Canalisation and plasticity on the developmental manifold of *Caenorhabditis elegans*

David J Jordan[1,2,✉] and Eric A Miska[1,2]

[1]Wellcome/CRUK Gurdon Institute, University of Cambridge, UK
[2]Department of Biochemistry, University of Cambridge, UK

How do the same mechanisms that faithfully regenerate complex developmental programs in spite of environmental and genetic perturbations also permit responsiveness to environmental signals, adaptation, and genetic evolution? Using the nematode *Caenorhabditis elegans* as a model, we explore the phenotypic space of growth and development in various genetic and environmental contexts. Our data are growth curves and developmental parameters obtained by automated microscopy. Using these, we show that among the traits that make up the developmental space, correlations within a particular context are predictive of correlations among different contexts. Further we find that the developmental variability of this animal can be captured on a relatively low dimensional *phenoptypic manifold* and that on this manifold, genetic and environmental contributions to plasticity can be deconvolved independently. Our perspective offers a new way of understanding the relationship between robustness and flexibility in complex systems, suggesting that projection and concentration of dimensional can naturally align these forces as complementary rather than competing.

✉ Correspondence: dj333 (at) cam.ac.uk

Introduction

Biological systems are remarkable for their ability to generate reproducible macroscopic dynamics from the complex interactions of large numbers of microscopic components. For example, in animals, the development of an entire organism from a single cell proceeds faithfully each generation even in the presence of environmental fluctuations and molecular noise. Such robustness arises at many spatial and temporal scales e.g., gene expression patterns give rise to reproducible cell differentiation [1], neural and muscular activity generate locomotion [2], and interactions between individuals of different species give rise to surprisingly reproducible ecological dynamics [3]. This robustness is called *canalisation* [4] and dynamics that are canalized are said to be *homeorhetic* [5].

While robust, canalised processes nevertheless allow for important variability; stem cell populations generate diverse tissue types, behaviors respond to stimuli and environmental cues, and populations adapt to changing environments. The structure of this macroscopic variability, however, is much more constrained than the variability intrinsic to the microscopic processes that underlie it. Although gene expression determines the dynamics of the cell cycle, variations in these dynamics are largely insensitive to stochastic fluctuations in the levels of the thousands of proteins. Thus, the ways in which these macroscopic phenotypes can vary is relatively "low-dimensional" compared to the ways in which the individual components can vary.

This projection of variations into a lower dimensional space provides a way for biological systems to be both robust and flexible. Robustness arises because most variations manifest as excitations onto relatively few phenotypic modes, and flexibility is permitted along these modes. As an example, consider how facial diversity can be generated by the combination of relatively few *eigenfaces* [6]. Here the eigenfaces are the modes in this system, and varying the weights of each mode can generate many

diverse faces. However, even large variations in these weights are unlikely to produce "non-face" images. Low dimensional representations of phenotypic diversity are ubiquitous in living systems. Variations in phenotypes that tend to a stable state during development have been successfully represented in low-dimensional phenotypic spaces called "morpho-spaces", for example, morphological traits like the shapes of finch beaks [7], of coiled sea-shells [8], and even the shape of the influenza antigen [9]. Additionally, while the dimensionality of dynamic, time-varying, and responsive phenotypes is more difficult to define rigorously [10], it has been shown in some cases to be low-dimensional. Some recent examples include the crawling behavior of *C. elegans* [11] and the neuronal dynamics that underlie it [12], the swimming behavior of both eukaryotic (*Tetrahymena thermophila*) [13] and prokaryotic (*Escherichia coli*) single celled organisms [14], and the transcriptional trajectories of cells during fate determination [15]. Concentration of dimension may be an intrinsic property of systems where robustness and control of the macroscopic observables are required [16] and may arise from constraints imposed by steady state growth [17, 18], or trade-offs between phenotypic archetypes suited for different tasks [19].

Here we propose that the process of canalisation may be superseded by a more general process of phenotypic concentration of dimension. That is, genetic networks evolve such that most variations will be projected onto a low dimensional manifold, and that this in turn provides both canalisation and phenotypic plasticity. Concentration of dimensions manifests as the co-variation of traits in a high dimensional multi phenotype space. This structure of the co-variation is called *phenotypic integration* [20] and its geometric properties can be captured by a phenotypic manifold, i.e., a lower dimensional space that faithfully captures most of the variation observed in a higher dimensional space. The process of finding such a manifold is called dimensionality reduction. The determination of this space and how it is measured can be essential to finding the dominant modes of phenotypic variation, and thus for assessing the degree of canalization. A recent example comes from the developmental program of the wing of the fruit-fly *Drosophila melanogaster* [21], which used landmark free morphometrics to uncover a dominant mode of variation that was not apparent from traditional landmark based techniques.

In this work, we use a custom built automated imaging system to map the phenotypic space of growth and development of *C. elegans*. Using this system, we recorded 673 individual growth curves during the $\approx$ 70 hours of their development from eggs to reproductive adults, and manually recording the timing of egg hatching and reproductive maturation in single animals. To construct as complete a manifold as possible, we sample from extant variations in different genetic and environmental contexts, including both natural genetic variation and single gene mutants. While the most unbiased approach to sampling genetic variation would be random mutagenesis, because of the immense sampling space, efficient sampling cannot be done. For this reason, we chose to sample "wild isolates" of *C. elegans*, i.e. strains collected from nature, and whose collections of genetic changes have been subjected to natural selection. For environmental diversity, we chose to alter the animal's diet; *C. elegans* feed on bacteria, and we chose a collection of bacteria that included the standard laboratory bacteria *E. coli* as well as bacteria isolated from natural sites where *C. elegans* were also collected [22]. Measurements were collected for three *C. elegans* wild-isolates, each fed on one of four different bacterial diets, and two mutants of the *C. elegans* laboratory strain N2.

Using these data, we demonstrate that the space of developmental trajectories can be captured by a low-dimensional manifold and show a correspondence between the directions of fluctuations in a fixed context the directions in which populations will shift due to genetic or environmental changes. We find that the manifold obtained using nonlinear dimensionality reduction techniques captures developmental variations in a way that allows one to neatly decompose the contributions of genetics and environment independently, with the major mode of variation corresponding to environmental shifts and the second mode corresponding to genetic changes.

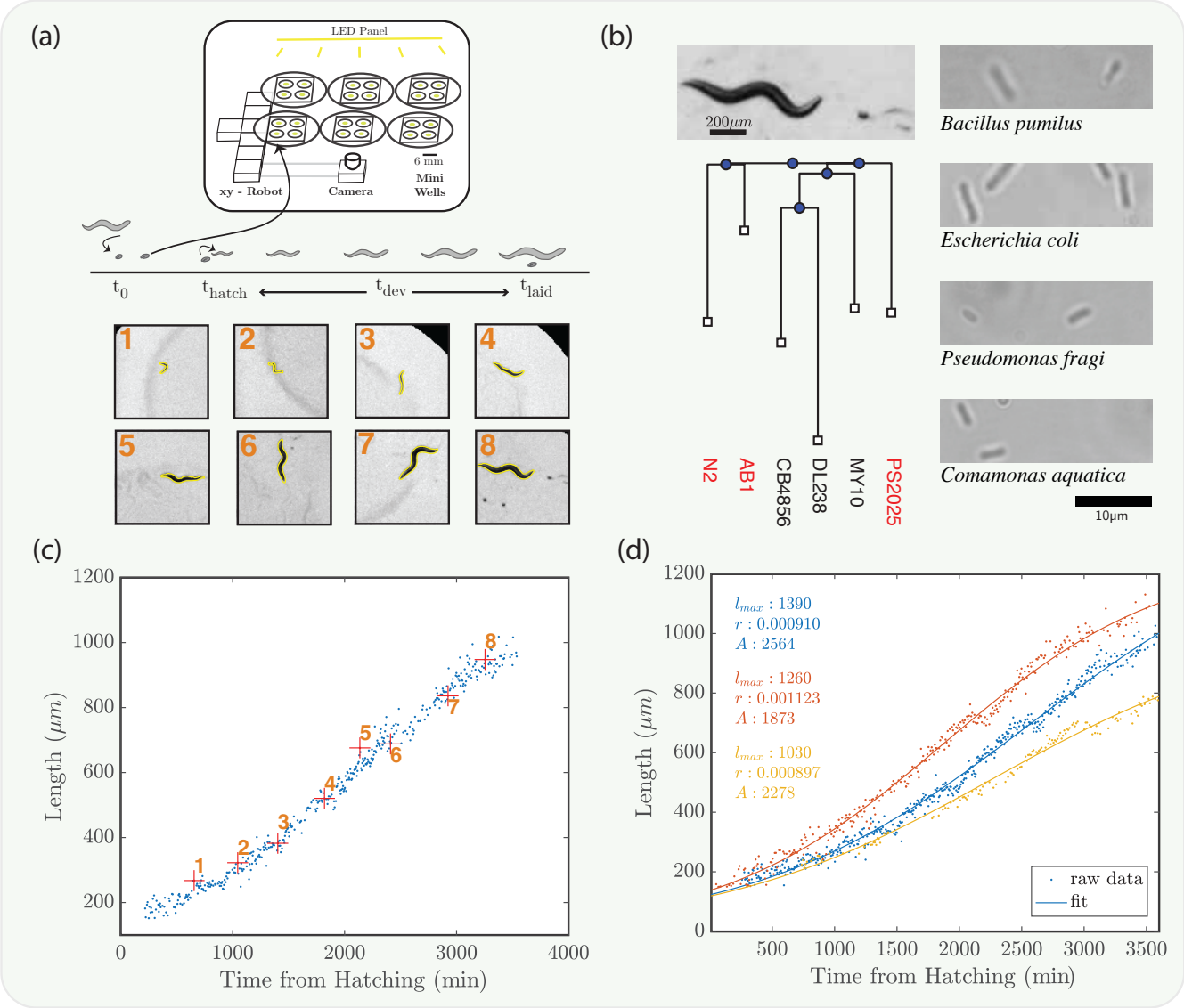


Figure 1: **(a)** A schematic of the imaging apparatus. Synchronized eggs are transferred into the mini wells at t_0. Hatching t_{hatch} and egg-laying time t_{laid} are manually recorded and give development time t_{dev}. Example images are shown for 10 time points (a, lower), with the computed outline of the worm shown (yellow). **(b)** An example image an adult *C. elegans* and an egg (scale bar 200μm). A phylogenetic tree pruned from CeNDR [23] database of some common wild isolates as well as the three strains used in this work, marked in red. In addition, micro-graphs of the bacteria used as food sources imaged at 100x magnification (scale bar 10μm). **(c)** The full developmental time course of an animal from hatching through adulthood. The blue points show calculated length over time, with 10 specific points highlighted (red plus) corresponding to each of the images in (a, lower). **(d)** shows the length measurements and the computed logistic best fit curves from three example time courses. The parameters of each logistic fit are shown to the left, with each color corresponding to the relevant data and curve.

