

RESEARCH ARTICLE

Head-to-head comparison of three experimental methods of quantifying competitive fitness in *C. elegans*

Timothy A. Crombie^{1,2}, Sayran Saber¹, Ayush Shekhar Saxena¹, Robyn Egan¹, Charles F. Baer^{1,3*}

1 Department of Biology, University of Florida, Gainesville, FL, United States of America, **2** Department of Molecular Biosciences, Northwestern University, Evanston, IL, United States of America, **3** University of Florida Genetics Institute, Gainesville, FL, United States of America

These authors contributed equally to this work.

* cbaer@ufl.edu



OPEN ACCESS

Citation: Crombie TA, Saber S, Saxena AS, Egan R, Baer CF (2018) Head-to-head comparison of three experimental methods of quantifying competitive fitness in *C. elegans*. PLoS ONE 13(10): e0201507. <https://doi.org/10.1371/journal.pone.0201507>

Editor: Denis Dupuy, INSERM U869, FRANCE

Received: July 12, 2018

Accepted: October 5, 2018

Published: October 19, 2018

Copyright: © 2018 Crombie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Data in [S2 Table](#) have also been deposited at Dryad with doi: [10.5061/dryad.bq3fn37](https://doi.org/10.5061/dryad.bq3fn37).

Funding: Support was provided by National Institute of Health (<https://www.nih.gov/>) grants R01GM107227 to CFB and E. C. Andersen, and S10100D012006 to CFB. SS was supported by a graduate fellowship from the Higher Committee for Education Development in Iraq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Organismal fitness is relevant in many contexts in biology. The most meaningful experimental measure of fitness is *competitive* fitness, when two or more entities (e.g., genotypes) are allowed to compete directly. In theory, competitive fitness is simple to measure: an experimental population is initiated with the different types in known proportions and allowed to evolve under experimental conditions to a predefined endpoint. In practice, there are several obstacles to obtaining robust estimates of competitive fitness in multicellular organisms, the most pervasive of which is simply the time it takes to count many individuals of different types from many replicate populations. Methods by which counting can be automated in high throughput are desirable, but for automated methods to be useful, the bias and technical variance associated with the method must be (a) known, and (b) sufficiently small relative to other sources of bias and variance to make the effort worthwhile. The nematode *Caenorhabditis elegans* is an important model organism, and the fitness effects of genotype and environmental conditions are often of interest. We report a comparison of three experimental methods of quantifying competitive fitness, in which wild-type strains are competed against GFP-marked competitors under standard laboratory conditions. Population samples were split into three replicates and counted (1) "by eye" from a saved image, (2) from the same image using CellProfiler image analysis software, and (3) with a large particle flow cytometer (a "worm sorter"). From 720 replicate samples, neither the frequency of wild-type worms nor the among-sample variance differed significantly between the three methods. CellProfiler and the worm sorter provide at least a tenfold increase in sample handling speed with little (if any) bias or increase in variance.

Introduction

In the context of evolutionary biology, *fitness* is the contribution of an individual to the next generation. Researchers working with *C. elegans* and related nematodes are often interested in

Competing interests: The authors have declared that no competing interests exist.

comparing the average fitness of different strains. The most straightforward way to quantify fitness is to count the total number of offspring produced by an individual over the course of its lifetime. The number of offspring produced over the lifetime of an individual i is its *absolute fitness* (usually depicted W_i). *Relative fitness*, w_i , is the absolute fitness of an individual scaled relative to that of a reference, usually either the most fit individual in the population, i.e., $w_i = \frac{W_i}{W_{MAX}}$, or the population mean, $w_i = \frac{W_i}{\bar{W}}$.

From the perspective of evolution, only relative fitness matters. All else equal, greater absolute fitness means greater relative fitness. However, all else is often not equal, for several reasons. First, demography matters: offspring produced early in an individual's life contribute more to fitness than offspring produced late in life [1]. More importantly, however, interactions between individuals can influence fitness in ways that will not be apparent if the different strains are not allowed to interact directly. Also, differences in relative fitness may often only be manifested under competitive conditions [2], because small differences in performance which would have no detectable effect on fecundity (e.g., sprint speed in gazelles) may translate into qualitative differences in fitness (e.g., which gazelle gets caught and eaten by the lion).

