Article

Distribution Analysis of the Lifespan Trait in Drosophila

Camila A. Yumuhova , Alexander V. Konopatov , Alexander A. Shtil [*] , Oleg V. Bylino [*]

Article

Distribution Analysis of the Lifespan Trait in Drosophila

Camila A. Yumuhova [1], Alexander V. Konopatov [1], Alexander A. Shtil [2,*] and Oleg V. Bylino [1,*]

[1] Department of Regulation of Genetic Processes, Laboratory of Molecular Organization of the Genome, Institute of Gene Biology RAS, Russian Academy of Sciences, 34/5 Vavilov St., Moscow 119334, Russia

[2] Blokhin National Medical Research Center of Oncology, 23 Kashirskoe Highway, Moscow 115522, Russia

* Correspondence: shtilaa@yahoo.com (A.A.S.); bylino@gmail.com (O.V.B.); Tel.: +7-(499)-135-60-89 (O.V.B.)

Abstract

Research into longevity and aging involves comparing the size of cohorts at certain points on survival curves. However, this analysis is oversimplified because it provides limited information about the sample structure and the distribution of lifespan as a trait within the studied population. Here, we introduce a method for estimating lifespan across the entire data range using distribution analysis. More specifically, we propose dividing the lifespan series into intervals, obtaining the frequencies of phenotypes by lifespan within the sample, followed by distribution analysis using the normality criterion. Additionally, to visualize the differences, we propose describing the resulting distributions formally using the normal distribution function and the β-function. We demonstrate that the proposed methodology enables to extract additional information from survival data, providing new insights into the processes that occur in populations in response to genetic interventions and shedding light on their impact on ontogenesis. The proposed approach adds a new layer of information to studies of longevity and aging and expands the toolkit of methods used to analyze survival data.

Keywords: lifespan; distribution analysis; Drosophila; aging; longevity

---

'I think it may fairly be assumed, in the light of what we now know, that no other measure will, statistically speaking, furnish so delicate and precise a measure of the general constitutional fitness of individuals as will their duration of life.'

Raymond Pearl, 1923

1. Introduction

Lifespan is one of the most complex quantitative traits. Like all quantitative traits, lifespan has a distribution [1,2]. It is fundamental to understand how the distribution of lifespan in the study populations will change in response to various interventions designed to increase lifespan (genetic, pharmacological, environmental), and to be able to correctly assess, understand, and interpret these changes. Studying of regularities of changes in lifespan distribution is a prerequisite for extending human lifespan as a species [3]. This kind of research is conveniently carried out using short-lived model organisms, such as Drosophila.

Studies of longevity and aging are inherently linked to the generation of Kaplan-Meier survival curves [4]. A survival curve reflects the decline of a cohort (individuals born at the same time) or population over time. Survival curves are commonly used in clinical studies, for example, to monitor the survival of patients after chemotherapy, as well as in longevity studies, to track the mortality of short-lived organisms in response to in response to a variety of interventions. To assess the differences between survival curves, a comparison of two samples based on median (50th percentile of mortality) and maximum (90% of mortality) lifespans is usually used. Appropriate statistical tests, such as the Kolmogorov-Smirnov two-sample test [5] or the log-rank/Mantel-Cox test [6], are

commonly used to estimate the former, while the Wang-Allison test [7] and the chi-squared test [8,9] are suggested for the latter. However, these tests only provide information on the significance of differences between samples at specific points in time. Thus, the survival curve does not provide detailed information on the frequencies of lifespan phenotypes within a sample. Rather, it only provides a general idea of how mortality increases over time within the study population. This requires the development of new approaches to the presentation and analysis of survival data that clearly show differences in the structure of the samples being compared.

Here, to solve the problem of sample structure, we converted the daily mortality data (deaths per day) into a lifespan series, where the unit of measurement is the lifespan of each individual. Then, to smooth out daily mortality fluctuations, we combined mortality data from the previous and following days. We divided the survival data into intervals/phenotype frequencies according to Sturges' rule and chose a convenient interval size of five days. Overlaying the probability series obtained in this way (i.e., phenotype frequencies by lifespan) for different samples and formally describing them using normal distribution curves made it possible to clearly compare the samples.

We analyzed the normality of the initial and obtained interval distributions using two tests: the Kolmogorov-Smirnov (KS) test for single samples and the Shapiro-Wilk (SW) test. Of all the tested options, the SW test on interval data proved to be the most effective, sensitive, and informative. The optimal conditions for applying the SW test to survival data were a sample size of up to 200 individuals and dividing the lifespan data of the compared samples into an equal number of intervals, calculated using Sturges' rule. This allows for fractional and unequal interval sizes, but requires an equal number of intervals between samples. To visualize the frequency distributions of the samples, we used the β-function in addition to the normal distribution function. Using the β-function, we demonstrated that distributions in Drosophila can take various forms, including a left-skewed β-distribution, a right-skewed β-distribution, a symmetric bell-shaped distribution similar to a normal distribution, and a uniform (linear) distribution. Using quantitative criteria of normality allows one to detect subtle changes in the sample structure, even when the shape and/or position of the survival curves of the compared cohorts are similar. It has been demonstrated that genetic interventions (introduction of the white mutation) are accompanied by changes in phenotype frequencies within the sample, changes in distribution type, and changes in quantitative parameters of normality. Additionally, it has been established that the distribution type in control WT-like lines does not necessarily correspond to the normal distribution. The distribution of these lines may be completely different. Thus, the presented study covers the entire spectrum of questions regarding the analysis of the distribution of lifespan in Drosophila.

2. Results

2.1. Dividing Survival Data into Intervals Enables an Effective Assessment of Phenotype Frequencies by Lifespan Within the Sample

To see the distribution of the lifespan trait in the studied samples, we converted the daily mortality series (Figure 1A) into a lifespan series where each value represents an individual's lifespan (Figure 1B). To understand the nature of the distribution, we applied normality tests to the resulting distributions using the standard Kolmogorov-Smirnov (KS) and Shapiro-Wilk (SW) tests for analyzing single samples. Unexpectedly, the results of the analysis showed that the SW test indicated a non-normal distribution for all 24 samples, while the KS test showed that the lifespan data for six of the 24 samples were normally distributed (see Tables 1,2 and 3, columns 6 and 10).
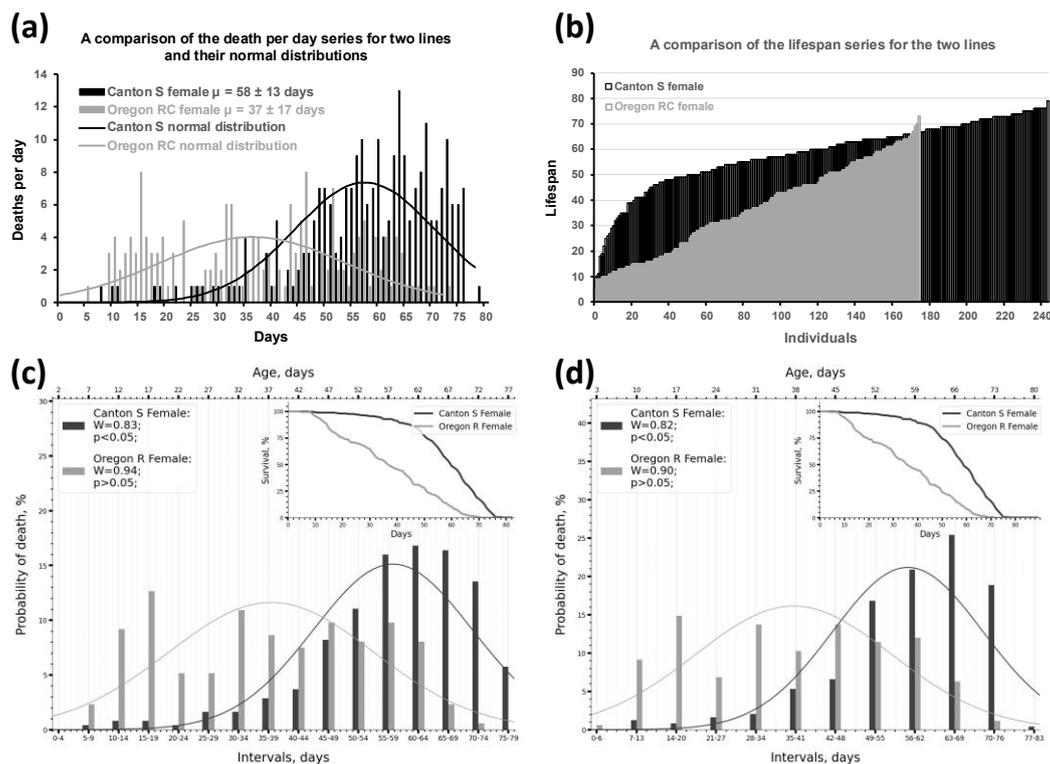
Figure 1. Original mortality distributions, mortality data converted to lifespan series and lifespan series converted to interval distributions/lifespan series. (**a**) Death-per-day data for females of two Drosophila WT-like laboratory lines with different genetic backgrounds (Canton S, black bars; Oregon RC, gray bars) are presented, along with normal distributions (bell-shaped curves) for these lines. Data from the w^{67c23} allele experiment (2012) were used for the analysis. The x-axis represents days (0, 1, 2, 3, etc.), and the y-axis represents the number of dead individuals per day. The numbers in the legend are μ (mean lifespan) ± standard deviation values. The normal distributions for the depicted death-per-day series were calculated as described in Section 3.3 of the Materials and Methods section. (**b**) The data from (a) were converted to a lifespan series. (**c**) The data from (b) were partitioned into five-day intervals using Sturges' rule. The partitioning used was as follows: day 0–4, day 5–9, day 10–14, etc. Phenotype frequency distributions (bars) with superimposed normal distribution curves are shown. Each bar shows the probability of dying in a given five-day interval. Normal distributions were calculated as in (a). W is the Shapiro-Wilk (SW) test statistic and p is the SW test p-value. The inset in the upper right corner of the graph shows the survival curves for the sample data. This graph was obtained using the Python script/program attached to the article (see Appendix #4). (**d**) The graph is similar to (c), but the data is divided into 7-day intervals. Data from (a) was used.

Table 1. Lifespan measurement experiment with the w^{1118} allele (2024).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
				SW test w/o intervals		SW test intervals		KS test w/o intervals		KS test intervals		Two sample	Number	Sample size
Genotype	Sex	μ, days	σ, days	W	p-value*	W	p-value*	D_n	p-value*	D_n	p-value*	KS test (WT vs. mut.)	of intervals	n
Canton S	M	34.8	± 12.7	0.99	0.036 not norm.	0.89	0.091 norm.	0.05	0.285 norm.	0.20	0.646 norm.	–	13	404
	F	33.1	± 13.5	0.98	0.001 not norm	0.94	0.478 norm.	0.07	0.091 norm.	0.15	0.917 norm.		13	308
w^{1118} Canton S	M	31.6	± 11.2	0.97	<0.0001 not norm.	0.80	0.005 not norm.	0.10	0.018 not norm.	0.26	0.307 not norm**	p<0.001***	14	220

Genotype	Sex	μ, days	σ, days	W	p-value*	W	p-value*	Dn	p-value*	Dn	p-value*	Two sample KS test (WT vs. mut.)	Number of intervals	Sample size n
	F	25.2	± 10.7	0.92	<0.0001 not norm.	0.84	0.019 not norm.	0.12	0.011 not norm.	0.30	0.197 not norm**	p<0.001	13	194
Oregon RC	M	35.2	± 10.4	0.96	<0.0001 not norm.	0.78	0.005 not norm.	0.10	<0.0001 not norm.	0.31	0.212 not norm**	–	12	472
	F	27.9	± 10.6	0.98	0.0001 not norm.	0.92	0.324 norm.	0.06	0.115 norm.	0.19	0.834 norm.		11	400
w^{1118} Oregon RC	M	33.7	± 10.9	0.95	<0.0001 not norm.	0.85	0.034 not norm.	0.11	0.001 not norm.	0.20	0.702 norm.	p>0.05	12	337
	F	25.1	± 8.7	0.98	0.0004 not norm.	0.88	**0.106 norm.**	0.08	**0.019 not norm.**	0.21	0.725 norm.	p<0.001	11	373

* at a significance threshold of p-value ≥ 0.05, the distribution was considered normal. ** not normal with Lilliefors adjustment for small sample sizes and normal without Lilliefors adjustment. *** the differences in survival curves between mutant and control flies of the same sex are shown.

