

# Large standard deviations and logarithmic-normality

## The truth about hemocyte counts in *Drosophila*

Richard Paul Sorrentino

Notre Dame Integrated Imaging Facility (NDIIF) and Department of Biological Sciences; University of Notre Dame; Notre Dame, IN USA

**Key words:** hemocyte, CHC, concentration, logarithm, normal, log-normal, blood

**Abbreviations:** CHC, circulating hemocyte concentration

While many quantifiable biological phenomena can be described by making use of an assumption of normality in the distribution of individual values, many biological phenomena are not accurately described by the normal distribution. An unquestioned assumption of normality of distribution of possible outcomes can lead to misinterpretation of data, which could have serious consequences. Thus it is extremely important to test the validity of an assumption of normality of possible outcomes. As it turns out, the logarithmic-normal (log-normal) distribution pattern is often far more accurate in describing statistical biological phenomena. Herein I examine large samples of values for circulating blood cell (hemocyte) concentration (CHC) among both wild-type and mutant *Drosophila* larvae, and demonstrate in both cases that the distribution of individual values does not conform to normality, but does conform to log-normality.

### Introduction

“The *standard deviation* and the *standard error of the mean* measure two different things and are often confused. Most medical investigators summarize their data with the standard error of the mean because it is always smaller than the standard deviation. It makes their data look better. However, unlike the standard deviation, which quantifies the *variability in the population*, the standard error quantifies *uncertainty in the estimate of the mean*. Since readers are generally interested in knowing about the population, data should generally not be summarized with the standard of the mean.”<sup>1</sup>

Blood cell (hemocyte) counts can be simply and quickly obtained, and are highly informative indications of an effect of experimental manipulation of *Drosophila*. Depending on the needs of a given researcher, values can be obtained for total hemocytes per organism or for hemocyte concentration. When experiments include appropriate control classes, such indices can provide immediate clues about the comparative effects of treatments on hemocyte proliferation and differentiation or on the innate cellular immune response.

While the methods used to obtain such counts vary, the read-out of any experiment is the same: a sample of raw count (or concentration) values. Thus the method of statistical analysis of obtained values should be relatively uniform. However, it is not. Most reports on hemocyte count/concentration use of

the standard error of the mean,<sup>2-6</sup> while a minority employ the standard deviation.<sup>7-9</sup> Most reports apparently assume a normal distribution when comparing means, while a very few transform raw values prior to analysis.<sup>10-12</sup>

There are certainly multiple reasons for this lack of uniformity, but a critically important reason that will be addressed herein is the apparent misperception that large sample standard deviations (i.e., large with respect to the values of the mean) are at best visually unappealing and at worst necessarily erroneous. Consequently, a researcher may decide to present true—but visually unappealing—information in a secondary online supplementary file,<sup>13</sup> or use the standard error of the mean or even resort to the practice of “dropping outliers”.

The standard error of the mean is not a substitute for the standard deviation. The standard deviation is a measure of the spread of  $n$  raw values (from one experiment or from one pooling of multiple experiments). The standard error, however, is the standard deviation not of  $n$  raw values, but of  $n$  experiments. It is a measure of the spread of  $n$  mean values obtained in  $n$  experiments, of a “mean of means”. Thus the employment of the standard error, which (when dealing with whole nonnegative integers, such as CHC values) is always going to be less than the standard deviation (standard error = standard deviation divided by the square root of  $n$ ). The standard error is more visually appealing, and there is a greater likelihood of demonstrating statistically “significant” differences between experimental groups.

Correspondence to: Richard Paul Sorrentino; Email: Sorrentino.5@nd.edu

Submitted: 05/24/10; Revised: 07/23/10; Submitted: 08/03/10

Previously published online: [www.landesbioscience.com/journals/fly/article/13260](http://www.landesbioscience.com/journals/fly/article/13260)

DOI: 10.4161/fly.4.4.13260

However, such a practice can lead to a much greater chance of a false conclusion of a significant difference between means.