A standard laboratory assay of competitive fitness in many organisms, including *Caenorhabditis*, is to allow different strains of interest ("focal" strains, usually strain = genotype) to compete against a standard, marked competitor strain [3,4]. Experimental populations are initiated with a known number of focal and competitor individuals, and the population allowed to grow until a pre-defined endpoint, at which time the individuals are either enumerated in totality or sampled. The proportion of individuals of the focal strain at the assay endpoint is designated p ; the proportion of competitors is $1-p$. The ratio $p/(1-p)$ —the odds that an individual in the population is the focal type—is often called the "competitive index", CI [5], and provides an estimate of the fitness of the focal strain relative to the marked competitor. Relative competitive fitness of focal strains i and j against the same competitor strain can be assessed by comparison of the odds ratio p_i/p_j . There are several indexes of competitive fitness, all of which are fundamentally based on the proportion p [6].

Experimental measurement of competitive fitness in many organisms, including *Caenorhabditis*, has been greatly facilitated by the availability of heritable fluorescent markers, e.g., GFP, which can be scored in much higher throughput than traditional phenotypic markers such as *dumpy* or *unc* mutants. The simplest competitive fitness assay is to pick a known number of worms of the focal and fluorescently-marked competitor strains onto a seeded plate, incubate the plate until the food is consumed, then count the worms under transmitted light and fluorescent light. Focal and competitor worms alike will be visible under transmitted light, whereas only the competitor worms will be visible under the fluorescent light (Fig 1). Images of the plate can be captured under transmitted and fluorescent light and worms counted, either by eye or by means of image analysis software.

Two methods exist by which the throughput of competitive fitness assays can be significantly increased. First, worms can be washed into wells of a microtiter plate and a motorized stage used to automate the image capture, followed by automated counting by image analysis. Second, a large-particle flow cytometer (aka, a "worm sorter") can be employed. The latter two methods involve significant initial investment, especially the worm sorter. However, given that the relevant hardware is available, it is useful to know the time/accuracy trade-offs involved with the different methods.

Here we provide a head-to-head comparison of three experimental methods of quantifying competitive fitness in *C. elegans*. Method 1 is our standard "by eye" competitive fitness assay, in which nothing is automated [7]. Method 2 employs the image-analysis software CellProfiler [8,9] to automate worm counting. Method 3 involves using a Union Biometrica BioSorter