Table 2. An experiment to measure lifespan with the w^{67c23} allele (2024).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
				SW test w/o intervals		SW test intervals		KS test w/o intervals		KS test intervals		Two sample KS test	Number of	Sample size
Genotype	Sex	μ, days	σ, days	W	p-value*	W	p-value*	Dn	p-value*	Dn	p-value*	(WT vs. mut.)	intervals	n
Canton S	M	33.3	± 12.7	0.98	<0.0001 not norm.	0.83	0.010 not norm.	0.09	<0.0001 not norm.	0.22	0.443 norm.	–	15	817
	F	31.2	± 13.3	0.99	<0.0001 not norm.	0.90	**0.123 norm.**	0.11	**<0.0001 not norm.**	0.18	0.740 norm.		14	676
w^{67c23} Canton S	M	39.7	± 18.2	0.99	0.0001 not norm.	0.97	**0.756 norm.**	0.06	**0.020 not norm.**	0.12	0.960 not norm**	p<0.0001	17	552
	F	37.5	± 18.0	0.98	<0.0001 not norm.	0.94	**0.408 norm.**	0.07	**0.012 not norm.**	0.18	0.654 norm.	p<0.0001	16	493
Oregon RC	M	36.4	± 13.7	0.98	<0.0001 not norm.	0.94	**0.417 norm.**	0.08	**0.013 not norm.**	0.11	0.994 norm.	–	15	374
	F	26.7	± 10.5	0.99	0.0075 not norm.	0.92	0.232 norm.	0.05	0.254 norm.	0.17	0.839 norm.		13	363
w^{67c23} Oregon RC	M	45.6	± 14.2	0.97	<0.0001 not norm.	0.84	0.008 not norm.	0.06	0.019 not norm.	0.25	0.222 not norm**	p<0.0001	17	641
	F	41.7	± 14.9	0.94	<0.0001 not norm.	0.87	0.039 not norm.	0.12	<0.0001 not norm.	0.22	0.479 not norm**	p<0.0001	15	660

* at a significance threshold of p-value ≥ 0.05, the distribution was considered normal. ** not normal with Lilliefors adjustment for small sample sizes and normal without Lilliefors adjustment. *** the differences in survival curves between mutant and control flies of the same sex are shown.

Table 3. An experiment to measure lifespan with the w^{67c23} allele (2012).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
				SW test w/o intervals		SW test intervals		KS test w/o intervals		KS test intervals		Two sample KS test	Number of	Sample size
Genotype	Sex	μ, days	σ, days	W	p-value*	W	p-value*	Dn	p-value*	Dn	p-value*	(WT vs. mut.)	intervals	n
Canton S	M	47.9	± 14.9	0.95	<0.0001 not norm.	0.92	**0.251 norm.**	0.11	**0.003 not norm.**	0.19	0.720 norm.	–	14	274
	F	58.3	±	0.92	<0.0001	0.83	0.010	0.09	0.046	0.22	0.473		15	244

			13.2		not norm.		not norm.		not norm.		norm.			
w^{67c23}	M	50.7	± 13.8	0.92	<0.0001 not norm.	0.80	0.003 not norm.	0.15	<0.0001 not norm.	0.22	0.406 not norm**	p=0.005 ***	16	262
Canton S	F	52.2	± 14.7	0.93	<0.0001 not norm.	0.89	**0.063** **norm.**	0.12	**0.001** **not norm**.	0.20	0.594 norm.	p<0.0001	15	283
	M	44.6	± 16.5	0.92	<0.0001 not norm.	0.87	0.036 not norm.	0.13	0.003 not norm.	0.24	0.332 not norm**	–	15	190
Oregon RC	F	37.2	± 17.4	0.96	<0.0001 not norm.	0.94	0.407 norm.	0.09	0.098 norm.	0.18	0.773 norm.		14	175
w^{67c23}	M	39.4	± 17.4	0.96	<0.0001 not norm.	0.98	**0.946** **norm.**	0.10	**0.043** **not norm.**	0.11	0.996 norm.	p<0.001	13	193
Oregon RC	F	39.4	± 15.3	0.97	0.0002 not norm.	0.95	0.587 norm.	0.08	0.127 norm.	0.14	0.951 norm.	p=0.040	14	209

* at a significance threshold of p-value ≥ 0.05, the distribution was considered normal. ** not normal with Lilliefors adjustment for small sample sizes and normal without Lilliefors adjustment. *** the differences in survival curves between mutant and control flies of the same sex are shown.

The results obtained led us to analyze the reasons for the unsatisfactory outcome. A thorough review of the experimental data showed that the survival data often contain mortality dips (days with zero mortality) and mortality outliers (days with a sharp increase in mortality). A typical example is shown in Figure 1A. We hypothesized that fluctuations in daily mortality might affect the assessment of normality.

To mitigate the effect of mortality fluctuations, we combined mortality of adjacent days by dividing the lifespan series into intervals according to the method of Sturges (the procedure is described in section 3.2 of Materials and Methods), who proposed a rule of thumb for determining the size (class) of an interval [10]. We applied Sturges' formula to our data with further rounding to integers (according to rounding rules) and obtained the following interval sizes for different samples: 7,7,7,7,6,5,6,5 (w^{1118} 2024), 8,8,9,8,8,8,7,7 (w^{67c23} 2012), 7,6,8,8,7,6,8,7 (w^{67c23} 2024) (the order of the samples is the same as in Tables 1,2 and 3). For a correct comparison, the same interval size must be used. According to Sturges, the most "convenient" intervals are 2,5,10,20,50,100,200, etc. When the interval calculated by Sturges's rule does not match a convenient interval, the nearest convenient interval should be used. We settled on an interval size of 5 days (Figure 1C, Figure S1A), testing a range of 3-7 days. We can see how the distributions with 79 and 73 days of observation (Figure 1A, Figure S1A) transform into distributions with 15 and 14 intervals (Figure 1C). The interval size of 5 days ensured acceptable detail and similarity to the original death-per-day distributions (compare Figure 1A with Figures 1C and S1A). The seemingly convenient value of 7 days created an overly simplistic, rarefied picture of the graph (compare Figure 1A and Figure S1A with Figure S1D). Thus, by dividing the lifespan series into equal intervals, we obtained a series of probabilities, i.e., phenotype frequencies by lifespan.

In graphs, mortality can be expressed in absolute numbers (i.e., the number of individuals) or in relative units (i.e., percentages or probabilities). Comparing in relative units normalizes the size of the samples being compared and appears to be the most informative. If one of the samples has a shorter lifespan or if there is no mortality during the initial time interval (five days), additional intervals with zero probability of mortality can be added to the obtained probability series. This makes it possible to visually compare the two samples in graphs. However, it is important to note that all zero intervals, both at the beginning and end of the distribution, should be excluded from further KS and SW test calculations, as including them introduces distortions to the values of the test statistics and affects the p-value.

To facilitate visual comparison of the samples, we superimposed the original distributions of deaths per day (Figure 1A, Figure S1B) or the obtained distributions of lifespan probabilities (Figure 1C,D, Figure S1A) with the corresponding normal distribution curves (indicated by the bell-shaped curves in Figure 1A and some subsequent figures). These curves were calculated based on the average

lifespan and standard deviation of the original (non-intervalised) lifespan series. The procedure is described in section 3.3 of Materials and Methods. This allowed us to compare two simple-looking normal distributions rather than two complex-looking frequency series and formalize and quickly understand the difference between the two samples.

Thus, dividing lifespan data into intervals and superimposing normal distribution curves on the resulting probability series eliminates the problem of mortality fluctuation, clearly shows differences in phenotype frequencies by lifespan within a sample, and makes it possible to effectively compare samples with each other.

2.2. Using Kolmogorov-Smirnov and Shapiro-Wilk Normality Tests to Describe the Parameters of the Lifespan Distribution

Next, we analyzed the resulting intervalised distributions using the KS and SW test and compared the results with the results of the analysis of the original distributions (not divided into intervals) obtained using the same tests for all 24 samples (Tables 1,2 and 3). We determined test statistics (numerical quantitative description of the results) and normality (qualitative description of the results according to the "yes" and "no" principle). Similar to the SW test on non-intervalised data, the KS test on intervalised data was found to be unrepresentative and showed a pattern of normality for all 24 samples (Tables 1,2 and 3, column 12). Applying the Lilliefors correction for small samples for the KS test helped to reveal non-normality of the distribution for 8 samples (Tables 1,2,3, column 12). The intervalised data analyzed by the SW test differed in 4 of the 24 samples from the results on the intervalised data obtained using the Lilliefors-corrected KS test. Moreover, for 3 of these 4 samples, when the SW test identified the distribution as non-normal, the Lilliefors-corrected KS test considered it normal and only for 1 sample the opposite situation was observed (Tables 1,2 and 3, compare columns 8 and 12, cases of differences are marked with gray shading). Taking into account the fact that KS test without Lilliefors correction always shows normal distribution for intervalised data, and with Lilliefors correction in 3 out of 4 samples shows the second kind of error (does not find deviations from normality where they are, compared to SW test), we concluded that SW test is more accurate and representative when analyzing lifespan data divided into intervals.

Next, to clarify which type of data is preferred for analysis, we compared the results of the SW test on intervalised data with the results of the KS test on non-intervalised data. We found that there are 8 out of 24 samples where differences in the results of the two tests are observed (Tables 1,2,3, compare columns 8 and 10, cases of differences are in bold). In all these cases, when the SW test on intervalised data showed a normal distribution, the KS test on non- intervalised data showed a non-normal distribution. To understand why this is observed, we took a closer look at the performance of the KS test in these 8 samples.

The analysis showed that, in general, the estimation of the normality of the distribution using the KS test is strongly influenced by the variation in daily mortality. We observed that the point corresponding to the supremum of the Kolmogorov-Smirnov function (based on the value of which the test concludes whether the distribution is normal or non-normal) can correspond to i) a day with high mortality preceded by one or more days with low mortality (Figure S2D,F,G). In this case, there is a strong rightward shift of the function, and the difference between the experimental and control CDFs is negative, ii) a day with high mortality preceded by several days with increasing mortality (Figure S2A,B,C,E,H). In this case, there is a strong leftward shift of the function, and the difference between the experimental and control CDFs is positive. Thus, in all 8 cases considered, the KS test identifies the distribution as non-normal based on the violation of the smoothness of the function.

To gain a deeper understanding of the regularities of the KS test, we also performed additional experiments (see the Section 2.2 «Additional observations regarding the mechanics of the KS test» in Appendix #1) and found that the KS test is strongly affected by fluctuations in mortality that fall in the middle of the distribution rather than at its edges, and by the presence of repeated values that occur in the lifespan series (Figure S3, Table S1, Figure S4).

Thus, the results presented here and in the Appendix #1 allow us to conclude that it is better not to use the KS test to estimate the pattern of lifespan distributions, regardless of the type of data (intervalised or non-intervalised). The KS test has low sensitivity for intervalised data and too high sensitivity for non-intervalised data. Thus, to estimate the pattern of lifespan distributions and predict biological regularities, we further used the SW test on intervalised samples.

2.3. Peculiarities of SW Test Functioning Under Different Intervalisation Conditions and on Samples of Various Types

Before conducting experiments using the SW test to compare mutant and control line samples, we examined the methodology more closely. We tested the limitations of the SW test on two sets of data: 1) 24 real samples and 2) an ideal normal series constructed using experimental sample data (Konopatov et al., in preparation; Figure S5). The ideal normal series uses information about the initial and final mortality dates, as well as the sample size (N). Third, we examined the SW test in simulations. In these simulations, we used ideal normal series with lengths ranging from 50 to 84 days (the range of maximum lifespans in our experiments) and sample sizes ranging from 100 to 795 individuals. In all three cases, we shifted the start of intervalisation (and the initial interval) by one, two, three, or four days. In other words, we compared the results of the test on samples with initial intervals starting from zero to four days, one to five days, two to six days, three to seven days, and so on until the end of the sample. In every assay, we used an interval length of five days.

It turned out that for experimental samples that were initially normal (for which we determined a normal distribution when we performed intervalisation from day zero, Tables 1, 2, and 3), shifting the first initial day of intervalisation from 0 to 1, 2, 3, or 4 days led to a change in the distribution type to non-normal in 12.5% of cases (Table 2). For non-normal samples, shifting the initial day of intervalisation changed the distribution type in 10.7% of cases (Table 2). For an ideal normal series constructed based on experimental sample data, shifting the initial day of intervalisation from 0 to 1, 2, 3, or 4 days resulted in 75.8% of the 120 simulated samples (24*5 days) being normal and 24.2% being non-normal (Table S2). In these experiments, the number of intervals depending on maximum lifespan (MaxLS) varied from 10 to 17. Thus, even with ideally normally distributed data, a false negative result may occur in a certain percentage of cases when the initial day of the initialization is shifted (the normal distribution must be re-determined, but a non-normal distribution is determined instead).

Table 4. Reproducibility of the method depending on the initial day of intervalisation on experimental sample data.