Results

Characteristics that change over time are more difficult to quantify and compare than those that are static or near a steady state. Often, dynamic phenotypes such as these are measured and compared at a single fixed time, but in this case, proper alignment or synchronization can be challenging. Ideally, one would like to compare the time series of time-varying phenotypes and compare these directly. The growth of a multi-cellular organism is a good example of such a time-dependent phenotype. Ideally, one would like to compare growth curves aligned to an unambiguous starting time.

To this end, we designed and constructed a low-cost parallel imaging platform capable of measuring *C. elegans* growth for 60 individual animals simultaneously over the course of their ≈ 70 hour development at a temporal resolution of 0.001 Hz, resulting in a time series of ≈ 200 observations per animal. In addition to length and area measured automatically, egg hatching, and first egg-laying by mature adults are manually recorded. These not only provide estimates of the animals reproductive development, but also provide a fixed time to which the time series can be aligned. Reproductive development is in general correlated with growth but need not necessarily be. Because animals grow from approximately 0.2 mm to 1 mm in length during their lifetime, any wide-field system capable of imaging many isolated worms simultaneously would lack the necessary resolution. To solve this, we developed a system that uses a fixed lens and USB camera that is programmed to move between 6mm diameter custom made wells using an XY plotting robot. Sample images from the time series are shown (Fig. 1a) (inset, 1-8), along with the associated time series and the best fit logistic function of the form:

$$l(t) = \frac{l_{max}}{1 + e^{-r(t-A)}} \tag{1}$$

as determined by non-linear least squares fitting, giving three parameters for each curve (Fig. 1d). We used this system to record development in a collection of *C. elegans* isolated from the wild and fed on various bacteria, some of which were collected from sites where *C. elegans* were found. In total, we assayed five unique *C. elegans*genotypes fed on four different bacterial diets (Fig. 1b) and collected a total of 673 growth curves, in addition to developmental data, across these conditions. The developmental data consist of the duration of *ex-utero* embryonic development and the duration of reproductive maturation. These are measured as the time from egg-laying to egg-hatching t_{hatch}, and the time from hatching until the animal grows to reproductive age and lays its first egg $t_{dev} = (t_{laid} - t_{hatch})$ (Fig. 1a). While these durations were single scalar quantities, the growth curves consisted of hundreds of points. Dimensionality reduction could have been performed directly on these high-dimensional vectors after appropriate regularization, but instead these growth curves were fit with the logistic function (Eq. 1), which performed similarly well and whose parameters are easier to interpret (See Supplemental Information). The three parameters, which we call the maximum length l_{max}, the growth rate r, and the shift A, determine the saturation value of the growth curve, the growth rate at the inflection point, and the temporal shift of the inflection point, respectively.

For each growth curve we have an independent measurement of the reproductive development time t_{dev} and of the animals length at this time $l(t_{laid}) = l_{dev}$ (Fig. 2a, squares). We can use these additional parameters to rescale each growth curve. First, we divide the length as a function of time by l_{dev}, yielding:

$$\hat{l}(t) = \frac{l_{max}/l_{dev}}{1 + e^{-r(t-A)}}$$

Then, rescaling time $\hat{t} \to t/t_{dev}$

$$\hat{l}(\hat{t}) = \frac{l_{max}/l_{dev}}{1 + e^{-rt_{dev}(\hat{t}-A/t_{dev})}} \tag{2}$$

Thus, the normalization simply rescales the fit parameters from $[l_{max}, r, A] \to [l_{max}/l_{dev}, r{\cdot}t_{dev}, A/t_{dev}]$ yielding unit-less quantities for the fit parameters. The fitting procedure on the normalized curves recovers the rescaled fitting parameters from the unnormalized data (See Supplementary Information). Interestingly, rescaling the curves to an independently measured parameter, the reproductive developmental duration, seems to collapse the growth curves, e.g. (Fig. 2b). To compare the variance

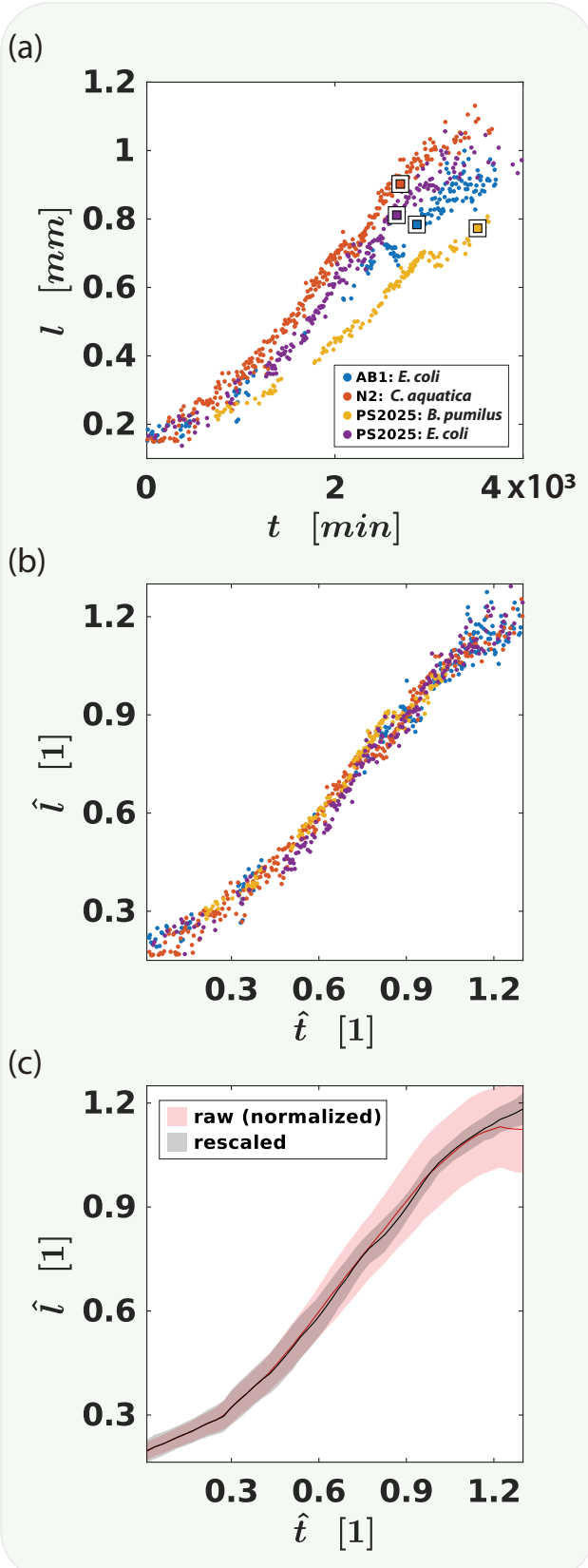
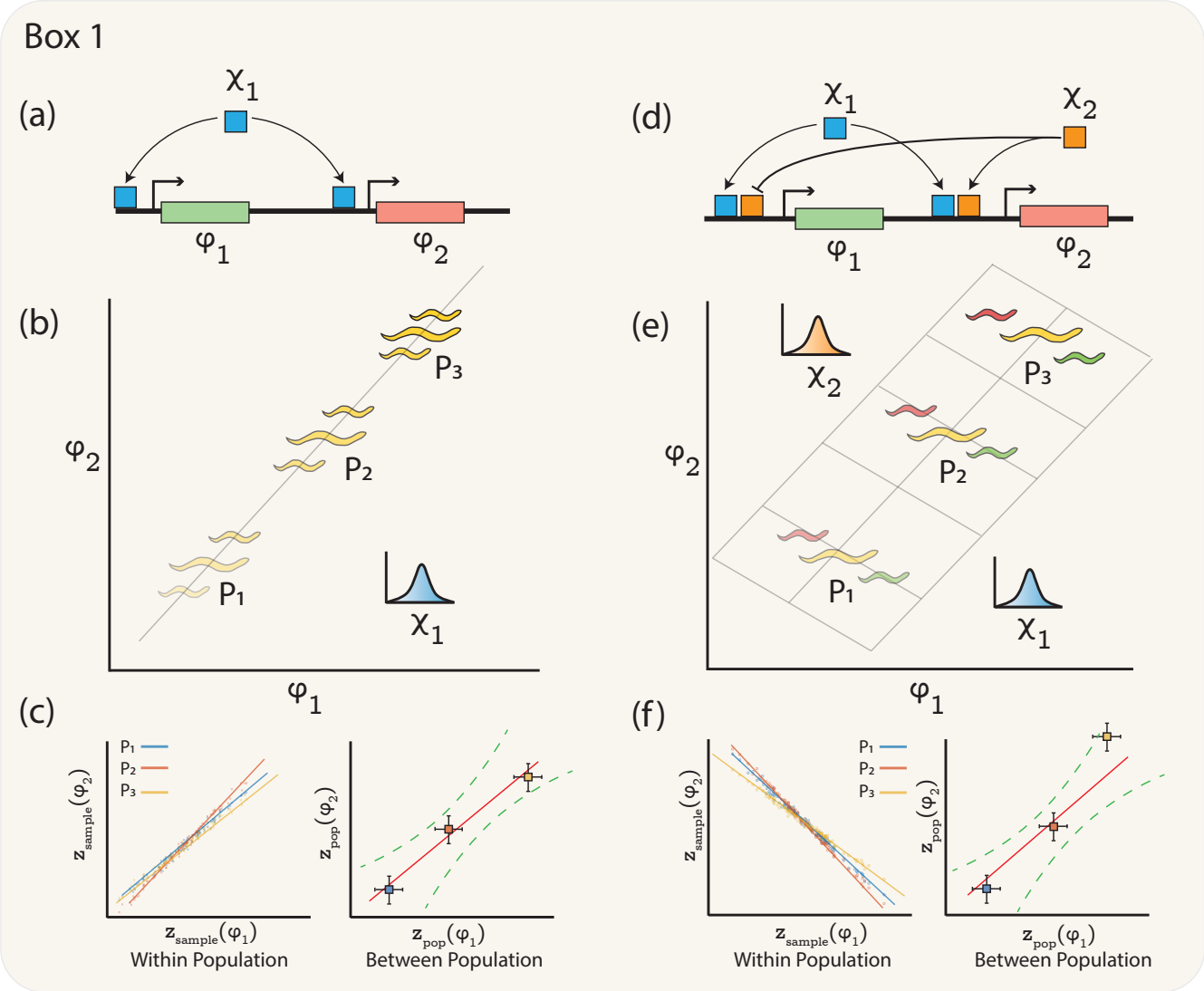