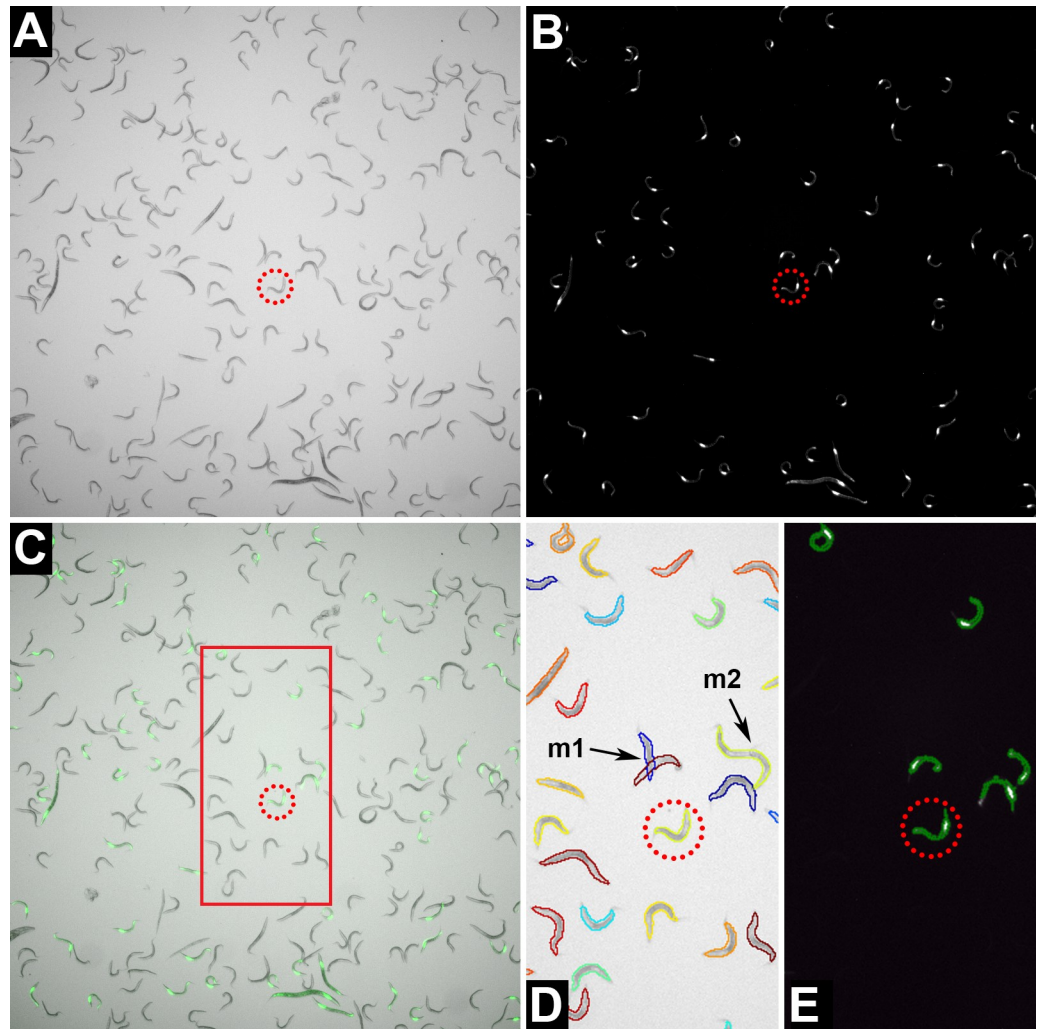


Fig 1. Images. Example of a comparison of the same sample of worms taken under bright-field (A) and under 470 nm fluorescent light (B). The bright-field and fluorescence images are merged showing that the images are exact overlays (C). Worms that appear under fluorescent light are the GFP-marked competitor; worms that do not fluoresce are the focal type. The frequency of the focal type, p , is the difference between the number of worms visible in (A) and (B) divided by the number of worms in (A). The red circles highlight the same individual in all images. Panels (D) and (E) show CellProfiler generated worm outlines and GFP objects respectively for the area bound by the red rectangle in (C). Occasional CellProfiler worm untangling errors are shown in (D); “m1” shows misaligned worm outlines for the overlapping worms, “m2” shows two worms mistaken as one.

<https://doi.org/10.1371/journal.pone.0201507.g001>

large particle flow cytometer to count worms. For each method worms are washed from agar competition plates into the wells of a 96-well plate to facilitate counts. Our primary interests are: (1) quantifying the variability of the data, (2) identifying potential bias, and (3) quantifying the time per datum collected.

Results

Variability

Measures of variance are often correlated with the mean, which can confound inferences regarding differences in variance when both quantities differ between groups [10,11]. To begin, we assessed three estimators of competitive fitness: the frequency of the focal type, p , the

competitive index $CI = p/(1-p)$, and $\log(CI)$. For each estimator of competitive fitness we calculated two measures of variability: the within-block standard deviation (SD) of a given focal strain/competitor strain/method combination and the mean within-block Median-Levene statistic [12], $\overline{Md} = \frac{1}{n} \sum_1^n |x_{ijkl} - \tilde{x}_{jkl}|$, where \tilde{x}_{jkl} is the block median of the estimator of competitive fitness of focal strain j against competitor strain k using method l , and n is the number of observations in block of a given focal strain/competitor strain/method combination.

Plots of the mean-variance relationships are shown in Fig 2. The correlations are weakest for $\log(CI)$, slightly greater for p and nearly perfect for CI . For each of the three measures of competitive fitness, the standard deviation (SD) is less correlated with the mean than is the Median Levene statistic (Md); the correlation is smallest for the SD of $\log(CI)$. Given these findings, our assessment of the three assay methods is based on $SD_{\log(CI)}$.

Box-plots of p and $SD_{\log(CI)}$ for the three methods, averaged over focal strains and competitor strains, are shown in Fig 3. The data are summarized in S1 Table and raw data are given in S2 Table. Averaged over focal strains and competitors, p does not differ significantly between the three methods ($F_{2,69,1} = 1.71, P > 0.18$), although it is slightly larger when estimated by CellProfiler. $SD_{\log(CI)}$ does not differ significantly between the three methods ($F_{2,49,5} = 2.23, P > 0.12$), but CellProfiler is slightly less variable than the other two methods. However, because