1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	18	19
			Internalization* from														
		μ, days	day 0			day 1			day 2			day 3			day 4		
Genotype	Sex		W	p-value**	Number of intervals	W	p-value	Number of intervals	W	p-value	Number of intervals	W	p-value	Number of intervals	W	p-value	Number of intervals
w^1118 (2024)																		
Canton S	M	34.8	0.89	0.091	13	0.91	0.153	14	0.87	0.038	14	0.91	0.146	14	0.92	0.271	13
	F	33.1	0.94	0.478	13	0.93	0.262	14	0.94	0.457	14	0.92	0.242	14	0.93	0.282	14
w^1118	M	31.6	0.80	0.005	14	0.82	0.008	14	0.81	0.006	14	0.78	0.004	13	0.81	0.008	13
Canton S	F	25.2	0.84	0.019	13	0.83	0.017	13	0.81	0.009	13	0.80	0.006	13	0.79	0.005	13
Oregon RC	M	35.2	0.78	0.005	12	0.81	0.014	12	0.81	0.009	13	0.80	0.007	13	0.76	0.003	12
	F	27.9	0.92	0.324	11	0.93	0.400	11	0.95	0.654	11	0.92	0.294	12	0.92	0.289	12
w^1118	M	33.7	0.85	0.034	12	0.85	0.034	12	0.83	0.023	12	0.81	0.017	11	0.82	0.020	11
Oregon RC	F	25.1	0.88	0.106	11	0.90	0.239	10	0.88	0.135	10	0.86	0.086	10	0.89	0.176	10
w^67c23 (2024)																		
Canton S	M	33.3	0.83	0.010	15	0.77	0.002	15	0.80	0.004	15	0.87	0.029	15	0.87	0.039	14
	F	31.2	0.90	0.123	14	0.77	0.003	13	0.74	0.001	13	0.93	0.305	13	0.90	0.157	13
w^67c23	M	39.7	0.97	0.756	17	0.95	0.456	17	0.93	0.254	16	0.92	0.158	16	0.95	0.521	16
Canton S	F	37.5	0.94	0.408	16	0.96	0.699	16	0.98	0.965	16	0.95	0.433	16	0.96	0.713	15

Group	Sex																
Oregon RC	M	36.4	0.94	0.417	15	0.93	0.314	14	0.93	0.298	14	0.94	0.381	14	0.96	0.650	14
	F	26.7	0.92	0.232	13	0.91	0.222	12	0.93	0.393	12	0.90	0.161	12	0.92	0.267	12
w^{67c23}	M	45.6	0.84	0.008	17	0.87	0.019	17	0.87	0.029	16	0.87	0.028	16	0.87	0.030	16
Oregon RC	F	41.7	0.87	0.039	15	0.87	0.035	15	0.88	0.055	15	0.86	0.027	14	0.89	0.088	14
w^{67c23} (2012)																	
Canton S	M	47.9	0.92	0.251	14	0.91	0.117	15	0.92	0.200	15	0.93	0.253	15	0.91	0.117	15
	F	58.3	0.83	0.010	15	0.82	0.007	15	0.80	0.004	15	0.83	0.009	15	0.81	0.004	16
w^{67c23}	M	50.7	0.80	0.003	16	0.81	0.003	17	0.82	0.004	17	0.81	0.004	16	0.82	0.005	16
Canton S	F	52.2	0.89	0.063	15	0.89	0.067	15	0.84	0.010	16	0.84	0.008	16	0.87	0.032	15
Oregon RC	M	44.6	0.87	0.036	15	0.87	0.031	15	0.89	0.063	15	0.89	0.072	15	0.89	0.066	15
	F	37.2	0.94	0.457	14	0.94	0.438	14	0.97	0.906	14	0.95	0.637	14	0.91	0.143	14
w^{67c23}	M	39.4	0.98	0.946	13	0.91	0.204	13	0.93	0.274	14	0.92	0.220	14	0.91	0.155	14
Oregon RC	F	39.4	0.95	0.587	13	0.92	0.290	13	0.95	0.608	13	0.96	0.772	12	0.92	0.287	13

*Interval length: 5 days. **With a significance threshold of $p \geq 0.05$, the distribution was considered normal. Green color in the table indicates normally distributed samples, red color indicates non-normally distributed samples.

During the simulations, we examined how sample size (N) and lifespan duration affected the SW test results. When studying the influence of N, we found that: As N increased from 100 to 795 (the maximum recorded in our experiments), the proportion of normally distributed samples decreased from ~95% at N = 100 to 34%-74% at N = 795, depending on when intervalisation started (days 0, 1, 2, 3, or 4) (Figure 2A). The average p-value decreased as N increased, indicating a trend toward greater non-normality (Figure 2B). Thus, as N increased, the number of false negative results increased. At the same time, the number of false negative results did not increase; it remained stable and was in a range of 1-6%, depending on the day on which the intervalisation was started with N of 100 to 200 individuals (Figure 2A). Therefore, to minimize SW test error on lifespan samples, the sample size should be between 100 and 200 individuals.
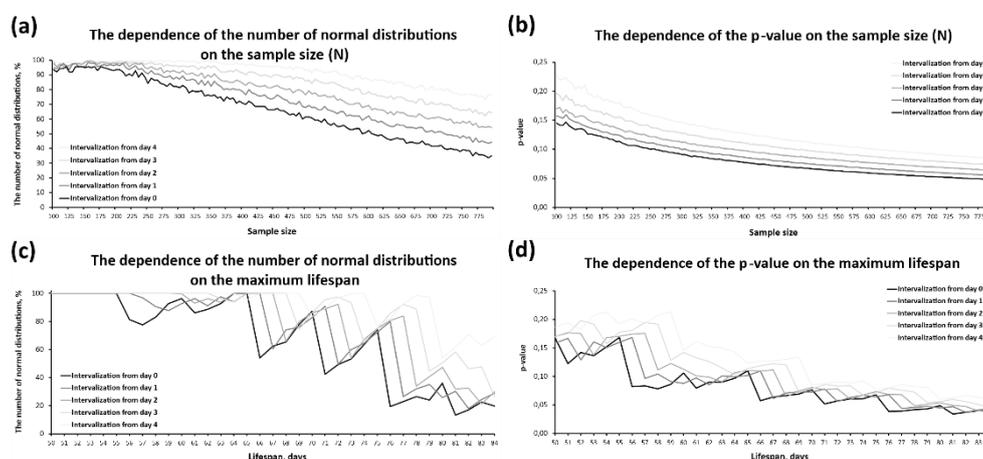
Figure 2. The figure shows the dependence of the frequency of preservation (retention) of normal distribution (**a,c**) and the p-value (**b,d**) on the sample size (number of individuals, N) (**a,b**) and the duration of lifespan (**c,d**) when performing the SW test. The dependencies were studied using an ideal normal series of lifespan of different lengths (different MaxLS) and with different N, intervalising this series from the 0th, 1st, 2nd, 3rd, and 4th days (marked on the graphs with different shades of gray). The interval size was five days. For each point on the graph, the results were averaged across all samples with varying N (from 100 to 795 in increments of five individuals), lifespan duration (from 50 to 84 days, the range of lifespans recorded in our experiments, in increments of one day), and the first day of death (the first death was recorded on the first, second, third, or fourth day). A total of 19,600 simulated samples were used (140N*35 lifespans*4 days of mortality onset).

When studying the effect of lifespan duration on the SW test result, the following was found: As the lifespan increased from 50 to 84 days (the range of MaxLS in our experiments), the proportion of

normally distributed samples decreased. Depending on the day when intervalisation was started, the proportion ranged from 19.8% (day 0) to 69.3% (day 4) (Figure 2C). The graphs in Figure 2C show a pronounced periodic alternation of sharp decreases and smooth increases in the number of normally distributed samples, with a fluctuation period of five days. This graph pattern is likely due to the formation of an additional interval when the lifespan exceeds the limit of the last interval. This results in a sharp decrease in the proportion of normally distributed samples as the number of intervals increases by one. The opposite trend, a gradual increase in the proportion of normally distributed samples after the sharp drop, can be explained by the gradual increase in size (height) of the last interval as lifespan increases. The average p-value decreased as lifespan increased, tending toward greater non-normality (Figure 2D). Thus, an increase in lifespan leads to the SW test more often "perceiving" the distribution as non-normal.

Thus, an increase in lifespan with a constant interval length of five days will be accompanied by a corresponding increase in the number of intervals. This leads to an increase in the proportion of samples with non-normal distributions. Consequently, as N increases (the number of intervals/frequencies of phenotypes in terms of lifespan in this case), the probability of determining the distribution as non-normal using the SW test increases. This assumption is also supported by the fact that changing the first day of intervalisation (0, 1, 2, 3, 4)—which is obviously accompanied by a reduction in the total number of intervals—also results in a decrease in the number of non-normally distributed samples (Figure 2C).

To investigate this issue, we conducted simulations and compared the results of the SW test when we intervalised an ideal normal series of varying lengths into 5-day intervals or when we intervalised this series into the optimal number of intervals according to Sturges' rule (see Section 3.2 of Materials and Methods). For each simulation, we took the result for the entire set of samples, as described in the legend for Figure 2. The only difference was that the N ranged from 100 to 940. We found that, when intervalising by five days, the number of normally distributed samples was 63.29%. Whereas when intervalising by Sturges, it is, depending on the type of rounding, 98.92% (rounding up, accompanied by a decrease in the number of intervals), 93.05% (rounding according to the rounding rule, giving an intermediate value), or 86.22% (rounding down, accompanied by an increase in the number of intervals). Thus, i) partitioning an ideal normal lifespan series into intervals according to Sturges' rule gives a higher percentage of preservation/retention of normal distributions than partitioning it into 5-day intervals, and ii) the result of the SW test is sensitive to the number of intervals.

In summary, the results of studying the SW test's functioning on small samples allow us to draw several conclusions: 1) The SW test result is influenced by lifespan duration (affected by the number of sample intervals) and sample size. The greater the number of intervals or N, the greater the number of false negative results. 2) Error in predicting normality caused by N can be minimized by using a sample size of up to 200 individuals. 3) Error arising from an increased number of intervals (e.g., intervalising every five days to obtain a frequency series of lifespan) can be minimized by intervalising according to Sturges' rule. The number of intervals will be smaller, and the interval size will be larger. At the same time, it is clear that the number of intervals in the compared samples must be the same.

2.4. Peculiarities of SW Test Functioning Under Different Intervalisation Conditions and on Samples of Various Types

The text continues here. We analyzed survival data for control lines and lines undergoing genetic interventions to investigate whether the normality criterion can be used to identify biological patterns and find differences between samples. The white mutation was used as a model. White mutants had the same genetic background as WT-like control flies. We used both quantitative (W, P) and qualitative (non-normal/normal) components of normality.

The effect of two different classical alleles of the gene – w^{1118} and w^{67c23} – on lifespan was investigated. Estimating the effect of the *white* gene on lifespan is an important task because, despite

more than 115 years of studying this gene [11,12], the physiological significance of *white* in Drosophila remains unclear [13–17].

The data obtained in the previous section clearly show that the number of intervals affects the normality assessment result for small samples using the SW test. Therefore, to use the SW test as a tool for identifying biological patterns and comparing two samples based on normality criteria, we should: 1) equalize the number of intervals between samples and 2) reduce the number of intervals. Accordingly, we abandoned the division of survival data into 5-day intervals and instead re-intervalised the samples according to Sturges' rule, using a fractional interval size (see Table 5). Next, we compared what would happen to the experimental samples in a paired comparison of mutant-control when we intervalised the data for five days and according to Sturges' rule. When equalizing the number of intervals, we used the number of intervals of the sample with the smallest value in each pair. As shown in Table 5, the same number of intervals for the mutant and control, coupled with a decrease in the total number of intervals, causes some non-normally distributed samples to become normally distributed when partitioned by Sturges' rule. However, the opposite pattern is sometimes observed.

Table 5. Comparison of the mutant and control when intervalising according to Sturges' rule (same number of intervals, different interval size) and intervalising according to 5 days (different number of intervals, same interval size).

