* wild isolates. Our first, broad goal is
113 to determine the realms of consistency and of idiosyncrasy in the mutational process in the
114 context of host-pathogen evolution. Our second, broad goal is to draw inferences about the
115 strength and pattern of natural selection acting on *C. elegans*' susceptibility to bacterial
116 pathogens, relative to the inferred strength of selection on competitive fitness.
117



118

119 **Figure 1**. Depiction of a hypothetical relationship between two traits, X and Y. The orange
120 ellipse represents the input of genetic (co)variance by mutation (**M**), the black dashed ellipse

121 represents the trajectory of bivariate directional selection ($\boldsymbol{\beta}$), and the blue ellipse represents the
122 genetic (co)variance matrix (**G**) at mutation-selection balance. After [20].

123

124 **Materials and Methods**

125 *Mutation Accumulation (MA) experiments – basic principles*

126 A mutation accumulation experiment is simply a pedigree, in which replicate populations derived
127 from a known, (ideally) genetically-homogeneous progenitor are propagated under conditions in
128 which natural selection is minimally effective [24, 25]. Typically, selection is minimized by
129 minimizing N_e; mutations with fitness effects $s<1/4N_e$ are effectively neutral [26].

130 MA experiments can be used to illuminate the workings of natural selection in two ways.
131 First, they provide the most straightforward and least assumption-laden way to estimate
132 mutational (co)variances, which can be compared to the standing genetic (co)variances as
133 outlined in the Introduction. Second, if selection is directional, or at least has a directional
134 component, mutational bias (i.e., $\Delta M \neq 0$) will generally point in the opposite direction of
135 selection. On average, mutations are deleterious. Thus, fitness declines under MA conditions.
136 Traits that are positively correlated with fitness should, on average, exhibit a negative mutational
137 bias; traits that are negatively correlated with fitness should, on average, exhibit a positive
138 mutational bias. By way of comparison, the same principle underlies inbreeding depression,
139 wherein rare recessive alleles produce a deleterious phenotype when homozygous.

140

141 *Mutation Accumulation (MA) lines*

142 Details of the mutation accumulation experiment are given in [27]. N2 is the standard laboratory
143 strain of *C. elegans*; PB306 is a wild isolate graciously provided by Scott Baird. The basic
144 protocol follows that of Vassilieva and Lynch [28] and is outlined in **Figure 2**. Briefly, replicate
145 populations (MA lines) were initiated from a cryopreserved stock of a highly-inbred ancestor at

mutation-drift equilibrium and propagated by transferring a single immature hermaphrodite at each generation (four-day intervals). Worms were maintained on 60mm NGM agar plates, spotted with 100 *µl* of an overnight culture of the OP50 strain of *E. coli* B, at a constant 20°C. The lines were propagated for 250 transfers (G_{250}), beginning in March, 2001 and culminating with a final cryopreservation in 2005.

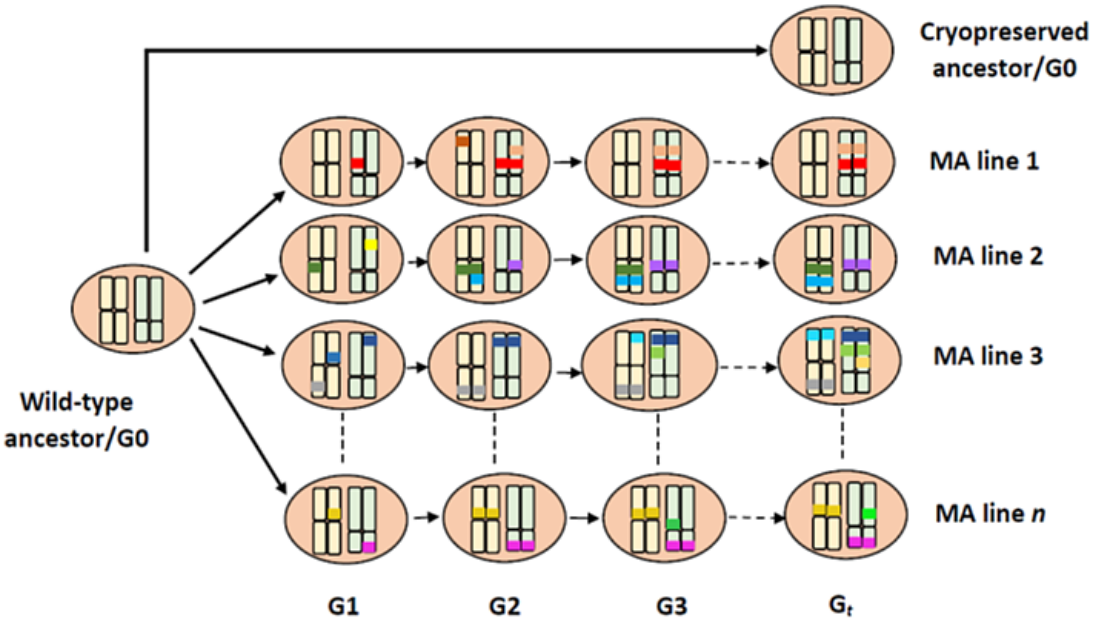


Figure 2. Schematic depiction of the MA experiment. Ovals represent individuals, with two pairs of chromosomes (yellow, green). Colored bars on chromosomes represent mutations unique to each MA line, which accumulate over the course of the *t* (≈250) generations of MA; the ancestor is taken to be genetically uniform and homozygous at all loci. At each generation (G1, G2, ...G_t), a single individual is propagated to a new plate to found the next generation.

At every transfer, the prior two generations of each MA line were kept at 20°C as backups; if the focal worm did not reproduce, the plate was reinitiated with an immature individual from the previous generation; we refer to this as "going to backup". If the focal worm reproduced but no offspring had reached the L2 stage, we kept the offspring in the experiment

to reproduce in the next generation; we refer to this delay in reproduction as "holding over". Holding over reduces the actual number of generations of evolution of an MA line below the maximum (G_{250}). In principle, going to backup should have no effect on the total number of generations of MA, because the backup worm should be a (double) first cousin of the worm that would have been transferred. However, there is some opportunity for overlapping generations on backup plates (e.g., if we had to go to backup twice or more in a row), so there is some ambiguity with respect to the actual number of generations of MA. The upper bound on the number of generations is the total number of possible transfers in the experiment (G_{250}); the lower bound is the number of successful transfers (G_{MIN}). Data on transfers and additional context is given in **Supplemental Table S1**.

Of the 100 N2 and PB306 MA lines initiated in 2001, 73 N2 MA lines and 71 PB306 remained available at the time these experiments were initiated in 2015. Of those lines, whole genome sequencing subsequently revealed that four pairs of the N2 lines and six pairs of the PB306 lines were cross-contaminated at some point in the MA process. Accordingly, we discarded one line from each pair, retaining the line for which the total sample size was larger. On average, each MA line carries about 65 spontaneous base-substitution and small indel mutations [29], with probably a few larger structural variants (Saxena and Baer, unpublished results).

The cryopreserved progenitor (G0) serves as a control in two ways. The difference between the trait mean of the G0 and MA treatments represents the cumulative effect of spontaneous mutations on the trait mean. If the G0 progenitor is subdivided into replicate "pseudolines" (PS) which are subsequently treated identically to the MA lines (**Figure 2**), the difference between the among-MA line and the among-PS line components of variance represents the cumulative heritable variance resulting from the accumulation of spontaneous mutations (and potentially epimutations, depending on the experimental design [30, 31]). PS lines were constructed by thawing a sample of the cryopreserved G0 progenitor onto an agar

plate. The next day, individual L4 stage hermaphrodites were picked singly to new plates, each of which was designated a PS line. PS lines were subsequently treated as if they were MA lines. Unfortunately, due to a management error on the part of the senior author (CFB), not all pathogen treatments include similar numbers of pseudolines (see below).

Competitive fitness assay

Competitive fitness of the PB306 MA lines and their G0 progenitor was assayed in two blocks, beginning in May, 2005. The details of that assay have been published elsewhere [32], and are reprised in the **Supplemental Material**. Fitness assay data are given in **Supplemental Table S2**.

Wild isolates

A collection of wild isolates of *C. elegans* was obtained from Erik Andersen (Northwestern University) in 2015 and cryopreserved in the Baer lab. A list of the wild isolates is given in **Supplemental Table S3**. The genome sequences of the wild isolates along with collection information are available at https://www.elegansvariation.org/.