Yet a second factor contributes to the observed large standard deviations: the erroneous assumption that raw CHC values must necessarily conform to a normal distribution. Yet this is an unreasonable expectation, as multiple factors work against such conformity: (1) CHC values cannot be negative; (2) other than zero, there is no lower limit to wild-type or mutant CHC; (3) there is no practical upper limit to wild-type and mutant CHC values. It ought not to be surprising that this confluence of factors leads to a skewed distribution pattern: higher frequencies between zero and the mean and lower frequencies, but more numerous frequency classes, above the mean. The drive to force CHC values into a normal distribution pattern is not unfounded: many biological indices (e.g., human lean body mass index<sup>14</sup>) do indeed conform to normality. But many do not (e.g., height and weight among domestic dogs<sup>15</sup>). Thus it is up to the researcher to determine whether the assumption of normality is valid.

In this report, I present evidence that the distribution pattern of a large sample of raw larval CHC values, either wild-type or mutant, violates normality, but is consistent with a logarithmic-normal (log-normal) pattern. An assumption of log-normality allows the calculation of standard deviations that are visually appealing, that accurately summarize raw data, and that do not unduly increase the risk of false negative and false positive interpretations.

## Results

In order to determine whether wild-type total CHC values conform to a normal distribution pattern, I performed an analysis of published ( $n = 50$ )<sup>12</sup> and as-yet-unpublished ( $n = 150$ ) individual CHC values (total:  $n = 200$ ; females: 90; males: 110) obtained for *Canton-S* larvae. All CHC values were obtained under the same experimental conditions. CHC values were obtained as described<sup>16</sup> from larvae that were  $120 \pm 4$  hr old, and had been raised at 25°C. A summary of statistical data for wild-type raw total CHC values are presented in **Table 1**.

Qualitative examination of the distribution pattern of raw wild-type CHC values (**Fig. 1A**) reveals a mixture of characteristics, not all of which are consistent with expectations of a normal distribution. For example, consistent with a normal distribution, the mean and median fall into the modal bin. However, further examination reveals contradictory information. Because CHC values cannot be negative, and because there is no symmetrically located upper limit to CHC, the histogram is noticeably skewed to the right. Statistical analysis confirms qualitative impressions. Skewness ( $\sqrt{b_1}$ ) was determined to be 1.6514, in excess of the critical value of 0.403 at  $p < 0.01$  (**Table 2**). Furthermore, the kurtosis ( $b_2$ ) value is 6.9516, which lies outside the upper boundary of the spread of critical values, 3.98 ( $p < 0.01$ ; **Table 2**). The observed skewness and kurtosis values indicate a significant difference from those expected for a normal distribution.

Additional strong support for non-normality is provided by a  $\chi^2$  goodness-of-fit test. The observed value is  $\chi^2_{0.001} = 566.0911$  (degrees of freedom:  $\nu = 9$ ), which is much greater than the

**Table 1.** Statistical data

Parameter	Data treatment	
	Raw	Transformed
Minimum	1,010	6.9177
Maximum	29,020	10.2757
Range	28,010	3.3580
Mean $\pm$ standard deviation	7,932 $\pm$ 4,895	8.8030 $\pm$ 0.6898
Coefficient of variance	0.6171	0.0693
<b>Back-transformation</b>	-	<b>6,654 x/ 1.8400</b>
Median	6,983	8.8512
No. of bins	10	10
Bin size	2,930	0.3649

Comparison of statistical parameters of raw and transformed *Canton-S* total CHC data.

**Table 2.** Tests of normality for raw *Canton-S* CHC data ( $n = 200$ )

Parameter	Observed value	Critical value	p at critical value	Reject null hypothesis?
$\chi^2$	566.0911	27.88	0.001	reject
$\sqrt{b_1}$	1.6514	0.403	0.01	reject
$b_2$	6.9516	2.37–3.98	0.01	reject

Null hypothesis: there is no difference between the distribution pattern of the observed data and the normal distribution. The distribution of raw CHC values fails three tests for normality:  $\chi^2$  goodness-of-fit test ( $\nu = 9$  degrees of freedom), skewness ( $\sqrt{b_1}$ ) test, and kurtosis ( $b_2$ ) test. Note that all three failures are at very significant levels of probability.