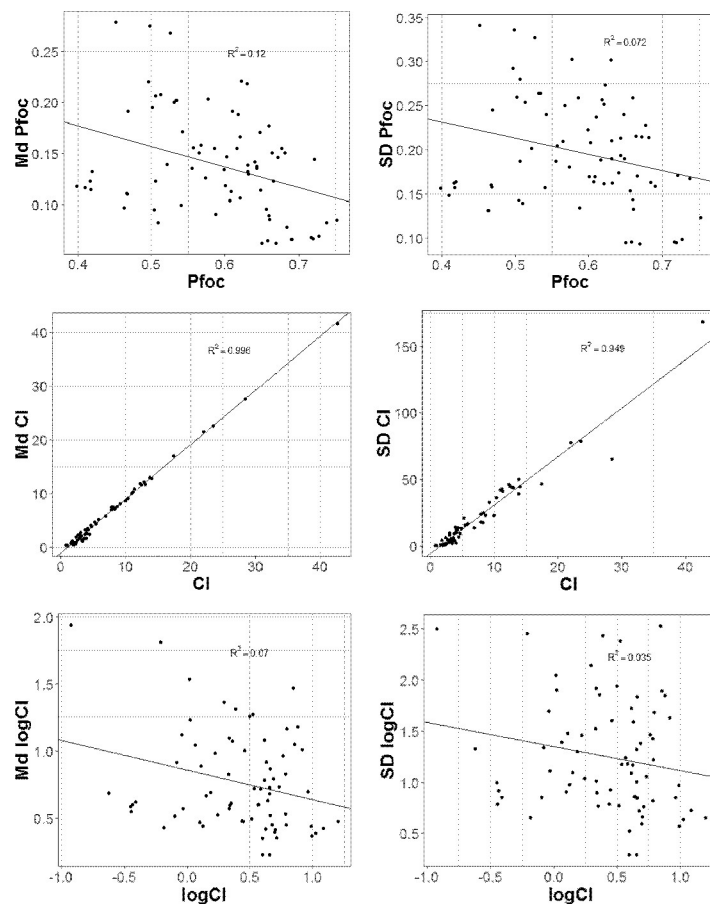


Fig 2. Variability of measures of competitive fitness. Plots of measures of variability (y-axis) vs. measures of competitive fitness (x-axis). Left panels show the Median-Levene statistic (Md) as the measure of variability, right panels show the standard deviation (SD) as the measure of variability. Top panels show the frequency of the focal type ("pfoc") as the measure of competitive fitness, middle panels show $CI = p/(1-p)$, bottom panels show $\log(CI)$.

<https://doi.org/10.1371/journal.pone.0201507.g002>

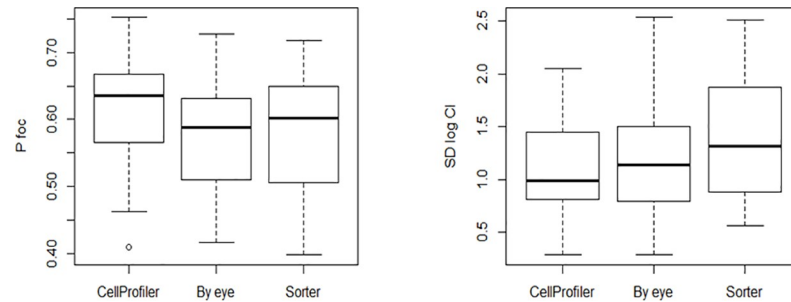


Fig 3. Variability of experimental methods. Left panel: boxplots of the frequency of wild-type individuals ("pfoc") for the three methods. Right panel: standard deviation of log(CI) for the three methods. Each data point is the average of a block/competitor strain/focal strain combination.

<https://doi.org/10.1371/journal.pone.0201507.g003>

of the weak negative correlation between mean and variance of $\log(CI)$, it is uncertain if CellProfiler is inherently slightly less variable than the other methods, or if the slightly lower variability would disappear in cases when the means were equal.

The repeatability of the "by eye" and CellProfiler methods can be assessed by counting the same image(s) twice (S3 Table). To quantify the repeatability of the by eye method, the same counter (SS) re-counted a subset of 59 images approximately one month after the original counts. The correlation between the two counts was $>99.9\%$ for both the total count and the GFP count. The mean absolute difference between the two counts, expressed as a fraction of the average of the two counts, was 0.73% for the total count and 0.46% for the GFP count. The correlation between the proportion of focal worms, p , between the two counts is 99.8%. We re-counted all 720 images counted by CellProfiler using identical software parameters on a different computer running the same version of CellProfiler (v2.2.0); the counts were exactly the same in every case.

Unlike images, worm sorter samples are ephemeral, in that a specific data point cannot be permanently associated with a given individual. To assess repeatability of the worm sorter counts, we split 336 samples from the same assay well into two wells on the counting plate, one of which contained approximately 1/4 of the sample by volume (\bar{n} worms = 143) and the other of which contained the remaining $\sim 3/4$ (\bar{n} worms = 551) (S3 Table). Of the 336 samples, 280 met the (arbitrary) criteria of n worms > 10 and $0.01 < p < 0.99$. The mean absolute difference in the proportion of focal worms, p , is 5.4%, and the correlation between p in the 1/4 and 3/4 split samples is 96.5%. The variation among the two sorter samples includes sampling variance of p (= technical variance), whereas the variation between the two samples by eye and by CellProfiler only includes counting error. The sample standard error of the binomial is $\sqrt{\frac{pq}{n}}$, so the difference in standard errors between the two samples equals $\sqrt{\frac{p^2}{n_1}} \approx \sqrt{\frac{551}{143}} \approx 2$.