Pathogen infection and survival assay

We used four strains of bacteria; three documented pathogens (the PA14 strain of *Pseudomonas aeruginosa*, the NCTC8325 strain of *Staphylococcus aureus*, the OG1RF strain of *Enterococcus faecalis*) and the standard laboratory worm food, the OP50 strain of *E. coli*. We use the term "pathogen" to refer to the three pathogenic species, and "bacteria" when OP50 is included. Each bacterial species was assayed in three or four experimental blocks. Pathogens were acquired from Erik Andersen and kept frozen at -80°C until use. The basic protocol is a variant of the Andersen lab "slow killing assay" [33] (**Supplemental Figure S1**), the variation being that the bacterial lawn was spread on the plates rather than spotted. Spreading

the lawn makes it less likely that worms can mitigate the harmful effects of the pathogen simply by avoiding it (although variation in behavior is obviously meaningful in its own right).

Preparation of slow-killing assay (SKA) plates - On day 5 of the assay, the bacterial species to be used in that block was thawed and spread onto LB plates and incubated at 37°C for two days. On day 7, a single bacterial colony was picked to inoculate ~28 ml of YT media, which was incubated in a shaking incubator at 100 RPM at 37°C for 24 hrs. After 24 hours (day 8), 5µl of bacterial culture was spread onto a 35 mm NGMA SKA plate containing 0.05% of 100mg/mL filter-sterilized fluorodeoxyuridine (FUDR) which prevents worm progeny from hatching by inhibiting DNA synthesis [34]. Assay plates were incubated in a closed plastic container at 37°C for 24 hours. On day 9, SKA plates were moved to a 25°C incubator.

Assay protocol - On day 1 of an assay block the MA lines and G0 ancestor were thawed onto 60mm agar plates (P0) seeded with OP50 and stored at 20°C. After 24 hours (day 2) a single L4 hermaphrodite (P1) was picked onto a new 60mm agar plate seeded with OP50 and incubated at the ultimate assay temperature of 25°C. Each MA line was replicated three-fold; the number of replicates from PS lines varied by block (**Supplemental Table S4**). On day 6, a single L4 hermaphrodite (F1) from each replicate was picked to a new plate and incubated at 25°C. On day 10, 30±3 L4 (sometimes L3) stage worms (F2) from each line were picked onto the SKA plates. Assay plates were assigned random numbers and subsequent handling was in random order. The number of living progeny on each plate was counted at 12 hour intervals for 120 hours.

Survival was determined via blue light exposure. Worms exhibit a phototactic response when exposed to blue light [35]. Non-motile worms were illuminated with a blue laser pointer for five seconds; worms that did not move after this time were categorized as dead. It is important to note that only worms that were observed moving were counted as alive, those that may have crawled off the plate, or buried themselves into the agar were not counted as alive.

Data Analysis

(i) Survival probability. Survival over the course of the assay was quantified by means of the chain binomial parameter, $\hat{p}$ [36]. For a single interval, the probability of survival from time *t* to time *t*+1 can be expressed as

$$P(N_{t+1} = n_{t+1} | N_t = n_t) = \binom{n_t}{n_{t+1}} p^{n_{t+1}} (1-p)^{n_{t+1}-n_t}, \qquad \text{Equn. 1}$$

where n_t and n_{t+1} denote the number of individuals alive at times *t* and *t+1*, respectively. The likelihood of survival over the entire course of the experiment (120 hrs at 12 hr intervals) is given by the equation

$$\mathcal{L}(p) = \prod_{i=1}^{T} \binom{n_{i-1}}{n_i} p^{n_i} (1-p)^{n_{i-1}-n_i}, \qquad \text{Equn. 2}$$

where *T* denotes the number of time steps. $\hat{p}$ is the value of *p* that maximizes the likelihood in equation 2, calculated by taking the natural logarithm of equation 2, differentiating with respect to *p*, and equating to zero.

We previously quantified survivorship over the course of a pathogen exposure assay with the statistic LT50 [33], but $\hat{p}$ is a more meaningful measure of survival than LT50 when a large fraction of individuals survive the assay, as is the case here for all bacteria except *PA14* (**Supplemental Figure S2**).

(ii) Mutational parameters. Three statistics were calculated to quantify the cumulative effects of mutation accumulation on pathogen susceptibility. First, we estimated the mutational bias (ΔM), the per-generation change in the trait mean due to mutation accumulation. ΔM is calculated as: $\Delta M = \frac{\bar{z}_{MA} - \bar{z}_0}{\bar{z}_0 t}$, where $\bar{z}_{MA}$ is the mean of the MA lines, $\bar{z}_0$ is the mean of the G0 progenitor, and *t* is the number of generations of MA (*t*=250) [37].

Second, we measured the mutational variance, V_M, which calculates the change in among line variance in the MA lines compared to the PS lines. V_M is calculated as:

$V_M = \frac{V_{L,MA} - V_{L,0}}{2t}$, where $V_{L,MA}$ is the among line variance of the MA lines, $V_{L,0}$ is the among line variance of the PS lines, and *t* is the number of generations of MA . Data were standardized by subtracting the block mean from each data point and dividing the difference by the block mean. The variance of the standardized data is the squared coefficient of variation, which provides a standardized measure of the evolvability of a trait [38]. We also standardized the data as a proportion of the residual (within-line) variance, V_E; the "mutational heritability" is calculated as $h_M^2 = \frac{V_M}{V_E}$.

Mutational covariances, Cov_M, and correlations, r_M, were calculated analogously to V_M, with the among-line components of variance replaced by the among-line components of covariance.

Genotypic variances and covariances of the wild isolates were calculated analogously to the mutational (co)variances. There is very little residual heterozygosity in the wild isolates [29], so wild isolates were treated as inbred lines, in which the genotypic variance, $V_G = \frac{V_L}{2}$, where V_L is the among-isolate component of variance [39]. Genetic covariances and correlations were calculated analogously. Broad-sense heritability, $H^2 = \frac{V_G}{V_P}$, where V_P is the total phenotypic variance.

Statistical Analysis

To begin, we considered each combination of *C. elegans* strain (N2 or PB306) and bacterial species independently. Means and variances were estimated by the general linear model (GLM) $y_{ijk}=\mu+a_i+L_{j|i}+e_{jk|i}$, where y_{ijk} is the (mean-standardized) value of $\hat{p}$ in a replicate, μ is the overall mean (near zero except for a small random deviation because of the mean-standardization), a_i is the fixed effect of treatment (MA or G0), $L_{j|i}$ is the random effect of line (MA or PS) within each treatment group, and $e_{jk|i}$ is the residual effect, assumed normally

distributed with mean 0 and variance independent between treatment groups (banded main diagonal covariance structure). Variances were estimated by restricted maximum likelihood (REML) with degrees of freedom determined by the method of Kenward and Roger [40], as implemented in the MIXED procedure of SAS v.9.4. The fixed effect of treatment was tested by F-test with Type III sums of squares.

Genetic covariance matrices (**M** and **G**, respectively) were estimated by REML, as follows. The diagonal elements of the covariance matrix are the variances in $\hat{p}_i$ for bacteria *i* and the off-diagonal elements are the covariances between $\hat{p}_i$ and $\hat{p}_j$ for bacteria *i* and *j*. (Co)variances can be partitioned into an among-line and within-line (residual) component. For each strain/treatment combination, the full GLM is $y_{ijk}=\mu+L_{j|i}+e_{jk|i}$, where y_{ijk} is the (mean-standardized) value of $\hat{p}$ in a replicate, μ is the overall mean (near zero except for a small random deviation because of the mean-standardization), $L_{j|i}$ is the random effect of line *j* (MA or PS) on bacteria *i*, and $e_{k|ij}$ is the residual effect. The $L_{j|i}$ are distributed (approximately) ~N(0,***G***), with (normalized) mean 0 and covariance structure ***G***, as described below. The $e_{k|ij}$ are assumed normally distributed ~N(0,***R***) with mean 0 and covariance structure ***R***, where the off-diagonal elements of ***R*** are constrained to equal 0 (banded main diagonal covariance), since the replicates assayed in the different pathogens are biologically independent samples.

For each strain/treatment (MA or G0) combination, we tested two (G0) or three (MA) hierarchical models of covariance structure, (1) the full model, with unstructured ***G***, (2) banded main diagonal ***G***, in which the off-diagonal elements are constrained to 0, and (3) ***G***=**0**, in which all elements of ***G*** are constrained to 0. The best model was decided based on the corrected AIC [41]. If the best model had more parameters than the next-best model, the models were compared by likelihood-ratio test (LRT); the models are nested, so twice the difference in the log(likelihood) is asymptotically chi-square distributed with degrees of freedom equal to the difference in the number of parameters in the two models. For the G0 PS lines, ***G*** is necessarily

banded main diagonal (if not **0**), because the PS lines are independent biological samples in the different pathogen assays.