Figure 2: **(a)** Developmental time course of selected individuals where the square indicates t_{dev} and l_{dev} for each. The data can be re-scaled to plot length l as a fraction of development length $\hat{l} = l/l_{dev}$ and time t as a fraction of development time $\hat{t} = t/t_{dev}$ as shown in **(b)**. Re-scaling the data in this way changes the fit parameters of each logistic function. $[l_{max}, r, A] \rightarrow [l_{max}/l_{dev}, r \cdot t_{dev}, A/t_{dev}]$. The re-scaling of the curves in **(b)** appears to collapse the growth curves. This seems to be true in general, as the growth curves re-scaled by their individual t_{dev} and l_{dev} show a smaller standard deviation (**(c)** black) than the raw data (**(c)** red) (which is normalized to the average t_{dev} and l_{dev} to facilitate plotting on the same scale).

of all of the rescaled growth curves to the variance of the un-rescaled growth curves, the raw data was normalized to the mean developmental duration and the mean and standard deviation of the resulting growth curves were compared, showing that the rescaled growth curves do "collapse" and that their resulting standard deviation is smaller than for the raw data normalized (Fig. 2c).

The growth curves of *C. elegans* of different genetic backgrounds and grown on differ food sources differ in their maximum length, their growth rate, and in the shift, as well in the durations of *ex-utero* development and reproductive development. However, we find that these parameters do not vary independently. For example, some recent work has shown a negative correlation between growth rate and developmental duration in *C. elegans* and we also observe this in our data. They suggest that this may be a mechanism for reducing variability in adult size in [24]. If correlations between traits arise from a mechanism for control such as this, we would expect variation in individuals to be correlated in the same way as variations between populations. Strikingly, we find that if parameters are correlated among individuals in one context, these correlations are more likely to appear also between populations in different contexts (To see how such correlations may arise, refer to Box 1). For example, there is a strong negative correlation between the maximum length and the growth rate among individuals in all combinations of conditions. Likewise, the mean values of these parameters among popula-



Box 1: Illustration of relationships between genetic architecture, dimensionality, and the correlation structure of traits. **(a)** A simple genetic circuit where green ϕ_1 and red ϕ_2 phenotypes are controlled by a single transcription factor χ_1. **(b)** Different populations (P1-P3) have different mean expression of χ_1 due to either genetic or environmental changes, changing the brightness of the yellow color. Fluctuations in χ_1 around these mean values lead to correlated noise within populations **(c, left)**. The mean values of ϕ_1 and ϕ_2 are also correlated between populations **(c, right)** , in the same way. This leads to a one dimensional phenotypic manifold **(b, grey line)**, which is consistent with the single knob χ_1 In **(d)**, however, there are two independent transcription factors (χ_1, χ_2) that control ϕ_1 and ϕ_2. In each population P1-P3 in **(e)**, χ_1 sets the mean value of both ϕ_1 and ϕ_2 in a correlated manner, changing the overall brightness. However, fluctuations in χ_2 introduce anti-correlated noise within populations **(f, left)**, shifting the spectrum to either green or red. The between population correlation is dominated by the changes in χ_1, resulting in correlated changes in both ϕ_1 and ϕ_2 **(f, right)**. The within population anti-correlation expands the dimensionality of the manifold **(e, grey plane)**. Dimensionality can be expanded also by uncorrelated noise both within and between populations (not illustrated).

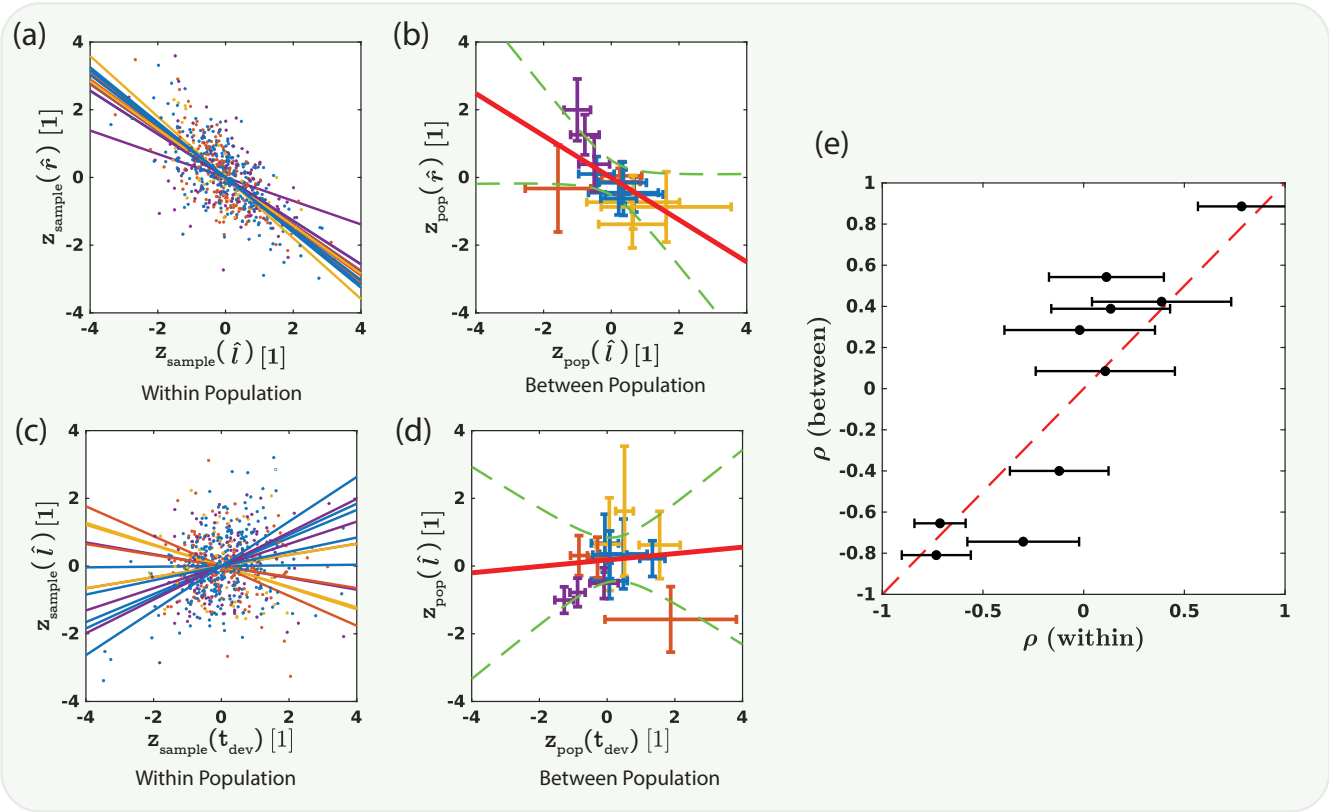


Figure 3: **(a)** An example showing a pair of developmental parameters $\hat{l}$ and $\hat{r}$ that are correlated within each populations measured. The scatter plot shows z-score of these parameters with respect to each conditions mean and variance, along with the best fit linear regression for that condition (colors indicate the food source). There is a significant within population negative correlation of these parameters in each condition. **(b)** The z-score is then calculated with respect to the mean and variance for all conditions taken together. These are then grouped by condition and the mean and standard deviation of the z-scores of both parameters are plotted (colors indicate the food source). The means are also correlated between conditions, and indicated by confidence interval (b, green dashed lines) of the slope of the linear regression (b, red line). For other traits, there is no significant within population correlation between traits, e.g. between the development time t_{dev} and $\hat{l}$ **(c)**. Similarly, in **(d)**, the between population means are also not significantly correlated . **(e)** Summarizes the within and between population correlation coefficient for all pairs or developmental parameters, the mean correlation coefficient for each of the within condition groups is plotted with error bars showing the standard deviation. For the correlation among the mean values, there is only a single value thus there are no vertical error bars. The red dashed line indicates equivalence, showing that the within population correlations are largely but not exclusively, predictive of the between population correlations.

tions grown in different conditions is also negatively correlated (Fig. 3 a,b). In contrast, the length and the duration of reproductive development are not strongly correlated among individuals, and neither are they correlated between populations (Fig. 3c,d). Computing the correlation coefficient (ρ) among individuals in a sample in fixed conditions, and plotting it against the value computed between population means among different conditions shows that this seems to be true in general for different pairs of parameters of *C. elegans* development (Fig. 3e).