The relative contribution of binomial sampling and biological variation among competition plates to the total variance is an empirical question. For the 1/4 volume sample, $\bar{p} = 0.553$ with standard error 0.0159; for the 3/4 volume sample, $\bar{p} = 0.524$ with standard error 0.0154. For the 1/4 volume sample, mean $\log(CI) = 0.153$ with standard error 0.0947; for the 3/4 volume sample, mean $\log(CI) = -0.0058$ with standard error 0.0946. Clearly, for the sample sizes employed in this study (> 100 worms per plate, on average), binomial within-sample variance is small relative to the biological variance among plates of worms.

Bias

If any of the methods are biased, the bias must be sufficiently small to have avoided detection, given the sample sizes. However, we suspect the slightly greater value of p (and thus CI and \log

(*CI*) estimated by CellProfiler relative to the other two methods (S1 Table) may in fact reflect a slight bias toward inflating the frequency of non-fluorescent worms, due both to false positives (non-worms called as worms under brightfield) and false negatives (failure to identify fluorescent worms as fluorescent). Successive reduction in edge-cropping of the images analyzed by CellProfiler (see Methods) leads to increasing values of p , relative to the value of p estimated by eye from the same (cropped) image (S1 Fig); presumably the reduction in p with increased cropping is due to the reduction in false positives.

Detecting a fluorescent worm boils down to a two-step process in CellProfiler: first, the worm body must be identified in the brightfield image, and then the fluorescent signal must be identified within that worm body in the associated fluorescent image. Occasionally, CellProfiler fails to identify the full worm (most often truncating the head or tail region), and if the fluorescent signal happens to fall in the missed region, the worm will be mischaracterized as non-fluorescent. We cannot quantify the magnitude of that effect beyond reiterating that it must be small relative to the other sources of variation in the data.

Time per datum

Each of the three methods has unique investments of time associated with it; we return to those in the Discussion. For the head-to-head comparison, each sample was first divided into aliquots (see Methods), thus the clock starts at the time the samples were divided. The "by eye" and CellProfiler methods both involve the analysis of images captured from samples in a 96-well plate. It took approximately 20 minutes to capture the 96(x2) images from one 96-well plate with the stage movement rate set to "slow" (faster image capture is possible), or about 12.5 seconds/sample.

An experienced counter (SS) took ~5–7 minutes to count both images for one sample by eye (bright field and GFP; $\bar{n}_{BF} \approx 176$, $\bar{n}_{GFP} \approx 77$). The rate at which images can be processed by CellProfiler will depend on the computer platform as well as the images. Working with a far-from-state-of-the-art Dell Optiplex 9020 PC running Windows 7, CellProfiler averaged 35.6 ± 16 minutes to process a 96-well plate, or about 22 seconds per sample.

As we applied it, the worm sorter took about 50 minutes to collect data from a 96-well plate, or about 30 seconds per sample. Once data are acquired by the sorter, counts can be obtained computationally essentially instantaneously.

Discussion

The results of this study are good news for worm biologists wanting reliable high-throughput estimates of competitive fitness using fluorescently-labeled strains. Automated image-analysis with CellProfiler and the BioSorter both provide estimates of competitive fitness that are nearly unbiased and no more inherently variable than the human eye. For samples of a few hundred worms, CellProfiler and the Biosorter can count worms at a rate that is approximately an order of magnitude faster than even a well-trained, well-motivated human, and the time-savings will increase as sample size increases.

A caveat is that both CellProfiler and the Biosorter require a non-trivial initial investment of time to optimize the data-collection protocol, whereas a sighted human can begin collecting data right away. In our experience the CellProfiler image analysis pipelines took nearly 40-man hours to develop and implement. However, less time will be required to optimize our existing pipelines (S1 File) for a different imaging system. Importantly, the time required to develop the CellProfiler pipelines could be reduced by standardizing imaging conditions so that block-specific pipelines and worm models are not required. For example, in our experiments variable magnification and lamp illumination settings across blocks required us to

develop unique pipelines for individual blocks, each of which required a few hours to prepare. We also chose to develop block-specific worm models because of variation in worm morphology across blocks. We believe that much of this variability could be avoided by standardizing levamisole treatments prior to imaging (see [Methods](#)). Levamisole is a potent cholinergic agonist that causes hypercontracted paralysis followed by relaxation and death, and the size and shape of worms change significantly during early exposure. We recommend incubating worms for at least 30 min in 5 mM levamisole prior to imaging, so that worms are consistently imaged in the relaxation phase following hypercontraction.