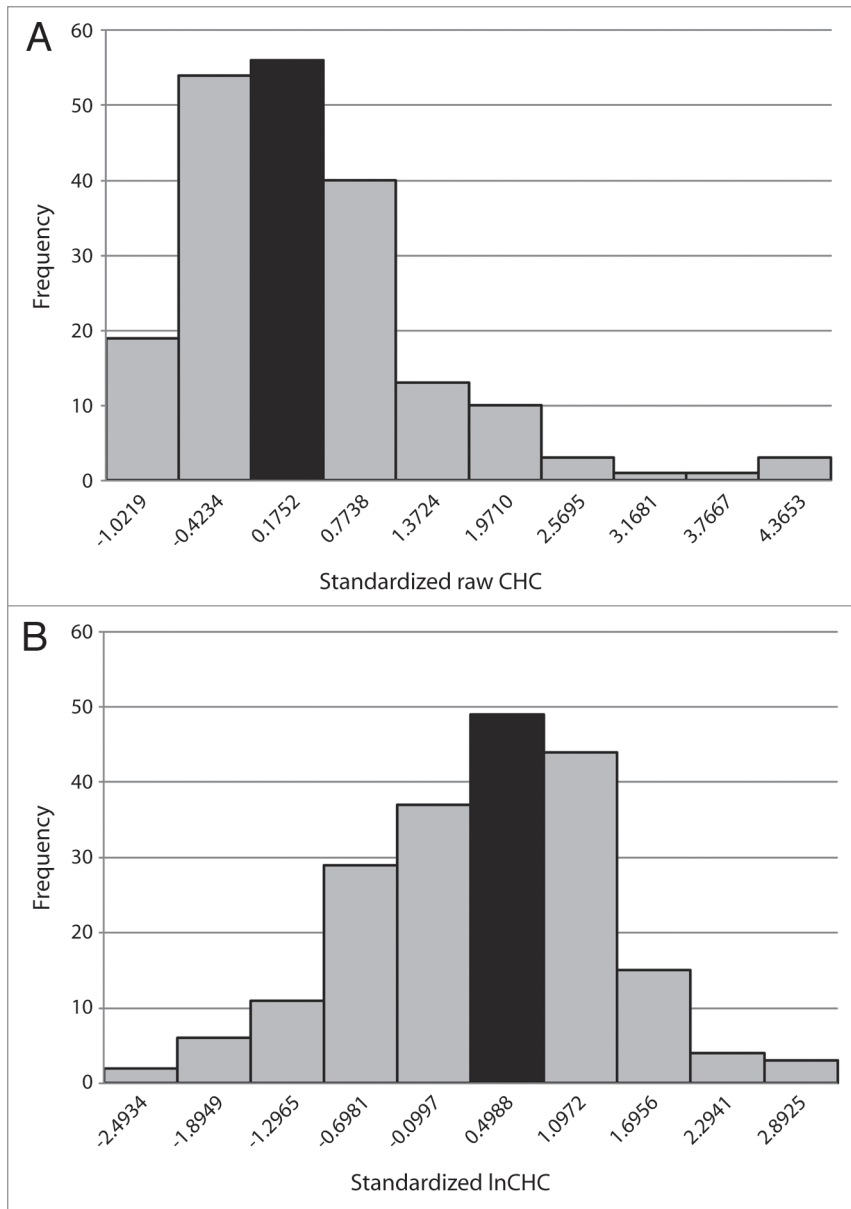
**Table 3.** Tests of normality for transformed *Canton-S* *ln*CHC data ( $n = 200$ )