Among-line variances and covariances of the wild isolates were estimated similarly to those of the MA and PS lines. Variances were estimated from $\boldsymbol{G_{WI}}$ with a banded main diagonal covariance structure and the model compared to a model with $\boldsymbol{G_{WI}}=\mathbf{0}$. Covariances were estimated jointly from the unconstrained model and compared to the model with $\boldsymbol{G_{WI}}$ constrained to banded main diagonal structure. Note the notational distinction between $\boldsymbol{G_i}$ (italicized, with a subscript) and **G** (no italics or subscript); the former represents the among-line variance-covariance matrix of group *i*, and the latter represents the genetic variance-covariance matrix.

The orientations of **M** and **G** in multivariate trait space were compared using the "genetic line of least resistance" approach, following Schluter [42]. We describe the method here; justification is provided in the Results. Data ($\hat{p}_i$) were mean-centered and divided by the block mean, averaged over all observations (MA and G0) in the block. This approach treats block as a fixed effect and puts all four variables on a common scale. The mutational and genetic (co)variance matrices (**M** and **G**, estimated as described above) were used as input data in a Principal Components Analysis (PCA), as implemented in the PRINCOMP procedure in SAS v. 9.4. The first principal component (PC1) is the axis of maximum genetic variance, and evolution will proceed most rapidly and easily along this axis, be it due to selection or drift [23, 43]. PC1 of **G** is commonly referred to as $\mathbf{g}_{max}$ [42]; we follow that convention and refer to PC1 of **M** as "$\mathbf{m}_{max}$". The deviation in orientation in trait space between $\mathbf{m}_{max}$ and $\mathbf{g}_{max}$ (or $\mathbf{m}_{max,1}$ and $\mathbf{m}_{max,2}$, where the subscripts refer to different sets of MA lines) is represented by the angle θ between the two vectors, where $\theta=\cos^{-1}(\mathbf{m}_{max}{}^{T}\mathbf{g}_{max})$ [42].

If $\mathbf{m}_{max}=\mathbf{g}_{max}$, θ=0. To test the hypothesis that $\mathbf{g}_{max} = \mathbf{m}_{max}$, we used a bootstrap resampling strategy [44]. We resampled MA lines with replacement, re-calculated the among-line (co)variance matrix, re-ran the PCA, and estimated $\mathbf{m}_{max}$ as before; we refer to the bootstrap estimate as $\mathbf{m}_{max}$**'**. We then calculated the angle θ' between the observed $\mathbf{m}_{max}$ and

the bootstrap estimate $\mathbf{m}_{max}$'. The fraction of bootstrap replicates (n=1000) in which θ' > θ is the approximate P-value of a test of the null hypothesis that $\mathbf{g}_{max}=\mathbf{m}_{max}$. Since this analysis takes $\mathbf{g}_{max}$ as estimated without error, it is formally a test of the question "How (un)likely is it that a mutational process with property $\mathbf{m}_{max}$ will produce **G** with property $\mathbf{g}_{max}$?". For a more conservative test, we can also sample wild isolates with replacement to quantify the variation in $\mathbf{g}_{max}$'.

Results

Survivorship of the N2 G0 progenitor on OP50 (food) sets the benchmark by which the effects of the other, pathogenic bacteria are assessed. Averaged over assay blocks, the N2 progenitor suffered approximately 11% mortality over the course of the 120 hr assay. Relative to that benchmark, survivorship was further reduced by as little as 1.5% in the PB306 G0 progenitor on OP50 and by as much as 84% in the N2 G0 progenitor on *PA14*. Averaged over strains and treatment groups (MA and G0), the relative pathogenicity of the three pathogenic bacteria is OP50<*E. faecalis*<*S. aureus*<<*PA14* (**Supplemental Figure S2; Supplemental Table S5**).

Mutation Accumulation (MA) lines

(i) Means. We expect that survival on OP50 (food) should be unambiguously beneficial, and both N2 and PB306 do exhibit a significant negative mutational bias on OP50 (**Table 1**; **Figure 3A**). Predictions about the direction of selection for survivorship on pathogenic bacteria are less clear, because fitness tradeoffs between disease resistance and traits such as growth rate and yield in the absence of pathogens are well-documented [45-48]. All else equal, however, living is better than dying.

Whether the deleterious effects of mutations are, on average, exacerbated under stressful conditions is a topic of active empirical investigation (e.g., [49-51]), but there is no theoretical reason to expect them to be [52]. Infection by a pathogen is *prima facie* a stressful

364 condition. If the average deleterious effect is increased under pathogen stress, we predict that
365 ΔM will be greater (more negative) in the pathogen treatments than in the OP50 treatment, and
366 moreover, that ΔM will be greatest (most negative) in the most virulent pathogen, *PA14*.

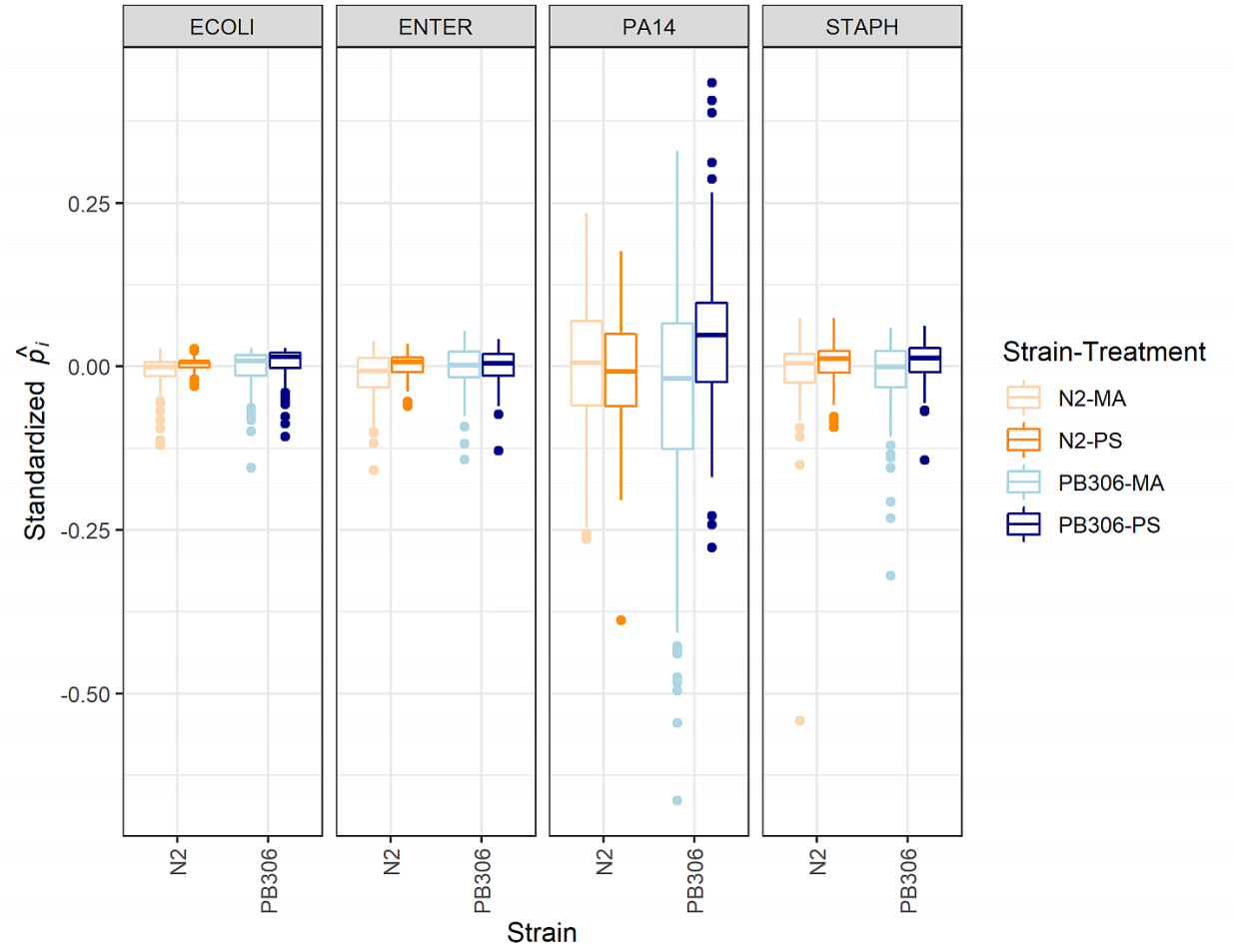


Figure 3. Box-plots of $\hat{p}$ standardized by the block mean (mean = 0). The four panels show the four different bacteria, from left to right (**A-D**): (**A**) *E. coli* OP50 (ECOLI), (**B**) *E, faecalis* (ENTER), (**C**) *P. aeruginosa* PA14 (PA14), and (**D**) *S. aureus* (STAPH). N2 is shown in orange, PB306 in blue; G0 PS lines shown in dark color, MA lines in light color. See Methods for details of $\hat{p}$ standardization.