Methods and materials

Strains

Three *C. elegans* strains (N2, PB306, and CB4856) were selected as focal strains for the competitive fitness assays. The competitor strains ST2 and VP604 were chosen based on bright, constitutive expression of GFP reporter alleles in all developmental stages. ST2 *ncls2* [*pH20::GFP* + pBlueScript] expresses GFP in nearly all neurons and VP604 *kbIs24* [*Pgpdh-1::dsRed2*; *Pmyo-2::GFP*; *unc-119*] X expresses GFP robustly in the pharynx.

Competitive fitness assay

Five replicate blocks of the competitive fitness assay were conducted in the spring of 2016. Blocks consisted of 24 replicate competitions for all 18 combinations of focal and competitor strains and assay methods, except in one block for which "by eye" data were not collected. Blocks were initiated by bleach-synchronization of strains that had been maintained as mix-staged populations on standard 100 mm NGM plates. The competition between strains began by transferring a single L4-stage focal hermaphrodite and a single L4-stage, GFP-marked competitor hermaphrodite to the same well of a 24-well plate. Each 24-well plate contained NGMA agar supplemented with nystatin (20 µg/ml) and streptomycin (50 µg/ml) and seeded with 10 µl of OP50-1 *E. coli*. Replicates for each of the six competitions were distributed equally among six 24-well plates, i.e. four replicates per 24-well plate. Competition between the focal and competitor strains persisted for 168 h at 20°C at which point the food source in nearly all wells had been completely consumed. The resulting nematode populations were then washed from the competition plates by adding 1.5 ml of M9 buffer to each well with a repeat pipettor, aspirating up and down three times with a disposable transfer pipet, then transferring the nematode suspension to a 2 ml capacity 96-well plate. Once filled, the 96-well plate was centrifuged for 1 min at 1,000 g, and the supernatant was aspirated with an 8-channel strip aspirator. In order to retain the nematode pellet, the strip aspirator was set to leave 100 µl in each well. The wells in the 96-well plate were then washed once more by adding 1.4 ml of M9 buffer, centrifuging, and aspirating, followed by the addition of 1.4 ml of M9. A 12-channel pipet was then used to gently mix the nematode suspension by pipetting up and down five times. Once mixed, a 110 µl sample of the resulting worm suspension was transferred into the appropriate rows of two 96-well, clear bottom plates. One of the 96-well plates was used to quantify competitive fitness using the image-based methods (manual counting and CellProfiler counting) while the other was used for sorter-based counting.

Imaging

Quantification of competitive fitness by image-based methods requires both a bright-field and paired GFP fluorescence image to identify focal animals (non-fluorescent) from competitors (fluorescent). To ensure that nematodes do not move between the bright-field and GFP

fluorescence images, levamisole was added to the wells at a final concentration of 5 mM. Bright-field and GFP fluorescence images (470 ± 20 nm excitation, 525 ± 25 nm emission) were captured at 20x or 30x magnification for each well using an automated epifluorescence microscope (IX-70, Olympus, Pittsburgh, PA, USA) fitted with a CCD camera (Retiga-2000R, QImaging, Surrey, BC, Canada), XYZ stage and focus motors (Prior Scientific, Cambridge, UK), and controlled by Image Pro-plus software (Media Cybernetics, Rockville, MD, USA). The bright-field and GFP fluorescence image pairs for each well were saved as a single stacked image before being used for the image-based competitive fitness quantification methods.

Quantification method

"By eye". Manual counts of focal and competitor worms were obtained as follows. Stacked images consisting of a bright-field slice and GFP fluorescence slice were imported into ImageJ software (<http://rsb.info.nih.gov/ij/>). A rectangle encompassing approximately 200 worms was superimposed on the bright field image slice and the worms within the rectangle counted (Fig 1A). We used the multipoint tool in ImageJ to facilitate counting worms. The multipoint tool allows the user to mark each worm on the image as it is counted and therefore reduces the chances of miscounting worms. The same rectangle was then superimposed on the GFP image slice and the fluorescent worms visible in that image counted using a different counter type in the multipoint tool (Fig 1B). The number of focal and competitor worms within a given well were then exported from the ImageJ results window and the frequency of the focal type, p , the competitive index $CI = p/(1-p)$, and $\log(CI)$ were calculated in excel.