Parameter	Observed value	Critical value	p at critical value	Reject null hypothesis?
$\chi^2$	7.2230	8.34	0.50	do not reject
$\sqrt{b_1}$	-0.2655	-0.280	0.05	do not reject
$b_2$	3.0774	2.59–3.40	0.10	do not reject

Null hypothesis: there is no difference between the distribution pattern of the transformed data and the normal distribution. The distribution of transformed *ln*CHC values passes three tests for normality:  $\chi^2$  goodness-of-fit test ( $\nu = 9$  degrees of freedom), skewness ( $\sqrt{b_1}$ ) test, and kurtosis ( $b_2$ ) test. Critical values those for the most significant tabulated p values<sup>19</sup> that were not surpassed by the observed test value.

critical value of 27.88 ( $p < 0.001$ ; **Table 2**). Thus, despite the qualitative indication of normality by the mean, median and mode, the distribution pattern in **Figure 1A** conclusively fails multiple tests for normality. The mean  $\pm$  standard deviation of raw *Canton-S* CHC is depicted in **Figure 2A**. Note that the magnitude of the standard deviation is over 60% that of the mean (**Table 1**, coefficient of variance; **Fig. 2A**).

In striking contrast, the distribution pattern of the natural logarithms of individual CHC values (*ln*CHC values; **Fig. 1B**) is obviously different. Again, the mean, median and mode fall into the same bin (**Table 1**). Yet the histogram appears far more similar to a normal pattern (**Fig. 1B**). Importantly, the results of tests for skewness, kurtosis and goodness-of-fit are all strongly consistent



**Figure 1.** (A) Histogram of raw *Canton-5* CHC values. Vertical axis: frequency. Horizontal axis: bins for ranges of CHC values. All bins are of equal width: 2,930 cells/microliter. Number under each bin represents the upper CHC boundary value of the bin (in standard units). The lower boundary value of the leftmost bin is zero. Black bar represents the modal bin and also contains the mean and median. (B) Histogram of transformed  $\ln$ CHC values. Vertical axis: frequency. Horizontal axis: bins for ranges of transformed  $\ln$ CHC values. All bins are of equal width: 0.3649. Number under each bin represents the upper  $\ln$ CHC boundary value of the bin (in standard units). The lower boundary value of the leftmost bin is 6.9177 (6.9177 is the minimum observed  $\ln$ CHC value). Black bar represents the modal bin and also contains the mean and median.

as lamellocytes only differentiate in large numbers in response to natural parasitic infection or a “tumorous” genetic background.<sup>21</sup> This raises the question of whether lamellocyte CHC values also conform to a log-normal distribution pattern. Analysis of lamellocyte CHC data from *u-shaped* (*ush*) mutant larvae ( $n = 40$ ; females: 20; males: 20),<sup>12</sup> which exhibit large numbers of lamellocytes in circulation, is presented in Tables 4–6. Once again, three separate tests demonstrate that untransformed data violate the assumption of normality (Tables 4 and 5), while data transformed by the natural logarithm conform to normality (Tables 4 and 6). Analysis of total CHC for these same larvae also demonstrated a violation of normality and a conformation to log-normality (data not shown).

## Discussion

with normality (Table 3). The mean  $\pm$  standard deviation of logarithmically transformed CHC values is presented in Figure 2B. Note that the magnitude of the standard deviation is less than 7% of that of the mean (Table 1, coefficient of variance; Fig. 2A and B). However, the unit of measurement in this case is an abstraction,  $\ln$ CHC. In order to express observations in useful real-world units (cells/microliter), a back-transformation is performed,<sup>20</sup> in which the exponentials of the mean and standard deviation are calculated (Table 1 and Fig. 2C). Note the difference between the mean raw (7,932 cells/microliter) and mean transformed (6,654 cells/microliter) values.

