LETTER

doi:10.1038/nature16550

The temporal scaling of *Caenorhabditis elegans* ageing

Nicholas Stroustrup[1], Winston E. Anthony[1], Zachary M. Nash[1]†, Vivek Gowda[1], Adam Gomez[1]†, Isaac F. López-Moyado[1]†, Javier Apfeld[1]†§ & Walter Fontana[1]§

The process of ageing makes death increasingly likely, involving a random aspect that produces a wide distribution of lifespan even in homogeneous populations[1,2]. The study of this stochastic behaviour may link molecular mechanisms to the ageing process that determines lifespan. Here, by collecting high-precision mortality statistics from large populations, we observe that interventions as diverse as changes in diet, temperature, exposure to oxidative stress, and disruption of genes including the heat shock factor *hsf-1*, the hypoxia-inducible factor *hif-1*, and the insulin/IGF-1 pathway components *daf-2*, *age-1*, and *daf-16* all alter lifespan distributions by an apparent stretching or shrinking of time. To produce such temporal scaling, each intervention must alter to the same extent throughout adult life all physiological determinants of the risk of death. Organismic ageing in *Caenorhabditis elegans* therefore appears to involve aspects of physiology that respond in concert to a diverse set of interventions. In this way, temporal scaling identifies a novel state variable, $r(t)$, that governs the risk of death and whose average decay dynamics involves a single effective rate constant of ageing, k_r. Interventions that produce temporal scaling influence lifespan exclusively by altering k_r. Such interventions, when applied transiently even in early adulthood, temporarily alter k_r with an attendant transient increase or decrease in the rate of change in r and a permanent effect on remaining lifespan. The existence of an organismal ageing dynamics that is invariant across genetic and environmental contexts provides the basis for a new, quantitative framework for evaluating the manner and extent to which specific molecular processes contribute to the aspect of ageing that determines lifespan.

Body temperature is a major determinant of lifespan in poikilotherms[3–5] that also influences mammalian ageing[6]. From 20 °C to 33 °C, the mean lifespan of *C. elegans* decreases 40-fold[7]. To explore the impact of temperature on the actual distribution of lifespans, we used our automated imaging technology[8] to collect highly resolved mortality data in multiple replicate populations placed across this temperature range (Methods). From these data we estimated the survival curve $S(t)$, which is the probability of being alive at time (age) t, and the hazard function $h(t) = -\mathrm{d}\log S(t)/\mathrm{d}t$, which is the instantaneous risk of death at time t (Supplementary Note 1.1 and Methods).

In many invertebrates, changes in temperature alter the rate at which the risk of death increases with time[4,5,9]. Our lifespan data, controlled for environmental heterogeneity (see statistical methods section in Methods), confirmed this effect. However, we further observed that changes in temperature appeared to shift $h(t)$ by an equal and opposite amount in magnitude and time when plotted on a log–log scale, suggesting that between any two temperatures T_0 and T_1, $\lambda h_{T_1}(t) = h_{T_0}(\lambda^{-1}t)$ independent of any particular parametric form of $h(t)$. This change in hazard corresponds to a simple stretching of the survival function along the time axis by a dimensionless scale factor λ: $S_{T_1}(t) = S_{T_0}(\lambda^{-1}t)$ (Supplementary Note 1.2). The sole effect of changes in body temperature on lifespan therefore appeared to be a temporal rescaling of mortality statistics.

To confirm this effect, we applied an accelerated failure time (AFT) regression model[10] in which lifespan distributions that only differed by temporal scaling would have identically distributed residuals (Supplementary Notes 1.3 and 1.4 and Methods). To identify any significant differences between AFT residual distributions, we applied a Kolmogorov–Smirnov test adapted to censored data (Supplementary Note 2). We identified no significant temperature-dependent deviations from temporal scaling within two thermal ranges: 19–30 °C and 30.5–33 °C (Fig. 1b–d and Extended Data Figs 1–3). Populations above 30.5 °C exhibited a more pronounced late-age deceleration (Fig. 1e, Extended Data Fig. 3 and Supplementary Note 1.4), consistent with an increased heterogeneity[11] (Supplementary Note 3). Yet, even at high temperatures, the observed hazard function appears to be dominated more by ageing (for example, a progressive increase in the hazard) than by chance events that would produce a constant hazard (that is, non-ageing).

We then asked whether other interventions could produce a temporal scaling. Since oxidative damage has been linked to ageing across taxa[12,13], we quantified the effect of the oxidant *tert*-butyl hydroperoxide (tBuOOH) and found that it quantitatively rescales lifespan distributions in a dose-dependent manner up to 3 mM (Kolmogorov–Smirnov $P > 0.02$) with significant deviations observed only at 6 mM (Kolmogorov–Smirnov $P = 9 \times 10^{-4}$; Fig. 1f–g and Extended Data Fig. 4).

To further explore the range of interventions that might yield temporal scaling, we considered three members of the insulin/IGF-1 pathway[5,9]: *daf-16*, a transcription factor required for lifespan extension by multiple signals[14], *age-1*, a regulatory kinase upstream of *daf-16*, and *daf-2*, the insulin/IGF receptor, all of which influence both lifespan and thermal stress resistance[7]. Each mutant population exhibited a lifespan distribution rescaled from the wild-type distribution, both at 20 °C (Kolmogorov–Smirnov $P > 0.015$; Fig. 2a–e) and at 33 °C (Kolmogorov–Smirnov $P > 0.017$; Extended Data Fig. 4). The insulin/IGF receptor *daf-2* influences the activity of the heat shock factor *hsf-1* (ref. 15), and disruption of *hsf-1* also shortens lifespan by temporal rescaling (Kolmogorov–Smirnov $P > 0.2$; Fig. 2c, f). Elimination of the hypoxia-inducible transcription factor *hif-1*, known to influence lifespan through *daf-16*-dependent mechanisms[16], behaved likewise (Kolmogorov–Smirnov $P > 0.2$; Extended Data Fig. 4).

Since changes in nutrition alter lifespan across taxa[17], we considered two modifications of *C. elegans* diet: ultraviolet inactivation of the bacterial food source[18] and disruption of feeding behaviour by the *eat-2*(*ad1116*) mutation[19]. Ultraviolet inactivation of bacteria extended

[1]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present addresses: Department of Microbiology and Immunology, School of Medicine, University of North Carolina at Chapel Hill, North Carolina 27599, USA (Z.M.N.); Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095, USA (A.G.); Division of Signaling and Gene Expression, La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA (I.F.L.-M.); Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA (J.A.).
§These authors jointly supervised this work.

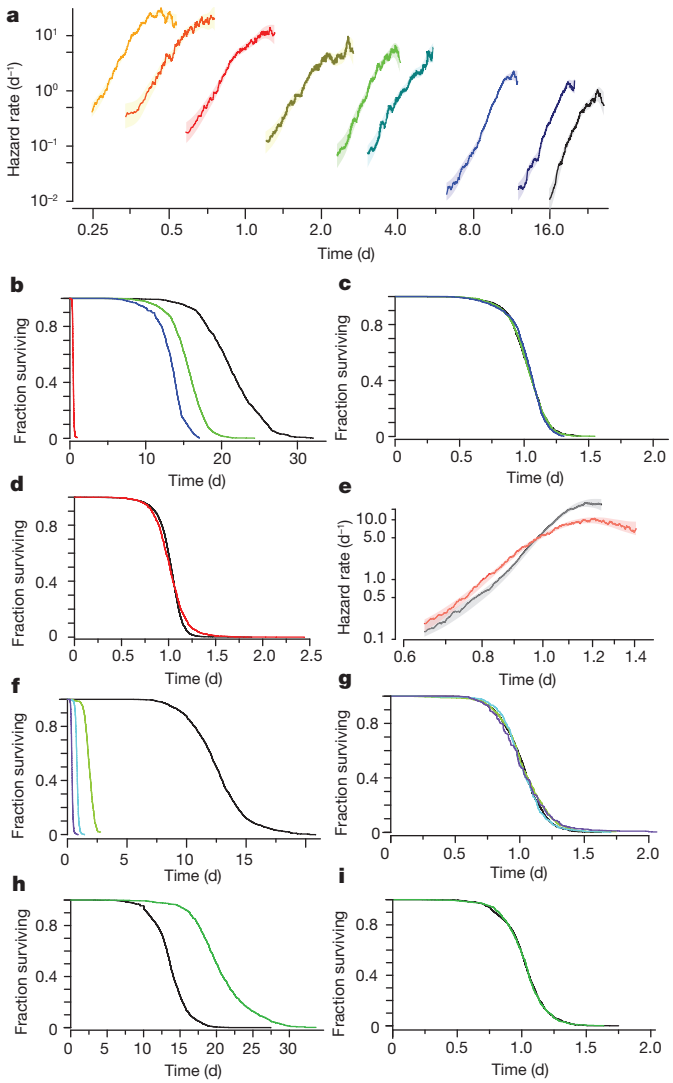


Figure 1 | Environmental determinants rescale *C. elegans* lifespan distributions. a, Populations grown at 20 °C were transferred on their second day of adulthood to a final temperature of (right to left) 20.1 °C (black), 23.7, 25.2, 29.1, 30, 30.9, 31.3, 32.5, and 32.6 (yellow). Individual lifespans were collected[7] and used to estimate the hazard function of each population using numerical differentiation of the Kaplan–Meier survival estimator (solid lines). The shaded areas represent the 95% confidence bands of the true hazard (Statistical methods). d, days. **b**, The lifespan of individuals living at 20, 25, 27, and 33 °C. **c**, The data in **b** were fitted with an AFT model $\log(y_i) = \beta x_i + \epsilon_i$ to remove differences in timescale (Methods and Supplementary Note 1.3). The AFT residuals $\exp(\epsilon_i)$ corresponding to populations at 20, 25, and 27 °C are plotted using the Kaplan–Meier survival estimator. **d**, The AFT residuals corresponding to populations held at 25 (black) and 33 °C (red) are plotted using the Kaplan–Meier survival estimator. **e**, Hazard functions were estimated from the 25 and 33 °C AFT residuals. **f**, The survival curves of populations exposed to 0 (black), 1.5 (blue), 3 (green), and 6 mM (purple) tBuOOH. **g**, The AFT residuals for the data of **f**. **h**, The survival curves of animals cultured on live *E. coli* (black) and ultraviolet-inactivated *E. coli* (green). **i**, The AFT residuals for the data of **h**.

lifespan via temporal scaling (Kolmogorov–Smirnov $P > 0.2$; Fig. 1h, i). In contrast, *eat-2*(*ad1116*) populations exhibited a significant deviation from temporal scaling (Kolmogorov–Smirnov $P = 5 \times 10^{-5}$), with a disproportionate increase in the standard deviation of lifespan compared with the mean (Fig. 2g, j). We also noted that *eat-2*(*ad1116*) populations exhibited a substantially increased variation in developmental timing. While such variation does not affect lifespan statistics based on manually synchronized young adults (Methods), it is possible that the causes of this developmental variation also underlie the increased variation of lifespan. We found that disruption of the mitochondrial complex I in *nuo-6*(*qm200*) populations produced analogous effects on developmental timing with a deviation from temporal scaling of lifespan similar to *eat-2*(*ad1116*) (Kolmogorov–Smirnov $P > 3 \times 10^{-18}$; Fig. 2h, k). Yet, populations with either allele exhibited temporally rescaled lifespan distributions in response to temperature changes (Kolmogorov–Smirnov $P > 0.2$; Fig. 2i, l and Extended Data Fig. 4). We conclude that while *eat-2*(*ad1116*), *nuo-6*(*qm200*), and shifts in temperatures from below to above 30 °C alter lifespan distributions outside the temporal scaling model, these interventions do not eliminate the ability of *C. elegans* to respond to subsequent interventions with temporal scaling. Temporal scaling thus appears to be a pervasive response to interventions of diverse modality and intensity.

Temporal scaling would arise if all physiological determinants of the risk of death in *C. elegans* acted as if they were jointly governed by a single stochastic process whose rate constant alone was altered by interventions (Supplementary Note 4). If the risk of death was determined in this way, we reasoned that transient interventions early in adulthood would produce a persistent temporal shift, not a scaling, of mortality statistics (Supplementary Note 4.3). To test this, we focused on temperature, which can be quantitatively, rapidly, and reversibly switched at any age between a baseline temperature T_0 and a transient temperature T_1 (Fig. 3a). We confirmed that transient exposure to higher temperatures produced a permanent shortening of lifespan[5] (Fig. 3b). We found that this shortening consisted of a temporal shift of the lifespan distribution (Fig. 3c, d) $S_{T_1}(t) = S_{T_0}(t - \Delta_\tau)$ that matches the magnitude of shift Δ_τ predicted if time were rescaled only for the period τ that animals were held at the transient temperature: $\Delta_\tau = \tau(1 - \lambda^{-1})$, with λ the scale factor relating populations always held at T_1 to populations always held at T_0 (Fig. 3e, f, Supplementary Note 4.3, Supplementary Table 2 and Extended Data Fig. 5). In a complementary experiment, we found that exposure to high temperature for different periods τ also gave shifts with the predicted magnitude (Extended Data Fig. 5). It appears, therefore, that the temporal scaling observed in Fig. 1a and the temporal shifting of Fig. 3 are compatible with a single model in which interventions alter the effective rate constant of a stochastic process governing those aspects of *C. elegans* physiology that determine risk of death. This process is evidently ongoing even very early in adulthood and is governed by the same rate constant as in late adulthood.

To clarify how molecular pathways contribute to temporal scaling, we quantified the magnitude of scaling produced by different intensities of intervention: that is, the scaling function. In the case of temperature, we applied an Arrhenius analysis[20,21] to interpret the change of λ (which in our framework rescales the rate constant of ageing) across the range 20–35 °C (Fig. 4a). We identified three distinct thermal regimes: I, 20–29.4 °C; II, 29.4–32.1 °C; III, 32.1–35 °C (Fig. 4b, Methods and Extended Data Figs 6 and 7) with regime I being further subdivided into Ia and Ib by a reproducible transition point at 24.4 °C.

Each scaling regime appears to correspond to a distinct molecular mechanism and barrier process dominating the timescale of ageing (Supplementary Table 1). Sharp decreases in lifespan have been observed to occur around 30 °C in *Drosophila melanogaster*[21], hinting at a more general phenomenon in poikilotherms. Notably, this transition coincided with a deviation from temporal scaling of lifespan distributions (Fig. 1e and Extended Data Fig. 3). Intriguingly, the scaling across the breakpoint between regimes Ia and Ib suggested that temporal scaling need not be disrupted by a change in the molecular mechanisms dominating the timescale of ageing.

Quantifying the effects of temperature on mutant strains, we found that the elimination of DAF-16 shorted lifespan by a rescaling of 28% in regime Ia and 25% in Ib (Fig. 4c, d). The *daf-16*(*mu86*) population exhibited the same slope in scaling function as wild type in Ia, and differed only by about 5% across regime Ib, suggesting that the mechanisms mediating the temperature dependence of lifespan

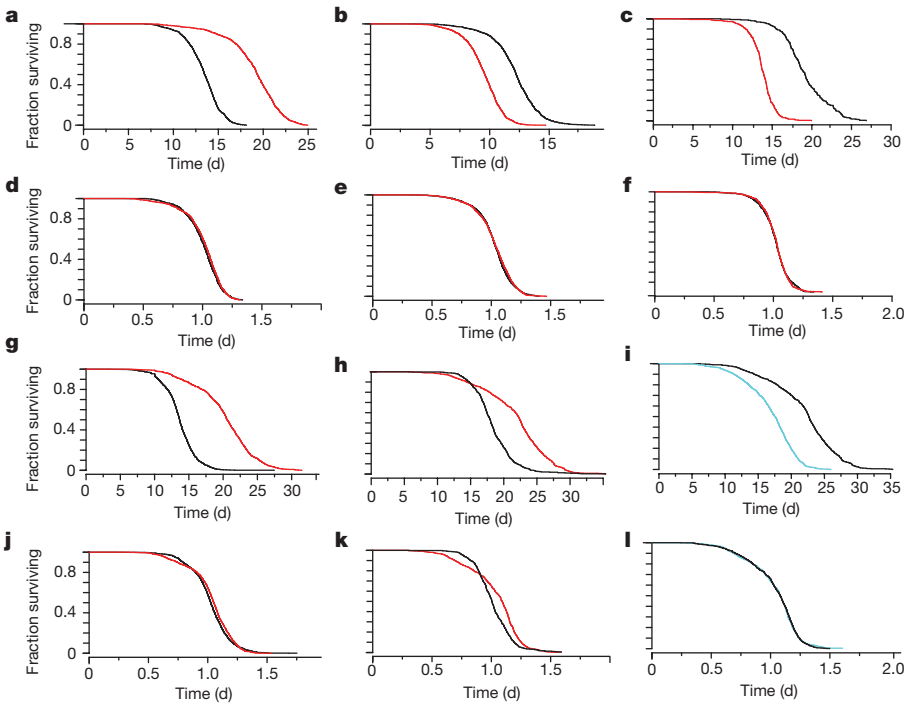


Figure 2 | Genetic determinants rescale *C. elegans* lifespan distributions. a–c, Survival curves are shown for *daf-2*(*e1368*) (red) and wild type (black) at 25 °C (**a**), *daf-16*(*mu86*) (red) and wild type (black) at 25 °C (**b**), and *hsf-1*(*sy441*) (red) and wild type (black) at 20 °C (**c**). **d–f**, The AFT residuals corresponding to the data in **a–c** respectively. Survival curves are shown for *eat-2*(*ad1116*) (red) and wild type (black) at 20 °C (**g**), *nuo-6*(*qm200*) (red) and wild type (black) at 25 °C (**h**), and *nuo-6*(*qm200*) populations held at 20 °C and 25 °C (**i**). **j–l**, The AFT residuals corresponding to the data in **g–i** respectively.

in regime I were not altered by elimination of DAF-16. In contrast, the hypomorphic alleles *daf-2*(*e1368*) and *age-1*(*hx546*) exhibit clear temperature-dependent effects across regime I (Fig. 4c, d). Both genes influence lifespan at 20 °C and 35 °C primarily by suppressing *daf-16* activity[22], which itself appears independent of temperature. Thus, *daf-2*(*e1368*) and *age-1*(*hx546*) alleles appear to be neomorphic in respect of the temperature dependence of their regulation of DAF-16.

We found that tBuOOH decreased lifespan at concentrations above 750 μM, with λ decreasing as a power law (Fig. 4e and Methods). This suggests an overall mass-action kinetics for the chain of events linking the direct targets of tBuOOH to the rescaling of the lifespan

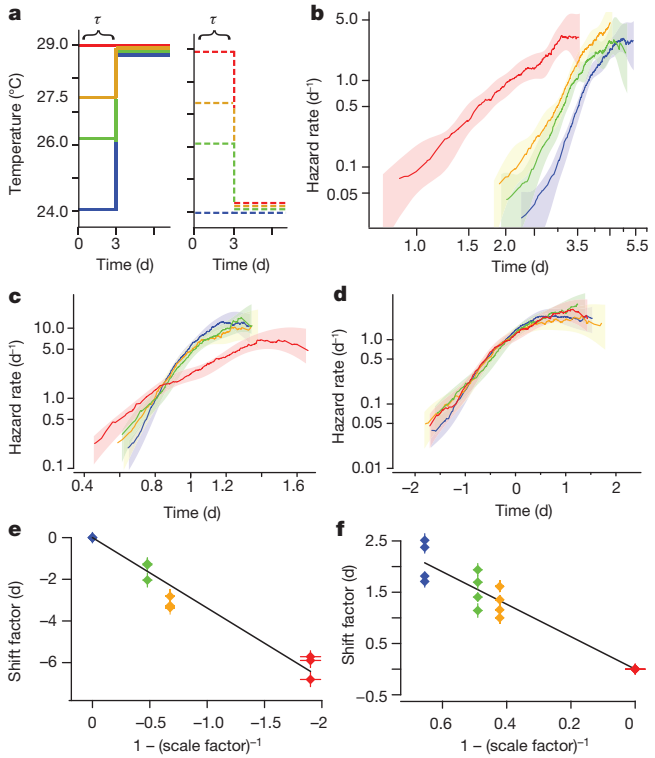


Figure 3 | Transient interventions during early adulthood shift the lifespan distribution. a, A schematic: populations were placed at 24 °C (blue), 26 °C (green), 27.5 °C (orange), and 29 °C (red). After $\tau = 3.2$ days, sub-populations were transferred to either 24 °C or 29 °C for the remainder of their lives. **b**, The hazard rate was estimated using the remaining lifespan of populations transferred to the final temperature of 29 °C. **c**, To test for temporal scaling between the populations shown in **b**, death times were fitted with the regression model $\log(y_i) = \beta x_i + \epsilon_i$, in which $\exp(\beta_i)$ is the best estimate for the scale factor λ. The residuals $\exp(\epsilon_i)$ are plotted as hazard functions in the colour scheme of **a**. **d**, To test for temporal shifts between the populations shown in **b**, death times were fitted with the regression model $y_i = \beta x_i + \epsilon_i$, in which β_i is the best estimate for the shift term Δ_τ. The residuals ϵ_i are plotted as hazard functions in the colour scheme of **a**. **e**, The shift term Δ_τ for populations transferred from each high temperature to 24 °C was plotted against $1 - \lambda^{-1}$, where λ is the scale factor relating populations always held at the corresponding high temperature to those always held at 24 °C. The prediction $\Delta_\tau = \tau(1 - \lambda^{-1})$ suggests that these points should fall along a line with a slope equal to τ in **a**. A linear regression on these points model estimates $\tau = 3.38 \pm 0.17$. **f**, As in **e**, but for populations transferred from lower initial temperatures to the final higher temperature of 29 °C, producing the estimate $\tau = 3.16 \pm 0.14$.

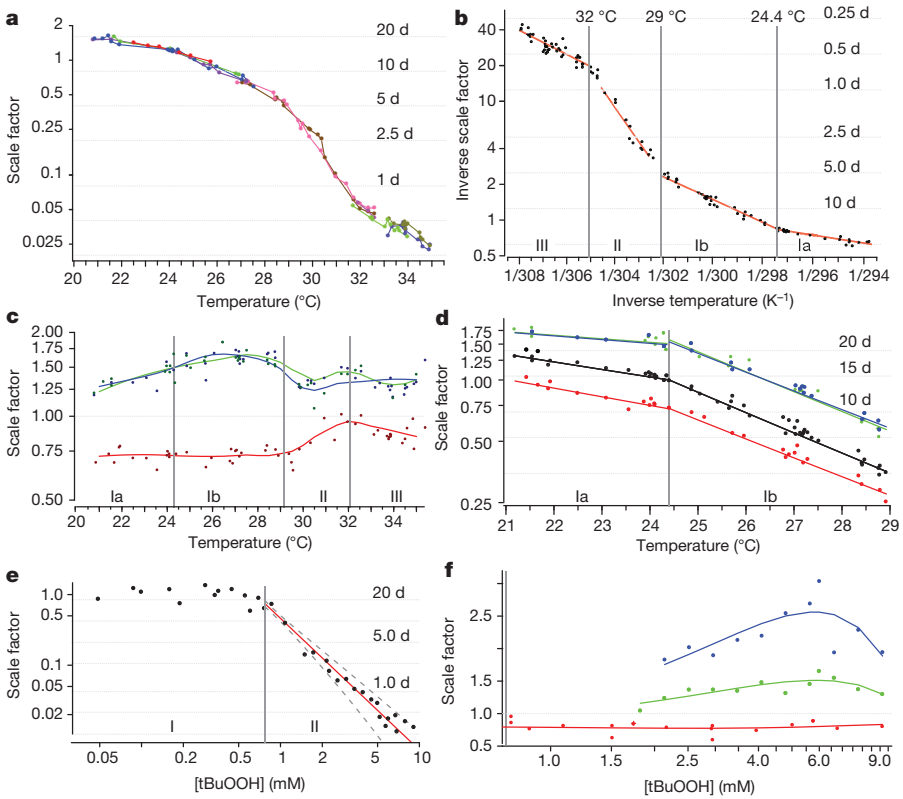


Figure 4 | Scaling functions. a, The magnitude of temporal scaling was estimated for wild-type populations held at fractional degree intervals across the range 20–35 °C. The scale factor λ of each population was estimated relative to a reference population at 25 °C. Grey lines mark the average lifespan of the reference population scaled by λ. Each replicate is shown as a separate colour, with each point corresponding to an aggregate population consisting of on average 130 individuals at the outset. **b**, The scale factor λ was determined for populations across the temperature range of **a**. The data points were fitted with a segmented Arrhenius model $\lambda(T)^{-1} = p_0 \exp(-p_1/RT)$ (red). **c**, The magnitude of scaling produced by *daf-16*(*mu86*) (red), *daf-2*(*e1368*) (green), and *age-1*(*hx546*) (blue) alleles relative to wild type was estimated at each temperature considered (points). Solid curves represent trends across temperature as fitted by a Loess regression. **d**, The combined magnitude of scaling produced by each allele and change of temperature was estimated relative to a single wild-type population 24 °C; colours as in **c**. Regimes II and III are shown in Extended Data Fig. 7. **e**, Wild-type populations at 20 °C were exposed to a series of tBuOOH concentrations ranging from 0 to 10 mM. For each population, λ was calculated relative to an unexposed population (0 mM). Data for concentrations above 0.75 mM were fitted by the model $\lambda([\text{tBuOOH}]) = p_2([\text{tBuOOH}])^{p_3}$ (red), yielding $p_2 = 0.47 \pm 0.02$ and $p_3 = -1.86 \pm 0.15$. **f**, As in **c**, but for the tBuOOH dosage series.

distribution. The distinct scaling functions of tBuOOH (power law) and temperature (multiple Arrhenius regimes) further suggest distinct molecular targets and mechanisms through which each type of intervention rescales the lifespan distribution.

As with temperature, the elimination of DAF-16 in the presence of tBuOOH reduced lifespan by a constant amount (Fig. 4f), 19.5 ± 8.8%, across all concentrations tested. Taken together with our temperature data in Fig. 4c, these results suggest that DAF-16 acts antagonistically but in parallel to the mechanisms through which tBuOOH and temperature shorten lifespan. DAF-16, tBuOOH, and temperature appear to affect ageing through their influence on risk determinants downstream of all three. For example, DAF-16 might attenuate or mitigate certain types of error or damage regardless of how the errors are created. The magnitude of temporal scaling produced both by *daf-2*(*e1368*) and by *age-1*(*hx546*) alleles varied across tBuOOH concentrations (Fig. 4g), which seems yet another aspect of a quantitative stress-dependent regulation of DAF-16 present in these strains but absent in wild type.

Disruption of *daf-2*, *daf-16*, *hif-1* or *hsf-1* produces distinct metabolic, cell-biological, and behavioural effects[15,23], as do changes in diet[24], temperature[25], and exposure to tBuOOH[26]. Yet, temporal scaling arises independently of the molecular targets specific to each intervention and requires that all risk determinants be affected to the same extent. This suggests that ageing in *C. elegans* can be described in terms of a whole-organism state variable r that completely determines all-cause mortality (Extended Data Fig. 9). State variables familiar from other contexts include temperature, pressure, and entropy, all of which describe the behaviour of a system resulting from the collective action of its many constituent elements without reference to their nature. In the same way, the change of the state over time, $r(t)$, describes the ageing process of *C. elegans* in terms of a collective action of all physiological determinants of risk. Where multiple risk determinants independently influence lifespan, temporal scaling requires that interventions simultaneously rescale, to an identical extent throughout life, the risk functions associated with each determinant (Supplementary Note 5.1). In models including dependencies among risk determinants, temporal scaling can emerge even when interventions act differentially across risk determinants (Supplementary Notes 5.2 and 5.3): dependencies can propagate the influence of interventions from one to all risk determinants, in effect producing a system-wide property that we call $r(t)$.