CellProfiler. We developed an image analysis pipeline using CellProfiler software to automatically quantify competitive fitness in a given well using the paired bright-field and GFP fluorescence images (S1 File). CellProfiler is a free, open-source software developed to automate quantitative measurement of phenotypes from large collections of images [9]. Implementation of the Worm Toolbox modules within CellProfiler allows for the automated detection of individual nematodes from bright-field images despite instances of nematodes clustering or overlapping [8]. Further processing of the paired GFP fluorescence images allows our pipeline to distinguish between focal (non-fluorescent) or competitor (fluorescent) phenotypes, and thus calculate all three estimators of relative fitness in a given image, i.e., frequency of the focal type, p , the competitive index $CI = p/(1-p)$, and $\log(CI)$. Quantification of relative fitness with CellProfiler begins by building a probabilistic model of nematode shape hereafter referred to as a "worm model". The first step in generating a worm model is to manually identify examples of single, non-overlapping nematodes using the UntangleWorms module from Worm Toolbox configured in the "train" mode. We used the "identifying_single_worms_for_worm_model.cpproj" project file to identify ≥ 50 single, non-overlapping nematodes from 5–10 random bright-field images within a particular block. The resulting "singleworm.png" binary output images were then used as input images for the "create_new_worm_model.cpproj" project file, which processes the single worms in the images to generate the worm model for a particular block. We generated a different worm model for each block due to variations in image magnification and nematode shape across blocks. We then used the "UntangleWorms-CountGFP_[Assay_#].cpproj" project files to quantifying the number of focal and competitor nematodes in all images for each block. These project files each use their own image processing pipeline and block-specific worm model. The image processing pipeline for blocks 10–13 include crop modules, while block 8 does not. The cropping is included to help correct for extremely uneven illumination in bright-field images, wherein well edges appear dark and centers appear bright. The effect of cropping modules on CellProfiler error is shown in S1 Fig. Image processing pipelines for all blocks include modules to detect and correct for any uneven

illumination that remains. Subsequent modules then identify putative nematode objects from the background via thresholding. Next, individual nematode outlines are identified from clusters of nematodes or debris using the block specific worm model. Using the paired GFP fluorescence image, fluorescent objects in each nematode outline are identified and counted by thresholding. The pipelines output a .csv file containing the number of nematode objects in a well, along with the number of children GFP objects counted within each nematode object. A particular nematode is assigned a focal genotype if it contains no GFP objects, or a competitor genotype if it contains one or more GFP objects (Fig 1D). Finally, the frequency of the focal type, p , the competitive index $CI = p/(1-p)$, and $\log(CI)$ are calculated for each well.

BioSorter. Our BioSorter competitive fitness analysis pipeline is presented in detail in Supplementary Appendix A1 of [7]; we reprise the basics here. After sample preparation, we aliquot 110 μ l of the sample into a well of a 96-well plate, which the LP Sampler aspirates to the sorter. The LP Sampler script is optimized for 100 μ l input, but there is a risk of incorporating bubbles in the fluid line if the sample is short. We include the extra 10 μ l for redundancy, and four wash steps with the LP Sampler to ensure we pick up most of the sample. In a typical wash step, the LP Sampler aspirates and dispenses water, cumulatively picking up most of the sample over the course of four wash steps. Potentially leftover samples are unlikely to introduce bias in a competitive assay.

The competitive fitness assay was extended until plates were starved of *E. coli* food (or nearly so), which greatly reduces the frequency of false positives in the sorter counts. *E. coli* tends to clump together and register as many small events (particles), decreasing the signal to noise ratio for smaller worms.

Data analysis. The dependent variable for purposes of quantifying variation is the standard deviation of $\log(CI)$, $SD_{\log(CI)}$. CI was measured from 24 replicate plates from each combination of focal strain (N2, PB306, CB4856), competitor strain (ST2, VP604), and quantification method (by eye, CellProfiler, Sorter) in five assay blocks. Thus, there are five estimates of $SD_{\log(CI)}$ for each unique combination of focal strain/competitor strain/method (four estimates for "by eye"). Method is the independent variable of interest; for the purposes of this study we are not interested in the variation among focal strains or competitor strains. However, because the sample sizes are small, we treat focal strain, competitor strain, and their interactions as fixed effects. Data were analyzed by general linear model (GLM) as implemented in the MIXED procedure of SAS v. 9.4. Variance components were estimated by restricted maximum likelihood (REML). The full linear model is: $y_{ijkl} = \mu + f_i + c_j + m_k + t_{i \times j} + u_{i \times k} + v_{j \times k} + w_{i \times j \times k} + \epsilon_{llijk}$, where y_{ijkl} is the estimate of $SD_{\log(CI)}$ in block l , μ is the overall mean, f_i is the effect of focal strain i , c_j is the effect of competitor strain j , m_k is the effect of method k , $t_{i \times j}$ is the effect of the interaction between focal strain i and competitor strain j , $u_{i \times k}$ is the effect of the interaction between focal strain i and method k , $v_{j \times k}$ is the interaction between competitor strain j and method k , $w_{i \times j \times k}$ is the effect of the three-way interaction, and ϵ_{llijk} is residual (among-block) variance. We initially estimated the residual variance separately for each focal/competitor/method combination, then pooled the residual variance over different combinations of groups, using the minimum corrected Akaike's Information Criterion (AICc) as the criterion for the best model. Similarly, competitor strain, focal strain, and their interactions were removed and the AICc calculated. The smallest AICc was given by the model with only method included as a fixed effect and the residual variance estimated separately for each method, pooling residual variance over focal and competitor strains within a method. Significance of fixed effects was assessed by F-test on type III sums of squares.