375 **Table 1**. Means (SEM). See Methods for derivation of $\hat{p}$ and details of standardization, and for calculation of ΔM. Values of ΔM in

376 bold font are significantly different from 0 (P<0.05); see Methods for details of statistical tests.

Strain	Treatment		OP50	*E. faecalis*	PA14	*S. aureus*
N2	G0	$\hat{p}$ (raw)	0.987 (0.0054)	0.971 (0.0047)	0.744 (0.0254)	0.949 (0.0078)
		$\hat{p}$ (std)	0.00369 (0.00124)	0.00590 (0.00176)	-0.00764 (0.00869)	0.00487 (0.00359)
	MA	$\hat{p}$ (raw)	0.980 (0.0067)	0.958 (0.0075)	0.764 (0.0079)	0.941 (0.0029)
		$\hat{p}$ (std)	-0.00305 (0.00178)	-0.00714 (0.00388)	0.01089 (0.00869)	-0.00228 (0.00299)
		ΔM (std)	**-2.70E-5** (0.87E-5)	**-5.22E-5** (1.70E-5)	7.41E-5 (8.69E-5)	-2.86E-5 (1.87E-5)
PB306	G0	$\hat{p}$ (raw)	0.986 (0.0061)	0.953 (0.0100)	0.868 (0.010)	0.948 (0.0105)
		$\hat{p}$ (std)	0.00608 (0.00270)	0.00142 (0.00238573)	0.04476 (0.01065)	0.01498 (0.00321)
	MA	$\hat{p}$ (raw)	0.976 (0.0025)	0.950 (0.0026)	0.757 (0.0474)	0.928 (0.0086)
		$\hat{p}$ (std)	-0.00326 (0.00270)	-0.00117 (0.00289)	-0.03275 (0.01671)	-0.00362 (0.00421)
		ΔM (std)	**-3.74E-5** (1.59E-5)	-1.04E-5 (1.50E-5)	**-31.00E-5** (7.93E-5)	**-7.44E-5** (2.12E-5)
Ave		ΔM (std)	-3.12E-05	-3.13E-05	-19.20E-5	-5.15E-05

In general, support for those predictions is weak. ΔM is negative in five of the six strain/pathogen combinations, although not significantly different from zero in three of the six. In two of the six cases, ΔM is less negative in the pathogen than in the N2 progenitor on OP50, and in one of the four cases in which the point estimate of ΔM is more negative than the N2 progenitor on OP50 (N2 on *S. aureus*), the estimate is not significantly different from 0. The prediction that ΔM should be greatest in *PA14* is borne out for PB306, but not for N2; ΔM in N2 is positive (albeit not significantly different from zero). However, survival of the N2 progenitor on *PA14* was already so low (~96% mortality by 120 hours; **Supplemental Figure S2**) that presumably there was little opportunity for mutations to make things worse.

(ii) (Co)variances. To begin, we estimated the among-line components of variance jointly over the four bacteria for the MA lines in each strain, i.e. $\boldsymbol{G_{MA}}$ was modeled with a banded main diagonal covariance structure (see Methods) and tested for significance by LRT against a model with $\boldsymbol{G_{MA}}$=**0**. $\boldsymbol{G_{MA}}$ differed significantly from **0** in both strains (N2, Chi-square=15.0, df=6, p<0.002; PB306, Chi-square=49.8, df=6, p<<0.0001). The among-MA line (co)variance $\boldsymbol{G_{MA}}$ includes the cumulative effects of mutation, as well as the effects of heritable, non-genetic variation [30]. The sum of the variances (the trace) is the overall size of $\boldsymbol{G_{MA}}$. $\boldsymbol{G_{MA}}$ is >5X greater in PB306 than in N2 (LRT chi-square=18.1, df=4, P<0.01), although the difference is less than twofold when *PA14* is excluded (LRT chi-square=7.3, df=3, P<0.07; **Table 2**).

Table 2. Variances in standardized $\hat{p}$ (SEM). Abbreviations are: V_L (among-line variance); V_E (environmental variance); V_M (mutational variance); h_m^2 (mutational heritability); $r_{M,W}$ (correlation between line mean $\hat{p}$ and competitive fitness; V_G (genetic variance); H^2 (broad-sense heritability).

Strain		OP50	*E. faecalis*	PA14	*S. aureus*
N2	V_L (MA)	0.000056 (0.000033)	0.000365 (0.000163)	0.001796 (0.001163)	0.000086 (0.000106)
	V_L (PS)	4.54E-6 (6.91E-6)	0	0.000128 (0.000357)	0.000254 (0.000118)
	V_E (MA)	0.000236 (0.000036)	0.000624 (0.000127)	0.008323 (0.001308)	0.001034 (0.000135)
	V_E (PS)	0.000114 (0.000016)	0.000334 (0.000046)	0.006918 (0.000961)	0.000664 (0.000111)
	V_M (std)	1.03E-07	7.30E-07	3.34E-06	0
	h_m^2	5.89E-04	1.52E-03	4.38E-04	0
PB306	V_L (MA)	0.00028 (0.000093)	0.000172 (0.000086)	0.01083 (0.003134)	0.000396 (0.000219)
	V_L (PS)	0.000026 (0.00003)	0.000019 (0.000061)	0	0.000041 (0.000096)
	V_E (MA)	0.000538 (0.000075)	0.000631 (0.000091)	0.01663 (0.002199)	0.001479 (0.00023)
	V_E (PS)	0.000408 (0.000059)	0.000521 (0.000092)	0.0143 (0.001809)	0.000789 (0.000147)
	V_M (std)	5.08E-07	3.06E-07	2.17E-05	7.10E-07
	h_m^2	1.07E-03	5.31E-04	1.40E-03	6.26E-04

	$r_{M,w}$	-0.001 (0.238)	-0.075 (0.311)	-0.114 (0.205)	0.431 (0.247)
Ave	V_M (std)	3.06E-07	5.18E-07	1.25E-05	3.55E-07
Ave	h_m^2	8.31E-04	1.03E-03	9.19E-04	3.13E-04
Wild Isolates	V_G	0.000282 (.000081)	0.000038 (.000028)	0.000765 (.000217)	0.000238 (0.0000675)
	V_E	0.00036 (0.000064)	0.000455 (0.000074)	0.001163 (0.000196)	0.000291 (0.000048)
	H^2	0.305	0.072	0.284	0.31

398

399 $\boldsymbol{G_{PS}}$ includes only the contribution of non-genetic heritable variation. The same analysis
400 for the PS lines has less power than for the MA lines because there are fewer PS lines than MA
401 lines (PB306, n=6,9,39,39 for *PA14*, *OP50*, *E. faecalis*, and *S. aureus*, respectively; N2,
402 n=9,9,39,39 for the same pathogens). For PB306, $\boldsymbol{G_{PS}=0}$ is the best model (ΔAICc = -4.4), and
403 for all four bacteria the point estimate of the among-line variance of the MA lines is an order of
404 magnitude greater than that of the PS lines (**Table 2**). In N2, however, the banded main
405 diagonal model is significantly better than the alternative model with $\boldsymbol{G_{PS}=0}$ (ΔAICc = -2.0; LRT
406 Chi-square=8.2, df=6, p<0.05). That result is driven by the variance on *S. aureus*, which is ~3X
407 greater among PS lines than among MA lines (**Table 2**).
408 In principle, the mutational covariance matrix $\mathbf{M}=\boldsymbol{G_{MA}}-\boldsymbol{G_{PS}}$ [53, 54]. Ideally, $\boldsymbol{G_{PS}=0}$ (i.e.,
409 there is no heritable (co)variance among PS lines), in which case $\mathbf{M}=\boldsymbol{G_{MA}}$. However, since we
410 have no estimate of the off-diagonal elements of $\boldsymbol{G_{PS}}$ (see Methods), we are left with two
411 choices: either assume that $\boldsymbol{G_{PS}=0}$ and forge ahead, or stop. Clearly, it is unjustified to assume
412 that $\boldsymbol{G_{PS}=0}$ in N2, whereas there is statistical justification for assuming that $\boldsymbol{G_{PS}=0}$ in PB306.
413 What we can say with some confidence is, if there are non-genetic factors that contribute to
414 heritable (co)variance, their cumulative effects are about an order of magnitude less than those
415 of mutation except in the case of N2 on *S. aureus*. Accordingly, we will forge ahead and
416 assume that **M** is as inferred from $\boldsymbol{G_{MA}}$ in PB306, although the tenuousness of that assumption
417 is manifest.
418 One of our primary goals was to compare $\mathbf{M}_{N2}$ to $\mathbf{M}_{PB306}$ (i.e **M** estimated from each set of
419 MA lines) to better characterize variation in the mutational process in the context of the
420 multivariate phenotype. The fact that $\boldsymbol{G_{PS}\neq 0}$ for N2 negates that possibility. However, $\boldsymbol{G_{MA}}$
421 represents the combined contribution of mutational and transgenerational epigenetic effects to
422 the heritable variation among MA lines. We redefine $\boldsymbol{G_{MA}}$ as the (epi)mutation matrix **M***, to
423 explicitly acknowledge the potential contribution of epigenetic factors in addition to the
424 cumulative effects of mutation, and similarly redefine PC1 of **M*** to $\mathbf{m}^*_{max}$. The assay of wild