The geometric structure of these correlations can be captured by a low dimensional manifold in the ambient phenotypic space. Non-linear principal components analysis [25] was used to determine the shape of this manifold from the ambient space of rescaled logistic parameters with the developmental durations included (Fig. 4a). The first two non-linear principal components capture 93% of the variance. While this dimensionality reduction may seem modest, in fact, the true dimensionality of the growth curve space is higher with some of the dimensions having already been compressed by the logistic fit. Within this embedding of growth curves, animals tend to cluster in ϕ_1 according to their food source (Fig. 4b) and in ϕ_2 according to their genetic background (Fig. 4c) (note only natural genetic variants are shown). This decomposition can be quantified with a linear regression model predicting either the NLPCA parameters (ϕ_1, ϕ_2) or the rescaled logistic fit parameters $(\hat{l}, \hat{r}, \hat{A})$ using the genetic or environmental conditions as independent variables. Linear regression models of the form:

$$y(G, E) = \beta_0 + \sum_i \beta_{1,i} G + \sum_j \beta_{2,j} E + \epsilon$$

were fit where ϵ, with $y \in (\phi_1, \phi_2, \hat{l}, \hat{r}, \hat{A})$, G and E are indicator variables which take the value of 1 or 0 depending of the strain i and food j in that condition, and ϵ are the residuals to be minimized. In Fig 4b and Fig 4c, the distributions grouped by environment and genotype respectively seem to have different means. We can quantify the this using the F-statistic, which is the variance of the between group means divided by the mean of the within group variances. Using this, we can see that grouping the data by environment does in fact give distinct distributions in ϕ_1 and grouping by genotype gives distinct distributions in ϕ_2 (Fig. 4e). However, a linear regression of the parameters before dimensionality reduction does not give a clean separation between genotype and environment.

The eigenfunctions of the non-linear principle components analysis cannot be derived analytically, but can be investigated by varying each component while keeping the others fixed. If ϕ_2 is fixed ($\phi_2 = 0$), the shapes of the resulting growth curves as ϕ_1 is varied reflect the strong anti-correlation between the parameters l_{max} and r. For positive values of ϕ_1, animals grow quickly during their maximum growth phase but are ultimately shorter (Fig. 4f, blue curve). In contrast, for negative ϕ_1, animals grow more slowly at their peak, but are longer as adults (Fig. 4f green curve). This may indicate a potential trade-off between the speed of growth during development and the ultimate size achieved by adult animals. In contrast, variation along ϕ_2 result both in growth that is slower and in animals that are shorter as adults (Fig. 4g, green curve) or both faster and longer (Fig. 4g, blue curve). Interestingly, variation *along* ϕ_1 does not seem to affect the duration of reproductive development as much as along ϕ_2 as shown by the color of the plotted points (Fig. 4d). Points away from the boundary in the positive ϕ_2 direction correspond to slower reproductive development.

Discussion

Biological systems are remarkable in part because they seem to be both incredibly robust and simultaneously flexible. Biological process faithfully regenerate complex developmental programs every generation, yet these same processes have given rise to an overwhelming diversity of complex matter. Canalisation and developmental plasticity seem to be opposing forces. One can imagine that these forces ebb and flow, acting in sequence. Cryptic variation which can be suppressed or revealed in certain circumstances, e.g. by the chaperone HSP-90 [26], is a good example of this, and it has been shown that disruption of the chaperone at the molecular level act as the switch from robustness to plasticity, even in complex phenotypes such as the morphology of an animal [27]. However, this may be only part of

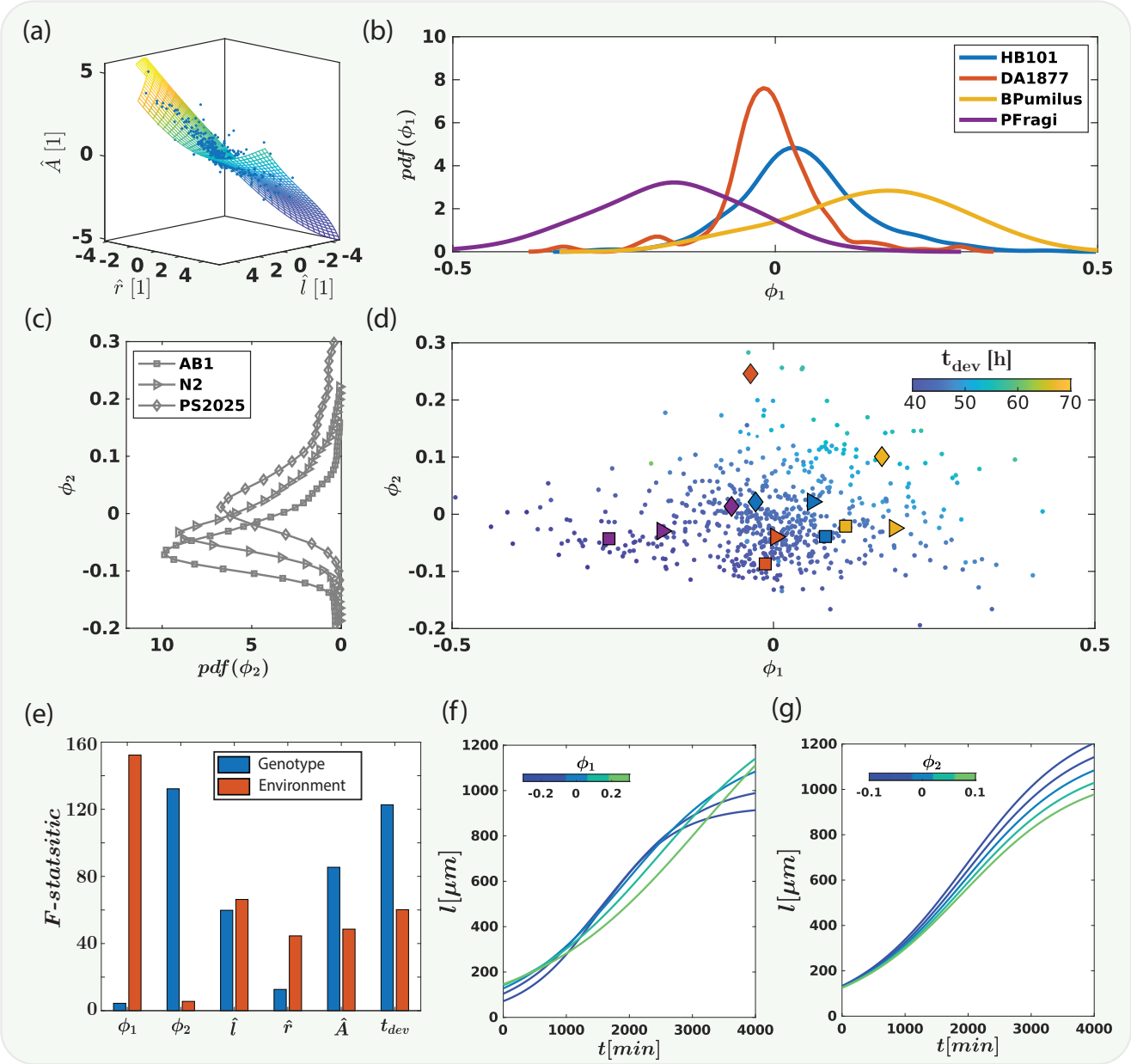
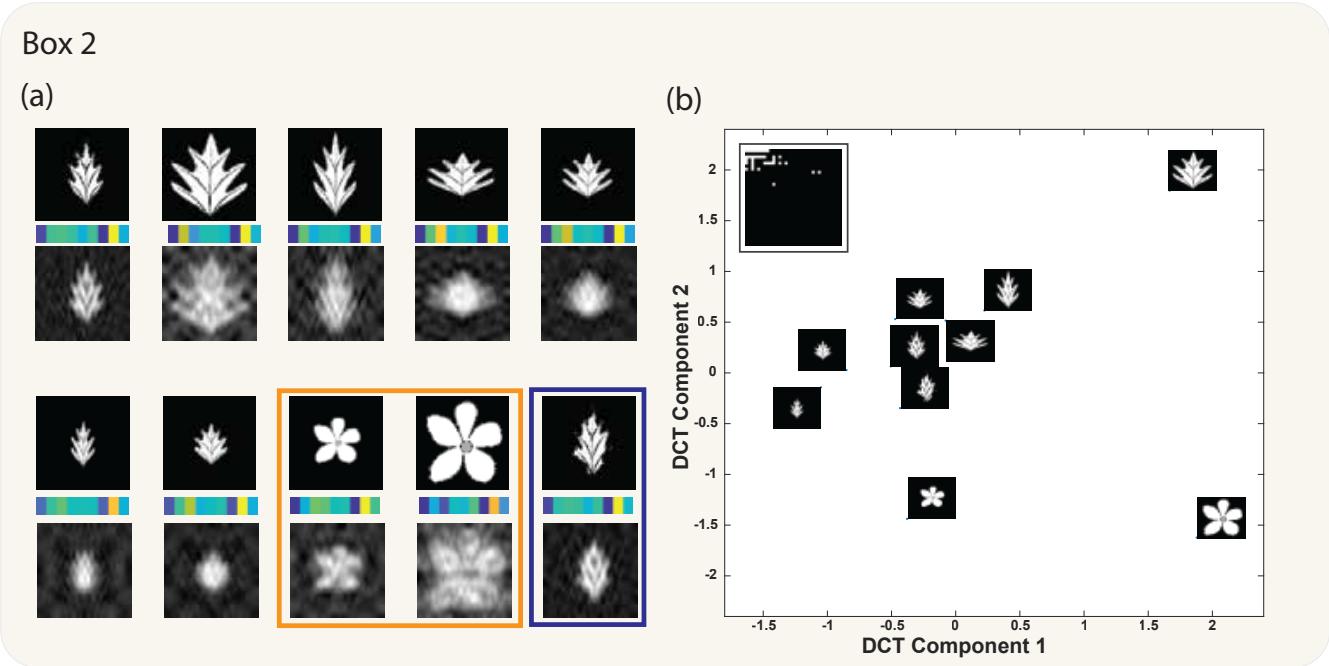