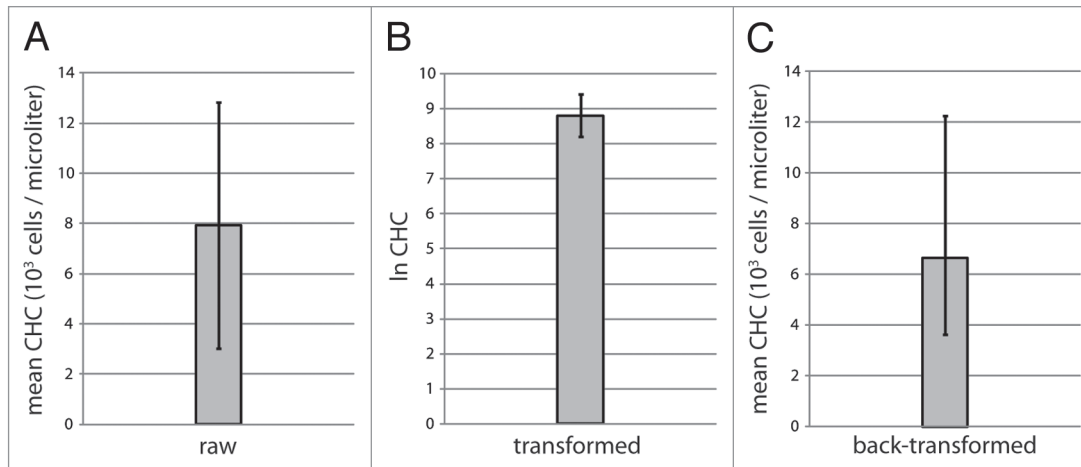
The data examined thus far represent total CHC in a wild-type genetic background, and no distinction was made among hemocyte classes. The hemocytes that were observed were almost entirely composed of plasmatocytes, with a small percentage (< 5%) of crystal cells. A third hemocyte class, that of the lamellocyte, is occasionally observed in wild-type hemolymph,

I have demonstrated that the distribution of a large sample ( $n = 200$ ) of wild-type total CHC values and that of *ush* mutant lamellocyte CHC values ( $n = 40$ ), despite initial qualitative tests, do not conform to normality. I further demonstrate that these same ranges of CHC values do not violate the log-normal distribution pattern. Two conclusions can be extracted from these observations: (1) the assumption of normality when comparing mean CHC values is erroneous; and (2) the assumption of log-normality is valid.

Thus I propose that analysis of raw hemocyte counts and concentrations be performed as follows:<sup>11,12,16,20,22</sup>

1. Assume log-normality: determine the  $\ln$  of each raw value (n.b.: when comparing not just total CHC, but CHC of hemocyte classes, sometimes a given sample might have a count of, say, zero lamellocytes; in such a case, add 1 to each raw value in all classes;  $\ln 1 = 0$ ).

2. Use the transformed values obtained in Step 1 to obtain the mean and standard deviations.



**Figure 2.** Bar graph presentations of mean and standard deviations of *Canton-S* CHC. (A) Raw CHC values. Mean  $\pm$  standard deviation. Vertical axis:  $10^3$  cells/microliter. (B) Transformed CHC values ( $\ln$ CHC). Mean  $\pm$  standard deviation. Vertical axis:  $\ln$ CHC. (C) Back-transformed CHC values. Mean  $\times$ /standard deviation. Vertical axis:  $10^3$  cells/microliter.

**Table 4.** Statistical data

Parameter	Data treatment	
	Raw	Transformed
Minimum	565	6.3370
Maximum	36,118	10.4946
Range	35,553	4.1576
Mean $\pm$ standard deviation	6,333 $\pm$ 8,154	8.2171 $\pm$ 1.0093
Coefficient of variance	1.2875	0.1228
<b>Back-transformation</b>	-	<b>3,704 <math>\times</math>/ 2.7436</b>
Median	3,544	8.1705
No. of bins	5	5
Bin size	8,345	1.0329

Comparison of statistical parameters of raw and transformed *ush* CHC data.