isolates was not designed to partition the heritable variance into genetic and non-genetic components, so we re-define **G** as **G***, the (epi)genetic variance-covariance matrix, and similarly redefine $\mathbf{g}_{max}$ as $\mathbf{g}^*_{max}$.

Turning first to V_M, mutational heritabilities ($h^2_M = \frac{V_M}{V_E}$) for all four pathogens are on the order of 0.5-1x10^{-3}/generation (**Table 2**), which is typical for a wide range of traits in a variety of organisms [55], including *C. elegans* [56]. Etienne et al. [33] previously estimated mutational parameters for survival on *PA14* of (almost) the same set of PB306 MA lines. The survival assay and measure of survivorship in the Etienne et al. study (spotted bacterial lawn, LT50) both differed from this study (spread lawn, $\hat{p}$), but the estimated h^2_M of LT50 was ~1x10^{-3}/generation, very close to the estimate from this study.

The (epi)mutational covariance/correlation matrices **M*** as inferred from $\boldsymbol{G_{MA}}$ are shown in **Table 3**. Two features of **M*** are evident: the correlations between survival on the actual pathogens (i.e., all except OP50) are large and positive (>0.5) and at least marginally significant, whereas the correlations between survival on OP50 and on the pathogens are much smaller.

The first principal component of **M*** ($\mathbf{m}^*_{max,PB306}$) accounts for ~99% of the variance among PB306 MA lines, and $\mathbf{m}^*_{max,N2}$ explains nearly as much in N2 (**Supplemental Table S6**). The appropriateness of drawing evolutionary inferences from comparisons of principal components of **G** has been criticized on the grounds that "...(evolutionarily) causal factors need not have orthogonal effects on the phenotype" [57], but when essentially all of the heritable variance falls along PC1, the "genetic lines of least resistance" approach seems reasonable, since there is hardly any unexplained variance about which the constraint of orthogonality could be misleading.

The difference in orientation of $\mathbf{m}^*_{max,PB306}$ and $\mathbf{m}^*_{max,N2}$ does not differ from the random expectation (θ=0.214, P>0.80). We therefore pooled samples over the two sets of MA lines and

450 **Table 3**. (Epi)mutational variance-covariance matrices **M*** (SEM). (Co)variances shown above/on the diagonal in dark shading;

451 correlations shown below the diagonal in light shading. See Methods for details of calculations.

Strain	Tr	Bacteria	OP50	*E. faecalis*	PA14	*S. aureus*
PB306	MA	OP50	2.82E-4 (0.94E-4)	0.21E-5 (0.65E-5)	-4.70E-4 (3.99E-4)	1.18E-4 (0.94E-4)
		E. faecalis	0.10 (0.30)	1.66E-4 (0.85E-4)	6.80E-4 (3.57E-4)	1.62E-4 (0.90E-4)
		PA14	-0.27 (0.22)	0.51 (0.24)	1.09E-2 (0.32E-2)	1.42E-3 (0.60E-3)
		S. aureus	0.34 (0.27)	0.62 (0.33)	0.67 (0.23)	4.15E-4 (0.22E-4)
N2	MA	OP50	5.60E-5 (3.30E-5)	-3.00E-5 (5.60E-5)	2.40E-5 (1.18E-4)	6.35E-6 (4.30E-5)
		E. faecalis	-0.19 (0.41)	3.41E-4 (1.61E-4)	8.80E-5 (3.28E-4)	1.68E-4 (0.95E-4)
		PA14	0.075 (0.37)	0.11 (0.42)	1.80E-3 (1.16E-3)	-1.20E-4 (2.69E-4)
		S. aureus	0.95 (0.65)	1 (-)	-0.31 (0.72)	7.90E-5 (1.06E-4)

452

453

re-calculated $\mathbf{m^*_{max}}$ for comparison to $\mathbf{g^*_{max}}$. The difference in orientation between $\mathbf{m^*_{max}}$ and $\mathbf{g^*_{max}}$ also does not differ from the random expectation (θ=0.409, P>0.64). In hindsight, the similarity of orientation in trait space between **M*** and **G*** was predictable, given that PA14 explains the largest component of the variance in both MA lines and wild isolates (**Table 2**).

(iii) Mutational correlations with relative competitive fitness. Under the MA conditions (20°C, NGM plates, fed on OP50), the PB306 MA lines decline in competitive fitness at about 0.1%/generation (ΔM_w= -1.13$\pm$0.25x10^{-3}/gen), very close to the estimate of ΔM_w for N2 in the same assay (-1.09/gen [32]). Mutational correlations (r_M) between $\hat{p}_i$ on pathogen *i* and relative competitive fitness are shown in **Table 2**. r_M between $\hat{p}$ on OP50 and competitive fitness in the MA environment is ~0, consistent with the 27X greater (more negative) rate of mutational decay of competitive fitness than of $\hat{p}$ on OP50 (**Table 1**). Although there is significant mutational variance for survivorship on OP50 at 25°C over the course of 120 hrs, the relevant mutations evidently have no consistent effect on competitive fitness under the MA conditions. Similarly, r_M between relative fitness and $\hat{p}_i$ on *E. faecalis* and *PA14* is also near 0; only with *S. aureus* is there a marginally significant mutational correlation (r_M=0.43, P<0.1).

Wild isolates

The standing (epi)genetic (co)variances (**G***) and correlations are given in **Table 4**. The full (unconstrained) model of **G*** provided the best fit (ΔAICc=3.5, LRT chi-square=16.2, df=6, P<0.02), with highly significant V_G (p<0.001) for all pathogens except *E. faecalis* (LRT chi-square=2.6, df=1, P>0.10). Broad-sense heritability, H^2, is approximately 30% for all bacteria except *E. faecalis*, for which H^2=7%. The lower H^2 on *E. faecalis* is a result of lower V_G rather than greater V_E, which is not much greater than on OP50 or *S. aureus*, and less than for *PA14* (**Table 2**).

Heuristic comparison of **G*** with **M*** reveals several noteworthy features. First, the rank orders of the mutational and standing (epi)genetic variances (V_M and V_G) are similar, with *PA14*

480 having the greatest variance and *E. faecalis* the least, with OP50 and *S. aureus* intermediate.
481 This finding is expected. V_M typically predicts V_G with high confidence for a variety of traits in
482
483 **Table 4.** Standing (epi)genetic variance-covariance matrix, **G*** (SEM). (Co)variances shown
484 above/on the diagonal in dark shading; correlations shown below the diagonal in light shading.
485 See Methods for details of calculations.