Figure 4: **(a)** The z-score of the three re-scaled logistic fit parameters are shown in a 3-d scatterplot (blue dots). These points lie close to a curved 2-d manifold which was found by performing non-linear principle components analysis (NLPCA). The flattened manifold is shown in **(d)** as a scatterplot where the color of the points indicates t_{dev} for each individual. The mean ϕ_1 and ϕ_2 for each condition are shown as a combination of symbols (*C. elegans* strain) and color (Bacterial food source). On this manifold, ϕ_1 seems to separate the environmental conditions as shown in **(b)** by the marginal distribution over bacterial food sources. In contrast, the marginal distributions **(c)** over the *C. elegans* strains shows separation in ϕ_2. Marginal distributions were computed with a kernel density estimator. This separation in ϕ_1 and ϕ_2 can be quantified by computing the f-statistic for a linear regression model on taking genotype and environment as regressors **(e)**. ϕ_1 regresses primarily on environment and ϕ_2 on genotype. Interestingly, regressing the three logistic fit parameters without first performing NLPCA results in mixed mixed regression on both genotype and environment for each. In **(f)** to determine the effect of varying ϕ_1 on the shape of the growth curve, ϕ_2 was fixed to 0 and ϕ_1 was varied through a range, as indicated by the colorbar, with coordinates being converted back from the unit-less quantities. Similarly in **(g)**,ϕ_1 is fixed and ϕ_2 varied.

the story. Canalisation and flexibility may in fact be complementary properties of systems that are organized to generate low dimensional phenotypic manifolds. Concentration of dimensionality is an intrinsic property of complex high-dimensional systems, even if only trivially because of the Johnson-Lindenstrauss lemma [28]. However, the emergence of low dimensions in biological systems is often orders of magnitude more than random projection would predict [16]. At the heart of concentration of dimensionality lies projection and we propose that is natural to view robustness and plasticity as analogous to orthogonal and parallel projection of variation onto a low dimensional manifold. Given a distribution of variations that arise from environmental heterogeneity, stochastic fluctuations, and genetic mutations, whether a given phenotype or set of phenotypes is buffered or responsive will be will depend on the projection function from high dimensional chemical space to the low dimensional phenotype space. In cases where a phenotype is highly buffered to some set of variations, almost all those variations will be orthogonal to the phenotypic manifold, and variations for which the phenotype is plastic will have large projections onto the manifold. In this scenario one might envision the process of evolution as one of shaping the projection function, and thus the resulting phenotypic manifold, given the statistics of distribution of variations. The process of adaptation can then be viewed as the process of learning the optimal projection over the prior distribution of fluctuations, i.e. an environment to phenotype rather than genotype to phenotype map [29], although the environment to phenotype map is certainly a product of the architecture of the gene regulatory networks of the organism.

It is tempting to look for the genetic basis of emergence of low dimensional phenotypes, especially in *C. elegans*, which has many genetic tools available. However, we believe it is unlikely that a single gene or a genetic module will underlie this process. Rather, it will likely be a property of the entire network of molecular interactions that underlie complex phenotypes. Nevertheless, there may be important connections between the structure of phenotypic manifolds and evolutionary dynamics. It has been proposed that movement along, rather than away from such manifolds constitutes a path of "least resistance" for genetic changes which might fix phenotypic variations. While some evidence supports this hypothesis, e.g. a study of the integration of phenotypic and life-history traits in the flowering plant *Arabidopsis thaliana* [30], other evidence from the morpho-space of the greenfinch *Carduelis chloris* suggests that within population correlations may not predict between population correlations [31]. This discrepancy may be due in part to how traits are quantified, how the dimensionality reduction is performed, and also to whether the populations included in the analysis sample natural genetic variation, genetic mutations, environmental variations, or combinations of all three. While the question of when and in what conditions the "directions" of phenotypic plasticity are predictive of the directions of subsequent genetic evolution remains open, phenotypic plasticity in canalized traits has been shown to be associated with rapid evolutionary diversification. An example comes from the polyphenism in the feeding structures of nematodes, in which the acquisition of mouth form plasticity is associated with an increase in evolutionary rates. Interestingly, even if only one of the two alternative forms is subsequently fixed, the underlying genetic architecture seems to maintain an expanded phenotypic manifold which facilitates future exploration [32]. While traditional forward and reverse genetics may not be the ideal approach, artificial evolution for expanded phenotypic manifolds seems promising, especially with an automated system that can map individuals to the phenotypic manifold in real time and use this as a selection criteria.

In physics, statistical mechanics provides tools to connect microscopic and macroscopic dynamics and to describe the behavior of the macroscopic observables that arise from systems with large numbers of identical interacting components. Examples include the Navier-Stokes equation for fluid flow or the Fokker-Planck equation for diffusion. The coarse graining of many interacting degrees of freedom into a few dominant modes can be formulated precisely in terms of projection operators [33–35]. These projections often make use of time scale separations. For example, the random motion of a Brownian particle results from its many collisions with surrounding molecules, but these collisions occur very fast and can thus be treated as white noise.



Box 2: Illustration of robustness and flexibility resulting from projection onto a low dimensional manifold. **(a)** In this example, 64x64 pixel leaf images represent a high dimensional environment to be sensed. The discrete cosine transform is then used as a projection operator. Past evolutionary history has optimized this projection function to recognize the top left leaf in panel **(a)** and for this, the top 35 modes capture 90% of the energy. The retained modes are shown in **(b, inset)**. In **(a)**, we can see a variety of environmental inputs and their reconstructions from their low dimensional representations, the first 10 mode weightings are shown as color band below each image in **(a)**. Despite the fact that this projection operator evolved for the leaf in the top left, it reconstructs the different leaf images reasonably well. In particular, high frequency variation around the original leaf shape is buffered **(a, blue box)**, resulting in canalisation of the representation such fluctuations in the environmental input. Furthermore, even distinct environmental inputs such as the flowers in **(a, orange box)** are reconstructed fairly well, demonstrating that this projection operator is robust, even in very divergent contexts. **(b)** shows the embedding onto the first two discrete cosine transform components. From this, we can visualize both the robustness and the plasticity associated with this projection operator. The leaves cluster separately from the flowers, revealing flexibility, and the two leaves that only differ in high frequency modes cluster together, showing robustness, with the variation between reconstructed representations smaller than that of inputs (See Supplemental Information). Here the first component is correlated with overall size and the second seems to separate leaves from flowers. It is interesting to note that the projection operator was not optimized in any way to distinguish leave from flowers, and in fact, the eigenfunctions used are the generic ones from the DCT. Over evolutionary time, organisms could optimize not only which modes are retained to better capture the prior distribution over contexts, but also could optimize the shape of the eigenfunctions themselves. This panel depicts an environmental sensing mechanism, but equally important, though not shown here, would be the phenotypic execution mechanism. The structure of the network which maps the low dimensional representation of the environment back into a high dimensional phenotype will also be an important source of variability and affect both robustness and plasticity.

For near equilibrium systems, the rigorous relation between the fluctuations arising from the many unobserved degrees of freedom and the evolution of the system's macroscopic observables are known collectively as Fluctuation-Dissipation relations [36, 37]. While biological systems are characteristically far from equilibrium, similar relations have been observed between phenotypic variability and evolutionary response [38, 39]. In fact, even in systems far from equilibrium, relations of this sort can arise similarly when there is a separation of both observed and unobserved degrees of freedom and of time-scales [40]. However, the origin of low dimensional dynamics and of fluctuation-dissipation like relationships in biological systems remains a mystery. A recent intriguing proposal for the origin of such relations and the emergence of low dimensionality in biological systems is the evolution of robustness. If the same evolutionary forces that increase robustness to noise also tend to increase robustness to genetic perturbations this could account for both phenomena [17]. Recently, Murugan and colleagues have presented a model describing how mutational perturbations may be constrained by global epistasis to excite only a few slow or *soft* modes with examples from protein elasticity and gene regulatory dynamics [41]. Their work shows that such a relationship between mutational induced and physically induced deformations is expected mathematically for protein elasticity. The existence of such slow modes may also be related to the seemingly universal emergence of *stiff* and *sloppy* modes from parameter space compression [42, 43].

In *C. elegans*, development has been shown to be controlled by a massive gene expression oscillator that is comprised of of ≈ 3700 genes [44, 45]. The existence of such an oscillator evokes a direct analogy to projection operator techniques as this networks only has a few excitable oscillatory modes similar to a Fourier transform or the cosine transform employed in Box 2. Recently, the components of a gene regulatory network that comprises a central clock to control this oscillator has been identified [46]. It would be interesting to assess how mutations in these core components affect how fluctuations in gene expression are projected onto the main oscillatory modes that control molting, and how this in turn manifests on the developmental manifold of *C. elegans*.

In this work we have focused on variability resulting from combinations of dietary and genetic perturbations, and we have done so in only fixed environments. It would be interesting to perform experiments in which the developmental trajectory was perturbed by some impulse or step function and to observe the resulting relaxation. *C. elegans* development is highly temperature dependent, and it would be interesting to first assay they developmental trajectories at different temperatures and then to perturb development using various temperature shift protocols. In addition, if we could find perturbations that tend to generate displacement in specific directions on the phenotypic manifold, we could test quantitatively for fluctuation response type relationships.

The projection operator perspective leads one naturally to consider how organisms might learn their particular projection function. Evolution by mutation and selection will surely play a large part in this process, especially when an organism must adapt to changes to the prior distribution of variations, but there is no reason, in principle, that this operator cannot be shaped also by within generation "learning" and other non-genetic and epi-genetic feedback mechanisms.