The temporal scaling of lifespan distributions constrains the dynamics of the state variable $r(t)$: the single stochastic process determining *C. elegans* lifespan must be invariant to timescale transformations and follow an average dynamics governed by an effective rate constant: $dr/dt = -k_r F(r)$, where $F(r)$ is an unknown function of r that does not depend on k_r. In this formulation, temporal scaling arises when interventions change k_r into k_r/λ. These dynamics place constraints on any stochastic process proposed to describe organismal ageing, as its parameters must change in a coordinated fashion. For example, if $r(t)$ were described by a biased random walk[27], the drift coefficient and the square of the diffusion coefficient must remain in a fixed proportion under intervention (Supplementary Note 6).

The idea that ageing is driven by changes in an organismal physiological state has been variously framed in terms of notions such as organization, vitality, organ reserve or resilience[3,28,29]. The temporal scaling across interventions justifies this notion, allowing an initial formalization. We note that any aspects of *C. elegans* physiology that change over time but do not influence lifespan, influencing 'quality' rather than 'quantity' of life, need not change in concert with $r(t)$.

We know neither the physiological basis of the state $r(t)$ nor the specific dynamics by which it changes with age. Yet, we can expect a broad set of lifespan determinants to affect only k_r, including minimally all determinants that influence lifespan exclusively through DAF-16 (refs 14 and 30), HSF-1 or HIF-1, or through the mechanisms that mediate the effects of temperature and tBuOOH on lifespan. If most ageing mechanisms currently studied influence only k_r, then future studies directed at clarifying the physiological origins of r and its dynamics should identify novel ageing mechanisms $F(r)$.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 July; accepted 18 December 2015.
Published online 27 January 2016.

1. Pincus, Z., Smith-Vikos, T. & Slack, F. J. MicroRNA predictors of longevity in *Caenorhabditis elegans*. *PLoS Genet.* **7,** e1002306 (2011).
2. Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **419,** 808–814 (2002).
3. Shaw, R. F. & Bercaw, B. L. Temperature and life-span in poikilothermous animals. *Nature* **196,** 454–457 (1962).
4. Mair, W., Goymer, P., Pletcher, S. D. & Partridge, L. Demography of dietary restriction and death in *Drosophila*. *Science* **301,** 1731–1733 (2003).
5. Wu, D., Rea, S. L., Cypser, J. R. & Johnson, T. E. Mortality shifts in *Caenorhabditis elegans*: remembrance of conditions past. *Aging Cell* **8,** 666–675 (2009).
6. Conti, B. *et al.* Transgenic mice with a reduced core body temperature have an increased life span. *Science* **314,** 825–828 (2006).
7. Lithgow, G. J., White, T. M., Melov, S. & Johnson, T. E. Thermotolerance and extended life-span conferred by single-gene mutations and induced by thermal stress. *Proc. Natl Acad. Sci. USA* **92,** 7540–7544 (1995).
8. Stroustrup, N. *et al.* The *Caenorhabditis elegans* lifespan machine. *Nature Methods* **10,** 665–670 (2013).
9. Johnson, T. E., Wu, D., Tedesco, P., Dames, S. & Vaupel, J. W. Age-specific demographic profiles of longevity mutants in *Caenorhabditis elegans* show segmental effects. *J. Gerontol. A* **56,** B331–B339 (2001).
10. Swindell, W. R. Accelerated failure time models provide a useful statistical framework for aging research. *Exp. Gerontol.* **44,** 190–200 (2009).
11. Vaupel, J. W., Manton, K. G. & Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16,** 439–454 (1979).
12. Martin, G. M., Austad, S. N. & Johnson, T. E. Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nature Genet.* **13,** 25–34 (1996).
13. Tullet, J. M. *et al.* Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in *C. elegans*. *Cell* **132,** 1025–1038 (2008).
14. Landis, J. N. & Murphy, C. T. Integration of diverse inputs in the regulation of *Caenorhabditis elegans* DAF-16/FOXO. *Dev. Dyn.* **239,** 1405–1412 (2010).
15. Hsu, A. L., Murphy, C. T. & Kenyon, C. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* **300,** 1142–1145 (2003).
16. Leiser, S. F., Begun, A. & Kaeberlein, M. HIF-1 modulates longevity and healthspan in a temperature-dependent manner. *Aging Cell* **10,** 318–326 (2011).
17. Kirkwood, T. B. & Shanley, D. P. Food restriction, evolution and ageing. *Mech. Ageing Dev.* **126,** 1011–1016 (2005).
18. Garigan, D. *et al.* Genetic analysis of tissue aging in *Caenorhabditis elegans*: a role for heat-shock factor and bacterial proliferation. *Genetics* **161,** 1101–1112 (2002).
19. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sc. USA* **95,** 13091–13096 (1998).
20. Suda, H., Sato, K. & Yanase, S. Timing mechanism and effective activation energy concerned with aging and lifespan in the long-lived and thermosensory mutants of *Caenorhabditis elegans*. *Mech. Ageing Dev.* **133,** 600–610 (2012).
21. Atlan, H., Miquel, J., Helmle, L. C. & Dolkas, C. B. Thermodynamics of aging in *Drosophila melanogaster*. *Mech. Ageing Dev.* **5,** 371–387 (1976).
22. Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115,** 489–502 (2003).
23. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424,** 277–283 (2003).
24. Walker, G., Houthoofd, K., Vanfleteren, J. R. & Gems, D. Dietary restriction in *C. elegans*: from rate-of-living effects to nutrient sensing pathways. *Mech. Ageing Dev.* **126,** 929–937 (2005).
25. McColl, G. *et al.* Insulin-like signaling determines survival during stress via posttranscriptional mechanisms in *C. elegans*. *Cell Metab.* **12,** 260–272 (2010).
26. Oliveira, R. P. *et al.* Condition-adapted stress and longevity gene regulation by *Caenorhabditis elegans* SKN-1/Nrf. *Aging Cell* **8,** 524–541 (2009).
27. Aalen, O. O. & Gjessing, H. K. Understanding the shape of the hazard rate: a process point of view. *Stat. Sci.* **16,** 1–14 (2001).
28. Medawar, P. B. *An Unsolved Problem of Biology: An Inaugural Lecture Delivered at University College, London* (H. K. Lewis, 1951).
29. Gladyshev, V. N. The origin of aging: imperfectness-driven non-random damage defines the aging process and control of lifespan. *Trends Genet.* **29,** 506–512 (2013).
30. Samuelson, A. V., Carr, C. E. & Ruvkun, G. Gene activities that mediate increased life span of *C. elegans* insulin-like signaling mutants. *Genes Dev.* **21,** 2976–2994 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Alcedo for nematode strains, X. Manière for providing the NEC937 *Escherichia coli* strain, B. Ward for reading our manuscript, and D. Marks, C. Romero, T. Kolokotrones, D. Yamins, P. F. Stadler, E. Smith, and all members of the Fontana laboratory for discussions and encouragement throughout this project. Some strains were provided by the Caenorhabditis Genetics Center, which is funded by US National Institutes of Health (NIH) Office of Research Infrastructure Programs (P40 OD010440). This work was funded by the NIH through grant R01 AG034994 and by a Glenn Award from the Glenn Foundation for Medical Research.

Author Contributions N.S. conceived and analysed the experiments. N.S., J.A., W.E.A., V.G., A.G., Z.M.N., and I.F.L.-M. developed experimental protocols and performed experiments. N.S. and W.F. interpreted data and performed model calculations. N.S. and W.F. wrote the manuscript with input from J.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S. (nstroustrup@post.harvard.edu) or W.F. (walter@hms.harvard.edu).

METHODS

Experimental methods. The following nematode strains were used: QZ0 (wild type (Bristol)), TJ1052 (*age-1*(*hx546*)*II*), and QZ120 (*daf-2* (*e1368*)), QZ60 (*daf-16* (*mu86*)), QZ121 (*hsf-1* (*sy441*)), QZ414 (*eat-2* (*ad1116*)), ZG31 (*hif-1* (*ia4*)), and MQ1333 (nuo-6(qm200)).

Hermaphrodites were cultured under standard conditions[8,31], at either 20 or 25 °C as noted, on plates containing 100 μg ml^{-1} ampicillin and seeded at an absorbance at 600 nm of 20 with the *E. coli* strain NEC937 B (OP50 ΔuvrA; KanR)[32]. Before seeding, all bacteria were irradiated in batch culture with 4 J m^{-2} of 254-nm light in a UV Stratalinker (Stratagene). Age-synchronous cohorts were prepared by hypochlorite treatment[33]. In strains exhibiting an increased variation in developmental time, in particular *eat-2*(*ad1116*) and *nuo-6*(*qm200*), developmentally synchronized L4 larvae were manually selected at the final stage of vulval maturation. To control for any effects of temperature on *C. elegans* development, all populations were held at 20 °C until their second day of adulthood, at which time they were exposed to the interventions described (unless otherwise stated). Individuals were randomly assigned groups by obtaining aliquots from populations suspended in M9 buffer. Automated lifespan experiments were run, blinded, and validated according to published methods[8].

All populations were transferred on their second day of adulthood onto plates containing 22.5 μg ml^{-1} nystatin (Sigma N3503) to prevent fungal growth and 27.5 μg ml^{-1} 5-fluoro-2-deoxyuridine (FUDR, Sigma) to eliminate live progeny. Where live bacteria were used (Fig. 1h, i), 10 μg ml^{-1} FUDR sufficed to eliminate live progeny.

Standard NGM agar plates were poured and dried according to published methods[8]. For tBuOOH assays, the compound was added to molten agar immediately before pouring. All plates containing tBuOOH were seeded and placed in a fume hood until the bacteria was absorbed, approximately 1 h. We found no evidence of time-dependent degradation of tBuOOH in solid agar plates, nor any effect of the tBuOOH degradation byproduct tbutanol on *C. elegans* lifespan (Extended Data Fig. 4), although evaporation of tBuOOH needed to be strictly controlled.

Scanner temperatures were measured using thermocouples (ThermoWorks USB-REF) mounted on the bottom of empty Petri dishes loaded onto each scanner. Scanners were given several hours to reach their stable operating temperature, at which point measurements were taken every 10 s. Because operating scanners exhibit small, regular oscillations in temperature[8], the average of multiple cycles, each lasting 20 min, was taken.

Statistical methods. *Population sizes*. Supplementary Table 2 shows various summary statistics, including population size, for the population in each figure panel. No statistical methods were used to predetermine sample size. Initial experiments showed that populations larger than 500 individuals provided more robust survival estimates. Statistical power was estimated retrospectively (Supplementary Notes 2.2 and 6.3). Animals lost from observation were censored according to published methods[8].

Replicates. All replicates described are biological replicates, performed in separate weeks using separate populations. All replicates performed in the laboratory yielding informative population sizes and meeting self-consistency standards (for example consistent results across multiple scanners[8]) are shown.

Single AFT regression (all figures). Using our automated microscopy method[8], a single, very large, homogeneous population must be distributed across multiple scanners. The local environment characteristic to each scanner can influence *C. elegans* lifespan. This influence can be measured and controlled using the same AFT model we use to quantify temporal scaling. In most cases, our AFT models take the form

$$\log(y_i) = \beta X_i + \epsilon_i \quad (1)$$

where y_i is the lifespan of individual i, β is the parameter vector being estimated, and X_i is the categorical covariate coding for the label associated with each individual i. This label takes different values in different contexts, representing either the plate name (Figs 1, 2, 3b, d and 4b–f), scanner name (Fig. 4a) or animal genotype (Fig. 4c, f). We take the logarithm of lifespan, $\log(y_i)$, following the standard approach for evaluating covariates that act multiplicatively, producing fold-changes in lifespan.

Each AFT regression model has an intercept that determines the 'reference' lifespan in relation to which all parameter vectors β are scaled. The intercept is implicitly determined by the particular categorical encoding scheme used for X_i. In Figs 1, 2 and 3b–d a 'deviation' coding is used, placing the intercept at the grand mean lifespan of all individuals. In Fig. 4, a dummy coding is used, placing the intercept at the mean lifespan of a specific reference group whose identity is noted in each case. The coding scheme used for the categorical variable has no effect on the value of model residuals ϵ_i and determines only the values of the AFT parameters β.

All AFT regression models were estimated via Buckley–James regression (included in the R[34] package RMS). In Figs 1a and 3b and Extended Data Fig. 5a, j, l, n, p, r, t, v 'device-corrected death times'[8] were calculated to remove obscuring effects of temperature variation between scanners on the aggregate hazard rate. A dummy encoding for X_i coded for the scanner on which each individual was observed. The model intercept was then added to the residuals ϵ_i.

In many panels, AFT residuals ϵ_i are grouped and plotted according to some experimental condition: by temperature in Figs 1c, d, e and 2l and Extended Data Fig. 4i; by tBuOOH concentration in Fig. 1g and Extended Data Fig. 4a; by bacterial treatment in Fig. 1i; and by genotype in Fig. 2d, e, f, j, k and Extended Data Fig. 4e, g, k, m, o, q, s, u, w. We emphasize that in each case, the experimental condition used to group plots is not a covariate in the AFT model. Instead, the single AFT categorical covariate codes for an experimental unit—plate or scanner name—as detailed above. In this way, any temporal rescaling produced by variation in environmental conditions across these units will be estimated and reflected in the parameter vector β, and its effects minimized on the residual times ϵ_i. These residuals are then grouped according to the experimental variable during subsequent analysis (Supplementary Note 2).

Single AFT regression for additive models (Fig. 3d–f, h and Extended Data Fig. 5b, e, l). To account for temporal shifts in mortality statistics, Buckley–James regression (using the R[34] package RMS) was used to fit the model

$$y_i = \beta X_i + \epsilon_i \quad (2)$$

where X_i is the categorical covariate the initial (Fig. 3d–f and Extended Data Fig. 5b) temperature at which individual i was placed, the duration of time spent at that temperature (Extended Data Fig. 5d–g), or the number of switches used (Extended Data Fig. 5l). A 'deviation' coding was used (see equation (1)) in Fig. 3d and a dummy coding was used for Fig. 3e, f.

Significance (Figs 1 and 2). The probability (P value) that the observed differences in lifespan between two populations are explained entirely by temporal scaling was estimated using the two-sided modified Kolmogorov–Smirnov test to identify heteroscedasticity among AFT residuals. This approach is described, with additional power analysis, in Supplementary Note 2.

Hierarchical clustering of survival curves (Extended Data Fig. 3). A hierarchical clustering of survival curves was computed to identify groups of curves temporally rescaled in respect to each other. All pairs of populations were compared using the modified Kolmogorov–Smirnov test as described in Supplementary Note 2. The modified Kolmogorov–Smirnov distance (Supplementary Note 2) was used as the distance metric for clustering, using the R[34] hierarchical clustering implementation hclust.

Estimating hazard functions (Figs 1–3). The time-dependent hazard rate was estimated through numerical differentiation of the Kaplan–Meier cumulative hazard estimate. To generate confidence bands for the true hazard rate under the assumption that it is locally smooth, death times were fitted with a piecewise-polynomial B-spline hazard model using the R[34] package bshazard[35].

Estimating the magnitude of temporal scaling across temperatures (Fig. 4a, b). Buckley–James regression was used to fit the AFT model described in equation (1), with X_i as a categorical covariate coding the scanner name corresponding to each individual tested. By specifying a single scanner at 25 °C as the reference category of a reference coded categorical variable, the AFT parameter vector β becomes the best estimate of the scale factors relating lifespan on each scanner X_i to the 25 °C reference: $S_X(t) = S_{25\,°C}(\lambda^{-1}t)$.

Because the average temperature was measured for each scanner X_i, the corresponding λ can therefore be plotted in Fig. 4a, b as a function of the temperature on the scanner that produced it.

Identifying distinct scaling regimes (Fig. 4b and Extended Data Figs 6 and 7a, b). Because the temperature scaling of timescale λ did not appear uniform, we applied a linear segmented regression model using transformed variables to identify the number scaling regimes and estimate the boundaries between them. An Arrhenius model was fitted, assuming that within each segment

$$\lambda(T)^{-1} = Ae^{-B/T} \quad (3)$$

with $\lambda(T)$ as the scale factor at temperature T, and A and B as pre-exponential and exponential Arrhenius constants, respectively. One or more breakpoints were incorporated at temperatures $T = \{T_1, ..., T_{n-1}\}$ to produce the segmented model

$$\log(\lambda(T)^{-1}) = A_i - \frac{B_i}{T},\ T_{i-1} < T < T_i, i = 1, \ldots, n \quad (4)$$

with T_0 and T_n the fixed starting and end temperatures, respectively. A linear model was tested in the same fashion:

$$\lambda(T)^{-1} = A_i + B_i T,\ T_{i-1} < T < T_i,\ i = 1, \ldots, n \quad (5)$$

Model parameters were estimated using the R[34] package 'segmented'.

To generate best estimates for the B parameter within each Arrhenius scaling regime identified by segmented regression, death times within each regime were isolated and fitted using a nonlinear regression approach (R^{34} package nls2). In all segmented and nonlinear regression models performed on λ values, each λ was weighted according to the size of the population used in the AFT regression from which the λ was estimated. Additional estimates of Arrhenius parameters were obtained following the approach described in equation (6).

Estimating the effect of mutations relative to wild type, across temperatures (Fig. 4c). To estimate the effect of mutant alleles on lifespan relative to wild type at each temperature, a separate AFT model as specified in equation (1) was run at each temperature considered. In each regression, X_i was specified as the genotype of individuals, with the categorical variable coded to use wild-type populations as the reference.

Characterizing the temperature dependence of mutant lifespan across temperatures, relative to wild type (Fig. 4d and Extended Data Fig. 7c). To identify differences in the temperature dependence of lifespan of mutants relative to wild type, we considered each thermal regime separately (regimes Ia and Ib in Fig. 4d and regimes II and III in Extended Data Fig. 7). We considered the model

$$\log(y_i) = \beta_g X_i + \beta_r Y_i + \beta_T T_i + \beta_x X_i T_i + \epsilon_i \quad (6)$$

where X_i is a categorical variable encoding for the genotype of the individual, Y_i is the biological replicate in which the individual was observed (animals originating from the same hypochlorite treatment in the same week have the same Y_i), and T_i is a continuous variable representing the temperature (in inverse degrees Kelvin) at which the individual was placed. The interaction term $\beta_x X_i T_i$ captures any differential effect of temperature on the mutant strain relative to wild type. The null hypothesis, therefore, is that mutant individuals exhibit no difference in their response to temperature compared with wild-type individuals, which can be rejected if a significant non-zero value of β_x is observed, as calculated from the associated Wald Z score. The estimated model parameters are provided in Supplementary Table 3.

Multiple AFT regression across tBuOOH replicates (Fig. 4e). The absolute effect of tBuOOH on lifespan varies between replicates. Several factors may contribute to this, including a variation in temperature of the molten agar and the time spent pouring it, and a variation in duration time required to dry plates. The relative effect of different tBuOOH concentrations on lifespan within each replicate, however, appeared more robust. So, we estimated the relative effects of tBuOOH on lifespan using the multiple regression model

$$\log(y_i) = \beta_c X_i + \beta_r Y_i + \epsilon_j \quad (7)$$

with X_i as a categorical covariate coding for tBuOOH concentration, and Y_i as a categorical covariate coding for the biological replicate. The tBuOOH covariate was coded such that the parameters β_c were relative to the 0 mM category. The replicate name covariate was coded such that the parameters β_r were relative to a single reference replicate. In this way, the parameters β_c represent the best estimate of the relative effect of tBuOOH across replicates.

To determine the quantitative dependence of the scale factor on tBuOOH concentration, the values of β_c and the tBuOOH concentrations corresponding to each X_i were fitted by the polynomial model $\lambda = b[\text{tBuOOH}]^a$ using the R^{34} nonlinear regression package nls2.

A distinct statistical approach was attempted to confirm the above analysis in a different way. All death times were fitted with the single model

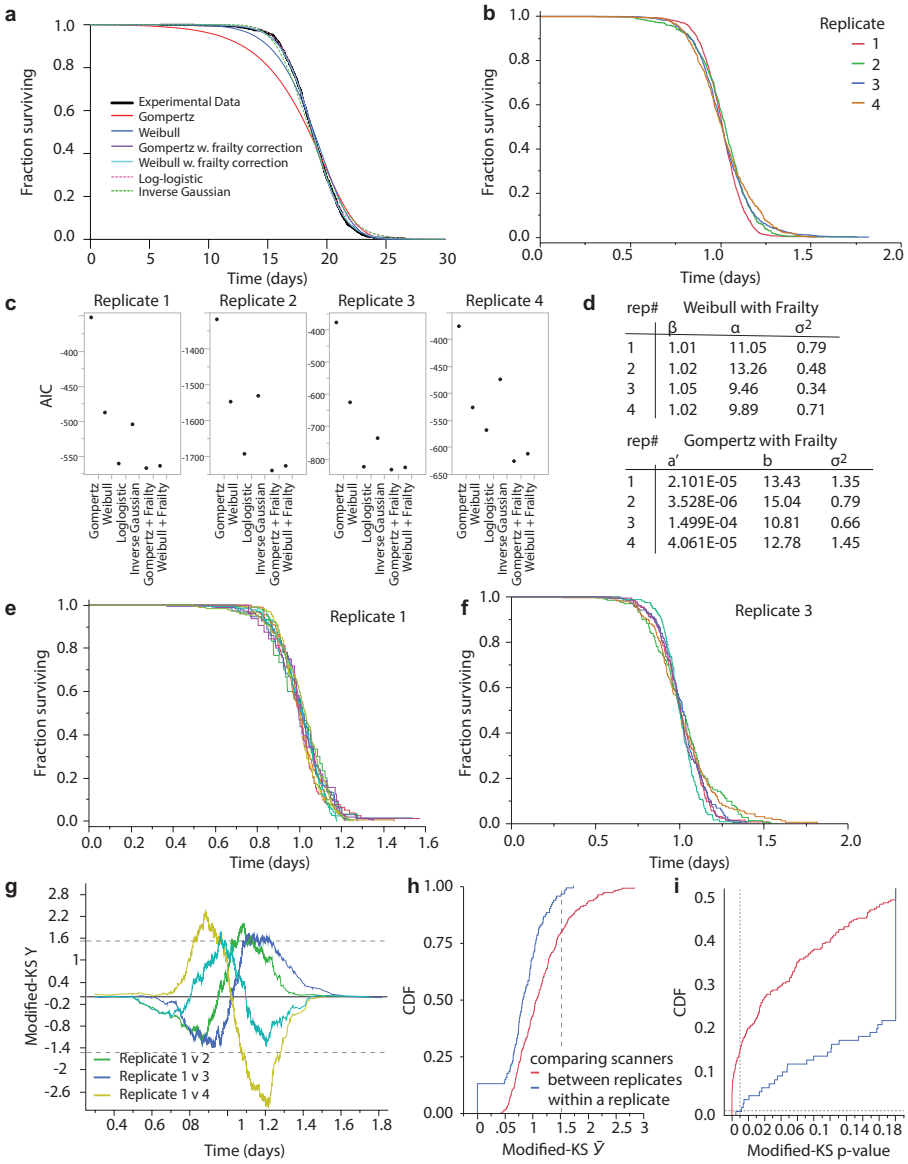
$$\log(y_i) = \beta_c B_i + \beta_r Y_i + \beta_x Y_i B_i + \epsilon_i \quad (8)$$

where B_i is the logarithm of the tBuOOH concentration to which each individual i was exposed, represented as a continuous covariate. This makes β_c the best estimate of the exponent of the power-law relationship between lifespan and tBuOOH concentration. Y_i is the biological replicate in which the individual was observed (animals originating from the same hypochlorite treatment in the same week have the same Y_i), and $Y_i B_i$ is a cross term to identify any systematic differences between the effect of tBuOOH between replicates. The estimated model parameters are provided in Supplementary Table 3.

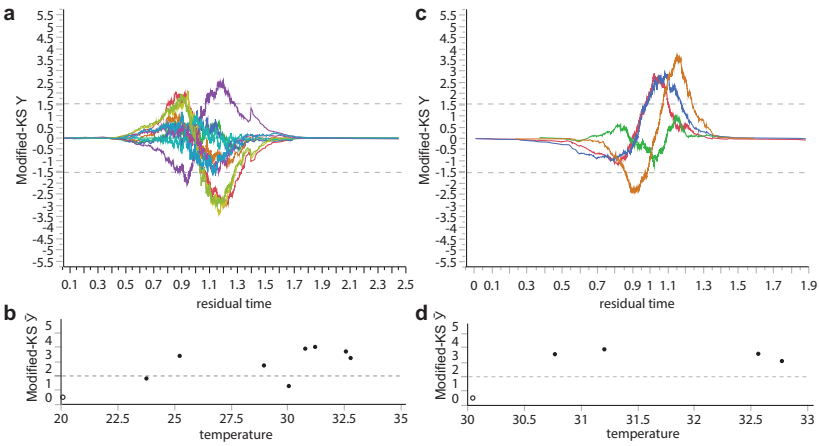
Single AFT regression on each tBuOOH replicate (Fig. 4e). To validate the multiple regression model described previously, the AFT model described in equation (1) was fitted separately on the data collected in each replicate, with X_i coded to correspond to the tBuOOH concentration. The dummy variables for the tBuOOH covariate were set up such that scale factors were relative to the 0 mM control group. Using these cofactors, the R^{34} nonlinear regression package nls2 was used to fit the polynomial relationship $\lambda = b[\text{tBuOOH}]^a$ for each replicate.

Estimating the effect of genotype relative to wild type, across tBuOOH concentrations (Fig. 4f). To measure the effect of mutant alleles on lifespan across t-BuOOH concentrations, a separate AFT model as specified in equation (1) was run on each data set collected at each concentration. In each regression, X_i was specified as the genotype of individuals, with the categorical variable coded to use wild-type populations as the model intercept.

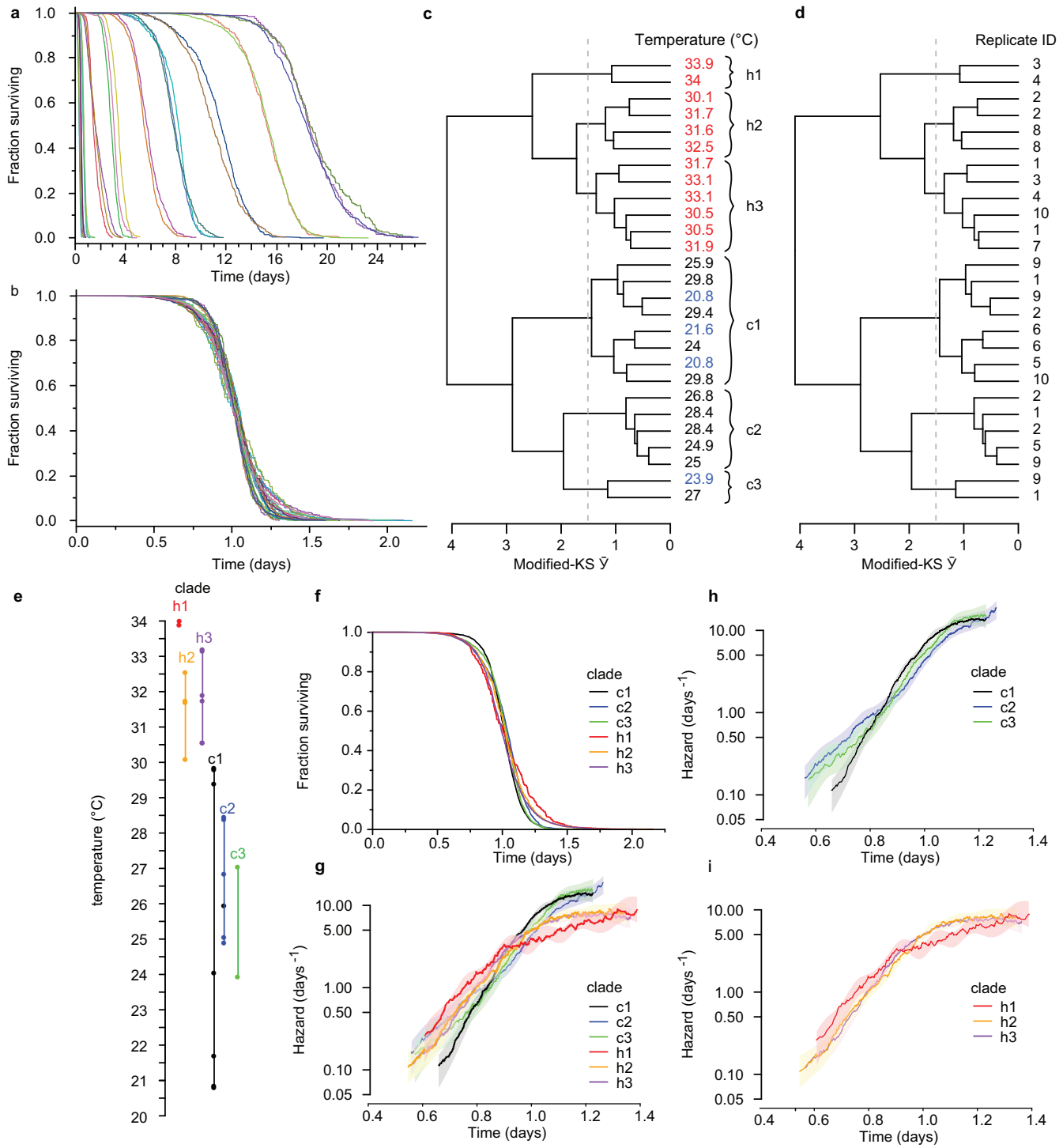
31. Stiernagle, T. in *WormBook* (ed. The *C. elegans* Research Community) http://dx.doi.org/10.1895/wormbook.1.101.1 (February 11, 2006).
32. Baeriswyl, S. *et al.* Modulation of aging profiles in isogenic populations of *Caenorhabditis elegans* by bacteria causing different extrinsic mortality rates. *Biogerontology* **11,** 53–65 (2010).
33. Wilkinson, D. S., Taylor, R. C. & Dillin, A. Analysis of aging in *Caenorhabditis elegans*. *Methods Cell Biol.* **107,** 353–381 (2012).
34. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2013).
35. Rebora, P., Salim, A. & Reilly, M. bshazard: a flexible tool for nonparametric smoothing of the hazard function. *R J.* **6,** 114–122 (2014).