1	2	3	4	5	6	7	8	9	10	11
Genotype	Sex	μ, days	Intervalisation according to Sturges's rule				Intervalisation using a five-day interval			
			W	p-value*	Number of intervals	Interval length	W	p-value	Number of intervals	Interval length
w^{1118} (2024)										
Canton S	M	34.83	0.89	0.221	9	7.17	0.89	0.091	13	5
	F	33.09	0.96	0.885	9	7.44	0.94	0.478	13	5
w^{1118} Canton S	M	31.59	0.80	0.025	9	7.29	0.80	0.005	14	5
	F	25.20	0.78	0.015	9	6.98	0.84	0.019	13	5
Oregon RC	M	35.21	0.83	0.033	10	6.07	0.78	0.005	12	5
	F	27.86	0.91	0.270	10	5.45	0.92	0.324	11	5
w^{1118} Oregon RC	M	33.69	0.85	0.065	10	5.85	0.85	0.034	12	5
	F	25.11	0.86	0.086	10	5.03	0.88	0.106	11	5
w^{67c23} (2024)										
Canton S	M	33.28	0.84	0.035	11	7.12	0.83	0.010	15	5
	F	31.21	0.75	0.006	10	6.44	0.90	0.123	14	5
w^{67c23} Canton S	M	39.70	0.94	0.521	11	7.91	0.97	0.756	17	5
	F	37.53	0.94	0.574	10	7.74	0.94	0.408	16	5
Oregon RC	M	36.36	0.97	0.901	10	7.02	0.94	0.417	15	5
	F	26.68	0.91	0.310	10	5.89	0.92	0.232	13	5
w^{67c23} Oregon RC	M	45.64	0.86	0.085	10	8.38	0.84	0.008	17	5
	F	41.70	0.87	0.108	10	7.47	0.87	0.039	15	5
w^{67c23} (2012)										
Canton S	M	47.85	0.94	0.579	10	7.64	0.92	0.251	14	5
	F	58.25	0.85	0.087	9	7.95	0.83	0.010	15	5
w^{67c23} Canton S	M	50.68	0.80	0.016	10	8.52	0.80	0.003	16	5
	F	52.22	0.86	0.095	9	8.06	0.89	0.063	15	5
Oregon RC	M	44.57	0.86	0.096	9	8.17	0.87	0.036	15	5
	F	36.96	0.96	0.817	9	7.70	0.94	0.457	14	5
w^{67c23} Oregon RC	M	39.37	0.90	0.280	9	7.35	0.98	0.946	13	5
	F	39.53	0.91	0.366	9	6.99	0.95	0.587	13	5

*With a significance threshold of p ≥ 0.05, the distribution was considered normal. Green color in the table indicates normally distributed samples, red color indicates non-normally distributed samples.
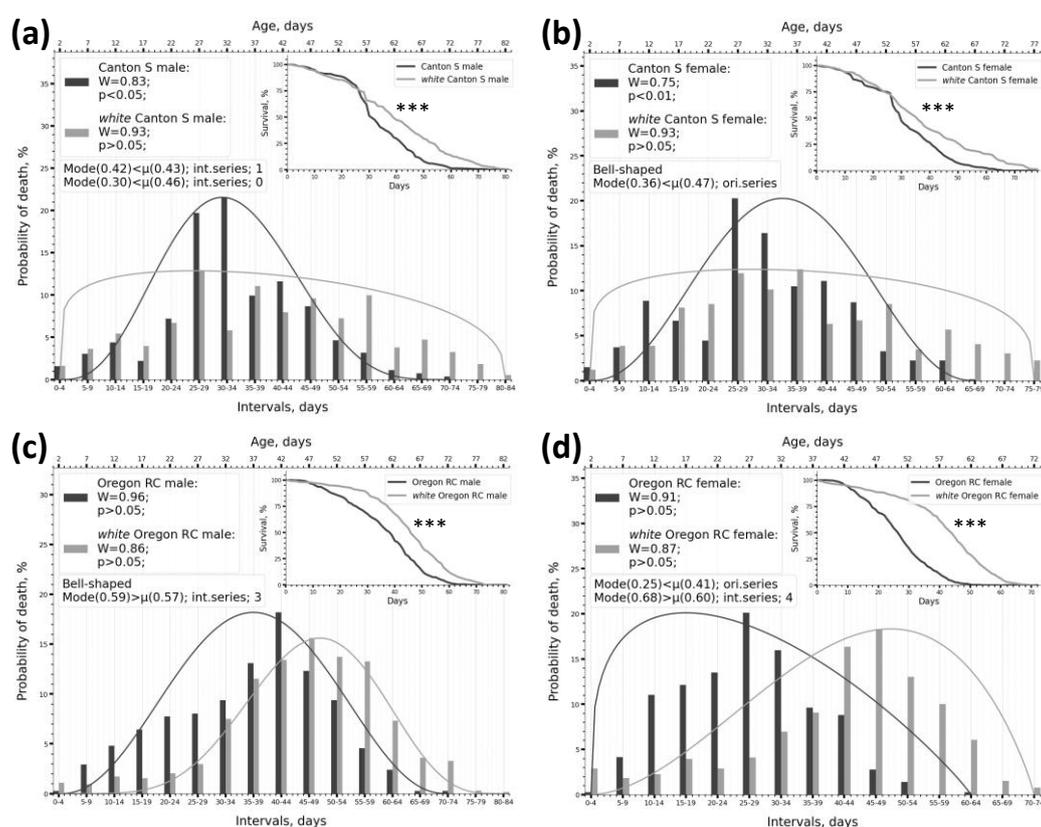
In an experiment with the w^{1118} 2024 allele (which reduces lifespan), it was found that the distribution changed significantly in three out of four cases when the *white* mutation was introduced (Figure 3A,B,D). The survival curves of the mutant and control differed significantly in all these cases according to the two-sample KS test (Table 1, column 13). A minor change in distribution shape also occurred in the Oregon RC male vs. white Oregon RC male pair. In this case, however, the two-sample KS test found no differences between the curves (Figure 3C). However, the SW test showed changes in W and P in all cases (Table 5, columns 4 and 5). And in 3 out of 4 cases (Figure 3A,B,C), there was a change in the type of distribution (normal/non-normal) (Table 5, columns 4,5). Thus, even when the survival curves did not differ from each other (as seen in the Oregon RC male vs. white Oregon RC male pair), it was possible to detect small changes in the distribution shape (the β-function helps in this) and quantitative indicators of normality.
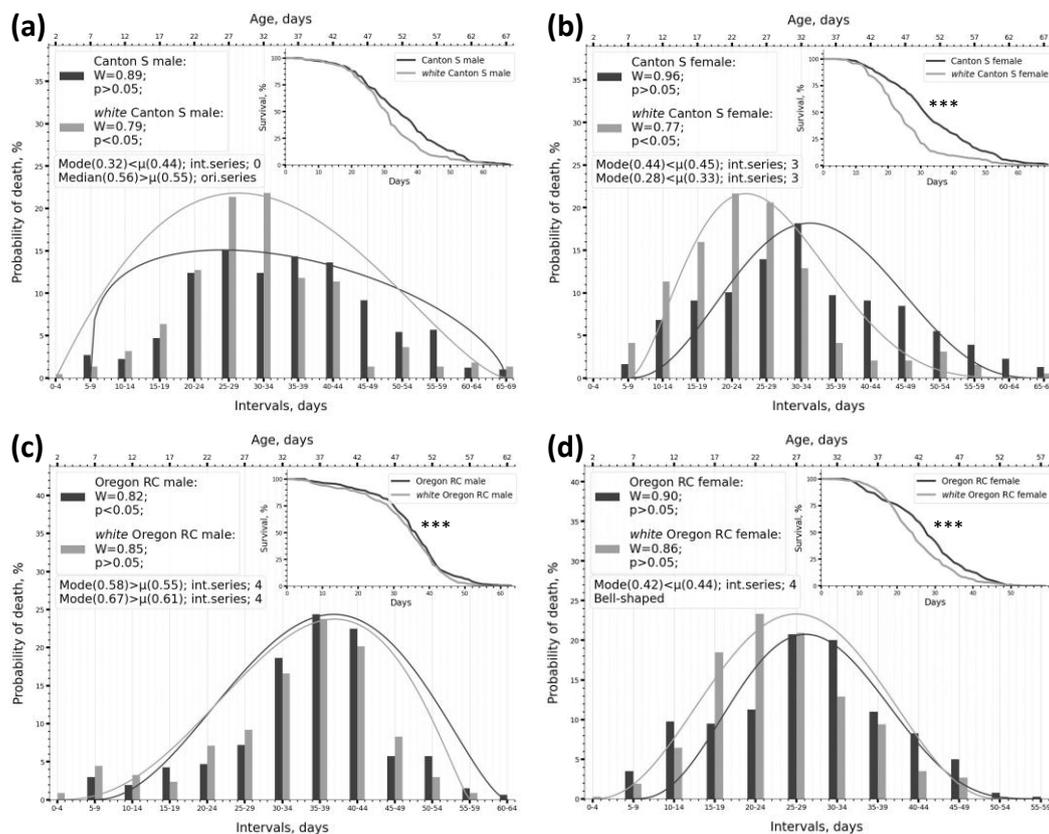
Figure 3. A comparison of the mutant and control in the experiment with allele w^{1118} (2024) is presented. The bars show the phenotype frequencies/probabilities by lifespan, with β-distribution curves superimposed. w^{1118} and control flies of the same sex were compared. Frequency distributions were obtained by dividing the original lifespan series into 5-day intervals. The inset in the upper left corner shows the results of the SW test calculation. W is the Shapiro-Wilk (SW) test statistic and p is the SW test p-value. The inset in the upper right corner shows the survival curves for the compared samples. The genotypes are labeled in the legend. *** – p<0.001. The lower inset in the top left corner shows the parameters used to construct the β-distribution graph (black and grey arcs). Mode, median, µ, int.series (interval lifespan series) and ori.series (original lifespan series). The number (0–4) indicates the calendar day from which the β-distribution graph construction begins (see the Python script instruction in Appendix #5 and the Python code in the supplementary file).

In the experiment with the w^{67c23} allele (which increases lifespan), the survival curves of the mutant and control groups differed significantly from each other according to the two-sample

Kolmogorov-Smirnov (KS) test in all cases (Table 2, column 13; Figure 4). The distribution changed in all cases and depended strongly on the genotype of the line. For instance, on the Canton S genetic background, the frequency distribution in both males and females was described by an arcuate β-distribution, which had a pronounced mode (Figure 4A,B). When the mutation was introduced, the arcuate β distribution transformed into a uniform β-distribution, and the frequency distribution lost its pronounced mode (Figure 4A,B). On the Oregon RC genetic background, the shape of the frequency distribution did not change significantly when the mutation was introduced, retaining a pronounced mode (Figure 4C,D). Changes in the shape of the frequency distribution were accompanied by changes in W and P in all cases (Table 5, columns 4,5). The distribution type changed to the opposite for the Canton S genotype but not the Oregon RC genotype (Table 5, columns 4,5). Thus, an increase in lifespan in response to genetic intervention is accompanied by genotype-dependent changes in the nature of frequency distribution, β-distribution curves, and quantitative measures of normality.

Figure 4. A comparison of the mutant and control in the experiment with allele w^{67c23} (2024) is presented. All designations as in Figure 3.

In an experiment with the w^{67c23} allele (2012), it was found that for the same lines in the collection and using the same algorithm for introducing the w^{67c23} allele into the lines, a different distribution pattern was observed than 12 years later (2024) (compare Figures 4 and 5). The 2024 and 2012 experiments were performed at slightly different temperatures: 25 °C and 23 °C, respectively. Therefore, it is not possible to directly compare the two experiments. However, it is possible to judge the changes that have occurred in the lines based on the ratio of effects between the control and the mutant. According to the two-sample KS test, for example, the positive effect of the w^{67c23} allele on lifespan was observed only in the 2012 experiment for the Canton S male vs. white Canton S male pair, whereas in 2024, the positive effect was observed for all genotypes studied (see Table 3, column 13, and compare with Table 2, column 13). In 2012, the frequency distributions of the mutant and control did not differ fundamentally in shape, and the mutation only modified the frequency

distribution without changing its shape. In contrast, the 2024 experiment revealed a sharp change in the frequency distribution's shape (at least for the Canton S genotype). The distribution type according to the SW test changed in 2012 for only one case (Figure 5A), while in 2024 it changed for two cases (Canton S genotype). The quantitative parameters of normality, W and P, changed when the mutation was introduced in both 2012 and 2024 (Table 5, columns 4 and 5). Based on the relationship between the survival curves and the frequency distributions of the mutant and control lines in 2012 and 2024, it can be assumed that the genetic background of the laboratory lines has changed so much that it has modified the effects of the w^{67c23} allele on lifespan distribution.
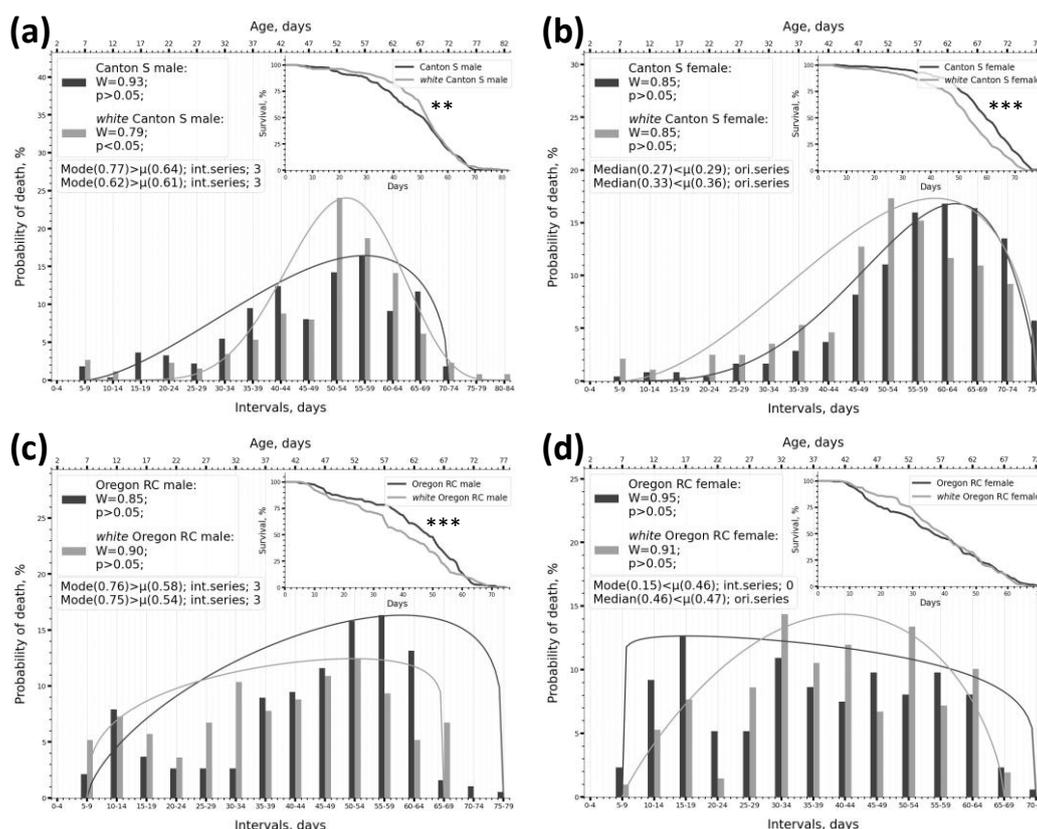
Figure 5. A comparison of the mutant and control in the experiment with allele w^{67c23} (2012) is presented. All designations as in Figure 3.