Conclusions

Biological systems are remarkably robust, and yet the same mechanisms that underlie this robustness have also generated the incredibly diversity of life on earth. In this work we attempt to reconcile these seemingly opposing forces by studying the structure of variability in the development of an animal in the contexts of different genetic backgrounds and environments. We have used high resolution imaging data of the growth of the nematode *C. elegans* in environmental and genetic contexts to map the variability of development. We find that traits which are correlated within a particular context predict whether the mean values of those traits will be correlated among different contexts. Correlation between traits indicate that the true dimensionality of the system may be lower and we find a parsimonious low dimensional representation of the variability, in

which the contributions of genetic variation on the phenotype are separable from those of environmental variation by a simple linear model. We present a framework in which the emergence of low dimensionality provides a basis for both robustness and plasticity. Concentration of dimensionality seems to an intrinsic property of the chemical and molecular networks that generate living systems, and also of complex dynamical systems in general. While beyond the scope of this work, we hope that the connection between projection operators in physical systems, which explain how reproducible coarse grained dynamics arise from the stochastic influences of innumerable microscopic components, may inform the theory in biology of how reproducible developmental dynamics arise from the interactions of similarly large numbers of bio-molecules.

Methods

Imaging hardware The imaging hardware consists of a single 3.2 MP monochome camera (Flea3, Teledyne FLIR) mounted to the arm of an XY-plotting robot (Eleksdraw, Eleksmaker) which moves the arm using two stepper motors controlled by a GRBL stepper motor controller that interprets G-code sent via a USB serial connection using the MATLAB (R2018b, Mathworks) fprintf command. Returning to the same position was accurate to with a few hundred microns, and wells were recentered periodically by fitting a circle to an intensity theshholded image of the well and zeroing the offset. Illumination was provided from above by an LED light panel. To maximize contrast, oblique illumination was blocked by a sheet of black acrylic in which an array of square holes was laser cut to allow for the direct illumination to pass though. A 50mm f/1.4 lens (NMV-50M1, Navitar) was attached to the camera via a 40mm lens tube (CML40, Thorlabs). In this system, one pixel in the image corresponded to $2.67\mu m$, giving an effective magnification of $0.93x$. Detailed instructions and a parts list with suppliers is available upon request.

Multiwell plates Multiwell plates were made by gently placing a pre-cut acrylic form into a standard 30mm Falcon petri dish filled with 5ml of liquefied NGM-Gelrite medium. The multi-well acrylic forms were 24mm square, with a 2x2 grid of 6mm diameter circular wells 4cm from the edge and 10 cm center to center, were cut from 3mm thick black acrylic using an LS 6090 PRO Laser Cutter (HPC Laser Ltd). After placing the multiwell plate into the molten solid media so that it rested on the surface, the plate was allowed to set and to dry for 1 hour covered at 20°C, and seeded with $2\mu l$ of bacteria corresponding to approximately 18 million cells, as measured using a Petroff-Hausser Counting Chamber (Hausser Scientific).

Temperature control Plates were kept in an insulated, temperature controlled box during imaging, which itself was in a temperature controlled room set to 20°C. The actual average temperature in the room was 19.7°C. Temperatures were measured with a custom thermometer; the signal from a linear thermistor (Omega Engineering) was difference amplified against a known voltage corresponding to 20°C. The amplified signal (Gain = 2) was recorded in MATLAB (Mathworks) from the analog input of a DAQ (Labjack). This was fed into a Proportional Integral Differential (PID) control script whose output was a voltage that controlled a Push-Pull current amplifier driving a Peltier effect element (Custom Thermoelectric). Fans were used to distribute air from the heat sinks on each face of the Peltier element either into the enclosure or as exhaust. Temperature was maintained to within 70 mK of the set point (20°C).

Image processing Images were captured directly into Matlab using the built-in video input object class. Moving objects were extracted from each image using background subtraction, to generate a region of interest with the largest amount of detected motion, as measured by the largest pixel difference. Within this region of interest, objects were detected by contrast with the background by applying a threshold to a laplacian of gaussian filtered image. Filtering was performed using the Matlab *imfilter* function with the *fspecial* function with filter size 15x15 and filter standard deviation of 1.5. Connected components were extracted from this threshholded image and the locations and properties of connected components were recorded from the resulting

black and white image. Area (in pixels), and centroid location were calculated using the MATLAB function *regionprops*, and length was computed using the MATLAB functions *bwmorph* to extract the skeleton, the function *bwgeodesic* to compute the length. These properties were used to classify each blob as a worm or not, based on the output of a pre-trained support vector machine. All images were also saved for later inspection, which was used to determine the egg-hatching and egg-laying times manually. All code used for analysis and data processing is available on Github, including data from processed images (Raw images are available on request).

Nematode culture and strains *C. elegans* was grown on NGM agar plates and fed *E. coli* HB101 for standard maintenance at 20 °C. Bristol N2 was used as wild-type strain [47]. In addition to N2, two wild isolate strains were used, AB1 isolated in Adelaide, Australia, and PS2025 isolated in Pasadena, California. In additon, two single gene mutants of N2 were used, *sid-2(mj465)*, a mutant that is not competent for RNAi by feeding, and *c28h8.3(mj649)* a catalytic mutant of a putative helicase.

Bacterial culture and strains In addition to *E. coli* HB101, three other bacteria were used as food sources. *Comamonas aquatica* DA1877 was used as it has been previously shown to increase the rate of *C. elegans* development by providing supplemental vitamin B12 [48, 49]. We also chose *Bacillus pumilus* and *Pseudomonas fragi* from the wild bacteria collection [22] as we had previously observed large developmental differences on these food sources.

Synchronization by coordinated egg laying To generate populations of synchronized animals without bleaching and starvation, young egg-laying adults (50-75 hours post hatching) were gently picked with a platinum wire to a fresh NGM plate seeded with the appropriate bacteria and allowed to lay eggs for a fixed duration, after which the adults were removed from the plate and the eggs were collected. The egg-laying rate of animals at this stage is ≈ 6 eggs/animal/hour. Synchronization could be tightened by shortening the duration of egg-laying, and the number of synchronized eggs could be increased by using more egg-laying adults.

Nematode growth media - gelrite NGM gelrite plates were made by replacing the Agar in normal NGM recipe with Gellan Gum (Sigma Aldritch), a polymer derived from algae. The recipe is given in Table 1. In addition, peptone and cholesterol were omitted to prevent bacterial growth on the plate, so that the only available bacterial food was that which was initially inoculated. Additive salts were prepared in 1 M stock solutions, and the KH_2PO_4 stock solution was adjusted to pH 6.

Table 1: NGM Gelrite

Ingredient	Amount
Autoclave Together	
NaCl	3.0 g/l
Gellan Gum	8.0 g/l
Add Aseptically	
$CaCl_2$	1 mM
$MgCl_2$	1 mM
KH_2PO_4	25 mM

Phylogenetic tree of wild isolates The phylogenetic tree of *C. elegans* wild isolates was generated from the full CeNDR phylogenetic tree [23] which was hard-filtered by isotype using the *prune* function in MATLAB.

Logistic fits Logistic fits were calculated using the MATLAB implementation of the Levenberg Marquardt [50] non-linear least squares algorithm within the *lsqcurvefit* package. Logistic fits were performed on the raw data as well as the rescaled data. The rescaled fits were nearly identical to the raw-fit parameters which were transformed according to Eq. 2 (See Supplemental Information Fig S1).

Non-linear PCA Nonlinear principle components analysis attempts to find an optimal auto encoder that can recreate the input data after passing it though a bottle neck layer with fewer components

than the input. The number of components in the bottle neck layer corresponds to the number of desired principle components. The NLPCA implementation we use is from NLPCA toolbox for MATLAB [25] and uses a multilayer perceptron architecture with a hyperbolic tangent activation function in the hidden layers.

Linear Regression Each phenotypic output can be decomposed as a genetic contribution, an environmental contribution, and some residual error. Linear regression was performed using the *fitlm* function in MATLAB. if y is the phenotypic variable of interest,

$$y(G,E) = \beta_0 + \sum_i \beta_{1,i} G + \sum_j \beta_{2,j} E + \epsilon$$

were fit where ϵ, with $y \in (\phi_1, \phi_2, \hat{l}, \hat{r}, \hat{A})$, G and E are indicator variables which take the value of 1 or 0 depending of the strain i and food j in that condition, and ϵ are the residuals to be minimized. *fitlm* uses an iteratively reweighted least squares algorithm. For example, the best fit linear model for ϕ_1 indicates that the average value of ϕ_1 is 0.022 and that changing to *Bacillus pumilus* 0.115 and switching to *Pseudomonas fragi* moves ϕ_1 -0.200. F-statistics were calculated by the MATLAB *anova* function, whichi si given by

$$F = \mathbb{V}[\mathbb{E}(y_i)]/\mathbb{E}[\mathbb{V}(y_i)]$$

again, calculated for each prediction variable $y \in (\phi_1, \phi_2, \hat{l}, \hat{r}, \hat{A})$ and decomposed according to environment or genotype $i \in E, G$.

Acknowledgements The authors would like to thank Marie-Anne Felix for providing us with wild-isolates of *C. elegans* as well as wild bacteria. We would also like to thank Stanislas Leibler, Bing Kan Xue, Maros Pleska, Jon Chuang, and Ricardo Rao for helpful discussion, and in particular S.L. for advice on the presentation of the material and B.K.X for suggesting the rescaling and the linear decomposition of the variance.

References

[1] S. Rulands and B. D. Simons, "Emergence and universality in the regulation of stem cell fate", Current Opinion in Systems Biology **5**, 57–62 (2017).

[2] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, "Dimensionality and dynamics in the behavior of c. elegans", en, PLoS Comput. Biol. **4**, e1000028 (2008).

[3] Z. Frentz, S. Kuehn, and S. Leibler, "Strongly deterministic population dynamics in closed microbial communities", Phys. Rev. X **5**, 041014 (2015).

[4] C. H. Waddington and H. Kacser, *The strategy of the genes: a discussion of some aspects of theoretical biology*, en (Allen & Unwin, 1957).

[5] J. S. Chuang, Z. Frentz, and S. Leibler, "Homeorhesis and ecological succession quantified in synthetic microbial ecosystems", en, Proc. Natl. Acad. Sci. U. S. A. **116**, 14852–14861 (2019).