Bacteria	*OP50*	*E. faecalis*	*PA14*	*S. aureus*
OP50	5.64E-4 (1.62E-4)	1.04E-4 (0.67E-4)	3.78E-4 (1.87E-4)	-0.70E-4 (1.22E-4)
E. faecalis	0.51 (0.31)	0.76E-4 (0.56E-4)	3.28E-4 (1.20E-4)	0.70E-4 (0.62E-4)
PA14	0.41 (0.17)	0.96 (0.32)	1.53E-3 (0.43E-3)	-0.40E-4 (2.07E-4)
S. aureus	-0.13 (0.23)	0.37 (0.32)	-0.045 (0.24)	4.76E-4 (1.35E-4)

486
487 disparate organisms [58-61]. The rank order is not an artifact of mean-scaling; it is the same for
488 the unscaled $\hat{p}$. Second, r_M and r_G are themselves uncorrelated (r = -0.14). That result
489 indicates that, at minimum, pleiotropy is not unbreakably strong. However, the standing genetic
490 correlation between $\hat{p}$ on *E. faecalis* and $\hat{p}$ on *PA14* is nearly 1 (r_G=0.96), although an inference
491 based on a correlation between two variables, one of which the variance is not significantly
492 different from zero (*E. faecalis*), is not robust.
493
494 **Discussion**
495
496 *Consistency and idiosyncrasy*
497 The average number of base-substitution and small indel mutations per genome is nearly
498 identical between the N2 and PB306 MA lines [29], and, although the data are limited, it
499 appears that the (probably small) number of larger structural variants is similar between the two

strains as well (A. S. Saxena and Baer, unpublished results). Thus, differences between the two strains in the cumulative effects of mutations are evidently the result of differences in the distribution of mutational effects (DME), rather than differences in the total number of mutations.

One reassuring consistency is the congruence of the results of this study with the findings of a previous assay of (nearly) the same set of PB306 MA lines on *PA14* [33]. Given the ubiquity of block effects (in this study, and in general), it was a distinct possibility that the outcome of two experiments in the same set of lines, seven years apart and performed by different people, would be very different. Although the details of the assays and the measures of survivorship differ between the two studies, h_m^2 and V_M are both within 20%. The mutational bias, ΔM, is about 3X greater (more negative) in this study than in the previous one, but the orders of magnitude (~10^{-4}/gen) are similar, and substantially greater than for the other, less virulent bacteria (**Table 1**).

Given the relatively consistent behavior of PB306 from one assay to the next, the discrepancy between PB306 and N2 on *PA14* is all the more striking. V_M is more than sixfold greater in PB306 than in N2, and the discrepancy cannot be attributed to inflated among-line variance in the G0 PS lines in N2 (**Table 2**). Contrast that result with those on *S. aureus*, in which the among-MA line variance in PB306 is ~4X greater than in N2, but the among-PS line variance in N2 is almost as large as the variance among MA lines.

That the among-PS line variances ($\boldsymbol{G_{PS}}$) are significantly different from **0** in N2 but not in PB306 warrants further scrutiny. The significant difference is entirely due to *S. aureus*; for the other three bacteria, the variance among the MA lines is ~10X greater than among PS lines, just as for PB306. Further, the among-PS line variance in N2 is not a result of just a couple of outlying lines; the distribution of PS line means has a distinct left tail reminiscent of the typical distribution of MA lines (**Supplemental Figure S4**). Type I error notwithstanding, the proximate cause of heritable variation among genetically identical lines is transgenerational epigenetic inheritance [62-65] which can have various ultimate causes, including starvation [66, 67],

various types of abiotic stress [67-69], and exposure to pathogens [70, 71], as well as random, spontaneous epimutation [31]. Any of those are plausible causes of variation among N2 PS lines. Whether N2 is inherently more prone to transgenerational epigenetic variation (in the lab environment) than PB306, or wild isolates in general, is not known, although epigenetic modifications have an underlying genetic basis.

Does the evidently different DME in N2 and PB306 provide additional insight into evolution? It is tempting to think that the same (average) number of mutations with different cumulative effects in different backgrounds implies that mutations must interact epistatically. However, the observable (not observed, in this instance) DME represents a sample from some underlying true DME. In principle, one could model the (multivariate) DME for both genotypes and compare the best model with a single underlying DME to the best model with two DMEs, one for each genotype. We do not pursue that line of inquiry here.

Natural selection

The finding that $\Delta M<0$ in all but one case (N2 on *PA14*) implies that mutations that reduce survival upon exposure to pathogenic bacteria are deleterious, on average. That is not surprising in hindsight, although it is certainly plausible that fitness tradeoffs could result in stabilizing selection on pathogen susceptibility. The lack of correlation between survival on pathogens with competitive fitness in the MA environment indicates that pleiotropy between mutational effects in the two contexts is not strong, but reveals little about selection in the natural environment.

To get a rough idea of how strong selection on these traits is, the neutral expectation provides a benchmark. For a neutral trait in a predominantly selfing organism such as *C. elegans*, at mutation-drift equilibrium, $V_G \approx 4N_e V_M$ [11]. N_e in *C. elegans* has been estimated to be on the order of 10^5 [72], so if $V_M \approx 5\times10^{-5}$, as for $\hat{p}$ on OP50, we expect $V_G \approx 2$, about a thousand times greater than the observed value. The discrepancy grows to about ten

thousand-fold for $\hat{p}$ on *PA14*. Although there is considerable uncertainty associated with all of these estimates, clearly, the ability to survive exposure to pathogenic bacteria is not a neutral trait.

If selection is sufficiently strong to be approximately deterministic (i.e., ignoring drift), at mutation-selection balance, $V_G \approx \frac{V_M}{s}$, where *s* represents the average strength of selection against a mutation affecting the trait [15]; see [17] for caveats. By that calculation, *s*≈0.002 for mutations affecting $\hat{p}$ on OP50, and about an order of magnitude stronger (*s*≈0.03) on the virulent pathogen *PA14*. Selection on the two less-virulent pathogens is intermediate, *s*≈0.009 on *E. faecalis* and *s*≈0.004 on *S. aureus*. By way of comparison, a similar calculation for competitive fitness (measured in the MA conditions) revealed *s*≈0.005 [32], which was in close agreement with a direct estimate of the average mutational effect $\bar{s}$≈0.0035, calculated by dividing the cumulative decline in competitive fitness of the MA lines by the average number of mutations carried by an MA line. These point estimates are obviously rough, but it is evident that mutations that affect susceptibility to pathogens experience significant selection in nature. Whether the selection is direct, imposed by exposure to pathogens in nature, or indirect, due to pleiotropic effects on fitness in the absence of pathogens, cannot be ascertained from these experiments. However, the absence of a mutational correlation between $\hat{p}$ and relative fitness (in the MA context) implies that it might be the former. Recent studies have revealed a complex transgenerational epigenetically inherited (TEI) behavioral response to the PA14 strain of *P. aeruginosa* that varies among wild isolates [73], which strongly reinforces the inference that PA14 imposes significant direct selection in nature.

Mutation-selection balance (MSB) is not the only possible mechanism by which genetic variation may be maintained; balancing selection (BS) is another. There are several well-documented balanced polymorphisms in *C. elegans* (summarized in [74]), and there are numerous examples from diverse organisms in which pathogens do contribute to BS [75-78].

An important theoretical consideration is that, if BS is the predominant mechanism responsible for the maintenance of genetic variation, there is no reason to expect V_G and V_M to be correlated. In fact, to the extent that anyone has looked, V_M and V_G are invariably highly positively correlated, as they are in this study. BS and MSB are often set up as alternative hypotheses for the maintenance of genetic variation, but it is realistic to expect that uniformly deleterious alleles will always contribute to trait variation, whereas balanced polymorphisms may or may not. If that scenario is generally true, the effect of BS would be to weaken the relationship between V_M and V_G, and that does not seem to be the case. If, in contrast, BS is the exception rather than the rule, the occasional trait experiencing BS would appear as a high outlier, with more V_G than predicted by its V_M, and selection as inferred from the ratio of V_M to V_G would appear anomalously weak. The strength of selection for survival on OP50 (food) is similar to the strength of selection on competitive fitness, and selection on the three pathogens is stronger still. Thus, although BS may contribute to maintenance of genetic variation for susceptibility to bacterial pathogens in some (or even many) cases, our results provide no support for that hypothesis.

Conclusions

1. With one exception (N2 on *PA14*), the cumulative effect of mutations is to increase susceptibility to pathogens. Moreover, survival upon exposure to pathogens is uncorrelated with competitive fitness under the MA conditions. Thus, there is little support for the existence of tradeoffs between susceptibility to pathogens and fitness in the absence of pathogens.

2. To the extent that **M*** and **G*** are reliable estimators of **M** and **G**, respectively, it is evident that survival on these bacteria is subject to effective selection (because $\mathbf{G} << 2N_e\mathbf{M}$), but also that **M*** is a sufficient predictor of the orientation of **G*** in multivariate trait space (because $\boldsymbol{g}^*_{max}$ is not significantly different from $\boldsymbol{m}^*_{max}$).