3. Use tests of significance (e.g., Student's *t*-test) using the values obtained in Step 2 and draw conclusions.

4. Back-transform the means and standard deviations obtained in Step 2: use the exponential function.

5. Express back-transformed mean and standard deviation as: mean  $\times$ /standard deviation (the term " $\times$ /" is expressed "times or divided by", just as " $\pm$ " is expressed "plus or minus").

Figures depicting comparisons of hemocyte counts/concentrations can still employ bar graphs and error bars (Fig. 2). Presentation of back-transformed values (Fig. 2C)<sup>12,21</sup> are ideal. However, transformed values can also be presented (Fig. 2B)<sup>11</sup> if appearance is a concern.

The comparison of CHC values for hemocyte classes such as lamellocytes is also a concern for many researchers. When all larval hemolymph samples in a given larval class produce non-zero lamellocyte CHC values (as they do in tumorous *ush* larvae) the log-normal transformation is valid (Tables 4–6; importantly, the same can be said of *ush* total CHC values; data not shown). It is suggested that a researcher nonetheless confirm log-normality,

**Table 5.** Tests of normality for raw *ush* CHC data ( $n = 40$ )

Parameter	Observed value	Critical value	p at critical value	Reject null hypothesis?
$\chi^2$	226.7998	18.47	0.001	reject
$\sqrt{b_1}$	2.5905	0.870	0.01	reject
$b_2$	9.5806	1.89–5.04	0.01	reject

Null hypothesis: there is no difference between the distribution pattern of the observed data and the normal distribution. The distribution of raw CHC values fails three tests for normality:  $\chi^2$  goodness-of-fit test ( $\nu = 4$  degrees of freedom), skewness ( $\sqrt{b_1}$ ) test, and kurtosis ( $b_2$ ) test. Note that all three failures are at very significant levels of probability.

**Table 6.** Tests of normality for transformed *ush*  $\ln$ CHC data ( $n = 40$ )

Parameter	Observed value	Critical value	p at critical value	Reject null hypothesis?
$\chi^2$	6.9487	7.78	0.10	do not reject
$\sqrt{b_1}$	0.3788	0.450	0.10	do not reject
$b_2$	2.7463	2.19–3.67	0.10	do not reject

Null hypothesis: there is no difference between the distribution pattern of the transformed data and the normal distribution. The distribution of transformed  $\ln$ CHC values passes three tests for normality:  $\chi^2$  goodness-of-fit test ( $\nu = 4$  degrees of freedom), skewness ( $\sqrt{b_1}$ ) test, and kurtosis ( $b_2$ ) test. Critical values are those for the most significant tabulated p values<sup>19</sup> that were not surpassed by the observed test value.

as different tumor-inducing genetic backgrounds may produce different distributions of lamellocyte CHC values. When lamellocyte populations are small (<3% of total CHC per larva) and not present in most larvae in a sample, then the distribution of values will include a very large zero bin and multiple small non-zero bins. While it is possible that a model for such a distribution may one day be described, I have not been able to find a probability distribution function that would apply to such a pattern. Furthermore, the distribution of just the non-zero lamellocyte CHC values in *Canton-S* larvae, both raw and transformed by the natural logarithm, fails

the test for normality (data not shown). Fortuitously, genetic backgrounds that produce lamellocytes in a minority of larvae tend to produce very low concentrations of lamellocytes. For instance, 26 of the 200 *Canton-S* larvae discussed herein exhibited at least one lamellocyte. Only one larva of the 26 exhibited a lamellocyte percentage of greater than 2.02%. Furthermore, I have neither observed nor read of genetic backgrounds that produce infrequent and large populations of lamellocytes; thus whether backgrounds that produce negligible infrequent and small lamellocyte populations actually need a test of significance to demonstrate how low they are is up to the researcher.

Of course, the values of log-normal transformations can be affected by the raw data units that are chosen. For example, transformation of 200 CHC values expressed in terms of cells per liter (as opposed to cells per microliter) produces an  $\ln$ CHC value of  $-5.0125 \pm 0.6098$ , which differ considerably from the values reported herein. But as long as a researcher can demonstrate or cite a log-normal distribution, then any tests of significance will be valid.