Extended Data Figure 1 | Characterizing the shape of wild-type lifespan distributions at 20 °C. a, The AFT residuals corresponding to the 20 °C wild-type population presented in Fig. 1 were fitted with a variety of parametric distributions (Supplementary Note 1.4). Fits made to AFT residuals, as opposed to absolute death times, are much less sensitive to any environmental heterogeneity existing between plates and scanners (statistical methods). **b**, The AFT residuals of four additional replicates at 20 °C to assess deviations from temporal scaling between replicates. **c**, The Akaike information criterion (AIC) was calculated for each parametric fit of each replicate's AFT residuals shown in **b**. Lower AIC values suggest preferred models. **d**, The parameters of Gompertz and Weibull distributions with frailty corrections are listed; both distributions were good fits across all replicates. **e**, **f**, The survival curves of populations collected in two biological replicates are shown, with one curve for individuals observed on each of 10 and 6 scanners, respectively. **g**, The modified Kolmogorov–Smirnov $Y(t)$ (Supplementary Note 2) is plotted for comparisons between replicate 1 and all others. Pairs for which $Y(t) > 1.51$ for some t exhibit statistically significant deviations from perfect scaling. In this case every replicate differed significantly from the first replicate. **h**, The distribution of modified Kolmogorov–Smirnov test scores, $\overline{Y}$, is plotted for comparisons between scanners within a replicate (blue) and between scanners in different replicates (red). Differences between replicates were larger than differences with replicates, suggesting that distance between survival curves observed between scanners cannot alone explain the distance between survival curves observed between replicates. **i**, The P values corresponding to each $\overline{Y}$ are shown, with values $P > 0.01$ considered statistically significant (grey line).

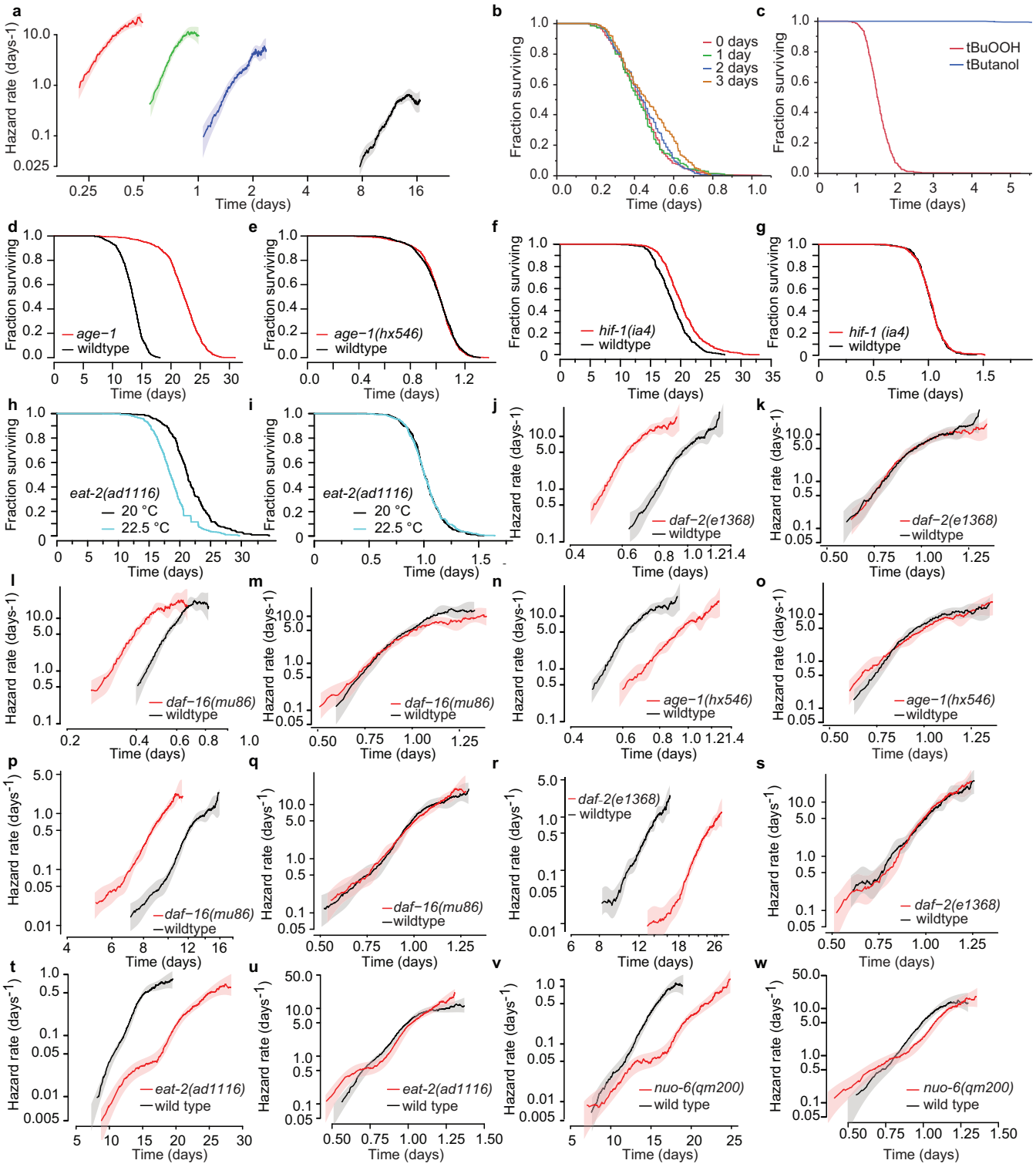


Extended Data Figure 2 | Apparent deviations from temporal scaling are observed when single replicates are performed at each temperature. **a**, For the data shown in Fig. 1a, the modified Kolmogorov–Smirnov score $Y(t)$ was calculated for AFT residuals (statistical methods and Supplementary Note 2) to compare the reference population at 20 °C with populations held at each of the other temperatures. Pairs for which $Y(t) > 1.51$ for some t (grey dashed line) exhibit significant deviations from perfect scaling. **b**, The modified Kolmogorov–Smirnov test scores $\overline{Y}$, corresponding to the maximum absolute value of $Y(t)$ observed at any time t, are shown for the comparisons in **a**, highlighting the statistical deviations observed between independent replicates performed at different temperatures. **c**, **d**, The same statistics were calculated when comparing all populations above 30 °C with the population at 30 °C.



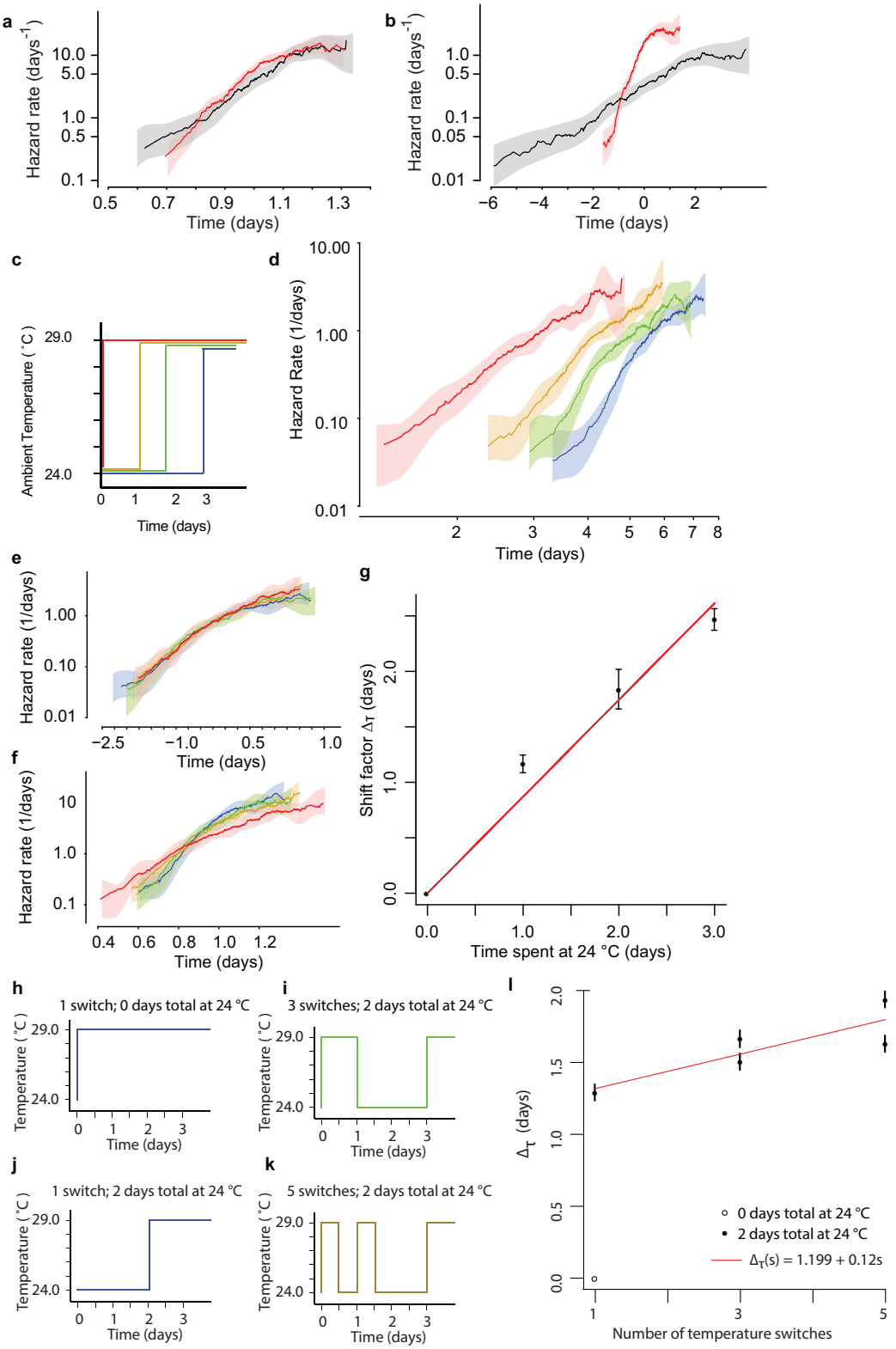
Extended Data Figure 3 | See next page for figure caption.

Extended Data Figure 3 | Independent replicates demonstrate that apparent deviations from temporal scaling within low- and high-temperature regimes arise from uncontrolled environmental variation. The lifespan of individuals from populations housed between 20 °C and 34 °C were collected using the lifespan machine (also shown in Fig. 4a, b). To characterize the effects of any uncontrolled experimental conditions specific to individual replicates, and identify any effects of temperature consistent across replicates, we divided the full temperature range into 2 °C intervals. Each 2 °C interval contained lifespan data collected in either two or three independent replicate experiments performed in separate weeks. **a**, Within each 2 °C range, all death times were fitted by an AFT regression model using plate name as a categorical covariate (Statistical methods). The device-corrected death times (the residual time plus model intercept[8]) were plotted, highlighting the changes in survival curve shape between replicates within each 2 °C range. **b**, All deaths across all temperatures were then fitted by a single AFT regression model with plate name as the categorical covariate. AFT residuals were grouped according to their replicate name and temperature range, and plotted to highlight the deviations from temporal scaling across all replicates at all temperatures. **c**, The modified Kolmogorov–Smirnov test (Supplementary Note 2) was applied on each pair of curves shown in **b**. The resulting Kolmogorov–Smirnov $\overline{Y}$ was used as a distance metric with which to perform a hierarchical clustering, shown as a dendrogram with each replicate population labelled by the temperature at which it lived. In this dendrogram, populations exhibiting smaller deviations from temporal scaling will have fewer branches between them. Clades that contain statistically significant deviations from temporal scaling have branches extending beyond the dashed grey line, indicating that $\overline{Y} > 1.51$ between branches. Six statistically distinct groups were identified, three above 30 °C and three below. **d**, The same dendrogram is shown with populations labelled according to the name of the replicate in which they were collected. Populations collected in single replicates did not fall into single clades. This suggests that some environmental factor variable within replicates, distinct from the particular temperature at which populations were placed, produced the observed deviations from temporal scaling. **e**, The statistically distinct clades identified by hierarchical clustering (**c**) are plotted on a temperature scale. Clades overlap at all temperatures except the 30 °C boundary, suggesting that only the 30 °C transition represents a true temperature-dependent deviation from temporal scaling. **f**, The aggregate survival curves containing the AFT residuals of all individuals in each statistically distinct clade are compared, to highlight the differences in shape between clades. **g**, The hazard rate plot of the AFT residuals of all individuals in each statistically distinct clade. **h**, Same as **g**, but showing only the hazard rate plots of populations kept at low temperature. **i**, Same as **g**, but showing only the hazard rate plots of populations kept at high temperature.



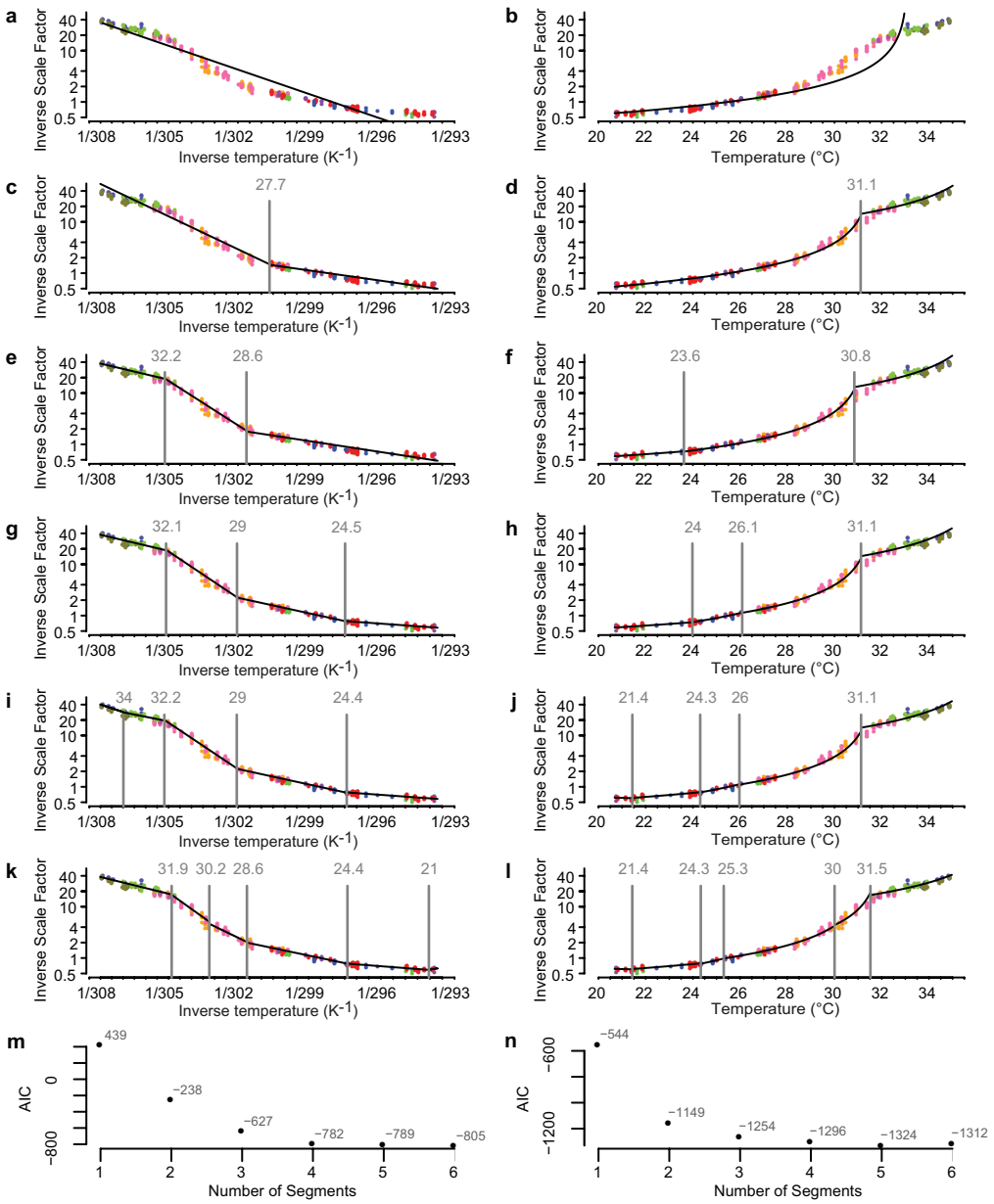
Extended Data Figure 4 | See next page for figure caption.

Extended Data Figure 4 | Additional survival curves and hazard plots. **a**, The hazard rates corresponding to the tBuOOH survival data presented in Fig. 1: 0 mM (black), 1.5 mM (blue), 3 mM (green), and 6 mM (red). **b**, To test for any effects of tBuOOH degradation and evaporation on lifespan, 9 mM tBuOOH plates were prepared and placed at 4 °C. On 4 consecutive days, a subset of plates were seeded with ultraviolet-inactivated bacteria and placed without *C. elegans* on scanners operating at 20 °C. In this way, four groups of plates were created, corresponding to 0, 1, 2, and 3 days of cumulative exposure to standard scanner conditions during which tBuOOH degradation and evaporation could potentially occur. A single age-synchronous population of 2-day-old adult *C. elegans* was then simultaneously distributed across all plates. The remaining lifespan of all worms at 20 °C was recorded using the lifespan machine. **c**, In a separate experiment, plates were prepared containing either 3 mM tBuOOH or 3 mM t-butanol, a degradation product of tBuOOH. On t-butanol, only a trivial fraction of individuals had died by the fifth day, so the experiment was terminated. **d**, **e**, The survival of wild-type and *age-1*(*hx546*) populations at 25 °C. **f**, **g**, The lifespan and AFT residuals for *hif-1*(*ia4*) and wild-type populations, calculated as in Fig. 2. **h**, **i**, The lifespan and AFT residuals for *eat-2*(*ad1116*) at 20 °and 22.5 °C. **j**–**o**, Age-synchronous mutant (red) and wild-type (black) populations were raised at 25 °C and then transferred to 33 °C on their second day of adulthood, where they remained until death. For each population at 33 °C, the hazard rate was estimated from the death times (**j**, **l**, **m**). The hazard rate was also estimated from the residuals of the AFT regression model $\log(y_i) = \beta x_i + \epsilon_i$ with genotype as a single categorical covariate (**k**, **m**, **o**). **p**, **q**, The hazard functions of death times and AFT residuals corresponding to the *daf-16*(*mu86*) data presented in Fig. 1. **r**, **s**, The hazard functions of death times and AFT residuals corresponding to the *daf-2*(*1368*) data presented in Fig. 1. **t**, **u**, The hazard functions of death times and AFT residuals corresponding to the *eat-2*(*ad1116*) data presented in Fig. 1. **v**, **w**, The hazard functions of death times and AFT residuals corresponding to the *nuo-6*(*qm200*) data presented in Fig. 1.

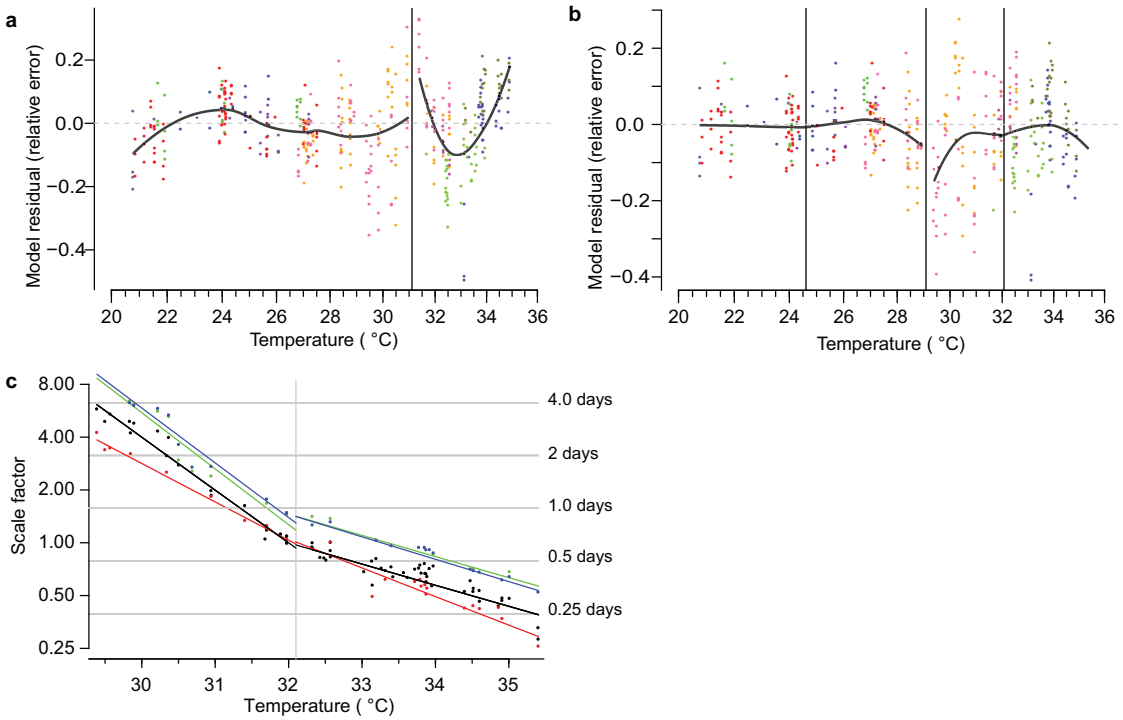


Extended Data Figure 5 | See next page for figure caption.

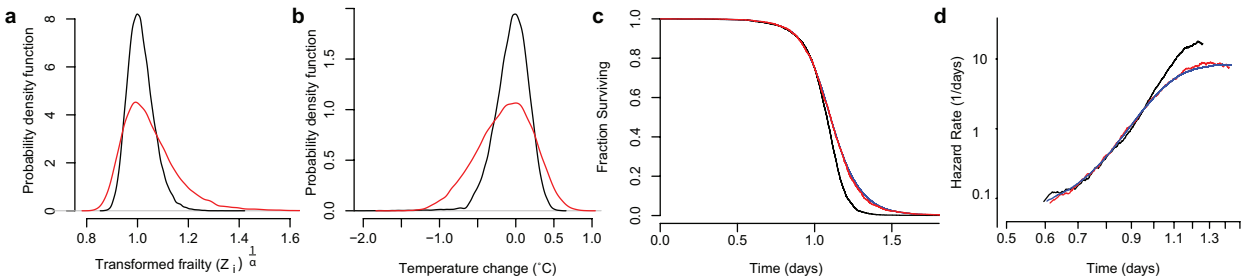
Extended Data Figure 5 | Additional temperature shift data. a, As a control for the temperature shift experiment shown in Fig. 3, the same regression as in Fig. 3c was run to test for temporal scaling between populations always held at 24 °C (black) and those always held at 29 °C (red). The residuals ϵ_i are plotted as hazard functions. **b**, The same regression as in Fig. 3 was run for the same populations as **b** here, to test for temporal shifts. **c**, To test for the effects of different durations spent at 24 °C before transfer to 29 °C. Age-synchronous, wild-type animals were grown at 20 °C and then transferred on their second day of adulthood to 24 °C. Subsets of these animals were then transferred to 29 °C on each of 3 consecutive days. **d**, For each population, the remaining lifespan was observed and the hazard functions estimated. All death times represent the number of days after the second day of adulthood. **e**, The residuals from a regression model with the duration at 24 °C as an additive categorical covariate $y_i = \beta x_i + \epsilon_i$. **f**, The residuals from a regression model with the duration at 24 °C as a proportional covariate $\log(y_i) = \beta x_i + \epsilon_i$. **g**, The shift values Δ_τ of the additive model are plotted along with a linear fit. **h**, To test for the effect of rapid temperature changes on lifespan, age-synchronous individuals were raised at 20 °C. On their second day of adulthood, a subset was transferred to 29 °C. **i**, Another subset of individuals remained at 24 °C for 2 days, after which they were transferred to 29 °C. **j**, Another subset was transferred to 29 °C, but switched down from 29 °C to 24 °C and then back again to 29 °C (filled circle; three shifts). **k**, A final subset was switched down and back twice (filled circle; five shifts). Note that all populations spent the same total duration at 29 °C, except for the aforementioned control population that was never switched. **l**, The data were fitted with an additive regression model $y_i = \beta x_i + \epsilon_i$ with the number of switchings as a categorical covariate. The encoding of this covariate was set so that all $\beta = \Delta_\tau$ represent each subpopulation's change in lifespan relative to the control population that was never switched.