[6] M. Turk and A. Pentland, "Eigenfaces for recognition", en, J. Cogn. Neurosci. **3**, 71–86 (1991).

[7] O. Campàs, R. Mallarino, A. Herrel, A. Abzhanov, and M. P. Brenner, "Scaling and shear transformations capture beak shape variation in darwin's finches", en, Proc. Natl. Acad. Sci. U. S. A. **107**, 3356–3360 (2010).

[8] D. M. Raup, "Geometric analysis of shell coiling: general problems", J. Paleontol. **40**, 1178–1190 (1966).

[9] D. J. Smith, A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. M. E. Osterhaus, and R. A. M. Fouchier, "Mapping the antigenic and genetic evolution of influenza virus", en, Science **305**, 371–376 (2004).

[10] W. Bialek, "On the dimensionality of behavior", en, Proc. Natl. Acad. Sci. U. S. A. **119**, e2021860119 (2022).

[11] A. C. Costa, T. Ahamed, D. Jordan, and G. J. Stephens, "Maximally predictive states: from partial observations to long timescales", en, Chaos **33**, 023136 (2023).

[12] C. Brennan and A. Proekt, "A quantitative model of conserved macroscopic dynamics predicts future motor commands", en, Elife **8** (2019).

[13]D. Jordan, S. Kuehn, E. Katifori, and S. Leibler, "Behavioral diversity in microbes and low-dimensional phenotypic spaces", en, Proc. Natl. Acad. Sci. U. S. A. **110**, 14018–14023 (2013).

[14]M. Pleška, D. Jordan, Z. Frentz, B. Xue, and S. Leibler, "Nongenetic individuality, changeability, and inheritance in bacterial behavior", en, Proc. Natl. Acad. Sci. U. S. A. **118** (2021).

[15]S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber, "Cell fates as high-dimensional attractor states of a complex gene regulatory network", en, Phys. Rev. Lett. **94**, 128701 (2005).

[16]J.-P. Eckmann and T. Tlusty, "Dimensional reduction in complex living systems: where, why, and how", en, Bioessays **43**, e2100062 (2021).

[17]K. Kaneko and C. Furusawa, "Direction and constraint in phenotypic evolution: dimension reduction and global proportionality in phenotype fluctuation and responses", in *Evolutionary systems biology: advances, questions, and opportunities*, edited by A. Crombach (Springer International Publishing, Cham, 2021), pp. 35–58.

[18]S. Klumpp and T. Hwa, "Bacterial growth: global effects on gene expression, growth feedback and proteome partition", en, Curr. Opin. Biotechnol. **28**, 96–102 (2014).

[19]O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, and U. Alon, "Evolutionary Trade-Offs, pareto optimality, and the geometry of phenotype space", Science **336**, 1157–1160 (2012).

[20]M. Pigliucci, "Phenotypic integration: studying the ecology and evolution of complex phenotypes", en, Ecol. Lett. **6**, 265–272 (2003).

[21]V. Alba, J. E. Carthew, R. W. Carthew, and M. Mani, "Global constraints within the developmental program of the drosophila wing", en, Elife **10** (2021).

[22]L. Frézal and M.-A. Félix, "C. elegans outside the petri dish", en, Elife **4** (2015).

[23]D. E. Cook, S. Zdraljevic, J. P. Roberts, and E. C. Andersen, "CeNDR, the caenorhabditis elegans natural diversity resource", en, Nucleic Acids Res. **45**, D650–D657 (2017).

[24]K. Stojanovski, H. Großhans, and B. D. Towbin, "Coupling of growth rate and developmental tempo reduces body size heterogeneity in c. elegans", en, Nat. Commun. **13**, 3132 (2022).

[25]M. Scholz, F. Kaplan, C. L. Guy, J. Kopka, and J. Selbig, "Non-linear PCA: a missing data approach", en, Bioinformatics **21**, 3887–3895 (2005).

[26]S. L. Rutherford and S. Lindquist, "Hsp90 as a capacitor for morphological evolution", en, Nature **396**, 336–342 (1998).

[27]B. Sieriebriennikov and R. J. Sommer, "Developmental plasticity and robustness of a nematode Mouth-Form polyphenism", en, Front. Genet. **9**, 382 (2018).

[28]W. B. Johnson and J. Lindenstrauss, *Extensions of lipschitz mappings into a hilbert space*, Providence, Rhode Island, 1984.

[29]B. Xue, P. Sartori, and S. Leibler, "Environment-to-phenotype mapping and adaptation strategies in varying environments", en, Proc. Natl. Acad. Sci. U. S. A. **116**, 13847–13855 (2019).

[30]M. Pigliucci and K. Hayden, "Phenotypic plasticity is the major determinant of changes in phenotypic integration in arabidopsis", en, New Phytol. **152**, 419–430 (2001).

[31]J. Merilä and M. Björklund, "Population divergence and morphometric integration in the greenfinch (carduelis chloris) – evolution against the trajectory of least resistance?", en, J. Evol. Biol. **12**, 103–112 (1999).

[32]V. Susoy, E. J. Ragsdale, N. Kanzaki, and R. J. Sommer, "Rapid diversification associated with a macroevolutionary pulse of developmental plasticity", en, Elife **4** (2015).

[33]H. Mori, "Transport, collective motion, and brownian motion", en, Progr. Theoret. Phys. **33**, 423–455 (1965).

[34]R. Zwanzig, "Memory effects in irreversible thermodynamics", Phys. Rev. **124**, 983–992 (1961).

[35]S. Nakajima, "On quantum theory of transport phenomena", Progr. Theoret. Phys. **20**, 948–959 (1958).

[36]H. B. Callen and T. A. Welton, "Irreversibility and generalized noise", Phys. Rev. **83**, 34–40 (1951).

[37]R. Kubo, "The fluctuation-dissipation theorem", en, Rep. Prog. Phys. **29**, 255 (1966).

[38]K. Sato, Y. Ito, T. Yomo, and K. Kaneko, "On the relation between fluctuation and response in biological systems", en, Proc. Natl. Acad. Sci. U. S. A. **100**, 14086–14090 (2003).

[39]Q.-Y. Tang and K. Kaneko, "Dynamics-Evolution correspondence in protein structures", en, Phys. Rev. Lett. **127**, 098103 (2021).

[40]G. Jung and F. Schmid, "Fluctuation-dissipation relations far from equilibrium: a case study", en, Soft Matter **17**, 6413–6425 (2021).

[41]K. Husain and A. Murugan, "Physical constraints on epistasis", en, Mol. Biol. Evol. **37**, 2865–2874 (2020).

[42]R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally sloppy parameter sensitivities in systems biology models", en, PLoS Comput. Biol. **3**, 1871–1878 (2007).

[43]B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, "Parameter space compression underlies emergent theories and predictive models", en, Science **342**, 604–607 (2013).

[44]G.-J. Hendriks, D. Gaidatzis, F. Aeschimann, and H. Großhans, "Extensive oscillatory gene expression during c. elegans larval development", en, Mol. Cell **53**, 380–392 (2014).

[45]M. W. Meeuse, Y. P. Hauser, L. J. Morales Moya, G.-J. Hendriks, J. Eglinger, G. Bogaarts, C. Tsiairis, and H. Großhans, "Developmental function and state transitions of a gene expression oscillator in caenorhabditis elegans", en, Mol. Syst. Biol. **16**, e9498 (2020).

[46]M. W. M. Meeuse, Y. P. Hauser, S. Nahar, A. A. T. Smith, K. Braun, C. Azzi, M. Rempfler, and H. Großhans, "C. elegans molting requires rhythmic accumulation of the Grainyhead/LSF transcription factor GRH-1", en, EMBO J., e111895 (2023).

[47]S. Brenner, "The genetics of caenorhabditis elegans", en, Genetics **77**, 71–94 (1974).

[48]E. Watson, L. T. MacNeil, A. D. Ritter, L. S. Yilmaz, A. P. Rosebrock, A. A. Caudy, and A. J. M. Walhout, "Interspecies systems biology uncovers metabolites affecting c. elegans gene expression and life history traits", en, Cell **156**, 759–770 (2014).

[49]L. T. MacNeil, E. Watson, H. E. Arda, L. J. Zhu, and A. J. M. Walhout, "Diet-induced developmental acceleration independent of TOR and insulin in c. elegans", en, Cell **153**, 240–252 (2013).

[50]D. W. Marquardt, "An algorithm for Least-Squares estimation of nonlinear parameters", Journal of the Society for Industrial and Applied Mathematics **11**, 431–441 (1963).

[51]A. J. Chang, N. Chronis, D. S. Karow, M. A. Marletta, and C. I. Bargmann, "A distributed chemosensory circuit for oxygen preference in c. elegans", en, PLoS Biol. **4**, e274 (2006).

[52]A. J. Bretscher, K. E. Busch, and M. de Bono, "A carbon dioxide avoidance behavior is integrated with responses to ambient oxygen and food in *Caenorhabditis elegans*", en, Proc. Natl. Acad. Sci. U. S. A. **105**, 8044–8049 (2008).

[53]S. Uppaluri and C. P. Brangwynne, "A size threshold governs caenorhabditis elegans developmental progression", en, Proc. Biol. Sci. **282**, 20151283 (2015).

[54]R. Laranjeiro, G. Harinath, J. E. Hewitt, J. H. Hartman, M. A. Royal, J. N. Meyer, S. A. Vanapalli, and M. Driscoll, "Swim exercise in *Caenorhabditis elegans* extends neuromuscular and gut healthspan, enhances learning ability, and protects against neurodegeneration", en, Proc. Natl. Acad. Sci. U. S. A. **116**, 23829–23839 (2019).