In all three experiments, the distribution of phenotype frequencies by lifespan can be formally described using either the β-function or the normal distribution function (Figures S6, S7 and S8). While this representation is more unified, as two bell-shaped curves of the same type overlap, it is less effective at showing the differences between samples than the β-distribution.

In summary, the results of the analysis of three experiments allow a number of conclusions to be drawn:

i) Determining quantitative indicators of normality (W and P) using the SW test (calculating the interval size according to Sturges' rule and equalising the compared samples by the number of intervals) enables subtle changes in the sample structure to be detected, even when the survival curves of the compared cohorts are almost identical in shape and do not differ significantly according to the results of statistical curve comparison tests (e.g. the two-sample Kolmogorov–Smirnov test; experiment with allele w^{1118} 2024).

ii) The characterisation of the distribution structure (frequency/probability series of phenotypes by lifespan) combined with a formal description of the distribution using the β-function or the normal distribution function, provides a clear picture of the processes occurring in the studied populations in response to genetic interventions.

(iii) The distributions of the lifespan of laboratory (highly inbred) Drosophila lines differ in form, depending on the genotype of the line and the year of the experiment, and change with the introduction of mutations. As can be assumed a priori, the distribution type in control Drosophila lines does not necessarily have to meet the normality criterion; the distribution of wild-type lines may be non-normal (experiments with alleles w[1118] 2024 and w[67c23] 2024).

3. Materials and Methods

3.1. Drosophila Lines and Survival Data for Analysis

For analysis, we used the *Drosophila melanogaste*r lifespan series: 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 8, 9, 9, 9, 9, 10, etc., where each number represents the lifespan (in days) of one individual fly. One hundred percent mortality was observed between days 50 and 84. We investigated WT-like laboratory control lines, Canton S and Oregon RC, as well as experimental Canton S and Oregon RC lines containing mutant alleles of the *white* gene (w[1118] and w[67c23]). The w[67c23] allele was isolated from the y[1]w[67c23] standard laboratory line chromosome by recombination with the control lines' chromosomes. The w[1118] allele was taken from line #3605 (Bloomington Collection). The genetic backgrounds of the mutant lines were equalized through 10 backcrosses with the control lines [18–24]. The backgrounds of control lines originated from the Drosophila collection at the Institute of Molecular Genetics of the Russian Academy of Sciences.

The analysis used lifespan data from three experiments: 1) Spring 2024: Canton S, w[1118] Canton S, Oregon RC, w[1118] Oregon RC lines. 2) Fall 2024: Canton S, w[67c23] Canton S, Oregon RC, w[67c23] Oregon RC. 3) Fall 2012: Canton S, w[67c23] Canton S and Spring 2013: Oregon RC, w[67c23] Oregon RC. The 2024 experiments were conducted at a temperature of 25 °C in a "warm" room, and the 2012–2013 experiments were conducted at a temperature of 23 °C "on the table." The experiments with the w[67C23] allele, conducted in 2012-2013 (hereafter referred to as 2012) and 2024, can be considered independent experiments on different genetic backgrounds. The genetic backgrounds of the lines changed significantly during the 12-year period they were kept in the collection, as seen in the shape and pattern of the survival curves. All experiments were conducted under natural light and humidity conditions. In each experiment, males and females were kept separately. Flies were mobilized for fresh food once every three days. The experiments were carried out in 100 x 25 mm polypropylene vials using the following medium: 66 g/l of pressed yeast, 35 g/l of semolina, 60 g/l sugar, 0.5% agar-agar, 0.5% propionic acid, and 30 g/l of raisins that had been twisted through a meat grinder.

3.2. Partitioning Survival Data into Intervals (Obtaining Frequency/Probability Series of Lifespan)

Sturges' rule [10] was used to determine the size of each interval:

$$C = \frac{R}{1+3,322*\lg N} = \frac{R}{1+\log_2 N} \qquad (1)$$

where, C is the optimal interval size (in our case, the number of contiguous days combined), R is the difference between the maximum and minimum sample values (in our case, maximum and minimum lifespans), N is the sample size (number of individuals), lgN is the decimal logarithm of N, $\log_2 N$ is the logarithm of N on base 2 (equivalent to 3.322*lgN).

The results of the intervalisation were adjusted to a uniform interval size of five days, and a series of phenotype frequencies by lifespan was constructed and plotted on graphs. To construct a frequency series of phenotypes by lifespan, we first found the number of individuals in a given interval (5 days). Then, we calculated the mortality rate in each interval as a percentage. Detailed instructions for dividing into intervals are provided in Appendix #2, Steps 2-3. The resulting frequency series were superimposed with either a normal distribution curve or β-distribution curve for illustrative purposes (see below).

To calculate the SW test, the samples were equalized by the number of intervals as follows:

First, the optimal interval size for each sample was calculated using Sturges' rule (Formula #1) to determine the number of intervals. If the number of intervals differed between the two samples, the smaller value was used in the calculations. The interval size for the second sample was

recalculated by dividing its lifespan by the number of intervals of the first sample. Thus, while the interval size for each sample differs (intervals may be fractional), the number of intervals in both samples remains the same, allowing for correct comparisons using the SW test.

Second, the number of individuals falling into each interval for each sample was determined. This was done as follows: From the first day of mortality, the interval size found in the previous step was measured. Next, the number of individuals whose lifespan was less than or equal to the obtained value was calculated. This process was repeated until the end of the lifespan series.

3.3. Formal Description of Frequency/Probability Series of Phenotypes by Lifespan and Mortality Series Using the Normal Distribution Function

To superimpose the normal distribution curve on the death per day series (Figure 1A), we calculated the value of the probability density function (PDF) for each day in the series (1, 2, 3, etc.) using the following formula:

$$F(\boldsymbol{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (2)$$

where, μ and σ are the mean and standard deviation, respectively, of the original (non-interval) lifespan series, σ^2 is the variance of lifespan, e is the exponent and π is the number Pi (3.14). The PDF can be calculated using the MS Excel function =NORM.DIST(x; μ; σ; FALSE), where FALSE is the PDF and TRUE is the cumulative distribution function (CDF).

μ and σ were calculated according to formulas 3 and 4, respectively, with raw mortality data for lifespan series without interval partitioning (the value of σ calculated from interval data will be considerably wider than that calculated from non-interval data):

$$\mu = \frac{\sum_{i=1}^{n} A_i * i}{N}, \qquad (3)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (i-\mu)^2 * A_i}{N-1}} \qquad (4)$$

where, $\sum_{i=1}^{n} -$ is the sum from the first to the last day of measurement (n); A_i is the number of individuals that died on day i; i is the day on which the individual died (1, 2, 3, …) (actually, the individual's lifespan); and N is the sample size (the total number of individuals).

The values of the normal distribution that were obtained (i.e. fractions of 1) were multiplied by the sample size (N). This was done when the normal distribution curve was superimposed on the graph showing mortality by day (see Figure 1A).In this case, the scale of the normal distribution curve correlated well with the size of the bars (see Figure 1A).

In cases where normal distribution curves were superimposed on the frequency series of phenotypes by lifespan, intervalised by five days, and graphs were constructed in MS Excel (Figure S1A), the obtained values of the normal distribution function were multiplied by 100 (the sum of mortality frequencies across all intervals) and then by the interval size (five). In this case, x in formula #2 was taken as the start day of the interval in the series of intervals (0, 5, 10…). This is related to a certain angularity of the curve in the graph in Figure S1A. The start date of the interval was taken as x, since MS Excel does not allow two different X-axes to be combined on one graph (but it does allow two Y-axes to be combined). This means it is impossible to combine the frequency series (bars) and the normal distribution curve plotted by day (1, 2, 3, etc.).

For graphs constructed using a Python script (Figure 1C and subsequent figures of a similar type) the values of the normal distribution function obtained were multiplied by 100 (the sum of the mortality frequencies across all intervals) and then by the interval size (5). In this case, the scale of the normal distribution curve correlated well with the scale of the frequency series (see Figure 1C and subsequent figures). For x in formula #2, in this case, the sampling day (0, 1, 2, 3, etc.) was used.

The normal distribution function never reaches zero and approaches it infinitely; therefore, on graphs, the PDF of the normal distribution may appear as an arc-shaped curve "suspended above the X-axis in the air" without ever touching the X-axis.

3.4. Kolmogorov-Smirnov Test Calculation for Single Samples

The one-sample Kolmogorov-Smirnov (KS) test was used to assess the normality of lifespan distributions. This test is based on identifying the supremum, or the point of maximum deviation between the experimental and expected cumulative distribution functions (CDFs), for a given sample [25,26] (Figure S1C).

The CDF of the test was calculated using the following formula:

$$F(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{5}$$

where, $F(x)$ is the function of lifespan dependence on time; x is the lifespan of an individual; μ is the average life span; σ is the standard deviation of mean (the square root of the variance); and *erf* is the error function, which can be calculated using the formula:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt \tag{6}$$

where, $\int_0^z$ is integral from 0 to z, π is the number π, dt is the infinitesimal increment of t, z is z-score (shows the deviation of the value x from μ, expressed as a quantity of σ), which can be calculated by the formula:

$$z = \frac{x-\mu}{\sigma\sqrt{2}} \tag{7}$$

The differences between the experimental CDF (black stepped curve) and the expected CDF (gray straight line) (Figure S1C) were found by subtracting the expected CDF values from the experimental CDF values at each point along the graph. The difference was considered positive when the experimental CDF deviated upward/leftward from the expected CDF and negative when it deviated downward/rightward. The maximum modulus of the difference between the experimental and expected CDFs was considered the supremum, or KS statistic, of the test (D_n, D experimental). The calculated D_n values were then compared with the critical D-value (D_{crit}) calculated for sample size N at the chosen significance level α (0.05). Critical values for N less than 35 were taken from Smirnov's table [27]. For N between 35 and 100, critical values were taken from Miller's table [28]. For N greater than 35, the formula $d_\alpha/\sqrt{n}$ was used, where $d_\alpha(N) = 1.36$ at $\alpha = 0.05$. Values for other significance levels were taken from [29,30]. When $D_n < D_{crit}$, the distribution was considered normal. If the exact sample size was unavailable in the tables, the p-value was found and compared to the chosen significance threshold, α. If the p-value was greater than α, the distribution was considered normal. We used the Kolmogorov's formula [25,26,31] to calculate the p-value, which is convenient for both small samples (intervalised lifespan series) (N<35) and large samples (non-intervalised original lifespan series):

$$P = 1 - \text{L}(z), \tag{8}$$

where, L(z):

$$\text{L}(z) = 1 - 2 * \sum_{y=1}^{\infty}(-1)^{y-1} * e^{-2*y^2*z^2} \tag{9}$$

where, $\sum_{i=1}^{\infty}$ is the sum from 1 to ∞. In practice, the sum of the first ten values of y is significant; the rest do not add precision to the measurement; e is the exponent; $z = D_n * \sqrt{n}$, $z^2 = (D_n)^2 * \left(\sqrt{n}\right)^2 = (D_n)^2 * n$, where n is the sample size.

The methodology for manually calculating the KS test on lifespan series for non-intervalised samples is provided in Appendix #2. Explanations of the test can be found in classic papers [26,29]. To automate the calculations, we created a Python script. The script description is provided in Appendix #4. The script is available at: https://colab.research.google.com/drive/1XUdqPQr6618-9LKY_RnviHubW6Dy0uUz?usp=sharing#scrollTo=wC_oo1z8WGpF. We also attach a PDF file with the customized code for this script. When using the script to calculate the KS test for small samples and large samples, D_n is calculated using the classical methodology described in Appendix #2 and the p-value is calculated using the classical Kolmogorov's formula.

3.5. Calculation of the Two Sample Kolmogorov-Smirnov Test

To assess the differences between the two survival curves, a two-sample KS test was used. The two-sample KS test calculation consists of three stages: 1) Find the lifespan as function of time (S_t) for each sample, 2) Find the difference between the two S_t functions (=$D_{n,m}$ statistic), 3a) Find D_{crit} and compare it with the Smirnov's percentage point table [27,28], or, instead, 3b) Find the p-value using Kolmogorov's formula (=result of the test). Dividing the value obtained using Kolmogorov's formula [25,26,31] by 2 gives the result of a one-sided test. However, there are other formulas for calculating the p-value of a one-sided test [32].