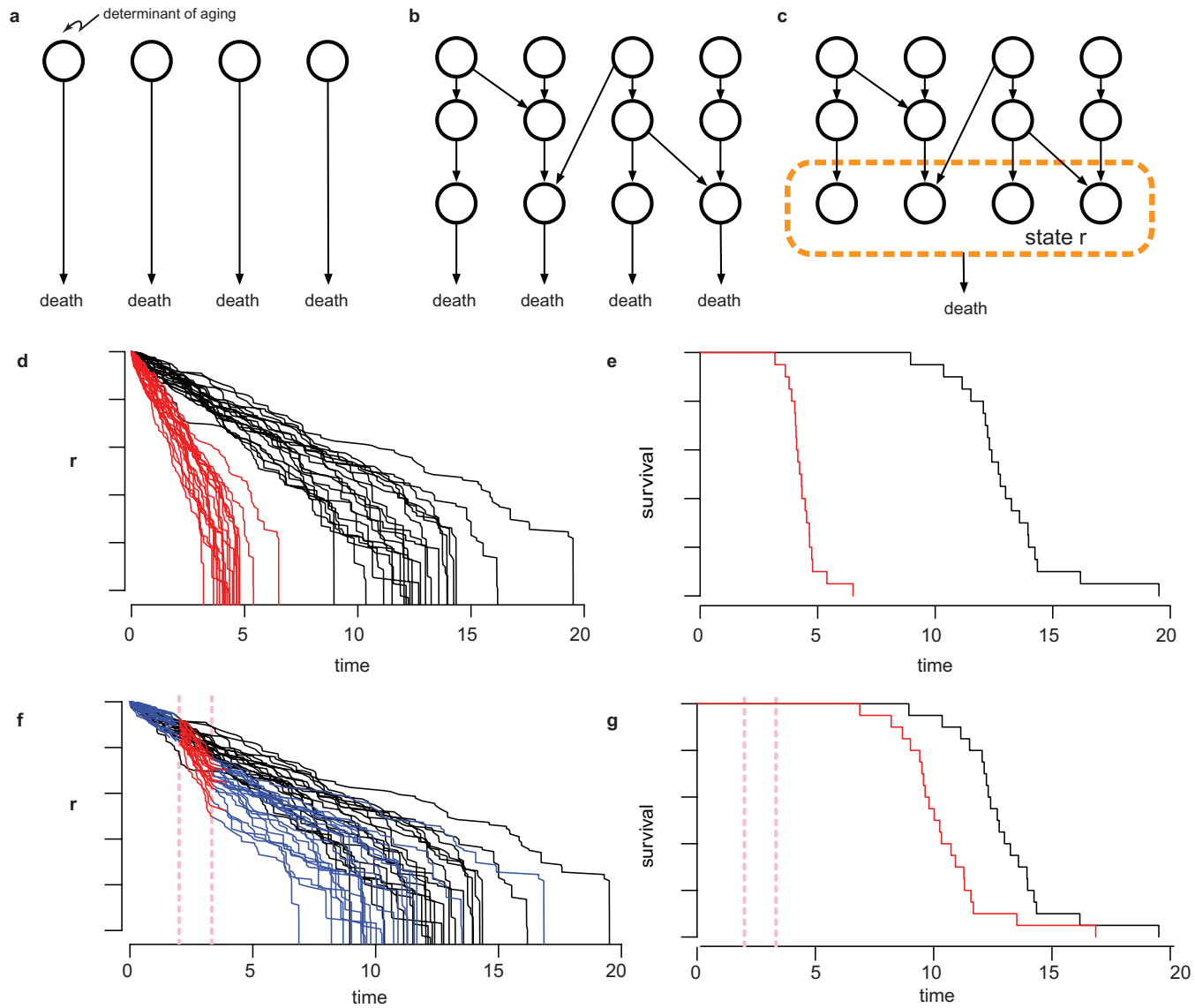
Extended Data Figure 6 | Identifying the number of thermal regimes, and the boundaries between them. The thermal scaling data presented in Fig. 4a were fitted with a segmented regression model (statistical methods) assuming that λ^{-1} relates to temperature either following an Arrhenius relationship, $\lambda(T)^{-1} = p_0 \exp(-p_1/RT)$ (**a**, **c**, **e**, **g**, **i**, **k**) or a linear relationship, $\lambda(T)^{-1} = p_2\,T + p_3$ (**b**, **d**, **f**, **h**, **j**, **l**; statistical methods). As usual we plot the Arrhenius relation on a log–log scale. To emphasize detail, the ordinate of the linear models is also plotted on a logarithmic scale, but the abscissa is kept linear. The model fits are plotted in black, with segment breakpoints shown as vertical lines. Colours correspond to independent biological replicates. Each model was fitted assuming a single segment (**a**, **b**), two segments (**c**, **d**), three segments (**e**, **f**), four segments (**g**, **h**), five segments (**i**, **j**), or six segments (**k**, **l**). **m**, The AIC corresponding to each number of segments for the Arrhenius is plotted. **n**, The AIC for the linear model is plotted. Because the Arrhenius and linear models are fit to distinct data sets (log-transformed and untransformed scale factors respectively, and inverse temperature and untransformed temperatures, respectively), AIC values cannot be compared between Arrhenius and inverse time models. Regime I can be adequately described either by one linear regime or two piecewise Arrhenius regimes, Ia and Ib. Across multiple replicates, the linear model consistently underestimated *C. elegans* lifespan around 25 °C, leading us to favour the piecewise Arrhenius model. Regimes II and III involve temperature ranges that are too narrow to distinguish between Arrhenius and linear models. Above 35.5 °C, lifespan is too short to be accurately measured with our time-lapse technique.



Extended Data Figure 7 | Additional data for the regression models in Fig. 4b, d. The residuals of the linear model (**a**) shown in Extended Data Fig. 6f. The residuals of the Arrhenius model (**b**) shown in Fig. 4b and Extended Data Fig. 6g are depicted, showing the deviation of the predicted value from the empirical data across each regime. Residuals are presented in the form of relative error, the ratio between the model's prediction and the empirical measurement. **c**, As in Fig. 4d, the response of each genotype, wild type (black), *daf-16*(*mu86*) (red), *daf-2*(*e1368*) (green), and *age-1*(*hx546*) (blue) to changes of temperature was estimated (statistical methods) in regime II and in regime III.



Extended Data Figure 8 | The potential effects of heterogeneity at 33 °C. a, Fifty-six thousand samples were drawn from the distribution of frailty effects $Z^{-1/\alpha}$ as described in Supplementary Note 3.1, where Z is a random variable, sampled from an inverse-gamma distribution with mean of 1 and a standard deviation corresponding to the value estimated from experimental data. Samples were drawn using the σ^2 estimated for populations at 25 °C (black) and at 33 °C (red), corresponding to the data presented in Fig. 1d. The probability density function of each population is shown, which can be interpreted as the variable effect of unknown factors on lifespan across individuals at each temperature. **b**, At each temperature, 25 °C (black) and at 33 °C (red), we estimated the distribution of temperature changes required to produce the distribution of frailties shown in **a**. This was accomplished using the temperature scaling function shown in Fig. 4b. **c**, 56,000 random samples were drawn from the transformed inverse gamma distribution of $Z^{-1/\alpha}$ with σ^2 set to the estimate of $\Delta\sigma^2$ in equation (15) of Supplementary Note 3.2. Each sample was multiplied by a death time drawn (with replacement) from the set of 25 °C residual times of Fig. 1d, shown here in black. These products constitute a 'transformed' set of death times, corresponding to the 25 °C residuals with additional frailty synthetically introduced. The residual death times of animals placed at 33 °C are shown for comparison (red).



Extended Data Figure 9 | The organization of lifespan determinants (schematic). a, A set of molecular determinants of risk of death (open circles) do not interact, as is assumed in a competing risks and weakest link models. **b**, Risk determinants might interact (arrows) in complex ways to determine lifespan. In this schematic, each risk of death is still determined by separate factors. **c**, Our data on temporal scaling suggest that the set of molecular determinants that determine risk of death (within the dotted circle) must change in concert when exposed to interventions in ageing. This set is therefore well described by a single state variable *r*. **d**, A cartoon of the stochastic decline of such a state variable (generated from the dependency model discussed in Supplementary Note 5.2). Each trajectory represents the values of *r* over time for each individual in a population. Interventions affect the dynamics of the state decline by rescaling the average dynamics of exposed individuals (red lines), which produces (**e**) a rescaling of the resultant survival curve. **f**, Transient interventions in young adults (applied within the red dotted vertical lines) transiently rescale the average dynamics, leading to (**g**) a shift in the lifespan distribution.

doi:10.1038/nature16550

Table of Contents

Note 1 — Quantitative descriptions of temporal scaling

Summary

This Note introduces the demographic terms we use to describe and quantify the phenomenon of temporal scaling in our experimental data. We then relate our perspective to common alternative approaches, specifically the proportional hazards (PH) model and parameterized regression models.

Contents of Note 1

1.1 Definitions and terminology

The empirical survival function $\hat{S}[t]$ reports the fraction of individuals remaining alive at time t. It is an estimate of the continuous survival function $S(t)$ from which experimental data are sampled. Time (age) t is defined relative to a synchronization event that marks $t = 0$, which in our case is the transition from L4 to adulthood as determined in Methods. In probabilistic language, $S(t)$ is a cumulative distribution; it reports the probability that an individual is still alive at time t. If T is the random variable denoting the time of death, then $S(t) = \text{Prob}(T > t)$. The probability density function, or pdf, $l(t)$ is the probability density that an individual dies between t and $t + dt$, $l(t) = \text{Prob}(t < T < t + dt) = -dS(t)/dt$. This is also referred to as the lifespan distribution. Lifespan distribution and survival function are related through the hazard function $h(t)$, which is the instantaneous rate of death events, or in probabilistic terms, the conditional probability of dying between t and $t + dt$, given survival up to t: $h(t) = \text{Prob}(t < T < t + dt \,|\, T > t)$. The basic demographic equation in continuous time for non-reproducing populations synchronized at $t = 0$ is then given by

$$l(t) = -\frac{d}{dt}S(t) = h(t)S(t). \tag{1}$$

The hazard, $h(t) = -d \log S(t)/dt$, can be seen as playing the role of a time-dependent rate factor in the survival kinetics, equation (1). The survival kinetics should not be confused with the aging dynamics, which is some stochastic process that governs survival by determining $h(t)$. In our experiments we directly observe the survival kinetics, but not the aging process. In Fig. 1 of the main text, we estimate $h(t)$ from lifespan measurements and identify a striking regularity in the way $h(t)$ responds to interventions. This, then, allows us to infer properties of the aging dynamics that generates $h(t)$.

We occasionally refer to survival, hazard, and lifespan distribution as "endpoint statistics" because the stochastic aging process ends in an absorbing boundary (death). In this sense, lifespan is the first passage time of an individual process to this absorbing boundary [1].

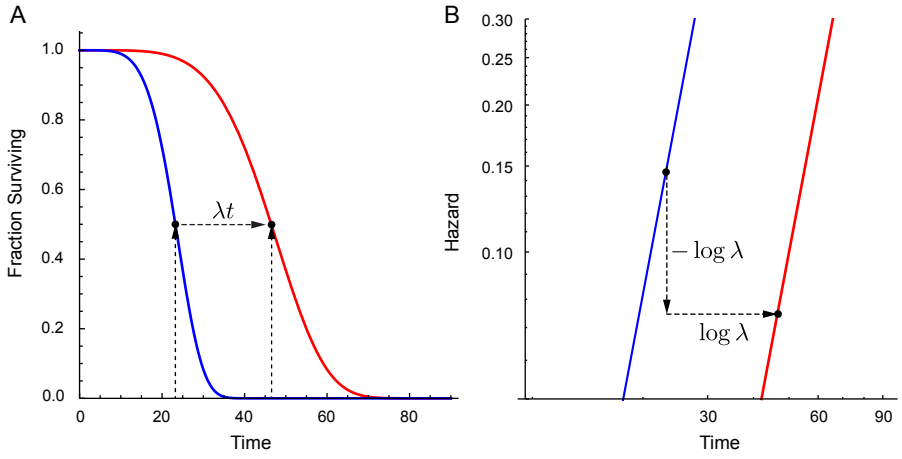
1.2 Temporal scaling

Temporal scaling refers to a situation in which the differences between lifespan distributions corresponding to populations in a reference state **0** and a perturbed state **1** consist in a rescaling of time, such that time t under condition **1** corresponds to time t/λ under condition **0** (Supp. Fig. 1-1), with λ a dimensionless factor:

$$S_1(t) = S_0(\lambda^{-1}t) \qquad \text{or} \qquad S_1(\lambda t) = S_0(t) \tag{2}$$

For example, if $\lambda = 2$, then 1 hour in system **1** corresponds to $1/2$ hour in system **0**. We chose the definition of λ such that a $\lambda > 1$ means that time has slowed down (lifespan is extended by a factor of λ) in the new state and $\lambda < 1$ means that time has accelerated (lifespan is shortened by a factor of λ). Temporal scaling leaves the shape of $l(t)$, $S(t)$, and $h(t)$ invariant. One aspect of shape is the coefficient of variation, defined as the ratio of standard deviation to mean lifespan.

As an illustration, consider the Gompertz survival function for some reference state **0**, $S_0(t) = \exp(-a(\exp(b_0^{-1}t) - 1))$. If an intervention alters the time scale b_0 into b_1, yielding $S_1(t) = \exp(-a(\exp(b_1^{-1}t) - 1))$, a dimensionless factor $\lambda = b_1/b_0$ rescales the time variable such that equation (2) holds: $\exp(-a(\exp(b_1^{-1}t) - 1)) = \exp(-a(\exp(b_0^{-1}\lambda^{-1}t) - 1))$.



Supplementary Figure 1-1: Scaling of survival and hazard functions (I). We speak of temporal scaling when two functions can be made identical by a dilation or contraction of the time axis. **A:** Each value of the blue survival function at time t, S_0, corresponds to the value of the red function S_1 at time λt: $S_0(t) = S_1(\lambda t)$. The point singled out in this example is the median and $\lambda = 2$. **B:** In Fig. 1a of the main text, we show the scaling relation not for the survival function but the hazard function instead. Temporal scaling in the hazard means that $h_1(t) = \lambda^{-1}h_0(\lambda^{-1}t)$ (see equation 3 below) or, equivalently, $h_1(\lambda t) = \lambda^{-1}h_0(t)$. In a log-log plot, we first shift a point on $\log h_0$ (blue curve) corresponding to a particular survival quantile, say the median (black dot), vertically by $-\log\lambda$ (this is the $\lambda^{-1}h_0(t)$ part on a log-log scale) and then horizontally along the time axis by $\log\lambda$ to obtain the corresponding point of $\log h_1$ on the red curve (i.e. the $h_1(\lambda t)$ part on a log-log scale). This is the origin of the remark in the main text that any two hazard curves are shifted by an "equal and opposite amount in magnitude and time." These statements are independent of any parametric form of the survival or hazard function. In this example we used for the purpose of illustration the Weibull distribution with survival function $S(t) = \exp(-(t/\beta)^\alpha)$ and hazard function $h(t) = \alpha/\beta\ (t/\beta)^{\alpha-1}$, which is linear on a log-log scale.

Using equation (2) in equation (1) yields:

$$-\frac{d}{dt}S_1(t) = h_1(t)S_1(t) \Rightarrow -\frac{d}{dt}S_0(\lambda^{-1}t) = h_1(t)S_0(\lambda^{-1}t)$$
$$\overset{\text{set } \tau=\lambda^{-1}t}{\Longrightarrow} -\frac{d}{d\tau}S_0(\tau) = \lambda h_1(\lambda\tau)S_0(\tau) \Rightarrow h_0(\tau) = \lambda h_1(\lambda\tau). \quad (3)$$

Hence, a rescaling of survival, $S_1(t) = S_0(\lambda^{-1}t)$, is equivalent to a rescaling of the hazard as

$$h_1(t) = \lambda^{-1}h_0(\lambda^{-1}t) \quad (4)$$

and, by a similar argument, of the lifespan density as $l_1(t) = \lambda^{-1}l_0(\lambda^{-1}t)$. An intervention that yields $\lambda = 2$ slows down the original system twofold with respect to survival, which is the same as saying that the new system will exhibit half the hazard rate after twice the time.

1.3 Accelerated Failure Time vs Proportional Hazards

In survival analysis, equation (2) is known as an Accelerated Failure Time (AFT) model, in which the lifespan random variable T of population **0**, living at condition C_0, is linearly transformed into the random variable λT of population **1**, living at condition C_1. An AFT regression model with a single categorical covariate can be expressed as:

$$\log t_i = \boldsymbol{\beta x_i} + \varepsilon_i, \quad (5)$$

where t_i is the observed lifespan of individual i, X an explanatory categorical variable and $\boldsymbol{x_i}$ a vector encoding this variable for individual i (e.g. whether it was living at C_0 or C_1); $\boldsymbol{\beta}$ is a vector of effects to be estimated, and ε_i is a residual representing the logarithm of the lifespan of individual i unexplained by the covariate. The relation to the scale factor λ is given by $\lambda = \exp(\boldsymbol{\beta x_i})$.

A rescaling of the lifespan random variable T into a new random variable λT alters the expected value of T, $E(T)$, into $\lambda E(T)$ and the variance $\sigma^2(T)$ into $\lambda^2\sigma^2(T)$. As a consequence, temporal rescaling leaves the coefficient of variation $\sigma(T)/E(T)$ unchanged.

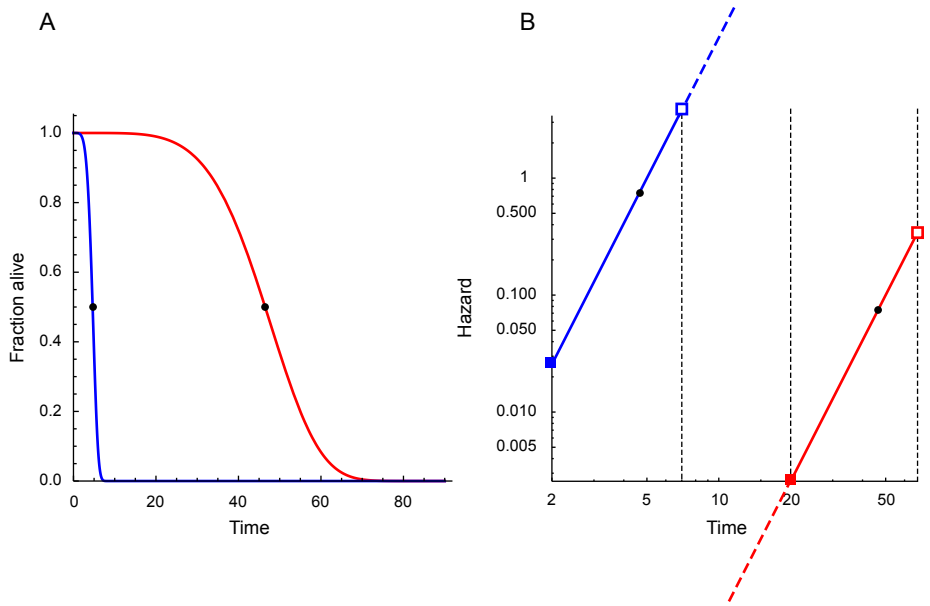
We apply the AFT model to link physical interventions in aging to their quantitative consequences on lifespan. When AFT models provide an excellent fit across sets of lifespan distributions, we conclude that the interventions producing such distributions result in temporal scaling. Temporal scaling suggests that interventions alter lifespan by altering the timescale of an aging *p*rocess that determines the risk of death $h(t)$.

This approach contrasts with the widely used proportional hazards (PH) framework, which takes the perspective that interventions act directly on the risk of death, such that

$$\log h_i(t) = \boldsymbol{\beta x_i} + \log h_0(t), \quad (6)$$

in which $h_i(t)$ is expressed in terms of a baseline hazard $h_0(t)$. In the case of two populations ($i = 0, 1$ and assumed to consist of identical individuals), the PH approach thus quantifies the consequences of exposure to condition **1** as a proportionality factor $\exp\beta$ [2].

In the special case that mortality follows a Weibull distribution, AFT and PH are mathematically equivalent descriptions, because a power law is a homogenous function. Yet, even when data are not well fit by a Weibull, PH and AFT models can both be reasonable characterizations of the effect of interventions. Our preference for the AFT model over the PH model is founded on three



Supplementary Figure 1-2: Scaling of survival and hazard functions (II). **A:** The panel shows a synthetic situation (Weibull survival) analogous to Supp. Fig. 1-1a but with a tenfold slow-down ($\lambda = 10$) in time from blue to red. In the blue population everything has already happened before something happens in the red population. For example, in our temperature experiments of Fig. 1 in the main text, the scale factor between highest and lowest temperature is about 40 (!). As in Supp. Fig. 1-1, the black dots indicate the survival median and we used a Weibull distribution for illustration (thus the linear hazard in the log-log plot of panel B). **B:** The panel illustrates the corresponding situation for the hazards. The solid lines (blue and red) depict the hazard trajectory in the time interval starting when 1% (solid square) of the population have died and ending when 99% (open square) have died. In this synthetic case, the mathematical functions continue beyond these intervals (dotted lines) with extremely high or extremely low values, but seen from an empirical standpoint there are no meaningful data outside the solid regime. Vastly increasing the population size would only have a marginal effect on this picture. The red and blue empirical hazard functions have effectively no overlap and can therefore not be quantitatively compared as in equation (6), i.e. one hazard cannot be expressed in terms of the other as the baseline hazard. There is no empirically justified sense in which they could be "proportional". This is not the case in situations characterized by relatively small effects, such as clinical drug applications, where the PH framework is routinely applied.

observations: (1) We find that our interventions have a permanent effect on lifespan when applied transiently early in life, before any deaths occur (Fig. 3b in the main text). This suggests that interventions act on some process determining T rather than on $h(t)$ directly. (2) In many of our experiments, populations differ so much in their aging timescale that there is no time interval during which both populations have survivors. While this presents no mathematical or computational problem, since $h_1(t) = \beta h_0(t)$ is well defined for all t, it does present an empirical problem, since $h_0(t)$ and $h_1(t)$ cannot be measured for any overlapping t (Supp. Fig. 1-2). (3) The parameter values obtained with the AFT model seem more physically realistic than those obtained with the PH model. For example, if one were to compare compare a population at 20 °C with one at 30 °C, the fold-increase in relative risk that we extrapolated for the PH model comes in at 4.5×10^5. It is hard to imagine how a physical intervention could *directly* impact risk ratios

related to aging almost a $1,000,000$-fold. It is more plausible that such an increase in risk is the *indirect* consequence of altering 18-fold the timescale of an underlying physical process, as assumed by the AFT model.

1.4 Scaling and parametric models

An often used method for characterizing the effect of interventions in aging is based on simple parametric distributions, such as the Gompertz, Weibull, log-logistic and inverse Gaussian distributions (see Note 1). For example, the Gompertz hazard has been used to classify interventions in *C. elegans* and other model organisms according to their effects on the parameters of the distribution [3–7]. Another function, quite common in the analysis of engineered systems, is the Weibull distribution, see Table 1 below.

1.4.1 On the parametrization of the Gompertz hazard

The wide-spread use of Gompertz warrants a comment on its parametrization. Scaling in the context of Gompertz fits is best understood using a different parametrization than the one typically employed. In the parametrization $h(t) = a\exp(t/b)$ (for example [3,4]), the parameters a and b are not independent, but a depends inversely on b. If b is interpreted as a time scale parameter (or $1/b$ as a rate parameter), then a change from b to λb requires a change from a to $\lambda^{-1}a$, see equation 4. Another way of seeing this dependence is to calculate the survival function associated with the parametrization in question: $S(t) = \exp(-ab(\exp(t/b) - 1))$. The argument of the exponent should be a dimensionless quantity, which means that the time scale parameter b should only occur in conjunction with the time variable t. For this to be the case, the factor ab must be independent of b, which can be achieved by using a new parameter $a' = ab$, and thus $h(t) = (a'/b)\exp(t/b)$, where a' is now a constant that is independent of b, with associated survival function $S(t) = \exp(-a'(\exp(t/b) - 1))$. This second parameterization is common in some statistical packages and areas of study, but does not seem to be widely used in studies of aging. In the parametrization $h(t) = a\exp(t/b)$, any intervention that acts on b must necessarily affect a, confounding attempts to map effects to distinct parameters. Changes in the Gompertz a—vertical drops in a lin-log plot—should be interpreted only after the dependence on b is accounted for.

1.4.2 The parametric form of wild-type *C. elegans* hazard

Our quantification of deviations from temporal scaling (Note 2) does *not* proceed by fitting empiric data with simple parametric functions. This is a significant methodological advantage. Nevertheless, it is useful to identify a "best" parametrized form for the *C. elegans* hazard function when attempting to quantify hazard heterogeneity in the population (Note 3.2). To this end, we surveyed several parametric hazards derived from the Gompertz, Weibull, log-logistic, log-normal, and inverse Gaussian distributions, as well as Weibull and Gompertz hazards with a frailty term. These functions are listed in Table 1. Model parameters were estimated in R using the `flexsurv` package. Across several replicates containing between 650 and 1,200 wildtype individuals at 20 °C, we confirmed previous observations [8,9] that neither the Gompertz nor the Weibull hazard without frailty provide good fits for *C. elegans* populations (Extended Data Item 1a). However, in all cases the Weibull distribution performs distinctively better than the Gompertz distribution (Extended Data Item 1b, c). The log-logistic and inverse Gaussian distributions both predict an intrinsic deceleration of the population hazard, and for that reason fit the data better than either the two-parameter Weibull or Gompertz distributions (Extended Data Item 1b–c). We find that incorporating a frailty correction reconciled the Gompertz and Weibull hazards with the experimental data; either frailty model provided excellent fits across all replicates. In all cases, the Weibull model required a smaller frailty term σ^2 (corresponding to the

assumption of a lower degree of hidden heterogeneity between individuals) to match the data than Gompertz (Extended Data Item 1b–d). This motivates our preference for the Weibull hazard with frailty for data analysis and simulation whenever a parametric form is needed.

$$\text{Gompertz:} \quad h(t \mid a, b) = \frac{a}{b} \exp\left(\frac{t}{b}\right)$$

$$\text{Gompertz with frailty:} \quad h(t \mid a, b, \sigma) = \frac{\frac{a}{b} \exp\left(\frac{t}{b}\right)}{1 + \sigma^2 \, a \exp\left(\frac{t}{b} - 1\right)}$$

$$\text{Weibull:} \quad h(t \mid \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha - 1}$$

$$\text{Weibull with frailty:} \quad h(t \mid \alpha, \beta, \sigma) = \frac{\frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha - 1}}{1 + \sigma^2 \left(\frac{t}{\beta}\right)^{\alpha}}$$

$$\text{Log-normal:} \quad h(t \mid \mu, \sigma) = \frac{\phi\left(\frac{\log t - \mu}{\sigma}\right)}{\sigma t \left(1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right)}$$

$$\text{Log-logistic:} \quad h(t \mid \alpha, \beta) = \frac{\frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha - 1}}{1 + \left(\frac{t}{\beta}\right)^{\alpha}}$$

$$\text{inverse Gaussian:} \quad h(t \mid \mu, \lambda) = \frac{\sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda (t - \mu)^2}{2 \mu^2 t}\right)}{1 - \Phi\left(\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right) \phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right)}$$

Table 1: Distributions used for fitting. ϕ and Φ denote the PDF and CDF, respectively, of the standard normal distribution; μ and σ denote the mean and standard deviation (in the case of the log-normal, the mean and standard deviation of the logarithm of x); λ, α, and a are shape parameters; β and b are scale parameters. In the case of frailty, individual hazards $h_i(t)$ are related to a baseline hazard by a random factor Z that follows a Gamma distribution with mean 1 and variance σ^2.

Note 2 — Quantifying temporal scaling

Summary

We quantify deviations from temporal scaling by applying a modified Kolmogorov-Smirnov (KS) test to the residuals of Accelerated Failure Time (AFT) regression models (Fig. 1c-d). The KS test we deploy was generalized to work with censored data by Fleming et al. [10]. Like the standard KS procedure, it tests whether the AFT residuals come from the same underlying distribution, in which case we conclude that all statistical differences between two sets of lifespan data are indeed accounted for by a scale factor and the two distributions are therefore related by temporal scaling. The KS test also provides a distance metric between survival curves. We note that many familiar statistical methods for evaluating differences between lifespan distributions, for example the log-rank test, perform poorly when survival curves cross, a situation perhaps rare in raw empiric data but essentially guaranteed among AFT residuals which, by construction, will have nearly identical means and differ only at higher moments.

Contents of Note 2

2.1 Adapting the Kolmogorov-Smirnov two-sample test

We use the test statistic proposed by Fleming et al. [10] to confirm or reject the null hypothesis that two survival datasets are sampled from the same underlying distributions (and thus related by temporal scaling). This test statistic is also useful as a distance metric for clustering survival curves (Extended Data Item 3).