Supplemental Information

Logistic fit rescaling Best fit logistic function parameters could be obtained from either the raw data or the rescaled data to yield equivalent results when the parameters fit from the raw data were adjusted according to Eq 2. This can be seen in the 3-d scatterplot of the raw adjusted parameters (Fig. S1a), where the fit parameters for the rescaled data are shown as orange circles and the fit parameters from the raw data adjusted by Eq 2 are shows as blue dots. If we qwuantify the mean squared error betweeen the fits fromt hese two different procedures, we find that most have 0 deviaiton and the maximum deviation is less than $2 \cdot 10^{-3}$ (Fig S1b).

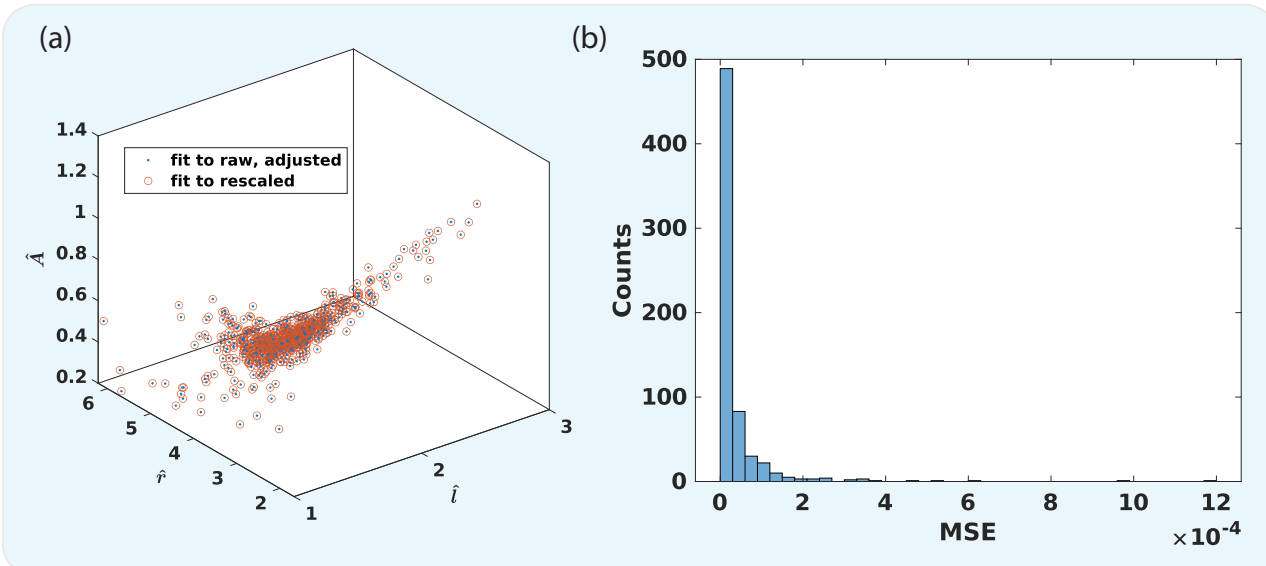


Figure S1: **(a)**

Fraction on Food Analysis It has been noted that the development of *C. elegans* in environments where they are completely surrounded by food is more rapid than on plates where they can be on or off food. It is also known that the tendency for animals to be on food varies among strains [51, 52]. In these experiments, animal behavior can have an influence on growth, as the experimental design provides an opportunity for the worms to leave and enter the food patch. Micro-fluidic chamber based systems are another alternative [53] in which bacterial density can be precisely controlled, however, the aqueous environment can result in "thrashing" behavior, which has been shown to affect the animals physiology [54]. Because, in our systems, animals can freely move on and off food, it possible that behavioral differences could account for developmental changes. To asses this, we quantified whether animals that spend more time on food were more likely to develop more quickly. We find that this does not seem to be the case, as there is no significant correlation between developmental duration and the fraction of time animals were found on food. In addition to individuals, the average development time was calculated for each combination of environment and genotype, and plotted against the total fraction of time spent by all individuals on food and this also does not show a significant correlation.

In this analysis, on food is taken to be less than 1.33 mm from the center of the well. This roughly corresponds to the region on the mini plat where the 2 μl of bacterial suspension is placed. To calculate the on food percentage, the radial distance of each detected worm centroid is computed by subtracting the well center according to $r = \sqrt{(x - x_c)^2 + (y - y_c)^2}$. A radial distribution function (rdf) was then estimated with a histogram using equal area annular bins. The overall (rdf) for each condition is shown in Fig S3(a)

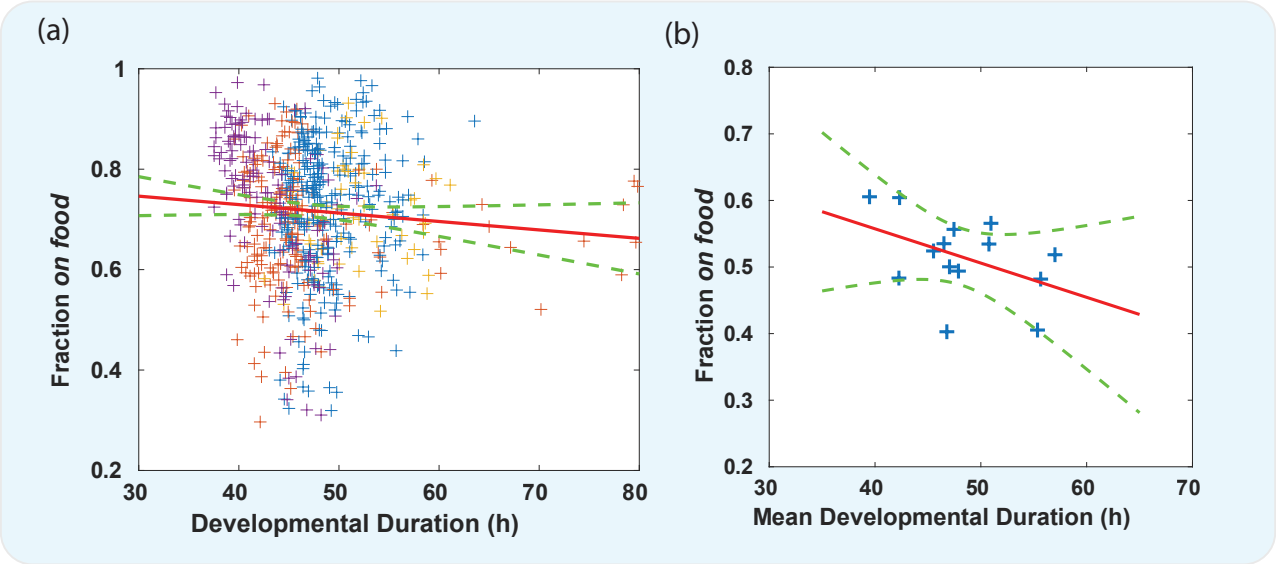


Figure S2: **(a)** For each individual, the fraction on food is plotted against that individuals developmental duration. Fraction on food is the fraction of frames the animal id recorded less than 1.33mm from the center of the mini well. In addition, **(b)** shows the mean fraction on food and the mean developmental duration for all individuals in a condition. The best fit linear regression (red line) with is shown with its 95% confidence interval (green dashed lines). The slope of the regression is not statistically significantly different from 0 for either the individuals or the means in each condition. Colors represent the type of bacterial food for those individuals.

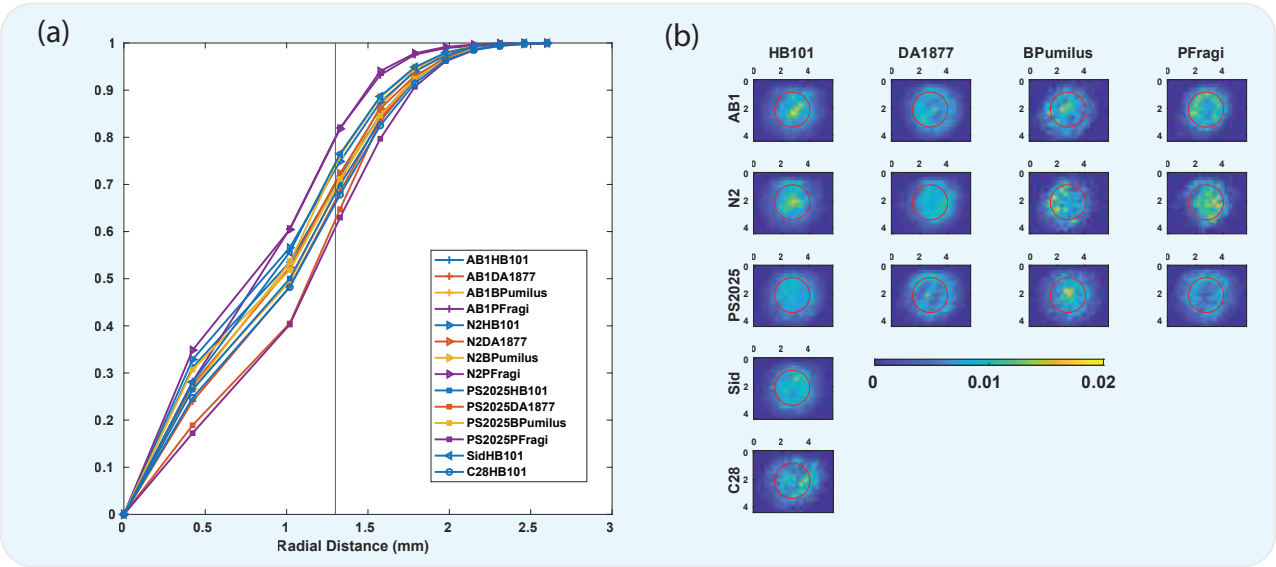


Figure S3: **(a)** Cumulative radial distribution functions of the centroid locations of all animals in a given condition as denoted by the combination of color (food) and symbol (genotype). The blue line represents the demarcation between on food ($< 1.33mm$) and off food ($\geq 1.33mm$). **(b)** Each heatmap shows the 2-d spatial distribution functions for all individuals in a given condition, estimated by a histogram. Inside of the red circle indicates the on food region.

as a cumulative distribution function, in addition to the 2-d spatial distribution functions with the on food area denoted by a red circle.