It must be pointed out that it is possible that there is at least one distribution pattern other than the log-normal that is an even better description of possible CHC values. For instance, regression analysis by way of the trendline function in Microsoft Excel can produce fourth- and fifth-order polynomial functions that fit far more closely ( $r > 0.99$ ) to the data than does the log-normal assumption. However, such functions, when extended even one or two standard deviations to the right of the boundary of data reported herein, predict huge frequencies in the hundreds and thousands (unpublished observations). Such predictions, in the absence of any supportive information, must be rejected for now.

It is also worth noting that not every published analysis of hemocyte counts/concentrations that has assumed normality is necessarily incorrect in its conclusions. Depending on the observed values, two means may have been separated widely enough that even the assumption of normality would not interfere with the detection of a difference. Importantly, the power of the justified assumption of log-normality lies in the ability to reveal significant differences between groups that would not have been evident otherwise.

In summary, the visually and intellectually unappealing large standard deviations that are obtained when analyzing hemocyte counts/concentrations should not be surprising in

the examination of biological systems. Large values for standard deviations of CHC are not necessarily a problem; the real problem is the unfounded and erroneous assumption of a normal distribution of raw values and a refusal to accept the nature and extent of variability in a given phenotypic character, even in an allegedly homogeneous laboratory population. Environment, chance and the as-yet-unexplained will always play roles in phenotype, including in hemocyte counts.

## Materials and Methods

**Fly stocks.** *Canton-S* served as a wild-type stock. *ush<sup>VX22/ush</sup>* larvae were obtained as described.<sup>12</sup>

**Egglays.** Flies were maintained using standard cornmeal-yeast-agar medium. All experiments took place at 25°C. Experimental egglays were allowed to take place in vials containing fly medium for four hours. After four hours had transpired, adults were moved to a new vial and any eggs laid during the egglay period were allowed to develop at 25°C. Hemolymph samples were obtained from individual feeding third-instar (L3) larvae at  $120 \pm 4$  hr after the initiation of the egglay.

**Determination of circulating hemocyte concentration (CHC).** CHC was determined as described.<sup>16</sup>

**Statistical analysis.**  $n$  = sample size. Minimum = lowest value in sample. Maximum = greatest value in sample. Range = maximum - minimum. Mean = (sum of all individual CHC values)/(number of CHC values). Standard deviation =  $[\sum(\text{individual CHC value} - \text{mean CHC})^2]/(n - 1)$ . Median = middle ordinal value or mean of two middle ordinal values, when all CHC values are listed in increasing order. Bin size:  $b = [3.5 \times (\text{standard deviation})]/n^{1/3}$ .<sup>17</sup>  $b$  was rounded down to the nearest integer. All bins have a width of  $b$ . Modal bin = most populous bin.  $\chi^2$  goodness-of-fit test was performed as described.<sup>18</sup> Skewness ( $\sqrt{b_1}$ ) and kurtosis ( $b_2$ ) were calculated as described.<sup>19</sup> Backtransformation: the exponentials of the values of the mean and standard deviation of  $\ln$ CHC values.<sup>20</sup>

## Acknowledgements

I thank Robert A. Schulz for supplies, equipment and laboratory space. I thank Chiyedza N. Small for reviewing the manuscript. CHC values were obtained in the execution of work that was supported by a grant from the National Institutes of Health to Robert A. Schulz (HL071540).