We first fit two merged sets of lifespan data with the AFT model

$$\log y_i = \boldsymbol{\beta x_i} + \epsilon_i \tag{7}$$

where the categorical covariate x_i codes for the experimental grouping of individuals, for example plate ID or scanner ID, as detailed in our Statistical Methods section. The resulting residuals ϵ_i are then grouped according to the experimental variable in which they differ, for example genotype or environmental condition. Given populations a and b, we refer to the CDFs of these residuals with $S_a^\epsilon(t)$ and $S_b^\epsilon(t)$, respectively, although these survival curves are never explicitly calculated.

In the absence of censoring, the canonical Kolmogorov-Smirnov test involves estimating

$$F_{a,b}(t) = S_a^\epsilon(t) - S_b^\epsilon(t) \tag{8}$$

and then computing the KS statistic

$$D_{a,b} = \sup |F_{a,b}(t)|, \tag{9}$$

where $\sup$ is the supremum (maximum) function, to calculate a p-value pertaining to the null hypothesis. Given the nature of the KS test and our AFT pre-treatment, the null hypothesis *affirms* temporal scaling.

Fleming et al. propose an analogous statistics for censored data, with

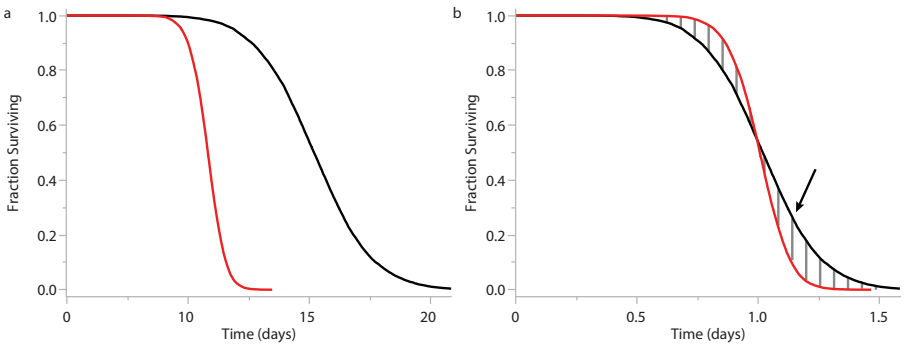
$$Y_{a,b}(t) = \left(\frac{S_a^\epsilon(t) + S_b^\epsilon(t)}{2} \right) \left(\beta_a^\epsilon(t) - \beta_b^\epsilon(t) \right), \tag{10}$$

where $\beta^\epsilon[t]$ is the cumulative hazard at time t for each set of residuals, and the modified KS statistic

$$\bar{\mathcal{Y}}_{a,b} = \sup |Y_{a,b}(t)|. \tag{11}$$

We will refer to $Y_{a,b}(t)$ as $Y(t)$ and $\bar{\mathcal{Y}}_{a,b}$ as $\bar{\mathcal{Y}}$ when the two populations a and b are obvious from context. $\bar{\mathcal{Y}}$ quantifies the degree of heteroscedasticity in the AFT covariate X_i. A significant p-value corresponding to $\bar{\mathcal{Y}}$ suggests a significant deviation from perfect scaling between the two survival curves.

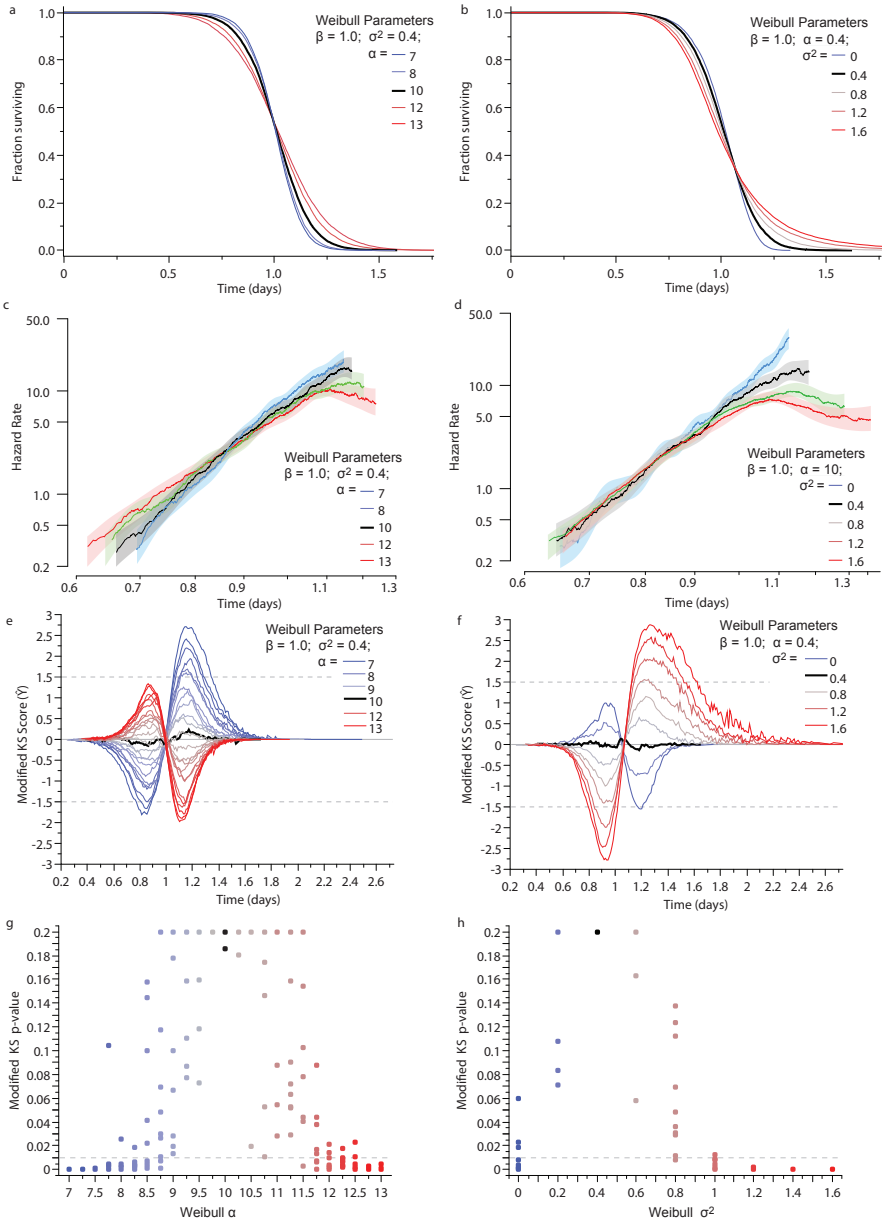
The p-value calculated from this modified KS test depends both on the significance level desired, the $\bar{\mathcal{Y}}$ obtained, and on the fraction of animals remaining alive at the end of an experiment (equation 2.4.iv in Flemming et al. [10]). Since this fraction is essentially zero in all experimental (and simulated) data sets presented here, the KS p-value simplifies to $\exp(-2 * \bar{\mathcal{Y}}^2)$. Accordingly, any $\bar{\mathcal{Y}} \geq 1.51$ or $\bar{\mathcal{Y}} \leq -1.51$ will be significant at the $p \leq 0.01$ level. Because of how $\bar{\mathcal{Y}}$ is defined, any $Y(t)$ rising above above 1.51 or below -1.51 indicates a significant deviation from temporal scaling. We use this threshold throughout the Supplement when interpreting the values of $Y(t)$ and $\bar{\mathcal{Y}}$.



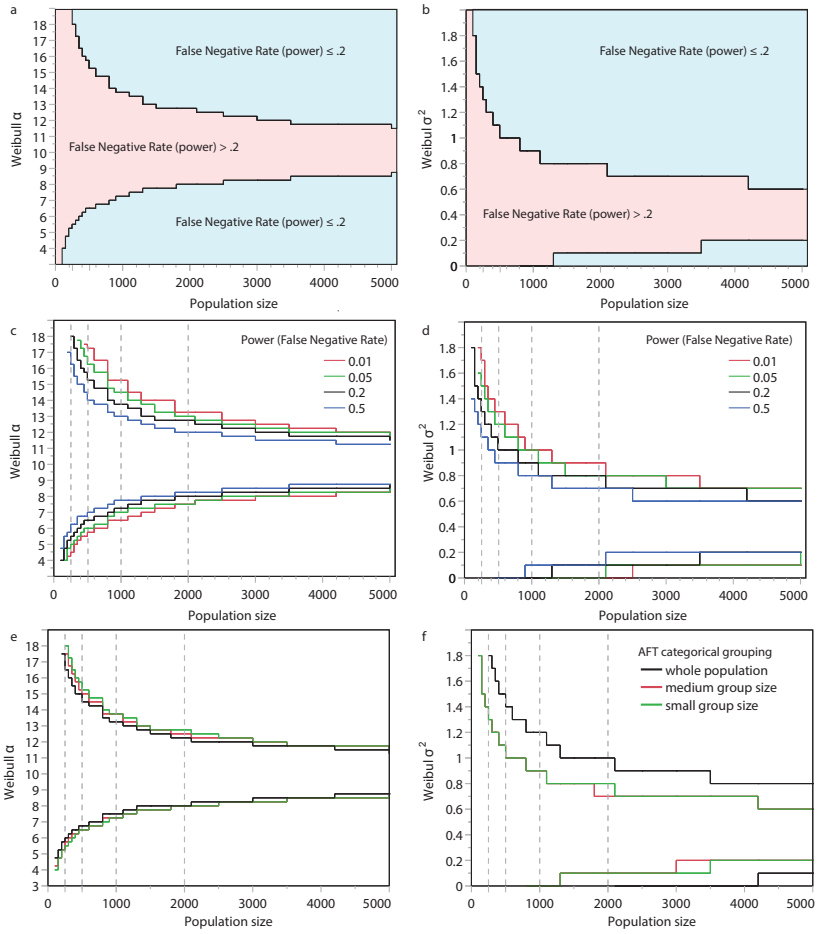
Supplementary Figure 2-1: Quantifying deviations from perfect scaling (schematic). **a:** To quantify deviations from perfect scaling, we first apply an Accelerated Failure Time model, to obtain the residuals of each curve (panel b). This removes any differences in time scale. If such differences are all there is, we have perfect scaling, which is to say that both residuals come from the same distribution. This is the null hypothesis that we test by quantifying the deviation from perfect scaling using a generalized Kolmogorov-Smirnov (KS) test [10]. **b:** The KS test evaluates the distance between two survival curves of AFT residuals (vertical gray bars) by assessing the probability of observing their maximal distance (arrow) by chance. This provides a way of quantifying the overall deviation from perfect scaling between the two curves.

2.2 Evaluating the performance of the generalized KS test

To evaluate the sensitivity of the AFT+KS-test combination, we compared sets of simulated lifespan distributions differing by progressively smaller parameter values (Fig. 2-2). In populations of 1,500 individuals, the test statistic reliably identified changes in the Weibull α of about 2.5 and increases in σ^2 of about 0.6. In this case and in subsequent power calculations, we censored 10% of death times to match the level typical in our experimental data.



Supplementary Figure 2-2: Evaluating the AFT+KS-test combination with simulated data. Sets of death times containing $1,500$ samples were drawn from a Weibull distribution with frailty correction((1.4). Baseline parameters were chosen to approximate those observed in the AFT residuals of wild-type *C. elegans* lifespan distributions, $\alpha = 10$, $\theta = 0.4$, and $\beta = 1$ (α and β refer to parameters of the Weibull distribution, not the AFT regression). Any change in the shape parameter α or frailty parameter σ^2 will cause simulated populations to deviate from the baseline population, as is visible in the survival curves (panels **a–b**) and hazard plots (panels **c–d**) of the AFT residuals calculated for each population. **e–f:** Each simulated population was compared to the baseline population, plotting the generalized KS distance $Y(t)$, equation (10). **g–h:** Comparison of each population to the baseline across replicate trials with a calculation of the modified KS p-value to estimate the likelihood that observed differences occurred by chance. All values $p > 0.2$ were cropped at $p = 0.2$.



Supplementary Figure 2-3: Power of the AFT+KS-test combination with simulated data. **a:** The phase-space plot indicates the minimum population size required to detect a given effect size (a difference in the Weibull α to the reference population) at a statistical power of 0.8 (a false negative rate of 0.2) and significance level (false positive rate) of 0.01. The blue (red) area indicates effect and population size combinations that yield a statistical power greater (less) than 0.2. Populations of $1,000$ individuals are required to detect differences in α of 2.5; increasing populations beyond $1,500$ individuals appears to yield diminishing returns. **b:** Similar conclusions can be drawn from a phase-space plot for the frailty σ^2. Increasing the population size beyond $1,500$ individuals is met with diminishing returns in the capacity to discriminate differences in frailty. **c, d:** Same as in panels a and b, but for a statistical power of 0.5, 0.8, 0.95 and 0.99. **e, f:** The AFT model preceding the KS test can be run on different population size "units" afforded by the Lifespan Machine: plates (≈ 40 animals), whole scanners (≈ 320 animals), or the entire population. Here, each sample of the same unit size is a categorical variable in the AFT regression. After running the regression on all units, the residuals are pooled to form the distribution that is compared against the reference distribution with the generalized KS test. Depending on whether the differences are in α (panel e) or σ^2 (panel f), pooled populations will be subject to different distortions. It is therefore meaningful to check how such pooling impacts power. The results are shown in panels e and f for AFT pre-processing on subpopulations that were small (40), medium (320) or large (750). The size of subpopulation pooled does not have a big effect on power with regard to α (panel e), but it has some effect with regard to frailty (panel f).

To assess the statistical power of the AFT+KS-test combination, we generated sets of simulated death times as for Fig. 2-2, but varied the population size between 50 and $5,000$ individuals; the Weibull α was varied between 4 and 18 and, separately, the frailty σ^2 between 0 and 2. Populations corresponding to each parameter set were compared to a baseline population of equal size with Weibull $\alpha = 10$ and $\sigma^2 = 0.4$, which closely matches *C. elegans* distributions. The p-value from each comparison was collected across 750 replicates. For each set of replicates and each population size, the fraction of significant p-values ($p \leq 0.01$) was then computed to estimate the power of the KS test at each population size (Fig. 2-3). Fig. 2-3 shows the α or σ^2 (the effect size) that can be detected at a given population size with a given false negative rate. We found that a minimum of 500 individuals per population were required to detect even substantial deviations from temporal scaling, but that increasing population sizes beyond 1,500 individuals provided diminishing returns in statistical power. As expected, substantial increases in population size were required to achieve greater statistical power: decreasing the false-negative rate of the test from 0.5 (half of all true deviations from survival curve shape are missed) to 0.05 (5% of all true deviations from survival curve shape are missed) required at least 2.5-fold larger populations at any effect size.

Note 3 — Heterogeneity and deviations from temporal scaling

Summary

We observe a change in hazard shape when animals are transferred from what we might call the "low temperature regime" (LTR, 20 °C–30 °C) to the "high temperature regime" (HTR, 30 °C–35 °C), across the 30 °C mark. This change appears to break scaling due to an increased deceleration of the hazard function in the HTR as compared to the LTR (Fig. 1d, e of the main text). A deceleration of the hazard can be accounted for in terms of "hidden" (i.e. not measured) heterogeneity also known as "frailty" in the demographic literature [11]. Such an approach distinguishes hazard functions that pertain to individuals from the hazard function that pertains to the population. From this perspective, an increased hazard deceleration could be due to an increased heterogeneity in the population. Furthermore, if heterogeneity were due to differences between individuals that nonetheless conform with a temporal rescaling of their aging process, the population hazard would still be affected in a way that deviates from temporal scaling. However, such a deviation would be incidental rather than fundamental. Heterogeneity of this kind might be an unavoidable side-effect of our experimental protocols, and a possible candidate are temperature inhomogeneities in our apparatus. We therefore sought to quantify the magnitude in heterogeneity that would be required to explain the change in hazard shape observed between 25 °C and 33 °C. We first quantify heterogeneity using a parametrized hazard function and then map it to a corresponding heterogeneity in temperature, since heterogeneity of this nature would be consistent with scaling at the individual level. The results of this analysis suggest that temperature inhomogeneities in our apparatus are unlikely to account for the change in hazard across the 30 °C transition. We cannot exclude other sources, some of which may be compatible with scaling of individual hazards, others may not. In sum, our experimental data do not allow us to adjudicate whether the change in hazard we observe is or is not a true scale breaker, but they suggest rejecting the hypothesis that potential temperature inhomogeneities in our apparatus are the cause of the change in hazard shape across the 30 °C transition.

Contents of Note 3

3.1 Background: Weibull heterogeneity (frailty)

In the demographic literature,"frailty" refers to unobserved (in the sense of not measured) heterogeneity among individuals within a population [11]. The assumption is often made that individuals or subpopulations differ from one another with respect to hazard in a fashion that can be characterized by a PH model:

$$h_i(t) = Z_i\, h_0(t), \tag{12}$$

where $h_i(t)$ is the hazard of individual i, $h_0(t)$ is the common baseline hazard, and Z_i is the value of a random variable Z, often assumed to follow a Gamma distribution with mean 1 and variance σ^2.

It has been shown (see any textbook, such as [1]) that under these assumptions the population

hazard $h(t)$ behaves like

$$h(t) = \frac{h_0(t)}{1 + \sigma^2 H_0(t)}, \tag{13}$$

where $H_0(t)$ is the cumulative baseline hazard, $H_0(t) = \int_0^t h_0(\tau)d\tau$.

For example, if the hazard follows a Weibull form, $\alpha/\beta(t/\beta)^{\alpha-1}$, the individual hazards in the frailty model just described become:

$$h_i(t \mid \alpha, \beta) = \frac{\alpha}{Z_i^{-\frac{1}{\alpha}}\beta}\left(\frac{t}{Z_i^{-\frac{1}{\alpha}}\beta}\right)^{\alpha-1} \tag{14}$$

In this case, the proportional frailty model, equation (12), is equivalent to a model in which individual differences (Z_i) multiply the timescale parameter β: in a Weibull frailty model, hidden heterogeneity within a population impacts lifespan through temporal scaling. The Weibull with frailty population-level hazard is listed in Table 1.

3.2 Quantifying heterogeneity

A deceleration in population hazard has been noted in model organisms and in human populations [6, 9, 12, 13]. Several distinct but mutually non-exclusive explanations for the phenomenon have been offered. One explanation invokes heterogeneity, which can give rise to a hazard deceleration because the more frail individuals die early, leaving a population with a declining hazard rate. This type of heterogeneity could be intrinsic to populations, or be produced by some extrinsic environmental heterogeneity, for example the slight temperature differences we can measure across different regions of our imaging apparatus [8]. Heterogeneity due to temperature differences at different locations in our apparatus would cause deviations from temporal scaling at the population level even though they only cause temporal rescaling at the individual level.

An alternative explanation is that Markov processes with an absorbing boundary (death) have, in general, absorption (hazard) rates that become constant in the long time limit, even if all individuals are identical initially. Such stochastic processes represent aging as a decline in a physiological state space. They reach a quasi-stationary distribution over physiological states [1, 14–16], which causes the hazard to level off. All examples in Supplementary Note 6, in which we illustrate scaling at the process level, exhibit this phenomenon. In particular, the inverse Gaussian distribution arises from a class of first-passage time models (drift-diffusion) [15] that fit our *C. elegans* data reasonably well. A Weibull model with frailty provides a better fit to our empiric data (Extended Data Item 1b–d), but a small frailty correction to the Gaussian model would entirely remove this advantage.

We next set out to determine whether the changes we observe in hazard curve shape could be reasonably explained by temperature heterogeneity. If the heterogeneity required to produce the observed degree of hazard deceleration were improbably large, we might reject that hypothesis.

In our data sets, the largest deviation from temporal scaling produced by temperature occurs at 30 °C. To estimate the magnitude of change in environmental heterogeneity required to explain this deviation from temporal scaling, we compared populations held at 25 °C and at 33 °C (Fig. 1d) by fitting both with a Weibull model with a frailty correction term (section 3.1). Instead of estimating model parameters independently for each populations and then comparing them, we

apply a single model to both populations simultaneously and quantify changes between them as relative factors. This enables more meaningful confidence intervals and therefore more informative estimates for the significance of differences. Our model represented the two curves as

$$h_{25^\circ}(t \mid \alpha, \beta, \sigma) = \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{1+\sigma^2\left(\frac{t}{\beta}\right)^{\alpha}} \qquad h_{33^\circ}(t \mid \alpha, \beta, \sigma, \Delta\alpha, \Delta\sigma^2) = \frac{\frac{\alpha\Delta\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha\Delta\alpha-1}}{1+\sigma^2\Delta\sigma^2\left(\frac{t}{\beta}\right)^{\alpha\Delta\alpha}} \tag{15}$$

with $\Delta\alpha$ and $\Delta\sigma^2$ representing the effect of high temperature on the shape and scale of the curve.

We estimated model parameters using a maximum likelihood approach (R-package `flexsurv`). At 25 °C, $\alpha = 10.5$ (95% confidence interval $ci = [10.2 - 10.9]$) . At 33 °C populations exhibited a 10% decrease in α, with $\Delta\alpha = 0.91$ (95% $ci = [0.86 - 0.96]$). At 33 °C exhibited a large increase in σ^2, relative to 25 °C, with $\Delta\sigma^2 = 2.7$, (95% $ci = [2.2 - 3.38]$). A model assuming a fixed σ^2 across temperatures ($\Delta\sigma^2 = 1$) provided a substantially worse fit ($AIC = -7385$) to the data than a model assuming a fixed α across temperatures ($\Delta\alpha = 1$, $AIC = -7461$). Were the difference in survival curve shape caused by only a single parameter, it is more likely to result from an increase in σ^2.

To map the magnitude of this increase in σ^2 to a biologically informative scale, consider that frailty (the distribution of the Z_i in equation 14) acts multiplicatively on the timescale β and can thus be interpreted as a distribution of individual scale factors λ_i in the population. A larger σ^2 suggests that individuals experience a greater diversity in their individual λ_i, as shown in Extended Data Item 8 a. At each set point, 25 °C and 33 °C, the data in Fig 4.b of the main text allow us to convert this heterogeneity into a corresponding distribution of temperatures accounting for the distribution of λ_i (and thus the estimated σ^2). This approach suggests that a relatively small heterogeneity in temperature is sufficient to explain σ^2 (Extended Data Item 8 b): at 25 °C only 25% of individuals would need to experience a temperature 0.18 °C less than average and only 25% would need to experience a temperature 0.10 °C greater than average. To explain the increased σ^2 at 33 °C, only 25% of individuals would need to experience a temperature 0.37 °C less than average, and only 25% would need to experience a temperature 0.31 °C greater than average.

Although these temperature differences seem small, we need to consider that the AFT algorithm used to estimate the σ^2 already accounts for any difference in temperature between scanners and between plates. Thus, these differences would have to correspond to thermal gradients *within* each plate. Seen from this angle, the putative temperature differences suddenly seem unlikely large. In addition, since individuals more or less randomize their location on plates over time, it seems improbable that a temperature gradient (of however unlikely magnitude) could produce a substantive heterogeneity-effect. We conclude, therefore, that static temperature heterogeneity does not account for the observed deviations from temporal scaling between populations held at 25 °C and 33 °C. Other sources of heterogeneity, for example a differential induction of the heat-shock response among individuals [17], must be invoked. Some might be compatible with temporal scaling at the individual level, others not.

Our frailty model assumes a baseline Weibull hazard. To determine whether our quantification of heterogenity depended on the specific model used, we ran a simulation that artificially injected heterogeneity into experimental lifespan distributions. We randomly selected $56,000$ death times from AFT-residuals of the data collected at 25 °C (Figure 1.d) and multiplied each sample by a random number drawn from a Gamma distribution corresponding to the increase in σ^2 observed

between populations at 25 °C and 33 °C (the $\Delta\sigma^2$ of Eq. 15). If the difference in shape observed between 25 °C and 33 °C shown in Fig. 1d of the main text resulted exclusively from an increase in frailty, then the distribution of the so-transformed residual death times should be equivalent to that measured at 33 °C. This is indeed what we see (Extended Data Item 8 c, d), suggesting that the deviations from perfect temporal scaling between 25 °C and 33 °C are consistent with an increase in heterogeneity within populations.

In final analysis, temperature differences within our apparatus seem not to account for the heterogeneity we observe across the 30 °C transition. It is therefore possible (though not certain) that whatever causes this heterogeneity is reflective of changes in mechanistic aspects of the stochastic aging process and thus breaks temporal scaling at the individual level.

Note 4 — The effect of transient perturbations

Summary

In Figure 1 of the main text, we identify interventions that, when applied throughout life, produce a temporal scaling of lifespan distributions. For such interventions, we derive expressions that predict the effects on the lifespan distribution when these interventions are applied transiently. We highlight the specific case where interventions are applied before any deaths occur in a population, in which case interventions are predicted to produce a temporal *shift* of the lifespan distribution of the affected population relative to a control. We then explain how these predictions are met by the experimental data presented in Fig. 3 of the main text.

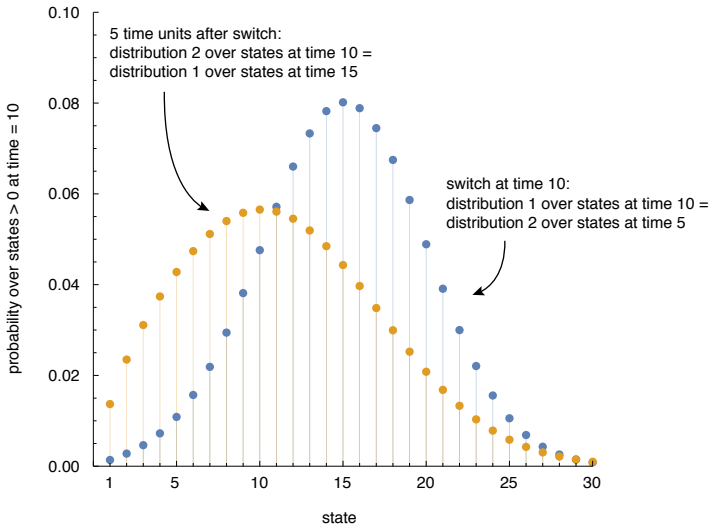
Contents of Note 4

4.1 Intuition

We observe that diverse interventions produce a temporal scaling of lifespan distributions. Perhaps the simplest way for such scaling to come about would be if, throughout adulthood, all changes within a small time increment dt of those aspects of *C. elegans* physiology that determine the risk of death in control individuals occurred within a rescaled time increment λdt in exposed individuals. The intervention would then be producing a temporal scaling of a stochastic aging *process* that determines risk and generate, through this rescaled process, a correspondingly rescaled lifespan distribution. In other words, the cumulative physiological decline of individuals under intervention would accrue on a rescaled timescale (faster or slower, depending on the value of λ) than for individuals in the control population. This speed-up or slow-down in the physiological decline process determining risk would have to occur even if specific underlying molecular mechanisms changed as a result of the intervention.

If at some point the intervention were stopped and conditions reverted to those of control animals, then, from that moment onward, no additional differences would accumulate between populations. However, the existing differences would persist, making an exposed population "older" or "younger" (in terms of risk) than the control population, and, as a result, causing its lifespan distribution to be *shifted*, not scaled, toward earlier or later chronological times, respectively, compared to the control.

Within any population some individuals live longer than others. In our experiments, we leverage this variation within apparently homogeneous populations to test whether interventions are in fact rescaling a lifelong stochastic process that determines lifespan. We show that the variation in lifespan observed within homogenous populations is indeed shifted to a predictable extent by interventions that produce a temporal scaling of the underlying stochastic aging process. We explain the extent of this shift informally in the caption to Fig. 4-1; we derive it formally for expected lifespan in sections 4.2–4.5 and for the case of distributions in section 4.6).