The values of the S_t function were found using the formula:

$$S(t) = \prod_{i=0}^{t} \frac{(n_i - d_i)}{n_i} \tag{10}$$

where, n_i is the number of individuals that lived until the day t, d_i – is the number of individuals that died at day t, $\prod_{i=0}^{t}$ is the result of multiplying all terms of the series from zero to the t-th term. In MS Excel, it is more convenient to find the value of the S_t function for each day by multiplying the previously found value of the function (S_{t-1}) by the value $\frac{(n_t - d_t)}{n_t}$ for day t.

The difference between the S_{t1} and S_{t2} functions was calculated using the formula:

$$D_{n,m} = \sup_t \left| S_{1,n}(t) - S_{2,m}(t) \right| \tag{11}$$

where, $D_{n,m}$ is the observed D (D statistics); $supS$ is the maximum difference between subsets S_1 and S_2; $S_{1,n}(t)$ and $S_{2,m}(t)$ are the values of the S_t function at each point of time t (day, in this case) for samples n and m, respectively.

The D_{crit} was found using the following formula:

$$D_{crit} = d_\alpha(N) * \sqrt{\frac{n+m}{nm}} \tag{12}$$

where, n and m это are the volumes of the first and second samples, respectively, $d_\alpha(N)$ is a tabulated value i.e., the coefficient used to find D_{crit}.

The value of $d_\alpha(N)$ is 1.358, 1.628, 1.731, 1.949 for significance levels α = 0.05, 0.01, 0.005, 0.001, respectively. The value of $d_\alpha(N)$ can be found for any percentile point using the formula:

$$C(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) * \frac{1}{2}} \tag{13}$$

The null hypothesis (H0), which states that the survival curves do not differ, is rejected at a significance level of α when $D_{n,m} > D_{crit.}$

The exact p-value was calculated using formulas #8 and #9. To find z, the following formula was used:

$$z = D_{n,m} * 1/\sqrt{\frac{n+m}{nm}} \tag{14}$$

For a one-sided test, the p-value can be found using the following formula [32]:

$$P = e^{-2z^2 - \frac{2z}{3} * \frac{m+2n}{\sqrt{mn(m+n)}}} \tag{15}$$

where, e is the exponent, $z = D_n * \sqrt{n}$, $z^2 = (D_n)^2 * \left(\sqrt{n}\right)^2 = (D_n)^2 * n$, where n and m are the sample sizes, n and m, respectively.

The attached MS Excel file on sheet Appendix #5 provides an example of instructions for calculating a two-sample KS test.

3.6. Shapiro-Wilk Test Calculation for Single Samples

The generalized SW test was used to evaluate the normality of lifespan distributions. Unlike the KS test, the SW test is based on a different principle. It assesses the symmetry of the data distribution by comparing the left and right parts of the distribution (Figure 1B). The coefficients used to calculate the SW test statistic (W) for small samples (≤50), i.e., intervalised lifespan series, were taken from Table 5 of the original paper [33]. Exact p-values for small samples were calculated using Royston's method [34], which proposes calculating P for any sample size without a percentage point table. As an alternative, approximate p-values can be found using Table #6 from the original work by Shapiro and Wilk [33]. For large samples, i.e., non-intervalised lifespan series, both W and P were calculated using Royston's method [34].

The p-value was calculated using the following final formula:

$$P = 1 - \Phi(z) \tag{16}$$

where, $\Phi(z)$ is the standard ($\mu = 0$, $\sigma = 1$) normal distribution function of z, μ is the mean lifespan, σ is the standard deviation (square root of the variance).

$\Phi(z)$ was found as follows:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \tag{17}$$

where, z is the z-score or standard score, which shows the deviation of the value of X from μ, expressed as a quantity of σ; and dt is the infinitesimal increment of t.

z was found as follows:

$$z = \frac{(X-\mu)}{\sigma} \tag{18}$$

The z-value can be found in the z-score table [35] or calculated using a normal distribution function. For example, the MS Excel function NORMDIST(X; μ; σ; TRUE) can be used, where TRUE is the cumulative function of the standard normal distribution (CDF).

The transformation of the observed z-distribution to the normal distribution (μ=0, σ=1) was performed using Royston's formula:

$$z = \frac{(g(X)-\mu-\sigma\mu_z)}{\sigma\sigma_z} \tag{19}$$

where, g(X) is the function that transforms the distribution of the original data into a normal distribution and μz and σz are the transformed μ and σ, respectively. Since μ = 0 and σ = 1, we obtain:

$$z = \frac{(g(X)-\mu_z)}{\sigma_z} \tag{20}$$

The formulas and coefficients for calculating g(X), μz, and σz according to Table 1 of Royston's original paper [34] differ for samples of size less than 12 and for samples between 12 and 2,000. We used the formulas and coefficients for samples ranging from 12 to 2,000 (see below), given in Table 1 of Royston's original paper, to calculate g(X), μz and σz. We performed these calculations for all cases, i.e. when the sample size was >12 and when the sample size was <12 (11 in our case), since we found that using the formulas and coefficients for sample sizes ranging from 4 to 11 (as given in Royston's original paper, Table 1) always yielded a p-value of zero. Therefore, the previously published formulas and coefficients are invalid for sample sizes ranging from 4 to 11.

$$g(X) = ln(1 - W), \tag{21}$$

$$\mu_z \approx 0.0038915 * (ln(n)^3) - 0.083751 * (ln(n)^2) - 0.31082 * ln(n) - 1.5861, \tag{22}$$

$$\sigma_z \approx e^{0.0030302 * (ln(n)^2) - 0.082676 * (ln(n)) - 0.4803} \tag{23}$$

where, *n* is the sample size, *e* is the exponent, and *W* is the statistic of the SW test. «The numerator of W is proportional to the square of the 'best' (minimum variance, unbiased) linear estimator of the standard deviation, and the denominator is the sum of squares of the observations about the sample mean» [36].

The SW test statistic was calculated using the following formula from the original paper [33]:

$$W = \frac{\left(\sum_{i=1}^{m} a_{n-i+1} * (x_{n-i+1} - x_i)\right)^2}{\sum_{i=1}^{n} (x_i - \frac{\sum x}{n})^2} \tag{24}$$

where, *x* is the sample, which is a sorted, ascending series of numbers showing the number of individuals that died in each interval; *n* is the sample size, or how many numbers are in the aforementioned series; *m* is n/2 for even *n* and (n-1)/2 for odd n; and $\sum_{i=1}^{m} -$ is the sum operator, where index *i* takes values from 1 to *m*. Xi is a member of the series, or sample, by count (1, 2, 3, etc.)....), and a_{n-i+1} are the coefficients from Table 5 for calculating the SW test statistic, as described in the original paper by Shapiro and Wilk [33]. W reflects the normality of the distribution. The value of W ranges from 0 to 1, and the closer W is to 1, the more probable it is that the data follow a normal distribution.

The methodology for the full manual calculation of the SW test on intervalised lifespan series is provided in Appendix #3. Explanations of the SW test can be found in [33,34]. We created a Python script to automate the calculations and included a description in Appendix #3. The program calculates *W* for small (intervalised) samples according to Shapiro and Wilk's original paper and calculates the p-value according to Royston's paper. For large (non-intervalised) samples, both *W* and p-value are calculated according to Royston's paper.

3.7. Calculating the β-Distribution of Lifespan

The advantage of formally describing frequency series using the PDF of a β-distribution is that when calculating the PDF of a β-distribution, the first and last days of mortality are specified. This makes the graph of the PDF of a β-distribution coincide with the boundaries of the frequency series. Additionally, the first and last values of the β-distribution PDF are 0, unlike the normal distribution PDF, so the initial and final values of the β-distribution graph coincide with the X-axis. The normal distribution function approaches 0 infinitely and therefore extends beyond the boundaries of the graph, giving the appearance of being suspended in the air.

The PDF of the β-distribution was found using the formula from Section 18.4, page 243 [37]):

$$\beta\ distribution = \frac{x^{p-1}(1-x)^{q-1}}{\beta(p,q)} \tag{25}$$

where, $\beta(p,q)$:

$$\beta(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \tag{26}$$

where, Γ(z) is the gamma function, the condition of which is that *z* must be a rational positive number ($R(z) > 0$). The gamma function was found using the following formula:

$$\int_0^\infty t^{z-1} e^{-t} dt \tag{27}$$

where, *t* are the values range that from zero to infinity on the X-scale, *z* is the argument for which the gamma function value is found, *e* is the exponent, and *dt* is the infinitesimal increment of *t*; the gamma function can be calculated using the =GAMMA(number) function in MS Excel, where (number) is the number on the X-axis for which the gamma function value is found on the Y-axis.

According to the law of β-distribution (see formula #25) PDF is defined in the interval *x*<1 and *x*>0 (0 < *x* < 1). Two parameters, *p* and *q*, determine its shape, with *p*>0, *q*>0. The values of *p* and *q* can

be expressed in consequently from the formulas for finding the Mean (μ) of the β-distribution and the Median (50th percentile of the sample, *Med*) of the β-distribution.

The following formula was used to calculate the μ of the β-distribution:

$$\mu = \frac{p}{p+q} \tag{28}$$

In order to express q through p and μ a number of transformations were carried out, resulting in the following:

$$q = \frac{p}{\mu} - p \tag{29}$$

In order to express p through q and μ, a number of transformations were carried out, resulting in the following:

$$p = \frac{\mu * q}{1-\mu} \tag{30}$$

The following formula was used to find the *median* of β-distribution:

$$Med = \frac{p - \frac{1}{3}}{p + q - \frac{2}{3}} \tag{31}$$

Finally, to express p through the median and μ, we substitute formula #29 for q in formula #31. This yields formula #32, which enables us to find p given μ and Med:

$$p = \frac{1}{3} * \frac{1 - 2*Med}{1 - \frac{Med}{\mu}} \tag{32}$$

Next, we similarly express q through Med and μ:

$$q = \frac{1}{3} * (\mu - 1) * \frac{1 - 2*Med}{Med - \mu} \tag{33}$$

As a result, knowing *Med* and μ, we can find p using the formula #32 and q using the formula #33. Then, we substitute the values of p and q into formula #25, calculating it for values of x in the interval from 0 to 1.

In order to construct a β-distribution, the dimension of the distribution must first be normalized. Without normalization, the graph will not function properly since the β-distribution exists between 0 and 1. To normalize the dimension, subtract the first day of mortality from the μ values, then divide by the difference between the last and first days of mortality. For example, $\mu = 35$ days and MaxLS = 84 days; the first death was recorded on day 8. Therefore, μ for constructing the PDF β-distribution = 35–8/(84–8).

In the simulations, the PDFs of the β-distribution were found using various specified values of Median and μ, or Mode and μ, ranging from 0 to 1. We used a decision tree to select between the Mode and the Median (Figure S9). Using only the Mode or Median to find p and q in the first approximation is more accurate from the perspective of experimental uniformity. However, it limits the range of possible distributions that can be represented, since some distributions can only be described using the Median. For example, the Mode cannot be used to describe a uniform distribution since the Mode value may be repeated several times in the lifespan series. Conversely, for some distributions, the Mode is more convenient since it is often located farther from μ, creating greater curvature of the graph and allowing it to better represent the frequency series.

If the equations p and/or q were calculated and the resulting values were less than 1 (the necessary condition for the equations to be valid is $p \geq 1$ and $q \geq 1$), then the standard values of p and q, which are both equal to 3.5, were used, resulting in a distribution that resembles a normal, bell-shaped curve.

In the Python script for visualizing the PDF of the β-distribution on graphs, we used μ, mode, and median (depending on the decision tree) from the original sample or the intervalised series and superimposed them on the frequency series on the graphs. The zero values of the β-distribution at its edges allow us to omit the edges of the distribution on the X-axis. The μ for the experimental sample was found using formula #4. The following formula was used to find the median (0.5):

$$F(m) = \int_0^m f(x)dx = 0.5 \tag{34}$$

where, *m* is the median (50th percentile of the sample), x is the upper limit of integration, f(x) is the function describing the lifespan at time *x*, and *dx* is an infinitesimal increment of *x*.

4. Discussion

4.1. The Situation with the Study of Lifespan Distributions in General

It is believed that human lifespan in heterogeneous populations is normally distributed in large data sets, but is skewed to the right [38]. At the same time, it is obvious that the rightward skew of human lifespan distributions (i.e., toward later ages of mortality) is a consequence of the overall level of development of human culture, living conditions, and medicine, and this situation has nothing to do with the survival curves of humans and apes in the wild [39]. In this regard, it is very difficult to understand any "natural" patterns of change in the distribution of lifespan by studying mortality statistics in human populations.

The distribution of lifespan in C. elegans is also considered normal. However, as with humans, it can be skewed depending on the conditions in which the worms are cultivated [40]. Therefore, C. elegans, being the shortest-lived multicellular model organism commonly used in laboratories, are a promising subject for studying lifespan distributions. However, aging in C. elegans is highly plastic and can be easily delayed, as these worms possess a programme of age-related suicide [41,42]. Perhaps aging in worms is even more complex, as it can be represented by two components that can be isolated using mathematical approaches and various cultivation conditions [40]. For these reasons, studying patterns of change in lifespan distribution in worms requires caution. On the other hand, the lifespan distribution in worms changes in response to the introduction of mutations, in a manner similar to what we observed in Drosophila under our experimental conditions [43].