We repeated the above analysis for the fraction of focal worms, p , with block included as an additional random effect and replicate (nested within block) as the unit of observation. The block for which we did not collect "by eye" data was omitted from the analysis.

Supporting information

S1 Fig. Cropping module effects on CellProfiler (data from Block 12). The proportion of wild-type worms (pfoc) estimated by CellProfiler is regressed against the same proportion estimated from the by eye count from the full crop of the same image. See [methods](#) for description of the cropping protocol.

(TIF)

S1 Table. An Excel spreadsheet containing summary statistics.

(XLSX)

S2 Table. An Excel spreadsheet containing the raw data.

(XLSX)

S3 Table. An Excel spreadsheet containing repeatability data.

(XLSX)

S1 File. This zip file contains sample bright-field and GFP fluorescence image pairs as well as CellProfiler project files. CellProfiler project files 1–2 are used to train and create worm models. CellProfiler project file 3 is used to score focal and competitor worms in image pairs and output scores as .csv files.

(ZIP)

Acknowledgments

We thank Joanna Dembek and Asher Shoucair for assistance in the lab, and the reviewers for their helpful comments.

Author Contributions

Conceptualization: Charles F. Baer.

Data curation: Sayran Saber, Ayush Shekhar Saxena.

Formal analysis: Timothy A. Crombie, Ayush Shekhar Saxena, Charles F. Baer.

Funding acquisition: Charles F. Baer.

Investigation: Timothy A. Crombie, Sayran Saber, Ayush Shekhar Saxena, Robyn Egan, Charles F. Baer.

Methodology: Timothy A. Crombie, Sayran Saber, Ayush Shekhar Saxena.

Project administration: Charles F. Baer.

Resources: Timothy A. Crombie, Ayush Shekhar Saxena.

Software: Timothy A. Crombie, Ayush Shekhar Saxena.

Supervision: Charles F. Baer.

Visualization: Timothy A. Crombie.

Writing – original draft: Charles F. Baer.

Writing – review & editing: Timothy A. Crombie, Sayran Saber, Ayush Shekhar Saxena, Charles F. Baer.

References

1. Charlesworth B (1994) Evolution in age-structured populations. Cambridge England; New York: Cambridge University Press. xiii, 306 p.
2. Agrawal AF, Whitlock MC (2010) Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends in Ecology & Evolution* 25: 450–458.
3. Morran LT, Parmenter MD, Phillips PC (2009) Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* 462: 350–352. <https://doi.org/10.1038/nature08496> PMID: 19847164
4. Chelo IM, Nédli J, Gordo I, Teotónio H (2013) An experimental test on the probability of extinction of new genetic variants. *Nature Communications* 4.
5. Knight GR, Robertson A (1957) Fitness as a measurable character in *Drosophila*. *Genetics* 42: 524–530. PMID: 17247713
6. Latter BDH, Sved JA (1994) A Reevaluation of data from competitive tests shows high levels of heterosis in *Drosophila melanogaster*. *Genetics* 137: 509–511. PMID: 8070661
7. Yeh S-D, Saxena AS, Crombie TA, Feistel D, Johnson LM, et al. (2018) The mutational decay of male-male and hermaphrodite-hermaphrodite competitive fitness in the androdioecious nematode *C. elegans*. *Heredity* 120: 1–12. <https://doi.org/10.1038/s41437-017-0003-8> PMID: 29234171
8. Wählby C, Kamensky L, Liu ZH, Riklin-Raviv T, Conery AL, et al. (2012) An image analysis toolbox for high-throughput *C. elegans* assays. *Nature Methods* 9: 714–716. <https://doi.org/10.1038/nmeth.1984> PMID: 22522656
9. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, et al. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* 7: R100. <https://doi.org/10.1186/gb-2006-7-10-r100> PMID: 17076895
10. Lande R (1977) On comparing coefficients of variation. *Systematic Zoology* 26: 214–217.
11. Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics* 130: 195–204. PMID: 1732160
12. Brown MB, Forsythe AB (1974) Robust Tests for the Equality of Variances. *Journal of the American Statistical Association* 69: 364–367.