Supplementary Figure 4-1: On shifting lifespan distributions. The distributions of a control population and a population that is transiently subject to an intervention will undergo a shift relative to each other, provided the intervention causes a rescaling of the lifespan distribution if applied throughout life. Consider a control population P_0 always held at condition C_0 and a test population P_1 held at C_1 for T time units, say from time $t = 0$ to $t = T$. As far as the aging process that governs risk of death is concerned, T time units pass for P_0, while T/λ units pass for P_1. (The scale factor λ has the same definition as in Note 1 and the main text. However, in our experiments we switch from a perturbed situation to a control and express time T in the exposed population relative to the control population; as a result, time T is divided by λ.) At time T, P_1 is switched to the same condition as P_0. From that moment on, both age at the same speed, maintaining the temporal difference between control and exposed population: $T(1-1/\lambda)$. For illustration, consider two biased random walks, RW1 and RW2, in a physiologic state space that governs risk. Over time, RW1 and RW2 drift towards the absorbing boundary at 0. RW2 proceeds twice as fast as RW1 and both walks started in the same state. The plot shows the probability distributions over live states of these rescaled walks at two times. At time 10, the parameter values of RW1 are changed to those of RW2. Because of the scaling assumption, 5 time units ago the distribution over states for RW2 must have been exactly the same that RW1 has at the switching time 10 (blue curve). Thus, from time 10 onward, the switched RW1 will evolve exactly like RW2, but with a delay of 5 time units, i.e. a pure shift (e.g. orange curve).

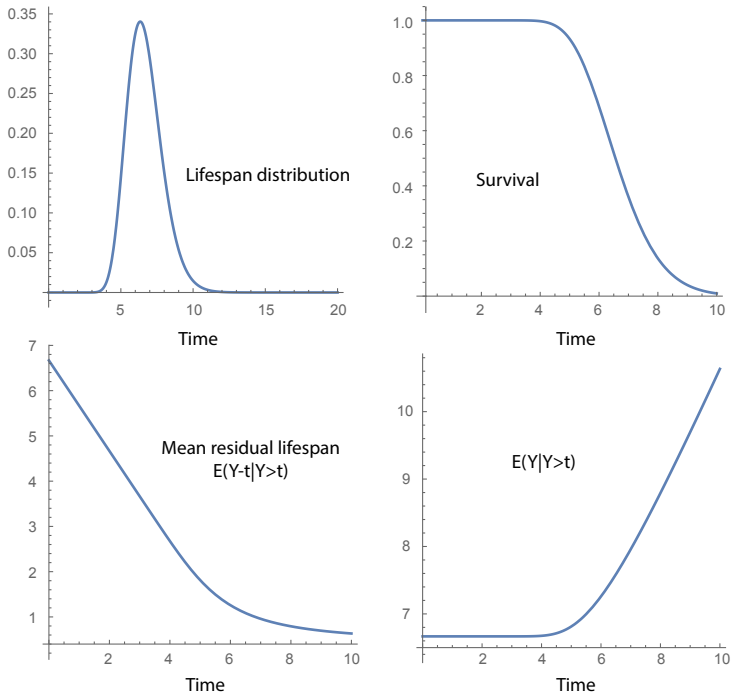
4.2 Definitions

Throughout this Note, Y denotes the lifespan random variable. To predict the consequences of transient interventions on mean lifespan, we need the notion of a conditional mean lifespan. The condition will be used in section 4.4 to represent the switching time. The mean *residual* lifespan (mrl) at time t, $E(Y-t\,|\,Y>t)$, is the expected remaining lifespan of an individual that is alive at time t. Since the expectation of the (linearly) transformed random variable $Y-t$ is the expectation of Y minus t, we can state:

$$E(Y-t\,|\,Y>t) = E(Y\,|\,Y>t) - t. \tag{16}$$

where $E(Y\,|\,Y>t)$ is the conditional expectation of lifespan (computed at age zero) given survival at t: $E(Y\,|\,Y>t) = \int_t^\infty \tau l(\tau)d\tau/S(t)$, where $l(t)$ is the probability density function and $S(t)$ is the survival function. To declutter notation, let us write $\bar{l}(t)$ for $E(Y-t\,|\,Y>t)$. Fig. 4-2

illustrates these notions graphically using the drift-diffusion process (section 6.2) as the underlying aging model.



Supplementary Figure 4-2: Mean residual lifespan at time t, $\bar{l}(t)$, and conditional expected lifespan $E(Y \mid Y > t)$ according to the diffusion-drift aging process with lifespan distribution (31) and survival (32).

4.3 No switching

First we consider the scenario in which individuals W_1 and W_0 are held throughout life at environmental conditions C_1 and C_0, respectively. (We shall use the indices 0 and 1 to refer to these conditions.) If we assume that scaling (equation 2) holds, then at time t, W_1 will have covered t/λ time units in its aging process compared to W_0, who has covered t time units. With this assumption, the mrl functions $\bar{l}_1(t)$ and $\bar{l}_0(t)$ of W_1 and W_0, respectively, can be related to each other:

$$\bar{l}_1(t) = \frac{\int_t^\infty \tau\, l_1(\tau) d\tau}{S_1(t)} - t = \frac{\int_t^\infty \tau/\lambda\, l_0(\tau/\lambda) d\tau}{S_0(t/\lambda)} - t = \frac{\int_{t/\lambda}^\infty x\, l_0(x) \lambda dx}{S_0(t/\lambda)} - t \tag{17}$$

$$= \lambda \left[\frac{\int_{t/\lambda}^\infty x\, l_0(x) dx}{S_0(t/\lambda)} - \frac{t}{\lambda} \right] \tag{18}$$

$$= \lambda \bar{l}_0(t/\lambda) \tag{19}$$

where, as before, $l_i(t)$, $i = 0, 1$ are the probability density functions and $S_i(t)$, $i = 0, 1$ the survival functions. We make use of temporal scaling in the second equality; in the third we make the variable transform $x = \tau/\lambda$; in the fourth we pull out λ. We therefore derive a scaling relation between mean residual lifespans: $\bar{l}_1(t) = \lambda \bar{l}_0(t/\lambda)$.

We can also calculate the difference in mean residual lifespan, $\Delta(t)$, between the populations

always living at C_0 and C_1:

$$\Delta(t) = \bar{l}_1(t) - \bar{l}_0(t) = \lambda\bar{l}_0(t/\lambda) - \bar{l}_0(t) = \lambda\left(E_0(Y \mid Y > t/\lambda) - t/\lambda\right) - E_0(Y \mid Y > t) + t$$
$$= \lambda E_0(Y \mid Y > t/\lambda) - t - E_0(Y \mid Y > t) + t \quad (20)$$
$$= \lambda E_0(Y \mid Y > t/\lambda) - E_0(Y \mid Y > t) \quad (21)$$

$\Delta(t)$ is pictured in Fig. 4-3 (left panel) for the case considered in Fig. 4-2. For early times t when the expected conditional lifespan does not effectively change, i.e. before any deaths occur in either population, we can write $E_0(Y \mid Y > t/\lambda) \approx E_0(Y \mid Y > t) \approx E_0(Y)$ and thus approximate the difference in mean remaining lifespan between the two populations by the simple relationship $\Delta(t) \approx (\lambda - 1)E_0(Y)$. For early times, $\Delta(t)$ is independent of t: $\Delta(t) = E_1(Y) - E_0(Y)$, and thus $E_1(Y) = \lambda E_0(Y)$, which is but the scaling relation for the mean life expectancy. This is illustrated on the left panel of Fig. 4-3 for the case $\lambda = 2$.

4.4 Switching

Now consider the scenario where W_1 is switched at time τ from condition C_1 to the condition C_0 at which W_0 is held. At τ, W_1 will have progressed in its lifespan by τ/λ relative to W_0, but after the switch both are progressing along the same process with the same speed. In this scenario, the difference in mrl is:

$$\Delta(\tau) = \bar{l}_0(\tau/\lambda) - \bar{l}_0(t) = E_0(Y \mid Y > \tau/\lambda) - \tau/\lambda - E_0(Y \mid Y > \tau) + \tau \quad (22)$$
$$= E_0(Y \mid Y > \tau/\lambda) - E_0(Y \mid Y > \tau) + \tau(1 - 1/\lambda). \quad (23)$$

For early switch time τ, before any individuals have died in either population, $E_0(Y \mid Y > \tau/\lambda) \approx E_0(Y \mid Y > \tau)$. In this case,

$$\Delta(\tau) = \tau(1 - 1/\lambda). \quad (24)$$

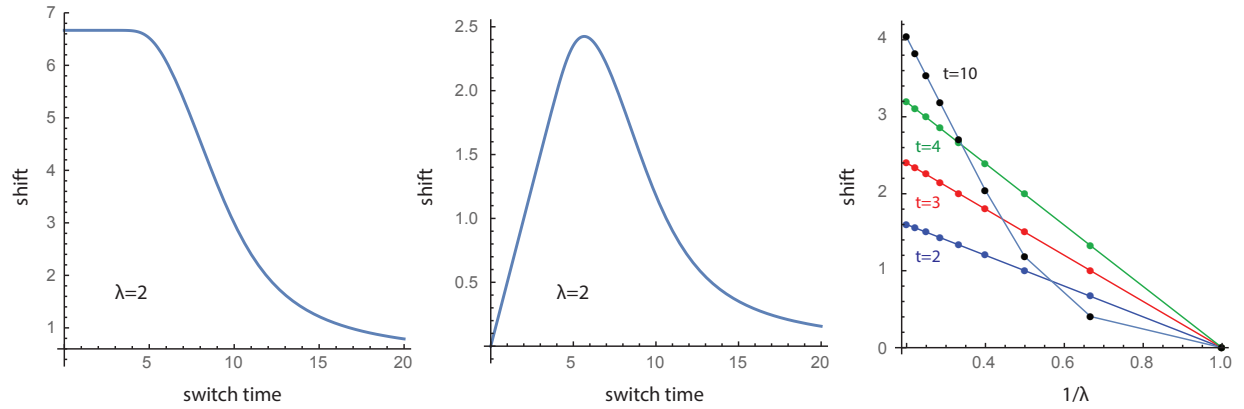
For late times τ, $E_0(Y \mid Y > \tau/\lambda)$ grows like τ/λ and $E_0(Y \mid Y > \tau)$ like τ. Thus $\Delta(\tau)$ will converge to $\tau/\lambda - \tau + \tau - \tau/\lambda = 0$. Indeed, a situation in which the population on condition C_0 is switched after everyone in it has died will be indistinguishable from one in which no switching occurred at all. In that case, the mean remaining lifespans are not related by a shift, but by scaling.

4.5 Predictions

Equation (24) makes three predictions: (1) interventions early in life produce temporal shifts in lifespan distributions; (2) the magnitude of the shift increases proportionally with the time τ at which the switch occurs; and (3) the magnitude of the shift is inversely proportional to the scale factor λ that relates the environmental conditions before and after the switch. These predictions are illustrated in supplementary Fig. 4-3. All three predictions depend on the temporal scaling assumption made in the first step of equation (23). Since they appear to be in excellent agreement with the data presented in Fig. 3 of the main text and the Extended Data Item 5, we conclude that temporal scaling is an appropriate characterization of the effect that temperature has on *C. elegans* aging.

4.6 The effect of switching on distributions

The general analysis of switching experiments in the previous section did refer to mean residual lifespan. The scale and shift factors we calculated for the mean residual lifespan also hold for the residual lifespan *distribution* by virtue of our assumption of temporal scaling. However, it is



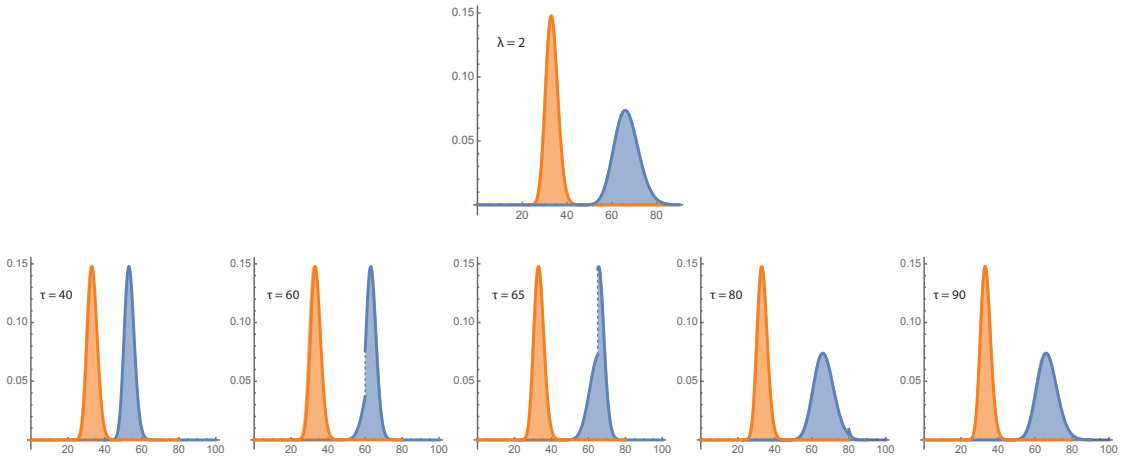
Supplementary Figure 4-3: Mean residual lifespan differences as a function of switch time τ (left and middle) and λ (right). Left panel: $\bar{l}_1(\tau) - \bar{l}_0(\tau)$ vs τ, without switching. This is the difference in mean residual lifespan at τ of W_1 always at condition C_1 and W_0 always at condition C_0. Middle: $\bar{l}_0(\tau/\lambda) - \bar{l}_0(\tau)$ vs τ, with switching from condition C_1 with $\lambda = 2$ to condition C_0 at time τ. Right: Difference in mean residual lifespan when switching occurs at the times indicated ($\tau = 2, 3, 4, 10$) vs $1/\lambda$, where λ is the scale factor of condition C_1 relative to C_0. The slope is $-\tau$, and the intercept is τ, when τ is small enough that $E(Y|Y > \tau) = E(Y|Y > \tau/\lambda) \approx E(Y)$. If λ is very large, such that $E(Y|Y > \tau) \approx \tau$ but $E(Y|Y > \tau/\lambda) \approx E(Y)$, the intercept becomes $E(Y)$ (and the slope is still $-\tau$). This is the case when λ is such that one can transfer W_1 before any member of its population, P_1, has died, but after all members of the population P_0 (represented by W_0) have died. The underlying aging model is the drift-diffusion process with lifespan distribution (31) and survival (32); section 6.2. See Fig. 4-2 for the survival function.

useful to numerically illustrate the situation for the full lifespan distribution.

For the purpose of illustration we use a simple drift-diffusion aging model. Consider two populations P_1 and P_0 that age according to equation 30 with parameters $C_1 = \{\mu_1, \sigma_1^2\}$ and $C_0 = \{\mu_0, \sigma_0^2\}$, respectively. We switch P_1 from C_1 to C_0 at time τ. Let the lifespan distribution of the shifted population be $p_{1_{\overset{\tau}{\to}0}}(X(t) = 0 \,|\, X(0) = c)$:

$$
\begin{aligned}
&p_{1_{\overset{\tau}{\to}0}}(X(t) = 0 \,|\, X(0) = c) = \\
&= \begin{cases} \int_0^\infty p_0(X(t) = 0 \,|\, X(\tau) = v) p_1(X(\tau) = v \,|\, X(0) = c)\, dv & \text{if } t \geq \tau \\ p_1(X(t) = 0 \,|\, X(0) = c) & \text{if } t < \tau \end{cases} \qquad (25) \\
&= \begin{cases} \int_0^\infty p_0(X(t-\tau) = 0 \,|\, X(0) = v) p_1(X(\tau) = v \,|\, X(0) = c)\, dv & \text{if } t \geq \tau \\ p_1(X(t) = 0 \,|\, X(0) = c) & \text{if } t < \tau \end{cases} \qquad (26)
\end{aligned}
$$

In (26) we made use of the fact that the process is time-homogeneous. $p_0(X(t-\tau) = 0 \,|\, X(0) = v)$ and $p_1(X(t) = 0 \,|\, X(0) = c)$ are given by (31) and $p_1(X(\tau) = v \,|\, X(0) = c)$ by (39). The integral can be expressed in "closed" form using error functions, but the lengthy expression is not particularly informative to write down explicitly. It is more useful to only present the result graphically, Fig. 4-4.



Supplementary Figure 4-4: Two populations P_1 (blue) and P_0 (orange) that are held at constant conditions C_1 and C_0, respectively, yield lifespan distributions that differ by a scaling factor (here $\lambda = 2$, blue relative to orange). In the second row, P_0 (a high-T population, say) stays at high T all the time, while P_1 is switched from low T to high T in different simulated experiments at times $\tau = 40, 60, 65, 80, 90$. The blue distribution is the resulting lifespan distribution for all of P_1. If the change in T occurs before any member of P_1 has died (up to about $\tau = 50$ in this example), one obtains a shift in the distribution relative to P_0, as seen in the case of $\tau = 40$. When the switch occurs at $\tau = 90$ (rightmost), there is nothing to switch, because all members of P_1 have already died, yielding P_1's original lifespan distribution—a scaling relative to P_0. In between, P_1's distribution will be piecewise—one piece resulting from any deaths that have occurred at low T before P_1 was switched and the other piece resulting from the deaths that occurred at high T after the switch. The piece of P_1's lifespan distribution at high T is a shifted version of the corresponding piece of P_0, while the piece coming from low T is a scaled version. The shift factors are the same as calculated for the mean residual lifespan case, $\tau(1 - 1/\lambda)$, regardless of the specific aging model.

Note 5 — Temporal scaling in phenomenological causal models

Summary

A mechanistic grounding of temporal scaling is hard to achieve at present, because little is known about the causes of death in *C. elegans* and the processes that lead up to them (such as bacterial pathogenesis [18] or germ-line proliferation [19], both of which seem to kill late in life). In this section we explore the conditions for temporal scaling in the context of an aging process that renders animals increasingly susceptible to the proximate causes that kill them. We analyze two types of phenomenological models. One type explores the failure of systems consisting of independent parts ("competing risks" model); the other explores the failure of networks consisting of interdependent parts (dependency network model). We numerically explore the constraints that each type of model places on temporal scaling, the main empirical phenomenon we observe in *C. elegans*. The dependency model in particular illustrates the concept of an emergent state variable whose dynamics determines lifespan. It also suggests how interventions that affect only a part of a system nonetheless impact the dynamics of the overall state variable in a fashion that produces temporal scaling. The two models we discuss here can be seen as quantitative versions of the schematics shown in Extended Data Item 9.

Contents of Note 5

5.1 Competing risk models

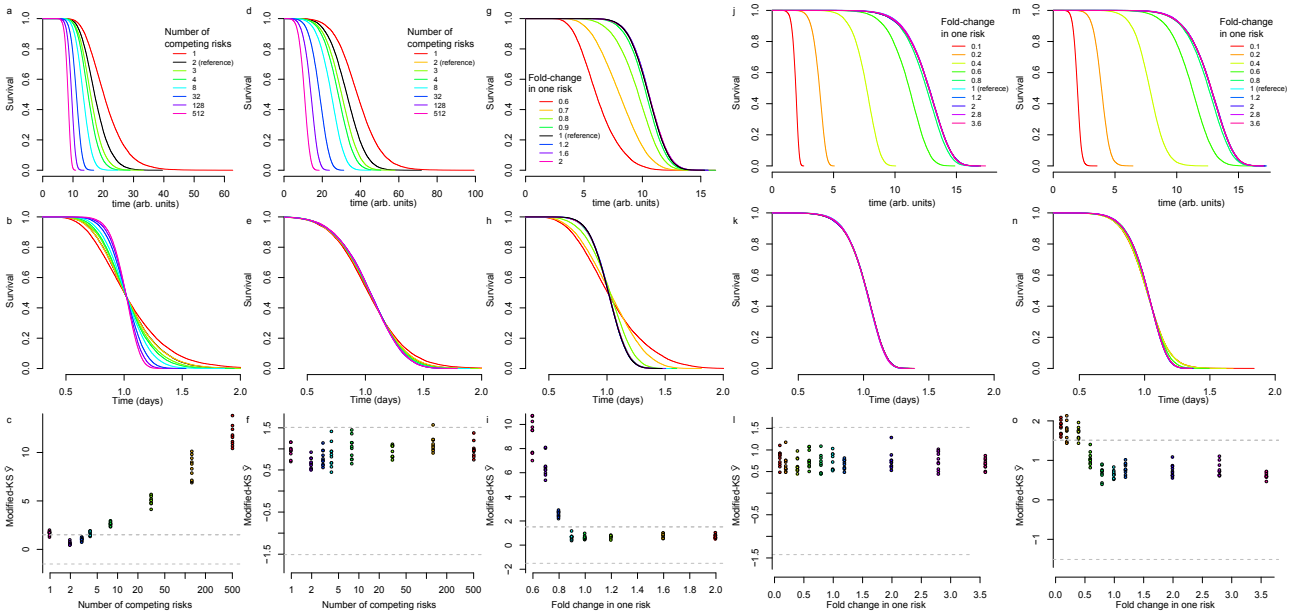
We first consider the possibility that animals might die from one of several possible causes that occur independently of one another. In the statistics literature, this assumption forms the basis of competing risks models [20]. In reliability theory, this assumption corresponds to an architecture in which systems rely on multiple critical components that operate in series, i.e. a "weakest-link" model. Both views posit that individuals/systems remain alive as long as no component has failed, and, conversely, that the first component to fail causes the entire system to fail. If each component i has survival function $S_i(t), i = 1, \ldots, n$, where n is the number of components, and hazard function $h_i(t), i = 1, \ldots, n$, the survival and risk functions of the whole system, $S(t)$ and $h(t)$, respectively, will be:

$$S(t) = \prod_{i=1}^{n} S_i(t) \qquad \text{and} \qquad h(t) = \sum_{i=1}^{n} h_i(t) \tag{27}$$

Obviously, systems made of strong components (i.e. components with greater mean lifespans) will fail later than systems made of weak components. It is also evident that systems with fewer components will last longer, all else being the same, than systems with a greater number of components, as each additional component introduces an additional possibility of failure. A component that is substantially weaker than all the others ($S_1(t) \ll S_i(t)$ for all t and all $i > 1$) will dominate the failure statistics of the system: $S(t) \approx S_1(t)$.

We explored the impact of parameters of the competing risks model with numerical simulations, summarized in Fig. 5-1. First, we considered the simple case where all components have the same distribution of failure times $S_i(t)$, specifically a log-normal distribution (Fig. 5-1a). In this case, the system's behavior conforms with Extreme Value Theory (EVT) for minima bounded below by zero (as is the case with lifespans). A central result of EVT is that the survival distribution $S(t)$ of a system consisting of a large number of components with independently and identically distributed (iid) failure time distributions will take the form of a Weibull distribution [21]. The Weibull distribution will arise regardless of the specific form of the probability density of failure describing the components.

In simulations, we confirmed that when the total number of components is small, increasing the number of components will shorten lifespan in a manner that deviates significantly from temporal scaling (Fig. 5-1b, c). Such a change is therefore inconsistent with our experimental results. When the total number of components is large, changes in the number of components have little effect on lifespan. This again disagrees with our experimental findings, where we show that substantial lifespan increases occur by temporal scaling. Additionally, any intervention that breaks the iid assumption will be inconsistent with our experimental results: Even with large numbers of components (i.e. independent death causes), for example $n = 50$, we observe that altering the shape of the risk of one single cause shortens lifespan in a manner that deviates from temporal scaling (Fig. 5-1g–i). This will be true in particular for any system whose competing death causes are not iid to begin with. Were any intervention to emphasizes one cause over any other, it would distort the overall lifespan distribution accordingly. In summary, we conclude that



Supplementary Figure 5-1: Independent competing risks. **a:** We presume a series of independent death causes (vital components) with log-normal distribution. Reducing the number of causes from two (black) to one (red) extends lifespan, while increasing the number of causes decreases lifespan (green-purple). **b:** As seen from the AFT residuals, altering the number of causes, yields a change in survival shape. **c:** The generalized KS-test confirms the significance of these changes; values of $|\bar{\mathcal{Y}}|$ greater than 1.51 (gray line) correspond to a significant difference in shape ($p < 0.01$). **d:** The same model as in panel **a** but all components failing according to a Weibull distribution with a frailty correction chosen to match experimental data. As before, changing the number of components will increase or decrease the mean lifespan (panel **e**), but with little effect on the survival curve shape (panel **f**). **g:** Assuming again a log-normal failure time distribution for each component and fixing the number of components at 50, we alter the μ parameter of a single component, changing its mean lifespan while keeping its coefficient of variation fixed. Changing μ 0.9-fold or less shortens the lifespan of affected populations (green-red), but increasing it 1.2-fold or more has little effect, as the remainder of shorter-lived components dominate. **h–i:** Decreasing the lognormal μ of a single component by 0.8-fold or less produces statistically significant deviations from perfect temporal scaling. **j–l:** If component failure follows a Weibull distribution with no frailty correction, decreasing the time scale β of a single component changes lifespan without producing any deviation from temporal scaling. **m–o:** When a frailty term is incorporated to allow the reference population to better match *C. elegans* data, significant deviations from temporal scaling can be seen at large effect sizes if the frailty term σ is not changed in proportion to the scale factor β. See text for details.

to reconcile our experimental data with a competing risks model requires that interventions alter the risk profiles associated with all independent causes of death in exactly the same manner.

There is one exception to the iid assumption, however. If the failure risks of all components happened to be Weibull risks, then an intervention that affected only the time scale parameter of a subset of such components would still exhibit temporal scaling at the systems level, (Fig. 5-1j–l), which would remain consistent with our experimental data showing scaling. Incorporating a frailty term so that the reference population more closely matches our empiric data requires that the frailty term σ^2 also be scaled in proportion with the Weibull β, or else significant deviations

from temporal scaling are produced (Fig. 5-1m–o; Note 6.4). Altering the number of components would produce effects consistent with our experimental data (Fig. 5-1d–f). All this reflects the fact that the sum of Weibull risks that only differ in their time scale parameters is still a Weibull risk. With regard to interpreting our experimental results, a component Weibull scenario would not be very helpful, as it still requires an explanation of why each system component exhibits an extreme value distribution (Weibull). This distribution arises from a large number of *independent* components that exhibit *identically* distributed failure times. It is not obvious why all biological risk factors should meet such criteria. Moreover, it is unclear why an intervention on any so-structured risk factor would affect only its time scale parameter. Finally, while lifespan-shortening interventions might be consistent with our results, any substantial increase in average lifespan would require that the risk profiles of many components be rescaled by the same intervention.

We believe the competing risks case to be overly simplistic. Indeed, for interventions to produce temporal scaling within a competing risks framework they would have to affect the likelihood of all death causes to the same extent. This would be the case, however, if there was a physiological state variable of the kind we postulate, whose stochastic decline determined lifespan.