Drosophila, another short-lived invertebrate model commonly used in laboratories, can be used to study the effects of various interventions (environmental—housing conditions/lifestyle, diet, as well as pharmacological and genetic, e.g., introduced mutations or transgenes) and is widely represented in studies of longevity and lifespan [2,18–20,22–24]. Surprisingly, the survival patterns in Drosophila mirror those in humans. This phenomenon was remarkably discovered at the beginning of the 20th century by the distinguished researcher Professor Raymond Pearl [44], who proposed using lifespan as the main indicator of fitness [45]. Although the number of offspring produced is currently considered the main criterion for genotype fitness [46–48], lifespan is nevertheless one of the main components of an individual's Darwinian fitness (life cycle traits), along with age of sexual maturity, fecundity, fertility and age-related survival dynamics [49,50]. Thus, lifespan is, if not the most important, then at least one of the most important components of overall fitness.

Although scientific literature contains many descriptions of experiments on the lifespan of Drosophila [51], accompanied by survival curves, there are virtually no descriptions of lifespan distribution in Drosophila. Even fewer studies have examined the patterns of change in these distributions. The literature contains only fragmentary references stating that the lifespan distributions in Drosophila are normal [52]. In this study, we demonstrate that this is not the case. Our data on the distribution of phenotype frequencies by lifespan show that the lifespan trait can be distributed very diversely in the studied samples, depending on the genetic background of the line. There are normal and non-normal distributions; distributions with high initial mortality and high late survival; and distributions resembling uniform or linear ones without a pronounced bell-shaped

rise in the center (see Figure 5D, for example). The frequencies of phenotypes by lifespan are distributed uniquely and characteristically for each line.

4.2. The Destabilisation of Ontogenesis and Changes in Lifespan Distribution Under Genetic Interventions

Having analysed the distribution of Drosophila lifespan series, we concluded that introducing mutations leads to changes in the frequencies of short- and long-lived phenotypes across the entire frequency range, relative to the control sample (see Figures 2, 3 and 4). This was observed when introducing mutations that increased (allele w^{67c23}) or decreased (allele w^{1118}) lifespan. Thus, a general destabilisation of ontogenesis is manifested, which is expressed in a change in the frequencies of phenotypes across the entire distribution.

Furthermore, our data show that genetic interventions (in this case, the *white* mutation) interact differently with the genotypes of different laboratory lines. Therefore, the mechanism by which the mutation affects lifespan may vary depending on the genetic background. Due to genetic drift in small, isolated populations, laboratory Drosophila lines accumulate genetic changes independently. This means that the interaction between a known mutation and a set of mutations unique to each line affects ontogenesis in its own way.

In this study, we did not specifically examine the asymmetry or kurtosis of distributions. This is a distinct task that requires its own research and in-depth analysis. The mathematical apparatus involved is quite complex [53], However, a brief examination of the graphs obtained using our methods reveals the following:

i) The introduction of the w^{1118} 2024 allele (reduces lifespan) shifts the distribution to the left relative to the control (Figure S6), whereas the introduction of the w^{67c23} 2024 allele (increases lifespan) shifts the distribution to the right (see Figure S7).

ii) When lifespan decreases, the distribution as a whole becomes narrower and higher (Figure S6A, B, D, σ decreases), or shifts without changing fundamentally (Figure S8B,D). Conversely, when lifespan increases, the distribution shifts but does not fundamentally change (Figure S7C), or it becomes wider and flatter (Figure S7A,B,D, σ increased). All of this reflects changes in the distribution in response to genetic interventions.

4.3. Selection of Tests for Analysing Lifespan Data to Determine Normality

The text continues here. At least 40 statistical tests have been developed to analyse the distribution of single samples [54]. The main criteria for the tests that are used are convenience (ease of calculation) and indicativeness (the ability to identify patterns). The most popular and widely used tests are the Kolmogorov–Smirnov test (KS test) [30,55] and the Shapiro–Wilk test (SW test) [34,56]. Both tests satisfy both criteria in their classical versions.

The KS test is a universal test that can be used to check whether a set of data belongs to a particular distribution, such as uniform, non-uniform, normal or non-normal, or exponential or linear. To assess normality or non-normality, the cumulative density function (CDF) should be used. Other functions are used for other types of distribution (e.g. the linear function). This test can be used with both discrete and continuous data.

The p-value of the KS test is typically calculated using the formula originally developed by Kolmogorov in 1933 [25,26,31]. As an alternative, a number of more contemporary and efficacious formulae (simpler to calculate) have been proposed in the literature. For instance, the formula advanced by Marsaglia [57]:

$$P = 2 * e^{-\left(2,000071 + \frac{0,331}{\sqrt{N}} + \frac{1,409}{N}\right) * N * (D^2)} \tag{35}$$

where, P is the p-value, *e* is the exponent, *N* is the sample size, and *D* is the KS test statistic.

We tested this formula and compared its performance with Kolmogorov's classic formula for intervalised and non-intervalised lifespan series for our experimental samples (Table S3). It turned out that in the case of non-intervalised series (samples with large N), the discrepancy in p-values

between the two formulas is observed in the third decimal place, i.e. both formulas are equivalent. However, for intervalised series (small samples), Marsaglia's formula sometimes gives p-values greater than 1 (e.g. 1.391) (Table S3). Therefore, Marsaglia's formula is not applicable when N<100; instead, Kolmogorov's formula should be used. For N<50, Lilliefors' percentage point table is required [58].

The SW test is based on Johnson's approximation method using the SB model, which approximates the observed distribution to a normal distribution [59]. In their 1965 paper, Shapiro and Wilk calculated critical P values for specific percentage points (0.01, 0.02, 0.05, 0.10, 0.50, 0.90, 0.95, 0.98 and 0.99) for N ranging from 3 to 50 [33]. This is a significant drawback of the original version of the SW test (generalized SW test) for lifespan studies since these studies involve analysing much larger data sets. Therefore, for samples of size greater than 50, the Royston method can be used. This method proposes a modified formula for calculating the SW test statistic W, new coefficients for calculating W and an algorithm for calculating the p-value for any sample size [34].

Like Johnson, Royston transforms the observed distribution z to normal, but uses a more sophisticated method of calculating W than in the generalised SW test. This method employs a function containing a polynomial to calculate W and P. The procedure for calculating W using Royston's method is the same for all sample sizes from 12 to 5,000. However, it differs for samples smaller than 12, for which other formulas are used to find g(X), μ_z and σ_z. However, despite Royston's method being widely used in online calculators, upon closer inspection, the Royston calculation procedure was found to be too cumbersome and does not meet the ease-of-use criterion. Therefore, the complexity of the calculation system and the associated risk of technical errors preclude the use of this method in thorough manual analysis of lifespan data series. Furthermore, when we calculated W and P for non-intervallised lifespan series using the Royston method, we found that all the analysed distributions were non-normal (see Tables 1, 2 and 3, columns 5 and 6). Therefore, for original lifespan data not divided into intervals (i.e. samples with large N), the SW test in Royston's modification is inconclusive. Therefore, in order to analyse lifespan series, we proposed dividing survival data into intervals and using the W calculation system from the standard generalised SW test. We combined this with the Royston method's procedure for accurately calculating P for any sample size. This approach overcomes the limitation of a sample size of 50 values and allows the use of a simple W calculation system based on the original work by Shapiro and Wilk.

The division of survival data into intervals and the pooling of mortality on adjacent days, which we propose as a means of overcoming the problem of daily mortality fluctuations, essentially corresponds to the counting of dead individuals every 2-3 days of life [18,60,61] or even once a week [62], which is acceptable in lifespan studies. In addition, combining mortality on adjacent days eliminates several shortcomings of daily counting:

i) inaccuracy in determining the date of death. For example, an individual may die immediately after testing on the same day but be recorded as having died the next day.

ii) the stochastic effect of transferring flies to new food. For example, if the food spoils prematurely, it is replaced earlier than after three days, which may affect lifespan. Clearly, combining the mortality of adjacent days corrects for these effects.

According to our data from 24 samples, the maximum lifespan range (i.e. the age at which 100% of individuals die) of laboratory Drosophila lines is 50–84 days at 25 °C. Therefore, when divided into 5-day intervals to determine frequencies, there will be 12-17 analysed segments/intervals/frequencies of phenotypes by lifespan. If one of the samples differs in the number of intervals, zero intervals (columns on graphs with a value of 0) may be introduced for the convenience of sample comparison on graphs. However, when calculating KS and SW tests, zero intervals (initial intervals that do not contain dead individuals) must be strictly excluded. For example, in the case of an additional zero interval(s) for the KS test, the value of D_n will change since the CDF values will shift relative to the Rank/N function, and D_{crit} will decrease because, as N increases, the value of $\sqrt{N}$ used to find the p-value will also increase (see Appendix #2). For the SW test, the values of the W test statistic and p-

value will change since an additional zero interval leads to a change in the number and composition of the coefficients used to calculate W and P (see Appendix #3).

When conducting simulations, we found that even when using perfectly normal data, frequency extraction (i.e. the intervalisation of the lifespan series) leads to the SW test producing some level of error when predicting normality (see section 2.3). Consequently, the SW test is not an absolutely ideal tool for accurately predicting normality in small samples. However, according to our data, the SW test is more accurate with intervalisation than without, and more accurate than the KS test with or without intervalisation (see section 2.3). We do not rule out the possibility that other tests for assessing normality for small samples may exist that are more efficient. This issue requires further research. Perhaps new tests based on principles of normality assessment other than the classic SW and KS tests need to be developed for small samples.

5. Conclusions

The present study offers a solution to the problem of representing survival data in a form alternative to conventional survival curves. A methodology for distribution assessment employing normality tests is proposed, complemented by visualization techniques based on the normal distribution function and the β-function. Representations of lifespan distributions were derived for laboratory lines of Drosophila melanogaster — including WT-like lines and WT-like lines subjected to a model genetic intervention (the white mutation). It has been demonstrated that genetic interventions can modify the frequencies of lifespan phenotypes, either by altering the distribution pattern or independently of such alterations.

The main results of the study can be summarised as follows:

1. Intervalised (e.g. 5-day) series of frequencies/probabilities of phenotypes by lifespan can be obtained from the initial lifespan series. Series of two samples superimposed on each other clearly show differences in the structure of the samples being compared. This method is an alternative to survival curves for displaying lifespan data. Dividing lifespan data into discrete intervals also minimises the impact of daily mortality fluctuations (eliminating mortality gaps and outliers), making the data suitable for further analysis using the normality criterion.

2. The best conditions for analysing lifespan data using the normality criterion are as follows: Use of the Shapiro-Wilk test on intervalised data for analysis (the Shapiro-Wilk test on original data, as well as the Kolmogorov-Smirnov test on intervalised and original data, are not indicative) 2) Intervalisation of lifespan data according to Sturges' rule with an equal number of intervals between the compared samples (the size of the interval between samples may differ) and a sample size of up to 200 individuals.

3. The frequency series of phenotypes by lifespan can be formally described using the PDF of a normal distribution (bell-shaped curve), or a PDF of a β-distribution (skewed, left- or right-sided curve, with a shifted center, or a symmetrical, bell-shaped curve, with parameters p and q equal to 3.5). This description allows for a quick and clear comparison of samples with each other («at a glance»).

4. The distributions of the lifespan trait in laboratory (highly inbred) Drosophila lines are diverse, depend on the genotype of the line, and change when mutations are introduced. Often, the distributions of Drosophila lifespan do not meet the formal statistical criterion of normality of distribution. Bell-shaped distributions account for the majority of distributions. However, there are also distributions resembling uniform/linear and left- and right-skewed β-distributions with a shifted center.

Thus, analysing survival data using distribution analysis is more diverse and sensitive than the standard method of comparing samples based on median and maximum lifespan on survival curves. The proposed approach to analysing lifespan distributions, and the use of the normality criterion for this purpose, could be valuable tools in lifespan and aging research, offering a new perspective on survival data. Understanding the distribution of phenotype frequencies by lifespan provides a more detailed picture of the processes affecting lifespan within the studied population, enabling more accurate tracking of the impact of genetic and pharmacological interventions on lifespan and ontogenesis.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Supplementary word file (contains Figures S1-9, Tables S1-S3), Supplementary Excel file (contain lists Appenix #2,3a,3b,5), Python script (.ipynb file), File .xml «Example data for calculations using a Python script» (contains two mortality rows for practicing script usage).

References

1. Falconer, D.S.; Mackay, T.F.C. *Introduction to Quantitative Genetics.*; Longman: London, UK, 1996, 1996;

2. Bylino, O.V.; Ogienko, A.A.; Batin, M.A.; Georgiev, P.G.; Omelina, E.S. Genetic, Environmental, and Stochastic Components of Lifespan Variability: The Drosophila Paradigm. *Int J Mol Sci* **2024**, *25*, 4482, doi:10.3390/ijms25084482.