602 3. The strength of selection against mutations underlying susceptibility to pathogens, as inferred
603 from the ratio V_M/V_G, is commensurate with selection against mutations affecting competitive
604 fitness. Our results provide no evidence that genetic variation for susceptibility to bacterial
605 pathogens is maintained by balancing selection.
606
607 **Acknowledgments**
608 Q. Allen, J. Dembek, D. Feistel, M. Puentes, A. Shoucair, and M. Snyder provided assistance in
609 the laboratory. J. M. Ponciano suggested $\hat{p}$ as the measure of survival. Support was provided
610 by NIH awards GM107227 to CFB, E.C. Andersen, and J. M. Ponciano, and GM127433 to CFB
611 and V. Katju.

References Cited

[1] Haldane, J.B.S. 1949 Disease and evolution. *La Ricerca Scientifica* **19**, 68-76.
[2] Hamilton, W.D. 1980 Sex versus non-sex versus parasite. *Oikos* **35**, 282-290. (doi:10.2307/3544435).
[3] Rockett, K.A., Clarke, G.M., Fitzpatrick, K., Hubbart, C., Jeffreys, A.E., Rowlands, K., Craik, R., Jallow, M., Conway, D.J., Bojang, K.A., et al. 2014 Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics* **46**, 1197-1204. (doi:10.1038/ng.3107).
[4] Barton, N.H. 1995 A general model for the evolution of recombination. *Genetical Research* **65**, 123-144.
[5] Lefebvre, A. & Palloix, A. 1996 Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: A case study, the interaction pepper - *Phytophthora capsici* Leonian. *Theoretical and Applied Genetics* **93**, 503-511.
[6] Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M. & Excoffier, L. 2013 Evidence for Polygenic Adaptation to Pathogens in the Human Genome. *Molecular Biology and Evolution* **30**, 1544-1558. (doi:10.1093/molbev/mst080).
[7] Hill, A.V.S. 2001 The genomics and genetics of human infectious disease susceptibility. *Annual Review of Genomics and Human Genetics* **2**, 373-400. (doi:10.1146/annurev.genom.2.1.373).
[8] Cogni, R., Cao, C., Day, J.P., Bridson, C. & Jiggins, F.M. 2016 The genetic architecture of resistance to virus infection in Drosophila. *Mol. Ecol.* **25**, 5228-5241. (doi:10.1111/mec.13769).
[9] Band, G., Le, Q.S., Clarke, G.M., Kivinen, K., Hubbart, C., Jeffreys, A.E., Rowlands, K., Leffler, E.M., Jallow, M., Conway, D.J., et al. 2019 Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat. Commun.* **10**. (doi:10.1038/s41467-019-13480-z).
[10] Kimura, M. & Ohta, T. 1971 Protein polymorphism as a phase of molecular evolution. *Nature* **229**, 467-&. (doi:10.1038/229467a0).
[11] Lynch, M. & Hill, W.G. 1986 Phenotypic evolution by neutral mutation. *Evolution* **40**, 915-935.
[12] Hudson, R.R., Kreitman, M. & Aguade, M. 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.
[13] Haldane, J.B.S. 1927 A mathematical theory of natural and artificial selection, Part V: Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838-844.
[14] Clayton, G. & Robertson, A. 1955 Mutation and quantitative variation. *American Naturalist* **89**, 151-158. (doi:10.1086/281874).
[15] Barton, N.H. 1990 Pleiotropic models of quantitative variation. *Genetics* **124**, 773-782.
[16] Kondrashov, A.S. & Turelli, M. 1992 Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* **132**, 603-618.
[17] Charlesworth, B. 2015 Causes of natural variation in fitness: Evidence from studies of Drosophila populations. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 1662-1669. (doi:10.1073/pnas.1423275112).
[18] Gould, S.J. & Lewontin, R.C. 1979 The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society Series B-Biological Sciences* **205**, 581-598. (doi:10.1098/rspb.1979.0086).
[19] Lande, R. 1980 The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**, 203-215.
[20] Phillips, P.C. & McGuigan, K.L. 2006 Evolution of genetic variance-covariance structure. In *Evolutionary Genetics: Concepts and Case Studies* (eds. C.W. Fox & J.B. Wolf), pp. 310-325. Oxford, UK, Oxford University Press.
[21] Kimura, M. 1962 On probability of fixation of mutant genes in a population. *Genetics* **47**, 713-719.
[22] Ewens, W.J. 1969 *Population Genetics*. London, Methuen and Company.

[23] Phillips, P.C., Whitlock, M.C. & Fowler, K. 2001 Inbreeding changes the shape of the genetic covariance matrix in Drosophila melanogaster. *Genetics* **158**, 1137-1145.
[24] Mukai, T. 1964 Genetic structure of natural populations of *Drosophila melanogaster*. 1. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1-19.
[25] Halligan, D.L. & Keightley, P.D. 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology Evolution and Systematics* **40**, 151-172. (doi:10.1146/annurev.ecolsys.39.110707.173437).
[26] Keightley, P.D. & Caballero, A. 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3823-3827.
[27] Baer, C.F., Shaw, F., Steding, C., Baumgartner, M., Hawkins, A., Houppert, A., Mason, N., Reed, M., Simonelic, K., Woodard, W., et al. 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5785-5790.
[28] Vassilieva, L.L. & Lynch, M. 1999 The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119-129.
[29] Rajaei, M., Saxena, A.S., Johnson, L.M., Snyder, M.C., Crombie, T.A., Tanny, R.E., Andersen, E.C., Joyner-Matos, J. & Baer, C.F. 2021 Mutability of mononucleotide repeats, not oxidative stress, explains the discrepancy between laboratory-accumulated mutations and the natural allele-frequency spectrum in *C. elegans*. *Genome Research*. (doi:10.1101/gr.275372.121).
[30] Johnson, L.M., Smith, O.J., Hahn, D.A. & Baer, C.F. 2020 Short-term heritable variation overwhelms 200 generations of mutational variance for metabolic traits in *Caenorhabditis elegans*. *Evolution* **74**, 2451-2464. (doi:https://doi.org/10.1111/evo.14104).
[31] Beltran, T., Shahrezaei, V., Katju, V. & Sarkies, P. 2020 Epimutations driven by small RNAs arise frequently but most have limited duration in *Caenorhabditis elegans*. *Nature Ecology & Evolution*. (doi:10.1038/s41559-020-01293-z).
[32] Yeh, S.-D., Saxena, A.S., Crombie, T.A., Feistel, D., Johnson, L.M., Lam, I., Lam, J., Saber, S. & Baer, C.F. 2018 The mutational decay of male-male and hermaphrodite-hermaphrodite competitive fitness in the androdioecious nematode *C. elegans*. *Heredity* **120**, 1-12. (doi:10.1038/s41437-017-0003-8).
[33] Etienne, V., Andersen, E.C., Ponciano, J.M., Blanton, D., Cadavid, A., Joyner-Matos, J., Matsuba, C., Tabman, B. & Baer, C.F. 2015 The red death meets the abdominal bristle: Polygenic mutation for susceptibility to a bacterial pathogen in *Caenorhabditis elegans*. *Evolution* **69**, 508-519. (doi:10.1111/evo.12585).
[34] Hosono, R. 1978 Sterilization and growth inhibition of Caenorhabditis elegans by 5-fluorodeoxyuridine. *Exp Gerontol* **13**, 369-374. (doi:10.1016/0531-5565(78)90047-5).
[35] Lee, K.H. & Aschner, M. 2016 A Simple Light Stimulation of Caenorhabditis elegans. *Curr Protoc Toxicol* **67**, 11.21.11-11.21.15. (doi:10.1002/0471140856.tx1121s67).
[36] Allen, L.J.S. 2010 *An Introduction to Stochastic Processes with Applications to Biology*. 2nd ed, Chapman and Hall / CRC.
[37] Lynch, M. & Walsh, B. 1998 *Genetic Analysis of Quantitative Traits*. Sunderland, MA, Sinauer.
[38] Hansen, T.F. & Houle, D. 2008 Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**, 1201-1219. (doi:10.1111/j.1420-9101.2008.01573.x).
[39] Falconer, D.S. 1989 *Quantitative Genetics*. Third ed. Essex, UK, Longman Scientific and Technical.
[40] Kenward, M.G. & Roger, J.H. 1997 Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983-997. (doi:10.2307/2533558).
[41] Hurvich, C.M. & Tsai, C.L. 1989 Regression and time-series model selection in small samples. *Biometrika* **76**, 297-307. (doi:10.2307/2336663).