5.2 Interdependency networks

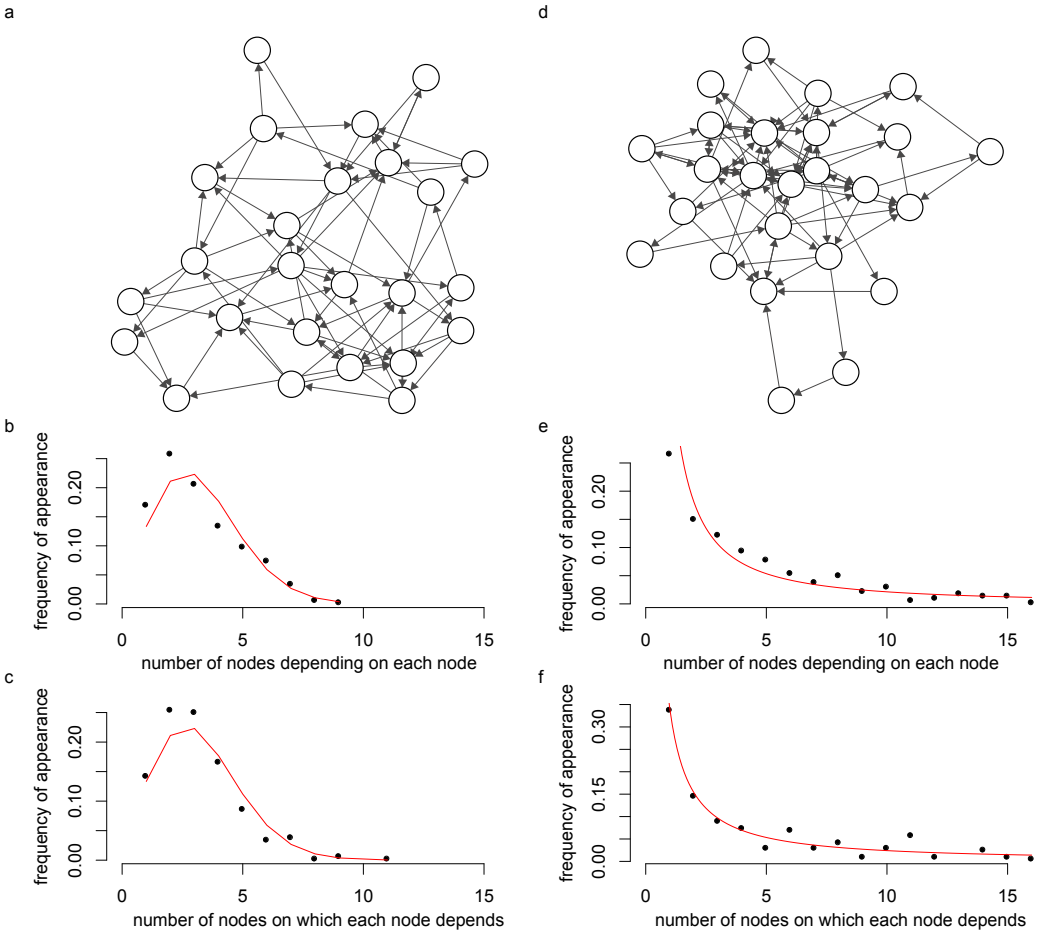
It seems unlikely that *C. elegans* die from causes that are independent. Interdependence is an essential aspect of life, as organ systems, tissues, cellular components and biochemical pathways all depend on each other. Although we do not understand these dependencies at a detail that allows us to suggest mechanistic models, progress might be made at a more abstract level by considering a recent aging model based on a dependency network [22].

In the Vural-Morrison-Mahadevan (VMM) model [22], nodes (representing functional units like cells, metabolic networks, organ systems) provide each other with physiological functions. A node has two states: functional or failed. A node fails with an intrinsic failure probability and, additionally, when a majority of nodes it depends upon have failed. This interdependency is represented as a set of directed edges that weave the nodes into a network. An arrow $i \rightarrow j$ means that node i contributes to maintaining the functionality that node j provides. Nodes may also have an intrinsic repair probability. The network fails as a whole when less than a specified fraction of nodes remains functional. The time to failure is the lifespan of an individual (network).

The VMM model requires that networks are initially connected, meaning that each node must depend on at least one other node, and each node must have at least one dependent. A node is subject to spontaneous (intrinsic) failure with a constant probability. Failed nodes can be repaired, again with a constant probability. A node experiences an extrinsic failure when more than half the nodes it depends upon have failed. This latter mechanism can cause networks to collapse by cascading failures. The entire network represents an individual. It fails when 99% of its nodes have failed. The model behavior is not sensitive to this threshold, as most networks collapse completely once more than 50% of nodes have failed.

AFT regression models and KS tests both assume that death times are not discretized. In order to use these statistical tools to analyze the results of simulations, we implemented the VMM model as a continuous-time Monte Carlo simulation [23]. We confirmed (data not shown) that our continuous-time version matched the behavior of the discrete version in [22].

Vural et al. demonstrate that specific parameter sets can cause interdependency models to yield hazard functions that correspond closely to empiric estimates of *C. elegans* lifespan distributions. Here, we do not address whether the VMM model fits our empiric data, because in Note 1.4 we show that a broad class of functions can provide reasonable fits to experimental data. Instead, we



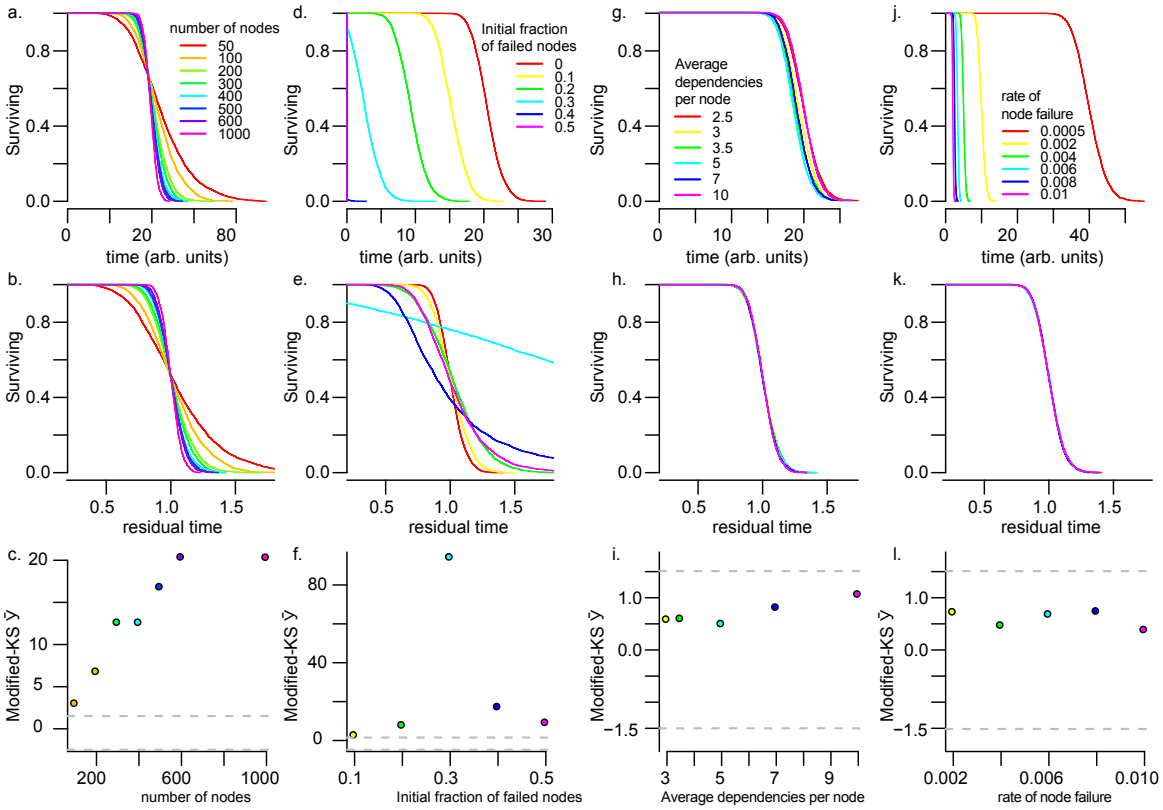
Supplementary Figure 5-2: Interdependency networks. **a, d:** Vural *et al.* [22] describe dependency networks with random and scale-free connectivity. See text for details. **b–c:** Random dependency networks exhibit a binomial distribution of incoming edges (number of neighbors on which any given node depends) and outgoing edges (number of neighbors that depend on any given node). **e–f:** Scale-free networks exhibit a power-law distribution of incoming and outgoing edges.

examine VMM models to determine which changes to network architecture and failure dynamics, if any, are capable of producing a temporal scaling of lifespan distributions.

5.2.1 Homogeneous interventions

We find that most changes to VMM network architecture and failure dynamics, including changes to the number of nodes, the initial number of failed nodes, or the number of dependencies per node, either do not produce a substantial effect on lifespan or alter lifespan distributions in a manner not consistent with temporal scaling (Fig. 5-3). In contrast, changes to the rate of intrinsic failure or repair of all nodes by the same factor produced a temporal rescaling of mortality statistics (Fig. 5-3). Regardless of how interdependent the nodes might be, a universal change of intrinsic failure rates simply speeds up or slows down the overall system dynamics. This temporal rescaling of system dynamics then produces a temporal rescaling of mortality statistics.

These results were consistent across both random and scale-free network architectures, and across multiple random permutations of network dependencies.

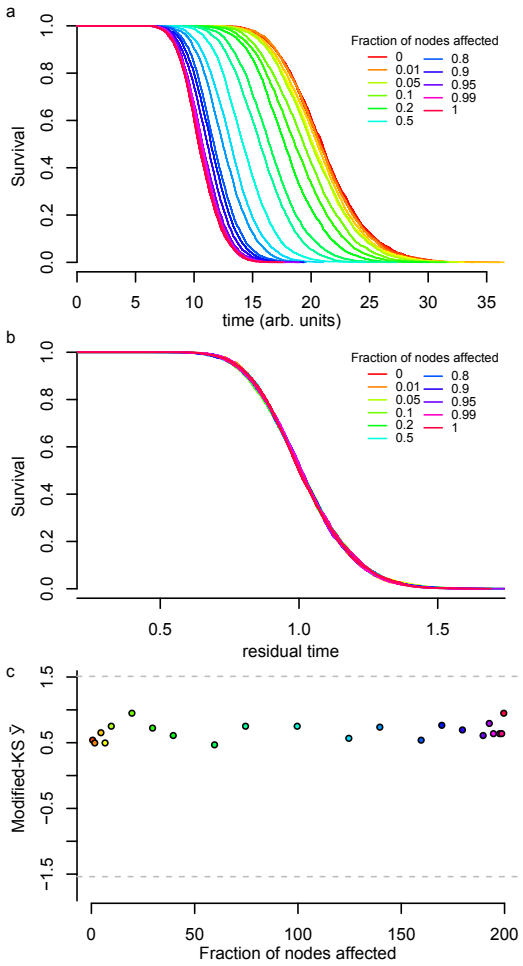


Supplementary Figure 5-3: Temporal scaling in interdependency networks. We recapitulate the perturbations by Vural *et al.* [22] for the purpose of studying their effect on temporal scaling. Populations were simulated by generating $1,500$ random networks of the same size with the same underlying distribution of node degrees and with all nodes sharing the same rate parameters. Each network underwent a continuous-time Monte Carlo analog of the dynamics defined in Vural *et al.* . Nodes fail (or are repaired) with a constant probability and subject to the majority rule described in the text. The lifespan of the network is the time it takes for all nodes to fail. **a:** Changing the number of nodes that constitute individuals did not alter the average lifespan of populations. **b:** However, it does alter the characteristic shape of population survival curves as can be see in he AFT-resisuals. **c:** All changes to the network size yielded a KS test statistic greater than 1.51 (grey lines), corresponding to a statistically significant ($p < 0.01$) deviation from perfect temporal scaling. **d:** Increasing the fraction of (randomly chosen) nodes that have failed at the outset did shorten the average lifespan of the population. **e–f:** Yet, it produced significant deviations from perfect scaling. **g:** Increasing the interdependence within networks only marginally increased lifespan. **h–i:** This intervention showed no significant deviation from temporal scaling. **j:** Changing the net failure rate of each node altered lifespan. **k–l:** As expected, such a homogeneous intervention produced perfect temporal scaling.

5.2.2 Heterogeneous interventions

We asked whether modifying the failure rate of only a subset of nodes would break temporal scaling. Such a change might mimic interventions in *C. elegans* that act only on a subset of lifespan determinants. We found that a two-fold increase in the failure rate of any fraction of

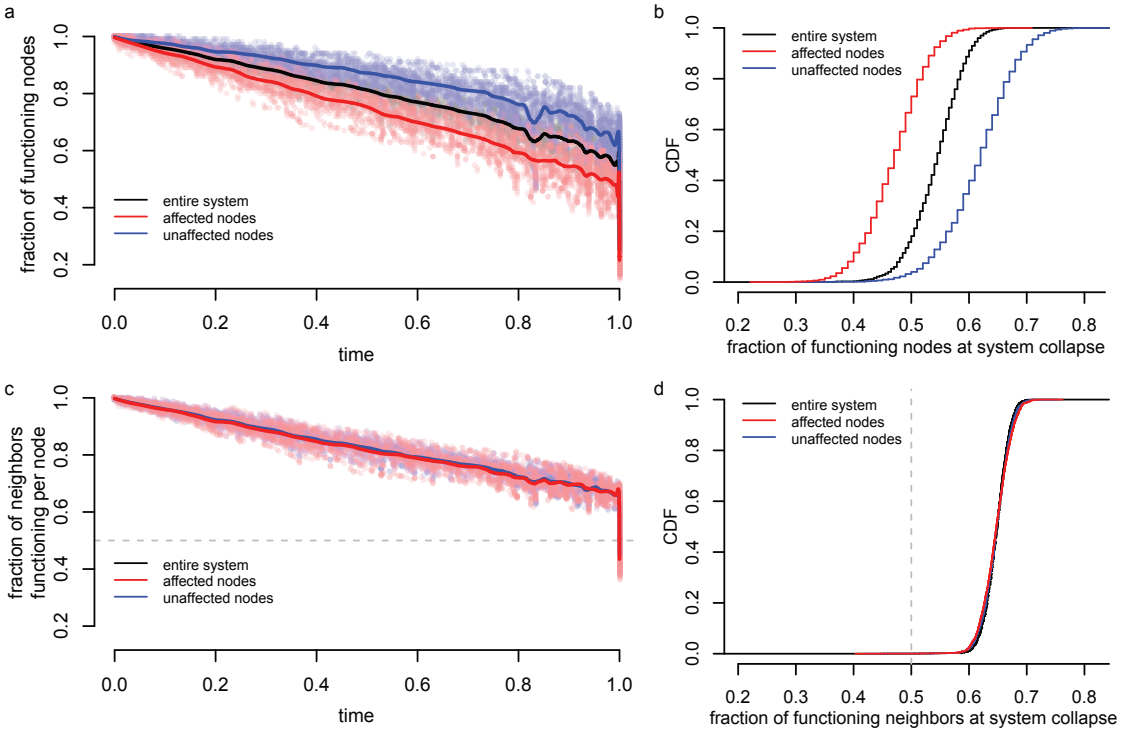
nodes altered lifespan distributions in a manner consistent with temporal scaling, with the magnitude of the scale factor depending on the fraction of nodes affected (Fig. 5-4). These results were consistent across both random and scale-free network architectures, and across multiple random permutations of network dependencies.



Supplementary Figure 5-4: Scaling of heterogeneous interventions in dependency networks. We investigated the effect of interventions that alter only a subset of all nodes. In populations of $1,500$ random networks, we increased the failure rate of a subset of nodes, thus "weakening" this subset. **a:** Not surprisingly, increasing the fraction of weakened nodes decreased network lifespan. **b, c:** Very surprisingly, the AFT residuals and the generalized KS test statistic indicated no significant deviations from temporal scaling ($p < 0.01$, gray lines)

We sought to explain this surprising result using the model introduced in the main text: that diverse interventions produce temporal scaling because they influence a single system-wide state variable that determines the risk of death. In VMM networks, networks collapse when the fraction of nodes still operational (henceforth the operational node fraction) drops below a minimum threshold. At the start of each simulation, this operational node fraction begins high and then exhibits an almost linear decline for a long period, decreasing steadily until a catastrophic failure [22] occurs. The operational node fraction, therefore, makes an obvious candidate for a state variable with which to describe the risk of death (network collapse). A reliability model [24] posits that aging is a decrease in the number of redundant system

components over time; perhaps older *C. elegans* have fewer functioning parts.



Supplementary Figure 5-5: State variables in dependency networks. **a:** The panel shows the dynamics of the fraction of functioning nodes (operational node fraction) in the entire system (black), the weakened subnetwork (red) and the non-weakened subnetwork (blue). All times were plotted as a fraction of lifespan, allowing the results of 500 simulations to be overlaid (colored circles) and summarized by a LOESS regression model (lines). **b:** As in panel a, but the quantity shown is the fraction of functional links per node (operational link fraction). All three subnetworks are indistinguishable with regard to this quantity. **c,d:** To better characterize the diversity within a population of the quantities tracked in panels a and b, we collected their value immediately before system collapse, leading to the death of the individual network, and plotted their cumulative distribution function.

Despite its intuitive appeal, we find that the operational node fraction is not the best state variable with which to describe the dynamics of VMM network aging. In Fig. 5-5, we show individual trajectories of the state variable that correspond to the lifespan distributions in Fig. 5-4, where half of the nodes operate with a two-fold increased failure rate. We find that the operational node fraction calculated specifically for the subsets of "weak" and "strong" nodes diverge over time. The operational node fraction of the entire network, as a state variable describing network dynamics, does not capture this divergence. Furthermore, though all networks exhibited a low operational node fraction immediately prior to network collapse, this fraction was highly variable, limiting its predictive value with regard to the timing of network collapse.

A different systemic property, the average fraction of dependencies remaining operational across all nodes (henceforth the operational link fraction), appears to be better suited. In VMM networks, a node fails when the fraction of nodes it depends on drops below 50% of its initial value. The operational link fraction might therefore be a useful state variable with which to characterize the rate of dependency-induced failure events in a network. We find that the

operational link fractions calculated for the subsets of "weak" and "strong" nodes do not diverge over time (Fig. 5-5d). While "weak" and "strong" nodes differ in their intrinsic failure probabilities, they are equally likely to have "weak" or strong" neighbors and therefore both groups must exhibit equivalent operational link fractions at all times. This makes for a systemic property that is consistent across all subnetworks. It also exhibits much less variability immediately prior to network collapse than the operational node fraction. This suggests that the operational link fraction is more predictive of lifespan than the operational node fraction (Fig. 5-5d).

5.3 Conclusions

In this study of dependency networks, we accomplish three tasks. First, we provide an example of how a single state variable, such as the operational link fraction, can describe the state of a complex system and predict its risk of collapse. Second, we illustrate that such a state variable can have a simple dynamics governed by a single time scale despite emerging from a complex network of heterogeneous components. Finally, we highlight one way in which interventions acting only on a subset of a system might produce a temporal scaling in the mortality statistics.

In our simulations of interdependency networks, temporal scaling phenomena arise from the basic assumption that nodes remain operational as long as *any* 50% of their dependencies remain operational. Because the dependencies of a node are fungible and contribute equivalently to a node's operation, the consequences of local perturbations can be distributed across the whole network. We do not know how such a homogeneity might arise in *C. elegans*.

Note 6 — Vitality/Resilience models based on diffusion

Summary

In this section, we explore how our experimental findings relate to a specific type of stochastic process often invoked to describe aging phenomenologically: random walks and diffusion in a physiological state space [1, 15]. In the context of these models, we explore the constraints that temporal scaling places on the effects of interventions on model parameters. A concrete process model also enables us to explore the sensitivity of temporal scaling to violations of these constraints and to gain intuition on how large populations would have to be to detect such violations with the statistical tools that we used to analyze our data.

Contents of Note 6

6.1 Semi-infinite random walk

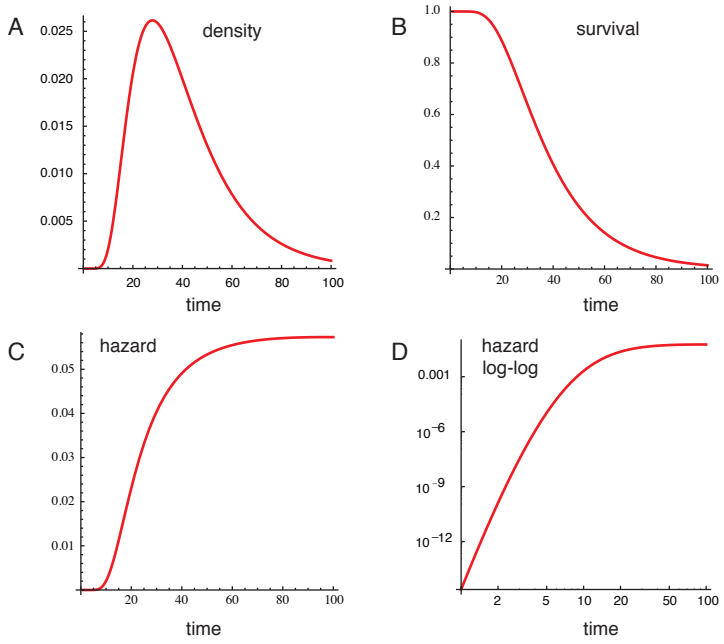
The aging process $X(t)$ considered here is continuous in time t and has infinitely many discrete states $i = 0, 1, 2 \ldots$ with an absorbing boundary at 0. At time $t = 0$ the random walker starts in state i: $X(0) = i$. The probabilities of moving up or down the state space are β_1 and β_{-1}, respectively, independent of i. If $\beta_{-1} > \beta_1$, absorption occurs with certainty.

The lifespan distribution (probability density of absorption times), $p(X(t) = 0 \,|\, X(0) = i)$ is given by [1, 25] (Fig. 6-1):

$$p(X(t) = 0 \,|\, X(0) = i) = i \left(\frac{\beta_{-1}}{\beta_1}\right)^{\frac{i}{2}} \frac{1}{t} \exp(-(\beta_{-1} + \beta_1)t)\, I_i(2t\sqrt{\beta_{-1}\beta_1}),\ t > 0 \tag{28}$$

with $I_i(x)$ the modified Bessel function (of the first kind) of order i, for which an integral representation is $I_i(x) = (1/\pi) \int_0^\pi \exp(x \cos\theta) \cos(i\theta) d\theta$.

To rescale time in the survival curve requires multiplying *each* rate parameter β_{-1} and β_1 with the same factor λ^{-1}. Consider the survival function $S(t)$:



Supplementary Figure 6-1: A: Lifespan density: $l(t) \equiv p(X(t) = 0 \,|\, X(0) = 20)$ with $\beta_{-1} = 1.5$ and $\beta_1 = 1.$ as per equation (28). **B:** Survival: $S(t) \equiv 1 - \int_0^t p(X(\tau) = 0 \,|\, X(0) = 20) d\tau$. **C:** Hazard: $l(t)/S(t)$. The hazard decelerates (in fact, asymptotes to a constant) because the process attains a quasi-stationary distribution over the states $j > 0$, $q_j = \lim_{t\to\infty} p(X(t) = j \,|\, X(t) > 0, X(0) = i)$, with $q_1 = (1 - \sqrt{\beta_1/\beta_{-1}})^2$ and $q_j = q_1\, j\, (\beta_1/\beta_{-1})^{(j-1)/2},\ \forall j > 1$. As $t \to \infty$, the hazard asymptotes to $h(\infty) = \beta_{-1} q_1 = (\sqrt{\beta_{-1}} - \sqrt{\beta_1})^2$, which, for the values used in this example, is $.0.0505\ldots$ **D:** Log-log plot of **C**. If we were to multiply β_{-1} and β_1 by 2, the survival curve **B** would look exactly the same up to a rescaling of the time axis by $1/2$.

$$S_1(t) = 1 - \int_0^t p(X(\tau) = 0 \,|\, X(0) = i) d\tau \tag{29a}$$

$$= 1 - i \left(\frac{\lambda^{-1}\beta_{-1}}{\lambda^{-1}\beta_1} \right)^{\frac{i}{2}} \int_0^t \frac{1}{\tau} \exp(-(\beta_{-1} + \beta_1)\lambda^{-1}\tau)\, I_i(2\sqrt{\beta_{-1}\beta_1}\lambda^{-1}\tau) d\tau \tag{29b}$$

integration by substitution: set $\tau' = \lambda^{-1}\tau$, $d\tau' = \lambda^{-1} d\tau \ldots$

$$= 1 - i \left(\frac{\beta_{-1}}{\beta_1} \right)^{\frac{i}{2}} \int_0^{\lambda^{-1}t} \frac{1}{\lambda\tau'} \exp(-(\beta_{-1} + \beta_1)\tau')\, I_i(2\sqrt{\beta_{-1}\beta_1}\tau')\lambda d\tau' \tag{29c}$$

$$= 1 - i \left(\frac{\beta_{-1}}{\beta_1} \right)^{\frac{i}{2}} \int_0^{\lambda^{-1}t} \frac{1}{\tau} \exp(-(\beta_{-1} + \beta_1)\tau)\, I_i(2\sqrt{\beta_{-1}\beta_1}\tau) d\tau \tag{29d}$$

$$= S_0(\lambda^{-1}t) \tag{29e}$$

Likewise for the probability density (28):
$p_1(X(t) = 0 \,|\, X(0) = i) = \lambda^{-1} p_0(X(\lambda^{-1}t) = 0 \,|\, X(0) = i)$.

6.2 Drift-diffusion process

The random walk with continuous state space is a drift-diffusion process with initial value c, drift coefficient $-\mu$ ($\mu > 0$), and diffusion coefficient σ^2:

$$X(t) = c - \mu\, t + \sigma\, W(t), \tag{30}$$

where $W(t)$ is a Wiener process, defined by $W(0) = 0$ and increments that are independent and normally distributed with $\mathrm{E}\{W(s+t) - W(t)\} = 0$ and $\mathrm{Var}\{W(s+t) - W(t)\} = s$. In terms of increments: $P(W(s+t) \in [y, y+dy] \,|\, W(t) = x) = N((y-x)/\sqrt{t}, 0, 1)/\sqrt{t}$, where $N(x, 0, 1)$ is the density of the standard normal distribution. Eqn. (30) is sometimes written as a stochastic differential equation, $dX(t) = -\mu\, dt + \sigma\, dW(t)$, with initial condition $X(0) = c$.

The lifespan distribution $p(X(t) = 0 \,|\, X(0) = c)$ is an inverse Gaussian [1]:

$$p(X(t) = 0 \,|\, X(0) = c) \quad = \frac{c}{\sigma\sqrt{2\pi}} \frac{1}{t^{3/2}} \exp\left[-\frac{(c - \mu t)^2}{2\sigma^2 t} \right]. \tag{31}$$

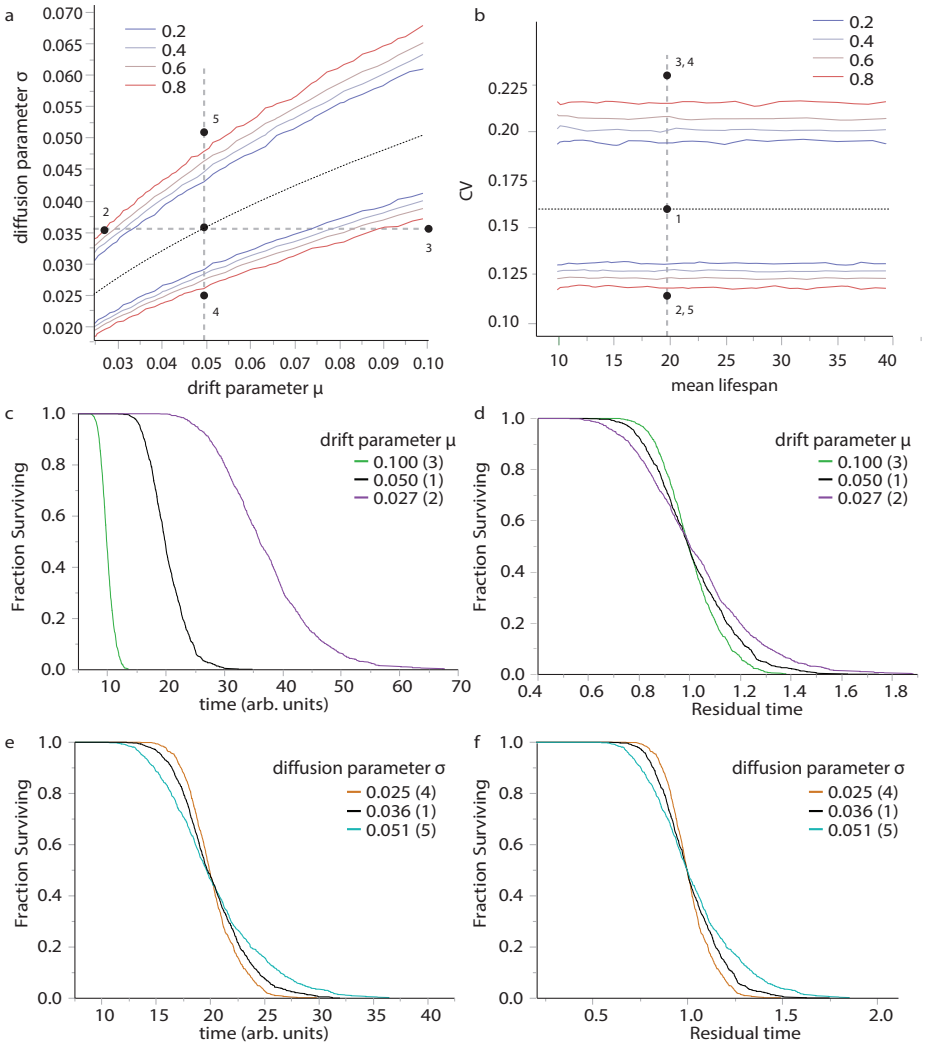
The survival function is $S(t) = 1 - \int_0^t p(X(t) = 0 \,|\, X(0) = c)$:

$$S(t) = \Phi\left(\frac{c - \mu t}{\sigma\sqrt{t}} \right) - \exp\left(\frac{2c\mu}{\sigma^2} \right) \Phi\left(\frac{-c - \mu t}{\sigma\sqrt{t}} \right) \tag{32}$$

with $\Phi(x)$ the cumulative standard normal distribution, $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} \exp(-t^2/2) dt$.