3. Olshansky, S.J.; Willcox, B.J.; Demetrius, L.; Beltrán-Sánchez, H. Implausibility of Radical Life Extension in Humans in the Twenty-First Century. *Nat Aging* **2024**, *4*, 1635–1642, doi:10.1038/s43587-024-00702-3.

4. Flatt, T.; Partridge, L. Horizons in the Evolution of Aging. *BMC Biology* **2018**, *16*, 93, doi:10.1186/s12915-018-0562-z.

5. 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* **1980**, *36*, 607–625, doi:10.2307/2556114.

6. Mantel, N. Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. *Cancer Chemother Rep* **1966**, *50*, 163–170.

7. Wang, C.; Li, Q.; Redden, D.T.; Weindruch, R.; Allison, D.B. Statistical Methods for Testing Effects on "Maximum Lifespan." *Mech Ageing Dev* **2004**, *125*, 629–632, doi:10.1016/j.mad.2004.07.003.

8. Moskalev, A.; Shaposhnikov, M. Pharmacological Inhibition of NF-κB Prolongs Lifespan of Drosophila Melanogaster. *Aging (Albany NY)* **2011**, *3*, 391–394, doi:10.18632/aging.100314.

9. McHugh, M.L. The Chi-Square Test of Independence. *Biochem Med (Zagreb)* **2013**, *23*, 143–149, doi:10.11613/BM.2013.018.

10. Sturges, H.A. The Choice of a Class Interval. *Journal of the American Statistical Association* **1926**, *21*, 65–66.

11. Morgan, T.H. Sex Limited Inheritance in Drosophila. *Science* **1910**, *32*, 120–122, doi:10.1126/science.32.812.120.

12. Green, M.M. 2010: A Century of Drosophila Genetics through the Prism of the White Gene. *Genetics* **2010**, *184*, 3–7, doi:10.1534/genetics.109.110015.

13. Myers, J.; Porter, M.; Narwold, K.; Bhat, K.; Dauwalder, B.; Roman, G. Mutants of the White ABCG Transporter in Drosophila Melanogaster Have Deficient Olfactory Learning and Cholesterol Homeostasis. *International journal of molecular sciences* **2021**, *22*, doi:10.3390/ijms222312967.

14. Sasaki, A.; Nishimura, T.; Takano, T.; Naito, S.; Yoo, S.K. White Regulates Proliferative Homeostasis of Intestinal Stem Cells during Ageing in Drosophila. *Nat Metab* **2021**, *3*, 546–557, doi:10.1038/s42255-021-00375-x.

15. Xiao, C.; Qiu, S. Frequency-Specific Modification of Locomotor Components by the White Gene in Drosophila Melanogaster Adult Flies. *Genes Brain Behav* **2021**, *20*, e12703, doi:10.1111/gbb.12703.

16. Mendoza-Grimau, V.; Pérez-Gálvez, A.; Busturia, A.; Fontecha, J. Lipidomic Profiling of Drosophila Strains Canton-S and White1118 Reveals Intraspecific Lipid Variations in Basal Metabolic Rate. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **2024**, *201*, 102618, doi:10.1016/j.plefa.2024.102618.

17. Rickle, A.; Sudhakar, K.; Booms, A.; Stirtz, E.; Lempradl, A. More than Meets the Eye: Mutation of the White Gene in Drosophila Has Broad Phenotypic and Transcriptomic Effects. *Genetics* **2025**, iyaf097, doi:10.1093/genetics/iyaf097.

18. Linford, N.J.; Bilgir, C.; Ro, J.; Pletcher, S.D. Measurement of Lifespan in Drosophila Melanogaster. *J Vis Exp* **2013**, 50068, doi:10.3791/50068.

19. He, Y.; Jasper, H. Studying Aging in Drosophila. *Methods* **2014**, *68*, 129–133, doi:10.1016/j.ymeth.2014.04.008.

20. Piper, M.D.W.; Partridge, L. Protocols to Study Aging in Drosophila. *Methods Mol Biol* **2016**, *1478*, 291–302, doi:10.1007/978-1-4939-6371-3_18.

21. Piper, M.D.W.; Partridge, L. Drosophila as a Model for Ageing. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2018**, *1864*, 2707–2717, doi:10.1016/j.bbadis.2017.09.016.

22. Ogienko, A.A.; Omelina, E.S.; Bylino, O.V.; Batin, M.A.; Georgiev, P.G.; Pindyurin, A.V. Drosophila as a Model Organism to Study Basic Mechanisms of Longevity. *Int J Mol Sci* **2022**, *23*, 11244, doi:10.3390/ijms231911244.

23. Louka, X.P.; Gumeni, S.; Trougakos, I.P. Studying Cellular Senescence Using the Model Organism Drosophila Melanogaster. *Methods Mol Biol* **2025**, *2906*, 281–299, doi:10.1007/978-1-0716-4426-3_17.

24. Na, H.-J.; Park, J.-S. Measuring Anti-Aging Effects in Drosophila. *Bio Protoc* **2025**, *15*, e5305, doi:10.21769/BioProtoc.5305.

25. Kolmogoroff, A. "Sulla Determinazione Empirica Di Una Legge Di Distribuzione." *"Giornale dell' Istituto Italiano degli Attuari.* **1933**, *4*, 83–91.

26. Feller, W. On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *The Annals of Mathematical Statistics* **1948**, *19*, 177–189, doi:10.1214/aoms/1177730243.

27. Smirnov, N. On the Estimation of the Discrepancy between Critical Curves of Distribution of Two Independent Samples. *Bulletin Mathématique de l'Université de Moscou.* **1939**, *2*, 1–16.

28. Miller, L.H. Table of Percentage Points of Kolmogorov Statistics. *Journal of the American Statistical Association* **1956**.

29. Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **1951**, *46*, 68–78, doi:10.1080/01621459.1951.10500769.

30. Facchinetti, S. A Procedure to Find Exact Critical Values of Kolmogorov-Smirnov Test. *Statistica Applicata* **2009**, 337–359.

31. Smirnov, N. Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics* **1948**, *19*, 279–281.

32. Hodges, J.L. The Significance Probability of the Smirnov Two-Sample Test. *Ark. Mat.* **1958**, *3*, 469–486, doi:10.1007/BF02589501.

33. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611, doi:10.2307/2333709.

34. Royston, P. A Toolkit for Testing for Non-Normality in Complete and Censored Samples. *Journal of the Royal Statistical Society. Series D (The Statistician)* **1993**, *42*, 37–43, doi:10.2307/2348109.

35. Warne, R.T. *Statistics for the Social Sciences: A General Linear Model Approach*; 2nd ed.; Cambridge University Press, 2020; ISBN 978-1-108-89431-9.

36. Royston, P. Remark AS R94: A Remark on Algorithm AS 181: The W-Test for Normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1995**, *44*, 547–551, doi:10.2307/2986146.

37. Cramer, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, 1957; Vol. Vol. 7.;

38. Robertson, H.T.; Allison, D.B. A Novel Generalized Normal Distribution for Human Longevity and Other Negatively Skewed Data. *PLoS One* **2012**, *7*, e37025, doi:10.1371/journal.pone.0037025.

39. Skulachev, V.P.; Shilovsky, G.A.; Putyatina, T.S.; Popov, N.A.; Markov, A.V.; Skulachev, M.V.; Sadovnichii, V.A. Perspectives of Homo Sapiens Lifespan Extension: Focus on External or Internal Resources? *Aging* **2020**, *12*, 5566–5584, doi:10.18632/aging.102981.

40. Stroustrup, N.; Anthony, W.E.; Nash, Z.M.; Gowda, V.; Gomez, A.; López-Moyado, I.F.; Apfeld, J.; Fontana, W. The Temporal Scaling of Caenorhabditis Elegans Ageing. *Nature* **2016**, *530*, 103–107, doi:10.1038/nature16550.

41. Ezcurra, M.; Benedetto, A.; Sornda, T.; Gilliat, A.F.; Au, C.; Zhang, Q.; van Schelt, S.; Petrache, A.L.; Wang, H.; de la Guardia, Y.; et al. C. Elegans Eats Its Own Intestine to Make Yolk Leading to Multiple Senescent Pathologies. *Curr Biol* **2018**, *28*, 2544-2556.e5, doi:10.1016/j.cub.2018.06.035.

42. Kern, C.C.; Srivastava, S.; Ezcurra, M.; Hsiung, K.C.; Hui, N.; Townsend, S.; Maczik, D.; Zhang, B.; Tse, V.; Konstantellos, V.; et al. C. Elegans Ageing Is Accelerated by a Self-Destructive Reproductive Programme. *Nat Commun* **2023**, *14*, 4381, doi:10.1038/s41467-023-40088-1.

43. Chen, J.; Senturk, D.; Wang, J.-L.; Muller, H.-G.; Carey, J.R.; Caswell, H.; Caswell-Chen, E.P. A Demographic Analysis of the Fitness Cost of Extended Longevity in Caenorhabditis Elegans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **2007**, *62*, 126–135, doi:10.1093/gerona/62.2.126.

44. Pearl, R.; Parker, S. Experimental Studies on the Duration of Life. VI. A Comparison of the Laws of Mortality in Drosophila and in Man. *The American Naturalist* **1922**, *56*, 398–405, doi:10.1086/279879.

45. Pearl, R. Duration of Life as an Index of Constitutional Fitness*. *Poultry Science* **1923**, *3*, 1–10, doi:10.3382/ps.0030001.

46. Pfenninger, M. The Effect of Offspring Number on the Adaptation Speed of a Polygenic Trait 2017.

47. Krohs, U. Darwin's Empirical Claim and the Janiform Character of Fitness Proxies. *Biol Philos* **2022**, *37*, 15, doi:10.1007/s10539-022-09847-0.

48. Viblanc, V.A.; Saraux, C.; Tamian, A.; Criscuolo, F.; Coltman, D.W.; Raveh, S.; Murie, J.O.; Dobson, F.S. Measuring Fitness and Inferring Natural Selection from Long-Term Field Studies: Different Measures Lead to Nuanced Conclusions. *Behav Ecol Sociobiol* **2022**, *76*, 79, doi:10.1007/s00265-022-03176-8.

49. Orr, H.A. Fitness and Its Role in Evolutionary Genetics. *Nat Rev Genet* **2009**, *10*, 531–539, doi:10.1038/nrg2603.

50. Flatt, T. Life-History Evolution and the Genetics of Fitness Components in Drosophila Melanogaster. *Genetics* **2020**, *214*, 3–48, doi:10.1534/genetics.119.300160.

51. Clancy, D.; Chtarbanova, S.; Broughton, S. Editorial: Model Organisms in Aging Research: Drosophila Melanogaster. *Front Aging* **2023**, *3*, 1118299, doi:10.3389/fragi.2022.1118299.

52. Izmaïlov, D.M.; Obukhova, L.K. Analysis of the life span distribution mode in 128 successive generations of D. melanogaster. *Adv Gerontol* **2004**, *15*, 30–35.

53. Azzalini, A. The Skew-Normal Distribution and Related Multivariate Families*. *Scand J Stat* **2005**, *32*, 159–188, doi:10.1111/j.1467-9469.2005.00426.x.

54. Arnastauskaitė, J.; Ruzgas, T.; Bražėnas, M. An Exhaustive Power Comparison of Normality Tests. *Mathematics* **2021**, *9*, 788, doi:10.3390/math9070788.

55. Simard, R.; L'Ecuyer, P. Computing the Two-Sided Kolmogorov-Smirnov Distribution. *Journal of Statistical Software* **2011**, *39*, 1–18, doi:10.18637/jss.v039.i11.

56. Mara, U.T. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. **2011**, *2*, 21–33.

57. Marsaglia, G.; Tsang, W.W.; Wang, J. Evaluating Kolmogorov's Distribution. *Journal of Statistical Software* **2003**, *8*, 1–4, doi:10.18637/jss.v008.i18.

58. Lilliefors, H.W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association* **1967**, *62*, 399–402, doi:10.1080/01621459.1967.10482916.

59. Johnson, N.L. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* **1949**, *36*, 149, doi:10.2307/2332539.

60. Chakraborty, T.S.; Gendron, C.M.; Lyu, Y.; Munneke, A.S.; DeMarco, M.N.; Hoisington, Z.W.; Pletcher, S.D. Sensory Perception of Dead Conspecifics Induces Aversive Cues and Modulates Lifespan through Serotonin in Drosophila. *Nat Commun* **2019**, *10*, 2365, doi:10.1038/s41467-019-10285-y.

61. Lin, Y.-C.; Zhang, M.; Wang, S.-H.; Chieh, C.-W.; Shen, P.-Y.; Chen, Y.-L.; Chang, Y.-C.; Kuo, T.-H. The Deleterious Effects of Old Social Partners on Drosophila Lifespan and Stress Resistance. *npj Aging* **2022**, *8*, 1, doi:10.1038/s41514-022-00081-2.

62. Bushey, D.; Hughes, K.A.; Tononi, G.; Cirelli, C. Sleep, Aging, and Lifespan in Drosophila. *BMC Neuroscience* **2010**, *11*, 56, doi:10.1186/1471-2202-11-56.