[42] Schluter, D. 1996 Adaptive radiation along genetic lines of least resistance. *Evolution* **50**, 1766-1774. (doi:10.2307/2410734).
[43] Lande, R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402-416. (doi:10.2307/2407630).
[44] Baer, C.F. & Lynch, M. 2003 Correlated evolution of life-history with size at maturity in *Daphnia pulicaria*: patterns within and between populations. *Genetics Research* **81**, 123-132. (doi:10.1017/s0016672303006098).
[45] Biere, A. & Antonovics, J. 1996 Sex-specific costs of resistance to the fungal pathogen *Ustilago violacea* (*Microbotryum violaceum*) in *Silene alba*. *Evolution* **50**, 1098-1110. (doi:10.2307/2410650).
[46] Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M. & Bergelson, J. 2003 Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74-77. (doi:10.1038/nature01588).
[47] Duncan, A.B., Fellous, S. & Kaltz, O. 2011 Reverse evolution: selection against costly resistance in disease-free microcosm populations of *Paramecium caudatum*. *Evolution* **65**, 3462-3474. (doi:10.1111/j.1558-5646.2011.01388.x).
[48] Kraaijeveld, A.R. & Godfray, H.C.J. 2008 Selection for resistance to a fungal pathogen in *Drosophila melanogaster*. *Heredity* **100**, 400-406. (doi:10.1038/sj.hdy.6801092).
[49] Katju, V., Packard, L.B. & Keightley, P.D. 2018 Fitness decline under osmotic stress in *Caenorhabditis elegans* populations subjected to spontaneous mutation accumulation at varying population sizes. *Evolution* **72**, 1000-1008. (doi:10.1111/evo.13463).
[50] Matsuba, C., Ostrow, D.G., Salomon, M.P., Tolani, A. & Baer, C.F. 2013 Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biology Letters* **9**. (doi:10.1098/rsbl.2012.0334).
[51] Yun, L. & Agrawal, A.F. 2014 Variation in the strength of inbreeding depression across environments: Effects of stress and density dependence. *Evolution* **68**, 3599-3606. (doi:10.1111/evo.12527).
[52] Agrawal, A.F. & Whitlock, M.C. 2010 Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends in Ecology & Evolution* **25**, 450-458. (doi:10.1016/j.tree.2010.05.003).
[53] Lynch, M. 1985 Spontaneous mutations for life-history characters in an obligate parthenogen. *Evolution* **39**, 804-818. (doi:10.1111/j.1558-5646.1985.tb00422.x).
[54] Lynch, M. & Walsh, B. 1998 *Genetics and Analysis of Quantitative Traits*. Sunderland, MA., Sinauer; xvi, 980 p.
[55] Houle, D., Morikawa, B. & Lynch, M. 1996 Comparing mutational variabilities. *Genetics* **143**, 1467-1483.
[56] Davies, S.K., Leroi, A.M., Burt, A., Bundy, J. & Baer, C.F. 2016 The mutational structure of metabolism in *Caenorhabditis elegans*. *Evolution* **70**, 2239-2246. (doi:10.1111/evo.13020).
[57] Houle, D., Mezey, J. & Galpern, P. 2002 Interpretation of the results of common principal components analyses. *Evolution* **56**, 433-440.
[58] Houle, D. 1998 How should we explain variation in the genetic variance of traits? *Genetica* **103**, 241-253.
[59] Houle, D., Bolstad, G.H., van der Linde, K. & Hansen, T.F. 2017 Mutation predicts 40 million years of fly wing evolution. *Nature* **548**, 447-452. (doi:10.1038/nature23473).
[60] Farhadifar, R., Ponciano, J.M., Andersen, E.C., Needleman, D.J. & Baer, C.F. 2016 Mutation is a sufficient and robust predictor of genetic variation for mitotic spindle traits in *Caenorhabditis elegans*. *Genetics*. (doi:10.1534/genetics.115.185736).
[61] Lynch, M., Latta, L., Hicks, J. & Giorgianni, M. 1998 Mutation, selection, and the maintenance of life-history variation in a natural population. *Evolution* **52**, 727-733.

[62] Rechavi, O. & Lev, I. 2017 Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans. *Current Biology* **27**, R720-R730. (doi:10.1016/j.cub.2017.05.043).
[63] Jablonka, E. & Raz, G. 2009 Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology* **84**, 131-176. (doi:10.1086/598822).
[64] Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T. & Lehner, B. 2017 Transgenerational transmission of environmental information in *C. elegans*. *Science* **356**, 316-319. (doi:10.1126/science.aah6412).
[65] Baugh, L.R. & Day, T. 2020 Nongenetic inheritance and multigenerational plasticity in the nematode *C. elegans*. *eLife* **9**. (doi:10.7554/eLife.58498).
[66] Rechavi, O., Houri-Ze'evi, L., Anava, S., Goh, W.S.S., Kerk, S.Y., Hannon, G.J. & Hobert, O. 2014 Starvation-Induced Transgenerational Inheritance of Small RNAs in *C. elegans*. *Cell* **158**, 277-287. (doi:10.1016/j.cell.2014.06.020).
[67] Jobson, M.A., Jordan, J.M., Sandrof, M.A., Hibshman, J.D., Lennox, A.L. & Baugh, L.R. 2015 Transgenerational effects of early life starvation on growth, reproduction, and stress resistance in *Caenorhabditis elegans*. *Genetics* **201**, 201-+. (doi:10.1534/genetics.115.178699).
[68] Chow, Y.L. & Sato, F. 2019 Transgenerational lipid-reducing activity of benzylisoquinoline alkaloids in *Caenorhabditis elegans*. *Genes to Cells* **24**, 70-81. (doi:10.1111/gtc.12657).
[69] Kishimoto, S., Uno, M., Okabe, E., Nono, M. & Nishida, E. 2017 Environmental stresses induce transgenerationally inheritable survival advantages via germline-to-soma communication in *Caenorhabditis elegans*. *Nat. Commun.* **8**. (doi:10.1038/ncomms14031).
[70] Rechavi, O., Minevich, G. & Hobert, O. 2011 Transgenerational Inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248-1256. (doi:10.1016/j.cell.2011.10.042).
[71] Belicard, T., Jareosettasin, P. & Sarkies, P. 2018 The piRNA pathway responds to environmental signals to establish intergenerational adaptation to stress. *BMC Biol.* **16**. (doi:10.1186/s12915-018-0571-y).
[72] Gilbert, K.J., Zdraljevic, S., Cook, D.E., Cutter, A.D., Andersen, E.C. & Baer, C.F. 2021 The distribution of mutational effects on fitness in *Caenorhabditis elegans* inferred from standing genetic variation. *bioRxiv*, 2020.2010.2026.355446. (doi:10.1101/2020.10.26.355446).
[73] Moore, R.S., Kaletsky, R., Lesnik, C., Cota, V., Blackman, E., Parsons, L.R., Gitai, Z. & Murphy, C.T. 2021 The role of the Cer1 transposon in horizontal transfer of transgenerational memory. *Cell* **3**, 00881-00883.
[74] Lee, D., Zdraljevic, S., Stevens, L., Wang, Y., Tanny, R.E., Crombie, T.A., Cook, D.E., Webster, A.K., Chirakar, R., Baugh, L.R., et al. 2021 Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nature Ecology & Evolution*. (doi:10.1038/s41559-021-01435-x).
[75] Hawley, D.M. & Fleischer, R.C. 2012 Contrasting epidemic histories reveal pathogen-mediated balancing selection on Class II MHC diversity in a wild songbird. *PLoS One* **7**. (doi:10.1371/journal.pone.0030222).
[76] Horger, A.C., Ilyas, M., Stephan, W., Tellier, A., van der Hoorn, R.A.L. & Rose, L.E. 2012 Balancing selection at the Tomato RCR3 guardee gene family maintains variation in strength of pathogen defense. *Plos Genetics* **8**. (doi:10.1371/journal.pgen.1002813).
[77] Fumagalli, M., Cagliani, R., Pozzoli, U., Riva, S., Comi, G.P., Menozzi, G., Bresolin, N. & Sironi, M. 2009 Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Research* **19**, 199-212. (doi:10.1101/gr.082768.108).
[78] Quéméré, E., Hessenauer, P., Galan, M., Fernandez, M., Merlet, J., Chaval, Y., Morellet, N., Verheyden, H., Gilot-Fromont, E. & Charbonnel, N. 2021 Pathogen-mediated selection favours the maintenance of innate immunity gene polymorphism in a widespread wild ungulate. *Journal of Evolutionary Biology* **34**, 1156-1166. (doi:https://doi.org/10.1111/jeb.13876).

800

801