Rescaling time means multiplying μ and σ^2 with the same factor λ^{-1}:

$$S_1(t) = \Phi\left(\frac{c - \mu\lambda^{-1}t}{\sigma\sqrt{\lambda^{-1}t}} \right) - \exp\left(\frac{2c\mu}{\sigma^2} \right) \Phi\left(\frac{-c - \mu\lambda^{-1}t}{\sigma\sqrt{\lambda^{-1}t}} \right) = S_0(\lambda^{-1}t). \tag{33}$$



Supplementary Figure 6-2: Detecting deviations from perfect scaling. **a:** The plot shows the parameter combinations (μ, σ) for which we can detect significant deviation from scaling relative to a reference set $(\mu_0 = 0.049, \sigma_0 = 0.036)$ with a given statistical power (color) for population size $1,000$. See text for details. Contour colors indicate increasing power. The area between the blue (inner) contours, corresponds to parameter combinations with a less than 20% probability, at $p < 0.01$, of detecting a deviation from scaling using our approach. Parameter combinations outside the red (outer) contours have a greater than 80% probability, at $p < 0.01$, of being detected. The dotted curve plots the line along which populations exhibit a coefficient of variation equal to that of the reference population (point 1) **b:** We plot the same results in terms of the coefficient of variation of lifespan vs the mean lifespan $1/\mu$, instead of directly the process parameters. (Scaling implies a constant coefficient of variation.) **c:** The panel shows survival curves corresponding to the marked positions 1 and 2 in panels a–b. **d:** The panel shows the corresponding AFT residuals on which we ran the Kolmogorov-Smirnov test. **e–f:** As in panels c and d, but for the marked positions 3 and 4 in panels a–b.

6.3 Statistical power and sensitivity of scaling in drift-diffusion processes

The drift-diffusion process we describe in section 6.2 exhibits a first passage time distribution that is the so-called inverse Gaussian distribution, equation (31). Without loss of generality, we

can fix $c = 1$. The mean lifespan then is given by μ^{-1}, its variance is σ^2/μ^3, and its coefficient of variation is $\sigma/\sqrt{\mu}$. One can see that, to produce temporal scaling, a change from μ to μ/λ must be accompanied by a corresponding change from σ to $\sigma/\sqrt{\lambda}$.

The extent to which our combined experimental and statistical approach can constrain drift-diffusion models depends on its capacity to detect deviations from scaling, which in turn depends on the size of the changes in each parameter and the size of the populations used to detect them. In other words, the issue is not to detect just changes in, say, μ (a wilcoxon or log-rank will work well even with relatively small populations of a hundred individuals), but to detect *disproportionate* changes in μ relative to a given σ. This requires much larger population sizes. Figure 6-2 shows the results for a population size of $1,000$, which corresponds to our typical experiments.

In Figure 6-2, the parameters μ and σ always refer to the parameters of the drift and diffusion component of the *process*. (The mean and the standard deviation of the inverse Gaussian then are $1/\mu$ and $1/\sigma$.) We generated a reference sample (μ_0, σ_0) of $1,000$ death times sampled from an inverse Gaussian probability density with (process) parameters $\mu = 0.049$ and $\sigma = 0.036$. These parameters were obtained by fitting an inverse Gaussian to our experimental data to obtain realistic parameters. (This parametric form is not the best fit for our experimental data—we prefer a Weibull distribution; but the inverse Gaussian provides a reasonable fit, as seen in Extended Data Item 1a–d.) We then generated sets (μ, σ) of $1,000$ death times each, sampled from inverse Gaussians whose μ and σ parameters were systematically varied over a range ($\mu \in [0.025 - 0.1]$ and $\sigma \in [0.017 - 0.071]$). We compared each parameter setting (μ, σ) to (μ_0, σ_0) using the AFT+KS combination (section 2.1) to calculate a p-value with which each pair (μ_0, σ_0) and (μ, σ) could be said to exhibit a significant deviation from perfect scaling at a significance level 0.01. By generating 750 replicates of each comparison, we determined the fraction of significant p-values, which provides us with an estimate of the statistical power at population size $1,000$, and thus the effect sizes that can be distinguished.

Because we apply the same statistical test to evaluate temporal scaling in both simulated models and experimental data, figure 6-2 can be used to assess our ability to draw strong inferences from our data. The only assumption we make here is that *C. elegans* lifespan distributions can be reasonably approximated by inverse Gaussian distributions.

Given a population of $1,000$ individuals, we ask how much the drift parameter μ of a drift-diffusion process must change, at constant σ, for our statistical approach to reliably detect deviations from perfect scaling at a significance 0.01 and power 0.8. The answer from Fig. 6-2 is that the drift parameter must be altered to shorten or extend lifespan by about two-fold or more. Tests for smaller effects will yield a higher rate of type II errors, and thus be less reliable. Increasing population sizes to $5,000$ individuals (requiring heroic efforts in the lab) would allow one to identify deviations from perfect scaling resulting from a constant σ when lifespan is altered only 0.5-fold.

This simulation suggests that, were aging determined by a Wiener process with drift, a fixed σ can be reasonably excluded for interventions that have a large effect on lifespan, for example exposure to tBuOOH at 1.5 mM, exposure to live bacteria (main Fig. 1 g-h), exposure to temperatures above 27 °C (main Fig. 1.b,c), and the *daf-2(e1368)* allele. For interventions with smaller effects on lifespan, data obtained with the population sizes we consider serve to limit the magnitude with which interventions affect σ. For example, in the case of *hif-1(ia4)*, our data would constrain the ratio of σ and μ in mutants to within 0.6 and 1.4 of wildtype. If this does not seem impressive, remember that our data also exclude other types of change to the aging process that might occur outside the Wiener process model (for example a disproportionate change in one risk determinant

as in Fig. 5-1d-f that produces a more prominent change in the survival distribution, e.g. a deviation from the hazard function associated with an inverse Gaussian lifespan distribution).

This all serves to illustrate a more general point: when scaling interventions produce large effects on lifespan (e.g. a 50% increase or decrease in a population comprising more than $1,000$ individuals), they can be used positively to suggest that the lifespan extension itself emerges from a simultaneous effect on multiple aspects of the aging process, σ and μ in the present model or all risk factors of equation (27). Interventions with smaller effects on lifespan are best interpreted as providing limits on the magnitude of any disproportionate effect across the aging process, such as causing a large change in σ without a corresponding change in μ, or adding or removing a risk factor in equation (27).

6.4 Drift-diffusion process with drift heterogeneity

Here we consider the diffusion process of the previous section, but introduce hetergenity in the drift coefficient μ. Let $p_\mu(V(t)=v)$ be the probability of being in state $v>0$ given a drift parameter $-\mu$ chosen from a normal distribution with mean $\bar{\mu}>0$ and deviation σ_μ:

$$p_\mu(X(t)=v)=\int_{-\infty}^{\infty} p(X(t)=v\,|\,\mu)\,p(\mu)\,d\mu. \tag{34}$$

A tedious calculation gives

$$p_\mu(X(t)=v)=\frac{1}{\sqrt{2\pi t\left(\sigma^2+\sigma_\mu^2 t\right)}}\left[1-\exp\left(-\frac{2cv}{\sigma^2 t}\right)\right]\exp\left(-\frac{(v-c+\bar{\mu}t)^2}{2t(\sigma^2+\sigma_\mu^2 t)}\right) \tag{35}$$

Integrating over $v>0$ yields the survival curve of the drift-diffusion process with drift heterogeneity:

$$\begin{aligned} S(t)&=\int_0^\infty p_\mu(X(t)=v)\,dv= \\ &=\Phi\left(\frac{c-\bar{\mu}t}{\sqrt{t(\sigma^2+t\sigma_\mu^2)}}\right)-\Phi\left(-\frac{c\sigma^2+t\bar{\mu}\sigma^2+2ct\sigma_\mu^2}{\sigma^2\sqrt{t(\sigma^2+t\sigma_\mu^2)}}\right) \\ &\quad\times\exp\left(-\frac{\sigma^2\sigma^2(c-t\bar{\mu})^2+(c\sigma^2+t\sigma^2\bar{\mu}+2ct\sigma_\mu^2)^2}{t\sigma^2\sigma^2(\sigma^2+t\sigma_\mu^2)}\right) \end{aligned} \tag{36}$$

Again, temporal scaling obtains when the parameters σ^2, $\bar{\mu}$, and σ_μ are scaled with λ^{-1}. Note that this agrees with our competing risk simulations in Fig. 5-1, where we present the phenomenon quantitatively.

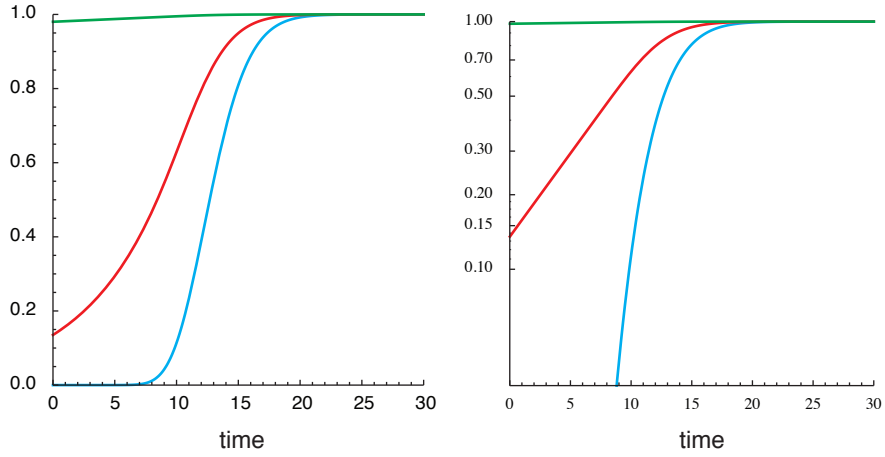
6.5 Strehler-Mildvan with vitality drift-diffusion

In 1960, Strehler and Mildvan proposed a now-prominent model of aging [26,27] that posits a separation of the aging process from the proximate causes of death. The former generates the (putatively) linear loss of vitality (represented by the physiological state variable $X(t)$ in the above processes), while the latter are events sampled from an age-independent distribution of (internal or external) challenges that cause death when their magnitude exceeds the vitality of the process at a given time.

Specifically, the waiting time τ between challenges has exponential density, $p(\tau) = \gamma \exp(-\gamma\tau)$. Thus, $P(\tau \geq t) = \exp(-\gamma t)$, $t \geq 0$. The magnitude m of a challenge is also exponential with parameter ν: $p(m) = \nu \exp(-\nu m)$ and $P(m \geq M) = \exp(-\nu M)$. If the vitality process $V(t)$ is deterministic, these assumptions imply for the hazard $h(t)$ [27]:

$$h(t) = \gamma\, P(m \geq V(t)). \tag{37}$$

That is, $h(t)$ is the average rate at which shocks occur (γ) times the probability that a shock kills.



Supplementary Figure 6-3: Hazard function of Strehler-Mildvan model with with vitality drift-diffusion. **A:** The plots show equation (41) as a function of time and ν (blue: $\nu = 1$; red: $\nu = 0.1$; green: $\nu = 0.001$). Parameters: $c = 20$, $\mu = 1.5$, $\sigma^2 = 1$, $\gamma = 1$. **B:** The lin-log plots illustrate the decidedly Gompertzian flavor of the model due to the imposed shock statistics.

We generalize equation (37) for the case in which vitality $X(t)$ is undergoing a drift-diffusion process as in section 6.2:

$$h(t) = \gamma \int_0^\infty P(m \geq v \,|\, X(t) = v)\, p(X(t) = v)\, dv. \tag{38}$$

The probability density that a drift-diffusion process that started at location c is at location $v > 0$ at time t is given by [1]:

$$p(X(t) = v) = \frac{1}{\sigma\sqrt{2\pi t}} \left\{ \exp\left(-\frac{(v - c + \mu t)^2}{2\sigma^2 t} \right) - \exp\left(\frac{2c\mu}{\sigma^2} \right) \exp\left(-\frac{(v + c + \mu t)^2}{2\sigma^2 t} \right) \right\}. \tag{39}$$

By integrating eqn (39) over v we obtain the probability density that the process has not yet been absorbed in state 0, which is the inverse Gaussian survival curve (32); $p(X(t) = v)$ loses probability into the absorbing state and we therefore need to have a separate term accounting for the probability accumulating in the absorbing 0 state. As usual, the integral can be expressed in

terms of cumulative standard normals (error functions):

$$h(t) = \gamma \int_0^\infty P(m \geq v \,|\, X(t) = v)\, p(X(t) = v)\, dv \tag{40}$$

$$\begin{aligned} &= \gamma \left[\exp\left(\frac{1}{2}\nu(-2c + 2\mu t + \nu\sigma^2 t) \right) \right. \\ &\quad \times \left\{ \Phi\left(\frac{c - (\mu + \nu\sigma^2)t}{\sigma\sqrt{t}} \right) - \exp\left(\frac{2c(\mu + \nu\sigma^2)}{\sigma^2} \right) \Phi\left(-\frac{c + (\mu + \nu\sigma^2)t}{\sigma\sqrt{t}} \right) \right\} \\ &\quad \left. +1 - \Phi\left(\frac{c - \mu t}{\sigma\sqrt{t}} \right) + \exp\left(\frac{2c\mu}{\sigma^2} \right) \Phi\left(-\frac{c + \mu t}{\sigma\sqrt{t}} \right) \right]. \end{aligned} \tag{41}$$

Therefore, for temporal scaling to occur in Strehler-Mildvan models, all parameters—μ, σ^2, and also the hit rate γ, but not ν, the parameter of the exponential density of hit magnitudes—must be multiplied with λ^{-1}.

References

[1] Aalen, O., Borgan, O. & Gjessing, H. *Survival and Event History Analysis : A Process Point of View* (New York, NY : Springer New York, 2008).

[2] Cox, D. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B* **34**, 187–220 (1972).

[3] Atlan, H., Mique, l. J., Helmle, L. & Dolkas, C. Thermodynamics of aging in *Drosophila melanogaster*. *Mechanisms of Ageing and Development* **5**, 371––387 (1976).

[4] Mair, W., Goymer, P., Pletcher, S. & Partridge, L. Demography of dietary restriction and death in Drosophila. *Science* **301**, 1731–1733 (2003).

[5] Wu, D., Rea, S., Cypser, J. & Johnson, T. Mortality shifts in *Caenorhabditis elegans*: remembrance of conditions past. *Aging Cell* **8**, 666–675 (2009).

[6] Johnson, T. E., Wu, D., Tedesco, P., Dames, S. & Vaupel, J. W. Age-specific demographic profiles of longevity mutants in *Caenorhabditis elegans* show segmental effects. *J Gerontol A (Biol Sci Med Sci)* **56**, B331–339 (2001).

[7] Samuelson, A., Carr, C. & Ruvkun, G. Gene activities that mediate increased life span of C. elegans insulin-like signaling mutants. *Genes Dev.* **21**, 2976–94 (2007).

[8] Stroustrup, N. *et al.* The *Caenorhabditis elegans* Lifespan Machine. *Nature Methods* **10**, 665–70 (2013).

[9] Vanfleteren, J., De Vreese, A. & Braeckman, B. Two-parameter logistic and Weibull equations provide better fits to survival data from isogenic populations of *Caenorhabditis elegans* in axenic culture than does the Gompertz model. *J Gerontol A Biol Sci Med Sci* **53**, B393–403; discussion B404–398 (1998).

[10] Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. & Harrington, D. P. Modified kolmogorov-smirnov test procedures with application to arbitrarily right-censored data. *Biometrics* **36**, 607–625 (1980).

[11] Vaupel, J., Manton, K. & Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454 (1979).

[12] Vaupel, J. W. *et al.* Biodemographic trajectories of longevity. *Science* **280**, 855–860 (1998).

[13] Baeriswyl, S. e. a. Modulation of aging profiles in isogenic populations of *Caenorhabditis elegans* by bacteria causing different extrinsic mortality rates. *Biogerontology* **1**, 53–65 (2010).

[14] Aalen, O. & Gjessing, H. K. Understanding the Shape of the Hazard Rate: A Process Point of View. *Statistical Science* **16**, 1––22 (2001).

[15] Weitz, J. & Fraser, H. Explaining mortality rate plateaus. *Proc Natl Acad Sci USA* **98**, 15383–15386 (2001).

[16] Steinsaltz, D. & Evans, S. Quasistationary Distributions For One-Dimensional Diffusions With Killing. *American Mathematical Society* **359**, 1285–1324 (2007).

[17] Wu, D., Rea, S., Yashin, A. & Johnson, T. Visualizing hidden heterogeneity in isogenic populations of C. elegans. *Exp Gerontol.* **3**, 261–70 (2006).

[18] Garigan, D. *et al.* Genetic analysis of tissue aging in *Caenorhabditis elegans*: a role for heat-shock factor and bacterial proliferation. *Genetics* **161**, 1101–1112 (2002).

[19] Riesen, M. *et al.* MDL-1, a growth- and tumor-suppressor, slows aging and prevents germline hyperplasia and hypertrophy in C. elegans. *Aging* **2**, 98–117 (2014).

[20] Kalbfleisch, J. D. & Prentice, R. L. *The Statistical Analysis of Failure Time Data. Second Edition.* Wiley Series in Probability and Statistics (Wiley, 2002).

[21] Rinne, H. *The Weibull Distribution: A Handbook* (Chapman and Hall/CRC, 2008), 1 edn.

[22] Vural, D. C., Morrison, G. & Mahadevan, L. Aging in complex interdependency networks. *Phys. Rev. E* **89**, 022811 (2014).

[23] Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **81**, 2340–2361 (1977).

[24] Gavrilov, L. & Gavrilova, N. The reliability theory of aging and longevity. *Journal of Theoretical Biology* **213**, 527–545 (2001).

[25] Heathcote, C. R. & Moyal, J. E. The Random Walk [In Continuous Time] And Its Application To The Theory Of Queues. *Biometrika* **46**, 400–411 (1959).

[26] Strehler, B. & Mildvan, A. General Theory of Mortality and Aging. *Science* **132**, 14–21 (1960).

[27] Wagner, P. Vitality heterogeneity in the Strehler-Mildvan theory of mortality. *MPIDR Working Papers* **WP-2011-012** (2011).

doi:10.1038/nature16550

Supplementary Table 1: Activation energies derived from fits shown in Figure 4b of the main text.

Temperature Range (°C)	Activation Energy (kJ/mol)	Activation Energy (eV)	Example Reaction type
20.7 - 24.3	56 ± 5	0.58 ± 0.05	metabolic
24.3--29.2	168 ± 3	1.74 ± 0.03	metabolic
29.7--31.8	552 ± 27	5.72 ± 0.28	unfolding of proteins
32.3--35.1	186 ± 9	1.93 ± 0.1	metabolic

± show standard error

Supplementary Table 2: Population Statistics, scale factors, and Kolmogorov-Smirnov Statistics corresponding to the comparisons presented the main text and extended data items.

Figure	Panel	Reference Condition	Reference Population Size	Test Condition	Test Population Size	Scale factor	KS Y Statistic	KS p-value	Additional Metadata
1	b-c	20C	1280	25C	1026	0.733	0.50	0.0258994	wild type
1	b-c	20C	1280	27C	2029	0.628	1.47	0.2	wild type
1	d-e	25C	3756	33C	4643	0.042	4.19	1.18E-15	wild type
1	f-g	0 mM	1588	1.5 mM	878	0.145	1.26	0.0852	wild type
1	f-g	0 mM	1588	3 mM	846	0.064	1.49	0.0234	wild type
1	f-g	0 mM	1588	6 mM	1129	0.030	1.96	0.0009	wild type
1	g-h	UV-killed Bacteria	2462	Live Bacteria	1252	0.667	0.85	>0.2	wild type
2	a,d	wild type	1223	*daf-2(e1368)*	1163	1.414	1.57	0.0145079	25C
2	b-e	wild type	1608	*daf-16(mu86)*	1491	0.780	1.28	0.0745341	25C
2	c-f	wild type	993	*hsf1(sy441)*	1174	0.723	1.22	0.1027819	20C
2	g-j	wild type	1365	*eat-2(ad1116)*	2462	1.463	2.30	4.95E-05	20C
2	h,k	wild type	1140	*nuo-6(qm200)*	1038	1.211	4.52	3.43E-18	25C
2	i,l	20C	583	25C	1038	0.760	0.73	>0.2	nuo-6
Ex. 4	d-e	wildtype	1223	*age-1(hx546)*	1227	1.647	1.00	>0.2	
Ex. 4	f-g	wild type	501	*hif-1(ia4)*	544	1.085	0.45	>0.2	25C
Ex. 4	h-i	20C	646	22.5C	474	0.862	0.65	>0.2	eat-2
3	c	24C for 3 days	1065	26C for 3 days	547		1.13	0.015877	29C scale
3	c	24C for 3 days	1065	27.5C for 3 days	573		1.56	0.0157039	29C scale
3	c	24C for 3 days	1065	29C for 3 days	563		5.29	1.08E-24	29C scale
3	d	24C for 3 days	1065	26C for 3 days	547		0.83	>0.2	29C shift
3	d	24C for 3 days	1065	27.5C for 3 days	573		0.75	>0.2	29C shift
3	d	24C for 3 days	1065	29C for 3 days	563		0.74	>0.2	29C shift
Ex. 4	j-k	wildtype	1539	*daf-2(e1368)*	1228	1.473	0.80	>0.2	
Ex. 4	l-m	wildtype	1680	*daf-16(mu86)*	1320	0.762	1.35	0.0535522	
Ex. 4	n-o	wildtype	1574	*age-1(hx546)*	1314	1.387	1.55	0.016599	
Ex. 5	a-b	24C always	541	29C always	563		1.62	0.010686	scale
Ex. 5	b-c	24C always	541	29C always	563		6.01	9.22E-32	shift
4 and Ex. 3	a-b and a-i	wildtype	16804	*n/a*	*n/a*				
4	c-d	wildtype	2921	*age-1(hx546)II.*	1140				
4	c-d	wildtype	2921	*daf-2(e1368)*	1059				
4	c-d	wildtype	2921	*daf-16(mu86) I*	1402				
Ex. 7	c	wildtype	5839	*age-1(hx546)II.*	1657				
Ex. 7	c	wildtype	5839	*daf-2(e1368)*	584				
Ex. 7	c	wildtype	5839	*daf-16(mu86) I*	1770				
4	e	wildtype	10811	*n/a*	*n/a*				
4	f	wildtype	5624	*age-1(hx546)II.*	3028				
4	f	wildtype	5624	*daf-2(e1368)*	2880				
4	f	wildtype	5116	*daf-16(mu86) I*	3963				

Scale factors were estimated using an accelerated failure time (AFT) regression

KS test statistics and p-values were estimated using the modified Kolomorogov-Smirnov test (Statistical Methods and Supplementary Note 2)

Figure	Panel	Group	Population Size	Group	Population Size
Ex. 1	a-i	Replicate 1	1249	Replicate 2	1388
		Replicate 3	816	Replicate 4	1230
1 and Ex. 2	a and a-b	20.1 °C	1733	30.9 °C	1734
		23.7 °C	4069	31.3 °C	1801
		25.2 °C	4207	32.5 °C	1653
		29.1 °C	1573	32.6 °C	4765
		30 °C	1743		
Ex 4	b	0 days	381	2 days	466
		1 day	294	3 days	388
Ex 4	c	t-Butanol	530	t-BuOOH	580
Ex 5	d-g	0 days	501	2 days	529
		1 day	485	3 days	477
Ex 9	l	1 shift	903	5 shifts	975
		1 shift	1127	5 shifts	952
		3 shifts	797		
		3 shifts	905		

Supplementary Table 3: Parameter estimates from the temperature and tBuOOH scaling functions.

		AFT Regression Parameters							Arrhenius Energy Estimates				Predicted Temporal Scaling Across Thermal Range					
20-29 °C (Figure 4.d)	Regime	β_T	β_T stderr	β_g	β_g stderr	β_x	β_x stderr	p(β_x) <.001 ?	Wild Type (KJ/Kmol)	Mutant (KJ/Kmol)	Wild Type (eV)	Mutant (eV)	Offset used to center T	T_1	T_2	wild type lambda	mutant lambda	mutant relative to wild type
daf-16 (mu86)	I.a	7298.3	235	-0.31	0.005	1073.4	399	N	61	70	0.63	0.72	-3.4E-03	20	24.4	0.69	0.66	0.95
daf-2(e1368)	I.a	7438.9	245	0.34	0.006	-3995.0	407	Y	62	29	0.64	0.30	-3.4E-03	20	24.4	0.69	0.84	1.22
age-1(hx546)	I.a	7350.0	231	0.34	0.011	-3643.8	683	Y	61	31	0.63	0.32	-3.4E-03	20	24.4	0.69	0.83	1.20
daf-16 (mu86)	I.b	19280.7	270	-0.31	0.005	-1586.4	380	Y	160	147	1.66	1.52	-3.3E-03	24.4	29.1	0.36	0.40	1.09
daf-2(e1368)	I.b	18815.2	287	0.46	0.006	-309.7	358	Y	156	154	1.62	1.59	-3.3E-03	24.4	29.1	0.37	0.38	1.02
age-1(hx546)	I.b	20206.4	272	0.48	0.006	-1740.7	358	Y	168	154	1.74	1.59	-3.3E-03	24.4	29.1	0.35	0.38	1.10
29-35 °C (Extended Data 6c)																		
daf-16 (mu86)	II	62846.2	642	-0.18	0.008	-15033.0	766	Y	523	398	5.42	4.12	-3.3E-03	29.1	32.1	0.13	0.21	1.63
daf-2(e1368)	II	66551.5	772	0.24	0.010	1039.2	1268	N	553	562	5.73	5.82	-3.3E-03	29.1	32.1	0.11	0.11	0.97
age-1(hx546)	II	65173.8	760	0.32	0.010	-1853.2	1202	N	542	526	5.62	5.46	-3.3E-03	29.1	32.1	0.12	0.13	1.06
daf-16 (mu86)	III	28838.0	515	-0.11	0.008	4541.2	658	Y	240	278	2.48	2.88	-3.3E-03	32.1	35	0.41	0.36	0.87
daf-2(e1368)	III	28893.2	531	0.34	0.012	2629.5	890	Y	240	262	2.49	2.72	-3.3E-03	32.1	35	0.41	0.38	0.92
age-1(hx546)	III	28876.9	496	0.30	0.008	3479.1	666	Y	240	269	2.49	2.79	-3.3E-03	32.1	35	0.41	0.37	0.90

Details on the regressions are provided in Statistical Methods

tBuOOH (Figure 4e)	β_C	β_B stderr	β_r_1	β_r_1 stderr	β_r_2	β_r_2 stderr	β_X_1	β_X_1 stderr	β_X_2	β_X_2 stderr
wild type	-1.84	0.0219	0.06	0.04	0.23	0.0	0.22	0.22	0.126	0.126

Details on the regressions are provided in Statistical Methods