## References

1. Glantz SA. How to summarize data. In: Malley J, Lebowitz H, Davis K, eds. *Primer of Biostatistics*, 6th Ed. USA: McGraw-Hill 2005; 37.
2. Eslin P, Prévost G. Hemocyte load and immune resistance to *Asobara tabida* are correlated in species of the *Drosophila melanogaster* subgroup. *J Insect Physiol* 1998; 44:807-16.
3. Remillieux-Leschelle N, Santamaria P, Randsholt NB. Regulation of larval hematopoiesis in *Drosophila melanogaster*: a role for the multi sex combs gene. *Genetics* 2002; 162:1259-74.
4. Kurucz E, Zettervall CJ, Sinka R, Vilmos P, Pivarcsi A, Ekengren S, et al. Hemese, a hemocyte-specific transmembrane protein, affects the cellular immune response in *Drosophila*. *Proc Natl Acad Sci USA* 2003; 100:2622-7.
5. Nappi AJ, Vass E, Malagoli D, Carton Y. The effects of parasite-derived immune-suppressive factors on the cellular innate immune and autoimmune responses of *Drosophila melanogaster*. *J Parasitol* 2004; 90:1139-49.
6. Moreau SJM, Guillot S, Populaire C, Doury G, Prévost G, Eslin P. Conversely to its sibling *Drosophila melanogaster*, *D. simulans* overcomes the immunosuppressive effects of the parasitoid *Asobara citri*. *J Insect Physiol* 2005; 29:205-9.
7. Qiu P, Pan PC, Govind S. A role for the *Drosophila* Toll/Cactus pathway in larval hematopoiesis. *Development* 1998; 125:1909-20.
8. Sorrentino RP, Carton Y, Govind S. Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. *Dev Biol* 2002; 243:65-80.
9. Brückner K, Kockel L, Duchek P, Luque CM, Rorth P, Perrimon N. The PDGF/VEGF receptor controls blood cell survival in *Drosophila*. *Dev Cell* 2004; 7:73-84.
10. Zettervall CJ, Anderl I, Williams MJ, Palmer R, Kurucz E, Ando I, et al. A directed screen for genes involved in *Drosophila* blood cell activation. *Proc Natl Acad Sci USA* 2004; 101:14192-7.
11. Sorrentino RP, Melk JP, Govind S. Genetic analysis of contributions of dorsal group and JAK-Stat92E pathway genes to larval hemocyte concentration and egg encapsulation response in *Drosophila*. *Genetics* 2004; 166:1343-56.
12. Sorrentino RP, Tokusumi T, Schulz RA. The Friend of GATA protein U-shaped functions as a hematopoietic tumor suppressor in *Drosophila*. *Dev Biol* 2007; 311:311-23.
13. Chiu H, Ring BC, Sorrentino RP, Kalamarz M, Garza D, Govind S. dUbc9 negatively regulates the Toll-NFκB pathways in larval hematopoiesis and drosomycin activation in *Drosophila*. *Dev Biol* 2005; 288:60-72.

- 
14. Nevill AM, Holder RL. Body mass index: a measure of fatness or leanness? *Br J Nutrition* 1995; 73:507-16.
  15. Greer KA, Canterbury SC, Murphy KE. Statistical analysis regarding the effects of height and weight on life span of the domestic dog. *Res Vet Sci* 2007; 82:208-14.
  16. Sorrentino RP, Schulz RA. A protocol for determining circulating hemocyte concentration of individual tumorous *Drosophila* larvae. *Dros Inf Serv* 2006; 89:103-7.
  17. Scott DW. On optimal and data-based histograms. *Biometrika* 1979; 66:605-10.
  18. Crow EL, Davis FA, Maxfield MW. In: *Statistics Manual*. Mineola, NY: Dover Publications 1960.
  19. Sachs L. Further test procedures. In: *Applied Statistics*. 2<sup>nd</sup> Ed. New York, NY: Springer-Verlag 1984; 325-6.
  20. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. *BioScience* 2001; 51:341-52.
  21. Sorrentino RP, Gajewski KM, Schulz RA. GATA factors in *Drosophila* heart and blood cell development. *Semin Cell Dev Biol* 2005; 16:107-116.
  22. Tokusumi T, Sorrentino RP, Russell M, Ferrarese R, Govind, Schulz RA. Characterization of a lamellocyte transcriptional enhancer located within the *misshapen* gene of *Drosophila melanogaster*. *PLoS ONE* 2009; 4:6429-38.